Complying With the Clean Water Act Using Aquatic Biota to Set Water Quality Standards[1]

Richard B. Shepard, PhD
Applied Ecosystem Services, Inc.

April 5, 2017

Abstract

The basis for setting water quality standards is decades out of date, given our current understanding of environmental data and availability of recently developed statistical models. The use of a single maximum concentration limit (MCL) for individual chemical elements does not reflect natural ecosystem function nor provide accurate indications of whether regulated industrial activities adversely impact the specific designated beneficial uses of surface or ground waters at specific locations. Water is a complex mixture of compounds, not individual ions, and concentrations vary with temperature, pH, binding and release with inorganic and organic substrata, and other factors. A sample of water represents a snapshot at a specific time and place. This is why aquatic ecologists have established data collection standards to minimize variability when measuring physical and chemical parameters of flowing and standing waters.

Aquatic biota are much more reliable indicators of ambient water quality than are concentrations of chemical elements. The EPA considers aquatic life to be the highest and best use of water (that is, the use most sensitive to anthropogenic disturbance). Aquatic biota exist with the abiotic physical and chemical environments to form natural ecosystems.

Natural ecosystems are highly complex; we cannot have complete knowledge of their variability and interactions among all components. About 50 years ago, when environmental laws began to be created, ecologists were moving from qualitative descriptions of ecosystems, communities, and populations to quantative measures of their dynamics. Also, appropriate statistical models did not exist, and computers were not as widely (or easily) used as they are today. To implement these statutes regulators had to assess and compare natural ecosystems in attempts to determine anthropogenic effects. The approach used then was to create methods producing a single numerical value assumed to summarize ecosystem quality and separate "good" from "bad" conditions. These species diversity and biotic integrity indices still are used today. And they still fail to describe ecosystem complexity, t quantify inherent natural variability, and to separate natural and anthropogenic changes to these systems. These faillings are overcome by applying appropriate, modern statistical models to biotic data.

An important benefit of robust statistical analyses of ecosystems is that they integrate components of each drainage basin and its stream network. This integration provides insights that regulators and other stakeholders can use to make informed decisions. These statistical analyses do not produce a dichotomous decision point (less than this number is good, greater than this number is bad), but allow the use of Best Professional Judgment and adjusted as more data and knowledge become available.

This monograph describes application of these ideas and protocols to several stream systems that drain the operational areas of the Jerritt Canyon Mine in the southern Independence Mountains, Elko County, Nevada. Each basin is described and characterized individually because they all differ. Inter-basin analyses could be accomplished with more data and the results would explain why the basins differ. The data available for each stream range from 4 to 8 years. Some statistical models could not be used because too few data were available. This conservative behavior discourages decision-making on weak or insufficient data. Over all the stream networks there was moderate to high variability in functional feeding group component ratios and explanatory variables. Any anthropogenic influences were within the variability range and did not modify biotic compositions in any distinctive way.

The ideas, models, and analyses described in this document can be usefully applied in any drainage with suitable data; their power for regulators and the regulated public justifies the efforts and costs for obtaining such data for baseline and permit compliance monitoring.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The intent and purpose of the Clean Water Act (CWA) is to maintain designated beneficial uses of surface and ground waters. Attainment of a designated beneficial use can be statistically analyzed (Sanchez, 2013) but the process is complicated and not commonly applied. The effort and cost might be appropriate for some water bodies and situations within a state, but assessment of water quality to prevent degradation of designated beneficial uses can be more readily calculated using other approaches.

This document examines in detail the approaches traditionally used to regulate water quality through aquatic chemistry and biota, then presents a robust statistical process that is consistent, technically sound, and legally defensible. Regulators can apply this same process everywhere while producing results that are specific to projects, locations, and designated beneficial uses. The data used in these analyses were collected over several decades from seven streams in the Independence Mountains in northern Elko County. Five of these streams drain basins into the South Fork Owyhee River basin on the west side of the range and two streams drain basins into the North Fork Humboldt River basin on the east side. All seven basins drain portions of the Jerritt Canyon Mine property; some basins had previous surface mine and exploration disturbances and others host current underground mines.

Federal and state statutes define designated beneficial uses. Setting water quality standards for some of these are comparatively easy using qualitative or water chemical constituent concentrations:

- Watering of livestock. The water must be suitable for livestock to safely drink.
- Irrigation. The water must not stunt plant growth, lower food values of the edible parts, or reduce yields.
- Recreation involving contact with the water. There must be no evidence of man-made pollution, floating debris, sludge accumulation, or similar detriments to human health.
- Recreation not involving contact with the water. The water must be free from:
  - Visible floating, suspended or settled solids arising from human activities;
  - Sludge banks;
  - Slime infestation;
  - Heavy growth of attached plants, blooms or high concentrations of plankton, discoloration or excessive acidity or alkalinity that leads to corrosion of boats and docks;
  - Surfactants that foam when the water is agitated or aerated; and
  - Excessive water temperatures.
- Municipal or domestic supply. The water must be capable of being treated by conventional methods of water treatment in order to comply with drinking water standards.
- Industrial supply. The water must be treatable to provide a quality of water suitable for the intended use.

However, there are three designated beneficial uses for which setting water quality standards are much more difficult because they involve natural ecosystems:

- Aquatic life. The water must be suitable as a habitat for fish and other aquatic life existing in a body of water. This does not preclude the re-establishment of other fish or aquatic life.

- Propagation of wildlife. The water must be suitable for the propagation of wildlife and waterfowl without treatment.
- Waters of extraordinary ecological or æsthetic value. The unique ecological or aesthetic value of the water must be maintained[1].

It is reasonable to assume that if a water body can be shown to meet the aquatic life designated beneficial use it will also support the propagation of wildlife and protect extraordinary ecological or æsthetic values. The intent of this document is to provide regulators with the background and justification for the application of specific statistical models and analyses that allow the agency to demonstrate when the aquatic life designated beneficial use has been, or continues to be, attained.

A technically sound and legally defensible process of setting water quality standards appropriate for aquatic life must be able to detect change, separate inherent natural variability from change caused by anthropogenic activities, and be based on robust mathematical and statistical foundations. Such a process is presented in this document.

1.1 Analyzing Environmental Data

There are three statistical model frameworks applied to environmental data to answer theoretical or applied questions. They are not equivalent and each is most appropriate to certain questions using defined types of data.

Classical statistics are part of the broad category called *frequentist.* These are the statistics usually presented in introductory courses. All the models in this category are designed to calculate a probability (the p value) of obtaining the outcome of a testable hypothesis; for example, that two independent samples were taken from the same population (the null hypothesis). One major drawback of this approach comes from its reliance on the probability of a series of outcomes that did not happen (at the tails of the probability distribution), and which depend on the way the experiment (or sampling) was designed. The frequentist approach fits the collected data to a defined statistical model. A second major drawback of this framework is the probability threshold of 5% (written as "$p < 0.05$"). There is nothing magical about the 5% Type I error rate; it continues to be the threshold only because it has been used for approximately 80 years. Dr. R.A. Fisher, the Scottish agronomist who invented modern statistics and hypothesis testing in the early 1930s, fought during the latter years of his life against setting a single threshold for determining when the null hypothesis should be rejected. He kept arguing that there is no mathematical justification for doing so, but it was such a useful simplification of a complex problem that his warnings were ignored and this probability level became set in concrete as the decision criterion.

A modification of the frequentist paradigm called *maximum likelihood estimation* (MLE) is used by most modern statistical models. For a particular statistical model, MLE finds the set of parameters that makes *the observed data most likely to have occurred.* Based on both deterministic and stochastic aspects of the data, the MLE function computes the likelihood (the probability of the observed outcome) given a particular choice of parameters. The function then finds the parameters that maximizes the likelihood and uses this estimate as the best fit to the parameters. The MLE approach fits the statistical model to each specific set of data. The MLE paradigm is often used to analyze water chemistry data when there are values below laboratory method detection limits.

Frequentist statistics assume there is a "true" state of the world which gives rise to a distribution of possible outcomes, one of which is what was observed with the data. The *Bayesian* framework resolves many of the conceptual problems of frequentist statistics: Bayesian answers depend on the actual observations and not on a range of hypothetical outcomes. Because of this, legitimate statements can be made about the probability of different hypotheses or parameter values. The major objection to Bayesian statistics is the need to specify *a priori beliefs* about the probabilities of different hypotheses, and these beliefs actually affect the answer. This is not as difficult as it might seem and the Bayesian approach is very useful with vertebrate population data such as fish and wildlife counts as well as with other types of environmental data.

1.2 Aquatic Chemistry

States with delegated responsibility for CWA compliance regulate industrial point and nonpoint discharges by issuing National Pollution Discharge Elimination System (NPDES) permits for point source discharges and Total

[1] While ecological value might be quantifiable, æsthetic value is a subjective linguistic variable that cannot be directly measured.

Maximum Daily Loads (TMDL) permits for nonpoint source discharges. Permit condition compliance is most commonly assessed by comparing concentrations of chemical constituents in samples of discharged waters to threshold maximum concentration levels (MCL). Concentrations above these MCLs violate permit conditions and may result in fines, penalties, or demands for remedial actions. Concentration threshold values may be appropriate for chemical constituents known to have acute or chronic toxicities to humans, livestock, and wildlife (e.g., metals and organic pesticides) but are likely not appropriate when the chemical constituents are non-toxic (e.g., total dissolved solids, sulfate, sodium) or when the designated beneficial uses are cropland irrigation, livestock, and the broad category of *aquatic life* or *fishable and swimable*. It is not common to find data that links the chemical constituent concentrations at the point of discharge or permit boundary with concentrations of those constituents at the point of withdrawal for a designated beneficial use, nor is it a standard practice to determine whether high concentrations occur naturally or are due to anthropogenic activity.

Water chemistry concentrations are snapshots in space and time and do not represent ambient conditions in the receiving water body. Furthermore, the reported concentration values lack environmental context; that is, it is not possible to determine if the observed concentrations are due to natural conditions (e.g., wildland fire in the drainage basin) or anthropogenic activities such as those associated with mining. Most permits require the holder to collect water samples at differing frequencies (weekly, monthly, quarterly) and without associated explanatory variables such as weather (rainfall, snow depth, temperature), hydrology (stream flow), or geomorphology (basin parameters, riparian cover). Explanatory variables are used to determine the extent of natural variability and assist in interpreting laboratory results of the response variables of interest. For statistical analyses, aquatic chemistry data are samples representative of the "population" of water from which they were taken. There is an infinite number of possible concentrations between zero and saturation so the population has a continuous distribution of potential values. Continuous distributions have defined statistical parameters such as the central value, variance, and skewness. With continuous data that fit the normal (Gaussian, or bell-curve) distribution we can use Student's *t*-test to determine if two samples are from the same population and analysis of variance (ANOVA) to compare multiple sets of response variables. However, aquatic chemistry concentrations frequently have a small percentage of very high concentrations and may have up to 80% of observations below the analytical laboratory's method detection limits. Because a normal distribution can extend below zero, and chemical concentrations cannot be less than zero, environmental chemical concentrations are very rarely normally distributed. Therefore, nonparametric and MLE statistical models should be applied to all environmental chemical data; these models are equally robust as are the more familiar parametric models. Testing hypotheses that two or more samples of an aquatic chemistry constituent come from the same population does not explain why the concentration has the measured value. Without evaluating the influence of explanatory variables it is not possible to determine if the response values and their variances are natural or the result of anthropogenic activities.

While laboratory analyses of water chemical constituent concentrations have been the basis for permit conditions and compliance monitoring this approach is scientifically much weaker than are approaches that use data on assemblages of aquatic benthic (i.e., bottom-dwelling or benthos) macroinvertebrates and fish. Using biota as the response variable and water chemistry, physical conditions, and drainage basin/stream network parameters as the explanatory variables allows measures of inherent natural variability and the ability to separate them from anthropogenic influences.

1.3 Aquatic Biota

Aquatic biota data are discrete counts, not continuous values, so the assumptions of parametric statistical models usually are not met. There are many nonparametric statistical models appropriate for biological data. All are equivalent to the more familiar parametric models and many others allow analyses not possible with parametric models.

With two sets of fish data we can use the Mann-Whitney *U*-test (also called the Wilcoxon signed-rank test) to determine if the two collections are from the same population. With multiple collections of fish data we can use the Kruskal-Wallis one-way analysis of variance to test the equality of median values from separate collections. Within the frequentist approach based on hypothesis testing, there is a nonparametric equivalent of every parametric model and the former are equally robust and more appropriate for statistical analyses of biological count data[2]. MLE and Bayesian approach models will be discussed later in this monograph. Applying any of these statistical hypothesis tests to benthic macroinvertebrates requires much care and is less useful for making decisions than are

[2]It is often possible, and highly useful, to transform non-linear data to allow use of linear models..

other approaches. All collections of benthic macroinvertebrates result in counts of mixed taxonomic levels. That is, some organisms are identified to species (*Pteronarcys californica*), more to genus (*Baetis* sp.), and most to the family (Chironomidae) level. Some can be identified only to the taxonomic level of order (e.g., Diptera, Ephemeroptera, Plecoptera, Trichoptera). Any statistical or ecological index (diversity, biotic integrity) *must* have all organisms at the same taxonomic level to produce usable results that can be compared over time and space.

Almost always, fish can easily be identified to species and individuals have an expected life span measured in years, so data collected at different times or different locations can validly be compared based on taxonomic identity. Benthic macroinvertebrates are very different: they have short life spans, can be difficult to identify below the family level as young juveniles, and tend to be present at a location over a limited time during each year. Benthic macroinvertebrate collections at different times or different locations cannot be assumed to be taxonomically equivalent for analytical purposes.

Aquatic biota represent overall ambient water quality, and this is particularly true for the benthic macroinvertebrates. Local populations are more limited in spatial and temporal distribution than are fish that inhabit the same stream or river reach. The use of appropriate statistical analyses of benthic macroinvertebrate and fish communities allow regulators to determine whether aquatic life in a stream, river reach, or lake has a desired condition. Most of these analytical models come from research in statistics and aquatic or numerical ecology. The differences between academic ecological research data and regulatory environmental compliance data are not always considered when determining the most appropriate analytical model and the interpretation of analytical results. This is a critical aspect in adopting a consistent process that produces site- and project-specific results suitable for state regulators to demonstrate satisfactory oversight of CWA compliance by the regulated public.

1.4 Ecosystem functions

Each organizational level in biology and ecology from cells to ecosystems exhibits *emergent properties.* These are functions that do not exist at one organizational level but emerge at a higher level. In humans, for example, the lowest organizational level is the cell. When the same type of cells are organized in tissues (liver, muscle, skin) different properties emerge. When tissues are properly organized as an individual still more different properties emerge that can be found only in the entire organizational structure and not in its component parts. At the human "ecosystem" level of organization (companies, government agencies) the emergent properties are the production of goods and services and the processing of information. The individuals that provide these functions may be of various sizes and shapes; the ecosystem functions are provided regardless of the individual structural components as long as all required capabilities are present.

The two ecosystem functions expressed by collections of benthic macroinvertebrates and fish (and useful for CWA compliance) are energy flow and nutrient cycling. Ecosystems are defined as assemblages (communities) of organisms, their biotic interactions, and the abiotic environment in which they live. Ecosystems are sustained by the continuous input of energy converted to organic matter by producers (terrestrial and aquatic plants) and consumed by various feeding (trophic) levels of consumers from herbivores (plant eaters) to carnivores (animal eaters). Nutrient cycling is the incorporation of organic and inorganic nutrients into living plants and animals for survival, growth, and reproduction and their release back into the abiotic environment through excretion and decay when organisms die. Rates and pathways of energy flow and nutrient cycling can be directly measured but this is not necessary since these functions can be indirectly assessed by the size and composition of benthic macroinvertebrate assemblages and fish populations, both of which reflect the ambient chemical composition and quality of the water in which they live.

1.5 Biotic assemblages: structure and function

1.5.1 Structure

Natural ecosystems, particularly aquatic ones, are complex and constantly changing. To understand, compare, and evaluate such complex and dynamic systems scientists, politicians, and regulators seek ways of reducing them to a single number. Historically, most of these simplifications are based on taxonomic identification (structure) using the number of species, their relative abundances, and their relationships to some index of goodness (diversity) or ecosystem integrity. There are two critical shortcomings to this approach:

1) The concepts behind many of these structural indices are frequently based on scientific or mathematical theory such as information theory (e.g., the Shannon and Simpson indices) or indices of biotic integrity. The applicability of these theories to assemblages of biological organisms have not been proven, only assumed. In addition, there is no objective criterion (mathematical, ecological, or biological) that separates a "good" index value from a "poor," "weak," or "bad" value Hurlbert 1971. The Index of Biotic Integrity (IBI) was developed by Karr (1987) as a measure of water quality and has been widely adopted across the US. A major weakness of IBI is that it is not a single, consistent equation relating a fixed set of structural components and numeric constants. Often, states develop a different IBI for each major river basin or a particular issue (agricultural, industrial, urban). The lack of both a single mathematical equation and objective criteria for assigning calculated numeric values to qualitative terms of goodness exposes the IBI to claims of being constructed to support an *a priori* decision rather than being demonstrably objective and applicable everywhere.

2) While it is relatively easy to distinguish most mammalian, avian, and fish to the taxonomic level of *species* this is not the case with aquatic insects and other macroinvertebrates. Biotic structure does not reliably work for the classification or assessment of streams and rivers for determining whether designated beneficial uses have been attained, for measurements of biodiversity, or for quantifying ecosystem function (Doledec et al. 2000, Resh et al. 2005, Chessman et al. 2007, Cuffney et al. 2007). In addition to mixed taxonomic levels being given the same weight in a diversity or integrity index, these metrics do not accommodate the range of benthic macroinvertebrate life history strategies. These life history differences mean that taxa collected at any site depends on the collection date (Lenat 1988, Clarke et al. 2002, Bruce 2002, Boyero 2005, Bogan and Lytle 2007). For example, at one extreme are the terrestrial locusts, the swarming phase of certain species of short-horned grasshoppers in the family Acrididae. Their populations emerge synchronously as reproductive adults in huge swarms over broad areas of land, quickly die after mating and egg laying, and the developing juveniles remain underground for 17 years before they pupate, emerge as flying adults, and start the next generation. At the other extreme are some species of biting aquatic black flies (order Diptera, family Simuliidae, genus *Simulium*) which may have multiple generations each year in some reaches of a stream and single generations in other reaches of the same stream. This means that different diversity and integrity indices can be calculated for the same stream or river reach depending on when and where the benthos are collected. In addition to the above issues there is nothing inherent in biotic structure represented by taxonomic identification and the ecosystem functions of energy flow and nutrient cycling. Therefore, functional attributes of the biota need to be used to determine whether designated beneficial uses are protected in compliance with the CWA.

The River InVertebrates Prediction and Classification System (RIVPACS) is a software package developed by the Institute of Freshwater Ecology (IFE) in the UK Wright 2000. The original application was to assess the biological quality of rivers within the UK; it has since been adapted and used in the US and other countries. RIVPACS offers site-specific predictions of the macroinvertebrate fauna to be expected in the absence of *major environmental stress*. The expected fauna is derived by RIVPACS using a small set of environmental characteristics. The biological evaluation is then obtained by comparing the fauna observed at the site with the expected fauna. RIVPACS is a major advance over diversity and biotic integrity indices. However, it retains the short-comings of all metrics based on taxonomy. Regardless, the basic concepts of identifying biological characteristics of specific sites or small drainage basins are sound and can form the foundation of a metric that identifies change in the benthic macroinvertebrate community beyond those inherent in natural flowing water ecosystems. It is possible to use RIVPACS to simulate faunal changes in response to environmental disturbance, provided that the disturbance directly involves the environmental variables used in RIVPACS predictions Armitage et al. 1983. These variables relate to channel shape, discharge and substratum. Many impacts, particularly those associated with pollution, will not affect these variables and therefore RIVPACS cannot simulate the effects of pollution. In a UK study RIVPACS was sensitive only to major changes in substratum. The author concluded that, because of the static nature of RIVPACS, it cannot respond to the dynamic effects and processes associated with environmental disturbance. Thus RIVPACS, while showing direction of change and indicating sensitive taxa, cannot be used to predict or forecast the effects of environmental impacts Armitage 2000. Recently, Ritz (2010) applied RIVPACS to determine if predictive models of diatom assemblages would provide an effective method to report on biological degradation in streams along the Central Coast of California. The author concluded that, "[T]he RIVPACS model did not perform well. The model suffered from low precision of reference site O/E[3] scores (mean SD = 0.22) and lack of accuracy to consistently predict low O/E scores at known degraded sites. However, the model was able to identify likely trends. For example, agricultural land use sites trended toward lower O/E scores indicating possible biological degradation." Published articles suggest that RIVPACS models might predict benthic macroinvertebrate communities but they

[3] Observed compared to Expected taxa

Table 1.1: Taxonomic levels and numbers in each of seven streams draining the Independence Mountains. Many fewer species than genera are the result of not being able to identify all organisms to species.

Taxon	Number
Class	11
Order	19
Family	86
Genus	208
Species	60

not be capable of detecting low-level anthropogenic changes if those changes are not in the physico-chemical data set characterizing reference sites (Turak et al. 1999, Clarke et al. 2002, Linke et al. 2005, Van Sickle et al. 2005, Yuan 2006).

Table 1.1 summarizes benthic macroinvertebrate taxonomic levels in collections over more than a decade from seven streams draining the southern Independence Mountains in northern Elko County, Nevada.

1.5.2 Function

The distribution of benthic macroinvertebrates is dependent on physical, chemical, and biological factors. The ecosystem changes along the stream or river network Vannote et al. 1980, and these changes are reflected by the organic materials (foods) within the stream channel. Organic particles vary in size and place within the channel and water column. These niches are exploited by organisms adapted to a particular particle size and method of feeding. These methods of feeding allow benthic macroinvertebrate taxa to be assigned to functional feeding groups which describe the trophic behaviors of the organisms and allow consistent useful insights to be drawn from the data (Cummins 1973, Faith 1990, Boyero 2005, Cummins et al. 2005).

Everyone has heard of food webs and food chains. These use the language of economics to divide organisms into three major categories: producers (plants), consumers (animals), and decomposers (fungi and bacteria). Plants convert sunlight into organic molecules and compounds which they use for growth and reproduction as well as providing energy and nutrients for the herbivore consumers who feed on them. Herbivores are also the source of energy and nutrients for the predators that eat them. In terrestrial ecosystems there are different plants with different growth and nutritional characteristics in different physical environments. Plant-eating animals exhibit different feeding strategies; browsers such as deer and elk eat leaves, bark, and stem ends of plants while grazers such as sheep and cattle clip grasses and forbs at or near ground level. By feeding on different types and parts of vegetation herbivores can co-exist in the same environment without competing for food resources. There is also a hierarchy of carnivorous predators. Some, like coyotes, feed primarily on small herbivores such as rabbits, mice, and voles while others such as wolves and cougars feed on larger herbivores (sheep, cattle, deer, elk). The same divisions exist in aquatic ecosystems.

The top of all freshwater aquatic food webs and chains are fish such as trout and bass. As in terrestrial ecosystems, top-level aquatic predators are highly mobile and occupy different habitats based on abiotic environmental conditions (temperature, dissolved oxygen, water flow velocity, channel width-to-depth ratios) and the available food resources. Benthic macroinvertebrates are less migratory than are the fish yet the do exhibit downstream movements ("drifting") dependent upon life stage and population pressures. Their food resources are relatively stable but are rearranged when the flow regime changes (during snow melt runoff, for example).

Aquatic producer organisms are algae, mosses, and vascular aquatic plants such as duckweek, cattails, reeds, and sedges. In areas where terrestrial vegetation is in or near the riparian zone leaves and twigs fall into the stream channel as coarse particulate organic matter (CPOM, particles >1mm) and are the primary energy and nutrient resource where the channel is shaded and has insufficient sunlight to support the growth of algae on the substrate. Aquatic consumer organisms (benthic macroinvertebrates) are divided into four main functional feeding groups (FFG, Cummins 1973): shredders, collectors (frequently divided into filterer collectors and gatherer collectors), scrapers (grazers), and predators. These groups are based on the location of their foods and the size range of organic particles they ingest.

Shredders feed on leaves, grasses, twigs, and other CPOM. Their function is to break down these large particles into smaller ones while obtaining the energy they need for survival, growth, and reproduction. Fine particulate organic matter (FPOM, particles <1mm and >45μm) has a large size range. Those particles small enough to be transported in the flowing water column are captured by filtering collectors for food; larger, loose FPOM that settles on the channel bottom, rocks, or large wood are fed upon by gathering collectors. In stream and river reaches open to sunlight the incoming solar energy is converted to organic molecules by algae and mosses (the producers) which are themselves food for scraping (grazing) consumers. Predators, too, have two feeding strategies: piercing and ingesting.

Using benthic macroinvertebrate functional feeding groups as the basis for setting water quality standards provides regulators with the most technically sound and legally defensible foundation for statistical modeling.

Chapter 2

Independence Mountains Drainage Basins and Streams

2.1 Introduction

The data analyzed in this monograph come from annual collections of macroinvertebrates and fish in five stream networks draining the southern Independence Mountains. One stream (Winters Creeks) drains eastward in the North Fork Humboldt River basin and four streams (Snow Canyon, Jerritt Canyon, Burns, and Starvation Creeks) drain westward in the South Fork Owyhee River basin (Figure 2.1). Each of the five streams has not been visited every year, nor for the same period of years, and data are reported from Sheep Creek in only 2012 and from California Creek in only 2013; Water Pipe Creek had no water chemistry available for use as explanatory variables.

As noted in Section 1.3, fish and benthic macroinvertebrate communities reflect abiotic environmental parameters such as basin, stream channel, and riparian characteristics, current velocity and volume, streambed composition, water temperature, dissolved oxygen, and water chemistry. Some of the more advanced approaches to characterizing benthic macroinvertebrate community structure incorporate geomorphic and other parameters as explanatory variables. For example, RIVPACS includes latitude, longitude, elevation, basin area, basin geology, channel slope, streambed substrate composition, alkalinity, data collection date, mean annual precipitation, and mean annual air temperature. And a study on the distribution of Lahontan cutthroat trout in northern Nevada uses stream channel width:depth ratio as an explanatory variable (Cade and Noon 2003).

Permit compliance monitoring differs greatly from academic or agency research in the amount of available data and the frequency of collection. However, characteristic drainage basin parameters are easily extracted from readily available digital elevation models (DEMs) which represent altitude at defined points on the Earth's surface at various size scales. The 10m resolution[1] DEMs, based on US Geological Survey 1:24000 topographic maps, are suitable for extracting drainage basins and characterizing them topographically and hydrologically. For small areas, or for answering questions requiring high resolution data, LIDAR[2] can provide 1-2cm horizontal cell size.

For interbasin comparisons of the seven Independence Mountain drainages 11 geomorphic parameters can be used as explanatory variables: area, perimeter, aspect, elevations (maximum, minimum, difference), shape factor, concentration time, total stream length, drainage density, and main channel slope. The stream order used to describe each stream network is based on Strahler's numbering. Strahler (1957) described a stream order system as a simple method of classifying stream segments based on the number of tributaries upstream. A stream with no tributaries (headwater stream) is considered a first order stream. A segment downstream of the confluence of two first order streams is a second order stream. Thus, a n^{th} order stream is always located downstream of the confluence of two $(n-1)^{th}$ order streams.

[1]Each cell measures 10 meters on a side.
[2]Light Distance And Ranging; similar to RADAR but using light waves rather than radio waves.

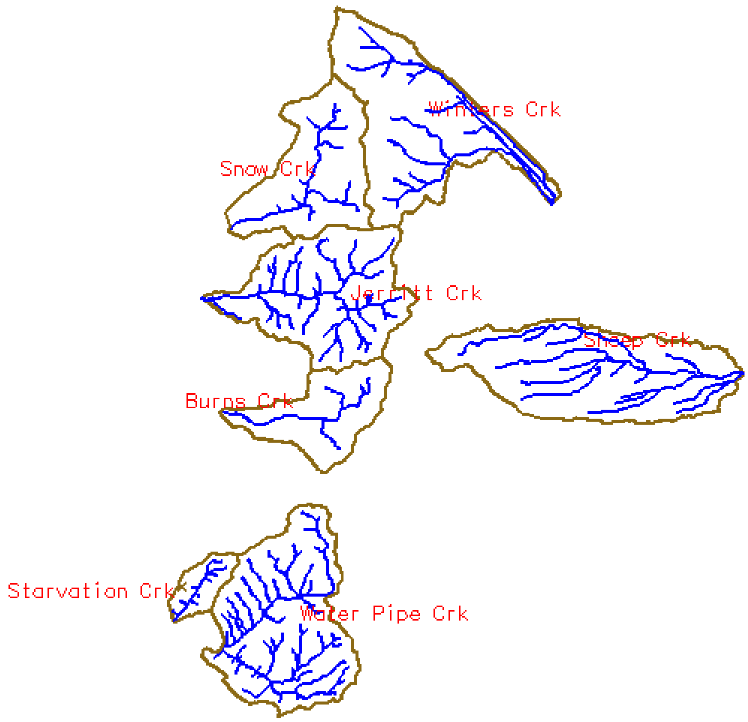

Figure 2.1: Drainage basins and stream networks for the seven streams for which biological data are available. There are meaningful differences in drainage basin and stream network characteristics among the seven areas.

2.2 Parameters

2.2.1 Basin area

Basin area is related to water discharge at specified occurrence frequencies. Bankfull discharge has a recurrence interval averaging 1.5 years and its relation to drainage basin area is given by the equation $Q \propto A^{0.75}$ where Q is discharge in volume per unit time and A is drainage basin area in equivalent units (e.g., Q in litres per second and A in kilometers squared). The mean annual discharge usually fills a stream channel to about one-third of its bankfull depth, and it tends to have a similar frequency of occurrence among streams of different features. This flow is equalled or exceeded about 25% of the time. Discharge is related to the distribution of streambed particle sizes and organic food particles used by benthic macroinvertebrates.

2.2.2 Basin perimeter

Perimeters of drainage basins, lakes, countries, and similar natural edges exhibit fractal behavior. Therefore, perimeter-based shape indices depend on the scale at which they are drawn. However, perimeters are used to develop shape indices that describe the characteristics and hydrological properties of the basin and influence the aquatic biota within the stream network. Computationally, the basin perimeter length is that of the contour if the basin is projected onto a flat surface.

2.2.3 Aspect

Aspect (terrain orientation) can affect snowmelt and the amount of evaporation, with south- and west-facing slopes more exposed to the sun's influence than are north- and east-facing slopes. Terrestrial (riparian) vegetation, soils, and hydrological responses to the compass orientation of the drainage basin influence channel and streambed characteristics and the aquatic biota living there.

2.2.4 Elevations (Maximum, Minimum, Difference)

Maximum and minimum elevations affect air and water temperatures, dates and extents of snowpack, and the growing season for riparian vegetation. The difference in elevations is related to the average slope of the main stream channel, current velocity, and discharge.

2.2.5 Basin shape factor

There are several factors describing basin shape; for example, the circularity ratio (the ratio between the area of the basin and the area of the circle having the same perimeter), the elongation ratio (the ratio between the diameter of the circle having the same area of the basin and the length of the main channel), the compactness coefficient (the ratio between the perimeter of the basin and the diameter of the circle having the same area), and the shape factor (the ratio between the area of the basin and the square of the length of the main channel). Because basin size and stream network topology vary so much in the Independence Mountain streams the shape factor reflecting basin area and main channel length is easier to interpret as an explanatory variable for the observed aquatic biota.

2.2.6 Concentration time

Concentration time relates drainage basin area, the length of the main channel, and the difference between the maximum and minimum elevations in the basin to calculate the time it takes water to reach the outlet. It is used to plan culvert and detention basin sizes and is a measure of current velocity and basin discharge. Considering the large differences in basin size, drainage density, and aspect concentration time is another potential explanatory variable for observed aquatic biota.

2.2.7 Total stream length

Total stream length affects habitat quantity and variability within the basin. The more complex the stream network the greater the number of low order tributaries. Low order tributaries are associated with steep and narrow valleys, abundant overhanging riparian vegetation, and a high percentage of shredder organisms and grazers capable of avoiding high velocity currents by hiding in coarse substrates or by having flattened body shapes that keep them in the low-velocity microlayer just above the substrate's surface.

2.2.8 Drainage density

Drainage density is the ratio between the total length of the stream network and the area of the drainage basin. A higher ratio (usually > 1.0) reflects more diverse habitats and flow conditions which support more aquatic biota to process the greater amounts of incoming energy and more efficiently use the nutrients in support of growth and reproduction.

2.2.9 Main channel slope

Main channel slope affects current velocity and the composition of the streambed. The steeper the average slope the faster the current and the more fine particulate organic matter (FPOM) entrained in the water column. This would support more filter-gatherers feeding on these food particles and support fewer collector-gatherers because comparatively less FPOM settles on the stream bed. In more shallow drainage basins the current velocity is slower and finer gravels and sands remain on the streambed where they support settled CPOM and FPOM as food for shredders and collector-gatherers.

2.3 Humboldt River basin

2.3.1 Winters Creek

The northeast corner of the Jerritt Canyon Mine property is within the Winters Creek basin on the east side of the Independence Mountains. Winters Creek is a fourth order stream network with 26 segments in a combination of dendritic and parallel drainage patterns. The basin has an unusual shape because the lower half of its area is

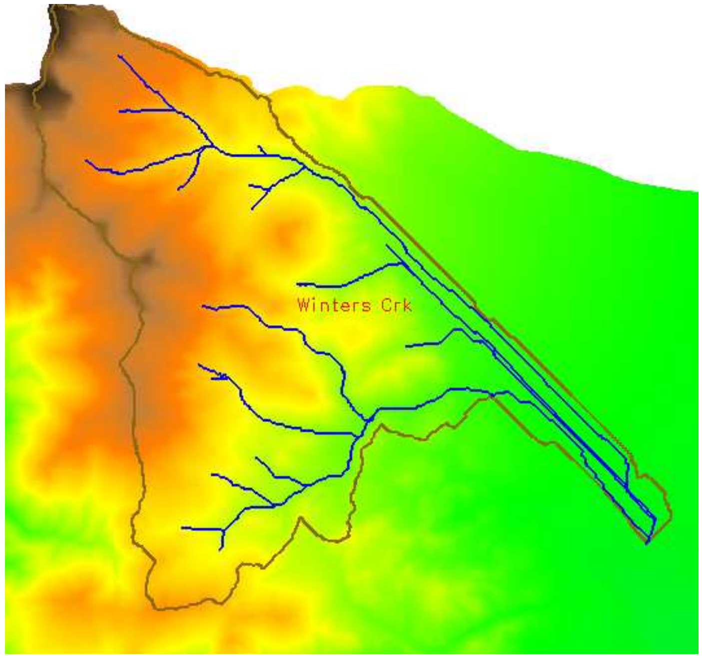

Figure 2.2: Winters Creek basin and stream network with elevations shown in color. The lowest elevation is in green and the colors change to yellow, orange, and brown as elevation increases.

close to the valley floor and stream channels run in parallel along that boundary (Figure 2.2). Geomorphometric parameters of the Winters Creek basin are shown in Table 2.1. The distribution of drainage basin area at each elevation is shown in Figure 2.3. Looking at this plot we see that half the area of the drainage basin is above 2,300 m elevation, which is approximately 67% of the elevation difference (Table 2.1); in other words, about half the area is in the upper $\frac{2}{3}$ of the basin.

2.3.2 Sheep Creek

Sheep Creek is approximately in the middle of the eastern side of the Jerritt Canyon Mine property. It is a third order stream network with 20 segments arranged in a dendritic drainage pattern within an oblong shape (Figure 2.4). About 75% of the basin area, and the stream network that drains it, is in low-gradient foothills (Figure 2.5). Geomorphic parameters of the Sheep Creek basin are shown in Table 2.2).

Table 2.1: Geomorphic parameters of Winters Creek drainage basin affecting hydrology and related to aquatic biota.

Parameter	Value
Area (km^2)	39.41
Perimeter (km)	133.34
Aspect (degrees)	124.2
Max. elevation (masl[3])	3143.27
Min. elevation (masl)	1921.50
Elevation difference (m)	1221.77
Shape factor	43.59
Concentration time (hrs.)	3.12
Total stream length (km)	51.13
Drainage density (km/km^2)	1.30
Main channel mean slope (%)	1.76

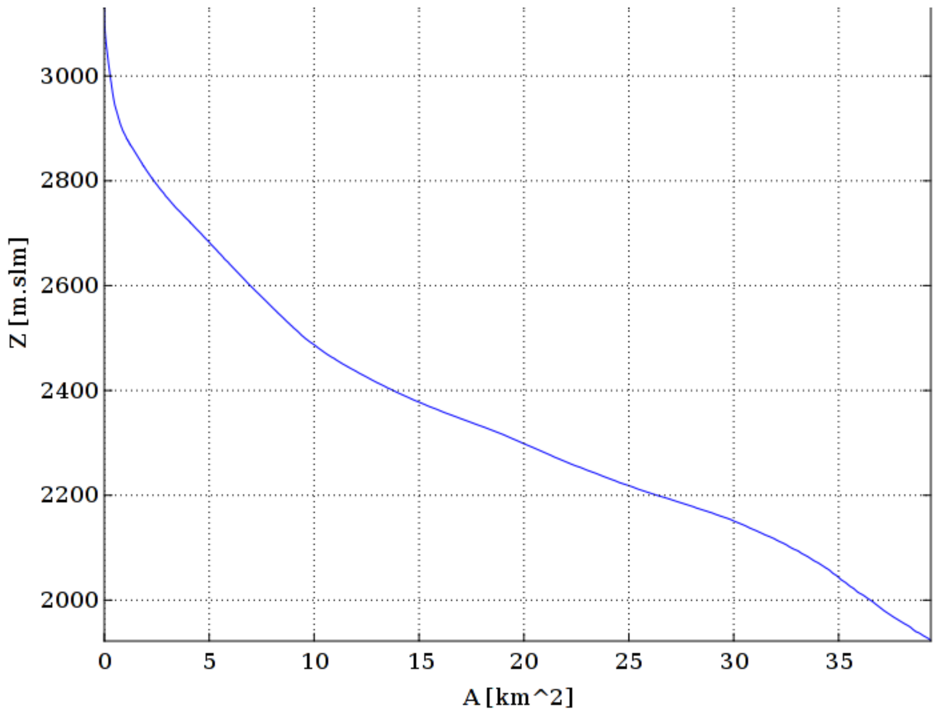

Figure 2.3: Winters Creek hypsographic curve showing the relationship of basin area to elevation. Between 2100–2400 m elevation the basin area decreases linearly from approximately 30 km^2 to 14 km^2 of a total basin area slightly less than 40 km^2.

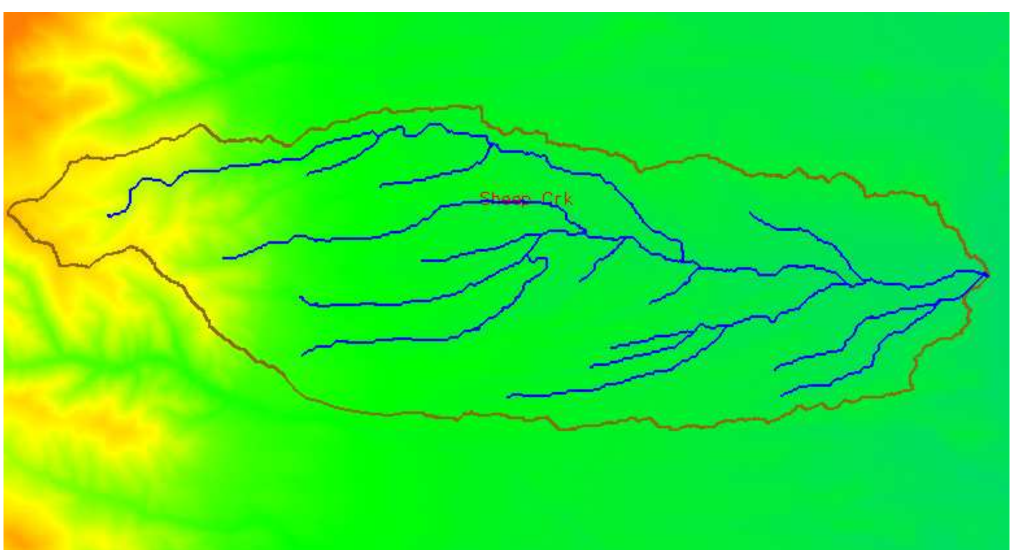

Figure 2.4: Sheep Creek basin and stream network with elevations shown in color.

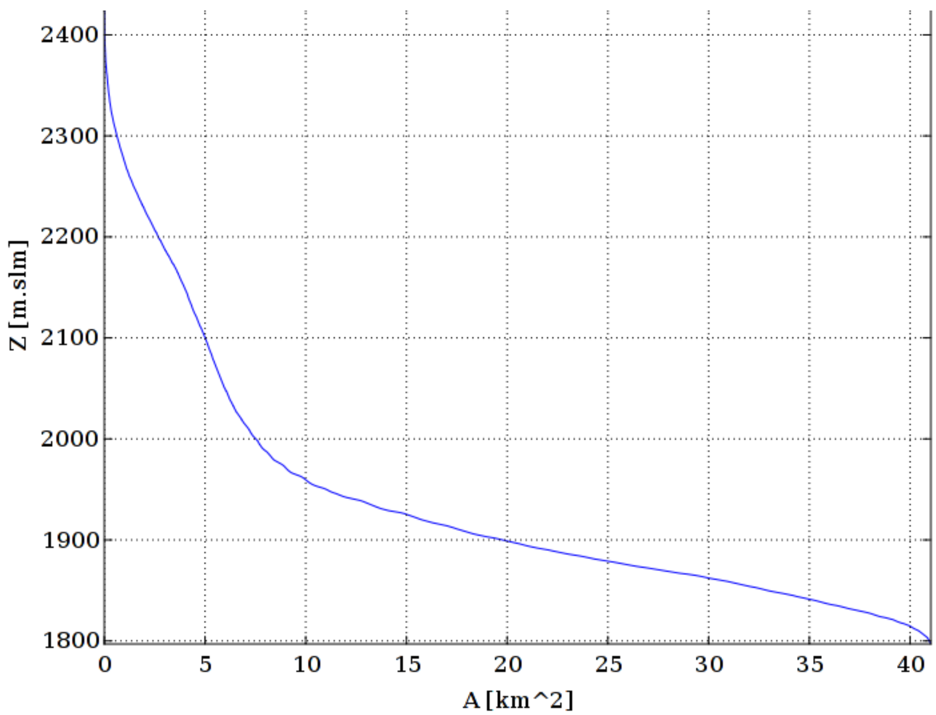

Figure 2.5: Hypsographic curve of the Sheep Creek drainage basin showing the amount of basin area at different elevations. The upper end is comparatively steep but the lower end is much more shallow.

Table 2.2: Geomorphic parameters of Sheep Creek drainage basin affecting hydrology and related to aquatic biota.

Parameter	Value
Area (km^2)	41.06
Perimeter (km)	123.57
Aspect (degrees)	93.9
Max. elevation (masl)	2431.47
Min. elevation (masl)	1796.78
Elevation difference (m)	634.69
Shape factor	60.24
Concentration time (hrs.)	4.97
Total stream length (km)	56.67
Drainage density (km/km^2)	1.38
Main channel mean slope (%)	0.59

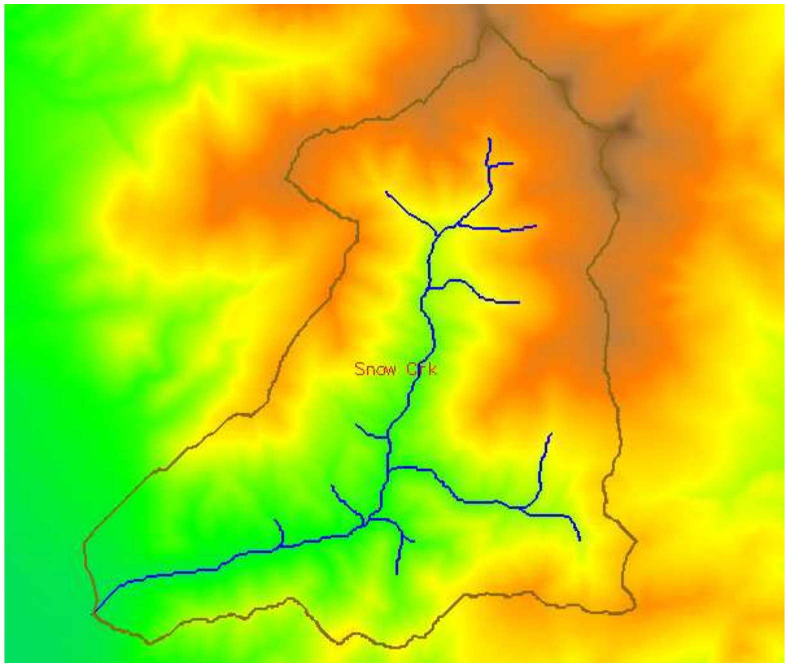

Figure 2.6: Snow Canyon Creek basin and stream network with elevations shown in color.

2.4 Owyhee River basin

2.4.1 Snow Canyon Creek

Snow Canyon Creek is the northernmost stream on the west side of the mountains. It is comparatively small and has a simple third order stream network with 11 segments in a trellis drainage pattern (Figure 2.6). Most of the basin is in the middle elevations; the highest and lowest elevations have comparatively little of the basin's area and the changes are quite steep at both ends (Figure 2.7). Geomorphic parameters of the Snow Canyon Creek basin are shown in Table 2.3).

Table 2.3: Geomorphic parameters of Snow Canyon Creek drainage basin affecting hydrology and related to aquatic biota.

Parameter	Value
Area (km^2)	28.52
Perimeter (km)	127.54
Aspect (degrees)	220.4
Max. elevation (masl)	2942.83
Min. elevation (masl)	1743.26
Elevation difference (m)	1199.57
Shape factor	79.33
Concentration time (hrs.)	3.35
Total stream length (km)	25.23
Drainage density (km/km^2)	0.88
Main channel mean slope (%)	2.92

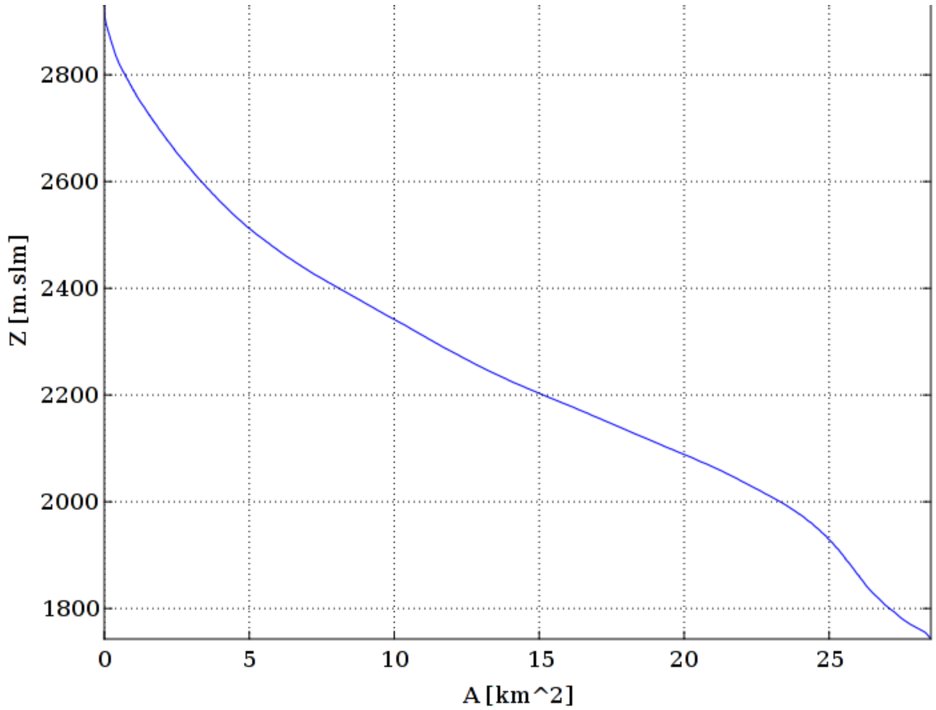

Figure 2.7: Hypsographic curve of the Snow Canyon Creek basin showing the amount of basin area at different elevations.

2.4.2 Jerritt Canyon Creek

Jerritt Canyon Creek basin is in the middle of the three adjacent westside drainages, and has a third order stream network with 22 segments in a dendritic drainage pattern (Figure 2.8). The amount of drainage basin area at each elevation level is shown in Figure 2.9. Geomorphic parameters of the Jerritt Canyon Creek basin are shown in Table 2.4).

2.4.3 Burns Creek

The Burns Creek drainage basin is small with a second order stream network having 6 segments in a simple dendritic drainage pattern (Figure 2.10). Approximately 70% of the basin's area lies between 2,200–2,400 m elevation; above and below that range the amount of basin area decreases rapidly (Figure 2.11). Geomorphic parameters of the Burns Creek basin are shown in Table 2.5).

2.4.4 Starvation Canyon Creek

Starvation Creek basin is the smallest of the seven basins (about 5 km^2) with a second order stream network of 10 segments in a trellis drainage pattern at the south end of the Jerritt Canyon Mine property (Figure 2.12). The basin is oblong withe a broad central valley. The stream network has a trellis pattern with short tributaries off each side of the main channel and the distribution of basin area by elevation level is similar to other streams draining these mountains (Figure 2.13). Geomorphic parameters of the Starvation Creek basin are shown in Table 2.6).

2.4.5 Water Pipe Creek

The Water Pipe Creek basin is large (about 40 km^2) and drained by a fourth order stream network of 53 stream segments in a dendritic pattern. It is adjacent to the east side of the Starvation Canyon basin and both streams flow under Highway 226 into Taylor Creek (Figure 2.14). The highest elevations are in the north and most tributaries there flow to the southeast. Because only the northern most area of the basin is at a high elevation the

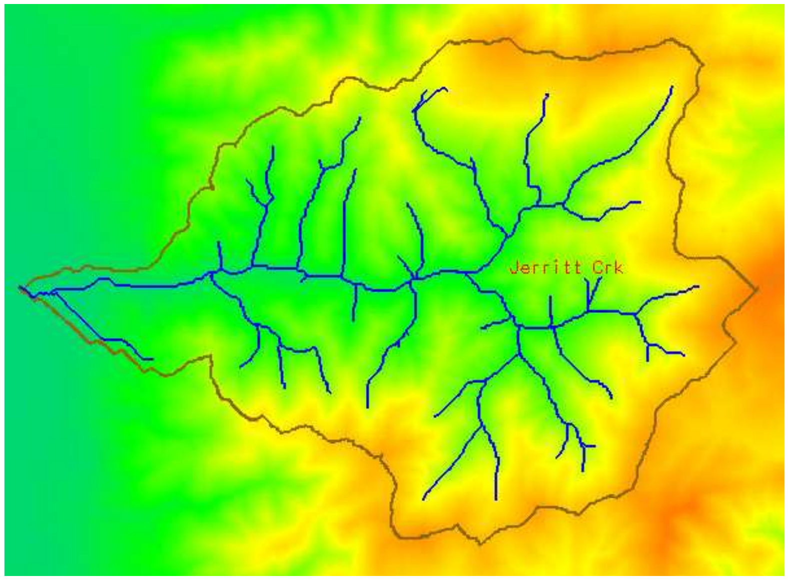

Figure 2.8: Jerritt Canyon Creek drainage basin and stream network with elevations shown in color.

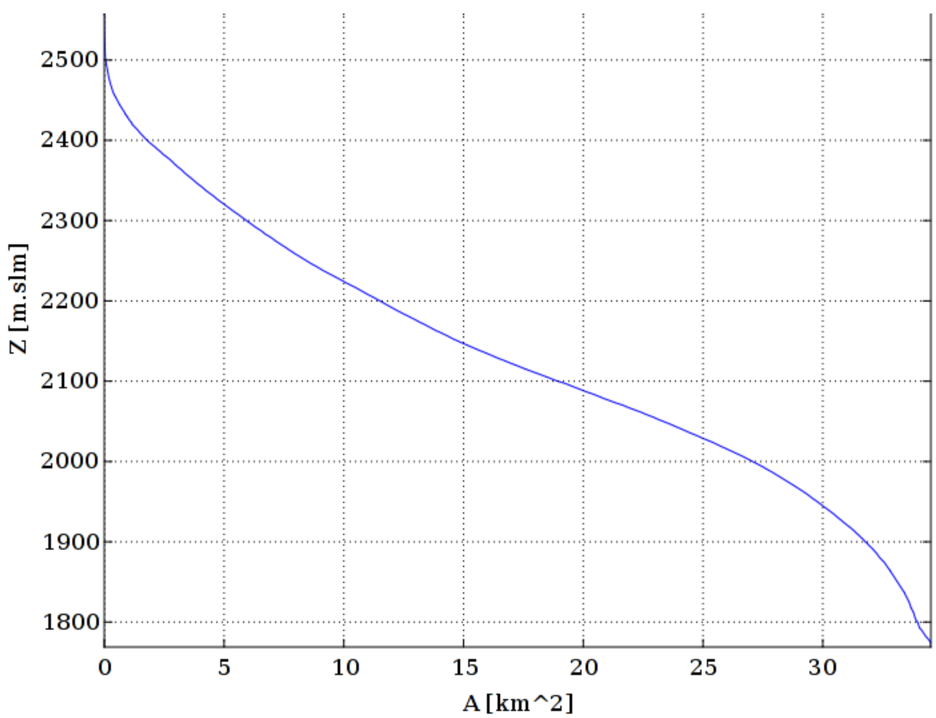

Figure 2.9: Hypsographic plot of basin area at each elevation level for the Jerritt Canyon Creek basin.

Table 2.4: Geomorphic parameters of Jerritt Canyon Creek drainage basin affecting hydrology and related to aquatic biota.

Parameter	Value
Area (km^2)	34.53
Perimeter (km)	106.14
Aspect (degrees)	254.4
Max. elevation (masl)	2566.42
Min. elevation (masl)	1769.18
Elevation difference (m)	797.24
Shape factor	32.20
Concentration time (hrs.)	3.26
Total stream length (km)	29.50
Drainage density (km/km^2)	0.86
Main channel mean slope (%)	1.32

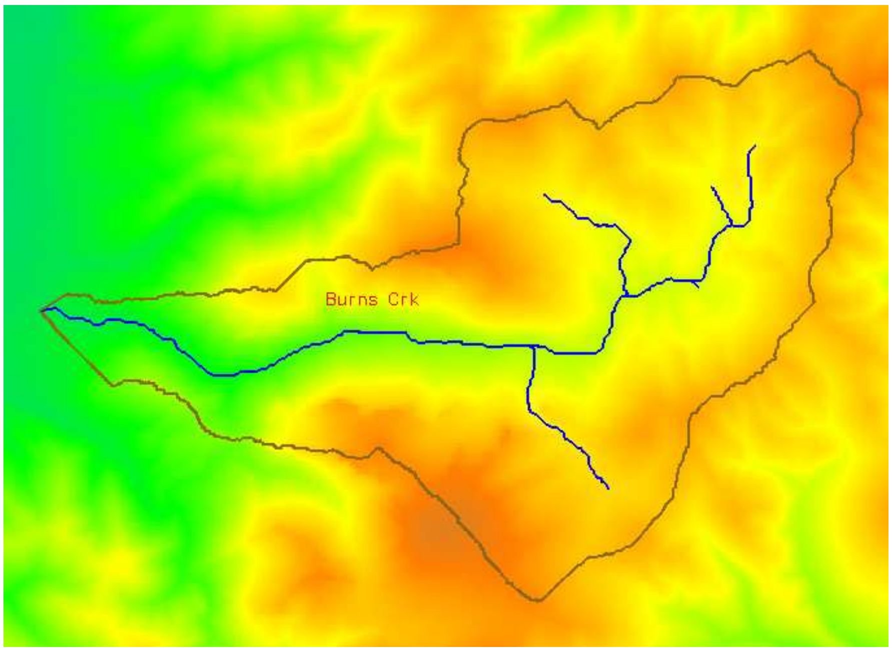

Figure 2.10: Burns Creek drainage basin and stream network with elevations shown in color.

Table 2.5: Geomorphic parameters of Burns Creek drainage basin affecting hydrology and related to aquatic biota.

Parameter	Value
Area (km^2)	17.30
Perimeter (km)	103.69
Aspect (degrees)	257.0
Max. elevation (masl)	2685.17
Min. elevation (masl)	1750.18
Elevation difference (m)	934.98
Shape factor	98.35
Concentration time (hrs.)	3.21
Total stream length (km)	16.44
Drainage density (km/km^2)	0.95
Main channel mean slope (%)	1.23

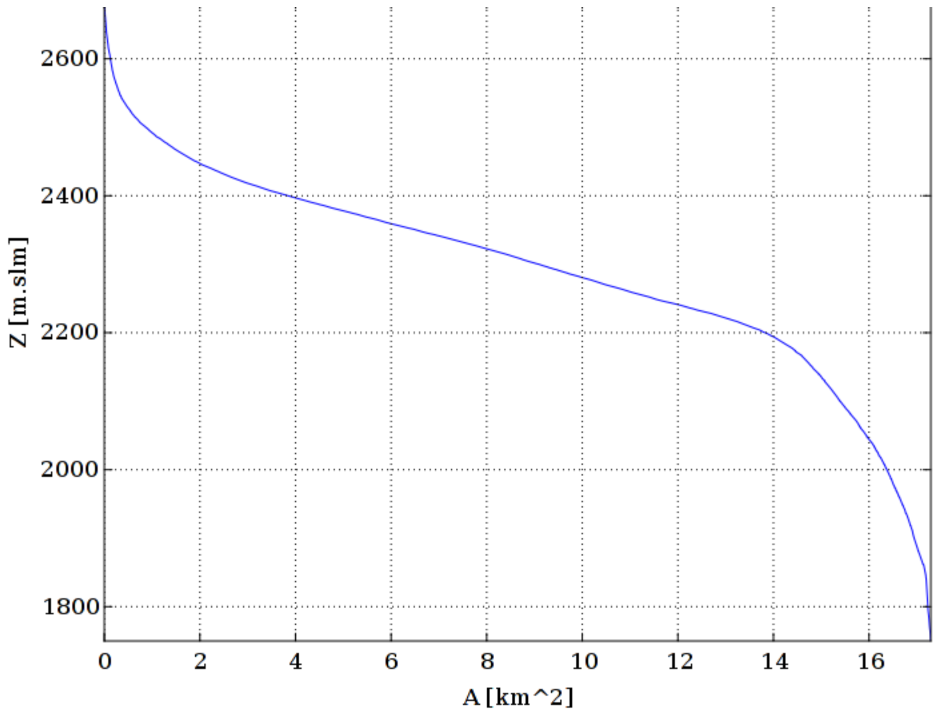

Figure 2.11: Hypsographic curve of Burns Creek basin showing the amount of area at each elevation level.

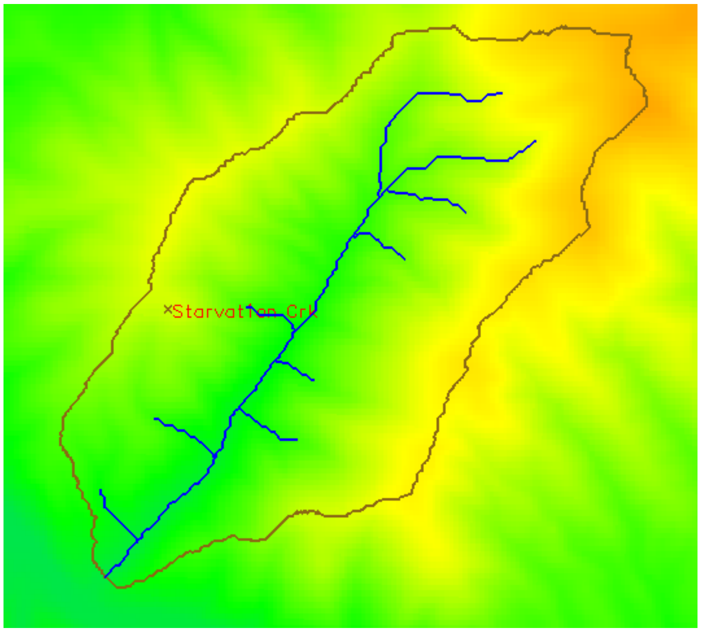

Figure 2.12: Starvation Creek drainage basin and stream network with elevations in color.

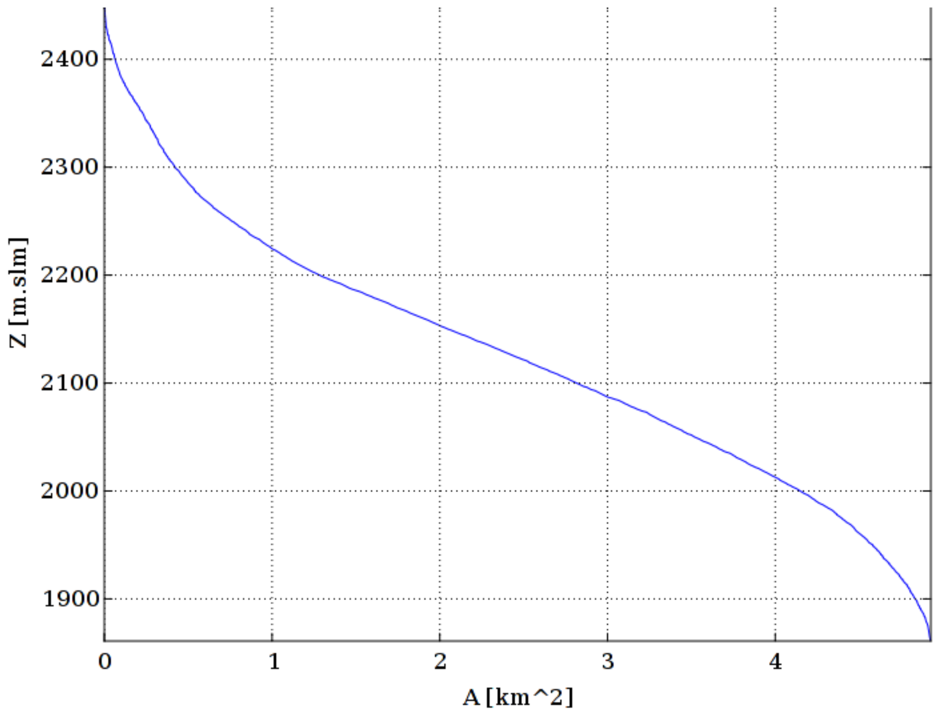

Figure 2.13: Hypsographic plot of Starvation Creek showing the basin area at each elevation level.

Table 2.6: Geomorphic parameters of Starvation Creek drainage basin affecting hydrology and related to aquatic biota.

Parameter	Value
Area (km^2)	4.93
Perimeter (km)	34.87
Aspect (degrees)	219.4
Max. elevation (masl)	2454.25
Min. elevation (masl)	1860.53
Elevation difference (m)	593.73
Shape factor	29.64
Concentration time (hrs.)	1.39
Total stream length (km)	5.84
Drainage density (km/km^2)	1.19
Main channel mean slope (%)	2.72

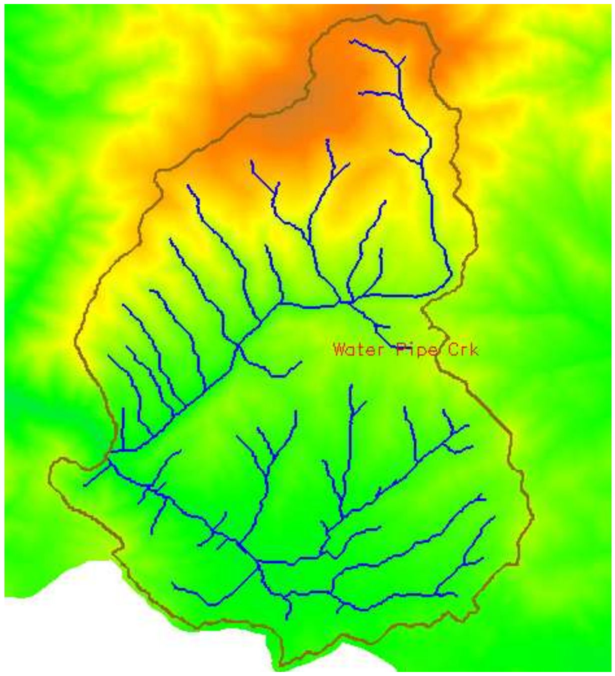

Figure 2.14: Water Pipe Creek basin and stream network with elevations shown in color.

hypsographic curve has a concave shape, unlike the other drainages (Figure 2.15). Geomorphic parameters of the Water Pipe Creek basin are shown in Table 2.7).

Table 2.7: Geomorphic parameters of Water Pipe Creek drainage basin affecting hydrology and related to aquatic biota.

Parameter	Value
Area (km^2)	39.52
Perimeter (km)	110.00
Aspect (degrees)	210.5
Max. elevation (masl)	2759.56
Min. elevation (masl)	1884.18
Elevation difference (m)	875.39
Shape factor	30.83
Concentration time (hrs.)	3.27
Total stream length (km)	60.00
Drainage density (km/km^2)	1.52
Main channel mean slope (%)	2.04

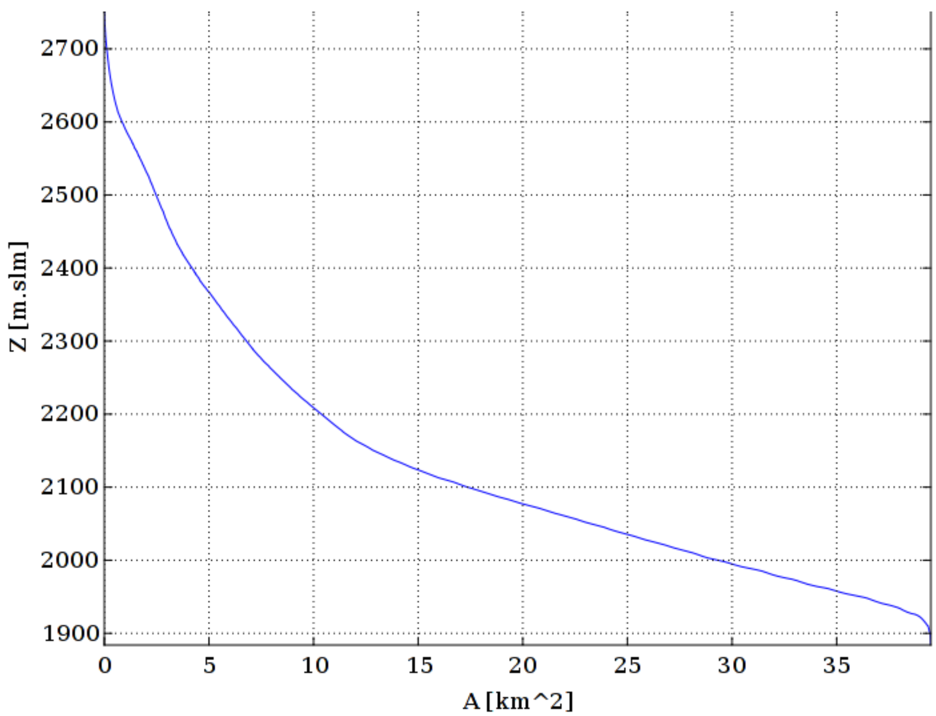

Figure 2.15: Hypsographic curve of the Water Pipe Creek drainage basin showing the distribution of area at each elevation level.

Chapter 3

Aquatic Biota

3.1 Introduction

Fish were collected in only four of the seven streams and were identified as trout (Lahontan cutthroat, redband, brook), Paiute sculpin, and speckled dace. There are too few fish to analyze. They will not be further considered.

Regardless of the sampler and protocol used to gather macroinvertebrates from a stream or river bed the results are a collection and not a statistical sample. The difference is important in selecting analytical methods and statistical models that are technically sound (that is, robust and appropriate) and legally defensible. In general terms, a sample is a subset of the entire population and it can be shown that the sample is representative of that population. A single cookie selected from a freshly baked batch is a sample because it is reasonable to assume that the dough was thoroughly and evenly mixed so any one cookie is indistinguishable from any other cookie. Terrestrial biological data are more easily determined to be samples extracted from the entire population but this is not the case with aquatic organisms, particularly the benthic macroinvertebrates.

Every sampler used to collect benthic macroinvertebrates has a fixed mesh size. Invariably, some very young juveniles will be smaller than the holes in the mesh and will be swept through and not be represented in the collection. Larger individuals can move out of the net or be carried out the mouth by hydraulic back pressure against a small mesh size in fast flowing waters. Other individuals may have a flat body shape that allows them to be tightly pressed against the surface of gravels and cobbles and not detached into the net while other taxa will burrow deep within the substrate even as that material is being stirred up to release the organisms into the water column. Since the number and taxonomic identification of all these potential losses cannot be known the resulting data represents a collection and not a true sample representative of the entire benthic macroinvertebrate community in the collection area. Also, the life cycle of benthic macroinvertebrates is variable and not synchronous so the taxa depend to a certain degree on when they are collected.

The overall pattern of relative abundances among the functional feeding groups is similar in all streams (Figure 3.1). Collector gathers are the most abundant FFG with exceptions in some streams at some locations and dates, but the pattern reveals that most of the organic food for benthic macroinvertebrates in these steams are fine particles deposited on the substrate. Shredders have a wide range of relative numbers reflecting the variability among individual drainage basin and stream network characteristics. There are proportionally fewer shredders in the three streams (California, Winters, and Sheep Creeks) draining the east side of the mountains than in the five streams draining the west side.

Because most collection efforts have occurred during the late spring and early summer individual sites at several streams might be dry so no data are available. The irregularity of collections might contribute to the perceived variabilities seen in the data. The date variability of the single annual samples is another justification for using functional feeding groups rather than taxon as the analytical units.

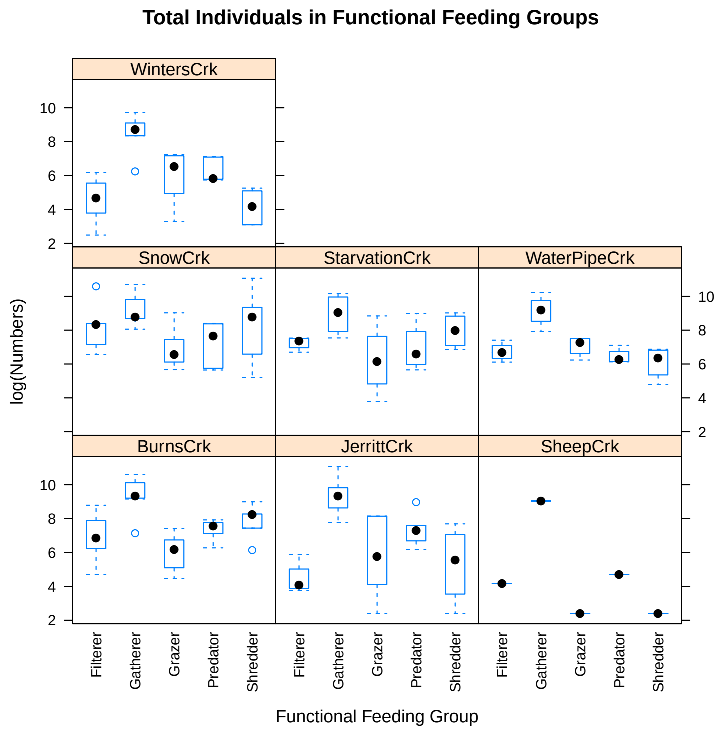

Figure 3.1: Boxplots showing the distribution of individuals in each functional feeding group for all data from each stream.

3.2 Humboldt River basin

3.2.1 Winters Creek

Benthic macroinvertebrates were collected from Winters Creek in six years: 2004, 2005, 2006, and 2011–2013 (Table 3.1) . In the first 3-year period the relative proportions of each FFG slowly changed from the first year (2004) to the third year (2006); recently, the year-to-year changes are much greater (Figure 3.2).

3.2.2 Sheep Creek

Sheep Creek has benthic macroinvertebrate data from only one date: June 1, 2006 so these data cannot be statistically analyzed. As more data from this stream become available patterns will emerge that describe the functional dynamics of this drainage basin. The number of taxa by FFG is shown in Table 3.2 and displayed in Figure 3.3.

3.2.3 California Creek

Like Sheep Creek, California Creek has benthic macroinvertebrate data for only one date: July 9, 2013. These data cannot be statistically analyzed. The number of taxa by FFG is shown in Table 3.3 and displayed in Figure 3.4.

3.3 Owyhee River basin

3.3.1 Snow Canyon Creek

Fish were collected in Snow Canyon Creek at stations SC and SC-100 in 2011–2013. These data are presented in Tables3.4.

Benthic macroinvertebrate data from Snow Canyon Creek are available from 2005, 2006, and 2010–2013 (Table and 3.5). Changes in proportions of FFGs, both number of individuals and number of taxa, were greater year-

Table 3.1: Winters Creek: number and percentage of taxa in each functional feeding group.

	2004		2005		2006		2011		2012		2013	
FFG	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Filterer	3	6.67	3	6.12	3	4.35	4	4.40	3	5.00	1	1.61
Gatherer	27	60.00	28	57.14	43	62.32	54	59.34	26	43.33	39	62.90
Grazer	3	6.67	3	6.12	7	10.14	6	6.59	4	6.67	2	3.23
Predator	11	24.44	13	26.53	15	21.74	24	26.37	22	36.67	18	29.03
Shredder	1	2.22	2	4.08	1	1.45	3	3.30	5	8.33	2	3.23
Totals	45		49		69		91		60		62	

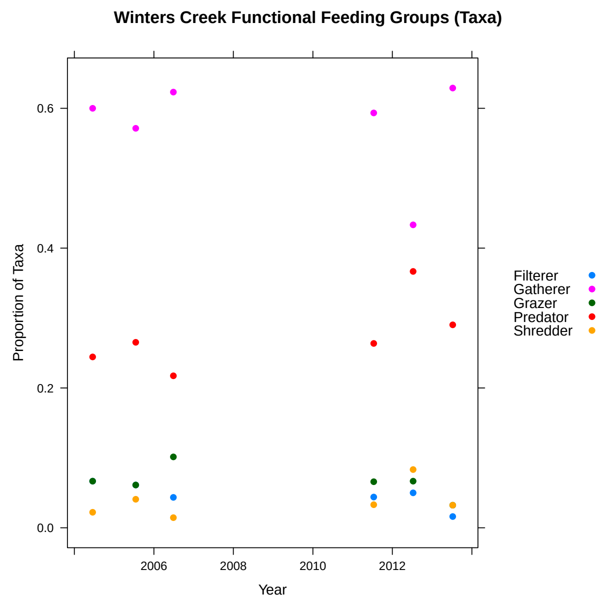

Figure 3.2: Relative proportions of taxa in each functional feeding group in Winters Creek in 2004–2006 and 2011–2013.

Table 3.2: Sheep Creek: number and percentage of taxa in each functional feeding group.

FFG	Number	Percent	Taxa	Percent
Filterer	65	0.75	3	10.71
Gatherer	8,418	97.72	17	60.71
Grazer	11	0.13	1	3.57
Predator	109	1.27	6	21.43
Shredder	11	0.13	1	3.57
Total	8,614		28	

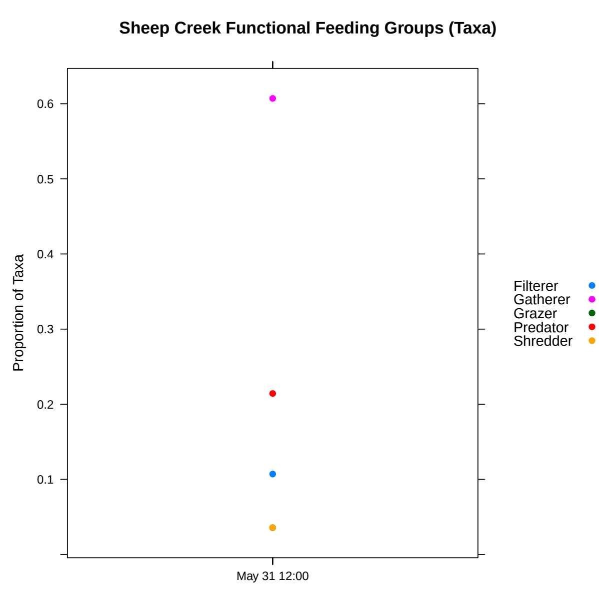

Figure 3.3: Proportion of individuals and taxa FFG in Sheep Creek in June 2006.

Table 3.3: California Creek: number and percentage of taxa in each functional feeding group.

FFG	Taxa	Percent
Filterer	3	8.57
Gatherer	20	57.14
Grazer	2	5.71
Predator	9	25.71
Shredder	1	2.86
Total	35	

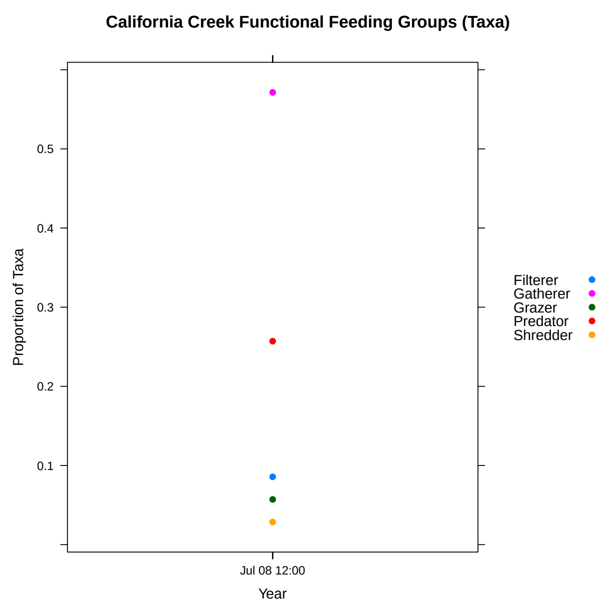

Figure 3.4: Proportion of individuals and taxa FFG in Sheep Creek in June 2006.

Table 3.4: Fish collected from Snow Canyon Creek.

Site	Date	Common Name	Count
SC	2011-07-14	Paiute sculpin	65
SC	2011-07-14	Redband trout	16
SC	2012-07-12	Paiute sculpin	114
SC	2012-07-12	Redband trout	25
SC	2013-07-10	Paiute sculpin	133
SC	2013-07-10	Redband trout	62
SC-100	2012-07-10	Redband trout	1
SC-100	2013-07-09	Redband trout	6

Table 3.5: Snow Canyon Creek: number and percentage of taxa in each functional feeding group.

FFG	2005		2006		2010		2011		2012		2013	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Filterer	3	5.56	2	3.70	5	7.35	3	1.30	4	5.48	1	2.56
Gatherer	29	53.70	31	57.41	32	47.06	111	48.26	38	52.05	21	53.85
Grazer	9	16.67	4	7.41	7	10.29	96	41.74	4	5.48	5	12.82
Predator	9	16.67	7	12.98	13	19.12	16	6.96	18	24.66	10	25.64
Shredder	4	7.41	2	3.70	11	16.18	4	1.74	9	12.33	2	5.13
Totals	54		46		68		230		73		39	

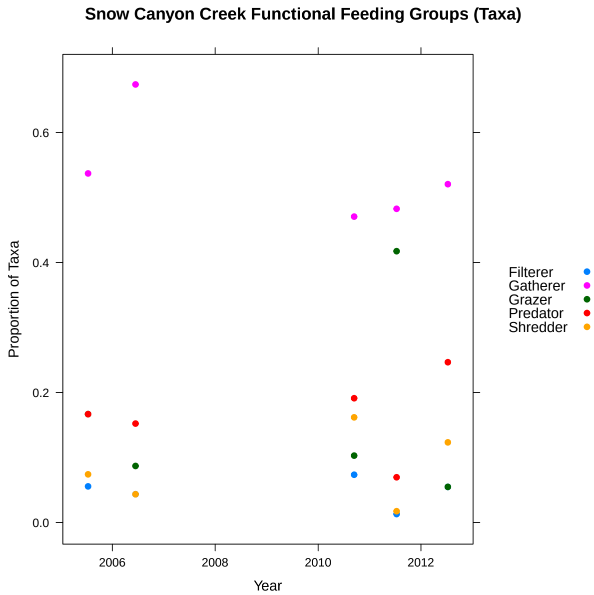

Figure 3.5: Proportions of individuals (left) and taxa (right) in each FFG in Snow Creek. There are no data from 2007–2009.

to-year in the 2010–2012 data than in the 2005–2006 data (Figure 3.5). The proportion of grazer individuals were approximately the same in 2010 and 2011 before sharply increasing in 2012 while the number of grazer taxa increased greatly 2010 to 2011 then decreased in 2012 to a level lower than in 2010. This indicates more individuals in each grazer taxon in the most recent data year.

3.3.2 Jerritt Canyon Creek

Data on benthic macroinvertebrates in Jerritt Canyon Creek are available for the years 2004–2006 and 2010–2013 (Table 3.6) . As with the other westside creeks the changes in proportions of taxa, particularly in the predominant collector/gatherer FFG, were less extreme in the three early years compared to the most recent four years (Figure 3.6).

3.3.3 Burns Creek

Burns Creek has had the largest fish collection efforts over the years. These data are presented in Table 3.7.

Burns Creek has benthic macroinvertebrate data from eight years: 2000, 2003, 2005, 2006, and 2010–2013 (Table3.8). Collector/gatherers are the dominant FFG in number of taxa (Figure 3.7). With seven data sets from this stream we can see the inherent natural variability in numbers of individuals and taxa within each functional feeding group.

3.3.4 Starvation Canyon Creek

Benthic macroinvertebrate data were obtained from Starvation Canyon Creek in 2006 and 2010–2013 (Table 3.9). The relative proportions of individuals and taxa in each FFG are presented in Figure 3.8. Whereas grazers are the second most common FFG for both numbers of individuals and taxa in other streams, shredders and collector/filterers are second in individual abundance in 2006, 2011, and 2012 while grazers consistently have the second highest proportion of taxa.

3.3.5 Water Pipe Creek

Fish were collected in Water Pipe Creek in 2006, 2010–2012. These data are presented in Table 3.10.

Table 3.6: Jerritt Canyon Creek: number and proportion of taxa in each functional feeding group.

FFG	2004		2005		2006		2010		2011		2012		2013	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
lterer	2	4.44	1	1.82	3	3.70	2	8.70	0	0	0	0	10	45.45
atherer	31	68.89	31	56.36	50	61.73	14	60.87	82	55.03	11	42.31	1	4.55
razer	1	2.22	5	9.09	6	7.41	2	8.70	14	9.40	2	7.69	0	0.00
redator	10	22.22	17	30.91	20	24.69	5	21.74	46	30.87	13	50.00	10	45.45
hredder	1	2.22	1	1.82	2	2.47	0	0	7	4.70	0	0	1	4.55
otals	45		55		81		23		149		26		22	

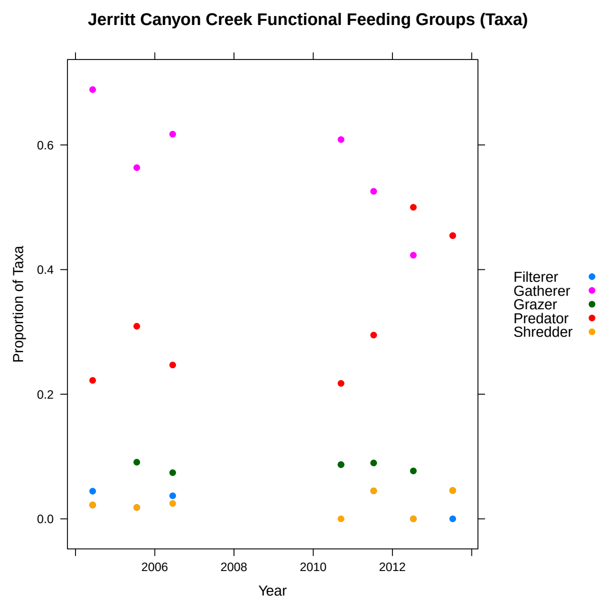

Figure 3.6: Relative proportions of each FFG in Jerritt Canyon Creek in the years 2004–2006 and 2010–2012. Collector/gathers dominate both the numbers of individuals and the numbers of taxa in both periods.

Table 3.7: Fish collected in Burns Creek.

Site	Date	Common Name	Count
B(W)	2000-06-08	Brook trout	16
B(W)	2000-06-08	Redband trout	9
B(W)	2003-06-09	Brook trout	16
B(W)	2003-06-09	Redband trout	4
B(W)	2006-06-28	Brook trout	8
B(W)	2010-09-14	Brook trout	30
B(W)	2010-09-14	Redband trout	16
B(W)	2011-07-15	Brook trout	43
B(W)	2011-07-15	Redband trout	6
B(W)	2012-07-11	Brook trout	79
B(W)	2012-07-11	Redband trout	10
B(W)	2013-07-11	Brook trout	83
B(W)	2013-07-11	Redband trout	18
B-02	2012-07-11	Brook trout	13
B-02	2012-07-11	Redband trout	1
B-02	2013-07-11	Brook trout	17
B-02	2013-07-11	Redband trout	2
BC-1	2000-06-08	Brook trout	18
BC-1	2000-06-08	Redband trout	6
BC-1	2003-06-09	Brook trout	5
BC-2	2000-06-08	Brook trout	17
BC-2	2000-06-08	Redband trout	6
BC-2	2003-06-09	Redband trout	1
BC-2	2010-09-14	Brook trout	9
BC-2	2011-07-13	Brook trout	18

Table 3.8: Burns Creek: number and percentage of taxa in each functional feeding group.

	2003		2005		2006		2010		2011		2012		2013	
Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
5.50	8	7.34	13	11.61	6	10.00	7	7.78	8	8.79	10	10.42	6	7.23
55.96	67	61.47	64	57.14	28	46.67	55	61.11	52	57.14	51	53.13	46	55.42
7.34	2	1.83	4	3.57	9	15.00	4	4.44	6	6.59	8	6.25	5	6.02
22.94	25	22.94	19	16.96	8	13.33	17	18.89	25	27.47	23	23.96	22	22.61
8.26	7	6.42	12	10.71	9	15.00	7	7.78	0	0	6	6.25	4	4.82
	109		112		60		90		91		96		83	

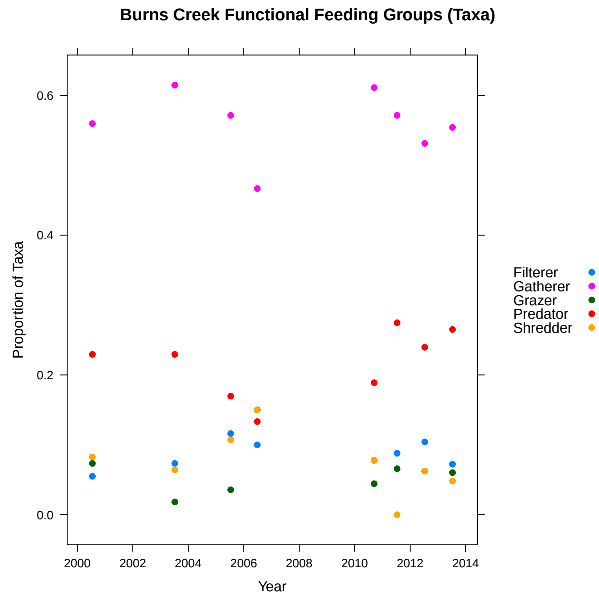

Figure 3.7: Burns Creek: proportions of taxa in each FFG.

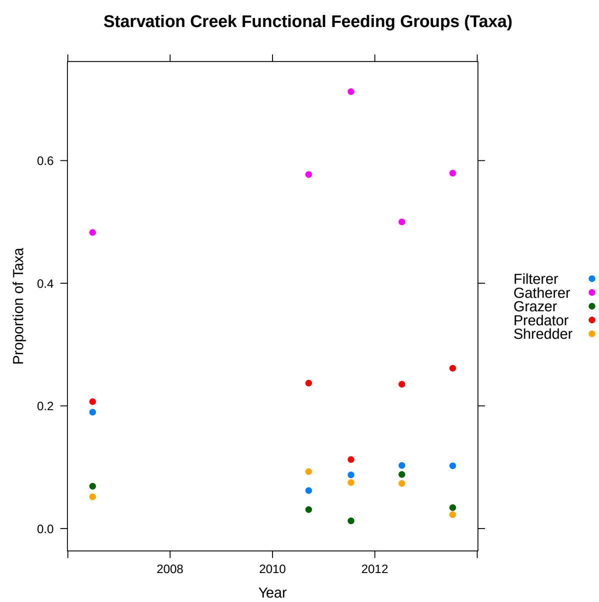

Figure 3.8: Starvation Creek: proportions of taxa in each FFG.

Table 3.9: Starvation Creek: number and percentage of taxa in each functional feeding group.

	2006		2010		2011		2012		2013	
FFG	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Filterer	11	18.97	6	6.19	7	9.75	7	10.29	9	10.23
Gatherer	28	48.28	58	57.73	57	71.25	34	50.00	51	57.95
Grazer	4	6.90	3	3.09	1	1.25	6	8.82	3	3.41
Predator	12	20.68	23	23.71	9	11.25	16	23.53	23	26.14
Shredder	3	5.17	9	9.28	6	7.50	5	7.35	2	2.27
Totals	58		97		80		68		88	

Table 3.10: Fish collections from Water Pipe Creek.

Site	Date	Common Name	Count
WP-1	2006-06-27	Redband trout	36
WP-1	2010-09-16	Redband trout	67
WP-1	2011-07-12	Redband trout	29
WP-1	2012-07-12	Redband trout	81
WP-2	2006-06-27	Redband trout	32
WP-2	2010-09-13	Redband trout	20
WP-2	2012-07-10	Redband trout	131

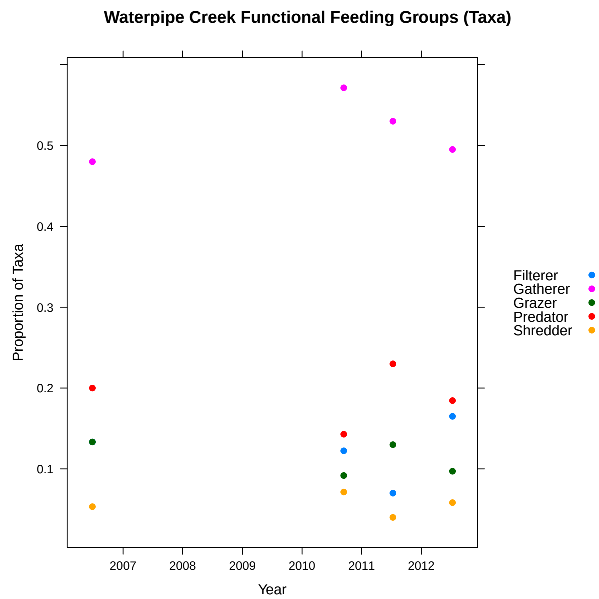

Figure 3.9: Waterpipe Creek: Proportions of taxa in each FFG.

Water Pipe Creek is immediately east of Starvation Canyon Creek at the southern end of the Independence Range so benthic macroinvertebrate data from this stream are from the same years as above; that is, 2006 and 2010–2013 (Table 3.11). The relative proportions of taxa in each FFG are presented in Figure 3.9).

Table 3.11: Water Pipe Creek: number and proportion of taxa in each functional feeding group.

	2006		2010		2011		2012		2013	
FFG	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Filterer	10	13.33	12	12.24	7	7.00	17	16.50	17	16.50
Gatherer	36	48.00	56	57.14	53	53.00	51	49.51	51	49.51
Grazer	10	13.33	9	9.18	13	13.00	10	9.71	10	9.71
Predator	15	20.00	14	14.29	23	23.00	19	18.45	19	18.45
Shredder	4	5.33	7	7.14	4	4.00	6	5.83	6	5.83
Totals	75		98		100		103		103	

Chapter 4

Analyses

4.1 Analytic approaches

4.1.1 Overview

The two categories of models applied to biotic environmental data, numeric indices and statistical, were developed by researchers in academia and government. The focus of these researchers was ecological: describing relationships between observed biota and explanatory variables. Ecological research begins with an idea expressed as a testable hypothesis. Data collection protocols are designed to provide abundant data from appropriate spatial and temporal intervals that meet the assumptions and requirements of the index or statistical model.

Environmental data analyses have a completely different foundation: data types, their collection locations, and collection frequencies have no underlying ecological or statistical basis. They are often arbitrary and frequently changed. This makes it more important that the analytical models be fit to the available data, not the other way around.

Indices, such as Ohio's Community Index (CI), the EPA's Rapid Bioassessment Protocol (RBP), the Index of Biological (Biotic) Integrity (IBI) developed at the University of Illinois, and the UK's RIVPACS are either adjusted to suit the system of interest or aim to compare a data set to a reference or expected data set. They also assume undesired anthropogenic impacts (pollution) and do not account for inherent natural variability.

Statistical models are developed by academic mathematicians, but are commonly modified to accommodate irregular environmental data. For example, time series models that do not require a constant data frequency, mixed effects regression models that accommodate non-numeric explanatory variables (such as month, river name, and other nominal variables), and compositional data models that work with components expressed as proportions or percentages (the situation with almost all environmental data).

In this monograph the units of analysis are the functional feeding group (FFG) proportions in each stream at each collection date. There are several justifications for using FFG proportions for analyses. FFGs are indicators of stream benthic macroinvertebrate community function and are less variable than are measures of benthic macroinvertebrate community structure represented by counts of individuals and mixed taxonomic levels. Biotic collections in compliance with permit conditions tend to be infrequent and do not provide the data density of research studies. Therefore, analyzing data using the family taxonomic level smooths much of the high frequency variability observed with measures based on lower taxonomic levels. Collecting biota in aquatic ecosystems is much less inclusive than are collections in terrestrial ecosystems. Fish are highly mobile and may be in a different stream reach than that being sampled. Benthic macroinvertebrates escape collection by being too small or too large for the mesh size of the collection net, by digging into the stream substrate to avoid capture, or by inhabiting a portion of the stream channel not subject to collection efforts. As a result of these uncertainties, it is not known just how representative the collections are of the actual populations present in the collection area. Using proportions of the five main FFGs acknowledges that there may be other, un-analyzed groups or uncollected individuals. The analyses are based on relative abundance rather than absolute numbers. This conservative approach is less likely to be successfully challenged when the analytical results are used to make regulatory or policy decisions.

The characteristics of each stream are summarized by descriptive analysis of each compositional set: compositional mean, metric standard deviation, and covariance displayed as ternary diagram matrices.

Multivariate regression models use each stream's compositional data set as the response variable and the available geomorphic, hydrologic, and chemical data as potential explanatory variables. This important step uses geologic, hydrologic, physical, spatial, and chemical values to explain the observed biological values and is critical for understanding ecosystem dynamics.

When data are available from multiple drainage basins and stream/river networks it might seem to be useful to apply a clustering model to the compositional data. This groups streams based on the similarity/dissimilarity of their biotic compositional data. Such grouping and classification is frequently useful for ecological researchers. However, regulators need to make decisions on specific stream or river reaches with reach-specific designated beneficial uses so the analytical process is applied to a single stream network in a drainage basin, independent of other basins and stream networks.

4.1.2 History

Early efforts to replace diversity and integrity indices with statistical models used mixed taxonomic levels with tools such as cluster analysis (e.g., Shepard 1984). During the past few decades multivariate and mixed effects ecological statistical models such as multiresponse permutation procedures (mrpp), nonmetric multidimensional scaling (NMDS), and ordination have joined clustering as potential tools for answering ecological and environmental questions using benthic macroinvertebrate data. While several models might be applied to any given data set careful thought must be given in advance to the specific questions to be answered.

Currently, two of the most appropriate statistical tools for environmental data analyses are quantile regression (for explaining observed and predicted response variable values with explanatory variable values) and compositional data analysis (CoDA). Both of these are exceptionally well suited for analyzing benthic macroinvertebrate collections and associating them with water quality and changes both natural and anthropogenic.

Better than methods based on community structure (taxonomy) are methods based on community functions of energy processing and nutrient spiraling. Functional feeding groups–first described and used by Cummins (1973) and subsequently by Cummins (1974), Cummins et al. (1989), and Cummins et al. (2005)–reflect these community functions. This makes them well suited to examining relationships between aquatic biota and water quality.

To measure ecosystem restoration efforts in Florida's Kissimmee River and its floodplain Merritt et al. (1996) evaluated assignment of benthic macroinvertebrates to functional groups based on the organisms' foods, habits, and other factors. They found FFGs useful in measuring the ecosystem attributes they considered. Faith (1990) evaluated functional feeding groups as summaries of benthic macroinvertebrate communities for the Upper La Trobe river system of Victoria, Australia. His research addressed two related questions: first, "what is the form of response of functional groups to environmental gradients?" and second, "are observed group patterns significant insofar as they are unlikely to arise in randomly defined groups of taxa?" His questions are directly applicable to use of benthic macroinvertebrate functional feeding groups for setting water quality standards. Uwadiae (2010) assessed benthic macroinvertebrate FFGs to assess environmental conditions in the Epe lagoon in southeast Nigeria. His study applied cluster analysis and correlated FFGs with total dissolved solids (TDS) and sediment composition (organic matter, sand, and mud). Uwaidae's conclusions are that functional feeding group ratios can detect changes in organic matter processing and be a surrogate for water quality.

The seven drainage basins described in this monograph differ from the reports cited above by the extent of their geographic focus. In each of the above the focus was a large basin and the river network draining it: Kissimmee River (Florida) channels and floodplains; Upper La Trobe river system (Victoria, Australia); and a portion of the Epe lagoon in south-west Nigeria. Data in this monograph are from seven comparatively small drainages and stream networks (Figure 2.1) and each is a separate unit for regulatory decision-making. Analyzing individual drainage basins provides regulators with a consistent process that can be used to set water quality standards anywhere.

Functional feeding group analyses for the Independence Mountains streams do not include the categories of omnivore and parasite because they are not reported from all streams at all collection dates and these two groups account for fewer than 1% of the individuals reported. Aquatic ecological research articles do not normally include omnivores or parasites and there is no additional insight into Independence Mountain stream ecosystem functions by including them here.

4.2 Location/time and taxonomic analyses

Community ecology research is designed to answer questions about either observations (locations or times) or variables (taxonomic data at define levels). Differences among observations are called Q analyses. Differences among variables are called R analyses. For the purposes of setting water quality standards and assessing compliance, each drainage basin and its stream network should be considered as a unit. This approach allows scaling from a small basin such as Starvation Creek to much larger ones such as the mainstem Humboldt River. While the specific taxa collected vary by collection date, functional feeding groups will be more consistent because they reflect ecological function rather than structure. Functional feeding groups proportions will change along the length of a stream (or a larger river such as the Humboldt or Truckee Rivers as channel characteristics and food resources change (Vannote et al. 1980). Support of regulatory decisions is provided by statistical models of the variability in benthic macroinvertebrate communities. In small drainage basins the focus should be on overall stream network collections (that is, how variable is the community over time). For large drainage basins (e.g., the main Humboldt River) the focus could be variability over time at specific locations or variability of locations along the river at the same time.

4.3 Analytical purpose

The purpose of environmental data analyses is to support regulatory decisions. Regulators need robust and technically correct analyses of baseline data for environmental impact assessments, operational permit issuance, and compliance with water quality discharge permit conditions. Setting defensible water quality standards requires identifying the extent of inherent natural variability, recognizing when variability exceeds that range, and determining whether those exceedences result from the regulated activity. This is especially important when the designated beneficial use is that of aquatic life, also expressed as fishable and swimable. There are two general considerations about environmental data and its analyses and interpretation that should be understood: the differences from ecological research data and how to understand complex, multivariable data.

4.3.1 Ecological vs. environmental data

To make informed regulatory decisions it is necessary to understand differences between ecological and environmental data. Analyses of environmental data historically used models developed by numerical ecologists for ecological data collected by academic and research agency scientists. These numeric and statistic models require well-structured data collected to fit assumptions and requirements of the models. This works for researchers who identify a question to be answered and work forward from that to determine when, where, and how much data need collecting to answer that question. The research approach of fitting data to models has leaked into the analyses of environmental data gathered in response to statutory and regulatory requirements. Most often, the results are mis-leading or incorrect. Regulatory decisions based on these results are ineffective at best or economically and socially harmful at worst.

Environmental data are messy and unstructured, collected to support environmental permit applications and monitor compliance with permit conditions. Locations change over time, data collection frequency is irregular, and chemical or biological data elements can cease being collected and re-instated at a future time. Such data cannot be fit to research models such as species diversity, indices of biotic integrity (IBI) or community (CI), predictive models based on expected taxa (RIVPACS), hydroelectric fish passage models (CRiSP), or pit lake water quality (PITLAKQ). For real-world environmental regulatory decision-making it is necessary to fit the model to the data.

It is difficult (or impossible) to get reliable, consistent, generally applicable analytical results of environmental data from numeric models. Therefore, an appropriate statistical model is used. There is such a large choice of statistical models (the R project alone has over 6,000 application-specific model packages for analyzing data of every type) that one appropriate for regulatory decisions based on environmental data can be identified and used to produce technically sound and legally defensible results.

It is common to read a report submitted to regulators comparing sets of water chemistry data using analysis of variance (determining if they are similar because they come from the same population) when the regulator wants to know whether the permitted operation has an undesired negative effect on water quality; that is, why the measured concentrations have the values they do. Providing an answer that does not answer the regulator's question can have severe consequences for the permit holder.

In practice, one of the largest differences between analyzing ecological and environmental data is that many of the most appropriate models for the latter are relatively unknown or recently developed. Among these statistical models are those for quantile regression and compositional data analysis. Quantile regression measures the relationships of explanatory variables on different portions of the observed range of the response variable (not just the mean response variable as is the case with linear regression). Compositional data analysis analyzes parts of a whole; for example, some chemical constituents in a medium with many chemicals or functional feeding groups of benthic macroinvertebrates.

4.3.2 Decision-making using complex, multivariate data

All ecosystems are complex: natural, economic, and societal. Governments report a range of official statistics (e.g., unemployment, cost of living, credit interest, purchasing parity, stock market indices) that represent a broad range of explanatory (predictive) variables that measure the economy's status and trends. We do not see a single number claiming to summarize this complexity, and would likely be puzzled on how to interpret that simplification of such a complex system. For some reason, however, too often simplifications of highly complex natural ecosystems into a single number (diversity or integrity index) is accepted as having interpretive value. Regulatory and policy decisions are made based on these single values without robust justification that they are meaningful in an environmental, not ecological, context. The better approach is to adopt mathematically sound statistical analytical models that are fit to the range of geographic, physical, chemical, and biological data that explain and predict the response variables in which we are interested.

With apologies for using a worn cliché, an analytical paradigm shift needs to be adopted by regulators and policy makers. A single number cannot effectively describe the complexities found in environmental data, therefore, multivariate statistics must be applied to bridge the gap between environmental data and regulatory decision-making.

Evaluating variability of benthic macroinvertebrate functional feeding groups with the effects on these compositions of geographic, hydrologic, and water chemical concentrations reflects the dynamics of the drainage basin and stream network. This insight is compared to knowledge of specific designated beneficial uses to determine if those uses are impaired. Such impairment is generally obvious: fish kills, wildlife or livestock morbidity or mortality, crop loss or stunted growth. Decisions based on multivariate data incorporate environmental, economic, and societal values and are sensitive to changes that justify operational adjustments.

4.4 Water quality standards

Fishable and swimable can be defined by the presence of fish safe for human consumption and the lack of pathogenic bacteria or parasites in the water column. Wildlife and cattle use can be defined like fishable and swimable waters, and irrigation can be defined by crop yield and growth rates. The more general term, aquatic life, has no consensus definition; the definition used in this document is the range of natural variability in the relative proportions of taxa in each functional feeding group. This range of variability is expected to differ by basin, particularly between those draining the western slopes of the Independence Mountains and those draining the eastern slopes. A combination of geomorphic, hydrologic, and chemical explanatory variables are analyzed for their contribution to the observed relative proportions of functional feeding groups.

The two specific questions the data analyses are to answer are,

1. What is the normal range of variability of functional feeding group proportions in a stream network based on available data?
2. What geomorphic, hydrologic, and chemical variables explain the observed patterns?

When analyses yield results beyond the normal range it is necessary to examine the individual explanatory variables to determine which have changed and to identify the reason for that change.

4.5 Functional feeding groups

4.5.1 Characterizing communities

Two useful questions for assessing biological communities germane to water quality status and change are:

1. What are the differences from one collection time to the next?
2. What environmental variables explain the observed patterns of biological community function?

There are many measures of association (commonly called similarity measures or coefficients) that might be applied to biological communities[1]. For the purpose of setting water quality standards and detecting changes that adversely impact the designated beneficial uses, changes from one collection time to the next, and over all collections, limits use to certain association measures. Since only five functional feeding groups are commonly used, differences in relative proportions among these groups, and the environmental explanatory variables that affect them, can be used to set water quality standards that are specific to each stream ecosystem and defined designated beneficial uses.

The composition of the biotic macroinvertebrate data available for analysis represent only a portion of those taxa in the stream system. As previously explained, the collections are not samples in the statistical sense of accurately reflecting the number of taxa and individuals in the entire population. Therefore, the statistical models often applied to ecological data for plants and vertebrates yield incorrect results when used on benthic macroinvertebrate data. Correct results are provided by statistical models based on compositions as explained in Section 4.5.3.

These compositions have values in the closed interval [0, 1][2]. These proportional values are used to measure distances between observed sets of functional feeding groups; almost all statistical models can be applied to these compositional data.

4.5.2 Proportions of taxa in each functional feeding group

The numbers of taxa collected will always vary because of both natural and anthropogenic causes. Not all possible feeding strategies are represented because some, such as parasites and omnivores, are both relatively rare and infrequently identified and counted. Relative proportions of taxa in each major functional feeding groups are less variable than the counts themselves and contain all the important information about the biotic communities' function.

Because functional feeding groups reflect energy processing and nutrient spiraling along the stream network the ecological interpretation based on proportions of taxa is relatively straight-forward. The numbers (and proportions) of individuals at any given collection event will vary according to their life cycles, but the proportion of taxa will be more constant because FFGs reflect the range of available niches at the collection location. That is, not all individuals actually present on the sampled substrata will be collected, and those collected cannot be assumed to represent their proportion in the entire benthic macroinvertebrate population. But, all habitats will have representatives filtering, gathering, shredding, and scraping (grazing) of fine particulate organic matter and there will be predators feeding upon the other organisms.

4.5.3 Compositional data analysis

Ecological distances, measured by Bray-Curtis dissimilarity coefficients (e.g., Irvine et al. 2011; Reiss et al. 2010) can be useful in comparing biological communities by space and time when organisms can be reliably identified to species and the collected data are samples representing the proportions of species in the entire population. These coefficients are distinct from statistical distances which are applied to CoDA data sets of functional feeding groups.

The concept of compositional data analysis originates from work by Ferrers in 1866. In 1879, the British mathematician and biometrician Karl Pearson discussed the complexity of its theoretical properties and indicated that in the practice of compositional data analysis, the "sum to unity" constraint was consciously or unconsciously ignored. Some traditional statistical methods designed for "unconstrained data" (i.e., those not limited to values between 0.0 and 1.0) were often misused and that led to seriously incorrect results and poor economic or financial decisions.

The first systematic research on compositional data was given by John Aitchison in his 1986 book, *The Statistical Analysis of Compositional Data*, in which he describes studies on the logistic normal distribution and log-ratio transformations of compositional data. In addition to economic and other "official" statistics released by governments, CoDA development was quickly adopted by mathematical geochemists to examine the chemical composition of

[1] Those interested in the mathematical and ecological differences are referred to Chapter 7 in Legendre and Legendre (1998).

[2] A closed interval includes the end points; an open interval includes values other than than the end points.

rocks (including ores) and soils. Much more recently, it has been applied to pollution analyses and biological/ecological data.

CoDA is applied to data sets consisting of counts, proportions, percentages, and concentrations that are "closed"; that is, each row in the data set has the same sum of 1.0 or 100. The information that supports decisions is contained in the ratios of the data set's components, not the size of the sample or collection. When water is analyzed for toxic metals, other minerals, or organic compounds in the volume of water collected is immaterial as all concentrations are scaled to mg/L or $\mu g/L$. Benthic macroinvertebrate functional feeding groups have meaning in the proportion of each group (reflecting the organic food resources available), not in the total number of organisms collected. The mathematical foundations of statistical compositional data analysis are complex and still developing (e.g., Aitchison 1994; Filzmoser and Hron 2010) yet the tools have been effectively applied to ecological and environmental community studies (de Valpine and Harmon-Threatt 2013; Jackson 1997; Johnson et al. 2006; Pendleton et al. 1998).

Much real world data (environmental, economic, political, psychological) have their important meaning in the ratios of the components, not the individual components. Environmental data, while often considered as individual chemical constituents or single species, are actually only portions of the whole. There are always chemical constituents and biota not measured or observed. Counts of biological data, such as the number of taxa in functional feeding groups of benthic macroinvertebrates, vary spatially and temporally making classical statistical analyses inappropriate because they are based on Euclidean[3] distances. For example, the difference between 0.1 and 0.2 is the same as the difference between 0.5 and 0.6; 0.1 units. However, the ratio of those differences are not the same: 0.2 is 2 times 0.1 while 0.6 is only 1.2 times 0.5. When data represent portions of the whole it is the ratios between them that contains all the important (and useful) information, not the raw values themselves.

Because compositional data take various forms there are multiple options for converting the raw data from Euclidean geometry to the Aitchison simplex geometry[4] (Aitchison 1994) upon which statistical compositional data analyses are performed. Both continuous data (such as the components of TDS) and count data (aquatic biota) can be analyzed. Almost all "classical" statistical models can be applied to compositional data, including regression (and quantile regression), principal components analysis (PCA), time series, and clustering. Because of CoDA's robustness and appropriateness for relating aquatic biota to water quality designated beneficial uses, these statistical models will be applied to the five drainage basins (for which there are multiyear data) of the southern Independence Mountains as a template for the process regulators can apply to water quality oversight in all drainage basins.

Because compositional data include selected components from all possible components of a chemical or biological system the standard descriptions of their center (mean) and spread (variance and its square root, standard deviation) are misleading. These data are analyzed as the ratios of their parts, specifically, the natural logarithm of the ratio. For environmental data the isometric log-ratio (ilr) preserves the geometric relations of the components and is used to calculate the mean value of each component (called the Aitchison mean) in such a way that it allows comparison of different data sets; the Aitchison composition mean is equivalent to the geometric mean of the raw data. Because variance of compositional data is difficult to interpret, particularly in the ecological contexts appropriate for environmental data, the metric standard deviation (msd) is used as a measure of the distance of log-ratio transformed data points from the Aitchison mean, back-transformed to the original units (proportions of each functional feeding group in this monograph).

Most water pollution control permits, including those issued under the federal NPDES program, require compliance monitoring based on water chemistry, not aquatic biota. The southern Independence Mountains data is one of the more extensive biological sets and provides a base upon which future data can be added. As more data become available and are analyzed using these protocols the interpretations will become more refined and the quality of operational and regulatory decisions will increase. The results will benefit industries, regulators, and society.

4.5.4 Cause and effect: explaining observations

Prior to determining whether a permitted activity has adverse impacts on a specific designated beneficial use it is necessary to evaluate inherent natural variability. Regression analyses is the class of statistical models that

[3]The straight-line distance between two points on a dimensional surface such as a map.

[4]In geometry a simplex is the generalization of a triangle. Compositional data is viewed in a ternary plot; a triangle where each point represents a component.

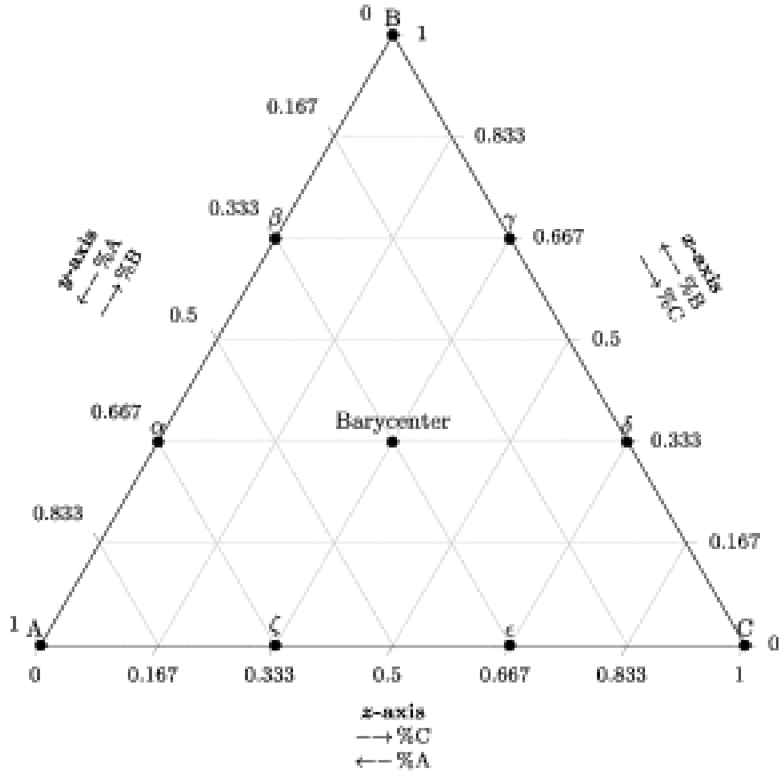

Figure 4.1: Example of a ternary diagram. The barycenter is the center of mass of all components.

measure the relationship between the mean value of a response variable and the measured range of explanatory variables. Regression models may be linear or non-linear, univariate (a single explanatory variable) or multivariate (many explanatory variables). Compositional data, which are multivariate themselves, may be either a response or an explanatory variable. For the purpose of setting water quality standards and evaluating compliance with them using benthic macroinvertebrate FFGs, they are the response variable in regression models. Regression equations can be used to forecast future values of the response variable assuming that the relationship to the explanatory variables remains constant. Time series analyses are also usefully applied to compositional data.

To demonstrate this analytical approach three chemical parameters[5]: nitrate nitrogen (NO_3^-), sulfate (SO_4^{--}), and hydrogen ion concentration (pH) are the explanatory variables. When more years of data are available for analyses basin parameters can be included as explanatory variables. These are all continuous variables; nominal or categorical variables can also be used in the model. Because chemical data were not provided for 2012 and 2013 the median value is used for these two years.

Several types of data plots (figures) can be used to describe and understand compositional data and their relationships to explanatory variables. Of these, ternary diagrams and form biplots illustrate data distributions useful for regulatory decision making. These are not commonly encountered in regulatory compliance reports and are explained here.

4.5.5 Ternary diagrams

Ternary diagrams are used to plot the position of a point based on three (or more) variables in 2-dimensional space, such as a printed page. They are common in geology and geochemistry but much less common in displaying environmental data. Compositional data components are proportions of a constant sum (1.0 or 100) and ternary diagrams visually display the relationships of the proportions among the parts.

A ternary diagram is shown in Figure 4.1. The component's label is on the corner where its value is 1.0, and the position of a point represents the proportional contribution of each component relative to those corners.

[5]The value in mg/L for the sampling date that is close to the summer biotic collection date.

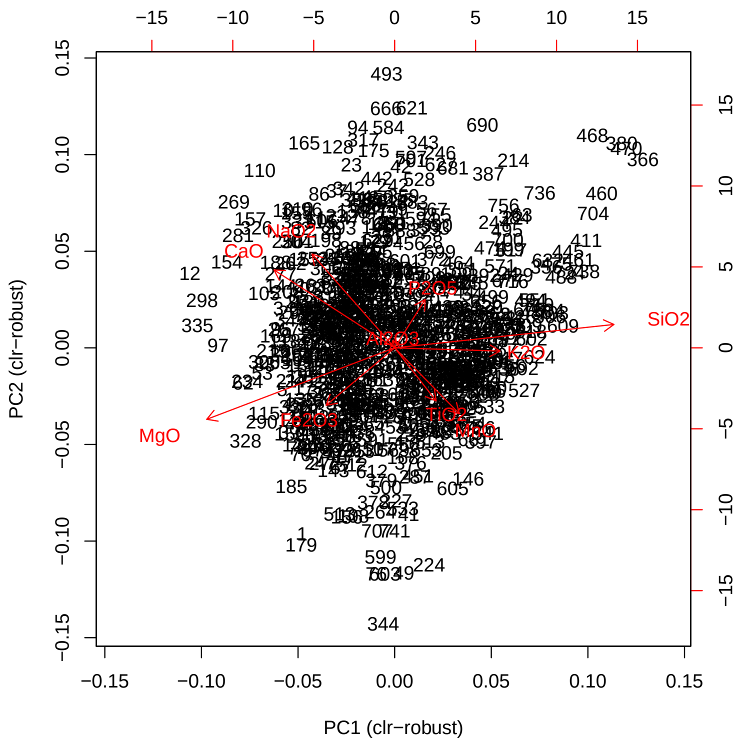

Figure 4.2: Covariance biplot of selected topsoil chemicals (arrows) with proportions at many sampling sites (numbers). Data are not from the five Independence Mountains streams and are in an unpublished manuscript.

Because a ternary diagram can display only three components (one at each corner) a matrix of ternary diagrams is necessary when a composition contains more components. These ternary diagram matrices are used to display the patterns of covariance among the five functional feeding group taxa in each of the five Independence Mountains streams. These diagrams have 95% confidence intervals shown in red on each plot. Interpretation of the ternary plots for each stream is presented in Section 4.6.

4.5.6 Biplots

Biplots are visual displays of tabular data that present multiple variables and multiple observations simultaneously and in relation to one another. The biplot simplifies the complexity by reducing this large number of variables to two principal components (the process is called Principal Components Analysis, PCA) and displaying the position of each observation as a point and each category as a vector in the cloud of data (Figure 4.2). The covariance biplot is the version appropriate for summarizing benthic macroinvertebrate functional feeding group proportions over the range of observation (years) and can be used when sufficient data are available.

The vector lengths are proportional to the variance of each variable, and the cosine of the angle between any two arrows reflects their correlation corefficient. Two uncorrelated variables have orthogonal (i.e., right angle) vectors; three or more variables on a common line have links of 0°or 180°and are, therefore, perfectly correlated (positively or negatively, respectively). The positions of the numbers represent data observations relative to the two principal components (bottom and left axes) and their distance from each other.

These patterns describe the simultaneous distributions of observations (table rows) and variables (table columns) by reducing that complexity to two dimensions. Evaluating the patterns as more observations are made reveals their variability over time.

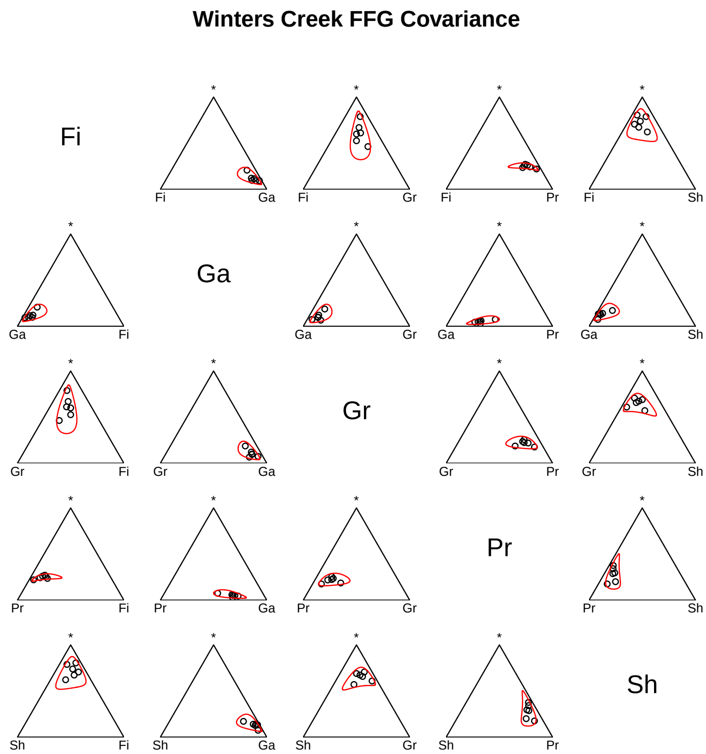

Figure 4.3: Winters Creek: functional feeding group covariance.

Table 4.1: Winters Creek: number of taxa in each functional feeding group by year.

Date	Filterer	Gatherer	Grazer	Predator	Shredder
2004	3	27	3	11	1
2005	3	28	3	13	2
2006	3	43	7	15	1
2011	4	54	6	24	3
2012	3	26	4	22	5
2013	1	39	2	18	2

4.6 Summary statistics for stream benthos

4.6.1 Winters Creek

The functional feeding group counts for the 6 years in the data set for Winters Creek is presented in Table 4.1. The Aitchison mean proportion (Table 4.2) indicates that gatherers comprise more than half all taxa over the range of years in the data set and predator taxa are almost one-third of the total. The metric standard deviation (msd) for each row in the data set is 0.3971 which indicates moderate variability.

4.6.1.1 Ternary diagram

Figure 4.3 displays the pair-wise covariance of the functional feeding group proportions over the period of collection. The dominance of gatherers is obvious along the second row; all data cluster close to the gatherer corner and the 95% confidence intervals (the red ellipses) are relatively small. There are no outliers (points outside the confidence limit). The covariance of filterers and grazers, however, is dominated by the other three functional feeding groups because grazers (who scrape organic materials off substates such as gravels and wood) always are a small component of the FFGs. A useful test of change is to recalculate this ternary diagram matrix after each

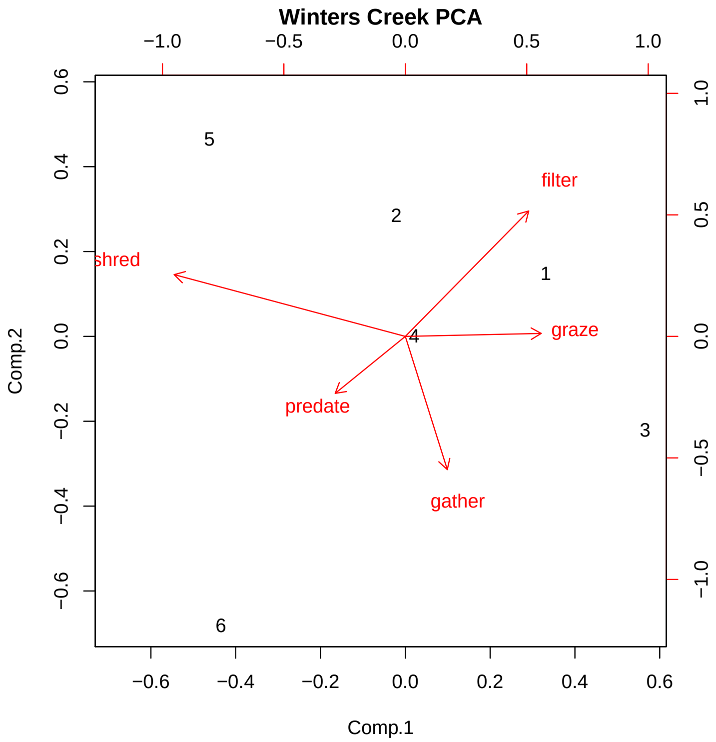

Figure 4.4: Winters Creek: covariance biplot.

Table 4.2: Winters Creek: Aitchison mean proportions of functional feeding groups for all years.

Filterer	Gatherer	Grazer	Predator	Shredder
0.04387	0.5827	0.0636	0.2766	0.0331

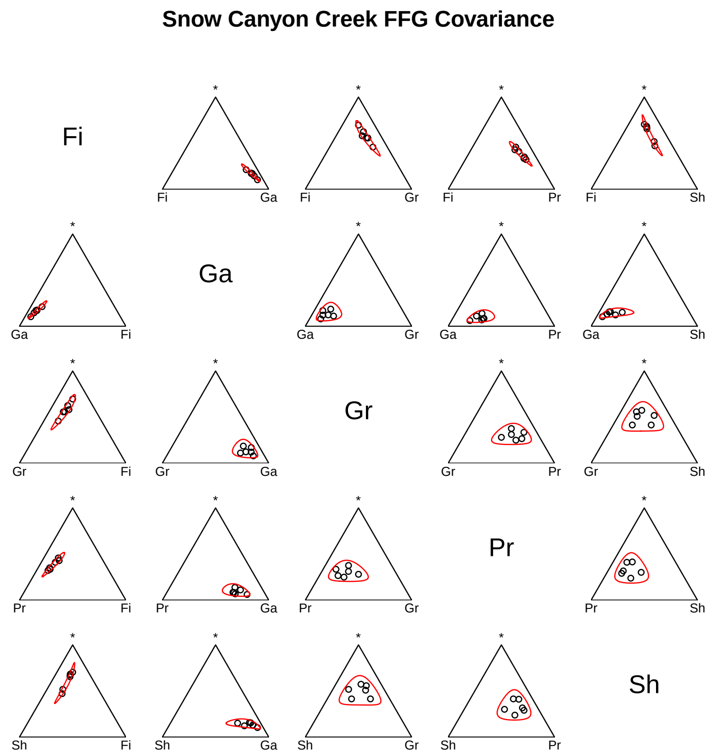

Figure 4.5: Snow Canyon Creek: functional feeding groups covariance.

year's collection is processed; expansion of the 95% confidence interval ellipsis, or outlying points, indicate that something has changed and the explanatory variables need to be examined to determine why.

4.6.1.2 Covariance biplot

The Winters Creek covariance biplot (Figure 4.4) displays more variability in the number of shredder taxa than any other functional feeding group. The numbers of filterer and predator taxa are negatively correlated (the vectors for those two variables increase in opposite directions). Shredder taxa numbers are negatively correlated with the other four groups with the strongest negative correlation to grazers. There is less variance in the number of predator taxa than in any other functional feeding group. The length of each vector (the line with the arrow head on the end) approximates the standard deviation of values for that variable.

There is little year-to-year similarity among the 6 years of data. However, along the first principal component (the x-axis), years 5 and 6 are similar and years 2 and 4 are similar. Along the second principal component (the y-axis) the observations for each year are spread across the range of values. In general, the proportions of taxa in each functional feeding group form two clusters along the first principal component (years 5 and 6 in one cluster and years 1 through 4 in the other cluster. Along the second principal component years 1, 2, and 5 have positive values, years 3 and 6 have negative values, and year 4 has a value approximately zero.

4.6.2 Snow Canyon Creek

The functional feeding group counts for the 6 years in the data set for Snow Canyon Creek are presented in Table 4.3. The Aitchison mean proportions indicate gatherers comprise more than half all taxa over the range of years and predators account for one-fifth of the total (Table 4.4). The metric standard deviation (msd) for the set of proportions is 0.3130 which indicates moderate variability.

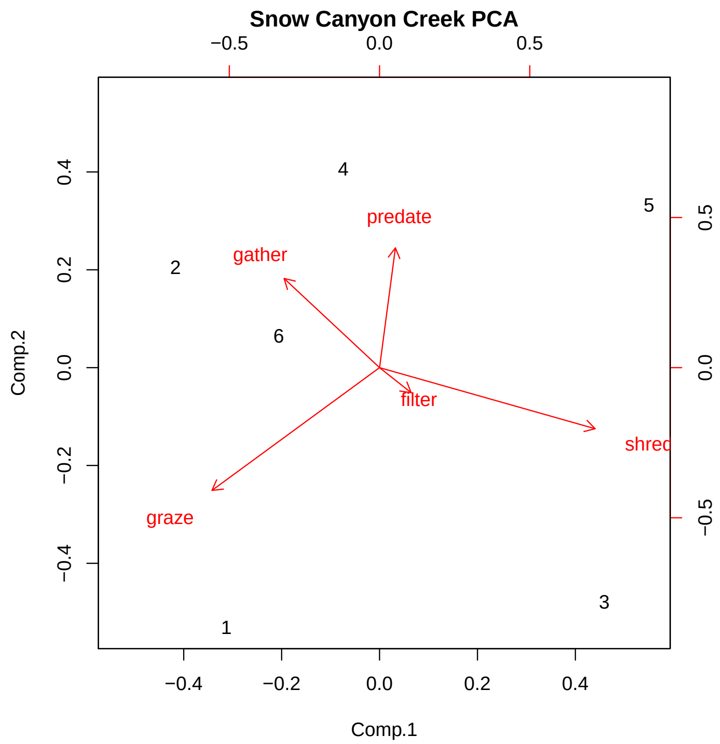

Figure 4.6: Snow Canyon Creek covariance biplot.

Table 4.3: Snow Canyon Creek: number of taxa in each functional feeding group by year.

Date	Filterer	Gatherer	Grazer	Predator	Shredder
2005	3	29	9	9	4
2006	2	31	4	7	2
2010	5	32	7	13	11
2011	3	37	5	16	4
2012	4	38	4	18	9
2013	3	33	7	15	4

Table 4.4: Snow Canyon Creek: Aitchison mean proportions of functional feeding groups for all years.

Filterer	Gatherer	Grazer	Predator	Shredder
0.0540	0.5597	0.0966	0.2082	0.0814

Table 4.5: Jerritt Canyon Creek: number of taxa in each functional feeding group by year.

Date	Filterer	Gatherer	Grazer	Predator	Shredder
2004	2	31	1	10	1
2005	1	31	5	17	1
2006	3	50	6	20	2
2010	3	14	2	5	0
2011	7	83	14	46	7
2012	0	11	2	13	0
2013	0	10	1	10	1

Table 4.6: Jerritt Canyon Creek: Aitchison mean proportions of functional feeding groups for all years.

Filterer	Gatherer	Grazer	Predator	Shredder
0.0293	0.5674	0.0656	0.3164	0.0213

4.6.2.1 Ternary diagram

Figure 4.5 displays the pair-wise covariance of the functional feeding group proportions. Some pairs have a narrow range within their 95% confidence limits (e.g., gatherere and grazers, gatherers and predators) while other pairs have much larger 95% confidence limits (grazers and others except gatherers).

4.6.2.2 Covariance biplot

The Snow Canyon Creek covariance biplot (Figure 4.6) displays slightly greater variability in the proportion of grazer taxa than of shredder taxa; there is very low variability in the proportions of filterer taxa. Gatherers and filterers are highly negatively correlated, grazers are uncorrelated with the former two groups, and predators and shredders are somewhat correlated with the other three functional feeding groups.

Among the yearly observations of proportions of taxa in each functional feeding group, years 1, 2, 6, and 4 form one cluster along the first principal component with years 3 and 5 forming a second, close cluster along that same axis. Along the second principal component axis, years 1 and 3 are very similar and years 2, 6, 4, and 5 cluster in a separate group with greater inter-year distances than seen in years 1 and 3.

4.6.3 Jerritt Canyon Creek

Tables 4.5 and 4.6 show functional feeding group counts and Aitchison mean proportions. Gatherers comprise more than half of all taxa and predators are about one-third of all taxa. The high variability of these data are seen in the metric standard deviation (msd) of 0.5729. This high value might reflect the missing years of data; variability could decrease if data collections continue.

4.6.3.1 Ternary diagram

Figure 4.7 displays the pair-wise covariance of the functional feeding group proportions. The second row demonstrates the dominance of gatherers and the small 95% confidence limits (the red ellipses). Other pairs (e.g., filterer-grazer and grazer-predator) have larger covariances with broad 95% confidence limits.

4.6.3.2 Covariance biplot

The display of all observations and variables in the Jerritt Canyon Creek covariance biplot (Figure 4.8) reveals that the gatherer and filterer functional feeding groups are almost completely positively correlated, and the standard deviation of the former group is negligible compared with the other four groups. Shredders and filterers have the greatest standard deviation to approximately the same degree. Predators are strongly negatively correlated with gatherers and filterers and moderately positively correlated with shredders and grazers.

There is very low similarity among each year's observations. No distinct clusters are seen along either principal component axis. This high variability among observations and variables suggests that Jerritt Canyon Creek is functionally highly dynamic when measured by proportions of benthic macroinvertebrates in each FFG.

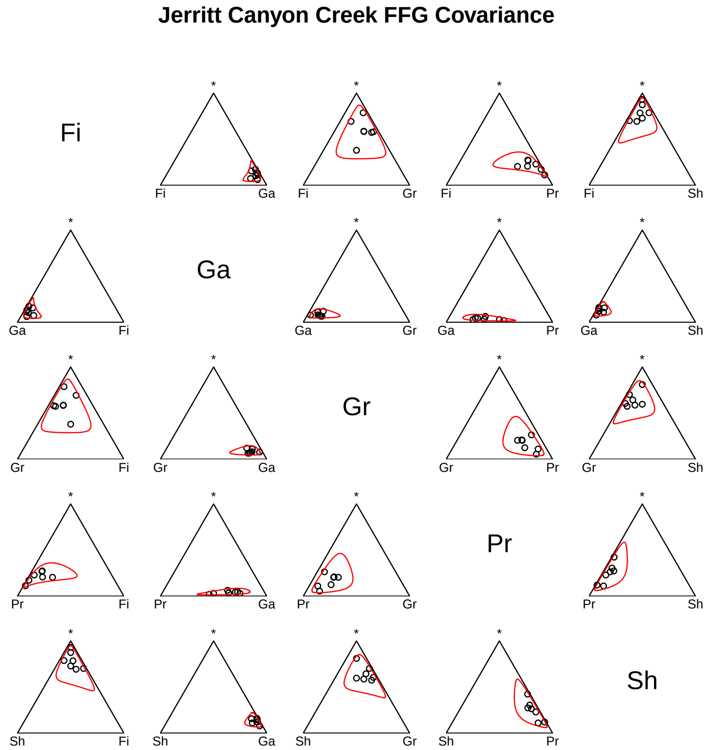

Figure 4.7: Jerritt Canyon Creek: functional feeding groups covariance.

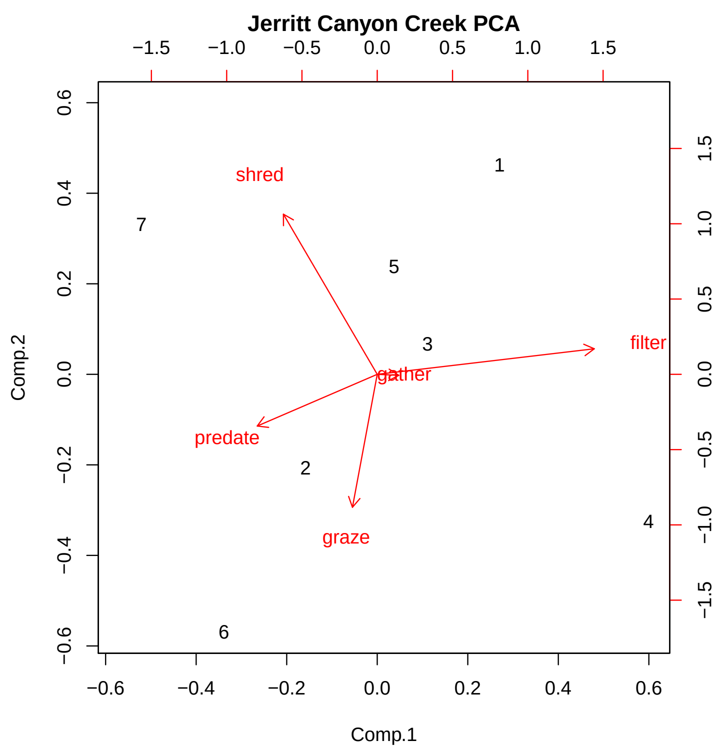

Figure 4.8: Jerritt Canyon Creek covariance biplot.

Table 4.7: Burns Creek: number of taxa in each functional feeding group by year.

Date	Filterer	Gatherer	Grazers	Predator	Shredder
2000	6	61	8	25	9
2003	8	67	2	25	7
2005	64	4	19	12	6
2006	28	9	8	9	7
2010	7	55	4	17	7
2011	8	51	6	25	0
2012	10	21	6	23	6
2013	6	46	5	22	4

Table 4.8: Burns Creek: Aitchison mean proportions of functional feeding groups for all years.

Filterer	Gatherer	Grazer	Predator	Shredder
0.0867	0.5783	0.0566	0.2185	0.0598

4.6.4 Burns Creek

The functional feeding group counts for the 8 years in the data set for Burns Creek is presented in Table 4.7. The mean proportions reveal that gatherers comprise more than half all taxa over the range of years in the data set and predator taxa are about one-fifth of all taxa. (Table 4.8). The metric standard deviation (msd) for the set of proportions is 0.5488 which is high variability.

4.6.4.1 Ternary diagram

Figure 4.9 displays the pair-wise covariance of the functional feeding group proportions over the period of collection. Like the previous streams, gatherers are a much higher percentage of taxa than are other groups and the pairs have a very narrow 95% confidence limits. Other pairs have greater ranges of variance and much broader 95% confidence limits. This may reflect high variability in the availability of food resources such as large particulate organic matter (LPOM for shredders), epilithic (attached) organic matter for grazers, and a variable number of predatory taxa.

4.6.4.2 Covariance biplot

The biplot of PCA of observations and variables of the Burns Creek data is shown in Figure 4.10. The filterer and predator functional feeding groups are completely positively correlated, and the gather FFG is very strongly positively correlated with the former two groups. Grazers and shredders are both slightly negatively correlated with the first three groups, and essentially uncorrelated with each other. Shredders are the most variable, grazers slightly less variable, and filterers the least variable.

Along the first principal component axes, year 6 is quite different from the rest of the years which are in a narrow cluster. Along the second principal component axis, years 2 and 4 are the most different and the other years are narrowly clustered near the center of the range.

Both observations (yearly proportions of taxa in each functional feeding group) and variables (the functional feeding group names) display quite different patterns from the stream systems to the north, on both sides of the Independence Mountains.

4.6.5 Starvation Canyon Creek

The functional feeding group counts for the 5 years in the data set for Starvation Canyon Creek is presented in Table 4.9. The mean proportions reveal that gatherers comprise almost 60% of all taxa over the range of years in the data set and predator taxa are about 21% of all taxa. (Table 4.10). The metric standard deviation (msd) for the set of proportions is 0.4921 which indicates high variability.

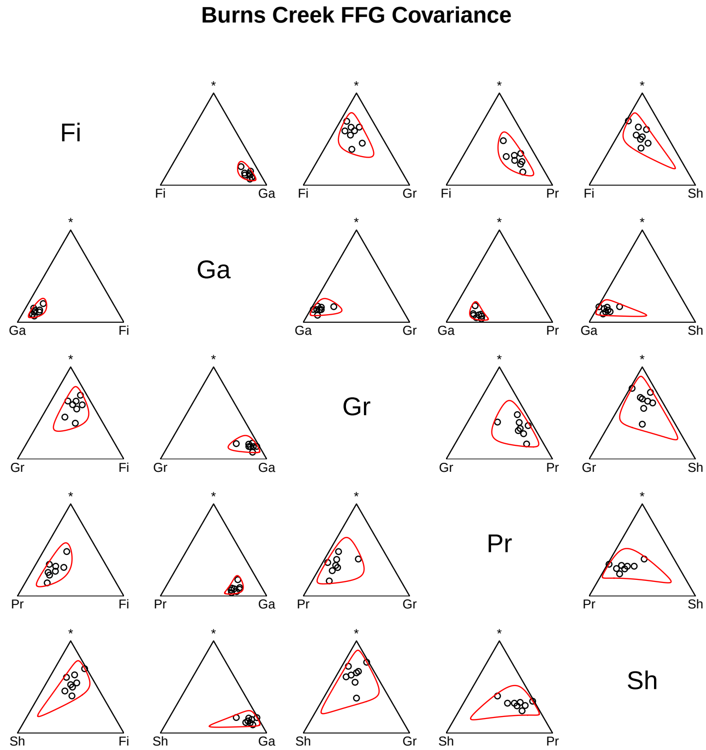

Figure 4.9: Burns Creek: functional feeding groups covariance.

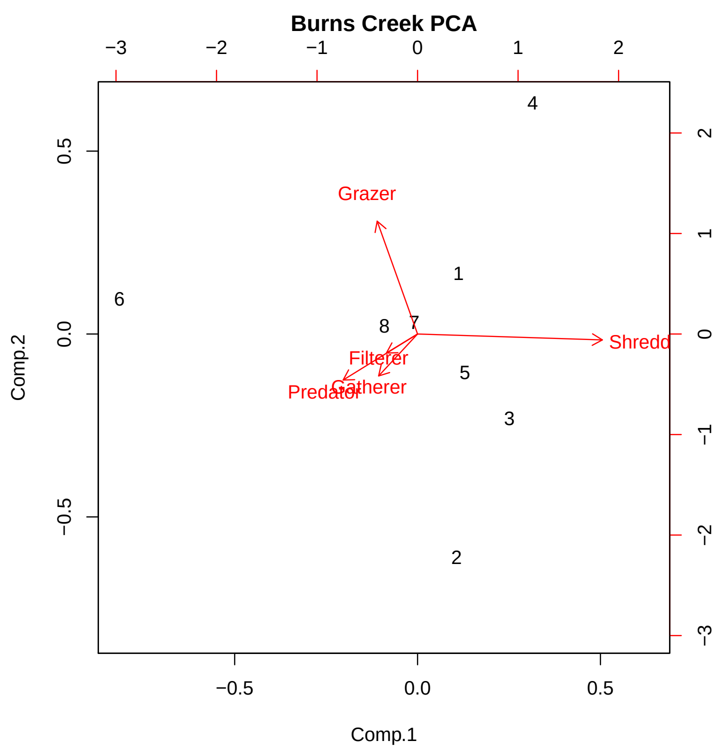

Figure 4.10: Burns Creek covariance biplot.

Table 4.9: Starvation Canyon Creek: number of taxa in each functional feeding group by year.

Date	Filterer	Gatherer	Grazer	Predator	Shredder
2006	11	28	4	12	3
2010	6	56	3	23	9
2011	7	57	1	9	6
2012	7	34	6	16	5
2013	9	51	3	23	2

Table 4.10: Starvation Canyon Creek: Aitchison mean proportions of functional feeding groups for all years.

Filterer	Gatherer	Grazer	Predator	Shredder
0.1054	0.5861	0.0395	0.2099	0.0591

4.6.5.1 Ternary diagram

Figure 4.11 displays the pair-wise covariance of the functional feeding group proportions over the period of collection. The covariance of filterers and other FFGs is larger in this stream than in the previous streams. Gatherers, however, still dominate and those pairs have narrow 95% confidence limits. Proportions of the grazer-shredder pair are very small compared with the amalgamated other three groups as indicated by the data points and 95% confidence limit ellipse being near the top of the triangle.

4.6.5.2 Covariance biplot

The Starvation Canyon Creek covariance biplot of observations and variables is in Figure 4.12. The FFG proportions have a wide range of variability: the standard deviation for the predator FFG is quite low, while that of the shredder and grazer FFGs is much higher. The predator FFG is inversely correlated with the shredder FFG, grazers and gathers are almost inversely correlated, and filterers are strongly positively correlated with the predators.

The five years of observations divide into two groups along the first principal component: years 1, 4, and 5 are close to each other and near the negative end of the axis range; years 2 and 3 are on the positive side of the axis range but more widely separated than are the first group. Along the second principal component, only years 2 and 4 are moderately close, and all years are quite different from all the others.

4.6.6 Water Pipe Creek

The functional feeding group counts for Water Pipe Creek are shown in Table 4.11. The mean proportions reveal that gatherers comprise 52% of all taxa over the range of years in the data set and predator taxa are about 19% of all taxa. (Table 4.12). The metric standard deviation (msd) for the set of proportions is 0.2519 which indicates relatively low variability.

4.6.6.1 Ternary diagram

Figure 4.13 displays the pair-wise covariance of the functional feeding group proportions over the period of collection. This stream system is quite different from the others. The proportions tend to be near the center of the ternary plot (other than for pairs with gatherere), and all 95% confidence limit ellipses are extremely narrow. This suggests there is comparatively greater variability among taxa over the years but less variability when considered pair-by-pair.

Table 4.11: Water Pipe Creek: number of taxa in each functional feeding group by year.

Date	Filterer	Gatherer	Grazer	Predator	Shredder
2006	10	36	10	15	4
2010	12	56	9	14	7
2011	7	53	13	23	4
2012	17	51	10	19	6

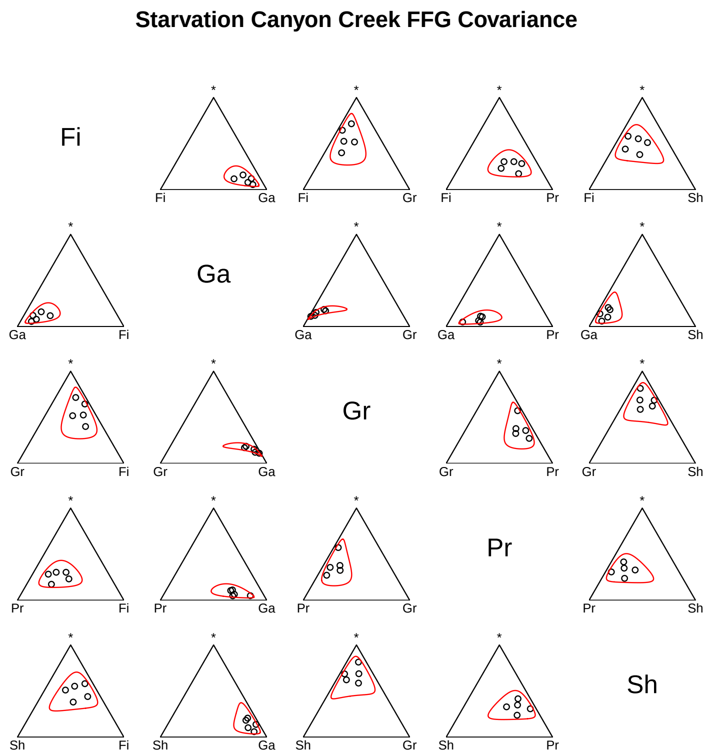

Figure 4.11: Starvation Canyon Creek: functional feeding groups covariance.

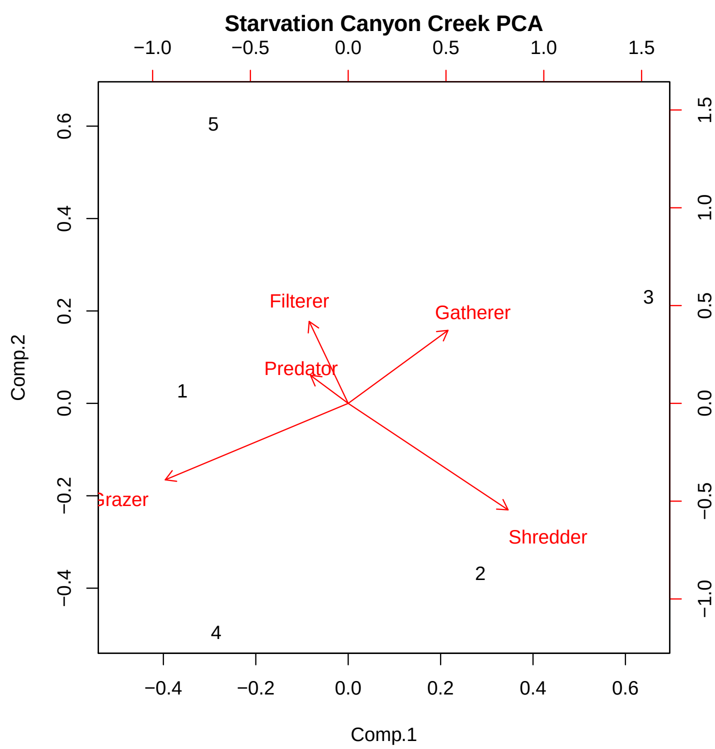

Figure 4.12: Starvation Canyon Creek covariance biplot.

Table 4.12: Water Pipe Creek: Aitchison mean proportions of functional feeding groups for all years.

Filterer	Gatherer	Grazer	Predator	Shredder
0.1186	0.5243	0.1129	0.1889	0.0553

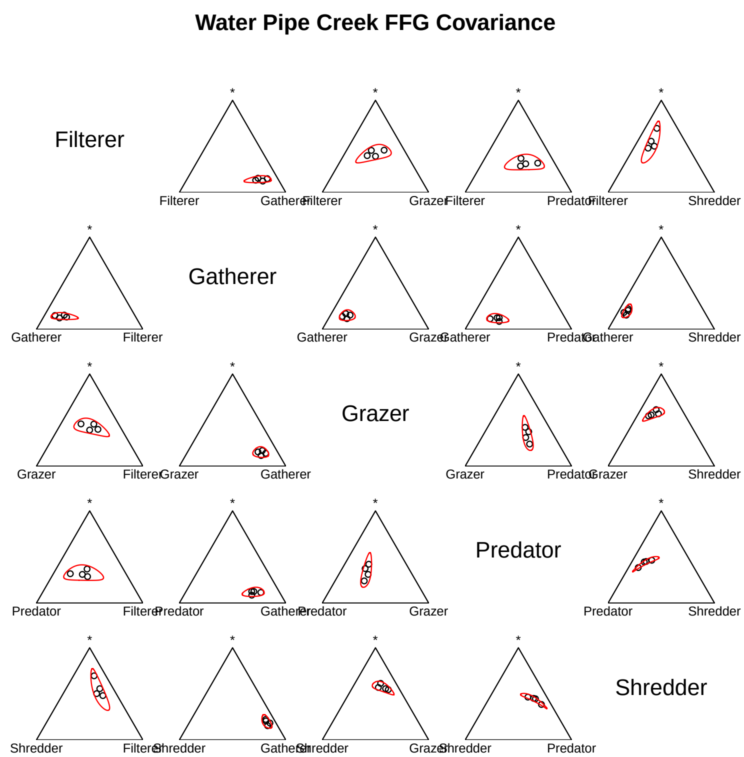

Figure 4.13: Water Pipe Creek: functional feeding group covariances.

4.6.6.2 Covariance biplot

No biplot can be drawn for Water Pipe Creek because there are only 4 years of data collection and 5 functional feeding groups. Principal components analysis (PCA) and biplots require that the data matrix have at least the same number of rows and columns; more rows than columns is the usual case. Every statistical model requires a minimum number of data points to produce valid and useful results. If additional data become available, the model can easily be re-run and the data set visually described.

4.7 Cause and effect

Keep in mind that statistical significance is not necessarily ecological or environmental significance. Macroinvertebrate collections from riffles and pools along a stream might have statistically significant differences in their FFG component ratios that are not ecologically significant because of habitat differences. Each collection is well adapted to the substrata and hydrology in which it lives.

The three chemical constituents[6] used in this example (NO_3, SO_4, and pH) are continuous potential predictor variables which might explain compositions of FFGs in each stream. Nitrate nitrogen concentrations might result from blasting activities, sulfate might stunt plant growth, and pH might indicate acidification from mining activities or affect the solubility of ions. Appropriate plots and linear regression models of FFG compositions as a function of these explanatory variables would reveal statistically significant relationships between aquatic biota and water chemistry. If there also are no significant ecological relationships the aquatic life designated beneficial use has been achieved across the range of chemical concentrations measured in the stream.

4.7.1 Continuous explanatory variables

Plots of continuous variable values against the log-ratio-transformed FFG proportions (Figures 4.14 through 4.26) illustrate variability in the relative proportions of each group in the data set for that stream. The columns display the three chemicals (nitrate, sulfate, pH) and the rows display the FFGs in alphabetical order. While there is a regression line (in red) on each plot, the predictive value is low because of high variability and clumping. More data likely will increase the predictive value of the fitted regression line.

Analying environmental data with multiple potential explanatory variables requires considering interactions among them. For example, determining if pH affects concentrations of NO_3 or SO_4, whether one of the later two influences the value of the other. and whether the year influences any of the three.

4.7.2 Discrete explanatory variable

Another explanatory variable is the year of data collection. Normally, the relationships between FFG compositions and non-continuous explanatory variables are evaluated using Analysis of Variance (ANOVA) rather than regression. However, there is a model combining ordinal (years in this case) and ratio (chemical concentrations) variables in a regression model.

The relationships of FFGs and year of collection is shown in ternary plots. These plots present the same variables as those in Section 4.6 but with each year as a separate symbol and color rather than as the same symbol within the 95% confidence interval for all years.

Another representation of the relationships between FFGs and years is provided by a matrix of boxplots (Figure 4.16 and similar). The x-axis of each panel is the year of data collection and the y-axis is the log-ratio of the two components. In each stream, and for each pair of FFG components, there is very low variability, seen in the individual box heights and whisker lengths being very small. When the centers of the boxes overlap the median proportions of the two FFGs are not significantly different. Each stream has at least one year when a particular pair of FFGs were significantly different from other years. Overall, however, these boxplots show the consistency in relative proportions of each FFG supporting the conclusion that there is no discernable anthropogenic effects on the benthic macroinvertebrate communities in these streams.

[6] Any number of chemical constituents can be used as explanatory variables.

Table 4.13: Winters Creek: explanatory variable values for analytical model.

Year	Stream Length	Drainage Density	Channel Slope	NO_3	SO_4	pH
2004	51.13	1.30	1.76	0.50	170	6.81
2005	51.13	1.30	1.76	0.27	180	8.26
2006	51.13	1.30	1.76	0.30	150	7.78
2011	51.13	1.30	1.76	0.10	150	7.84
2012	51.13	1.30	1.76	0.14	120	7.70
2013	51.13	1.30	1.76	0.14	120	7.70

4.7.3 Regression models

Understanding relationships among potential explanatory variables allows meaningful interpretation of the effects of the explanatory variables (chemicals, basin parameters, and year) on the response variable (relative proportions of the functional feeding groups). Parallel plots are used to visualize dependence of a composition on continuous variables such as water chemical constituents. These are matrices where the explanatory variables are the columns and the compositions are the rows. To display a discrete variable (year in this monograph) there are two choices: parallel plots or the use of color and/or size on the continuous variable plots.

The model for multivariate regression when compositional data is the response variable is given by this equation:

$$ilr(Y_i) = ilr(a) + X_i ilr(b) + ilr(\epsilon_i)$$

where a and b are unknown compositional constants, Y_i is a random composition, X_i is a real explanatory variable, *ilr* is the isometric log-ratio, and the error term $\epsilon_i \sim N(0_{D-1}, \sum_{ilr})$ has a mean of zero and a standard deviation of 1. The output is the estimated parameters $\hat{a}$ (the intercept) and $\hat{b}$ (the slope) in ilr coordinates. These parameters are converted back to the original units and displayed as the components of the composition. Interpret the intercept, **a,** as the expected composition when X = 0. The slope, **b**, is interpreted as the perturbation (amount of change) to the composition if X increases by one unit. The output makes little sense when expressed in ilr units because it tests each component separately and not as a whole composition so an ANOVA is applied to test for the joint significance of X.

The column headings for regression results are the functional feeding groups: Fi = filterers, Ga = gatherers, Gr = grazers, Pr = predators, and Sh = shredders.

As the number of years increase for which data have been collected the analytical results become more robust. This is seen in the individual chemical ANOVA significance levels increasing from 10% to 1% as the years of data collection increase to from 4 to 8. However, these years are too few to model the interactions among all explanatory variables and the compositional response variable.

4.7.4 Winters Creek

The three basin and three chemical parameters used as explanatory variables in the analytical model are in Table 4.13.

4.7.4.1 FFG composition dependence on continuous variables

Figure 4.14 shows the chemical explanatory variables against the centered log-ratio (clr) coefficient of the FFG components. Filterers and grazers increase proportionately with nitrate and sulfate concentrations, but decrease with increasing pH. Predators and shredders decrease relative to increasing nitrate and sulfate concentrations, but their responses differ with increasing pH. Predator response to pH increases is relatively flat, similar to that of grazers, while shredders show increased relative proportions with increasing pH. Grazers maintain almost no change in relative proportions with increases in nitrate, sulfate, or pH; they are the largest FFG component for all years. While Figure 4.14 does not identify the year of each data point, there are several years with the same or similar concentrations of nitrate, sulfate, and pH yet the effects of those concentrations on the relative proportion of the FFG varies from very small (e.g., sulfate and predators) to very large (e.g., nitrate and gatherers). From a regulatory perspective this high variability and lack of consistent patterns strongly suggests that inherent natural variability is high and if there were any influences from past or current mining activities in those drainages those effects are within the range of natural variability.

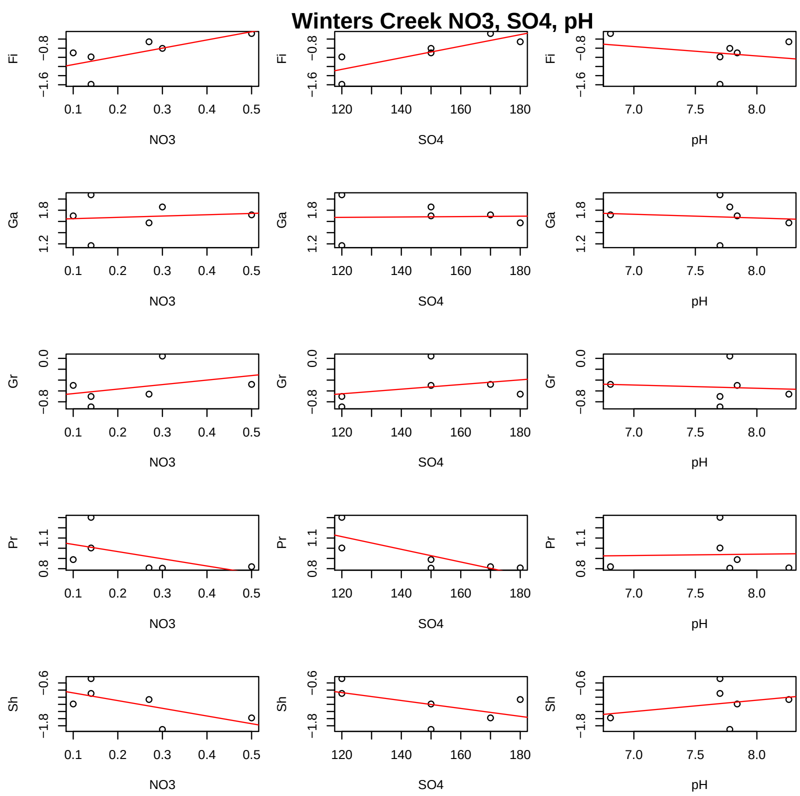

Figure 4.14: Relationships of continuous chemical explanatory variables to the different functional feeding groups in Winters Creek for all years. Fi = filterers, Ga = gatherers, Gr = grazers, Pr = predators, Sh = shredders.

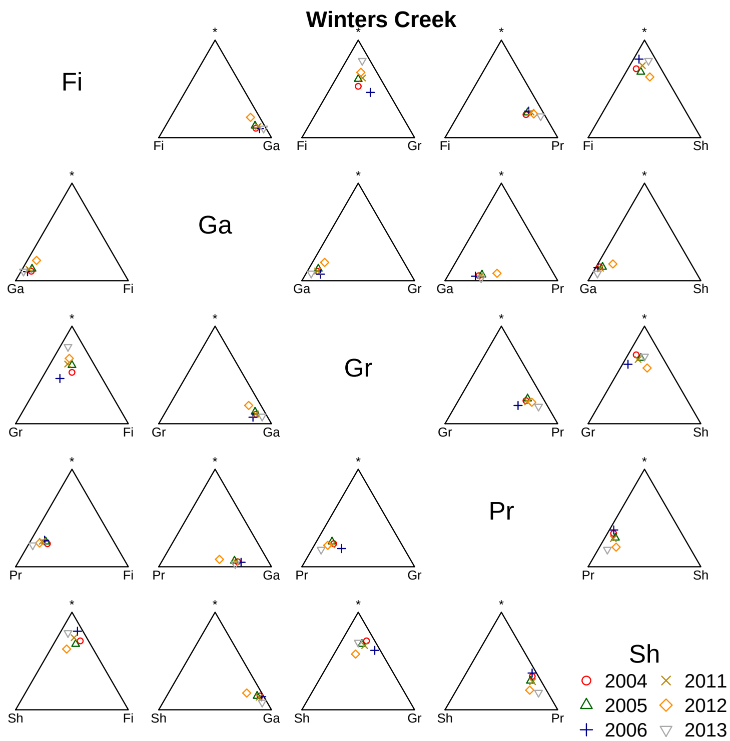

Figure 4.15: The year of data collection is the discrete (categorical) potential explanatory variable for functional feeding groups in Winters Creek.

4.7.4.2 FFG composition dependence on discrete variables

Figure 4.15 shows relationships of the discrete explanatory variable (year of data collection) on FFG concentrations. A different view of the relationships between functional feeding groups and year data collection is seen in the boxplots of Figure 4.16.

4.7.4.3 Regression and ANOVA results:

Nitrate (NO_3)

```
              Fi     Ga     Gr     Pr     Sh
(Intercept) 0.0280 0.5426 0.0513 0.3231 0.0550
log(NO3)    0.5914 0.1238 0.2250 0.0485 0.0113
Analysis of Variance Table
            Df Pillai approx F num Df den Df Pr(>F)
(Intercept) 1 0.99877  202.297      4      1 0.05268 .
   log(NO3) 1 0.93626    3.672      4      1 0.37065
Residuals 4
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Sulfate (SO_4)

```
              Fi      Ga      Gr      Pr      Sh
(Intercept) 0.0046   0.3789  0.0231  0.4760  0.1174
log(SO4)    0.2025   0.2001  0.20081 0.1988  0.1978
```

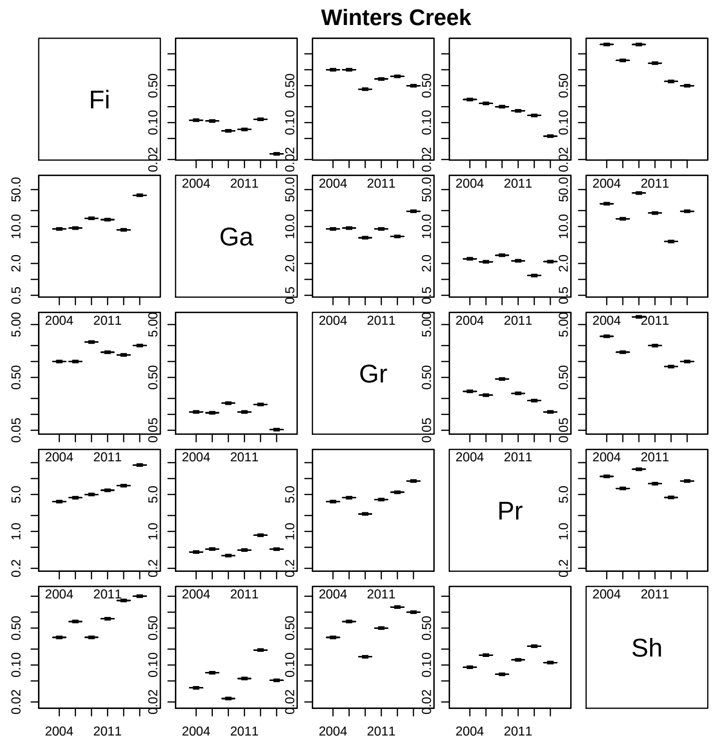

Figure 4.16: Boxplots of FFG log-ratios as a function of data collection year in Winters Creek.

```
Analysis of Variance Table
             Df  Pillai approx F num Df den Df  Pr(>F)
(Intercept)  1 0.99804  127.443      4      1 0.06633 .
log(x2)      1 0.97971   12.072      4      1 0.21221
Residuals    4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

pH

```
            Fi     Ga      Gr     Pr     Sh
(Intercept) 0.1388 0.6312 0.0655 0.1626 0.0019
pH          0.1602 0.1842 0.1854 0.1994 0.2708
Analysis of Variance Table
            Df  Pillai approx F num Df den Df  Pr(>F)
Intercept)   1 0.99918  304.229      4      1 0.04297 *
pH           1 0.89346    2.096      4      1 0.47223
Residuals    4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In Winters Creek, nitrate and sulfate do not influence FFG compositions while pH does.

4.7.5 Snow Canyon Creek

The three basin and three chemical parameters used as explanatory variables in the analytical model are in Table 4.14.

Table 4.14: Snow Canyon Creek: explanatory variable values for analytical model.

Year	Stream Length	Drainage Density	Channel Slope	No_3	SO_4	pH
2005	25.23	0.88	2.92	0.60	816	7.87
2006	25.23	0.88	2.92	0.40	224	7.59
2010	25.23	0.88	2.92	0.10	571	7.81
2011	25.23	0.88	2.92	0.52	130	7.42
2012	25.23	0.88	2.92	0.42	363	7.79
2013	25.23	0.88	2.92	0.42	363	7.79

4.7.5.1 FFG composition dependence on continuous variables

Figure 4.17 shows the chemical explanatory variables against the centered log-ratio (clr) coefficient of the FFG components.

4.7.5.2 FFG composition dependence on discrete variables

Figure 4.18 shows the discrete (years) explanatory variable against FFGs. An alternate view of the relationships between FFGs and data collection year is seen in the boxplots of Figure 4.19.

4.7.5.3 Regression and ANOVA results

Nitrate (NO_3)

```
            Fi     Ga     Gr     Pr     Sh
(Intercept) 0.0419 0.6042 0.0961 0.2106 0.0472
log(NO3)    0.1056 0.2992 0.3235 0.2337 0.0380
Analysis of Variance Table
             Df  Pillai approx F num Df den Df  Pr(>F)
(Intercept)  1 0.99983  1460.68      4      1 0.01962 *
log(NO3)     1 0.80424     1.03      4      1 0.62037
Residuals    4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Sulfate (SO_4)

```
            Fi     Ga     Gr      Pr    Sh
(Intercept) 0.0115 0.6990 0.0084 0.2751 0.0058
log(SO4)    0.2000 0.1999 0.2001 0.1996 0.2001
Analysis of Variance Table
            Df  Pillai approx F num Df den Df Pr(>F)
(Intercept)  1 0.99997   9850.9      4      1 0.007556 **
log(SO4)     1 0.96773      7.5      4      1 0.2666
Residuals    4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

pH

```
            Fi           Ga         Gr           Pr         Sh
(Intercept) 2.020507e-05 0.93314952 4.343049e-06 0.06682589 4.153889e-08
pH          1.844553e-01 0.06203955 2.427573e-01 0.07681852 4.339293e-01
Analysis of Variance Table
             Df  Pillai approx F num Df den Df  Pr(>F)
(Intercept)  1 0.99978  1142.38      4      1 0.02219 *
```

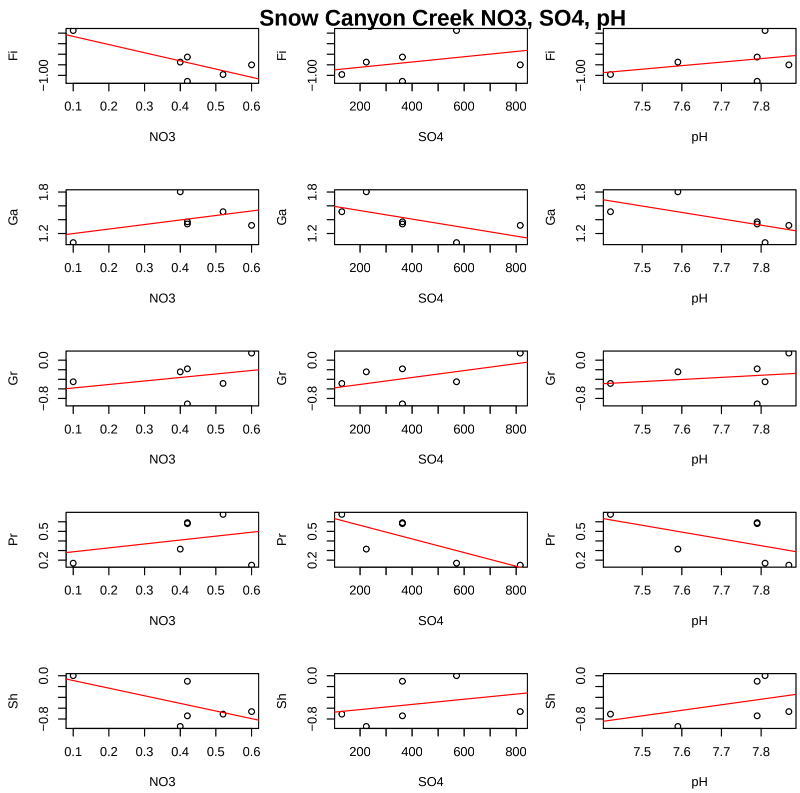

Figure 4.17: Relationships of continuous chemical explanatory variables to the different functional feeding groups in Snow Creek for all years. Fi = filterers, Ga = gatherers, Gr = grazers, Pr = predators, Sh = shredders.

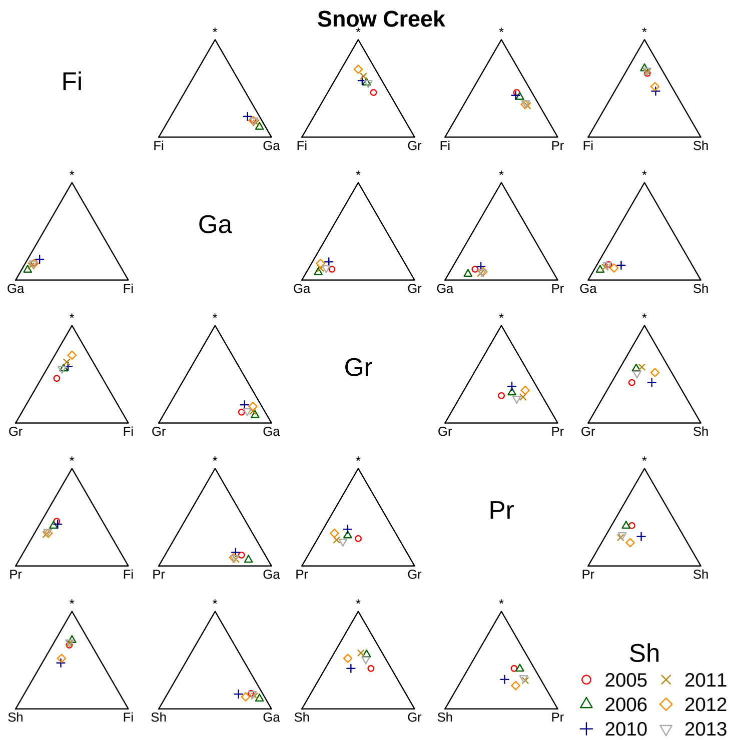

Figure 4.18: The year of data collection is the discrete (categorical) potential explanatory variable for functional feeding groups in Winters Creek.

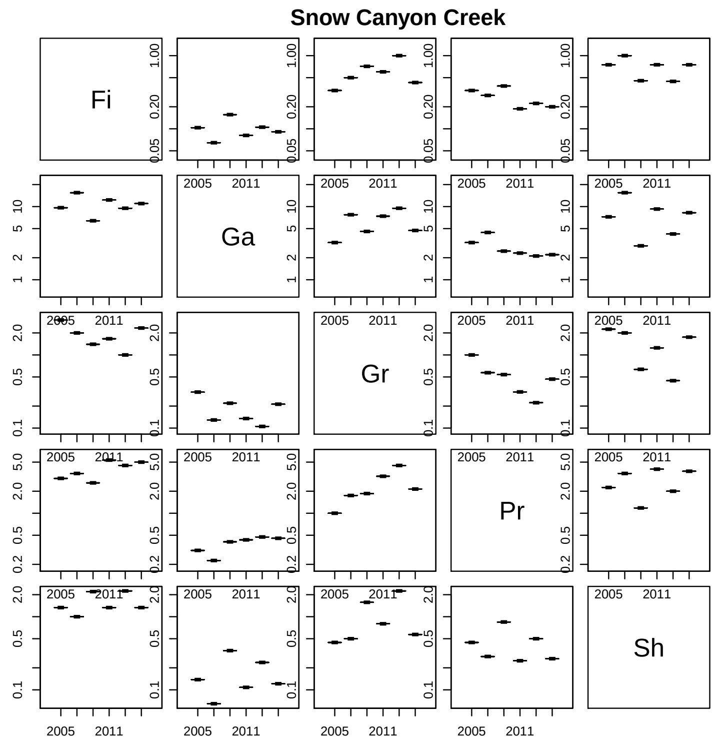

Figure 4.19: Boxplots of FFG log-ratios as a function of data collection year in Winters Creek.

```
pH           1 0.83764     1.29      4      1 0.57170
Residuals    4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In Snow Canyon Creek nitrate, sulfate, and pH all influence the relative proportions of the FFGs in the collections.

4.7.6 Jerritt Canyon Creek

The three basin and three chemical parameters used as explanatory variables in the analytical model are in Table 4.15.

Table 4.15: Jerritt Canyon Creek: explanatory variable values for analytical model.

Year	Stream Length	Drainage Density	Channel Slope	NO_3	SO_4	pH
2004	29.50	0.86	1.32	1.70	2200	8.70
2005	29.50	0.86	1.32	2.50	5000	8.43
2006	29.50	0.86	1.32	1.80	6670	8.57
2010	29.50	0.86	1.32	0.54	4000	8.00
2011	29.50	0.86	1.32	2.70	4300	8.47
2012	29.50	0.86	1.32	0.76	595	8.21
2013	29.50	0.86	1.32	0.76	595	8.21

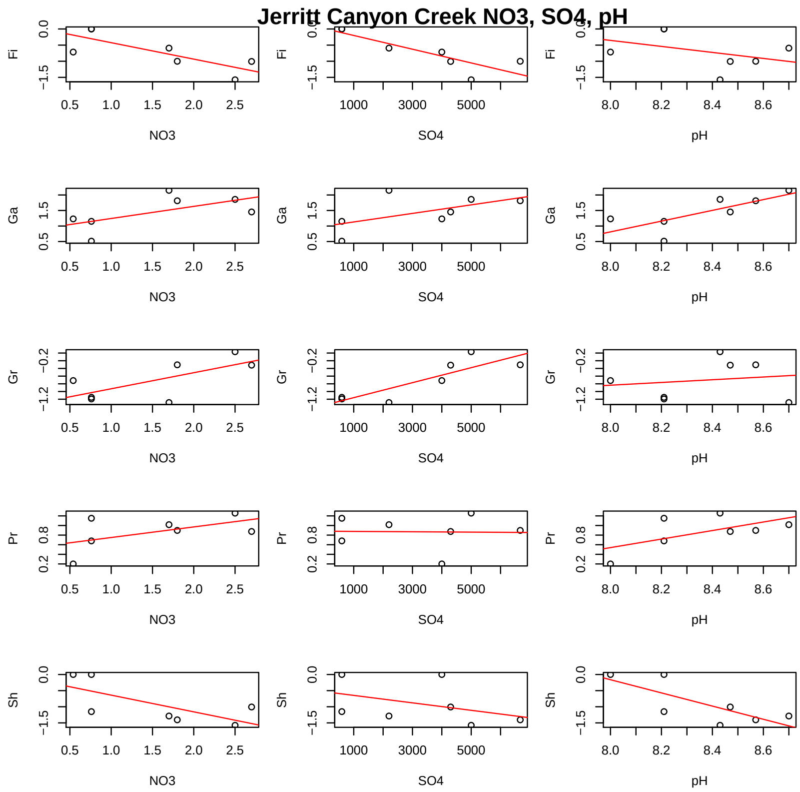

Figure 4.20: Relationships of continuous chemical explanatory variables to the different functional feeding groups in Snow Creek for all years. Fi = filterers, Ga = gatherers, Gr = grazers, Pr = predators, Sh = shredders.

4.7.6.1 FFG composition dependence on continuous variables

Figure 4.20 shows the chemical explanatory variables against the centered log-ratio (clr) coefficient of the functional feeding group components. There is high variability and clumping of data which makes the fitted regression line useful only as a suggestion of trend over time.

4.7.6.2 FFG composition dependence on discrete variables

Figure 4.21 shows the discrete (years) explanatory variable and FFG compositions. An alternate display of relationships between FFGs and year of data collection is shown in Figure 4.22.

4.7.6.3 Regression and ANOVA results

Nitrate (NO_3)

```
            Fi     Ga     Gr     Pr     Sh
(Intercept) 0.1734 0.3736 0.0414 0.2705 0.1411
log(NO3)    0.1104 0.2717 0.2792 0.2292 0.1095
Analysis of Variance Table
            Df  Pillai approx F num Df den Df  Pr(>F)
(Intercept)  1 0.98774   40.296      4      2 0.02436 *
log(NO3)     1 0.76439    1.622      4      2 0.41571
Residuals    5
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

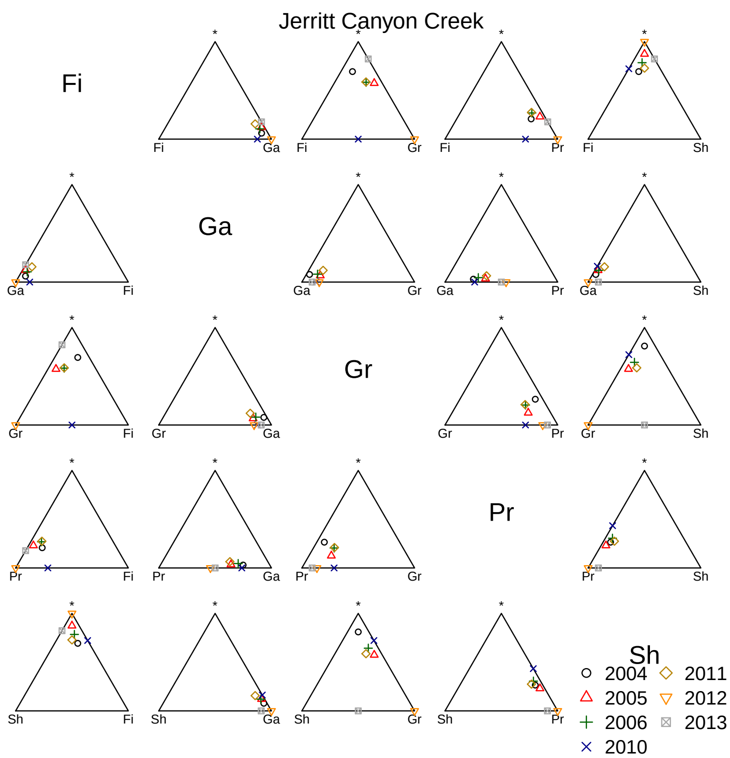

Figure 4.21: The year of data collection is the discrete (categorical) potential explanatory variable for functional feeding groups in Winters Creek.

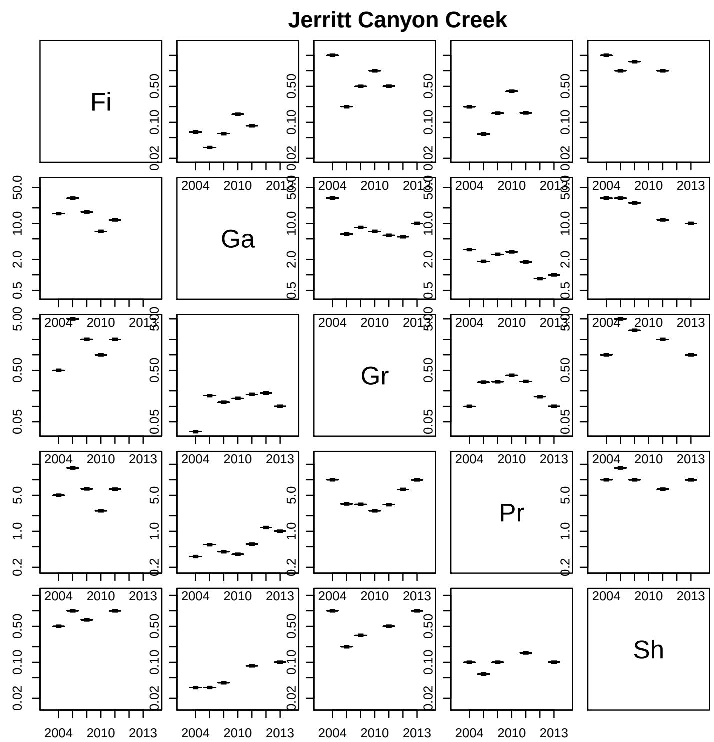

Figure 4.22: Boxplots of FFG log-ratios as a function of data collection year in Winters Creek.

Sulfate (SO_4)

```
              Fi     Ga      Gr     Pr     Sh
(Intercept) 0.1452 0.3873 0.0370 0.3462 0.0843
log(SO4)    0.1200 0.2000 0.2000 0.2000 0.2000
Analysis of Variance Table
            Df  Pillai approx F num Df den Df  Pr(>F)
(Intercept)  1 0.99351   76.558      4      2 0.01294 *
log(SO4)     1 0.95350   10.252      4      2 0.09085 .
Residuals    5
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

pH

```
              Fi           Ga           Gr           Pr           Sh
(Intercept) 0.0001122987 2.012035e-13 2.367091e-09 1.323694e-10 0.99988770
pH          0.0392941399 5.635809e-01 1.421352e-01 2.420748e-01 0.01291498
Analysis of Variance Table
            Df  Pillai approx F num Df den Df  Pr(>F)
(Intercept)  1 0.98931   46.270      4      2 0.02127 *
pH           1 0.69289    1.128      4      2 0.51990
Residuals    5
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In Jerritt Canyon Creek nitrate, sulfate, and pH all influence the relative proportions of the FFGs in the composition. This result is likely because Jerritt Canyon Creek has 7 years of data, more than any of the other streams

Table 4.16: Burns Creek: explanatory variable values for analytical model.

Year	Stream Length	Drainage Density	Channel Slope	NO_3	SO_4	pH
2000	16.44	0.95	1.23	0.29	238	7.61
2003	16.44	0.95	1.23	0.50	40	8.35
2005	16.44	0.95	1.23	0.30	194	8.36
2006	16.44	0.95	1.23	0.02	179	6.52
2010	16.44	0.95	1.23	0.17	182	8.22
2011	16.44	0.95	1.23	0.36	240	8.11
2012	16.44	0.95	1.23	0.19	59	8.20
2013	16.44	0.95	1.23	0.19	59	8.20

except for Burns Creek (below). This illustrates the importance of continuing data collection and analyses to establish the functional dynamics in a stream system.

4.7.7 Burns Creek

The three basin and three chemical parameters used as explanatory variables in the analytical model are in Table 4.16.

4.7.7.1 FFG composition dependence on continuous variables

Figure 4.23 shows the chemical explanatory variables against the centered log-ratio (clr) coefficient of the FFG components.

4.7.7.2 FFG composition dependence on discrete variables

Figure 4.24 shows the discrete explanatory variable (years) and the ratios of each FFG pair. A different view of the relationships between FFG proportions and year of data collection is shown in Figure 4.25.

4.7.7.3 Regression and ANOVA results

Nitrate (NO_3)

```
            Fi     Ga     Gr     Pr     Sh
(Intercept) 0.0957 0.5066 0.1306 0.1694 0.0976
log(NO3)    0.1150 0.2867 0.0062 0.4651 0.1271
Analysis of Variance Table
            Df  Pillai approx F num Df den Df   Pr(>F)
(Intercept)  1 0.99235   97.315      4      3 0.001664 **
log(NO3)     1 0.92458    9.194      4      3 0.049437 *
Residuals    6
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Sulfate (SO_4)

```
            Fi     Ga     Gr     Pr     Sh
(Intercept) 0.0739 0.5658 0.0075 0.3438 0.0090
log(SO4)    0.1997 0.1997 0.2004 0.1996 0.2006
Analysis of Variance Table
            Df  Pillai approx F num Df den Df   Pr(>F)
(Intercept)  1 0.99165   89.030      4      3 0.001899 **
log(SO4)     1 0.68776    1.652      4      3 0.354465
Residuals    6
---
```

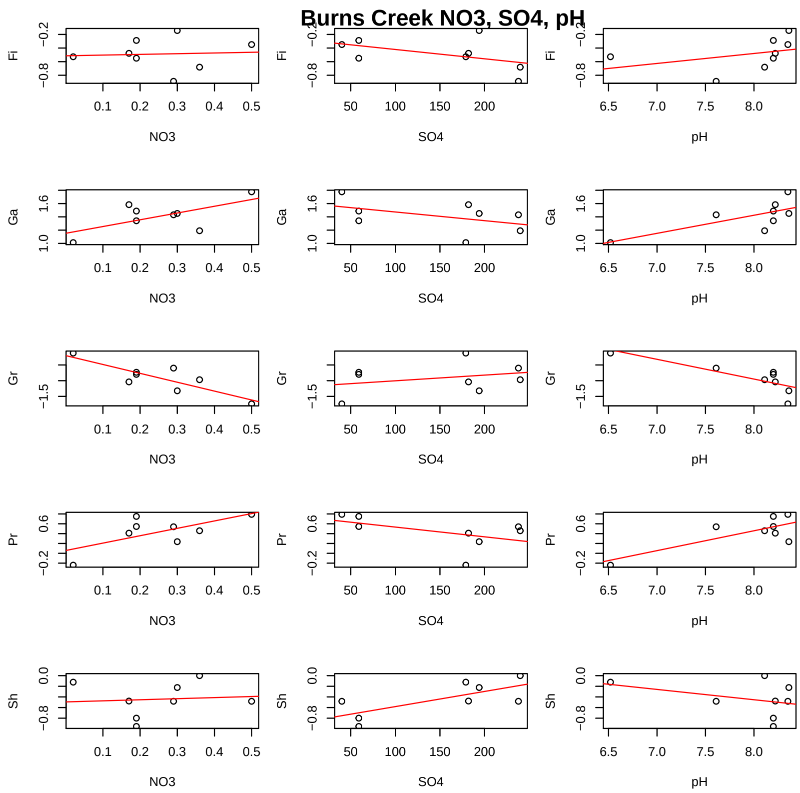

Figure 4.23: Relationships of continuous chemical explanatory variables to the different functional feeding groups in Snow Creek for all years. Fi = filterers, Ga = gatherers, Gr = grazers, Pr = predators, Sh = shredders.

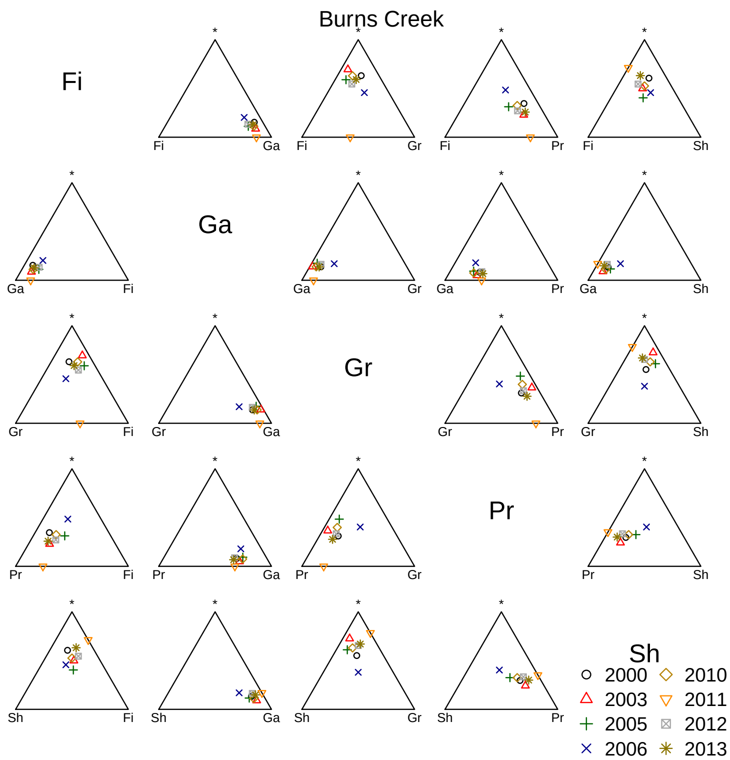

Figure 4.24: The year of data collection is the discrete (categorical) potential explanatory variable for functional feeding groups in Winters Creek.

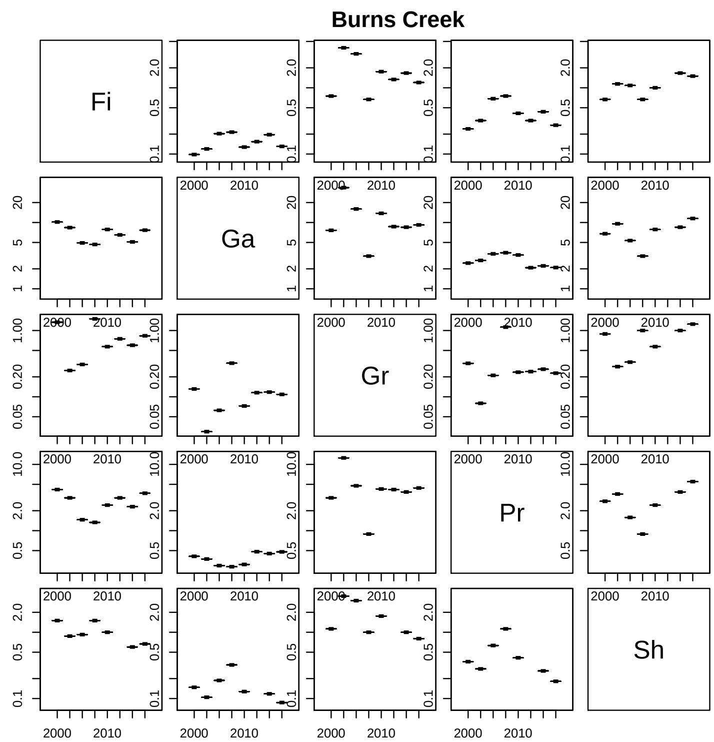

Figure 4.25: Boxplots of FFG log-ratios as a function of data collection year in Winters Creek.

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

pH

```
              Fi     Ga     Gr     Pr     Sh
(Intercept) 0.0031 0.0075 0.9400 0.0010 0.0479
pH          0.2174 0.2466 0.1002 0.2812 0.1547
Analysis of Variance Table
            Df  Pillai approx F num Df den Df   Pr(>F)
(Intercept)  1 0.99149   87.343      4      3 0.001954 **
pH           1 0.84115    3.971      4      3 0.143197
Residuals    6
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In Burns Creek all three continuous chemical variables significantly influence the composition of the functional feeding groups. Statistical models of Burns Creek have 8 years of data as inputs which makes the results more accurate. Statistical models are conservative in limiting significance conclusions when data are sparse, and by not producing output when there are insufficient data for a model.

4.7.8 Starvation Canyon Creek

The three basin and three chemical parameters used as explanatory variables in the analytical model are in Table 4.17.

Table 4.17: Starvation Canyon Creek: explanatory variable values for analytical model.

Year	Stream Length	Drainage Density	Channel Slope	NO_3	SO_4	pH
2006	5.84	1.19	2.72	0.31	61	8.00
2010	5.84	1.19	2.72	0.31	61	8.00
2011	5.84	1.19	2.72	0.21	64	8.06
2012	5.84	1.19	2.72	2.00	58	8.00
2013	5.84	1.19	2.72	0.31	61	8.00

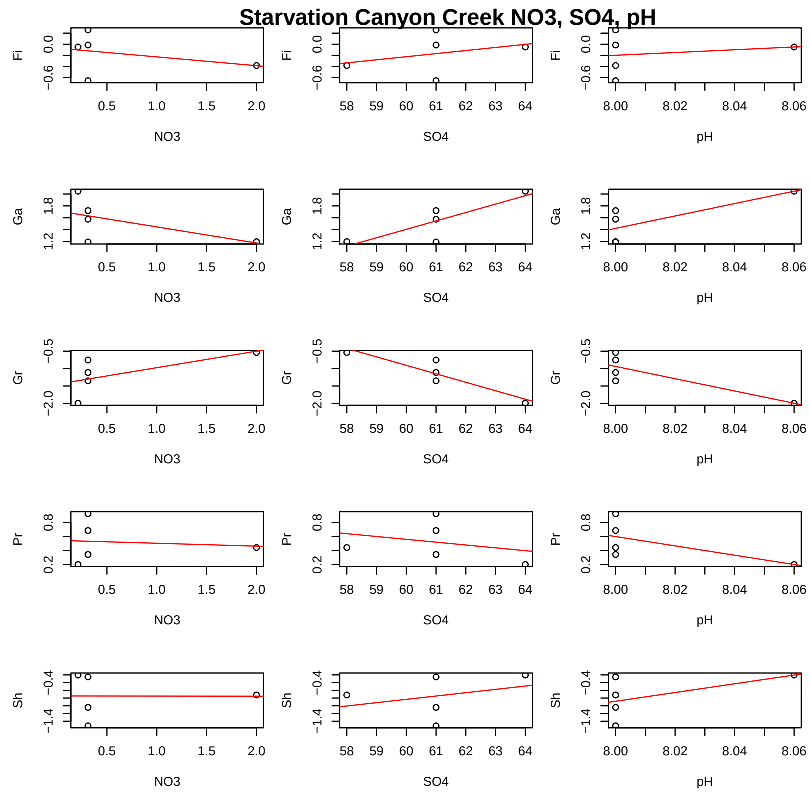

Figure 4.26: Relationships of continuous chemical explanatory variables to the different functional feeding groups in Snow Creek for all years. Fi = filterers, Ga = gatherers, Gr = grazers, Pr = predators, Sh = shredders.

4.7.8.1 FFG composition dependence on continuous variables

Figure 4.26 shows the chemical explanatory variables against the centered log-ratio (clr) coefficient of the FFG components.

4.7.8.2 FFG composition dependence on discrete variables

Figure 4.27 shows the discrete explanatory variable of data collection years and the FFGs. An alternative perspective of the relationships between FFGs and year of data collection is shown in Figure 4.28.

4.7.8.3 Regression and ANOVA results

Nitrate (NO_3)

```
              Fi     Ga     Gr     Pr     Sh
(Intercept) 0.1057 0.5260 0.0670 0.2367 0.0641
log(NO3)    0.1648 0.1472 0.3103 0.1853 0.1924
```

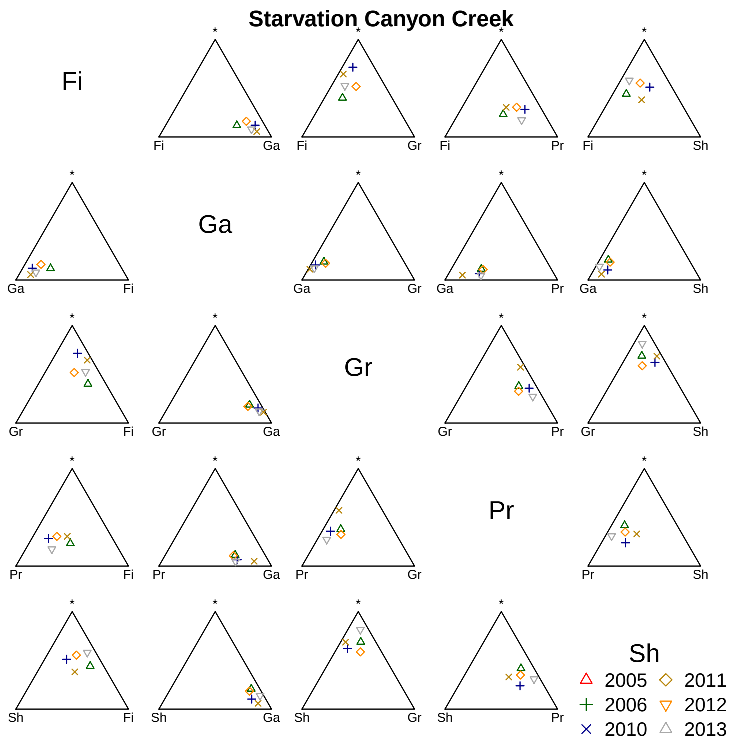

Figure 4.27: The year of data collection is the discrete (categorical) potential explanatory variable for functional feeding groups in Winters Creek.

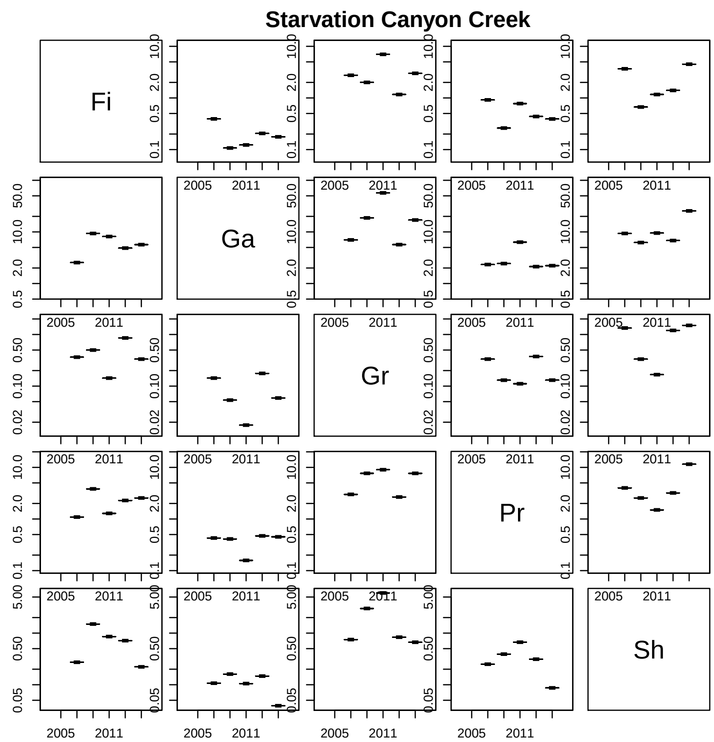

Figure 4.28: Boxplots of FFG log-ratios as a function of data collection year in Winters Creek.

There is too little variation in the concentrations of nitrate values to be analyzed in a regression or ANOVA model.

Sulfate (SO_4) There is too little variation in sulfate values to be modeled in a linear regression or ANOVA. Regression model results are meaningless.

pH There is too little variation in pH values to be modeled in a linear regression or ANOVA. Regression model results are meaningless.

In Starvation Creek nitrate and sulfate concentrations and pH likely do not influence the composition of the functional feeding groups based on the few years of data currently available for analysis.

4.7.9 Water Pipe Creek

No water chemistry data are available for analysis for years 2006–2013. Therefore, there are no regression analyses for this stream system.

Chapter 5

Conclusions

Water quality standards based on maximum concentration limits of ions without explicit documentation how that adversely affects a specific designated beneficial use at a defined location are ineffective and antiquated. These values do not accommodate inherent natural variability or separate natural processes from anthropogenic impacts. Water chemical analyses are snapshots of ionic concentrations at a specific time and location and do not represent ambient conditions.

The biota found in a water body reliably represent ambient conditions. Fish, when occupying stream and river reaches or lakes, indicate that the channel and basin characteristics, physical, and chemical conditions are acceptable. However, fish are not present in all portions of a stream or river network because access might be denied, the substrate not suitable for that stage of its life history, or the current velocity too low or too high. They are also not evenly distributed in standing waters of ponds, reservoirs, and lakes. Benthic macroinvertebrates are ubiquitous and sensitive to physical, hydrological, and chemical conditions in all reaches of a river network from headwaters to mouth. This abundance makes them suitable measures of ambient environmental conditions and identifiers of anthropogenic impacts.

Aquatic ecosystems are highly complex and dynamic, and environmental data collection efforts occur at variable time intervals and low frequencies (e.g., once to a few times per year). Benthic macroinvertebrate taxa vary spatially and temporally and are characteristic of local substrates, flow conditions, and energy sources (external vegetation or in-stream photosynthesis). Over the past 40 years various models have been proposed to summarize aquatic ecosystem complexity in a single number (diversity and biotic integrity indices are good examples). Taxa represent ecosystem structure and not function. Species richness does not necessarily reflect any anthropogenic influences and diversity indices have no objective criteria for determining a value that is "good" or "bad." Diversity indices do not allow quantitative comparisons of different locations or times other than that one value is higher than another value. Biotic integrity indices are created for individual river systems or landscape segments of the river and are tuned to best "fit" that system or segment. There is no objective way to compare biotic integrity values by location or time for the same reasons as diversity indices cannot validly be compared.

Community structure is a poor measure of environmental conditions and is difficult to analyze statistically. Community function, however, is a robust reflection of environmental conditions. Benthic macroinvertebrate functional feeding groups reflect the trophic (feeding) levels that are present throughout the river continuum from headwaters to mouth. While the organic plant base of the biotic community changes from external plant materials in the smaller, upper tributaries to internal photosynthesis by algae, moss, and vascular plants in the open lower reaches, every location has taxa that feed on all available foods. This makes the ratios of functional feeding groups as components suitable for statistical analyses using compositional data analysis models.

Statistical models are more robust and legally defensible than are structural indices. More importantly, the model is consistent yet produces data-specific results. This means the same analysis can be applied to different stream and river networks, or to the same ones at different times, and quantitative differences are meaningful and reflect natural variability, anthropogenic impacts, or both. They can also be used to forecast future conditions, making them valuable for project planning, environmental impact assessments, and monitoring on-going projects to objectively determine whether there are changes due to operations.

The amount of data from Independence Mountains streams used in these analyses are too few to be analyzed by multiple regression models. This is actually a strength of this approach rather than a weakness. Making policy or compliance decisions based on sparse data will be less certain than if such decisions wait until more data have been collected and statistical tests of cause-and-effect can be completed. While regression and analysis of variance models could not be used for all of the data available for inclusion in this monograph, there are sufficient data for five of the stream systems to describe and summarize them. Results of these analyses strongly suggest that the variability in each stream is such that it is not possible to detect any anthropogenic changes in the observed patterns.

While it will take time to shift statutes to a more modern, realistic basis for determining whether industrial activities adversely impact appropriate designated beneficial uses, regulators can start assembling the data needed by requiring permit holders to measure hydrological, physical, chemical, and biological data at the same time and location, regardless of collection frequency. Because benthic macroinvertebrate taxa can be placed into the appropriate functional feeding group at the taxonomic family level much effort, time, and money will be saved in processing the biotic data. And the results are certain to be technically sound and legally defensible.

References

Aitchison, J. (1994). Principles of copositional data analysis. In *Multivariate Analysis and Its Applications*, volume 24 of *Monograph Series*, pages 1–9. American Mathematical Society.

Armitage, P. (2000). The potential of rivpacs for predicting the effects of environmental change. In Wright, J., Sutcliffe, D., and Furse, M., editors, *n International Workshop held in Oxford, UK*, pages 93–111. Freshwater Biological Association.

Armitage, P., Moss, D., Wright, J., and Furse, M. (1983). The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running water sites. *Water Research*, 17(3):333–347.

Bogan, M. and Lytle, D. (2007). Seasonal flow variation allows 'time-sharing' by disparate aquatic insect communities in montane desert streams. *Freshwater Biology*, 52:290–304.

Boyero, L. (2005). Multiscale variation in the functional composition of stream macroinvertebrate communities in low-order mountain streams. *Limnetica*, 24(3-4):245–250.

Bruce, J. (2002). Characterization and analysis of temporal and spatial variations in habitat and macroinvertebrate community structure, fountain creek basin, colorado springs and vicinity, colorado, 1998–2001. Technical Report Water-Resources Investigations Report 02-4093, US Geological Survey.

Cade, B. and Noon, B. (2003). A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and Environment*, 1(8):412–420.

Chessman, B., Williams, S., and Besley., C. (2007). Bioassessment of streams with macroinvertebrates: effect of sampled habitat and texonomic resolution. *Journal of the North American Benthological Society*, 26(3):546–565.

Clarke, R., FUrse, M., Gunn, R., Winder, J., and Wright, J. (2002). Sampling variation in macroinvertebrate data and implications for river quality indices. *Freshwater Biology*, 47:1735–1751.

Cuffney, T., Bilger, M., and Haigler, A. (2007). Ambiguous taxa: effects on the characterization and interpretation of invertebrate assemblages. *Journal of the North American Benthological Society*, 26(2):286–307.

Cummins, K. (1973). Trophic relations of aquatic insects. *Annual Review of Entomology*, 18:183–206.

Cummins, K. (1974). Structure and function of stream ecosystems. *BioScience*, 24:631–641.

Cummins, K., Merritt, R., and Andrade, P. (2005). The use of invertebrate functional feeding groups to characterize ecosystem attributes in selected streams and rivers in south Brazil. *Studies on Neotropical Fauna and Environment*, 40(1):69–89.

Cummins, K., Wilzbach, M., Gates, D., Perry, J., and Taliaferro, W. (1989). Shredders and riparian vegetation: leaf litter that falls into streams influences communities of stream invertebrates. *BioScience*, 39(1):24–30.

de Valpine, P. and Harmon-Threatt, A. (2013). General models for resource use or other compositional count data using the dirichlet-multinomial distribution. *Ecology*, 94(12):2678–2687.

Doledec, S., Olivier, J., and Statzner, B. (2000). Accurate description of the abundance of taxa and their biological traits in stream invertebrate communities: effects of texonomic and spatial resolution. *Archiv fur Hydrobiologie*, 148(1):25–43.

Faith, D. (1990). Benthic macroinvertebrates in biological surveillance: monte carlo significance tests on functional groups' responses to environmental gradients. *Environmental Monitoring and Assessment*, 14:247–264.

Filzmoser, P. and Hron, K. (2010). Robust methods for compositional data. In *Proceedings of COMPSTAT'2010*, pages 79–88.

Hurlbert, S. (1971). The nonconcept of species diversity: a critique and alternative parameters. *Ecology*, 52(4):577–586.

Irvine, K., Dinger, E., and Sarr, D. (2011). A power analysis for multivariate tests of temporal trend in species composition. *Ecology*, 92(10):1979–1886.

Jackson, D. (1997). Compositional data in community ecology: the paradigm or peril of proportions? *Ecology*, 78(3):229–240.

Johnson, D., Hoeting, J., and Poff, N. (2006). Biological monitoring: A bayesian model for multivariate compositional data. In Upadhyay, S., Singh, U., and Dey, D., editors, *Bayesian Statistics and Its Applications*, pages 15–48. Anamaya Publishers, New Delhi, India.

Karr, J. (1987). Biological monitoring and environmental assessment: a conceptual framework. *Environmental Management*, 11(2):249–256.

Legendre, P. and Legendre, L. (1998). *Numerical Ecology, Second English Edition*. Elsevier Sciences B.V.

Lenat, D. (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.

Linke, S., Norris, R., Faith, D., and Stockwell, D. (2005). Anna: A new prediction method for bioassessment programs. *Freshwater Biology*, 50:147–158.

Merritt, R., Wallace, J., Higgins, M., Alexander, M., Berg, M., Morgan, W., Cummins, K., and Vandeneeden., B. (1996). Procedures for the functional analysis of invertebrate communities in the kissimmee river-floodplain ecosystem. *Florida Scientist*, 59(4):215–274.

Pendleton, G., Titus, K., DeGayner, E., Flatten, C., and Lowell, R. (1998). Compositional analysis and gis for study of habitat selection by goshawks in southeast alaska. *Journal of Agricultural, Biological, and Environmental Statistics*, 3(3):280–295.

Reiss, P., Henry, M., Stevens, H., Shehzad, Z., Petkova, E., and Milham, M. (2010). On distance-based permutation tests for between-group comparisons. *Biometrics*, 66:636–643.

Resh, V., Beche, L., and McElravy, E. (2005). How common are rare taxa in long-term benthic macroinvertebrate surveys? *Journal of the North American Benthological Society*, 24(4):976–989.

Ritz, C. (2010). Effectiveness of rivpacs predictive models to evaluate diatom response to nutrient stress in coastal california streams. Master's thesis, California State University Monterey Bay.

Sanchez, G. (2013). PLS Modeling with R. Documentation with the R package 'plspm'.

Shepard, R. (1984). The logseries distribution and Mountford's similarity index as a basis for the study of stream benthic community structure. *Freshwater Biology*, 14:53–71.

Strahler, A. (1957). Quantitative analysis of watershed morphology. *EOS Transactions, American Geophysical Union*, 38:912–92–=0.

Turak, E., Flack, L., Norris, R., J.Simpson, and Waddell, N. (1999). Assessment of river condition at a large spatial scale using predictive models. *Freshwater Biology*, 41:283–298.

Uwadiae, R. (2010). Macroinvertebrates functionsl feeding groups as indices of biological assessment in a tropical aquatic ecosystem: implications for ecosystem functions. *New York Science Journal*, 3(8):6–15.

Van Sickle, J., Hawkins, C., Larsen, D., and Herlihy., A. (2005). A null model for the expected macroinvertebrate assemblage in streams. *Journal of the North American Benthological Society*, 24(1):178–191.

Vannote, R., Minshall, G., Cummins, K., Sedell, J., and Cushing, C. (1980). The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciencess*, 37:130–137.

Wright, J. (2000). An introduction to rivpacs. In Wright, J., Sutcliffe, D., and Furse, M., editors, *Assessing the Biological Quality of Fresh Waters: RIVPACS and other Techniques*, pages 1–26. Freshwater Biological Association.

Yuan, L. (2006). Theoretical predictions of observed to expected ratios in rivpacs- type predictive model assessments of stream biological condition. *Journal of the North American Benthological Society*, 25(4):841–850.