Simplifying Benthic Macroinvertebrate Collection and Analysis
Using Multivariate Statistics

by
Autumn S. Pickett

A Thesis
Submitted in partial fulfillment of the requirements for the degree
Master of Environmental Studies
The Evergreen State College
June 2012
©2012 by Autumn S. Pickett. All rights reserved.
ABSTRACT

Simplifying benthic macroinvertebrate collection and analysis using multivariate statistics

Autumn S. Pickett

Biological assessment (bioassessment) is a direct way to evaluate, track changes and prioritize management actions in ecosystems. Benthic macroinvertebrates are often subjects of bioassessment because they are relatively easy to collect and identify, and have been studied extensively. Bioassessments involve a variety of statistical models that integrate the information collected using different methods. In particular, multivariate models compare the expected occurrence with observed, or ordinate species data to express the observed occurrence of taxa in “species space.” The purpose of this thesis investigation is to use multivariate statistical models to see if there may be meaningful but simpler ways to characterize patterns found in a large macroinvertebrate dataset, and if these summary patterns might simplify the way biological data collection can be conducted in the future.

A large dataset of benthic macroinvertebrates in the Wenatchee Basin was analyzed using multivariate ordination software (PC-ORD 5.32) to compare reference to non-reference sites. The data were examined as abundance and richness of species, higher taxonomic levels and functional feeding groups to see if patterns emerged when compared against selected environmental gradients. It appeared there were several characterizations that did no worse in distinguishing between reference and test sites than the full analysis of raw species. These characterizations were of richness and abundance of functional feeding groups and richness, abundance and presence/absence at higher taxonomic levels.

Importantly, simplifying the classification of macroinvertebrates could allow for identification in the field so that insects could be returned alive to their habitat. Simplified methods may also prove more efficient, less costly and less time-intensive while maintaining the quality of results. More investigation is needed to determine if these simplified methods can be applied to other streams and datasets prior to widespread use.
Contents

ABSTRACT ........................................................................................................................................................................4
Contents ........................................................................................................................................................................ iv
Figures and Images ................................................................................................................................................................vii
Tables ........................................................................................................................................................................viii
Acknowledgements ........................................................................................................................................................ xi
Introduction ....................................................................................................................................................................... 1
Bioassessment ................................................................................................................................................................... 1
  Stream Communities ....................................................................................................................................................... 2
  Integrity and reference sites ........................................................................................................................................ 3
  Stream macroinvertebrates ......................................................................................................................................... 4
Macroinvertebrate Assemblages and Ecosystem Integrity .......................................................................................... 4
Simple Community metrics ......................................................................................................................................... 5
Tolerance measures .......................................................................................................................................................... 8
Rare species .................................................................................................................................................................... 8
Ephemeroptera, Plecoptera and Trichoptera: %EPT and EPT richness ................................................................. 9
Functional Feeding Groups ....................................................................................................................................... 9
Multimetric and Multivariate analysis ....................................................................................................................... 14
  Multimetric Indices of Biological Integrity (IBI) ........................................................................................................ 14
  Metric assignment ....................................................................................................................................................... 15
  Multivariate Models ................................................................................................................................................... 16
Observed vs. Expected (O/E) Models (RIVPACS) ....................................................................................................... 18
Ordination ....................................................................................................................................................................... 19
  Non-metric Multidimensional Scaling (NMDS) ......................................................................................................... 20
Multivariate Models: Important Considerations ....................................................................................................... 21
  Distance measures ...................................................................................................................................................... 21
  Transformation and relativization .............................................................................................................................. 22
  Hypothesis testing ...................................................................................................................................................... 23
Materials and Methods .................................................................................................................................................. 25
  Study Area ................................................................................................................................................................. 25
Washington State Bioassessment ...................................................................................................................................... 28
  Data acquisition ........................................................................................................................................................ 28
  Data collection methods ........................................................................................................................................ 29
  Descriptive Attributes of Sites .................................................................................................................................. 30
Figures and Images

Figure 1. Map of study area showing sample sites of three studies ........................................ 25
Figure 2. "Scree" plot showing stress at different dimensions of all sites with raw species data. 44
Figure 3. NMDS Ordination graph of all sites with all species, showing separation in populations
between wilderness and non-wilderness sites, with mean annual precipitation and longitude as
the main physical drivers. .................................................................................................... 45
Figure 4. NMDS Ordination graph of all sites with all species, showing separation in populations
between wilderness and non-wilderness sites, with mean annual precipitation, longitude and
watershed area as the main physical drivers. ........................................................................ 46
Figure 5. NMDS Ordination graph of all sites with all species showing separation in populations
between reference and test sites. Solid circles are reference sites with lower precipitation, shaded
circles are reference with higher precipitation and open triangles are non-reference sites....... 46
Figure 6. The NMDS ordination of all the sites showing separation of macroinvertebrate
assemblage by stream order ................................................................................................. 47
Figure 7. Assemblages distinguished in ordination space by reference and wilderness
designations with communities defined by their macroinvertebrate order richness. Solid
triangles are wilderness or references, hollow circles are non-reference and non-wilderness. ... 48
Figure 8. NMDS ordination of assemblages defined by functional feeding group richness shown
designated by wilderness and reference condition. Areas of non-reference and non-wilderness
(hollow triangles) are shown occurring in the direction of increasing watershed area (WSAREA).
........................................................................................................................................ 49
Tables

Table 1. Richness and relative abundance of some macroinvertebrate orders and average tolerance related to the divisions in stream order, elevation, month and study. Asterisks (*) denote significant differences based on regression or ANOVA. Groups denote the division by the three smaller studies, EMAP, WC and WEN. ................................................................. 37

Table 2. Stream order composition by study ................................................................. 38

Table 3 Average richness and abundance of macroinvertebrate orders in EMAP, WC and WEN groups (number of different groups or number of individuals). ................................................................................. 39

Table 4. Comparison of physical attributes between reference and non-reference and wilderness and non-wilderness designations for the EMAP, WC and WEN studies. All pairs are significantly different except WEN stream order and EMAP watershed area and stream order for both reference and wilderness ........................................................................................................... 40

Table 5. Number and percentage of references and wilderness sites within each study .......... 40

Table 6. Community metrics differing in reference/non-reference and wilderness/non-wilderness sites. Species richness (S), evenness (E), Simpson’s diversity index (D’) and Shannon’s diversity index (H) values are shown. Asterisks (*) denote significant differences based on ANOVA ................................................................................................................. 42

Table 7. Differences in order-level richness values for two reference designations and their non-reference counterparts. Asterisks (*) denote significant differences based on ANOVA .......... 42

Table 8. Functional Feeding Group (FFG) richness differences for reference or wilderness designations, and between stream order categories (1 - 6). Feeding groups are filterer-collector (FC), gatherer-collector (GC), omnivore (OM), parasite (PA), piercer (PI), predator (PR), scraper (SC) and shredder (SH). Asterisks (*) denote significant differences based on ANOVA .......... 43

Table 9. Community metrics differences (p values) between reference or wilderness designations for each study. Species richness (S), evenness (E), Simpson’s diversity index (D’) and Shannon’s diversity index (H) values are shown. Asterisks (*) denote significant differences. ................. 44
Table 10. MRPP results for the entire dataset and macroinvertebrate community structure compared among a variety of temporal and physical stream variables. The dataset was reorganized by lower taxonomic specificity and a variety of simple community metrics. Values represent A (chance-corrected within-group agreement; effect size) and p-values (the probability that the groups differ by chance alone). Shaded cells denote non-significant results or the inability of a simpler community metric to distinguish among communities. ........................................50

Table 11. MRPP results for the entire dataset and macroinvertebrate community structure compared among a variety of temporal and physical stream variables. The dataset was reorganized by lower taxonomic specificity and a variety of simple community metrics. Values represent A (chance-corrected within-group agreement; effect size) and p-values (the probability that the groups differ by chance alone). Shaded cells denote non-significant results or the inability of a simpler community metric to distinguish among communities. ........................................51

Table 12. MRPP results for each study (EMAP, WC, WEN) separately which compare macroinvertebrate community structure among a variety of temporal and physical stream variables. The dataset was reorganized by lower taxonomic specificity and a variety of simple community metrics (either by species abundance, family and order abundance; functional feeding group richness (FFG1 is using primary designation, FFG2 is primary and secondary designations); richness of the groups EPTD&C, Family, and Order; and lastly presence/absence of Order and Family). Values represent A (chance-corrected within-group agreement; effect size) and p-values (the probability that the groups differ by chance alone). Shaded cells denote non-significant results or the inability of a simpler community metric to distinguish among communities. .......................................................................................................................... 52

Table 13. MRPP results for each study (EMAP, WC, WEN) separately which compare macroinvertebrate community structure among a variety of temporal and physical stream variables. The dataset was reorganized by lower taxonomic specificity and a variety of simple community metrics. Values represent A (chance-corrected within-group agreement; effect size) and p-values (the probability that the groups differ by chance alone). Shaded cells denote non-significant results or the inability of a simpler community metric to distinguish among communities. .......................................................................................................................... 53
Table 14. MRPP results for each study (EMAP, WC, WEN) separately which compare macroinvertebrate community structure among a variety of temporal and physical stream variables. The dataset was reorganized by lower taxonomic specificity and a variety of simple community metrics. Values represent $A$ (chance-corrected within-group agreement; effect size) and $p$-values (the probability that the groups differ by chance alone). Shaded cells denote non-significant results or the inability of a simpler community metric to distinguish among communities. .............................................................. 54

Table 15. MRPP results for difference in total richness between reference designations separated by high elevation (> 900 m) and low elevation (< 900 m) sites. Asterisks (*) denote significant differences between means. ........................................................................................................ 55
Acknowledgements

Much thanks to my "reader," Dr. Carri LeRoy for her help in planning this thesis, help with the software and the many edits it took to get done. But most of all I want to acknowledge her endless patience. Also I owe her thanks for introducing me to the world of ecological statistics through an elective she offered to MES.

Thanks also to members of the Environmental Assessment Program of the Washington Dept of Ecology who allowed me to use their data. This appreciation goes especially to George Onwumere and Glenn Merritt who gave me so much information on the background of the data projects.

I also appreciate the help of the graduate writing tutors at TESC, especially Jim Ayers.

I derived immense support from my MES cohort that has endured beyond our time in the program. They also have been a great source of inspiration.

My deepest appreciation is for Paul, who encouraged and supported me through this entire process, even when it slowed to a halt. That has meant everything to me.
Introduction
Macroinvertebrates are highly varied and consist of a rich array of species and varied life stages or forms, and they comprise one of the central foci of stream studies (Hauer and Resh 2006). A large dataset of benthic macroinvertebrates was collected by the Washington State Department of Ecology (WDOE) over many years in one large watershed, the Wenatchee Basin. The studies were conducted using the same method of targeting riffles in tributaries of the Wenatchee River, and samples were sent to a lab for subsample identifications of up to 500 individuals. My thesis analyzes these data using multivariate methods to determine how reference sites compared to non-reference sites. Data are examined as raw abundances by species, but also converted into different taxonomic levels, by tolerance levels and functional feeding groups. In addition, the functional feeding groups and higher taxa designations were used as both abundance and richness to see if any patterns emerged when compared against several environmental gradients.

Most bioassessment studies involve considerable investment in time and resources. The purpose of this thesis investigation is to use multivariate statistical models to see if there may be meaningful but simpler ways to characterize patterns found in a large dataset, and if these summary patterns might simplify the way biological data collection can be conducted in the future.

Bioassessment
Preserving and protecting ecosystems is increasingly important as undisturbed natural areas are becoming more scarce and some types of disturbances are irreversible, such as human-induced extinctions. In the United States, this imperative is law: restoring or maintaining biological integrity is part of the requirement of the Federal Water Pollution Control Act (Clean Water Act 1972). Other laws include the Endangered Species Act (1973) which protects species and their habitat. In contrast to chemical sampling which has dominated environmental studies in the past, and only describes a single moment, biological assessment (or bioassessment, bio-monitoring) are increasingly being used by regulatory agencies to assess ecosystems, track changes and prioritize actions (Davis 1995).

Bioassessment is a direct way to collect information about a biological system or ecosystem. A biological description offers a holistic view of the state of an ecosystem because biological communities reflect environmental conditions over time and space (Karr et al 1986).
Bioassessment is conducted by sampling biological assemblages in a structured way in order to ascertain changes, especially due to anthropogenic sources (Karr et al. 1986). A community can be measured and compared by indicators, or metrics, that are found to reflect the condition of that community. Other methods of assessing biological conditions include multivariate statistics, which try to assess many variables at once, finding important patterns and correlations. Both mathematical models use matrices of species and physical variables from each collection location.

The purposes for bioassessments are varied and include characterizing how populations change across environmental gradients, such as elevation, distance, or substrate and how these variables interact. Another purpose is to establish baseline conditions for future comparisons. Many bioassessments, especially by regulatory agencies, are performed to distinguish between an impaired site and a site in natural undisturbed conditions and for characterizing the level of impairment. In some cases of known impairment, the analysis may be used to discern the biological effect, such as assemblage change, due to that type of disturbance. Learning the effect disturbances have on natural communities help guide decision making about land use and restoration.

Bioassessment and monitoring projects are usually based on collecting, enumerating and analyzing samples of populations. Physical and chemical conditions can also be included in studies. The living things that are chosen for a bioassessment study could be singular genres—mammals, birds, fish, insects, plants, periphyton, microbes etc., or could be a community made of multiple trophic levels and groups. It is common for aquatic macroinvertebrates to be used for studies of stream ecosystems.

**Stream Communities**

Whether a particular species is present or not in a stream ecosystem is influenced by a combination of factors, including chemical and physical variables and biological interactions (Carter et al. 2006). The taxa present at any location will depend on their interactions with habitat (substrate, flow, turbulence, presence of woody debris, etc.), riparian conditions, their habit (how organisms move and feed), and the seasonal timing and food supply available (Merritt et. al 2008). Since there are many complex interactions (including physical, historical and biological factors; Holt, 1993) that account for which assemblage occurs in a specific place at a specific time, characterizing conditions in a way that expresses why certain biota are there and
what is natural or disturbed is difficult. The richness, abundance, biomass, specific taxa present and assemblage characteristics of a community may differ in degree over space and time, (Statzner 1987) and the reasons are not always discernible. Yet sampling the assemblage can provide a holistic view that gives clues as to ecosystem functioning and health.

**Integrity and reference sites**

Biological integrity is defined as a living community that exhibits a composition and function that is comparable to natural conditions (Hughes et al. 1998) or a system that is balanced, integrated, stable or adaptive with a full range of ecosystem elements and processes expected in areas with no or minimal human influence (USEPA 2005). This implies that the ecosystems are resilient following disturbance and are more or less self-sustaining. The species assemblage present will be partly the product of evolutionary forces that were shaped by prior natural disturbance of varying degrees and frequencies, and therefore resilience will be part of natural cycles (Resh et al. 1988). Stream sites that exhibit these qualities can be defined as “reference” sites.

The concept of “reference condition” is used to compare sampling sites to the condition that would be expected were there no ecosystem degradation. Careful classification of reference sites is important because it can help define the variation found in stream communities and allow the distinction between variations caused by natural causes and by anthropogenic disturbances (Mazor et al. 2006). The methods used to determine and choose reference sites vary and are sometimes ambiguous. In some cases there are no unaltered areas to study so “best condition available” or “possible” is used. Reference sites can be chosen through statistical analysis of metrics using the sites that exhibit the theoretically best values for some important metrics, (often *a posteriori*) or, they can be chosen *a priori* by their history; those that have been least or undisturbed over some known measure of time and space. Once reference sites have been chosen they can be used in models.

It is best to use a reference site for comparison with others within the smallest scale feasible, like the same watershed, where abiotic and biotic influences are similar. Too few reference sites makes a study difficult, and finding enough valid reference sites at the smallest geographical scale may become expensive. Luckily, it has been established that an “ecoregion” is a fairly good first scale for comparing reference conditions (Feminella 2000, Hawkins et al. 2000), and can be used as the geographic extent for comparing sites. A huge advantage of testing and recording reference sites is that they can be used in future studies “as is” without the need to
determine which part of a set of new sites is eligible for that distinction each time.

Stream macroinvertebrates

Bioassessments use the description of biological entities and assemblages, and although there are many choices, macroinvertebrates are commonly chosen for stream studies. First, because they are a major constituent of all streams with important roles such as cycling nutrients and consuming algae, and second, because they are the main source of food for important fish and other vertebrate species (Merritt et al. 1996). Importantly, aquatic macroinvertebrates are relatively easy to collect in a standardized way, can be identified using available keys, and have been studied extensively so that their tolerance for certain conditions and likelihood of occurrence in a particular place is often known (Haurer and Resh 2006). Additionally they have long enough life cycles to be reflective of disturbances over a longer period of time, not just the moment of measure (unlike physical and chemical data which represent only a snapshot of conditions). Finally, because macroinvertebrates are constrained to specific habitats, aquatic macroinvertebrates reflect disturbance spatially, although “drift” of some species can interfere with analysis. Drift is a natural process where macroinvertebrates that are normally benthic (found on surfaces beneath water) will enter the water column either actively or passively to move and colonize downstream (Smock 2006). Many of these attributes make macroinvertebrates especially useful as indicators for ecosystem conditions. Although macroinvertebrate communities are complex and change spatially and temporally, patterns can still be discerned (Southwood 1996).

Macroinvertebrate Assemblages and Ecosystem Integrity

Macroinvertebrate assemblages can be used to assess streams and rivers by defining their biological integrity. Because biological Integrity is an abstract concept with no concrete definition, it is a description and not a diagnosis (Karr et al. 2000). In other words, the assemblage found in a particular healthy and well-functioning stream at a particular time will describe the integrity for that stream and there should not be expectations of certain assemblage characteristics for defining stream integrity. The composition can vary but a stream will still be healthy. Other "healthy" streams will have unique assemblages that describe their integrity and the characteristics of the range in community composition found in these streams are what is used to compare. Although a macroinvertebrate assemblage can be used to define
and detect the lack of integrity, it is not possible to use it to ascertain the exact cause of a problem. This is due to the complex interactions involved that determine a macroinvertebrate assemblage.

The varying assemblages of aquatic macroinvertebrates depend on many interacting biological and physical factors but a few of these factors have stronger influences and can be used when modeling community structure. For investigations in streams, it has been established that two of the most important physical factors for defining spatial characteristics of biotic occurrence (especially when using macroinvertebrates), besides large-scale geographic area, are stream order, which defines the position of the stream from its source, and elevation (Cereghino et al. 2003). Climate and temperature are greatly influenced by elevation and seem to have an important influence on species presence. There is also a seasonal difference in which organisms may be found in a stream. Life cycles of macroinvertebrates are varied and certain stages will not be present in the stream at certain times. Physical or chemical disturbances can alter the kinds of species found in particular places although the presence of a species could be from recruitment from nearby where conditions are not disturbed. This downstream “drift” occurs with different ease for different taxa. Overall, habitat type and condition may be the main drivers of community composition. Poff (1997) suggests that abiotic factors have more influence on assemblage occurrence than biotic, and that the adaptive traits to survive flooding, drying, local shear stress, temperatures and human pollution are key. Poff et al. (2006) studied the correlation of many traits and trait states for some common stream invertebrates to how they occur over multiple environmental gradients. They found potential in defining some macroinvertebrate traits (uncorrelated ones or groups of traits that occur together) that are robust for predictive power for disturbances and changes and can be used in stream studies. Lamouroux et al. (2004) found that species occurrence depended more on adaptation to physical habitat than food availability. They found that some of the important traits were body form, mode of attachment, feeding habits, reproduction and lifespan. Therefore it is a very complicated network of interactions that drive the composition of the macroinvertebrate assemblage. Yet, there are discernable patterns in the assemblages that can be found and used to assess the condition of a stream.

**Simple Community metrics**

Macroinvertebrate assemblages may be seen and described through different metrics which are
ways of organizing the community and include taxon abundance (the total number of individuals), richness (the number of different species) and evenness (a measure of how evenly distributed individuals are across species, tolerance and functional feeding groups; Carter et al. 2006). Discovering patterns in subsets of the biota and defining them as “metrics” or indicators can be used in lieu of a full sample list for describing an assemblage. Identifying individuals to species level is difficult and time-consuming, so alternatives like identification to higher-level taxonomy are often used. Metrics that work will be those that vary with disturbance and are more or less predictable among similar environments. Metrics are chosen to emphasize a particular distinction of some kind, such as disturbed sites compared to undisturbed. Members of a community can also be described by characteristics like how they feed, (their functional feeding group FFG), how they tolerate conditions (tolerance values), by their habit (if they swim, burrow, etc.), or combinations of these distinctions. These attributes can be substituted for species abundances in the community description to answer specific questions about community structure and function and will be explained further in the next section.

Abundance, a very useful and intuitive metric, is the number of living organisms in a sample, either by category or collectively. Species, or any taxon, can be measured for abundance by count, density, frequency, or biomass (total weight of class or group). Relative abundance or composition involves the ratio of one type to others. Individuals could be identified for abundance at species or higher taxonomic levels, or by classes like age, size or life history stage.

Richness (denoted by "S") is the number of different species or types of entities in a sampling unit. Richness increases with sample size (and size of area sampled) so comparisons must control that variable (Hurlbert 1971). Richness can be the entire number of different species but it can also be used to describe other subgroups, like the richness of important taxonomic family groups, or richness within distinct kinds of functional feeding groups. Richness is fairly easy to calculate and is used in many studies. Species richness makes a good metric for describing assemblages because it reflects a combination of many influences. The physical heterogeneity, the productivity and the geological history of a stream all can be reflected in the richness of biota (Southwood 1995; Statzner 1987). The different ways in which animals acquire food, move, reproduce and grow, and the conditions that are needed to produce their food, will determine where they can exist. For a stream environment, more physical complexity can create an array of micro-habitats, which can accommodate varying needs. A more complex system, both physically and by its community structure, will therefore be able to sustain more types of
organisms and will be reflected in a measure of richness. Richness is generally believed to be a positive attribute of a community and a measure of a thriving healthy system. A high richness reflects a complex system that may be better able to recover from disturbance. Interestingly, richness is highest in intermediate stages of disturbance (Statzner 1987, Southwood 1996) and intermediate stream reaches where physical factors fluctuate from up and downstream influences. High richness may also signal redundancy of biological function where several taxa use or produce a resource together or in competition. Pavluk et al. (2000) showed that in the ideal case, the trophic structure of aquatic ecosystems tends toward the greatest richness in trophic niches or guilds present. When there is high species richness and redundancy of function in a community, and something eliminates or suppresses a species, there are still others that can and will fill its trophic role. Therefore high richness can buffer a community from the negative impacts of some disturbances, although it will not protect stream function from all disturbances.

Evenness, is the degree that abundances of species are equal in a community (Poole 1974), or the probability of encountering a different species in a sample (Hurlbert 1971). This important metric is a "feature of species-abundance relations independent of any single way of measurement or any theoretical abundance distribution" and has many mathematical definitions (Alatalo 1981). Evenness (for non-zero entities) is defined in PC-ORD as
\[ E = \frac{H'}{\ln(S)} \]
which expresses if there is heavy dominance by a small number of species, where \( H \) is Shannon's diversity index (see next section). Since diversity measures like Shannons's \( H' \) are created using a measure of "evenness," there might be some confusion or circularity about evenness measures. Another consideration is that evenness is overestimated as sample size increases because of sampling bias (where richness is underestimated), therefore as with richness, comparison of sites for evenness will require controlling the size (but not the area of collection) of the sample (Hurlbert 1971). Evenness is generally believed to be higher for more mature and stable communities that will exhibit less dominance by one or a few species than by communities in earlier successional stages (Cao et al. 1998), although this may be true for some types of assemblages (like macroinvertebrates) and not for others (like plant communities, i.e. sphagnum bogs) at some scales of observance (landscape and temporal).

Diversity is a calculated function of richness and evenness, or the predominance of species in a sample unit (Hurlbert 1971) and captures patterns of species distribution. Diversity indices are dimensionless statistics that integrate species richness and abundances in a sample and differ mainly by how they represent rare species. Two commonly used indices are
Shannon's entropy \( (H') \) and Simpsons (D). Diversity measures can relate to stability, maturity, productivity, evolutionary time, predation pressure, and spatial heterogeneity (Hill 1973, Hurlbert 1971). For instance, a stream with many microhabitats that has been functioning for a long time would be expected to have high diversity and a recently physically disturbed stream or one that is adjusting to an invasive species would exhibit a lower diversity. But high richness, evenness and diversity measures should not always be interpreted as better. Increased richness might indicate that invasive species have become established. Similarly, increased evenness, for which a higher value is usually considered positive, may indicate the loss of rare species.

**Tolerance measures**

Macroinvertebrates tolerate poor stream conditions at various levels. For different ranges of temperature, turbidity, chemical factors and other variables often associated with detrimental or anthropogenic disturbances, some macroinvertebrates have known and measured tolerances. When conditions are poor, or integrity low, then assemblages with a higher proportion of "tolerant" macroinvertebrates will appear. Studies use measures like count or percent of tolerant species, or one of several structured indices using tolerance information that have been developed (i.e. Hilsenhoff Biotic Index; Hilsenhoff 1988). While tolerance information is useful for bioassessment, it has a few drawbacks. Importantly, tolerance values are often assigned by "expert opinion" derived from where they are found, not from controlled experiments (Carter et al. 2006). And, many tolerance values are determined at higher taxonomic levels that might ignore differences in tolerance for specific genera or species. When species level tolerance is known, the samples need to be identified to the lowest possible taxonomic level which can be difficult using available keys. In addition, tolerance values for some species differ between regions. Some of the tolerance values are derived for specific stressors and so are not applicable to assessing all situations. And finally, some taxa have not been studied and assigned tolerance values, but in most cases tolerance values summarize and reflect a general condition (Carter et al. 2006). Many of the most common species have known tolerances to specific stressors and their predominance in an assemblage can be informative.

**Rare species**

In some analyses using diversity indices like Shannon's and Simpson's, it is recommended to exclude rare species from samples because they might cause confusing “noise,” although in
other analyses rare species are thought to contain important information (McClune and Grace 2002). Obviously, omitting rare species will affect diversity metrics and certainly reduce richness metrics. Cao and Williams (1998) conclude that rare species are “critical for bioassessment.” Their study showed that excluding rare species affected the richness metric at the least impacted sites (that often have a high number of species present) while not affecting the metric of the most impacted sites (that often exhibit low richness) which led to much less sensitivity for detecting the differences between the reference and test sites. But they noted that the effect of excluding rare species on multivariate analysis needs more study (Cao and Williams 1998).

**Ephemeroptera, Plecoptera and Trichoptera: %EPT and EPT richness**

Commonly used metrics to assess streams often include the richness and the relative abundance (percentage of a population) of individuals from the families of Ephemeroptera, (mayflies) Plecoptera, (stoneflies) and Trichoptera (caddisflies) (EPT). This is because these families are very important, prominent and prevalent in most aquatic systems and have known sensitivities to disturbance. Generally they have low tolerances to many disturbances and are therefore indicators of high quality waters. A study in France (Cereghino et al. 2003) used unimpaired rivers and neural network models to successfully correlate EPT and Coleoptera richness with only 4 environmental gradients. They found elevation and stream order to be the most predictive of the EPT richness in their region of study. In addition, these researchers found that at larger spatial scales other environmental factors affected which species were present, but despite this, richness in each of these orders was still predictable (Cereghino et al. 2003). Blocksom (2003) discusses the richness of Ephemeroptera, Plecoptera and the functional group of collector-filterer taxa that vary with catchment area and how they should be considered when used as metrics for rating stream condition. Another study (Baptista et al. 2007) successfully included %Diptera (higher values representing degradation) and %Coleoptera (representing primary production).

**Functional Feeding Groups**

The motivation for assigning and studying guilds or functional feeding groups is to make ecosystem analysis easier by creating a framework that incorporates and defines all species. Grouping helps reduce a community into a smaller dimension that can be more easily understood. Taxonomy by itself is not only a lot of information but it does not reflect how a
community functions and interacts. Communities can be populated by vastly different species; however many of these species can be similar in terms of function. It is important to find a meaningful way to group species when performing community analysis and bioassessment, one that will help elucidate important functional relationships in the community.

There are several important ways organisms function in a community, and many ways to characterize and identify these. Trophic status reflects how an organism derives its energy in the food chain. Organisms can be predators, prey, primary consumers, producers or detritivores. Guild is a confusing term that originally was meant to organize creatures by how they used resources. If resources are viewed mainly as food, guilds could be quite similar to trophic status. But the guild concept also considers methods of food acquisition. In the case of macroinvertebrates, these can be by filtering, scraping and piercing for example. Niche is another term with ambiguous usage, but generally is meant to describe the physical and resource requirements needed for a class of organisms to survive (Simberloff and Dayan 1991, Southwood 1996, Loeschcke 1987). An interesting account of the history, differing opinions and usages of these terms can be found in Simberloff and Dayan (1991). Nonetheless, feeding methods of organisms show adaptations to niches and can be used to characterize macroinvertebrate communities (Merritt and Cummins 2006).

Assignment of organisms in a community into groups, guilds or niches may be a difficult and imprecise task making their use as indicators questionable. In the case of macroinvertebrates, many appear to be flexible enough in their habits to survive by more than one rigid manner. Multivariate quantitative analysis can help resolve some of the ambiguity, incorporating the complexity of species, and defining the classes of resources (Simberloff and Dayan 1991). In addition, assignment to groups depends on the definitions used, but the definitions for resources used and ways of using them can be ambiguous or defined at different levels of specificity. Simberloff and Dayan (1991) suggest that the term "Functional Group" should be used to describe members of a community that use similar resources in a similar way and might be in competition. But, use of a similar resource does not preclude some kind of resource partitioning or separation by acquisition method used that avoids competition. Using a guild or group as an indicator can be risky because of the finer differences that change the meaning of the group relationships.

Functional feeding groups (FFGs) are based on 4 food categories (coarse and fine particulate organic matter, periphyton and prey) and the morphological mechanics and behavior
associated with acquiring the food (Merritt and Cummins 2006) resulting in broad categories of predator, parasite, collector-filterer, collector-gatherer, shredder, piercer and scraper. The mouthparts of the macroinvertebrates determine the easiest mechanism for ingesting food, such as scraping periphyton from surfaces or piercing and sucking juices from plant cells or other animals, and the size and shape of the animal can influence where it can forage.

Although often useful, FFGs are not always an accurate way to organize a community. The usefulness of this structure can be compromised because categorization of species into FFGs is sometimes difficult. For instance, what is actually ingested can change as the macroinvertebrate grows and seasons change. The food resources are either plants, animals, detritus or a combination, but there are divisions in these resources to consider; some herbivores will eat live primary producers from within the stream system which can be algae or other plants, but others will eat from the riparian edges. A "piercer" can be a predator or an herbivore. And it is not always clear if the organism eats fresh or decaying food. Groups that are assigned may not always reflect many other important characteristics like the size category of the food that is eaten or the body type that dictates where exactly the food is eaten from. Another problem of using FFGs in an analysis is that the FFG for a species is not always well-defined. Tomanova et al. (2006) studied taxa in neotropical streams to determine more accurate categorizations for tropical species. They found some significant differences from the assumed and assigned FFGs from their study of gut contents. They suggest that the genus may adapt and utilize what is abundant and alter its FFG in order to survive (Tomanova et al. 2006), so assignments to FFG can differ by region. Likewise, in streams with strong currents, species may adapt to eat things that will allow them to avoid browsing on unstable surfaces. So although macroinvertebrates can express a dominant feeding morphology, the complexity and flexibility of what they eat (which can change between seasons, rivers and habitats) can make FFG assignment difficult. Yet, functional feeding groups are an important way to categorize macroinvertebrates and are a more simple, useful and valid way to describe a community than many other methods.

By disregarding taxonomic relationships which do not express how organisms interact as a community, FFG designations may paint a better picture of community structure and function because unrelated taxonomic groups can exhibit the same functional traits (Poff et al. 2006, Merritt and Cummins 2006). Using taxonomy alone could result in grouping species for analysis that might either compete or be otherwise mutualistic. Using FFGs allows the study of those
functional groups that interact in a community without unnecessary noise from a large
taxonomic list. Because the FFGs partially reflect stream condition, using them may lend more
information to an analysis of a stream ecosystem.

Functional feeding groups reflect both the geomorphic and the overall biotic conditions
of a stream and provide insight into the food resource base at both site-specific and general
levels. The proportions of FFGs present will reflect the available food because of the
morphological and behavioral mechanisms of food acquisition by the organisms and the
diversity of FFGs show the degree that a community is dependent on different food resources
(Merritt and Cummins 2006). It has been suggested that the richness and composition of FFGs at
one trophic level may affect groups at lower levels (Jonsson and Malmqvist 2005; Vannote 1980).
Uwadie (2010) found that FFG communities changed predictably with habitat size, where small
forested streams were dominated by shredders and gatherers, medium streams were dominated
by scrapers and gatherers, and larger streams by gatherers and filterers. He concluded that it
was possible to more easily get important information using FFG ratios instead of species. His
work confirmed the predictions of the River Continuum Concept (Vannote et al. 1980).

The River Continuum Concept of Vannote et al. (1980) describes a predictable model of
biotic assemblage occurrence transition from the river headwaters to the mouth. Physical
channel factors have much to do with the biotic continuum (Statzner 1987) and account for
differences between regions (Resh et al. 1988). But in general, species exploit the environment
in the most efficient way to maximize energy consumption and this creates a predictable series
of species assemblages. As resources are processed, some are stored and some released
downstream where they are utilized. In general, this manifests as shredder groups being most
common in lower order streams with higher riparian edge input, which are gradually replaced by
grazers in the mid-reaches and then dominated by collectors where the rivers become large and
wide. There will be changes in the balance over a season due to the shifts in resources available
and their processing by fluctuating populations of macroinvertebrates, and this dynamic will
continue to evolve over years. Therefore when looking at an assemblage, the species identity
may not matter as much as its function. Additionally, higher functional group richness could
increase the stability of a community if the stream conditions (biotic, temperature, substrate,
flow, food and riparian condition) provide enough diversity to sustain a diverse community.
Higher richness will support more species in the detrital processing chain in the stream and
therefore is an overall positive attribute for stability of a system (Jonsson and Malmqvist 2005).
Therefore, functional feeding group can provide important information about the integrity of a stream.

Studying the diversity or richness and the presence or absence of FFGs can provide interesting analysis. High richness of these groups can be a signal for ideal health (Pavluk et al. 2000). The Index of trophic completeness (ITC) is a group of indices using “functional trophic relations.” A study by bij de Vaate and Pavluk (2004) concluded that the theory, which suggests all trophic guilds will be present in healthy systems, is correct. The study identified and compared 12 FFGs that were based on food resources, food size and method of food acquisition which required identification to species level. The complete set of guilds should be present in streams despite the differences in species abundance and composition due to seasons, substrates, velocities and other physical situations (Cummins 1973). Even natural disturbances will not alter this for very long as recovery takes place over time according to the different species life cycles and environmental fluctuations on many scales (Statzner 1987, Resh et al. 1988). The physical structure of the stream will be "reset" and the biota present might exhibit alternative (facultative) feeding behaviors at first (Statzner 1987). Only in truly disturbed streams (caused by pollution or a harmful physical alteration) will guilds be eliminated or missing (bij de Vaate and Pavluk 2004).

Snyder and Johnson (2006) confirmed this in their study of Blue Ridge Mountain (VA, USA) streams disturbed by catastrophic floods. The physical changes due to flooding were reflected in trophic structure based on total macroinvertebrate density, but the communities were stable and diverse. A study in Nigeria in lagoons correlated the percentages of four FFGs with environmental parameters like total dissolved solids and total organic matter. Percentage differences in the FFGs and the loss of guilds were observed in highly disturbed locations (Uwadiae 2010). This study showed that various types of pollution and disturbance will affect the macroinvertebrate community structure by loss of guilds. In two anthropogenically disturbed rivers Uwadiae (2010) found increases in a species of predator that resulted in significant loss of guilds. But an “ITC” index cannot be used to identify the kind of pollution, although some kinds of pollution will cause a predictable response in some of the guilds (like anything reducing the growth of primary producers may have a negative effect on herbivores).

Macroinvertebrates, when correctly identified by taxon, functional feeding group or guilds and organized by richness or abundance can be used to assess rivers and streams. There are many interesting relationships in macroinvertebrate communities that can characterize
stream conditions. Statistical methods have been developed that use assemblages for assessing stressor effects, baseline conditions and changes over physical gradients in streams.

**Multimetric and Multivariate analysis**

Biological systems can be modeled and described in many ways. Multimetric and multivariate models are two of the most common statistical tools used to evaluate complex ecological systems. Both incorporate the variety of biological occurrences and physical conditions found in nature to categorize sites and both can be used in bioassessment. Both multimetric and multivariate methods are data-intensive, needing large datasets that include many variables organized as matrices of species and physical attributes for a collection of sites. Both of these methods use sites of known integrity (reference sites) to compare to others. Both of these models can distinguish differences in biological occurrence due to physical gradients (like elevation, stream channel substrate or riparian condition, etc.); however, neither multimetric nor multivariate models can be used to directly explain the cause of an aberration in the expected assemblage, they are descriptive only. Multimetric indices can rate the condition of streams, using the metrics (with values that were rated by condition) whereas multivariate models can characterize a stream based on a suite of variables all at once, displaying similarity of members among groups.

Both multimetric and multivariate models seem equally valid. Herbst and Silldorf (2006) compared multimetric IBI and multivariate software "RIVPACS" models with three collection and processing methods and found they were all very similar in describing streams communities. Stribling et al. (2008) found that multivariate Observed vs. Expected (O/E) models and an index of biological integrity gave very similar results for assigning impairment with very similar precision associated with 4 different sampling methods. Multivariate methods may best be used for exploratory analysis to generate testable hypotheses while carefully chosen metrics used in a multimetric index can be successfully used in biomonitoring (Fore et al. 1996). All of these types of models perform better when the samples are being compared within a small geographic area, such as the same ecoregion or river system.

**Multimetric Indices of Biological Integrity (IBI)**

Complex ecological systems can be approximately described using well-chosen metrics in a multimetric model. Multimetric models use biologically derived indicators (metrics) to rate
stream conditions. A model becomes an index that uses distilled metrics (like species richness or ratios of certain taxa occurrence) that best characterize a particular assemblage over a gradient or between reference and impaired sites. "Indices of biological integrity" (IBIs) are used to detect trends over time at a particular site and for general screening of sites. Sometimes these models are used to confirm results of multivariate assessments and to confirm stressors. As with other bioassessments, IBIs are affected by physical (chemical and landscape), temporal and historical factors, as well as collection techniques. A set of metrics should be specifically tailored to each geographic area studied. This method has particular appeal because the result is a simple index that can be visualized and understood by most people.

**Metric assignment**

The metrics, which are measures of biological occurrence, are derived from the data, evaluated for response to different conditions and non-correlation, and then chosen for the model. A species list is categorized by population abundance and richness at different taxonomic levels and by other functional or tolerance traits. These are then analyzed alone, combined or organized in several ways, for instance as a percentage of sample. The range in metric values are examined for consistency and precision in distinguishing reference streams from known degraded streams or for distinguishing points along a gradient of interest with the least overlap of values.

The group of metrics are then refined to represent a balance of several categories of biological measures (e.g. richness, presence and absence of indicator species and trophic functions), eliminating correlated measures so as not to double-count species, and to include as many different factors as apply. Ideally the metrics should also connect several conditions in the biological system (not represented directly by the chosen metrics). For example the species used in the metric may not be so important alone, but the fact that they hold an important place in a trophic web (perhaps as primary food for another important species), or that they respond to and represent a physical condition like stream bed material or water chemistry provides insight.

Once metrics are chosen, a judgment is made about the divisions in the values that will be used to define the quality (best to worst) or the gradient. Numerical values are used as qualitative descriptions for condition and are assigned to metric value ranges that align with site quality or portion of a gradient. For instance the range of each metric value found at reference sites could be assigned a score of "10" and called "excellent", while the range found in highly
degraded conditions could be assigned "1" and called "poor." Often there are only 3 or 4 scores (ranges). This number assignment criteria would be the same for each metric. The assigned metric values of the chosen group (often around 9 or 10) of metrics are then added together to produce one dimensionless score that approximately describes or rates the condition of a site and allows for comparisons among sites using standard statistical approaches. The scoring methods, which can vary, set the divisions in the values and the original assumptions used (i.e. for distinguishing references sites or other variables) and can affect the quality of the model (Blocksom 2003). The concept and detailed instructions for creating a multimetric IBI are described in Karr et al. (1986).

There are many examples of multimetric IBI-type models that successfully distinguish biological differences over a gradient of anthropogenic disturbances (Wiseman 2003; Mebane et al. 2003; Baptista et al. 2007). The state of Idaho uses a stream macroinvertebrate index in conjunction with a fish and habitat index in their assessments (Grafe 2002). The index “correctly” classifies 94 percent of the stressed sites below the 25th percentile of least impacted scores and has been developed for several ecoregions in their state (Grafe 2002). As an example, one of the models for macroinvertebrates had 9 metrics: four richness metrics (total taxa, Ephemeroptera taxa, Plecoptera taxa, and Trichoptera taxa), percent Plecoptera, percent scraper and clinger taxa, Hilsenhoff Biotic Index (HBI) (to incorporate tolerance) and percent dominant taxa. In all of these studies, metrics had to be carefully selected for area specific sensitivity to gradients and correlations.

**Multivariate Models**

As the name implies, multivariate methods are designed for complex situations when many variables need to be analyzed simultaneously. There are several types of multivariate statistical approaches. These include classification-type analyses like RIVPACS (River Invertebrate Prediction and Classification System) which compare the expected occurrence of taxa with what is observed. There are also clustering and ordination analysis that are used to group communities by similarity and visualize patterns in complex datasets. These can reduce the data to fewer dimensions making them easier to interpret and represent graphically.

For all of these methods, it is important to choose the correct distance measure and method of transforming the data prior to analysis (McGarigal et al. 2000). Tests of significant differences among groups such as Multi-response Permutation Procedure (MRPP) or
Permutational Multivariate Analysis of Variance (PerMANOVA) can also be applied to more closely examine natural patterns and determine differences in community structure among sites.

Non-parametric multivariate methods are often used in biological assessment because these kinds of analyses can accommodate the complexity of ecological interactions and can manage data that are correlated and non-normally distributed. Multivariate models avoid experiment-wise error where significant results can arise by chance (type 1 error) when univariate tests are used on complex data (McGarigal et al. 2000). Multivariate statistics work well and are robust for ecological data including community data and other parameters (e.g. niche-space and taxonomic data) for several reasons. Some of the assumptions of parametric statistical tests like normality of the data can be ignored. Because most multivariate methods are permutative, they still function when some of the variables are correlated (as ecological variables sometimes are). Finally, it is not necessary to assume or assign any of the variables as strictly independent or dependent. These multivariate models inherently take into account environmental gradients due to physical parameters.

Complex ecological systems can be represented in a multivariate model but not completely explained (Fore et al. 1996). Multivariate models are used when trying to account for differences in species occurrence or assemblage characteristics that are the result of both measurable physical, biotic and historical factors (like past disturbance) and other important factors that may not have been measured or are not represented in the dataset. For community assemblages, the physical habitat variables and temporal patterns are only part of the explanation for why a suite of species might be found at a particular site. Also important are the interactions among members of the community and many other complex factors, some that may never be understood.

The models for community analysis are usually set up as matrices that are typically sample units vs. species and/or environmental parameters. Matrix algebra is used for the underlying similarity calculations (Fielding 2007). A common design for multivariate models uses the presence/absence or abundance of each species in a sample compared to a gradient of physical attributes at each sample site (in a second matrix).

Multivariate models are both descriptive and inferential at the same time. A descriptive method like Exploratory Data Analysis, (EDA), sometimes know by unflattering terms like “data mining” or “dredging” or "data driven analysis," is a method that helps find patterns that were not predicted a priori (Fielding 2007). Exploratory data analysis makes sense when science
accumulates large quantities of data that likely present opportunities to find new ecological patterns. It takes advantage of new methods that are very computationally intensive. In this way multivariate tests can be used to generate new hypotheses and also to explore the possible causes of a pattern. When using multivariate statistics inferentially, significant results in the descriptive function will reveal the set of variables that best explains the evidence against a null hypothesis (McGarigal et al. 2000). Later, univariate statistical techniques can be used to further explore and test the significance of the patterns that were revealed. Studies can be designed to use these techniques in complex systems for finding answers that cannot be performed any other way. Whether inferential or exploratory, multivariate methods can be very powerful tools that deserve continued study.

Many multivariate methods have been developed that accommodate different types of data in better or worse ways due to their theoretical underpinnings. Much care needs to be taken when using any of these methods, choosing the model in the first place and organizing and characterizing the data so the interpretation of patterns has validity. As the ease of access to higher computational power increases, research might benefit from more exploration into multivariate statistics. Important new patterns might be found for identifying details of community interactions that will help in our understanding of ecological processes and interpretation of changes.

**Observed vs. Expected (O/E) Models (RIVPACS)**

Many O/E multivariate models, which are easy to interpret and have proven to be valuable tools to regulating agencies, have been developed and used (Hargett, et al. 2005; Hawkins, C. 2004). These models use physical attributes (i.e. ecoregion, latitude, longitude, day number) at reference sites as predictor variables to calculate the probabilities of certain taxa being present. These probabilities are then used with test sites for comparison in order to assign impairment ratings. The observed measure at a site is used to create a ratio where a ratio of 1 means all expected taxa are present as a measure "taxonomic completeness." This type of test can be used to assign a "reference rating" to a site or express relative degradation. Washington State uses the RIVPACS (River InVertebrate Prediction and Classification System) model. There are other multivariate O/E models designed to compare assemblages that were based on or derived from the original RIVPACS model (which was created in the UK in the 1980's; Wright 1994) including PREDATOR (Predictive Assessment Tool for Oregon; Hubler 2005), BEAST (BEthnic Assessment of
Sediment) used in parts of Canada (Reynoldson et al. 1995), and AusRivAS, (Australian River Assessment Scheme; Schiller 2003). The differences include how impairment or difference in community is derived, for instance, RIVPACS detects loss of expected taxa and BEAST uses changes in community composition derived from the location in ordination space (Mazor et. al. 2006). These models are powerful tools that account for physical gradient effects on communities and can be kept for future samples.

**Ordination**

Another powerful multivariate tool that is used primarily for visual detection of relationships of communities and physical attributes is ordination. This family of multivariate techniques lies within a larger group of techniques that include cluster analysis and discriminant analysis. These methods express samples by observed occurrence of taxa in multidimensional “species space.” The distances between samples in this type of space express their dissimilarity. Ordination types include Principal Components Analysis (PCA), Correspondence Analysis, Canonical Correlation Analysis, Detrended Correspondence Analysis, Non-metric Multidimensional Scaling (NMDS) and others. Ordination can quantify relationships of a large number of variables (none considered dependant) into a meaningful arrangement of fewer dimensions (components). It maximizes the variance through many iterations, and creates vectors that attempt to show the source of the variation. Principal Components Analysis (PCA) analyzes variables for correlations and creates new derived variables, or components, in decreasing order of their contribution to the variance of the original set. Principal Components Analysis does not use distance measures and is used primarily for exploration of data (Fielding 2007) or for the ordination of non-ecological data (e.g., genetic markers). In ecological analysis, PCA can be used with physical data but not community data because it assumes linear relationships among variables (Plotnikoff and Wiseman 2001). Another technique, Factor Analysis, is similar to PCA but focuses on correlations rather than variances. Canonical Correlation Analysis (CCA or CANCOR) discovers relationships among sets of variables. It is an extension of multiple regression and involves two or more sets of variables, one treated as independent and the other set as dependant. CCA creates combinations of composite variables (from weighting) so correlation is maximized. It uses the redundancy of data (things that affect the processes at the same time that produce the same effect) to sort out what best explains the structure (the “major independent variables”). The goal of CCA is to find a few gradients that will explain most of the variation in the dataset, (including community
assemblages), so as not to lose too much information (Plotnikoff and Wiseman 2001).

**Non-metric Multidimensional Scaling (NMDS)**

One ordination technique that is especially suited to ecological data is non-metric multidimensional scaling ordination (McCune and Grace 2002). Non-metric multidimensional scaling ordination does not assume that the data are normal or that there are relationships among the variables. It can accommodate any distance measure desired, and uses ranked distances (Fielding 2007). This lessens the “zero–truncation” problem with heterogeneous samples (McCune and Grace 2002) which is the phenomenon that species will exist along a gradient but cannot be found at all beyond certain limits of the gradient. NMDS and other MDS (multi-dimensional scaling) techniques accommodate another problem with ecological data that other multivariate tests encounter which is when the number of variables are greater than sampling units. This is avoided because the distance measures in these techniques do not distinguish between x or y in the matrix. A potential problem is that in these matrices, nonoccurrence of a species could be used to define similarity. Many datasets (including the one I am using) contain many zeros for species occurrence and are termed “sparse,” but the effect of many zeros on the results of the ordination is not clear (McCune and Grace 2002).

Non-metric multidimensional scaling uses an iterative algorithm that tries to preserve the rank distances between samples using the term “stress” to characterize how the newly reduced dimensionality describes the distances in the original structure (Fielding 2007). A "scree" plot of stress values can be examined to determine where there is a break in stress, where more dimensions offer little relief of stress. Stress, or distortion in the distance measures that happens with lower dimensionality is unavoidable, although it is more desirable if most of the stress is lost in the fewest number of dimensions (Fielding 2007).

Ordinations like NMDS are computationally intensive and slow but this problem is less an issue as computers improve. Not all software programs offer NMDS ordinations, but the program PC-ORD does (McCune and Grace 2002). After an ordination, the graph module of the software can produce a 2- or 3-dimensional graph where it is possible to inspect the contribution of individual explanatory variables (Grandin 2006). The plot can show the relationship of the ordination axis with species to any quantitative categorical explanatory variable. The sample site points can be coded by the categorical variable from the second matrix. If sites appear separated in the species space by this code, there may be a real community differences and similarities
that can be explained using that variable.

**Multivariate Models: Important Considerations**

*Distance measures*

Multivariate techniques are used to optimally summarize, order or partition the dataset to see the structure and separation in the data (McGarigal et al. 2000). Data similarity is determined by assessing the "distance" between entities (i.e. sample units or sites in "species space"). There are several general types of distance measures that are used: Euclidean, City-block, Correlation coefficients and Chi-square. Distances measures emphasize different features of the data (McGarigal et al 2000). Euclidean Distance is the straight line distance between the points in species space (found using the Pythagorean Theorem in the number of dimensions used in the model). A variant of this is Relative Euclidean Distance which smoothes the data to eliminate the difference produced by total abundance, focusing on relative abundances instead. Other distance measures are termed City Block Distance measures where the distances between points are measured along a path in a grid (of the number of dimensions of the data). Important City Block measures include Bray-Curtis and Sørensen’s distance measures (McCune and Grace 2002). Correlation coefficients are also used to determine how similar points are. These measure the cosine of the angle between radii on which the points are lying (which is the measure of the arc). Finally, Chi-square distance is another measure often used in ordination techniques especially Detrended Correspondence Analysis and Canonical Correspondence analysis (CCA). Chi-square distance is weighted to proportionalize the differences in frequencies of entities in the data (Fielding 2007).

The Sørensen, Bray-Curtis and Jaccard similarity and the Kulczynski disimilarity measures are special cases of the City Block distance measure that include proportional coefficients. The Sørensen, and the almost identical Bray-Curtis distance measures represent the overlap of species abundances along an environmental gradient, and are determined as the shared abundance (determined by adding the absolute differences between the counts) divided by the total abundance. Converting this number to a "dissimilarity" measure produces the Sørensen distance measure which can also be used in ordination. Relative Sørensen measures allow each sample unit to be equalized in the analysis using proportions rather than total abundance. The Sørensen distance measure is not compatible with most multivariate analyses (Discriminate
Analysis, Canonical Correlation, and Canonical Correspondence Analysis) but can be used in Bray-Curtis and NMDS ordination. McCune and Grace (2002) see City Block measures like the Sørensen distance as intuitively more desirable to use for ecological community data because they measure the distance along the grid "edges" of multidimensional space where most points of sample units are found when plotted. Most graphs of species space will show many points (sampling units) near the origin where occurrence of species are zero or low, and points further out will be mostly near one axis or the other where abundance of one species dominates (what they term the "dust bunny" distribution; McCune and Grace 2002). City Block distance measures will embody and emphasize the importance of the environmental gradients that affect more species. Bray-Curtis distance measures ignore variables that have zero occurrence between two samples and emphasize the variables that have higher values. Another advantage of City Block measures is that they de-emphasize the influence of outliers (Fielding 2007). Because City Block methods measure the distance to and along the axis between points, the distances in species space are longer than if they were measured in Euclidean distance, and adding City Block distances will be proportionally higher (McCune and Grace 2002).

Transformation and relativization

The nature of species data which can be very heterogeneous within and between samples poses problems for analysis that can be solved partly with mathematical techniques. The community of organisms found at sites respond to a very complicated set of gradients, mostly environmental, but many unknown, which manifests as most species being not normally distributed along environmental gradients. PCA and DA assume linear responses to gradients. The Gaussian ideal distribution assumes a bell-shaped curve with responses having a predictable mean and standard deviation. The curves that represent species distributions do not often meet these assumptions for several reasons. Species may occur along a gradient in an asymmetrical manner, or they may be polymodel where the curve peaks in more than one place. Some species occurrences are not continuous through space - there are gaps where they are not found. Another problem is that species frequency may display in a graph as “solid” where points occur in a pattern spread all over the space under a bell shape. This happens when the conditions over a gradient are less than optimal for other reasons. Or the data may have the “zero-truncation problem” which is that they only occur in part of a gradient, and are just completely absent beyond some range (hence "negative" occurrences cannot be calculated; McCune & Grace
For these reasons, much ecological data benefits from transformation that may render a more symmetrical or linear response curve.

The complexity of species occurrence and the presence of rare species contribute much to the distance between sampling units (McCune & Grace 2002; Fielding, 2007) that may exaggerate or obfuscate important relationships between samples. But including rare species helps distinguish sites that are of higher quality and have higher richness (Thorne et. al 1999). Also, it is common for one species to be much more abundant than others so using count data can produce a very skewed picture. Therefore, data in the species matrix should be relativized in some manner to avoid the dominance of one factor (species or other variables) over the others.

There are many methods for transforming data. Different data respond to each method differently and should be a serious consideration in any analysis (McCune and Grace 2002). Transformations include logarithmic, exponential and other mathematical functions. However, datasets with many zeros cannot be transformed easily (as by log normal transformation; McCune and Grace 2002). There are smoothing functions and other adjustments that can be made such as deleting rare species, or adding a small constant to all zero values. Other options include adjustments to rank, standard deviation, or a variety of weighting functions.

Relativization to species maximum (usually in the columns of a matrix) evens out the rare and abundant species at a site. Relativization by site (usually in the rows of a matrix) will even out the differences in populations (sample size) among sites.

Hypothesis testing

There are many non-parametric statistical tests for distinguishing among groups in ordination space. These include analysis of similarity (ANOSIM), Qb method, and two commonly used tests, Multi-Response Permutation Procedures (MRPP), and Permutative Multivariate Analysis of Variance (PerMANOVA). Discriminant analysis (DA) and MANOVA are the parametric equivalents of these last two. However the non-parametric techniques do not assume multivariate normality or homogeneity of variances. Non-parametric methods test for the lack of significant difference between two or more a priori groups, (like between reference and non-reference streams and between different stream orders). Appropriate distance measures, transformations and relativizations should be used for each test. MRPP assumes that sample units are independent, and that appropriate weighting or relativization was performed prior to calculating an appropriate distance measure (McCune and Grace 2002). Variations of MRPP allow for blocked
designs and other experimental innovations. PerMANOVA allows for one-way, two-way and nested MANOVA.

The results of MRPP and PerMANOVA produce test statistics and “p” values, which help the researcher determine how likely the result would have been due to chance alone. MRPP produces a chance-corrected, within-group agreement statistic, A, which describes within-group homogeneity compared to random expectation (which would produce a value of A = 0; McCune and Grace 2002). The highest value for A is 10.0 which occurs when all items in a group are identical. In the highly heterogeneous samples common in community ecology, an A =0.3 is considered very high and it is common to find A < 0.1 coupled with a very low p value (which shows a significant difference among test groups). A negative A value indicates a within group heterogeneity less than would be found by chance. With community data, a large sample size may show a significant p value but with a relatively low A value. False statistical significance can arise when sample sizes are very large so careful examination and interpretation of results should be made.

Multi-Response Permutation Procedures (MRPP) calculate distances among all observations within each group and generates a weighted average of distances (weighted by the number of observations within each group). The distances represent the “signal” and the smaller the average distance, the stronger the signal. Next, MRPP generates "noise" by randomly shuffling the variables (within the same column) within the dataset. The program again uses the weighted average of distances within the random groups to re-calculate (this is equivalent to “noise”), and this reshuffling or permutation procedure is repeated until there is a distribution of average distances. Multi-Response Permutation Procedures then calculate the probability of randomly getting a smaller distance than the average distances for the true groups, which is the p-value.

Permutative Multivariate Analysis of Variance (PerMANOVA) is the “sum of squared distances between points and their centroid divided by the number of points” (McCune and Grace 2002). Using PerMANOVA avoids having to meet the assumption of linear species responses and normally distributed errors. It assumes that the rows and columns are independent and have similar dispersions (wider or narrower dispersions of similar data will appear as different).
Materials and Methods

Study Area
The Wenatchee Basin is located in the “Northern Cascades” ecoregion. The watershed area is approximately 3548 km² of mostly high elevation forests, but includes world-renowned fruit orchards and a few small cities. The watershed runs from high peaks through agricultural development and then through lower elevations where the vegetation becomes shrub-steppe. The annual rainfall varies in this watershed from less than 200 mm in the lowest elevations (City of Wenatchee) to over 3600 mm in the Cascade crest. Most of the watershed, approximately 3444 km² (Figure 1), drains into the Wenatchee River which eventually empties into the Columbia River. The rest of the terrain in this basin drains directly into the Columbia River.

The Wenatchee watershed begins with two main streams (Little Wenatchee and White) which drain the high Cascade Mountain peaks (some above 1700 m). This area is steep and

Figure 1. Map of study area showing sample sites of three studies
contains glaciers. In fact, snow pack and glaciers are the main source of most of the water for all the rivers in this basin (WRIA 45 planning unit, 2006). These high streams feed Lake Wenatchee, a natural lake and the origin of the Wenatchee River. Several major tributaries flow into the Wenatchee River beyond the lake. These include Icicle and Nason Creeks, Chiwawa River, Peshastin, Mission and Chumstick Creeks. Together these contribute to a total of about 370 stream km. The Wenatchee River enters the Columbia River near the city of Wenatchee.

The Wenatchee Basin is located entirely within Chelan County. Eighty-six percent of the land is under forest production or wilderness use. Over 80 percent of the land is under federal or state jurisdiction, in either Washington State Forest or National Forest including Wilderness (Alpine Lakes and Glacier Peak Wilderness). About 36 km² of the land in the middle and lower elevations along the Wenatchee River are agricultural. Most farms are pear orchards and some have been in operation for nearly 100 years; however, agricultural lands represent only about 1% of the total watershed area. There is a very small area of other agricultural land, mainly devoted to stock, agricultural support, recreational use, and small business. Roads or railroads also cross most of the basin both through public and the privately held land. The private land and the urban areas are mostly in the valley bottoms at the lower elevations. Much of the future development growth is expected in these flat areas. In addition to the city of Wenatchee, there are a few smaller municipalities, mainly along the Wenatchee River: Leavenworth, Cashmere, and the smaller communities of Peshastin, Monitor, Dryden and Plain. The population of the entire watershed is about 243,000, although some are part-time residents. There is an increasing number of vacation homes being built in the higher elevations. Overall, the Wenatchee basin is a very beautiful and mostly natural area prized by recreational enthusiasts.

A rich native fauna can be found in this watershed including some threatened or endangered species. The Wenatchee basin is the home to the peregrine falcon, bald eagle, northern spotted owl, marbled murrelet, lynx and the larch mountain salamander. Some of the healthiest fish runs in the Columbia River originate here. A report by the Upper Columbia Salmon Recovery board reports that the Wenatchee basin holds the greatest diversity of salmonids (sockeye salmon, steelhead, bull trout, spring and summer Chinook, and others; Hillman, 2004).

In general the Wenatchee basin is ecologically sound but there are problems that are causing increasing habitat loss and degradation. As the human population continues to grow, development is altering the environmental functionality of the stream channels and floodplains.
Human activities and structures, like the extensive road and rail system, negatively impact streams, for instance erosion from these causeways adds sediments to the stream beds (Upper Columbia Regional Technical Team 2008).

One of the most contentious and important resource issues in the Wenatchee Basin is the use of the available water. Several interests compete for the use of the water: urban uses, agriculture, fire protection, tourism as well as what is needed for a healthy natural ecosystem. The demand for water is highest in late summer when the flow is the lowest, especially in the lower elevation areas that receive little precipitation. Agriculture, with its extensive and old canal system, and residential development in valley bottoms, has historically used withdrawals beyond the flow needed to insure that streams remain viable for wildlife. The WDOE has designated this basin, which is also known as Washington State Water Resource Inventory Area 45 (WRIA 45), as over-appropriated; having flows that are at times inadequate to support fish. This is a legal, as well as an environmental problem. The Wenatchee watershed is part of the land the Yakama Nation ceded to the United States in the Treaty of 1855. As part of this treaty, the Yakama have the right to “usual and accustomed” uses of the lands and waters for hunting and fishing. Therefore, the tribes argue that streams must have enough flow to support fish, including in upstream reaches. Unfortunately, the watershed, already stressed by these competing demands, may be stressed further by future loss of available water due to changes in climate.

Climate change models show that the Pacific Northwest will continue the trend of the last 100 years by becoming increasingly warmer and wetter (Mote and Salathé 2009) which should worsen the situation for future water resource use and protection. Increased precipitation is expected to fall during the normal rainy season, but less of it will fall as snow because of higher average temperatures. This may cause excessive flow or flooding during the rainy months and decreased snowpack. Decreased snowpack is also a problem because the melting snow provides stream flow in the summer when there is little rain. Spring melt may occur one to two months earlier with a similar delay in the fall for the return to normal flows. It is estimated that snowpack may be reduced by 28% in the next 10 years (Littell et al. 2009). Exacerbating this situation will be the warmer summers that will increase evapotranspiration and demand for irrigation for the orchards. Future development and resource protection will have to compete in an environment of decreasing water supply.

Monitoring projects could help detect trends that will guide efforts for management of
the water resources here. In the next section, two efforts that are providing data and studying trends in the Wenatchee Basin are described. Information from reference streams will define the goals for restoration or preservation and also document natural changes.

Washington State Bioassessment
Washington State Department of Ecology (WDOE) has been using bioassessment increasingly in recent years for monitoring and enforcement of the Clean Water Act. Collections and descriptions of biotic assemblages have been performed by the WDOE since 1993. The most common type of model created and used by WDOE with bioassessment are multimetric indexes of biological integrity (IBI) or observed/expected multivariate models like RIVPACS. A few multimetric models have been developed for a few of the approximately nine, level III ecoregions identified in the state (e.g. Puget lowlands and Cascades by Wiseman 2003). Since ecoregions are the foremost category used to predict a similar range of biotic occurrence in streams (Wiseman 2003; Omernik and Bailey 1997), distinct models are needed for each ecoregion. Following, two bioassessment studies that WDOE has conducted in the Wenatchee Basin are briefly described.

Data acquisition
The data analyzed in this thesis were obtained from two separate studies of biological assessment being conducted by WDOE from two unrelated projects, the Integrated Status and Effectiveness Monitoring Project (ISEMP) and the Environmental Monitoring and Assessment Program (EMAP) from the Wenatchee basin, or WRIA 45. The majority of the data is from the ISEMP project managed by National Oceanic and Atmospheric Administration (NOAA Fisheries Service), and funded by Bonneville Power Authority (BPA). The ISEMP project is a pilot project and includes three basins (Wenatchee, John Day and South Fork Salmon River basins). It was initiated in 2003 in response to NOAA's 2000 Biological Opinion, a document that guides federally owned dams regarding salmon and steelhead recovery. The purpose and design of ISEMP is to monitor fish populations and habitat and to test monitoring protocols, sampling designs and indicator metrics. Another purpose of ISEMP includes trend monitoring and effectiveness monitoring for habitat restoration projects (Merritt 2006). The project sampled fifty sites each year; twenty-five of these were randomly chosen and sampled once each year of the study along with twenty-five new randomly chosen sites. The WDOE collected data for
habitat quality, channel condition, riparian condition and reach characteristics using the protocols from Hillman (2004). The survey plan was specifically designed to be used in bioassessment. The design and protocols continue to be evaluated and improved upon with the intention of becoming the standard for the state. The samples for this thesis were taken in the years 2004 – 2007. The other study from which WDOE provided data was the EMAP Western Pilot (2000-2003). This is a federally directed assessment of 12 states and tribal lands which, in Washington, is partially administered by the WDOE (Washington DOE 2011). The study is one that attempts to assess stream conditions using extrapolation from randomized representative steams. Data collected included biological, chemical, and physical habitat information (Stoddard et al. 2005). One focus region in the EMAP project is the Wenatchee River Basin (WRIA 45) - 44 out of 90 sites from this study are within this basin. The plan for the project was to use an Observed/Expected multivariate model to assess the conditions of Washington state’s wadeable streams.

The EMAP and ISEMP studies were conducted in a way that made combining their data possible. The guidelines for describing the physical parameters were the same and they both used the same collection protocols; specifically ISEMP follows the EMAP project’s design (Hillman 2004). The collection season was from July 1 to Sept 30 of each year. The macroinvertebrates that were collected from both studies were delivered to a lab (Terraqua, Inc.) that was contracted to NOAA-Fisheries. This was where the sub-sampling and macroinvertebrate identification took place.

*Data collection methods*

Macroinvertebrate collections in these studies were made in wadeable perennial streams by the “targeted riffle” sample protocol for EMAP (Hillman 2004). Collections were made from 1 ft² kick samples collected randomly from up to 8 different riffles in a reach. The samples were consolidated into a composite sample for each reach. A reach was defined as a length of stream 20 times the bank full width (150 m minimum to maximum 500 m). The protocol that was followed takes into account how to sample to avoid disturbance in the area prior to sampling, where to sample in the riffles and what to do if there are fewer than 8 separate riffles in a reach. The method included holding the kick net steady, manually cleaning off each rock larger than a golf ball so any insects flow into the net, and visually checking the rock before placing it out of the area. The protocol directed the collector to kick a 1 ft² area
above the net for 30 sec. Sampling was continued until the net contents impeded flow, then the net was emptied into a container holding all the samples and the sampling continued for 8 ft² per reach. At the end of sampling a reach, the net was cleaned thoroughly into the collection, using tweezers if necessary. The combined samples were preserved in 70% ethanol. The samples from each reach was then sent to a lab that would randomly identify at least 500 benthic macroinvertebrates from each sample (Moberg 2007).

The database from ISEMP and EMAP had hundreds of sample sites. For this thesis sites that were in the Wenatchee Basin were sorted by GIS and used for this analysis. Many sites were rejected due to very small or very large samples sizes or lack of accompanying physical attribute data. Also rejected were duplicate samples that were taken in the same month and year at the same site. This left over 183 useable sample sites for analysis.

The data from the Wenatchee Basin were analyzed by physical and community metrics both as a whole set and also broken into three smaller groups that compared variables within and between. One of the smaller groups was the data from the EMAP study and the other two were from the ISEMP study (WC and WEN) and were separated by the coding used in the dataset. The dataset contained both biotic and physical attributes, but no explicit habitat descriptions.

**Descriptive Attributes of Sites**

Physical, temporal and qualitative attributes used in this analysis included date of collection (month and year), elevation, latitude, longitude, watershed area, slope, mean annual precipitation and sinuosity. Stream order was determined using GIS maps with the sample points and stream coverage. Two sets of reference/non-reference site designations were added. One set was chosen from a list of sites that were identified as "Reference" sites by DOE and were used to make the groups Reference and Non-Reference. The other set of sites were created using each site's location in or outside of the National Wilderness Area boundaries as a surrogate Reference, assuming that the protection of that designation would produce higher quality conditions. The WDOE deemed some sites in the National Forest as non-reference and there were many more DOE "reference" sites identified outside of this enclosure so the "wilderness reference" sites were fewer.
Community data

The samples of benthic invertebrates were mostly categorized at the species or genus level in the database. This taxonomic identification allowed for re-designation of taxa at family and order levels, by tolerance values and functional feeding group assignments (as both primary FFG and a primary/secondary designation). Community data were used as richness and abundance of different taxonomic levels and a composite of macroinvertebrate orders E, P, T, C and D, FFGs, and by tolerance values.

Collectively there were 376 separate species (some were with higher order designations but were counted as a separate species). The samples averaged 500 macroinvertebrates each from subsampling done at a laboratory. In all there were 91,354 macroinvertebrate specimens enumerated. Enumerating the taxonomic designations to higher levels reduced the data complexity to yield 86 families and 22 orders: Amphipoda, Basommatophora, Coleoptera, Copepoda, Diptera, Ephemeroptera, Haplotaxida, Lepidoptera, Lumbriculida, Megaloptera, Nematoda, Nematomorpha, Oligochaeta, Ostracoda, Plecoptera, Rhynchobdellida, Sarcoptiformes, Trichoptera, Tricladida, Trombidiformes, Turbellaria, Veneroida. Functional feeding groups were divided into 8 primary trophic divisions: Filterer-collector, Gatherer-collector, Omnivore, Parasite, Piercer, Predator, Scraper and Shredder.

Data analysis

The goal of this project was to find some meaningful patterns using exploratory data analysis with multivariate ordination models. The first major exploration involved reassigning the species list into several alternate designations to be used as surrogate for a species list. Statistical features in the ordination program PC-ORD (5.32) were used to determine if any of these biological metrics could predict assignment of a stream to reference condition status. Also tested were how these various re-designated assemblages differentiated along environmental gradients. While not identifying a specific stressor responsible for impairment, the trends found could help improve the assumptions that go into models for assessing stream condition.

Functional Feeding Group and Tolerance

Assignment into a functional feeding group (FFG) can be made at multiple taxonomic levels and therefore almost none of the species in the samples were omitted for lack of information. The designations of feeding group and tolerance values for each species or genus were found in
multiple sources. The FFG designations were obtained from Merritt et al. (2008) or Barbour et al. (1999), and describe the dominant role of each species in North America. Tolerance values were specifically calculated for the Pacific Northwest and found in Merritt et al. (2008) or, where an entry was missing, additional designations were found in Barbour et al. (1999) from the EPA website (http://water.epa.gov/scitech/monitoring/rsl/bioassessment/).

Richness measures
The focus in this analysis of macroinvertebrate assemblages was on the presence, proportions and richness of macroinvertebrates as metrics for describing stream conditions. Richness measures were used in three ways for analysis. First, the data was characterized completely by species present as abundance. This allowed for the calculation of total richness, and also Shannon's diversity index (Haurer and Resh 2006). Second, the richness measure of the macroinvertebrates by taxonomic order level was calculated for 5 important orders, E, P, T, C and D and again for all orders present. Third, the richness of FFG’s were calculated by the number of distinct species that could be categorized by a primary FFG and also by the addition of secondary FFG assignment.

Data characterization and transformation
The data were organized as two matrices, one of sites and their invertebrate abundances and one of sites and their physical attributes. These are used together in the ordination and associated statistical tests. In addition to using the species matrix as collected, raw species were transformed and condensed into sub-groups. The groups used for the analysis were: entire species list including rare taxa count (total abundance), Species presence/absence, Order presence/absence, Order abundance and Order richness, Family abundance and Family presence/absence, FFG richness, both with and without a secondary designation, and a group by tolerance scores as abundance. For analyzing the entire species list, these data were relativized by maximum of column (species) totals to smooth out the differences between very common and rare species. This means that for each species, the highest occurring value became "1" and the other values were represented proportionally less than one. When the coefficient of variability, or 100*(standard deviation/mean), is very large (> 300), it is highly recommended to make some relativization or transformation of data (McCune and Grace 2002). These data showed very high CV (CV > 500) for the entire species list due to the contrast between the high
number of absent and/or rare species and the large populations of common ones in most of the samples. Rare species were not eliminated to preserve diversity. Coefficients of variation at taxonomic levels above species were lower: order CV = 250 (22 Orders), family CV = 382 (86 families). Therefore the family abundance was relativized but not the order abundance. None of the other designations warranted any transformation.

Environmental data were not used in this analysis the same way as the species data which were used to describe the similarity of sites. The physical and temporal data in the multivariate model were used to group sites either in categories or along gradients. A few important environmental variables were used for this purpose including elevation, stream order and reference designation. Physical data were also used to explore some other bivariate relationships between the sampling sites.

Considerations of Characterizations

When analyzing the samples by full species designations, the similarities among sites might have lower values because none of the rare species were excluded. Different rare species showing up in each sample will make the assemblages appear much more different than if they were ignored. Weighting and transforming the species abundances, which was done in this analysis, can help with this potential problem. The bias produced by transformation does not negate any significant dissimilarity (McCune & Grace 2002). Issues that may lead to skewed or inaccurate conclusions about sample similarity include the way these data were sampled and recorded. More revealing than mere counts or proportions, “biovolume” measurements of the various groups and characterization by life stages (when they are largest and feeding for instance) were not recorded. A minor problem in these data was that some of the macroinvertebrates were identified at a much higher taxonomic level than others. Each entry recorded at higher levels (than species) was counted as an additional new species. Any effect on richness values from this were negligible because any false increase in richness (if the named organism matched an already identified one) would be balanced by the loss of richness by characterizing several new and different species by the same higher taxonomic designation.

Ordination

Non-metric multidimensional scaling (NMDS) ordination was performed on the matrices using PC-Ord (5.32). This type of ordination uses iterations and rankings to analyze sites by species
composition. The ordination produces a graph in two or three dimensions that illustrates the
similarity of entities. Individual rows of the main matrix become points in ordination space. In
this analysis, sites were plotted as points in a distilled two dimensional axis of "species space."
Sites that are closer together on the graph are more similar in species composition than samples
that are farther apart. This type of ordination is “constrained,” compared to unconstrained
ordination which gives patterns with no explanation for them. In NMDS it is possible to constrain
a set of variables by their relationship to another set which can give clues about the structure.
The constraining set of variables (in this analysis, the second set of matrices of physical variables)
is used to describe the first set (the population at each site). The sample site points can be
visually coded by a categorical variable from the second matrix. If sites appear separated in the
species space by this code, there may be a real community differences and similarities that can
be explained using that variable. If there is a strong effect in composition resulting from an
environmental gradient described in the second matrix, a vector will be drawn and labeled. For
instance, if species composition varies in a predictable way among sites along an elevation
gradient, an arrow is drawn pointing in the direction of increasing elevation in the cloud of
sample site points.

For the ordination in PC-ORD the "autopilot" setting on medium was used with Sørenson
(Bray-Curtis) distance measure which is commonly used for community data (McCune and
Mefford 1999) and species data that were transformed by relativizing by species maximum. This
setting uses 50 runs with real data (starting with a random configuration each time), stepping
down from 4 axes (dimensions) to one. The best starting configuration that produces the least
"stress" (for each dimension) from the real runs is saved to disk. Stress is a measure of the
difference in the ranked distances between entities in the original matrix "of column times rows"
dimensions and the distances in a reduced dimensional matrix. Then, NMDS in PC-ORD performs
250 runs with randomized data, shuffling the data within columns and using a different random
starting configuration before each run and collects these statistics. Next, the software chooses
the best (lowest stress) solution for each dimensionality from the real data. At each
dimensionality, the final stress must be lower than that for 95% of the randomized runs (i.e. p <=
0.05 for the randomization test). The stability criterion is 0.00001 and uses 15 iterations to
evaluate the stability. Instability is calculated as the standard deviation in stress over the
preceding iterations (15 in this case) (documentation from PC-ORD 5.32).
MRPP and summary statistics

Information from the second matrix can be used for parsing out differences in assemblages due to the physical placement of the streams as well as the designations like reference or non-reference. PC-ORD was used to calculate the multi-response permutation procedure (MRPP) for testing particular variables for significance in separating assemblages. This requires choosing categorical variables from the second matrix to create groups that can be compared. A p-value is produced that describes the likelihood that the similarities found in the assemblages (grouped by categorical variables) are not random. These groups were then considered insignificantly different. An "A" value is also produced in MRPP that expresses within-group agreement. For this thesis the variables used were reference/non-reference (two sets, reference chosen by different methods), Strahler stream order, elevation, month, year, slope, and divided into the three smaller studies (EMAP, WC and WEN). The community data groups categorized as richness and abundance at different taxonomic level were weighted by n/sum(n) for these MRPPs.

Diversity and evenness metrics, which are calculated by the PC-ORD software, were used along with other summary statistics for characterizing and comparing the physical variables and the community data with univariate statistics. Physical variables like elevation, stream order and Reference and Non-Reference were used to separate sites for comparison of some of the community metrics like richness and abundance. This was done for all the sites together and also when broken into the three smaller groups.

First, physical variables were compared with each other, and then with community data using univariate statistics (first as one combined group then separately for each of the three sub-projects). In addition, data were compared using the reference and wilderness designations. This was done to characterize the differences, if any, of some of the important variables. Second, NMDS ordinations were created to explore which factors might influence community structure. Third, MRPPs were used to test for significant differences in community structure based on physical or reference variables (first as one combined group then separately for each of the three sub-studies).
Results

Physical and community metrics

Although some of the physical variables at the sites were correlated, many (but not all) of the community statistics appeared independent of these variables because they did not react to all the variables. As one would expect, decreasing elevation was correlated with stream order increase \( (p < 0.0001) \), watershed area increase \( (p < 0.0001) \) and precipitation decrease \( (p = 0.0001) \). Effects on the macroinvertebrate community due to elevation, precipitation, watershed area and stream order were mixed. But in general, total macroinvertebrate richness did not differ with elevation \( (p = 0.26) \), stream order \( (p = 0.34) \) or watershed area \( (p = 0.12) \).

Richness of the macroinvertebrate orders Ephemeroptera (E) and Trichoptera (T) did not vary with elevation or stream order; however, Plecoptera (P) richness responded to this variable, increasing with increased elevation \( (p < 0.0001) \), and decreasing stream order \( (p = 0.0002) \) (Table 1). The effect of P richness was enough to drive variation when combined as EPT richness which increased as elevation increased \( (p = 0.006) \) and appeared as a curve (peaking at stream order 4) for stream order \( (p = 0.03) \). For the macroinvertebrate orders tested as relative abundance, %E decreased with decreasing watershed area \( (p = 0.03) \). For stream order and elevation, %E, %P, %T, and %EPT differed; %E increases with stream order \( (p = 0.03) \) but did not change significantly by elevation. %P decreases with decreasing stream order \( (p = 0.0001) \) and with higher elev. \( (p = 0.006) \). For %T and %EPT there was no significant influence of stream order or elevation but %EPT decreased with increasing watershed area \( (p = 0.035) \). In addition, %T increased as the month of sampling increased \( (p = 0.0019) \).

Average tolerance did not show significant differences between wilderness vs. non-wilderness sites \( p = 0.35 \) and was only close to being significant between reference and non-reference sites \( p = 0.11 \). This may be another clue that wilderness sites might not mimic reference sites because macroinvertebrates with higher tolerance would be expected to be more abundant in compromised sites. Average tolerance did not show significant differences over the gradient of elevation, between stream orders, slopes or between the three studies but did show a significant decrease over months \( (p = 0.044) \) potentially reflecting some seasonal differences in macroinvertebrate populations. Average tolerance appeared to increase somewhat with watershed area but the ANOVA test was not significant \( (p = 0.19) \).
Table 1. Richness and relative abundance of some macroinvertebrate orders and average tolerance related to the divisions in stream order, elevation, month and study. Asterisks (*) denote significant differences based on regression or ANOVA. Groups denote the division by the three smaller studies, EMAP, WC and WEN.

<table>
<thead>
<tr>
<th>Richness</th>
<th>Stream Order</th>
<th>Elevation</th>
<th>Months</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Richness</td>
<td>p = 0.34</td>
<td>p = 0.26</td>
<td></td>
<td>p = 0.67</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>increase p = 0.0001*</td>
<td>decrease p = 0.0001*</td>
<td></td>
<td>p = 0.005*</td>
</tr>
<tr>
<td>Diptera</td>
<td>p = 0.98</td>
<td>p = 0.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ephemeroptera</td>
<td>p = 0.29</td>
<td>p = 0.59</td>
<td></td>
<td>p = 0.52</td>
</tr>
<tr>
<td>Plecoptera</td>
<td>decrease p = 0.0001*</td>
<td>increase p = 0.0001*</td>
<td></td>
<td>p = 0.48</td>
</tr>
<tr>
<td>Trichoptera</td>
<td>p = 0.11</td>
<td>p = 0.184</td>
<td></td>
<td>p = 0.93</td>
</tr>
<tr>
<td>EPT</td>
<td>p = 0.02* peak at 4</td>
<td>increase p = 0.006*</td>
<td></td>
<td>p = 0.96</td>
</tr>
<tr>
<td>Trombidiformes</td>
<td>increase p = 0.0505*</td>
<td>p = 0.6</td>
<td></td>
<td>p = 0.023*</td>
</tr>
<tr>
<td>Average tolerance</td>
<td>p = 0.66</td>
<td>p = 0.87</td>
<td>decrease p = 0.044</td>
<td>p = 0.89</td>
</tr>
</tbody>
</table>

Relative Abundance

| %Ephemeroptera   | increase p = 0.03* | p = 0.15 |           | p = 0.89 |
| %Plecoptera      | decrease p = 0.006* | increase p = 0.006* |           | p = 0.77 |
| %Trichoptera     | p = 0.88 | p = 0.20 | increase p = 0.002* | p = 0.89 |
| %EPT             | p = 0.68 | p = 0.73 |           | p = 0.94 |
| %Coleoptera      | increase p = 0.0001* |           |           | p = 0.005* |
| %Trombidiforms   |           |           |           | p = 0.023* |

Physical, temporal and community characteristics of the 3 smaller studies

The 183 sites were separated into three smaller groups which brought out some unique characteristics like physical variability and community composition response. The groups were divided by study effort, "EMAP," "WC" and "WEN" which had 42, 70 and 71 samples sites, respectively. The years of sampling differed: EMAP did not overlap with the other 2 groups and was sampled from 2000 (1 site) to 2003. WC was sampled from 2004 to 2006 and WEN samples were collected from 2005 to 2007. The number of distinct sites that were sampled also differed among the groups and may have contributed to the differences or similarities in community
metrics for these groups. Each of the 42 EMAP sample sites were in a different location. The sites of WC were almost all sampled multiple years, usually about 3 years each, and contained approximately the same number of sites for each year (only 2 of the 70 sites were sampled once). In the end, WC yielded only 25 different sites. One third of the 71 WEN sites were sampled more than once (there were 58 distinct sites), and of these, most were sampled twice and many of these were the same year but on a different date. The average elevations for EMAP, WC and WEN were 883 m, 701 m and 754 m, respectively, with annual precipitations of 1683 mm, 1110 mm and 1171 mm, respectively. The sites also varied in the same manner for average watershed area (48 km², 252 km² and 140 km², respectively). In terms of the watershed area and the sampling strategies of each study by sampled stream orders, the EMAP sites are much more balanced than the other two and contain the most 1st and 2nd order streams (Table 2). The WC and WEN sites sampled a higher percentage of 5th and 6th order stream sites and relatively few 1st or 2nd order sites. Despite these differences, these sites appear well mixed in physical distribution over the study area (Figure 1).

Table 2. Stream order composition by study

<table>
<thead>
<tr>
<th>Stream Order</th>
<th>EMAP</th>
<th>WC</th>
<th>WEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st and 2nd</td>
<td>21.43%</td>
<td>5.71%</td>
<td>8.45%</td>
</tr>
<tr>
<td>3rd</td>
<td>26.19%</td>
<td>15.71%</td>
<td>19.72%</td>
</tr>
<tr>
<td>4th</td>
<td>21.43%</td>
<td>20.00%</td>
<td>30.99%</td>
</tr>
<tr>
<td>5th and 6th</td>
<td>30.95%</td>
<td>58.57%</td>
<td>39.44%</td>
</tr>
</tbody>
</table>

Although the physical characteristics differed, there were few differences among studies in simple community metrics. Average species, order and family richness values were similar among studies. When the community metrics were compared between the three smaller studies (EMAP, WC and WEN), there were no significant differences except for the richness and relative abundance of coleoptera and the richness of Trombidiforms.

For the EMAP, WEN and WC studies, there were respectively 267, 270 and 274 different species, 62, 61 and 61 different macroinvertebrate families and 16, 19 and 17 different orders represented (p > 0.05). Among studies (see Table 1), there were many other simple community
metrics that also did not differ: %E, %P, %T, %EPT, average tolerance, total richness, nor richness for E, P, T or EPT (p > 0.05). But there was a difference in richness of Coleoptera, (p = 0.005) and Trombidiiforms (p = 0.023), and % Coleoptera (p<.005) for which the averages for both were highest in the WC study and lowest in the EMAP study (Table 3).

Table 3 Average richness and abundance of macroinvertebrate orders in EMAP, WC and WEN groups (number of different groups or number of individuals).

<table>
<thead>
<tr>
<th></th>
<th>Richness</th>
<th></th>
<th></th>
<th></th>
<th>Abundance</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>P</td>
<td>T</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>EMAP</td>
<td>1.4</td>
<td>16.7</td>
<td>10.7</td>
<td>8.7</td>
<td>9.9</td>
<td>13.9</td>
<td>114.6</td>
<td>170.5</td>
</tr>
<tr>
<td>WC</td>
<td>2.9</td>
<td>15.9</td>
<td>10.5</td>
<td>7.7</td>
<td>9.7</td>
<td>36.5</td>
<td>110.1</td>
<td>168.7</td>
</tr>
<tr>
<td>WEN</td>
<td>2.4</td>
<td>16.7</td>
<td>10.1</td>
<td>8.6</td>
<td>10.0</td>
<td>27.9</td>
<td>114.6</td>
<td>156.2</td>
</tr>
</tbody>
</table>

Reference vs. non-reference, wilderness vs. non-wilderness
Reference sites and wilderness sites showed some clear differences in physical site variables in addition to their comparatively undisturbed condition. Both reference and wilderness designations were much more common in higher elevations. Watersheds of over 150 km² found at lower elevations had no reference or wilderness sites at all. Reference and wilderness sites also differed significantly from non-reference and non-wilderness sites by stream order composition (p < 0.001 for both). Some physical relationships between the reference and non-reference sites are that reference sites tend to have higher precipitation (mm), are smaller in area (km²), and are comprised of lower stream orders (Table 4).

Reference sites within studies
Each group had approximately the same percentage of reference sites for both reference designations (Table 5). Wilderness sites were less abundant than reference sites. Wilderness sites represented 29%, 24%, and 26% of the total for WAP, WEN, and WC studies, respectively. Reference sites represented 33%, 34%, and 36% of the total for WAP, WEN, and WC studies, respectively.
Table 4. Comparison of physical attributes between reference and non-reference and wilderness and non-wilderness designations for the EMAP, WC and WEN studies. All pairs are significantly different except WEN stream order and EMAP watershed area and stream order for both reference and wilderness.

<table>
<thead>
<tr>
<th></th>
<th>WC</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reference</td>
<td>Non-Reference</td>
<td>Wilderness</td>
<td>Non-Wilderness</td>
</tr>
<tr>
<td>Avg. Precip mm</td>
<td>1683</td>
<td>830</td>
<td>1878</td>
<td>845</td>
</tr>
<tr>
<td>Avg. Area km$^2$</td>
<td>29</td>
<td>361</td>
<td>34</td>
<td>328</td>
</tr>
<tr>
<td>Avg. Str. Order</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Avg. Elev. M</td>
<td>957</td>
<td>577</td>
<td>1006</td>
<td>596</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>WEN</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reference</td>
<td>Non-Reference</td>
<td>Wilderness</td>
<td>Non-Wilderness</td>
</tr>
<tr>
<td>Avg. Precip mm</td>
<td>1305</td>
<td>1103</td>
<td>1600</td>
<td>1037</td>
</tr>
<tr>
<td>Avg. Area km$^2$</td>
<td>34</td>
<td>194</td>
<td>50</td>
<td>168</td>
</tr>
<tr>
<td>Avg. Str. Order</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Avg. Elev. M</td>
<td>904</td>
<td>677</td>
<td>982</td>
<td>681</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>EMAP</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reference</td>
<td>Non-Reference</td>
<td>Wilderness</td>
<td>Non-Wilderness</td>
</tr>
<tr>
<td>Avg. Precip mm</td>
<td>2027</td>
<td>1492</td>
<td>2265</td>
<td>1451</td>
</tr>
<tr>
<td>Avg. Area km$^2$</td>
<td>24</td>
<td>62</td>
<td>29</td>
<td>56</td>
</tr>
<tr>
<td>Avg. Str. Order</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Avg. Elev. M</td>
<td>1053</td>
<td>789</td>
<td>1108</td>
<td>793</td>
</tr>
</tbody>
</table>

Table 5. Number and percentage of references and wilderness sites within each study.

<table>
<thead>
<tr>
<th>Total sites</th>
<th>Group name</th>
<th>No. of sites</th>
<th>Percentage of sites</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reference</td>
<td>Wilderness</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>WAP</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>71</td>
<td>WEN</td>
<td>24</td>
<td>17</td>
</tr>
<tr>
<td>70</td>
<td>WC</td>
<td>23</td>
<td>18</td>
</tr>
</tbody>
</table>
Comparison of community metrics between reference and wilderness designations

Total richness of species did not differ between the reference and non-reference sites or the wilderness and non-wilderness sites by using ANOVA.

When broken into the 3 separate groups, total richness was again not different between reference or wilderness designations (Table 9). Again, no differences in any community metrics were seen between the Wilderness and non-Wilderness sites. The reference designation had mixed results when examined by study and two groups, EMAP and WEN, showed significant differences in all three remaining metrics while the group WC showed none. This pattern was repeated within studies when tested with MRPP for other community metrics.

The response of some community metrics showed differences between the two reference designations and their associated non-reference sites even though total richness and the richness and percentage of some individual orders were the same. The metric of total richness was heterogeneous enough in all tests performed to show no significant differences in any pair of variables used. But the community metrics of evenness, and both diversity measures showed significant differences for reference, but none for wilderness (Table 6).

Reference sites differed from non-reference sites by having higher richness of Plecopteran and Trichopteran families and lower Coleopteran families, while wilderness and non-wilderness sites varied by also having higher richness of Plecopteran families and lower Coleopteran families, but also showed lower richness of Dipteran families (Table 7.) The same diversity metric H (p = 0.02) differed significantly (D’ showed the same, but non-significant trend, p = 0.076), and all values of evenness and diversity were higher for reference sites. There were no significant differences between any of these metrics for sites designated as wilderness or non-wilderness.

Functional Feeding Group (FFG) Richness

Categorizing the macroinvertebrates by their main functional feeding group for richness was not adequate to distinguish sites as reference or non-reference or as wilderness or non-wilderness by ANOVA, but there was a significant effect seen in FFG composition by stream order (Table 8).
Table 6. Community metrics differing in reference/non-reference and wilderness/non-wilderness sites. Species richness (S), evenness (E), Simpson’s diversity index (D’) and Shannon’s diversity index (H) values are shown. Asterisks (*) denote significant differences based on ANOVA.

<table>
<thead>
<tr>
<th></th>
<th>Reference/Non-reference</th>
<th>Wilderness /Non-wilderness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p = 0.27</td>
<td>p = 0.61</td>
</tr>
<tr>
<td>S</td>
<td>p = 0.01*</td>
<td>p = 0.27</td>
</tr>
<tr>
<td>E</td>
<td>p = 0.08</td>
<td>p = 0.51</td>
</tr>
<tr>
<td>D’</td>
<td>p = 0.02*</td>
<td>p = 0.53</td>
</tr>
</tbody>
</table>

Table 7. Differences in order-level richness values for two reference designations and their non-reference counterparts. Asterisks (*) denote significant differences based on ANOVA.

<table>
<thead>
<tr>
<th></th>
<th>Reference/Non-reference</th>
<th>Wilderness /Non-wilderness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p-value</td>
<td>p-value</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>Lower 0.0001*</td>
<td>Lower 0.0001*</td>
</tr>
<tr>
<td>Diptera</td>
<td>0.21</td>
<td>Lower 0.022*</td>
</tr>
<tr>
<td>Ephemeroptera</td>
<td>0.35</td>
<td>0.09</td>
</tr>
<tr>
<td>Plecoptera</td>
<td>Higher 0.0038*</td>
<td>Higher 0.03*</td>
</tr>
<tr>
<td>Trichoptera</td>
<td>Higher 0.0049*</td>
<td>0.12</td>
</tr>
<tr>
<td>Trombidiforms</td>
<td>0.72</td>
<td>0.59</td>
</tr>
</tbody>
</table>

The richness of individual FFGs provided mixed but more useful results. Reference sites had significant differences in the richness of the FFGs of Filter-Collectors (FC), Gatherer-Collectors (GC), Omnivores (OM), Parasites (PA) and Predators (PR); but no significant difference for Scrapers (SC). Only FC, GC and PR richness were able to distinguish wilderness sites from non-wilderness. Reference appeared to be a much better division to detect differences in FFG richness than the wilderness designation. All but SC and GC were significantly different among stream orders.
Table 8. Functional Feeding Group (FFG) richness differences for reference or wilderness designations, and between stream order categories (1 - 6). Feeding groups are filterer-collector (FC), gatherer-collector (GC), omnivore (OM), parasite (PA), piercer (PI), predator (PR), scraper (SC) and shredder (SH). Asterisks (*) denote significant differences based on ANOVA.

<table>
<thead>
<tr>
<th>RICHNESS</th>
<th>Wilderness Reference</th>
<th>Reference</th>
<th>Stream Order</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F (1,181) p-value R²</td>
<td>F (1,181) p-value R²</td>
<td>F (5,177) p-value R²</td>
</tr>
<tr>
<td>FFG1</td>
<td>0.34 0.56</td>
<td>1 0.32</td>
<td>2.53 0.030* 0.06</td>
</tr>
<tr>
<td>FC</td>
<td>5.7 0.018* 0.03</td>
<td>9.61 0.002* 0.05</td>
<td>4.04 0.002* 0.10</td>
</tr>
<tr>
<td>GC</td>
<td>8.66 0.004* 0.05</td>
<td>6.82 0.034* 0.01</td>
<td>0.368 0.87</td>
</tr>
<tr>
<td>OM</td>
<td>0.2827 0.59</td>
<td>0.74 0.39</td>
<td>3.27 0.008* 0.08</td>
</tr>
<tr>
<td>PA</td>
<td>1.29 0.257</td>
<td>5.69 0.18 0.03</td>
<td>2.49 0.033* 0.07</td>
</tr>
<tr>
<td>PI</td>
<td>0.84 0.36</td>
<td>1.25 0.27</td>
<td>0.38 0.86</td>
</tr>
<tr>
<td>PR</td>
<td>8.627 0.004* 0.05</td>
<td>18.17 0.000* 0.09</td>
<td>6.49 0.000* 0.15</td>
</tr>
<tr>
<td>SC</td>
<td>0.199 0.655</td>
<td>0.07 0.79</td>
<td>1.168 0.327</td>
</tr>
<tr>
<td>SH</td>
<td>0.33 0.566</td>
<td>5.85 0.02 0.031</td>
<td>5.33 .0001* 0.131</td>
</tr>
</tbody>
</table>

Comparison of community metrics within studies

The wilderness groups showed no significant influence on any of the community metrics of richness, evenness or diversity for any of the 3 studies in isolation (although there is a trend toward higher total richness in wilderness sites for the EMAP study, p = 0.056). The WC study showed no significant differences between reference designations for any of the community metrics tested (Table 9). In contrast, the EMAP and WEN studies showed no significant difference in species richness, but significant differences in species evenness and the 2 diversity measures between reference designations.

Non-metric Multidimensional Scaling (NMDS) Ordination

NMDS ordinations were performed to see if there might be compositional differences in the macroinvertebrate assemblages based on some of the physical attributes. The ordination graphs revealed some underlying patterns. The scree plot (Figure 2) shows a lessening in the slope as
Table 9. Community metrics differences (p values) between reference or wilderness designations for each study. Species richness (S), evenness (E), Simpson’s diversity index (D’) and Shannon’s diversity index (H) values are shown. Asterisks (*) denote significant differences.

<table>
<thead>
<tr>
<th></th>
<th>Reference /Not Reference</th>
<th>Wilderness/ Non-Wilderness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EMAP</td>
<td>WC</td>
</tr>
<tr>
<td>S</td>
<td>0.17</td>
<td>0.34</td>
</tr>
<tr>
<td>E</td>
<td>0.05*</td>
<td>0.61</td>
</tr>
<tr>
<td>D’</td>
<td>0.03*</td>
<td>0.30</td>
</tr>
<tr>
<td>H</td>
<td>0.03*</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Figure 2. "Scree" plot showing stress at different dimensions of all sites with raw species data. Dimensionality is increased. Where the slope becomes less steep is the "break" that signals that increasing dimensions will not decrease the stress of the ordination appreciably. This shows that 2 dimensions provided sufficiently low stress reduction to represent the data for the ordination of all sites with all species. This was also true in all other ordinations presented here.
Species abundance

The ordination of species assemblages by study (EMAP, WEN and WC) shows that the composition of macroinvertebrates differed by study and that the mean annual precipitation (SiteMean) and longitude (LON_DD) were the driving forces for the differences (Figure 3). Full species assemblages also showed separation in the ordination by wilderness designation (Figure 4), and reference designation which is further coded and shown by high and low precipitation sites (Figure 5).

Viewing the same species assemblage ordination by stream order (Figure 6) and elevation (not shown) shows some patterns of separation. Note the community composition of stream orders 5 and 6 separating out and clumping (towards lower left) defined by increasing watershed area (WSAREA). Communities from streams of orders 3 and 4 also separate out (top right direction) in the direction of increasing annual precipitation (SiteMean). In contrast, orders 1 and 2 are more spread out but appear to be clumping in two separate areas. In both cases the mean annual precipitation and the watershed area were the strongest influences on community similarity.

**Figure 3.** NMDS Ordination graph of all sites with all species, showing separation in populations between wilderness and non-wilderness sites, with mean annual precipitation and longitude as the main physical drivers.
Figure 4. NMDS Ordination graph of all sites with all species, showing separation in populations between wilderness and non-wilderness sites, with mean annual precipitation, longitude and watershed area as the main physical drivers.

Figure 5. NMDS Ordination graph of all sites with all species showing separation in populations between reference and test sites. Solid circles are reference sites with lower precipitation, shaded circles are reference with higher precipitation and open triangles are non-reference sites.
Higher taxon and functional groups

Groups made by the same sets of physical variables can also be seen clustering in the ordination graphs, showing similarity when the communities are characterized at higher taxonomic levels and functional groups. Macroinvertebrate communities identified to the order level and used to calculate richness show patterns of separation like communities identified by species-level identifications for both reference and wilderness designations (Figure 7).

Similarly, when the ordination graphs were drawn again with communities identified by functional feeding group richness, clear clustering of the communities were again seen for reference and non-reference sites with watershed area (WSAREA) the strongest physical influence (Figure 8).

Multi-Response Permutation Procedures (MRPP)

Statistical validity of the visual separation of the ordinations was tested and confirmed with MRPP on selected variables and subsets of the data. Several different subsets of the data were used to test for differences in macroinvertebrate community structure: species-level abundance,
Figure 7. Assemblages distinguished in ordination space by reference and wilderness designations with communities defined by their macroinvertebrate order richness. Solid triangles are wilderness or references, hollow circles are non-reference and non-wilderness.

presence/absence, richness of the orders EPTC&D, abundance of individuals by tolerance score, richness and abundance of first FFG designation, as well as first with secondary FFG designation, and finally, FFG (first designation only) as presence/absence. Abundance requires the number of individuals of each taxon or type, richness requires the number of different kinds of each taxon or type, and presence/absence which was coded 0 or 1. The physical variables used to constrain
Figure 8. NMDS ordination of assemblages defined by functional feeding group richness shown designated by wilderness and reference condition. Areas of non-reference and non-wilderness (hollow triangles) are shown occurring in the direction of increasing watershed area (WSAREA).
the community variables were month, year, stream order, elevation, slope, and reference and wilderness designations.

The results from MRPP for all the sites together, for all the community characterizations (higher taxonomy, FFG, etc.) showed that the assemblages were distinct from each other with detectable significant differences apparent for most physical and temporal designations tested (month, year, stream order, slope, elevation and reference and wilderness designations) including the variable that separated the collection into the 3 smaller studies (Tables 10 and 11) which may be an effect of the very large sample size (183 sites). However, there were three exceptions: 1) for EPTC&D richness, there were no differences among studies; 2) FFG1 presence/absence could not distinguish either of the reference or wilderness designations; and 3) the community characterized by order presence/absence could not distinguish between reference and non-reference sites. The FFG P/A characterization was excluded from further testing. Slope and elevation showed the highest "A" values (within-group agreement) for many of the tests. These variables might contribute the most to making a group less heterogeneous, in other words, slope and elevation may be the strongest variables used here affecting community structure.

Table 10. MRPP results for the entire dataset and macroinvertebrate community structure compared among a variety of temporal and physical stream variables. The dataset was reorganized by lower taxonomic specificity and a variety of simple community metrics. Values represent A (chance-corrected within-group agreement; effect size) and p-values (the probability that the groups differ by chance alone). Shaded cells denote non-significant results or the inability of a simpler community metric to distinguish among communities.
Table 11. MRPP results for the entire dataset and macroinvertebrate community structure compared among a variety of temporal and physical stream variables. The dataset was reorganized by lower taxonomic specificity and a variety of simple community metrics. Values represent A (chance-corrected within-group agreement; effect size) and p-values (the probability that the groups differ by chance alone). Shaded cells denote non-significant results or the inability of a simpler community metric to distinguish among communities.

<table>
<thead>
<tr>
<th>Entire Dataset</th>
<th>Order richness A value</th>
<th>p value</th>
<th>Order Abundance A value</th>
<th>p value</th>
<th>Order Presence/Absence A value</th>
<th>p value</th>
<th>Family Count A value</th>
<th>p value</th>
<th>Family Presence/Absence A value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order</td>
<td>0.034</td>
<td>0.000</td>
<td>0.027</td>
<td>0.000</td>
<td>0.05</td>
<td>0.00</td>
<td>0.024</td>
<td>0.000</td>
<td>0.037</td>
<td>0.000</td>
</tr>
<tr>
<td>Month</td>
<td>0.021</td>
<td>0.000</td>
<td>0.014</td>
<td>0.003</td>
<td>0.01</td>
<td>0.03</td>
<td>0.014</td>
<td>0.000</td>
<td>0.027</td>
<td>0.000</td>
</tr>
<tr>
<td>Year</td>
<td>0.061</td>
<td>0.000</td>
<td>0.014</td>
<td>0.006</td>
<td>0.16</td>
<td>0.00</td>
<td>0.031</td>
<td>0.000</td>
<td>0.086</td>
<td>0.000</td>
</tr>
<tr>
<td>Slope</td>
<td>0.035</td>
<td>0.000</td>
<td>0.035</td>
<td>0.000</td>
<td>0.06</td>
<td>0.00</td>
<td>0.047</td>
<td>0.000</td>
<td>0.067</td>
<td>0.000</td>
</tr>
<tr>
<td>Elevation</td>
<td>0.080</td>
<td>0.000</td>
<td>0.086</td>
<td>0.000</td>
<td>0.07</td>
<td>0.00</td>
<td>0.055</td>
<td>0.000</td>
<td>0.068</td>
<td>0.000</td>
</tr>
<tr>
<td>Ref</td>
<td>0.017</td>
<td>0.000</td>
<td>0.010</td>
<td>0.001</td>
<td>0.00</td>
<td>0.11</td>
<td>0.009</td>
<td>0.000</td>
<td>0.011</td>
<td>0.000</td>
</tr>
<tr>
<td>Wilderness</td>
<td>0.014</td>
<td>0.000</td>
<td>0.013</td>
<td>0.000</td>
<td>0.01</td>
<td>0.01</td>
<td>0.010</td>
<td>0.000</td>
<td>0.013</td>
<td>0.000</td>
</tr>
<tr>
<td>Group</td>
<td>0.049</td>
<td>0.000</td>
<td>0.006</td>
<td>0.027</td>
<td>0.14</td>
<td>0.00</td>
<td>0.026</td>
<td>0.000</td>
<td>0.085</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Looking closer with MRPP at the differences within these separate groups of sites (EMAP, WC and WEN) by community and physical variables, it turns out that two of the groups responded similarly to the composite group to many of the variables, but one appeared to have within group similarity enough that it was difficult to distinguish even reference and non-reference using the full species list which was significantly different in the other two groups (Tables 12, 13, and 14). The lower sample size, reduces the power of the analysis, so larger differences are required for detection in these smaller datasets. In other words, a larger sample sizes allows these tests to distinguish more subtle effects or differences. The first column in Tables 12, 13 and 14, show MRPP results for each separate study using the species abundance data constrained by the same physical and temporal variables as the MRPP tests with the whole dataset (Tables 10 and 11). The rest of the columns show results for the populations described by distilled community metrics constrained by these same variables and were performed to see if there was agreement with the full species abundance results. When the results by the distilled community metrics are compared with full species abundance results, many had the same or similar significant differences, (including non-significance), but there were exceptions. The community differences by reference designation were as distinguishable with most distilled community metrics except for the WEN study: the full species list could distinguish reference sites but most distilled community metrics could not.
Table 12. MRPP results for each study (EMAP, WC, WEN) separately which compare macroinvertebrate community structure among a variety of temporal and physical stream variables. The dataset was reorganized by lower taxonomic specificity and a variety of simple community metrics (either by species abundance, family and order abundance; functional feeding group richness (FFG1 is using primary designation, FFG2 is primary and secondary designations); richness of the groups EPTD&C, Family, and Order; and lastly presence/absence of Order and Family). Values represent A (chance-corrected within-group agreement; effect size) and p-values (the probability that the groups differ by chance alone). Shaded cells denote non-significant results or the inability of a simpler community metric to distinguish among communities.

<table>
<thead>
<tr>
<th>Study</th>
<th>All Species Abundance</th>
<th>FFG1 Richness</th>
<th>FFG1 Abundance</th>
<th>FFG2 Richness</th>
<th>EPTD&amp;C Richness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A value</td>
<td>p value</td>
<td>A value</td>
<td>p value</td>
<td>A value</td>
</tr>
<tr>
<td>EMAP</td>
<td>Order</td>
<td>0.017</td>
<td>0.003</td>
<td>0.041</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>Month</td>
<td>0.011</td>
<td>0.013</td>
<td>0.042</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>Year</td>
<td>0.010</td>
<td>0.006</td>
<td>0.022</td>
<td>0.066</td>
</tr>
<tr>
<td></td>
<td>Slope</td>
<td>0.044</td>
<td>0.000</td>
<td>0.016</td>
<td>0.332</td>
</tr>
<tr>
<td></td>
<td>Elevation</td>
<td>0.021</td>
<td>0.007</td>
<td>0.045</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td>Ref</td>
<td>0.000</td>
<td>0.095</td>
<td>0.016</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td>My Ref</td>
<td>0.002</td>
<td>0.161</td>
<td>0.006</td>
<td>0.228</td>
</tr>
<tr>
<td>WC</td>
<td>Order</td>
<td>0.045</td>
<td>0.000</td>
<td>0.088</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Month</td>
<td>0.015</td>
<td>0.000</td>
<td>0.011</td>
<td>0.097</td>
</tr>
<tr>
<td></td>
<td>Year</td>
<td>0.002</td>
<td>0.269</td>
<td>-0.005</td>
<td>0.651</td>
</tr>
<tr>
<td></td>
<td>Slope</td>
<td>0.063</td>
<td>0.000</td>
<td>0.109</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Elevation</td>
<td>0.083</td>
<td>0.000</td>
<td>0.118</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Ref</td>
<td>0.027</td>
<td>0.000</td>
<td>0.070</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>My Ref</td>
<td>0.021</td>
<td>0.000</td>
<td>0.031</td>
<td>0.001</td>
</tr>
<tr>
<td>WEN</td>
<td>Order</td>
<td>0.01</td>
<td>0.00</td>
<td>0.006</td>
<td>0.276</td>
</tr>
<tr>
<td></td>
<td>Month</td>
<td>0.02</td>
<td>0.00</td>
<td>0.038</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>Year</td>
<td>0.01</td>
<td>0.00</td>
<td>0.004</td>
<td>0.249</td>
</tr>
<tr>
<td></td>
<td>Slope</td>
<td>0.04</td>
<td>0.00</td>
<td>0.072</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>Elevation</td>
<td>0.03</td>
<td>0.00</td>
<td>0.050</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>Ref</td>
<td>0.01</td>
<td>0.00</td>
<td>0.008</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>My Ref</td>
<td>0.021</td>
<td>0.000</td>
<td>0.015</td>
<td>0.022</td>
</tr>
</tbody>
</table>
Table 13. MRPP results for each study (EMAP, WC, WEN) separately which compare macroinvertebrate community structure among a variety of temporal and physical stream variables. The dataset was reorganized by lower taxonomic specificity and a variety of simple community metrics. Values represent A (chance-corrected within-group agreement; effect size) and p-values (the probability that the groups differ by chance alone). Shaded cells denote non-significant results or the inability of a simpler community metric to distinguish among communities.

<table>
<thead>
<tr>
<th>Study</th>
<th>all species Abundance</th>
<th>Family Abundance</th>
<th>Family P/A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A value</td>
<td>p value</td>
<td>A value</td>
</tr>
<tr>
<td>EMAP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order</td>
<td>0.017</td>
<td>0.003</td>
<td>0.025</td>
</tr>
<tr>
<td>Month</td>
<td>0.011</td>
<td>0.013</td>
<td>0.014</td>
</tr>
<tr>
<td>Year</td>
<td>0.010</td>
<td>0.006</td>
<td>0.013</td>
</tr>
<tr>
<td>Slope</td>
<td>0.044</td>
<td>0.000</td>
<td>0.063</td>
</tr>
<tr>
<td>Elevation</td>
<td>0.021</td>
<td>0.007</td>
<td>0.033</td>
</tr>
<tr>
<td>Ref</td>
<td>0.003</td>
<td>0.095</td>
<td>0.004</td>
</tr>
<tr>
<td>My Ref</td>
<td>0.002</td>
<td>0.161</td>
<td>0.001</td>
</tr>
<tr>
<td>WC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order</td>
<td>0.045</td>
<td>0.000</td>
<td>0.052</td>
</tr>
<tr>
<td>Month</td>
<td>0.015</td>
<td>0.000</td>
<td>0.017</td>
</tr>
<tr>
<td>Year</td>
<td>0.002</td>
<td>0.269</td>
<td>-0.006</td>
</tr>
<tr>
<td>Slope</td>
<td>0.063</td>
<td>0.000</td>
<td>0.081</td>
</tr>
<tr>
<td>Elevation</td>
<td>0.083</td>
<td>0.000</td>
<td>0.110</td>
</tr>
<tr>
<td>Ref</td>
<td>0.027</td>
<td>0.000</td>
<td>0.032</td>
</tr>
<tr>
<td>My Ref</td>
<td>0.021</td>
<td>0.000</td>
<td>0.021</td>
</tr>
<tr>
<td>WEN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order</td>
<td>0.045</td>
<td>0.000</td>
<td>0.021</td>
</tr>
<tr>
<td>Month</td>
<td>0.021</td>
<td>0.001</td>
<td>0.017</td>
</tr>
<tr>
<td>Year</td>
<td>0.034</td>
<td>0.000</td>
<td>0.012</td>
</tr>
<tr>
<td>Slope</td>
<td>0.009</td>
<td>0.016</td>
<td>0.058</td>
</tr>
<tr>
<td>Elevation</td>
<td>0.061</td>
<td>0.000</td>
<td>0.018</td>
</tr>
<tr>
<td>Ref</td>
<td>0.028</td>
<td>0.000</td>
<td>0.010</td>
</tr>
<tr>
<td>My Ref</td>
<td>0.010</td>
<td>0.003</td>
<td>0.008</td>
</tr>
</tbody>
</table>
Table 14. MRPP results for each study (EMAP, WC, WEN) separately which compare macroinvertebrate community structure among a variety of temporal and physical stream variables. The dataset was reorganized by lower taxonomic specificity and a variety of simple community metrics. Values represent A (chance-corrected within-group agreement; effect size) and p-values (the probability that the groups differ by chance alone). Shaded cells denote non-significant results or the inability of a simpler community metric to distinguish among communities.

<table>
<thead>
<tr>
<th>Study</th>
<th>all species abundance</th>
<th>Order Richness</th>
<th>Order Abundance</th>
<th>Order P/A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A value</td>
<td>p value</td>
<td>A value</td>
<td>p value</td>
</tr>
<tr>
<td>EMAP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order</td>
<td>0.017</td>
<td>0.003</td>
<td>0.053</td>
<td>0.011</td>
</tr>
<tr>
<td>Month</td>
<td>0.011</td>
<td>0.013</td>
<td>0.059</td>
<td>0.001</td>
</tr>
<tr>
<td>Year</td>
<td>0.010</td>
<td>0.006</td>
<td>0.032</td>
<td>0.016</td>
</tr>
<tr>
<td>Slope</td>
<td>0.044</td>
<td>0.000</td>
<td>0.055</td>
<td>0.077</td>
</tr>
<tr>
<td>Elevation</td>
<td>0.021</td>
<td>0.007</td>
<td>0.062</td>
<td>0.025</td>
</tr>
<tr>
<td>Ref</td>
<td>0.003</td>
<td>0.095</td>
<td>0.010</td>
<td>0.119</td>
</tr>
<tr>
<td>My Ref</td>
<td>0.002</td>
<td>0.161</td>
<td>0.010</td>
<td>0.131</td>
</tr>
<tr>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order</td>
<td>0.045</td>
<td>0.000</td>
<td>0.077</td>
<td>0.000</td>
</tr>
<tr>
<td>Month</td>
<td>0.015</td>
<td>0.000</td>
<td>0.007</td>
<td>0.163</td>
</tr>
<tr>
<td>Year</td>
<td>0.002</td>
<td>0.269</td>
<td>0.000</td>
<td>0.445</td>
</tr>
<tr>
<td>Slope</td>
<td>0.063</td>
<td>0.000</td>
<td>0.103</td>
<td>0.000</td>
</tr>
<tr>
<td>Elevation</td>
<td>0.083</td>
<td>0.000</td>
<td>0.151</td>
<td>0.000</td>
</tr>
<tr>
<td>Ref</td>
<td>0.027</td>
<td>0.000</td>
<td>0.041</td>
<td>0.000</td>
</tr>
<tr>
<td>My Ref</td>
<td>0.021</td>
<td>0.000</td>
<td>0.017</td>
<td>0.009</td>
</tr>
<tr>
<td>WEN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order</td>
<td>0.021</td>
<td>0.001</td>
<td>0.011</td>
<td>0.148</td>
</tr>
<tr>
<td>Month</td>
<td>0.034</td>
<td>0.000</td>
<td>0.005</td>
<td>0.211</td>
</tr>
<tr>
<td>Year</td>
<td>0.009</td>
<td>0.016</td>
<td>0.007</td>
<td>0.165</td>
</tr>
<tr>
<td>Slope</td>
<td>0.061</td>
<td>0.000</td>
<td>0.030</td>
<td>0.044</td>
</tr>
<tr>
<td>Elevation</td>
<td>0.028</td>
<td>0.000</td>
<td>0.062</td>
<td>0.000</td>
</tr>
<tr>
<td>Ref</td>
<td>0.010</td>
<td>0.003</td>
<td>0.007</td>
<td>0.088</td>
</tr>
<tr>
<td>My Ref</td>
<td>0.010</td>
<td>0.002</td>
<td>0.014</td>
<td>0.014</td>
</tr>
</tbody>
</table>
Characterizations in the WEN study that did not distinguish reference sites were EPTC&D richness, FFG1 richness and order richness, abundance and presence/absence. The EMAP study could not distinguish reference sites by the full species abundance nor any community designation, equally. The WC study could distinguish in all cases. Interestingly, the WC group did show a significant ability to distinguish reference and non-reference sites by "order P/A" when the full species abundance list with all 3 sites together could not. Other reorganizations of the sample assemblages had varying agreement with the full species abundance results for each of the other physical and temporal variables. But many of these were just as distinguishable with the distilled metrics as they were with complete species abundance (Tables 12, 13, and 14).

Additionally, the data of the composite group were divided into two smaller groups (by higher or lower elevation) to test the strength of the previous results within these two groups. This narrows the scope of one of the variables, elevation, which creates smaller and possibly more homogeneous community groups, so differences in communities will need to be stronger than with the entire group to show a significant difference. This division did not erase the difference detected for total richness between both reference and wilderness reference designations (Table 15).

Table 15. MRPP results for difference in total richness between reference designations separated by high elevation (> 900 m) and low elevation (< 900 m) sites. Asterisks (*) denote significant differences between means.

<table>
<thead>
<tr>
<th></th>
<th>Total richness elevation &gt; 900 m</th>
<th>Total richness Elevation &gt; 900 m</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A value  p value</td>
<td>A value  p value</td>
</tr>
<tr>
<td>Reference/non-reference</td>
<td>0.005  0.027*</td>
<td>0.015  0.000*</td>
</tr>
<tr>
<td>Wilderness /non-wilderness</td>
<td>0.009  0.002*</td>
<td>0.005  0.006*</td>
</tr>
</tbody>
</table>

Summary
Data characterizations in this thesis differed in ability to describe some of the environmental, temporal and qualitative groups tested. The most effective characterizations were deemed ones that agreed with results obtained using the community described by full species abundance, and by ranking the variables they agreed with as reference, elevation and stream order being more important than month or year or wilderness. The temporal designations were less useful because the month was a gross estimate of the time (day of the season were not accounted for,
thus the difference between two adjacent months could be 1 day up to 31 days) and the years had an uneven distribution among the groups. The wilderness designation was not found to be useful because it did not show much distinction in community structure using the statistical tests performed.

The best characterization, according to the standard, was found to be family abundance. After that, the one that had most agreement was FFG2 richness. The FFG1 abundance did almost as well and in some cases better than FFG1 richness. Next for quality in agreement was order abundance which did better than order richness except in the EMAP group, which was the group that had the least agreement with the species abundance for many characterizations. The reason this group differed is unclear but may be because of the small sample size or because the sites separated by Reference designation were similar enough that they were indistinguishable in most cases. Smaller samples like this may require a larger effect difference to show a significant p value. (McCune and Grace 2002).

Sub-study Analysis
The differences seen for how the communities responded to different variables in each sub-study may be understood by some of the physical and sampling attributes.

EMAP
The sites in the EMAP study (WAP) differed by having a much smaller average watershed area and higher elevation than the others. The MRPP were not able to use many of the assemblage characterizations to tell these sites apart by the physical variables. In fact, the EMAP group did not have community structure different enough (by the whole species list) to detect any differences between the reference or non-reference and the wilderness or non-wilderness sites. For this group, all new ways of expressing the community caused other variables to drop from significance except family count. Order richness only lost ability to detect the difference by slope.

This lack of distinction by MRPP may be because the communities were very similar, not exhibiting much real difference between reference and non-reference, or it might be because the sample size was too small and perhaps contained many unique or rare species. But the EMAP site communities were distinguishable when viewed by other physical and temporal variables with the full species list and by some of the re-named species sets (i.e. order and EPTDC Richness, (Table 12, 13, 14), which suggests that there just might not be enough truly
degraded sites to become "non-reference".

**WC**

The sites in this study were almost all sampled in multiple years, usually about 3 years each and the number of samples were evenly divided by the years. Therefore, this group had a lot of duplicate sites (although from different years), so the lowest effective number of different sites. This might be a cause for why this group, even using the full abundance list with MRPP, could not distinguish sites by the temporal category of year and in some cases, by month (for order abundance and P/A, family abundance and P/A and FFG2 richness; Tables 12, 13, and 14). Redundant information may not help distinguish groups, and more diverse samples might be important. All other physical variables were able to be separated by all the community surrogate metrics. Yet, this group had insignificant differences between the reference and non-reference site community metrics (Table 9.). This group had the lowest average elevation, largest watershed area and highest percentage of 5th and 6th order streams (over half).

**WEN**

In addition to the averages, the ranges of precipitation, elevation and watershed area also differed for WEN sites from the other two groups. The average stream order is the same (Strahler order 4) for all of the reference and non-references sites. WEN is the least physically diverse group of sites. This may account for the generally lower ability for the MRPP to distinguish community variable groups apart by the physical variables. WEN did not do well with order abundance, richness and P/A nor EPTCD richness or FFG1 richness for telling physical variables and reference sites apart. Since the physical attributes of this group are more similar, the assemblages might be less diverse. Also, some of the physical variables may have a strong influence on community structure and blur any differences that might exist over a more diverse landscape. A large sample may be needed in cases like this. Interestingly, WEN had the highest number of distinct sites of the 3 groups (58) yet in general, WEN had the most trouble replicating the results of the raw species abundance list with the higher taxa and alternate community metrics.

The response of these different groups suggests that in order to get a useful resolution of community differences between variables like reference designations, the sites chosen will have to be balanced by a combination of factors like site diversity and number of sites.
Particularly, sample sites that are mostly similar in some way, or sampled multiple times will need a large number of samples. A study with diverse landscape or with strikingly degraded and pristine reference sites may allow for many less sites. If number of samples are a limitation in a study, sites must be chosen to promote physical differences.

Discussion
What are the considerations needed to created a workable multimetric model? Finding and using the correct variables to characterize stream sample sites that help partition them by their macroinvertebrate communities is important. The ways in which the communities are described and characterized must also be carefully chosen. Choosing a multivariate statistical model requires additional consideration. In the end, there are a few things this thesis revealed that could be further explored or rectified in future work of this kind.

Habitat and Environmental Variables
The question of how much and which biotic or habitat and environmental data is useful in an analysis seems to have varying answers. In some cases relatively few variables may be sufficient but often more variety in variables (including specific habitat descriptions) improves multivariate analyses. Because of the way macroinvertebrate assemblages in this thesis responded differently to the same physical variables when using subsets of samples, data that included more descriptive physical variables and possibly habitat would likely improve the ability to categorize sites and assess disturbance. Yet the limited environmental and abiotic data used in this thesis seemed sufficient to find differences in site categories when used with the larger dataset and with many of the community variables for the subset studies. Past research corroborates these findings. Cereghino et al. (2003) successfully used only four physical variables in their study using a neural network analysis to predict EPTC richness. Additionally, Poff (1997) argues that abiotic landscape variables should be the framework for these kinds of studies because they are disconnected from the evolutionary relationships in assemblages, and the taxa present should be viewed by their adaptations to these more general variables. Hargett, et al. (2007) studied multivariate RIVPACS O/E models used in streams in Wyoming and found that the strongest predictive variables, of the very many tested, for macroinvertebrate communities were the log of the watershed area, the log of coarse substrate and the ecoregion. In addition, latitude, longitude, elevation and some geologic variables were among the best and 12 of these top 14
variables could be obtained using a GIS map (Hargett et al. 2007). The results in this thesis showed that slope, elevation and stream order had the strongest influences on macroinvertebrate assemblages (based on highest "A" values -within group agreement) shown through MRPP tests). Nevertheless, there certainly might be other physical variables which may perform better than those examined here, including other biotic or chemical descriptions.

Many other studies have shown significant prediction of macroinvertebrate assemblages using habitat data, and some show that compared to abiotic environmental variables, biotic variables can also be strong. In Portugal, Aguiar et al. (2002) found (using Canonical Correspondence Analysis) that riparian variables were much more important than other abiotic environmental characteristics (possibly because food types are related to riparian features). Using multivariate techniques, Haidekker and Hering (2008) found that the assemblages of the macroinvertebrate orders EPT&C could be predicted best with the physical variable stream temperature, which they relate to floodplain and land use, but that conductivity, substratum type and riparian cover percentage were also important. Another study (Chessman and Royal, 2004) used this idea in reverse, estimating the species assemblage that might be present under natural conditions using environmental filters alone, in lieu of the details at established reference sites. Their study was one that used a taxon pool with known tolerances and preferences to estimate the expected richness of macroinvertebrate families using a variety of environmental data including habitat descriptions (Chessman and Royal 2004). Halwas et al. (2005) used channel units successfully in their study as a surrogate measure of stream habitat. Overall, biotic or abiotic variables that echo habitat conditions are used in many, if not most studies and appear to be useful for macroinvertebrate assemblage characterizations.

Whether to use biotic or abiotic variables might not be an either/or choice. Instead, diverse variables of both types, or at least several, may better distinguish sites along a gradient. Hargett et al. (2007) suggest that many variables should be used together in order to improve how a model covers many diverse sites. The study by Lamouroux et al. (2004) found that the variability of the community, specifically the functional habits (including size, form, attachment, feeding, etc.) depended on filters at both large and small environmental scales; regional to microhabitat scale. There may be no universal set of variables needed to predict community composition, and which variables work for a particular place or research project might be unique.
Characterizing the assemblages

How to characterize a community, by taxon or functional description, and by abundance or richness are also questions with many answers. The varying responses of communities to the same variable are apparent in this thesis and other studies as well. Taxonomic and functional diversity are not equivalent (Poff et al. 2006). Similar functional diversity values can be formed using very different species, and taxon richness can also be expressed with different species, making these variables less sensitive to community (by species) differences. Cereghino et al. (2003) found assemblage composition and species richness made different classifications with the same data. Macroinvertebrate habit may be a better way to describe an assemblage than functional or taxonomic group. Habit categories (which were not used in this thesis) include those for positioning and movement, with designations like clingings, burrowers, skaters, etc. (Merritt et al. 1996). Fore et al. (1996) states directly that habit measures have been found to be more robust than functional feeding groups in some instances. The evidence for habit being a strong predictor in stream studies might make that characterization especially useful.

Abundance or Richness

Richness has been related to habitat diversity and stability, and may intuitively seem like a better metric to use than abundance. Cereghino et al. (2003) used richness in their study because they thought richness was predictable with the variables they used and that richness was a good indicator for disturbance. But as Jonsson and Malmqvist (2005) found, both the species populating a FFG as well as the richness of a FFG may have ramifications for the community structure and proportions of the different FFGs. In their experiment, different species composition and richness of shredders affected the particle quantity and quality (size) and, in turn significantly affected the growth of black flies. Perhaps this web of connections based on taxa presence is a stronger influence on the character of the assemblage than just the diversity of the separate groups. In this thesis, FFG1 and order abundance performed the same or better than FFG1 and order richness as a way to organize the groups by physical variables (except for the EMAP group). Designating the assemblage by a FFG with a secondary group (FFG2) made the analysis stronger also, probably because it expanded the number of groups. Richness may seem more important, but abundance appears to be a stronger categorization in some cases, including this thesis.

While a presence/absence metric may greatly simplify field collection, it may be of low
value in studies. It had the lowest ability of any of the metrics used in this thesis to group data, and Bowman and Bailey (1997) also found it provided mixed and inferior results compared to quantitative metrics for the ability to describe community structure with increasing taxonomic levels.

**Higher taxonomy**

Feminella (2000) found that higher taxonomic resolutions could provide acceptable distinction between sites among catchments. Bowman and Bailey (1997) compared the distance matrices calculated at different taxonomic resolutions (using data from 10 published studies) to see how much information was lost. They found that higher taxonomic levels maintained the distances and were not much different for describing the communities than using lower taxonomic labels. Metrics commonly used in a multimetric Index of Biological Integrity (IBI) are also derived with coarser taxonomic data. For instance, a collection where each subject is classified by order could provide many of the common metrics used (i.e. E, P and T ratios) and adding functional group or habit would make it possible to derive many more commonly used metrics (like predator ratios).

**Functional Groups**

Characterizing a macroinvertebrate population by functional groups including feeding and habit, and by tolerance scores may offer very different answers about group similarities. Poff et al. (2006) stresses that when using functional traits, it is important to find the correlations among traits when using many variables as multivariate methods become more common. They go on to explain how traits are linked as "trait states" or "syndromes" that can be used to describe members of macroinvertebrate communities. This approach holds much potential for predicting changes in both species and species assemblages along environmental gradients in terms of traits that are sensitive to local environmental conditions. While some traits are uncorrelated to phylogenetic relationships, others are, and using the more statistically uncorrelated designations makes for more robust multimetric analysis (Poff 2006). Poff (1997) describes using abiotic and environmental factors to explore macroinvertebrate functional groups or functional relationships, including tolerance. Relating these physical and habitat conditions to functional aspects of macroinvertebrates, rather than taxonomic, will be more general and not influenced by co-evolution of any of the species. Since they are independent of any evolutionary linkages between taxa, sites can be compared using the resulting assemblages derived by adaptation or
attributes of the species present which should improve comparisons for sites across larger geographic scales (Poff 1997).

**Multivariate considerations**
Multivariate modeling can improve studies of biotic communities with environmental variables for several reasons. First, because they allow the use of many types of variables at the same time, more information can be explored and interactions accounted for. Second, because multivariate models use permutative statistics, they are not required to meet strict assumptions. In addition, some types of multivariate models might actually favor the use of alternative classification or simplified, distilled metrics. One inherent problem of multivariate O/E models is with detecting taxa changes based on reference conditions without allowing the replacement of new taxa in the assemblage which would belie the reference status of an equally rich but different assemblage (Van Sickle 2008). But the quality and meaning of the results of multivariate models depend on many factors, so must be used with careful study beforehand and caution with interpretation.

**MRPP vs. univariate statistics**
It is interesting that the univariate tests did not show FFG1 richness as distinct between reference and non-reference when it was a strong effect on overall community structure when the assemblage of FFGs was tested. This may attest to the power of MRPP and multivariate methods in general. For FFG richness, MRPP uses the richness of each FFG designation together as a group, (the matrix has each column with a different FFG and each cell shows the abundance), while the univariate test only uses one number for total "FFG richness" to test the mean and variation between sample sites groups. Similarly, there was no difference in average tolerance between site groupings by ANOVA but there was using the assemblage of tolerance values. Again, here MRPP uses the number of tolerant individuals in each tolerance class instead of one number which is calculated as a mean. Since the abundance uses the number of individuals in each class (like FFG or tolerance) all at once, richness (of the whole class) is taken into account with the abundance, adding more information to the comparison.
Points to explore further

This thesis revealed a few considerations for the design of bioassessments that can use simpler or coarser community data with multivariate statistics. These include the variables measured and used to describe the sample sites, the number and qualities of the sites chosen, the kinds of characterizations to be used for the assemblages, and general study design and collection methods.

The importance and quality of many physical variables that can be used to group communities using species abundance or richness characterization is not explored in this thesis, but may have influenced the results. Multivariate models are powerful tools, and when combined with appropriate variables, can organize communities in many different categories for comparison and to test inclusion. If a model includes different, better or more physical data than used in this thesis, or environmental data like habitat characterizations which were not included in this analysis, it is likely that it may perform much better for assigning inclusion into groups like "reference" sites. Feminella (2000) found using PCA that variables of chemical description in streams were much more important than physical variables, and that variables related to geographic position, like elevation, were the least useful. If these variables of low potential use were some of the strongest in my analysis, there may be reason to believe that environmental data would greatly improve the analysis and results.

The quality and quantity of sampling sites appear important also. This thesis showed that perhaps the collection of sites needs to have a certain physical diversity for use in a multivariate model and this may be especially the case when the number of sites are low. The WEN study which had the highest number of different sites but contained the smallest range of environmental variables was inferior to the WC study (with fewer distinct sites but covering a larger environmental range) for its ability to use distilled metrics and reproduce the significant differences found when using full species abundance in MRPP tests. Although, it is not certain that site diversity was the reason this study was less differentiated with the variables used, and a closer exploration may prove useful. The effect of multiple samples from the same test site at different times is another issue to explore.

Whether this type of simplified collection method and analysis can be used broadly, in different locations is another question. Although the evidence points to this being true, it may be that a unique set of usable metrics from the data may need to be derived for each place, as it was shown in this thesis that the same metrics did not perform equally among the three sub-
studies. Or, there might be cause to search for some metrics that can be applied uniformly for use with data over a larger and more inclusive geographic area. Finding metrics that work best for each system should be part of a study design.

If field protocols for sampling and identification are to be used by regulatory agencies studying streams, methods that will produce comparable results to lab sampling can be explored, and then described and standardized for future collecting and field analysis. In Australia, there is already a "live-sort" collection protocol in use and being refined specifically for use with multivariate models (Schiller 2003). Haurer and Resh (2006) also describe field collection and identification methods for use in North American stream.

**Conservation and Management Implications**

There may be multiple benefits, including higher efficiency, lower costs and environmental impacts of bioassessments, if stream macroinvertebrate data can be collected using higher taxonomic or traits-based identity, but still provide a useful level of discrimination for multivariate models. Determining functional feeding group based on mouthpart morphology or feeding behavior or the identity of specimens by family or order levels are simple characterizations that may be performed in the field using published keys, perhaps in conjunction with established rapid biological assessment protocols. In this way, monitoring results might be obtained more quickly and at a reduced cost and time investment. Extra time may be spent in the field characterizing the sample, but lab sample identification is an even longer and expensive process. This efficiency may result in more sites being sampled, or sampled more often which may improve the quality of the results of bioassessment studies.

Results from the Wenatchee basin in this thesis show that data of higher taxonomic level or functional feeding group designation can be used with multivariate techniques to make distinctions in the data that do no worse than analysis based on raw total species abundance data. The characterizations that worked in this thesis involved the richness of functional feeding groups as well as richness and abundance at higher taxonomic levels (family and order). The more detailed FFG with secondary designations that did so well in this thesis investigation and tolerance values would require fully identified samples that would not be practical with a streamlined design and collection method. Investigations that needed diversity indexes or other kinds of studies involving species richness and taxa changes over time would also need full species identification. In a community characterization, it might be more ideal to have biomass
than abundance which may also be difficult in the field. But the need to classify a site for functioning as well as a similarity to a reference site may only require the kind of simplified community metrics described here.

Another benefit with using data that can be measured in the field is that insect samples can be returned to their habitat and not destroyed. Some of the taxa studied are quite rare as they show up once or twice in all the samples, and overall hundreds of thousands of insects were collected, so this altered procedure could be ecologically desirable. Therefore, this type of procedure might further the goals of the Environmental Protection Agency's “Rapid Bioassessment Protocols,” namely increasing cost effectiveness, allowing for many site investigations, accelerating data acquisition, and improving environmental effects of sampling (Barbour et al. 1999). Another advantage of using coarser community data is that it will reduce the "noise" of species diversity that can be found over larger geographic areas and therefore be able to link studies over a wider area.

**MES statement**

An MES thesis is designed to use interdisciplinary philosophy and lead us to incorporate a broader social framework to environmental issues. This is a valuable concept because viewing science through the "big picture" can open thinking into creative directions that can solve or prevent problems. In this case, I decided to use scientific exploration to hopefully aid a practical issue while incorporating a personal desire to learn to work with a large dataset and to be able to learn more about how to apply statistics to a real world project. I was generously offered the use of the dataset used in this thesis by the Washington State Department of Ecology. At that time, bioassessment was beginning to be more widely used and the WDOE was attempting a RIVPAC multivariate model with these data. After becoming familiar with the data given, I was struck by the huge undertaking of this collection and some implications. It seemed wasteful; many of the samples were not ever used (in any project) and samples that were used were often subsampled, leaving many invertebrates unidentified and unaccounted for. I did not like that so much life was mined out of stream reaches in the name of conservation. I wondered if this procedure was really necessary and if it were possible to change the way these data are collected. Realizing that patterns in the data might reveal something that could help with the analysis, my next step was to "play" with the data.

My thesis advisor encouraged me to use ordination and multivariate tests which she had
taught us in an MES elective class. I organized, sorted and re-categorized the data and tried different analyses, until I found something promising. It seemed that higher taxonomy and functional descriptions could be substituted for the fully identified samples. This was exciting because if it were possible, it could not only make bioassessment less expensive without cutting quality, which is important when budgets of environmental agencies are being cut nationally, but it would alleviate some of the disruption to the stream communities that collection caused.

There were many other studies that explored aspects of this idea, which confirmed the potential usefulness of using a simplified collection protocol.

The problem I encountered with this thesis was when to stop. Everything learned led to the desire to explore and refine further. It became obvious how wide and deep each exploration could go, but limits were important. In order to write a piece of the puzzle, a project needs to be focused. Also, many of the directions that could be pursued with these dataset would not necessarily fulfill the MES mission, spanning social or political context. Categorizing the patterns in stream communities may be interesting and important, but finding the connection to society changes the framework. I hope that bridging some of the theory of using coarser descriptions for characterizing stream communities to the methods and practices of regulating agencies research studies will improve conservation policies and effectiveness.

**Conclusion**

It may be useful for future and continuing studies in the Wenatchee Basin, and elsewhere, to consider using a simplified live, identify and release collection method as an addition to normally collected and preserved and lab identified samples, perhaps eventually reducing the number of the latter. Identifying macroinvertebrate samples as high as the order level of taxonomy, or by functional feeding group or habit may be desirable because there are just a few easily distinguished categories and yet they provide strong evidence for usefulness in community separation with few physical variables using multivariate methods.
References


http://water.epa.gov/scitech/monitoring/rsl/bioassessment/index.cfm


http://www2.state.id.us/deq.


Hubler, S. 2005. Predator: Development and use of RIVPACS-type macroinvertebrate models to assess the biotic condition of wadeable Oregon streams. Oregon Department of Environmental Quality Laboratory Division Watershed Assessment Section publication DEQ08-LAB-0048-TR.


http://www.epa.gov/emap/west/html/docs/wstream.html


