1-1-2006

Determining Biogeochemical Assemblages on the Stony River, Grant County, WV, using Fuzzy C-Means and k-Nearest Neighbors Clustering

M. Joseph Hughes

Determining Biogeochemical Assemblages on the Stony River, Grant County, WV, using Fuzzy C-Means and k-Nearest Neighbors Clustering.

Thesis submitted to the Graduate College of Marshall University.
in partial fulfillment of the requirements for the degree of
Master of Science in Physical Science

by
M. Joseph Hughes

Dr. Mike Little, Committee Chair
Dr. Dan K Evans
Dr. Scott Sarra

Marshall University

November 16, 2006

Abstract

Determining Biogeochemical Assemblages on the Stony River, Grant County, WV, using Fuzzy C-Means and k-Nearest Neighbors Clustering.

Periphyton assemblages were assessed on the Stony River, a high-gradient stream in the Potomac drainage of Grant County, WV. Periphyton samples were collected from nine sites along the mainstem and in two tributaries. Chlorophyll-a, dry weight, taxonomic identifications, and bioaccumulated metals concentrations data were compiled. These data were related to water quality parameters measured at each site during the study. Fuzzy C-means and k-nearest neighbor clustering on the combined, normalized dataset produced similar results. Clustering separated species occurring in each tributary from each other and those dominating the mainstem. Nearly every bioaccumulated metal was associated with one of these tributary clusters; phosphorus and silicon were exceptions with silicon being associated with diatoms. The remaining clusters formed a continuum of community composition along the mainstem different from the spatial arrangement of sites. Additionally, taxa occurring in small quantity force clusters to form near the center of the data-space, confounding results.

Acknowledgements

Special thanks go to Doug Kaylor for his support at every step; to my advisor and mentor, Dr. Mike Little; to Dr. Peter Villa for his help in defining the periphyton study; and to Dominion for access to data and supportive funding.

Table of Contents

1 Introduction

1.1 Periphyton

This paper describes an assessment of periphyton assemblages on the Stony River, a high-gradient stream in the Potomac drainage of Grant County, WV. This assessment includes an examination of periphyton assemblage composition and the effects that the bioaccumulation of metals and abiotic parameters have on assemblage composition.

Abiotic factors that affect periphyton populations can include light, temperature, water chemistry, nutrient availability, scouring from high currents, and grazing by macroinvertebrates. When periphyton communities established at a certain site are transferred to a different site they change rapidly, both in species composition and abundance, indicating that the absence of certain species is due not to a lack of introduction but to preferences for certain factors not present at the new site (Keithan and Barnese 1989).

Light is commonly a restrictive factor in small streams where light availability is limited in the summer when the forest canopy closes. Studies comparing algal growth in forested streams to algal growth in streams flowing through clear cut areas show that overall periphyton growth is positively correlated with light availability (Keithan and Lowe 1985; Lowe, Golladay and Webster 1986). Light levels can also influence periphyton community composition as certain species are adapted to either high or low light conditions. McIntire (1968, 1973) studied the effects of light on algal species in Oregon and found that diatoms were abundant at low light levels and at higher light levels filamentous green algae, yellow brown algae and cyanobacteria were more common.

Most studies on the effect of temperature on periphyton communities examine the composition of communities through seasonal variation in temperature. Seasonal changes in taxonomic composition are seen due to the preference of certain taxa for certain temperature ranges. For instance, warmer temperatures are partially responsible for the increase in green algae and cyanobacteria during the summer months (Whitton 1975). Studies measuring changes in chlorophyll concentrations indicate that total biomass does not vary greatly over the year. This condition is likely due to the natural progression of dominant algal taxa in periphyton communities as relates to the change in temperature exhibited during seasonal shifts (Marker 1976). However, larger fluctuations in temperature naturally exhibit larger effects on periphyton communities.

Water chemistry, more specifically pH, nutrient availability and dissolved oxygen levels, also play a role in the establishment and success of periphyton communities. In a three year study in the Adirondacks, Passy (2006) found changes in diatom community composition accompany fluctuations in pH in both chronically acidified streams as well as episodically acidified ones. Nutrients like phosphorus, nitrogen, and silica can be limiting factors as to the types and abundance of periphyton as well as to the primary productivity of the algal community; however this effect is mitigated in lotic systems because of nutrient transfer via current. Current also limits thermal stratification and ensures mixing of the water column and of the nutrient load in the water. Increased current velocity also increases the dissolved oxygen level in the water, increasing the oxygen available for cellular respiration.

In addition, high current velocity resulting from increased discharge results in scouring events which greatly affect periphyton densities and species composition. Periphyton taxa very widely in their methods of attaching to substrate and therefore differ in their ability to remain attached at increased current velocity. Scouring events also increase abrasion from tumbling substrate which removes periphyton from surfaces (Power and Stewart, 1987). At higher current velocities associated with flooding events, substrate may also invert, burying the algae in sediment thus limiting light exposure and nutrient availability (Robinson and Rushforth 1987).

Periphyton communities also change in accordance with metal concentration. Metals reported to affect benthic lotic ecosystems are iron, cadmium, lead, zinc, copper and aluminum. Both primary productivity and community respiration rates decrease with increasing metal concentrations, indicating that both heterotrophic and autotrophic components of periphyton communities are impaired (Hill et al. 1997). Abundance and morphological changes in various diatoms have been shown to be indicative of metal contamination and appear to be strongly correlated with different metals (Cattaneo et al. 2004, Macfarland et al. 1997). For instance, Vuori (1995) found that increased iron concentrations decrease cellular metabolism and osmoregulation, leading to a decrease in species diversity and abundance of periphyton, macroinvertebrates, and fish. This affects the quality of benthic habitats and the structure of benthic food chains. In addition to causing changes in morphology and abundance of periphyton, metals also bioaccumulate in periphyton species and then concentrate up the food chain through benthic macroinvertebrates ultimately into game fish and birds of prey (Besser et al. 2001, Vymazal 1984).

Natural communities of periphyton, such as occur at a given site, are typically composed of many species. Because sites are all connected, many species that occur at one site most probably occur at others; other species are unique to a specific site sampled, often because of abiotic factors at that site (Keithan and Barnese 1989). Species that occur at multiple sites might consistently occur with each other as an assemblage, allowing each site to be described as an aggregate of these assemblages.

Verb and Vis (2005) conducted an extensive study drawing in periphyton, metals, and water chemistry data from 56 stream sites. Their purpose was to associate periphyton species with metals and water chemistry parameters in order to use periphyton as a water quality predictor – a purpose very similar to this study. The large number of sites were ordinated on 16 axes representing periphyton indices, metals concentrations, nutrients concentrations, habitat indices, and other abiotic parameters. Similar sites were then found via various multivariate analysis techniques, including clustering. Periphyton taxa were indirectly associated with these parameters by way of the sites in which they are dominant. This study inverts that approach. Here periphyton species, metals, and abiotic factors make up our data points and are ordinated on axes representing a few sites.

The main goals of this investigation are 1) to describe the periphyton community and water chemistry on the Stony river, 2) determine if an ‘inverted’ clustering technique will yield results similar to the techniques employed by Verb and Vis (2004), and 3) to determine if periphyton, metals, and water chemistry assemblages exist on the Stony River.

1.2 Site Description

The Stony River is a high gradient stream in the Potomac drainage with a number of biochemical gradients primarily caused by acid mine drainage. Six sites were chosen to assess possible impacts from two tributaries; 4MR in Four Mile Run, 4M1 above the tributary, and 4M2 below. LRR, LR1, and LR2 have the same configuration around Laurel Run. Two additional sites were chosen, 0A near the Mount Storm Lake dam outfall, and SR4 under the Route 50 bridge, to assess non-point source impacts over a six mile section of the river. The final site, 0B, was chosen because of the different flow regime it experienced as a side channel. Map 1 shows the site groups.

1.2.1 Dam Outfall

Sites 0A and 0B are located near the outfall of the Mount Storm Lake dam (Map 2). 0A is located in the mainstem of the Stony River and receives water directly from the pool below the Mount Storm Lake outfall. As such, flow at 0A is sensitive to dam release. The bed of 0A is dominated by cobble and boulder, matted with diatoms. Unlike every other site downstream, adult fishes are occasionally seen in 0A, but no stable fish population exists on the Stony River and these fishes are simply washouts from the lake above.

Site 0B is a shallow side channel that typically remains connected to the mainstem of the Stony River only through ground water. The site receives flow via an overflow channel in the outfall pool and from a tributary consisting of seepage through the dam mixed with an intermittent stream. After heavy rains, 0B connects to the river through a marsh-like area. During large dam releases, flow at 0B is dominated by water directly from the lake, and may directly connect with the main stem during very large releases. Unlike the mainstem of the river, the bed of 0B is silty-sandy with boulder outcrops; obvious green algae cover the bed. Juvenile fishes are often seen in 0B prompting its inclusion to assess the site's suitability as refugia for larval fish.

1.2.2 Four-Mile Run

Four-Mile Run is a first order stream that joins the left side of the Stony River approximately one and a half miles downstream from the dam outfall (Map 3). Site 4MR is located in Four Mile Run approximately 15 m upstream from the mouth. The tributary runs through both active and abandoned mine lands and is heavily affected by acid mine drainage. Throughout the study, the tributary was treated with alkali. The pond seen on the map contains a lye solution to treat acidic run off; several of these ponds line the mining road on the north side of the river. It is assumed that a similar treatment is used for the tributary itself. On one occasion, during the summer of 2004, the tributary smelled strongly of ammonia. Because the tributary is being treated, which reduces the solubility of metals, the stream bed is covered in yellow-boy, $Fe(OH)_3$, an iron-based particulate.

Sites 4M1 and 4M2 lie above and below the tributary, respectively. Both sites are similar in makeup, though 4M2 is streaked with yellow-boy whereas 4M1 is not. Like 0A, flow is mostly dependant on dam release, with very low flow when the dam is closed. For the greater portion of the study, the stream bed at both sites consisted mostly of large boulder and cobble, like 0A. However, at the end of the study, pebble and sand washed out from a re-mining operation near nearby completely covered the stream bed, and was accumulating into several barriers.

1.2.3 Laurel Run

Like Four Mile Run, Laurel Run is a first order stream that runs through active mine lands. The tributary connects with the Stony approximately two miles downstream from the outfall on the left bank (Map 4). Site LRR is located approximately 20 m upstream from the mouth. Laurel Run has cold, clear water with occasional patches of thick moss on the banks and filamentous green algae in the stream. Like most of the area, the bed is dominated by large boulders. Laurel Run's treatment status is unknown, though yellow-boy is absent.

Sites LR1 and LR2 frame the tributary, with LR1 upstream and LR2 downstream. Both sites are physically similar and both have a shallow run and a deep pool. The stream bed has more cobble and smaller boulder than the upstream sites. Like the sites upstream, flow is largely dependant on dam release.

1.2.4 Downstream

Downstream, the gradient is no longer steep, and the bed has few boulders and is mostly cobble and pebble. Due to the aggregate effect of several tributaries, flow is considerably greater than the upstream sites when the dam is closed. Site SR4 is located in the mainstem of the Stony River immediately downstream of the Route 50 bridge, approximately six miles from the dam outfall.

2 Materials and Methods

2.1 Data Collection

Sites 0A, 4M1, 4M2, LR1, LR2, and SR4 were all initially split into two subsites representing lighted and shaded areas. However, midway though the study, it became apparent that light did not limit periphyton growth on the stony river – shaded areas near the banks often had more biomass than light areas midstream. Growth was limited not by light, but scouring. Data collected from subsites were rolled into the parent site; the result being that those sites in the main stem has twice as many collections as the tributaries.

2.1.1 Periphyton

Periphyton collections were made on four dates, 4-Sep-2004, 7-Nov-2004, 15-Jan-2005, 5-Jun-2005 for each of the nine sites; separate collections were made on each date for the two subsites at each of the six sites in the mainstem of the river.

2.1.1.1 *Field Collections*

Specialized tools were created to collect periphyton in the field. A water-tight cylinder was cut from one inch diameter PVC pipe to approximately two inches high. One end of the cylinder was fit with closed-cell foam to enable to pipe to be pressed against an uneven rock without water leaking from the bottom. A brush was constructed by cutting the head from a plastic toothbrush and reattaching the end of the handle to the back of the head with modeling glue. The result is a small push broom-like brush than can be twisted in the cylinder to scrub periphyton from rocks. A large pipette was also required to transfer the resulting periphyton solution into sample containers; a kitchen baster was used.

Chlorophyll and dry weight collections required filters and vacuum filtering equipment. For dry weight collections, filters were weighed and individually stored in small plastic zipper bags; each weight was then recorded in permanent marker on the outside of the bag.

For each collection, three cobble were chosen haphazardly from the site, giving preference to removable cobble with at least one flat surface. Using the cylinder, brush, and baster, periphyton was removed from a one inch diameter circle on each cobble and placed in a 50 mL centrifuge tube. Additional river water was added to each tube to raise the total volume to 50 mL. Occasionally, due to large amounts of periphyton, the sample was allowed to settle, and water was pipetted from the top to make room for the remainder of the collection.

For chlorophyll and dry weight collections, 17 mL of well-mixed solution was removed from each sample container and filtered using a Micropore filter system and hand vacuum pump. This left 16 mL of solution in each sample, which was then combined into a single aggregate sample, fixed with 0.5 mL of Lugols solution, and then labeled with the site name.

2.1.1.2 *Sample Processing*

Composite periphyton samples were sent to Dr. Robert Verb at Ohio Northern University for taxonomic and density analysis.

Dry weight filters were removed from their plastic bags and placed on top of them. Bags and

filters were then placed on racks which were in turn placed in a drying oven and allowed to dry for at least 72 hours. Filters and sample were then weighed. Dry weight was calculated by subtracting the initial filter weight from this weight.

Samples to be processed for chlorophyll were placed in a 1" wide glass tube, approximately 8" long with 5mL of glacial 90% aqueous acetone. A broad tissue grinder, cut to match the bottom of the tube in shape, with a 10" bolt coming from the head, attached into a normal power drill. The sample and filter were then ground until the filter was completely pulverized. This solution was then decanted into a 15 mL centrifuge tube. The tube was then rinsed with 2 mL of the acetone solution and decanted into the tube. A final volume was then recorded using the gradations on the centrifuge tube; due to evaporation, this volume was most often between 5 and 6 mL.

Samples were spun in a centrifuge for 5 minutes at 1700 rpm. If the samples had not completely sedimented after 5 minutes, they were spun for 1 minute intervals at 1700 rpm until all samples had sedimented. No samples were ever spun more than 8 minutes. Then 3 mL of supernatant from each sample was pipetted into glass cuvettes and absorption measured at wavelengths of 664 nm and 750 nm; the acetone solution was used as a blank. Each cuvette was then treated with 0.1 mL of 0.1 N HCl solution to convert chlorophyll-a (and other similarly-structured magnesium containing pigments) to phaeophytin (or their equivalent phaeo-pigment); absorbance was then read at 665 nm and 750 nm. Because the ratio of chlorophyll-a before acidification to chlorophyll-a after acidification is 1.7, the amount of phaeophytin can be determined (Lorenzen, 1967). Cuvettes were rinsed with acetone solution and allowed to dry between samples.

Chlorophyll-a and phaeophytin, a chlorophyll-a degradation product, were calculated using Lorenzen (1967), which is also the method used by Standard Methods (1998):

$$Chlorophyll\text{–}a\left(\frac{\mu g}{L}\right) = 26.7\left(E_{664} - E_{665a}\right)\frac{V_{ext}}{V_{sample}} \qquad \text{Eq 1}$$

$$Phaeophytin\ \left(\frac{\mu g}{L}\right) = 26.7\left(1.7E_{665a} - E_{664}\right)\frac{V_{ext}}{V_{sample}} \qquad \text{Eq 2}$$

Also, because there are confounding factors arising from multiple pigments (Carlson and Simpson, 1996), Jeffery and Golterman's Total Chlorophylls was calculated (1971):

$$Chlorophyll\ \left(\frac{\mu g}{L}\right) = 11.0E_{664}\frac{V_{ext}}{V_{sample}} \qquad \text{Eq 3}$$

Numeric subscripts of E denote the absorbance wavelength, corrected by the turbidity reading at 750 nm; a in subscripts denotes absorbance readings after acidification. V_{ext} is the volume of the extracted sample; V_{sample} is the amount of liquid filtered, here 17 mL.

Because a known surface area of substrate was sampled, and not a specific volume of water filtered, the above values were corrected:

$$C\left(\frac{\mu g}{cm^2}\right) = C\left(\frac{\mu g}{L}\right)\frac{50mL}{6.4516cm^2} \qquad \text{Eq 4}$$

2.1.2 Bioaccumulated Metals

Metals bioaccumulation samples were collected from each of the 9 sites on 6-Aug-2005 by collecting large quantities of periphyton from points throughout each site. This was then filtered, dried, and sent to the Chemical Analysis Lab at the University of Georgia for metals analysis using ICP-Mass spectrometry. Of particular interest were aluminum, iron, and manganese, metals that typically dominate acid mine drainage streams. Also of interest, because periphyton compose the base of the food web in their ecosystem, were the bioaccumulated toxins arsenic, cadmium, mercury, and lead. However, a total of 28 elements, many nutrients – such as calcium, potassium, and phosphorus – were measured and used in analysis.

2.1.3 Water Quality

YSI datasondes measured four water quality parameters – temperature, specific conductance, dissolved oxygen, and pH –at all sites, with the exception of SR4, at 15 minute intervals for at least a three week period at each site. Data from site SR4 was associated with an independent study.

Datasondes were deployed at 0A and 0B during the summer of 2004 and 2005, at sites around Four Mile Run in the fall of 2004, and at sites around Laurel Run during the Winter of 2004/05, with all collections approximately matching periphyton collections.

Two measures were computed from each parameter for each site, the Median Daily Difference (MDD) and the Median Diurnal Change (MDC). The MDD is a measure of a parameter's consistency between consecutive days. The MDC measures how much a parameter changes within each day.

$$MDD = \text{median}\left(\frac{1}{n}\sum_{i}^{n}\left(Q_{i,j} - Q_{i,j+1}\right)\right)$$
$$MDC = \text{median}\left(\max\left|Q_j\right| - \min\left|Q_j\right|\right)$$

where $Q_{i,j}$ is a matrix with observations of a parameter at a given site at different times of day in rows and different days in columns, and where the max and min functions operate over the observations in a given day, returning a row vector.

2.2 Data Aggregation

At the heart, clustering methods are data reduction methods – taking continuous, multi-dimensional data as inputs and returning a categorical value. To be successful, the data set must be complete – to compare attributes, data must exist for the same time and place in both subsets. As such, only the intersection of the subsets can be used for clustering. Therefore, site SR4 was not used in the clustering due to a lack of water quality data. Additionally, because the composite periphyton collection for site LR2 was lost during a winter auto accident, site LR2 was also omitted from clustering.

The data set is a matrix of observations organized by attribute and site. In order for clustering to succeed, this data must be transformed so that the values of the observations more accurately reflect each site's contribution to defining the character of each attribute. Every data subset must also be standardized, that is, scaled so that all values fall between two arbitrary values so that they can be compared with the other subsets that have different natural scaling. All data was standardized to [0, 1]. The following describes the types of transformations performed on each subset of data. After these transformations were completed, all data was concatenated into a single matrix, A, with attributes on rows and sites on columns; this matrix was then centered row rise by subtracting row means in order to assess variance.

2.2.1 Periphyton

Species biovolume was used to quantify each species at each site. Biovolume is determined by multiplying the mean volume of individuals of a species by the counted number of individuals in a sample. This more accurately reflects the ecological impact a species has on a site over simple counts by adjusting for variations in size. To avoid having many species spuriously clustered together near the origin of the hyperspace simply because they always occur in amounts orders of magnitude less than a few highly prolific species, all data was log transformed to reduce the effect of these outliers. To maximize the range of the log scaling, data was linearly scaled prior to this transformation such that the minimum non-zero value is 1. After the log transformation, data was scaled by dividing all observations by the maximum observation, achieving well-spaced values on [0 1].

$$O_{i,j} = \frac{O_{i,j}}{\min|O > 0|}$$
$$P_{i,j} = \frac{\log\left(O_{i,j} + 1\right)}{\max\left|\log\left(O + 1\right)\right|} \qquad \text{Eq. 5}$$

2.2.2 Bioaccumulated Metals

Bioaccumulated metals were quantified as parts per unit of periphyton dry weight. Because the impact of two different metals on a site depends on more than simply the absolute amounts of those metals present, metals were scaled individually by dividing all observations by the maximum observation for each metal. This results in, for example, 0.016 ppm of mercury being equally associated with 0A as 110,500 ppm of iron is associated with 4MR. Had all data been scaled equally, the mercury data would have been reduced to insignificance.

$$M_{i,j} = \frac{O_{i,j}}{\max|O_i|} \qquad \text{Eq. 6}$$

2.2.3 Water Quality Parameters

Before clustering could be performed on the water quality parameters, the large number of observations for each parameter at each site needed to be reduced. The minimum, mean, and maximum observation for each parameter was taken to describe that parameter's global character at each site. Also, since each parameter showed diurnal fluctuation, each site was described as a

set of median diurnal means and extrema. For each full day of observations, the minimum, mean, and maximum observation was found. The median of each of these was then taken to describe that parameter's diurnal character at each site. MatLab Listing 1: ExtractData, contains the implementation of this characterization.

For periphyton and metals, clustering will result in a list of species and metals that occur together. Similar results could be had with water quality parameters, giving, for example, that species S, metals A and B, and high diurnal-maximum dissolved oxygen all appear together. It is more useful, though, to know how much dissolved oxygen appears. Clustering, however, only returns a group of attributes that occur with each other. In order to get around this limitation, a number of categorical variables were created for each water quality parameter, and each site was rated based on how much that categorical variable was represented by that site.

Each character was assigned four equally spaced centers covering its range of values. The exponential radial basis function was then applied to each character and its four centers. This function acts as an activation function, returning 1 when the value and the center are the same, and decreases to 0 as the distance between value and center increases. The shape parameter, ε, was chosen such that the activation for a value exactly half way between two centers will be 0.5. Given a center, c_i and characterization value for each site, v_j, the water quality attribute for each site:

$$W_{i,j} = \exp\left(\varepsilon(c_i - v_j)^2\right)$$
$$\varepsilon = -\frac{\ln(16)}{d^2} \qquad \text{Eq. 7}$$

Where d is the distance between two adjacent centers. MatLab Listing 2: BuildDataSet, contains this implementation.

2.3 Clustering

By assigning each site an orthogonal axis, attributes may be plotted as a point cloud in a hyperspace by using the observations as coordinates. If assemblages exist, the density of points in the space will be non-uniform; attributes with similar associations will occur in clusters. It is the purpose of clustering methods to find and enumerate these groups.

Three methods for determining these groups, and their strength, were explored. The first creates assemblages visually from a two dimensional projection using the singular value decomposition, a method similar to principle components analysis, a common dimensionality reduction method used by biologists. The second, fuzzy C-means, allows assemblage members to belong to multiple assemblages with varying degrees of membership, a condition that seems reasonable when dealing with ecological data. The last method, k-nearest neighbors, is a simple nonparametric clustering method to allow for assemblages that do not clump neatly into hyperspheres, another condition that seems likely.

2.3.1 Singular Value Decomposition Projection

Data in an 8-dimensional hyperspace is not easily visualized. As such, a view point through which this hyperspace can be flattened, such that a maximum of variability is preserved, is

needed. For example, if we project our own disk-shaped solar-system onto a flat surface, we would not want to look at it on-edge and see a band of stars. Rather, we would want to move to the 'top' so that a roughly circular collection of stars is spread out over our view, knowing that we are willingly sacrificing information about depth.

For an n-dimensional space, an optimal view can be selected by using the singular value decomposition (SVD). The SVD of a matrix will provide a diagonal matrix, Σ, and two unitary rotational matrices, U and V (Tefethen and Bau, 1997), such that:

$$A = U\Sigma V^{\mathsf{T}} \qquad \text{Eq 8}$$

The diagonal of Σ contains the lengths of the 8 semiaxes of the 8-D point cloud described by the data; the length of each semiaxis corresponds to the amount of variance in the data. By convention, Σ is ordered such that the longest semiaxis is in position *(1,1)*, the next longest in *(2,2)*, and so on with the shortest in *(n,n)*. By using only the first two semiaxes with the rotational data from U, we optimally project our data into two dimensions (Tefethen and Bau, 1997). That is, using a Cartesian plane:

$$x_i = U_{i,1}\Sigma_{1,1} \quad y_i = U_{i,2}\Sigma_{2,2} \qquad \text{Eqs 9}$$

This projection is similar to Principle Components Analysis (PCA); however, PCA performs an SVD on the covariance matrix instead of the matrix itself, allowing for additional statistical interpretations. The SVD method gives a different projection than the results from PCA (centering each observation by subtracting its mean would give a projection similar to PCA, but at a different scale). Only an optimal projection is required, and translating points from the projection-space back into an approximate location in the original data space is impossible using PCA, but is simple using the SVD method:

$$B = \begin{bmatrix} x & y & 0 \cdots 0 \end{bmatrix} V^{\mathsf{T}} \qquad \text{Eq 10}$$

Dimensionality projection methods such as PCA are best suited when groups are known, or strongly suspected, to exist in the data. The method then gives clear, well defined clumps. Because there is so much variation in our data, however, such separation is not expected. This method is primarily included to illustrate the usefulness of the other two clustering methods, and also to allow for a common point of comparison by translating the clusters given by the other methods onto a 2-dimensional space.

2.3.2 Fuzzy C-Means

K-means clustering attempts to position cluster 'centers' in the data space such that the sum of the distance squared between points and the nearest center is minimized. The number of clusters must be chosen prior to clustering and does not change while the algorithm runs. Because points are assigned to clusters by choosing the closest center, each cluster is hyper-spherical in shape. Also, because the objective function minimizes distance squared, clusters tend to be of similar size.

This type of clustering partitions a data space and assigns each attribute to exactly one cluster with similar data points belonging to the same cluster. Biological systems, however, do not respect sharp partitioning – organisms can be found in a variety of habitats and gradients exist,

therefore it is often useful to allow attributes to belong to multiple clusters with varying degrees of membership.

Fuzzy c-means is a widely used method derived from k-means clustering. Fuzzy c-means uses the same objective function as k-means, but instead of restricting the membership matrix to a sparse matrix with exactly one '1' per row, the matrix has values on [0, 1] with rows summing to one. Typically, a data point will have a high membership in one cluster, slight membership in a few other clusters, and no membership in the remainder. Similar data points will have high membership in the same cluster(s). Many points split among clusters typically means that more clusters are needed.

Fuzzy c-means clustering operates by minimizing the product of the distance between cluster centers (c_j) and a data point (x_i) by x_i's membership in that cluster (u_{ij}). This is achieved by minimizing the objective function:

$$J_m = \sum_{i=1}^{n}\sum_{j=1}^{k} u_{ij}^{m}\left\|x_i - c_j\right\|^2 \quad 1 \le m \le \infty \qquad \text{Eq 11}$$

where n is the total number of data points, and k is the total number of clusters. By convention, the 2-norm is used, though any norm is sufficient. A 'fuzziness' index, m, determines how sharply clusters are defined; a value of 1 results in traditional hard clustering and a value of 2 normalizes memberships linearly. Higher values relax clusters to the point of near non-usability (Bezdek, 1981). In this study, m=1.5 was used to achieve sharper clusters.

The objective function cannot be minimized directly, thus minimization is achieved via an iterative method in which memberships, U_{ij}, and cluster centers, c_j, are alternately calculated using equations 12 and 13 to approach a local minimum. Initial values for U are chosen from a uniform random distribution.

$$u_{ij} = \frac{1}{\sum_{r}^{k}\left(\frac{\left\|x_i - c_j\right\|}{\left\|x_i - c_r\right\|}\right)^{\frac{2}{m-1}}} \qquad \text{Eq 12}$$

$$c_j = \frac{\sum_{i=1}^{n} u_{ij}^{m} x_i}{\sum_{i}^{n} u_{ij}^{m}} \qquad \text{Eq 13}$$

Because the iterative method converges only to a local minimum, 100 sample clusterings were computed; the one that best minimized the objective function while separating cluster centers was chosen. To determine this, the Xie-Beni index (Xie and Beni 1991), was used with smaller values preferred:

$$S = \frac{J_m}{\min\left\|c_i - c_j\right\|^2 \cdot n} \qquad \text{Eq 14}$$

Fuzzy c-means clustering was implemented using MatLab Listing 3: fuzzyCluster. This function computes the Xie-Beni index from centers after minimizing J_m, not concurrently.

2.3.3 k-Nearest Neighbors

Unlike k-means clustering, which requires the number of clusters to be chosen and then determines optimal placement for their centers, k-nearest neighbors assigns a point membership in the same cluster as its nearest neighbors. The algorithm requires two inputs, k, the number of nearest neighbors to compare, and d_{min}, the minimum distance allowed between points in difference clusters. Assigning points to clusters by distance from the nearest point in a cluster, as opposed to distance from the cluster center, creates clusters in any shape, not just hyper-spheres; lines, abstract shapes, and even concentric circles can be identified (Gitman 1973).

The k-nearest neighbors algorithm first calculates a matrix, D, of Euclidean distances between each point to be clustered. Points within d_{min} of each other are initialized to the same cluster. The algorithm then sorts D, and, for each point i, finds the clusters of the first k entries in row D_i. The most represented cluster is the new cluster for point i, and for every other point in i's old cluster. Thus larger clusters are built from the consolidation of smaller clusters. If the most represented cluster is simply the cluster to which point i already belongs, nothing happens; when this occurs for every point in a cluster, that cluster is stable. The algorithm loops through each point repeatedly until there is no longer change in the cluster assignments.

Because the k-nearest neighbors solution-space has no local minima, no error term is calculated; the optimal solution for any given k and d_{min} is always found. However, the selection of k and d_{min} can dramatically change the solution. A value of k near $2\ln(n)$ is recommended by Wong and Lane (1983). Using this as a guideline, k and d_{min} were selected such that no cluster had fewer than five members, with the understanding that this limitation is arbitrary.

The implementation of the k-nearest neighbors is in MatLab Listing 4: kNearN.

2.3.4 Datasets suitable for the different clustering algorithms

Fuzzy C-means and k-nearest neighbors are suitable for identifying clusters in datasets with different topologies. Figure 2 and Figure 1 show contrived datasets that are appropriate for k-nearest neighbors and fuzzy C-means, respectively, but not for the other algorithm. In these figures, clusters identified by k-nearest neighbors are outlined, and clusters defined by fuzzy C-means are represented in shaded colors.

Determining which algorithm will best identify groups is trivial when all of the variance in the data can be visualized in two dimensions. However, the topology of multidimensional data is not easily determined. As such, both methods are used in this study and their fitness will be evaluated on how well they differentiate groups in the data that are known to differ from the others, such as the communities found at the two tributaries.

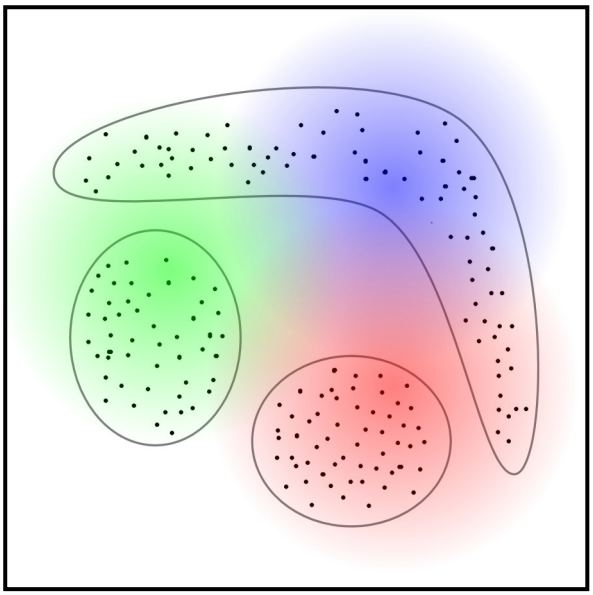


Figure 2. Contrived data points showing when k-nearest neighbors is more appropriate to fuzzy C-means.

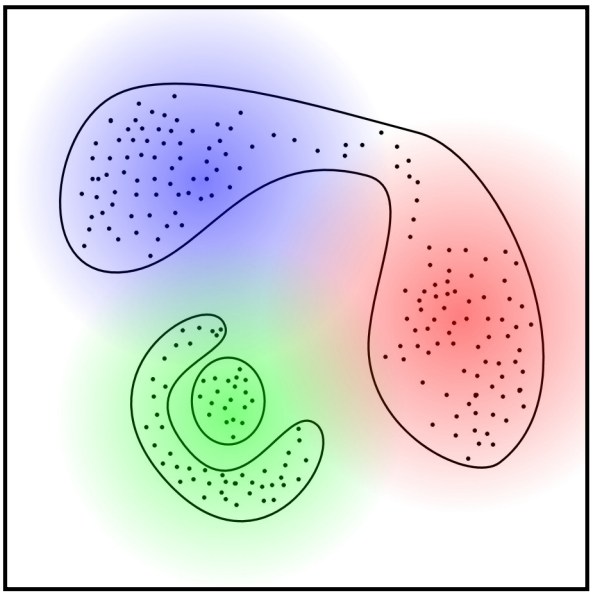


Figure 1. Contrived data points showing when fuzzy C-means is more appropriate to k-nearest neighbors.

3 Results

3.1 Natural History

3.1.1 Stony River Periphyton

Species richness and diversity of periphyton were all calculated using biovolume. This causes filamentous green algae, which are large compared to diatoms, to appear heavily represented. Using simple counts, *Achnathidium minutissimum*, becomes the dominate species for nearly all sites, because, as its name suggests, it is very small. Biovolume was chosen because it better measures site composition and function; however, it is very sensitive to constants given for species volume. Both *Mougeotia* and *Oedogonium* are listed at the genus level, and thus the biovolume is not species specific.

In collections made during September 2004, *Mougeotia* species, a filamentous green algae associated with low pH, was common in 0A and in sites between Four Mile Run and Laurel Run, and in Laurel Run itself, though not downstream. *Oedogonium* species, a genus of filamentous green algae, dominated 4M1. No species truly dominated LR2, though most were diatoms. *Audouinella hermannii*, a holdfast species resistant to high flows, dominated both 4MR and SR4. Site 0B also had no truly dominant species, and was composed mostly of green diatoms (Table 1). Despite the prevalence of filamentous green algae, only 0B and LRR appeared green during collection.

Shannon's diversity index was highest for sites 0B (3.62) and LR2 (3.20); these sites had high evenness, though LR2 had only moderate richness values. Site 0B also had the highest richness (92), though LR1, a well protected site, also had a large number of species (79). The sites around Four Mile Run had the lowest species diversity, though each were dominated by different species. The tributaries had the lowest richness, with 4MR having only 17 species, and most of those contributing under 1%; LRR had 26 species and considerably better evenness (Table 1).

Of the four sites with winter collections analyzed, all of them reduced in species richness, which was expected due to lack of light and colder temperatures. Species diversity increased for every site, largely due to die-off of the dominant filamentous green species. However, these

0A (1.33 / 43)	**88.1%**
Mougeotia spp.	73.2%
Unidentified diatoms	5.1%
Cymbella/Encyonema spp.	2.9%
Cosmarium botrytis	2.6%
Ulothrix spp.	2.2%
Achnathidium minutissimum	2.0%
0B (3.62 / 92)	**62.5%**
Denticula kuetzingii	13.3%
Achnathidium minutissimum	8.9%
Cymbella affinis	8.1%
Tabellaria flocculosa	6.2%
Pinnularia microstauron	4.5%
Synedra ulna	3.7%
Nitzschia spp.	3.6%
Unidentified diatoms	3.4%
Gomphonema spp.	3.0%
Aulacoseira granulata	2.8%
Surirella spp.	2.7%
Navicula spp.	2.3%
4M1 (0.90 / 66)	**86.9%**
Oedogonium spp.	84.8%
Vaucheria spp.	2.1%
4MR (0.43 / 17)	**92.8%**
Audouinella hermannii	92.8%
4M2 (0.57 / 60)	**91.4%**
Mougeotia spp.	91.4%
LR1 (1.03 / 79)	**83.2%**
Mougeotia spp.	83.2%
LRR (1.49 / 26)	**95.5%**
Mougeotia spp.	53.6%
Oedogonium spp.	18.7%
Microspora tumidula	13.5%
Geminella minor	3.6%
Klebsormidium rivulare	3.4%
Eunotia exigua	2.6%
LR2 (3.20 / 61)	**73.5%**
Synedra ulna	16.5%
Cosmarium sp.	13.0%
Achnathidium minutissimum	8.9%
Brachysira vitrea	5.1%
Unidentified diatoms	5.1%
Frustulia rhomboides	4.4%
Gomphonema spp.	4.2%
Aulacoseira granulata	4.1%
Gomphonema cf entolejum	3.7%
Synedra tenera	3.1%
Ctenophora pulchella	2.8%
Cymbella/Encyonema spp.	2.7%
SR4 (1.11 / 66)	**87.2%**
Audouinella hermannii	79.9%
Closterium acerosum	4.4%
Synedra ulna	2.9%

Table 1. Periphyton species with greater than 2% share in September sites. Shannon's Index and species richness follow site name.

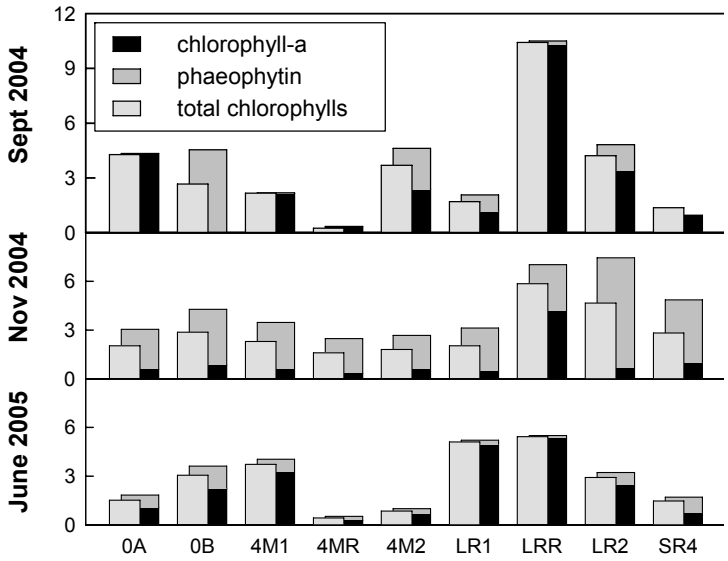


Figure 3. Mean pigment amounts in $\mu g \cdot cm^{-2}$ from spectrophotometry using Lorenzen's formula for chlorophyll-*a* and phaeophytin, and Golterman and Clymo's total chlorophylls over three collection dates. Bar scale is consistent for all dates.

0A (2.54 / 40)	**80.3%**
Mougeotia spp.	34.8%
Gomphonema parvulum	11.1%
Achnathidium minutissimum	10.0%
Synedra ulna	5.7%
Navicula lanceolata	5.6%
Encyonema silesiacum	3.9%
Fragilaria sp.	3.4%
Unidentified diatoms	2.9%
Ctenophora pulchella	2.9%
4M1 (2.10 / 58)	**78.5%**
Mougeotia spp.	52.0%
Synedra ulna	15.9%
Achnathidium minutissimum	3.4%
Pinnularia spp.	2.7%
Cymbella/Encyonema spp.	2.5%
Synedra parasitica	2.1%
4M2 (1.26 / 46)	**87.3%**
Oedogonium spp.	74.7%
Synedra ulna	7.3%
Phormidium amoenum	3.3%
Unidentified diatoms	2.1%
SR4 (3.16 / 55)	**70.1%**
Synedra ulna	24.3%
Diatoma tenue	8.9%
Achnathidium minutissimum	7.5%
Encyonema muelleri	5.4%
Fragilaria sp.	3.6%
Diatoma moniliformis	3.4%
Gomphonema spp.	3.3%
Nitzschia spp.	3.0%
Brachysira vitrea	3.0%
Aulacoseira granulata	2.9%
Pinnularia spp.	2.8%
Cymbella/Encyonema spp.	2.1%

Table 2. Periphyton species with greater than 2% share in January sites. Shannon's Index and species richness follow site name.

species still represented the most biovolume of all species present. Total biovolume also decreased over seven-fold from the summer to the winter; though this could also be explained by the duration between collection and the last scouring event (Table 2).

Chlorophyll-a from spectrophotometry is an approximation, and various formulae exist. Golterman and Clymo's (1971) total chlorophylls is simply a scaling of the absorbance at 665 nm; whereas Lorenzen's (1967) formula attempts to separate chlorophyll from its degradation product, phaeophytin. Lorenzen's chlorophyll-a is a scaling of absorbance at 665 nm minus the absorbance after acidification; thus, though called 'chlorophyll-a' it measures all chlorophylls minus their degradation products. Because of this, when Lorenzen's chlorophyll-a and phaeophytin are added, they should approximately equal Golterman and Clymo's total chlorophylls; however, the two together almost always exceed it (Figure 3).

Phaeophytin is created when chlorophyll degrades, either before or after collection, confounding any determination of site health from phaeophytin levels. However, assuming that post-collection degradation affects all samples in a collection evenly, relative site health may be predicted. From this, using data from all collections, one may conclude that LRR was a healthier site than 0B or LR2 during each collection (Figure 3); chemical data contradicts this however. Possible error may result from simply assuming that high chlorophyll levels are correlated with health. This assertion is often true in terrestrial systems, but with periphyton may simply indicate a lack of predation or the presence of well adapted species. Better to conclude that LRR supported a stronger periphyton population; combining this with other characters of that site

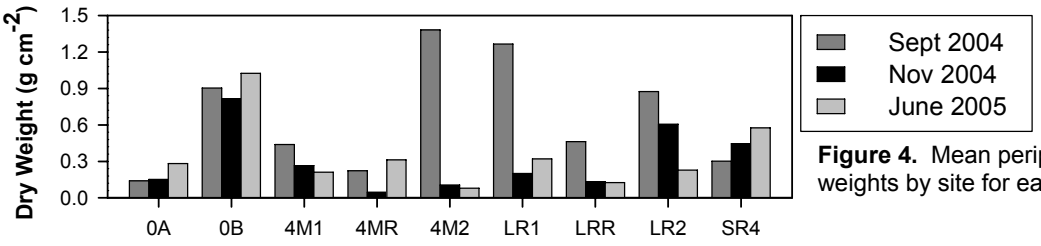


Figure 4. Mean periphyton dry weights by site for each season.

might help to identify potential periphyton species that themselves indicate poor site health where they flourish.

Dry weights of samples collected in June 2005 were destroyed when attempting to determine ash-free dry weight. Figure 4 shows the dry weights of all collections, however. Because diatoms are heavy due to their silica casings, these values are not useful for determining biomasses comparable between sites. However, this data does clearly show the stability of some sites, such as 0B versus others, such as 4M2 and LR1.

3.1.2 Bioaccumulation of Metals

Figure 5 shows bioaccumulated concentrations of metals that tend to dominate AMD streams. Iron is the dominant metal in all sites except for LRR which is dominated by manganese; There is a strong introduction effect of iron around 4MR, with concentrations increasing from 2.7 mg/g at 4M1 to 14.5 mg/g at 4M2. No introductory effects are seen around LRR, possibly because, due to the structure of the river, periphyton at LR1 is shielded from scouring, persists longer, and so bioaccumulates more metals than periphyton at LR2.

Figure 6 shows the bioaccumulated concentrations of selected toxins; Hg concentrations were negligible and are omitted. 4MR had the highest toxin bioaccumulation, followed by LR1. It seems most likely that the metals concentration at 4MR is highly affected by yellow boy covering the periphyton. The high concentrations at LR1 are also likely caused by the persistence of periphyton at that site. However, there is a general increase in bioaccumulated metals as sites move downstream.

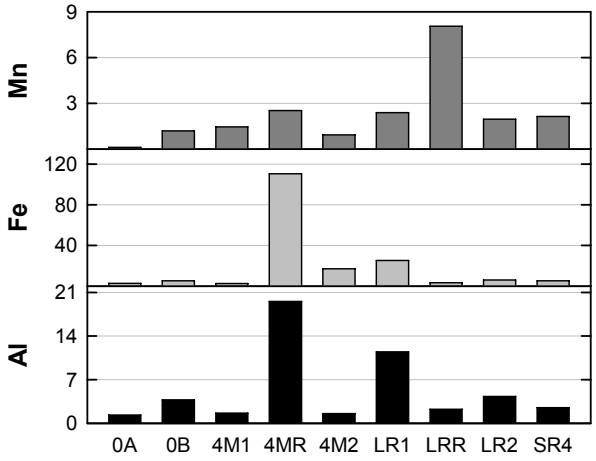


Figure 5: Concentrations (mg/g) of metals common in acid mine drainage affected streams.

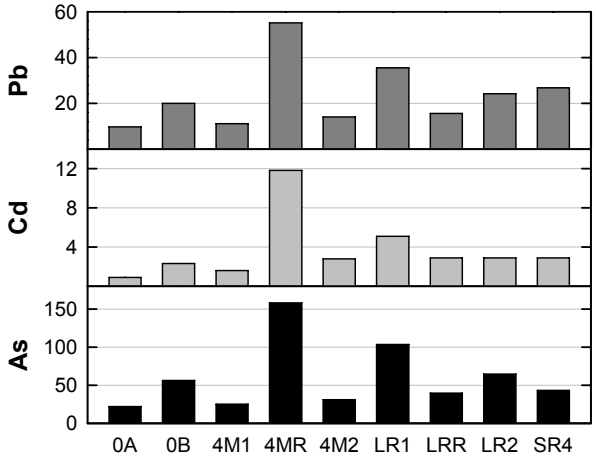


Figure 6. Concentrations (µg/g) of selected toxins.

3.1.3 Stony River Chemistry

Figure 5 shows means taken over the entire sampling period for water quality parameters at all sites. Sites 0A and 0B were sampled during the summer, sites in and around Four Mile Run during the fall, and sites in and around Laurel Run during the winter.

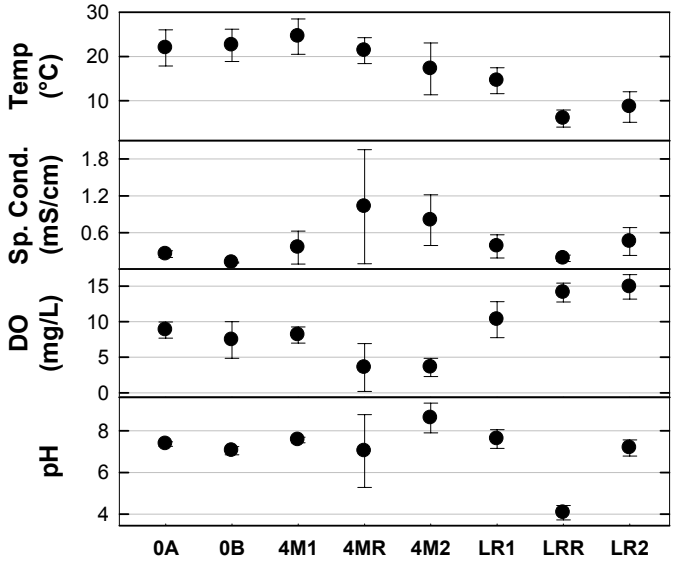


Figure 5. Means taken over the entire sampling period of water quality parameters of each site. Error bars cover one standard deviation.

Temperature is well-correlated with season, with mean summer temperatures of 22 C and mean winter temperatures of 10 C. The LRR tributary is also cooler than the mainstem, which is expected for the smaller stream.

Specific conductance is much higher and much more volatile in 4MR than other sites, with a possible introduction effect seen in 4M2. The datasonde at 4MR was in an area of the tributary with constant flow, and thus concentration effects from evaporation cannot explain the wide range of values measured.

Elevated dissolved oxygen in sites in and around Laurel Run are consistent with higher solubility at lower temperatures and slowed metabolisms/ less respiration during winter. The very low dissolved oxygen in 4MR and 4M2, however, are unlikely to be physically induced, as 4M2 is structurally the same as 4M1. Indeed, this appears to be another introduction effect (possibly indirect) from 4MR.

The majority of sites on the Stony River are mildly alkaline, possibly due to over-treatment of AMD. While collecting topological data at Four Mile Run in a separate study, the smell of ammonia was prevalent; presumably from alkaline treatment. The volatility of pH at 4MR suggests that this treatment is intermittent. Laurel Run appears to be untreated with mean pH of 4.07. Additionally, 4M2 is highly alkaline with mean pH of 8.61. The sampling period for 4M2 was shorter than that for 4MR; similar volatility may have been seen if the sampling periods were concurrent.

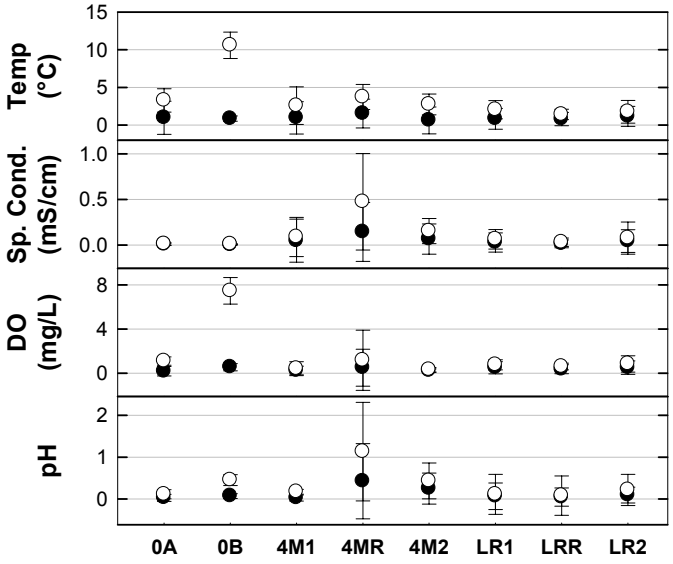


Figure 6. Median daily differences (filled circles) versus median diurnal changes (open circles) for all sites. Error bars cover one standard deviation.

Figure 6 shows the median daily difference (MDD) and the median

diurnal change (MDC) of water quality parameters with their standard deviations. Sampling periods were the same as for Figure 5.

The MDD is the median of the differences between consecutive days and gives a measure of a site's day to day stability. Values near zero indicate either a stable site or a site in dynamic equilibrium; sites with large MDD values experience large sudden changes. The MDC is the amount of change that occurs within any given day. MDC values may never be less than the MDD values; if the two are equal, then the diurnal change is simply a reflection of the day to day trend. However, if the MDC is larger than the MDD, the site experiences at least some diurnal cycle. In Figure 6, these are sites where the circles are well separated. If the standard deviations of these values do not intersect, then diurnal cycles dominate the site's chemical trends.

A diurnal cycle of temperature is indicative only of solar heating. All sites show some separation of the MDD and the MDC, indicating that they have some diurnal cycle; well shaded and winter sites exhibited less. Only 0B, which is very open, had trends dominated by diurnal heating. Only sites around Four Mile Run show a separation that suggests a diurnal trend of specific conductance; however, no site's specific conductance was dominated by diurnal trends. Dissolved oxygen follows a diurnal trend only at 0A and 0B; pH only at 0B.

3.2 Clustering

3.2.1 Singular Value Decomposition Projection

The first two eigenvectors of UΣ, and thus the PCA plot, represent only 54% of the variance in the combined-data matrix, A. The x-axis represents 34% of that variance. Because nearly half of the variance is not represented, the PCA/SVD plot is not a useful tool to accurately establish clusters within our dataset. It is, however, still useful for the visualization of high-dimensional data in 2 dimensions.

3.2.2 Fuzzy C-Means

The dataset was separated using fuzzy c-means into six clusters (J1.5=44.34 and Xie-Beni index of 0.0090). The black cluster is the most central (cluster center 0.09 from the origin). Blue is also highly central, but skewed toward 0B. A central cluster center is indicative of traits shared by many sites equally. The black and blue clusters are also the most central on the SVD projection. The green cluster is centrist, but not strongly so. It captures characteristics shared by 0AL, 0B, and 4M1 but which are not present in downstream sites and the two tributaries.

Memberships of individual periphyton species, water quality parameters, and metals in each cluster are visualized in the figures of Appendix 3. Black areas represent species, parameters, or metals that have memberships in a cluster near 100%. The sum of a given species, parameter or metal's memberships in all clusters equals 100%.

The blue, black, and green clusters are all largely composed of periphyton taxa and approximately 92% of periphyton taxa have the bulk of their membership in those three clusters. For the most part, these clusters represent the 'long tail' of periphyton taxa with very low densities and which were perhaps found at only one site. Taxa associated with

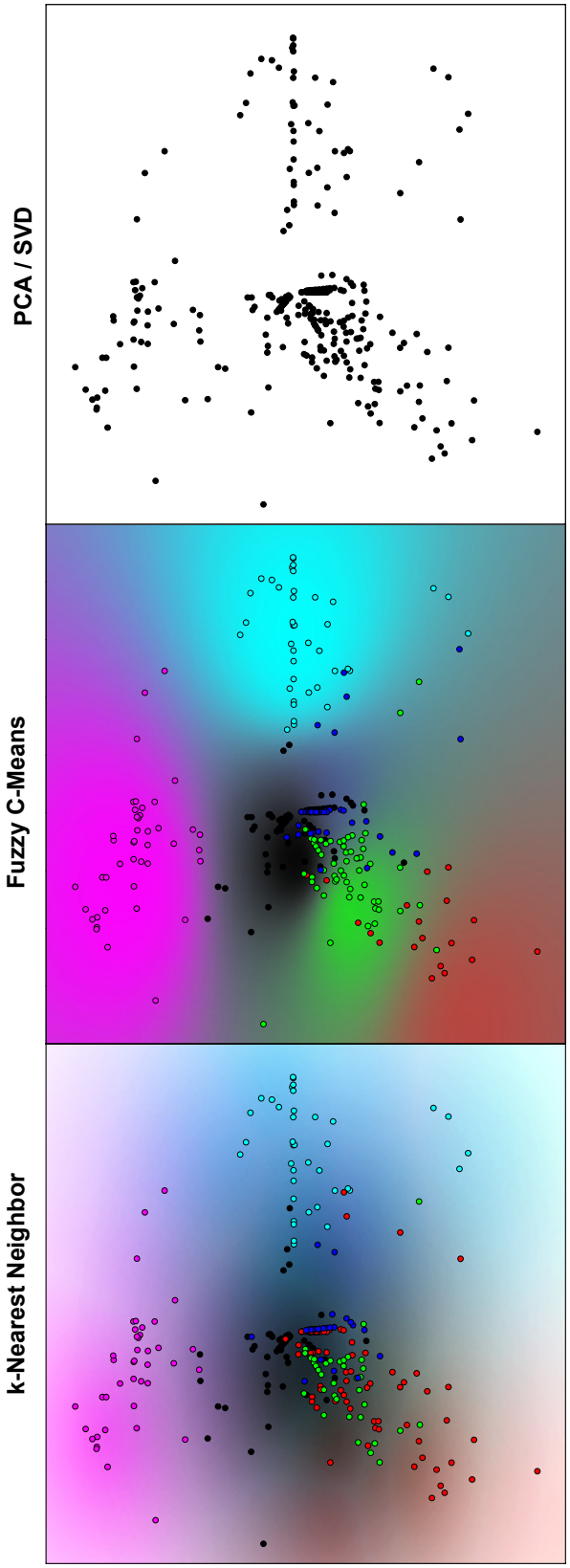


Figure 9. SVD Projection plots of the data, and each clustering result. Fuzzy-C-means colors determined using the highest membership.

	0AL	0AS	0B	4M1	4MR	4M2	LR1	LRR
blue	-0.02	0.02	0.20	-0.03	-0.06	-0.03	-0.03	-0.04
black	-0.00	0.06	-0.02	-0.01	-0.03	0.04	-0.00	-0.04
green	0.00	0.07	0.03	0.23	-0.12	-0.03	-0.08	-0.11
red	0.32	-0.02	0.19	0.36	-0.35	-0.34	0.17	-0.34
cyan	-0.09	-0.08	-0.05	-0.10	-0.08	-0.13	-0.06	0.59
magenta	-0.22	-0.22	-0.15	-0.18	0.66	0.12	0.13	-0.15

Table 3 Fuzzy c-means cluster centers.

many sites at a similar density (regardless of magnitude), such as *Oedigonium*, *Mougeotia*, *Fragilaria*, and *Zygnema*, have cluster memberships somewhat equally distributed over these clusters. The blue cluster has no water chemistry regime strongly associated with it. The black cluster is associated with a medium-high specific conductance (median diurnal maximum of 0.65 mS/cm, global maximum of 1 mS/cm). The green cluster is associated with warm water, with a median diurnal mean of 22 C, and low specific conductance (global minimum of 0.15 mS/cm). Many water quality measures are distributed over the three clusters, however. Collectively, they define parameters that are moderate and well suited for a variety of taxa – pH between 6 and 7, DO between 5 and 8 mg/L, temperatures between 12 and 18 C, and low to moderate specific conductance. No metals are strongly associated with any of the three clusters individually or taken as a group. However, phosphorus and silicon are the strongest. Mercury is associated evenly with all clusters, most likely because very little was present in any sample.

The red cluster is strongly associated with sites 0AL and 4M1, and strongly disassociated with sites 4MR, 4M2, and LRR, suggesting a significant difference between sites upstream from both tributaries and the tributaries themselves. On closer examination, however, only *Ulothrix spp*. and *Cymbella affinis* are strongly associated with this cluster, and they were equally associated with the green cluster. A few other taxa have weak memberships in this cluster. Also, no bioaccumulated metals are associated with the red cluster. Instead, the red cluster has most of its memberships from the water quality parameters. Here, it represents slightly alkaline pH (7-8), high dissolved oxygen (8-11 mg/L), and moderate specific conductance (0.35 mS/cm). Given the method in which the water quality data was created, and the lack of taxa and metals clustered in this group, it is most likely that the red cluster represents the combination of water quality parameters that are not found at any Stony site.

The cyan cluster is strongly associated with LRR and moderately disassociated with every other site, suggesting that LRR is unique. The periphyton taxa dominant in LRR are , unsurprisingly, found in the cyan cluster – *Oedogonium spp*, *Mougeotia spp*, and *Microspora tumidula*. This cluster is associated with globally high dissolved oxygen (14 mg/L), low pH (4 to 6), and cold temperatures (4-10 C). It is also associated with sodium, manganese, magnesium, calcium, barium, and potassium bioaccumulation.

The magenta cluster is strongly associated with 4MR and somewhat associated with the downstream sites 4M2 and LR1, but disassociated with all sites upstream from 4MR and from LRR, which is out of the main stem. This suggests that 4MR significantly changes the biological and chemical make up of downstream sites. Only one periphyton species is associated with the magenta cluster, *Audouinella hermannii*, the species which is dominant in 4MR, and it shares membership with the centrist clusters. Biochemically, the magenta cluster represents alkaline water (pH = 9), very low dissolved oxygen (1 to 4 mg/L), and very high specific conductance (global max of 1.4 mS/cm and median diurnal maximum of 0.9 mS/cm). Additionally, all of the remaining metals are strongly associated with the magenta cluster.

3.2.3 k-Nearest Neighbors

The dataset was separated into six clusters using k-Nearest Neighbors with d_{min} of 0.55 and k equal to 11 (the number of neighbors suggested by Wong and Lang). Memberships in clusters are given in Appendix 4. Clusters produced using k-nearest neighbors correlate well with clusters produced using fuzzy c-means. Coloration in Figure 7 has been set such that clusters with similar constituents are colored the same. The cyan, magenta, and black clusters are closely correlated, whereas the green, red, and blue clusters differ.

The cyan cluster closely represents the periphyton community at LRR. Metals associated with this cluster are barium, calcium, potassium, magnesium, manganese, and sodium. Temperatures are low (globally measurements between 4 and 10 C), specific conductance is low (global and median diurnal means of 0.15 mS/cm), dissolved oxygen is high (global measurements between 10 and 14 mg/L), and pH is acidic but with circum-neutral events (median diurnal range of 4 to 5, but global range of 4 to 6).

The black and red clusters both have large numbers of periphyton species and few metals. Dominant species at site 0B are all found in the red cluster. Silicon is associated with black and mercury and phosphate is associated with red. The black cluster is associated with temperature events ranging from 12 to 17 C, and with unstable specific conductance that ranges from moderate to high (global range 0.20 to 1.00 mS/cm, median diurnal range of 0.34 to 0.65 mS/cm). Dissolved oxygen of approximately 11 mg/L and a stable pH of 8 is also associated with Black. The red cluster is a warm water cluster (global range of 16 to 31 C) with lower specific conductance tolerance (0.20 mS/cm). It is also associated with diurnally fluctuating dissolved oxygen (range of 5 to 12 mg/L) and a stable, neutral pH.

Neither the green nor blue cluster are associated with any metals. Periphyton species in green are not associated with any specific pH, but are associated with moderate daily temperature fluctuation of warm water (median diurnal range of 22 to 27). It is also associated with fluctuating specific conductance (0.15 to 0.40 mS/cm) and slightly alkaline pH (7-8). The blue cluster is not associated with any temperature or specific conductance regime, but is associated with medium low dissolved oxygen (4 to 6 mg/L) and slightly acidic pH (6).

Finally, the magenta cluster contains no periphyton species, but a large number of metals. These metals are associated with a wide range of water temperature (global range 8-24 C), high and unstable specific conductance (global range 0.30 to 1.40, median diurnal range of 0.57 to 0.90 mS/cm), low DO (range between 1 and 5 mg/L), and high stable pH (9); an apt description of site 4MR.

4 Discussion

4.1 Periphyton Species at Sites

Biovolume was chosen to represent taxa abundance in this study under the assumption that it is the amount of species that influences abiotic factors. Because of this choice, however, many sties were dominated by the macroalgae genera *Mougeotia* and *Oedigonium*, even though during collection diatoms were visually dominant at most sites. Counts have the opposite problems of artificially inflating the importance of small taxa. Verb and Vis (2005) use a relative importance value (RIV) in their calculations which incorporates simple counts and cell density as well as biovolume. They found that the RIV, however, was unable to describe ecological variance any better than traditional methods. Instead, it may be best to separate macroalgae from other periphyton when calculating dominance, and report values for both groups.

Table 4 matches Stony River sites to the multivariate groups described by Verb and Vis (2005) using pH, dissolved oxygen, and specific conductance, and gives the dominant species they share. Verbs groups were strongly defined by the pH and specific conductance axes, so the reduced parameters of comparison should not greatly influence a site's group placement. Periphyton taxa were predicted well for sites 0B, LRR and LR2, somewhat for site 0A, and poorly for the remainder. Although the lack of correspondence is partially due to the use of biovolume instead of RIV, Verb and Vis give a separate list of macroalgae associated with each group which did not include the dominants seen at the Stony sites.

Site	V&V Group	Shared Dominants
0A	IV	*Achnathidium minutissimum* *Cymbella/Encyonema spp.*
0B	IV	*Achnathidium minutissimum* *Cymbella affinis* *Synedra ulna*
4M1	IV	
4MR	III	
4M2	III	
LR1	IV	
LRR	V	*Eunotia exigua* *Klebsormidium rivulare* *Microspora tumidula* *Mougeotia spp.*
LR2	IV	*Achnathidium minutissimum* *Brachysira vitrea* *Cymbella/Encyonema spp.* *Synedra ulna*

Table 4. Comparison of dominant species from Stony River sites and multivariate groups reported by Verb and Vis (2005). Shared dominants are listed alphabetically.

The spottiness of the correspondences confirm that strong relationships strictly between abiotic parameters and taxa do not exist. Although they are linked and some correlation exists, these relationships are nebulous. Because the Verb and Vis study concerned itself only with southeastern Ohio streams, it may also be the case that different taxa form assemblages under the same abiotic parameters depending on locality or other less obvious factors.

4.1.1 Changes in Species September to January

Audouinella hermannii, a species associated with low pH was found in SR4 in September in very large amounts; no *A. hermannii* was found in the winter collection. This discrepancy could be due to either a shift in the mainstem pH, possibly due to treatment, or to the mainstem temperature shifting below the tolerance of *A. hermannii*. More data is necessary to draw conclusions.

In September, 4M1 had very high levels of *Oedigonium spp*, a genus associated with

circumneutral pH; 4M1 was dominated by *Mougeotia spp*, a genus associated with pH of approximately 3-6. However, in the January collection, the dominances for the two sites switched, suggesting that pH around Four Mile Run is prone to fluctuation. The pH tolerances of the two taxa overlap (pH 5.86 to 6.31), so both could survive concurrently, with dominance shifting after a scouring event. Such an event would free up habitat to be colonized by the genus for which pH is the most favorable at the time. It is likely that even though pH may fluctuate after the event as well, the currently dominant species remains dominant until the next scouring event provides free strata for colonization. This switch could also be explained by the seasonal temperature change, in which case the fluctuation may be a perennial dynamic.

4.2 Bioaccumulation of Metals

Vymazal found that metals uptake in periphyton communities is proportional to the concentration of metals in the water. Constants of proportionality differ by metal and community composition. However, periphyton communities reduced all metals to slight concentration in just four hours of exposure (Vymazal 1984). Clearly periphyton represent a major reservoir of metals in lotic systems. However, the concentration of metals in the water cannot be determined by the concentration of metals in periphyton because even low levels will bioaccumulate over long periods of time due to replenishment from flow.

Additionally, periphyton species that bioaccumulate metals will only exist in places where metals levels are below toxicity. Conversely, those species that occur in streams high in metals may be poor bioaccumulators. Therefore, periphyton bioaccumulation data does not directly tell us anything about the metals actually present in a lotic environment, though it may tell us something about species that can accumulate metals safely. As such, high levels of bioaccumulated metal may be indicative of metals concentrations below toxicity, whereas metals levels above toxicity destroy the periphyton and prevent bioaccumulation.

4.3 Data Models

The models that are chosen to represent parameter impacts directly influence the results of clustering. That is, the measure that represent a parameter's importance and the transform of parameter into a normalized form must take into account the way the parameter impacts its environment. Choosing these is not straightforward, as before impacts from any given parameter can be quantified, they must be clearly identified. Since the predominant paradigm is to view the environment as impacting taxa and not as taxa impacting the environment, this work is scarce in the literature, leaving us to use best reasonable guesses.

4.4 PCA/SVD as a Visualization Tool

As a clustering tool, PCA/SVD projection is not useful; the two reduced dimensions do not contain enough of the variability in the data to make spatial associations reliable. However, as a dimensionality reduction method for the visualization of clusters formed using other methods, SVD projection is well suited. Additionally, by examining the clustering results in the SVD projection, clusters that are very different are easily identified, as are those that are similar. The ability to visualize the extent of cluster regions is also important, allowing the researcher to see what points are on cluster boundaries – which is particularly interesting in 'hard' clustering methods. As such, although the clustering work may be done with a more complete algorithm,

reporting results as an SVD 'map' provides additional useful information.

4.5 Clustering

In both the fuzzy C-means and k-nearest neighbors clustering methods, the cyan cluster represented the community at LRR and the magenta cluster represented 4MR. The metals lists for both clusters were the same, as were the biochemical parameters. Both methods also emphasized that no periphyton taxa are truly associated with the conditions found at 4MR. The cyan and magenta groups are also the most distinct on the PCA/SVD projection. This suggests that the two tributaries are distinctly different communities than the mainstem of the Stony. This in and of itself is not surprising, but does provide confirmation that the clustering methods were able to find valid subgroups in the data.

In both clustering methods, the green, blue, and black clusters all contain taxa present in the mainstem of the river. It is most likely that these species form a continuum of tolerances and thus a continuum of community. Although clustering has drawn mathematically optimal lines to create groups, these groups are more than likely simply mathematical constructs. Data from more sites is necessary to discover assemblages within this super-community. Finally, these clusters include a large number of taxa that were found only once and in very small quantity. Many of these taxa are likely most strongly associated with some biochemical regime that is outside of the range of regimes found on the Stony River. As such, it bears reiterating that for these species, any results are only valid on the Stony River.

When clustering with fuzzy C-means, the red cluster emerged because some water quality parameters had a 'gap' between low and high values which was not represented on the Stony. However, because some taxa (*Ulothrix spp* and *Cymbella affinis*) were found at sites with both low and high values, they were clustered together with the middling values that were generated algorithmically. This is a clear success for using fuzzy C-means clustering for this purpose – it is capable of identifying groups that never explicitly occur in the data.

When clustering with k-nearest neighbors, again the red cluster showed the algorithm's strength. The k-nearest neighbors algorithm was able to differentiate the three sites outside of the mainstem of the Stony River from the three sites that were, and created clusters that represent composites of the mainstem sites. The red cluster represented the same biochemical regime in both clustering methods; however, using k-nearest neighbors, the red cluster also included periphyton taxa dominant at site 0B. It is likely that site 0B has a topology similar to the top cluster in Figure 2, and so was able to be identified using the nonparametric method.

5 References

1. Besser, JM; Brumbaugh, WG; May, TW; Church, SE; Kimball, BA. (2001) Bioavailability of Metals in Stream Food Webs and Hazards to Brook Trout (*Salvelinus fontinalis*) in the Upper Animas River Watershed, Colorado. Archives of Environmental Contamination and Toxicology 40(1): 48-59.

2. Bezdek, J. C. (1981). Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York

3. Carlson, R.E. and J. Simpson. (1996). A Coordinator's Guide to Volunteer Lake Monitoring Methods. North American Lake Management Society. 96 pp.

4. Cattaneo, A; Couillard, Y; Wunsam, S; Courcelles, M (2004). Diatom taxonomic and morphological changes as indicators of metal pollution and recovery in Lac Dufault (Quebec, Canada). Journal of Paleolimnology 32(2): 163-175.

5. Gitman, I. (1973) An Algorithm for Nonsupervised Pattern Classification. IEEE Transactions on Systems, Man, and Cybernetics, SMC-3: 66-74.

6. Golterman, H.L. and R.S. Clymo. (1971). Methods for Chemical Analysis of Fresh Waters. IBP Handbook No. 8. Blackwell Scientific.

7. Hill, BH; Lazorchak, JM; McCormick, FH; Willingham, WT (1997). The effects of elevated metals on benthic community metabolism in a Rocky Mountain stream. Environmental Pollution 95(2): 183-190.

8. Jeffrey S. W., Humphrey G. F. (1975). New spectrophotometric equation for determining chlorophyll a, b, c1 and c2. Biochem. Physiol. Pflanz 167: 194-204

9. Keithan, E; Barnese, L. (1989). Effects of pH and nutrients on periphyton colonization. Journal of Phycology 25 suppl: 8.

10. Keithan, ED and Low, RL (1985) Primary productivity and spatial structure of phytolithic growth in streams in the Great Smokey Mountains National Park, Tennessee. Hydrobiologia 12: 59-67.

11. Lorenzen C. J. (1967). Determination of chlorophyll and phaeopigments: spectrophotometric equations. Limnology and Oceanography 12: 343-346.

12. Lowe, RL, Golladay SW, and Webster JR. (1986) Periphyton response to nutrient manipulation in streams draining clearcut and forested watersheds. J. N. Benthological Society 5: 221-229.

13. Mcfarland, BH; Hill, BH; Willingham, WT (1997) Abnormal *Fragilaria spp*. [Bacillariophyceae] in streams impacted by mine drainage. Journal of Freshwater Ecology 12(1): 141-149.

14. Marker AFH (1976). The benthic algae of some streams in southern England. I. Biomass of the epilithon in some small streams. Journal of Ecology 64: 343-58.

15. McIntire CD (1968). Structureal characteristics of benthic algal communities in laboratory streams. Ecology 49: 520-37

16. McIntire CD, (1973). Periphyton dynamics in laboratory streams: a simulation model and its implications. Ecological Monographs 43: 399-420.

17. Passy SI. (2006). Diatom community dynamics in streams of chronic and episodic acidification: the roles of environment and time. Journal of Phycology 42: 312–323.

18. Power ME and Stewart AJ (1987). Disturbance and recovery of an algal assemblage following flooding in an Oklahoma stream. American Midland Naturalist 117: 333-345.

19. Robinson CT and Rushforn SR (1987). Effects of physical disturbance and canopy cover on attached diatom community structure in an Idaho stream. Hydrobiologica 154: 149-159.

20. Standard methods for the examination of water and wastewater. (1998.) 20th edition. American Public Health Association, Washington, D.C.

21. Tefethen L. N. and Bau D. (1997) Numerical Linear Algebra. SIAM, Philadelphia.

22. Verb RG, Vis ML (2005). Periphyton assemblages as bioindicators of mine-drainage in unglaciated western Allegheny plateau lotic stystems. Water, Air, and Soil Pollution 161: 227-265.

23. Vuori K-M. (1995). Direct and indirect effects of iron on river ecosystems. Annales Zoologici Fennici 32(3): 317-329.

24. Vymazal, J (1984). Short-term uptake of heavy metals by periphyton algae. Hydrobiologia 119(3): 171-179.

25. Whitton BA (1975). Algae, in Rover Ecology. University of California Press, Berkeley, CA pp 81-105.

26. Wong MA, Lane T. (1983) A kth Nearest Neighbour Clustering Procedure. Journal of the Royal Statistical Society, Series B (Methodological) 45(3): 362-368

27. Xie X. L. and Beni G. (1991). A validity measure for fuzzy clustering. IEEE Trans. Pattern Analysis and Machine Intelligence 13: 841-847.

Appendix 1: MatLab Listings

Listing 1: ExtractData

```
function [Q mn mx] = ExtractData(Files, ext, dir)
% [Q mn mx] = ExtractData(Files, ext, dir)
% Extract Data accepts a list of files to read (ext can specify a common
% extension, defaulting to '.csv' and dir a default directory).  The
% function then Reduces that data and stores it in a 3-D matrix.  The
% minimum and maximum median diurnal values are shown for each parameter
% taken over all files.
%
% The function returns those min and max median diurnal values as well as
% the reduced data from all files.  Columns of Q are min, mean, and max.
% Rows are data from each file/site, and pages are each parameter.

%% Initialize
if nargin < 2       ext = '.txt';
elsif nargin > 2    cd dir;
end

%% Reduce each file passed
for f = 1:length(Files)
    A = load([char(Files(f)) ext]);
    Q(:,:,f) = ReduceData(A, 96);
end

%% Return min/max data to choose centers
mn = min(Q,[],3)
mx = max(Q,[],3)

Q = permute(Q, [3 1 2]);


function P = ReduceData(A, t)
%P = ReduceData(A, t)
% A is a matrix with cyclic data in rows and parameters in columns.
% t is the period (96 for 15 minute data collection intervals).
% ReduceData returns matrix with the median diurnal min, mean, and max as
% rows for each parameter.

%% Find Global observations
P(1:3,:) = [min(A); mean(A); max(A)];

%% Remove any partial cycles
[n, m] = size(A);
r = mod(n, t);
if r > 0
    p = floor(r/2)+1;                       % if data is not on a cycle boundary
    q = ceil(r/2);                          % remove half of the remainder from the
    A = A(p:(n-q),:);                       % beginning, and half from the end
    n = length(A);
end

%% Reshape the matrix with t columns and m pages
A = shiftdim(A, -1);
A = reshape(A, t, [], m);

%% find median diurnal min, mean, max
P(4,:) = median(min(A));
P(5,:) = median(mean(A));
P(6,:) = median(max(A));
```

Listing 2: BuildDataSet

```
function [Z U] = BuildDataSet(Q, W)
%[Z U] = BuildDataSet(Q, W)
% Build dataset accepts the datamatrix from ExtractData, Q, as well as a
% command matrix detailing where centers for data comparison are.
%
% W has min, mean, and max on the rows, a beginning center, an ending
% center, and the number of centers on columns.  Pages have each
% parameter.
%
% Z is a linearized list of parameters/observation types/centers on rows and
% sites as columns.  U is a key to the linearization.

[m l n] = size(W);
t = size(Q);
t = t(1);

v = sum(sum(W(:,3,:)));                                 % number of 'observations'
R = zeros(v, t);
U = zeros(v, 3);

a = 1;
g = -log(16);                                           % this g allows a sample equally between % two
                                                        centers to evaluate to .5 for both.
for j = 1:n
    for i=1:m                                           % observation type
        p = a:(a+k-1);                                  % parameter
                                                        % range of Z to be built
        k = W(i,3,j);
        c = linspace(W(i, 1, j),W(i, 2, j),k);          % number of centers
        d = c(2)-c(1);                                  % k equally spaced centers
                                                        % distance between centers
        U(p, 1) = repmat(i, k, 1);
        U(p, 2) = repmat(j, k, 1);                      % index of observation type
        U(p, 3) = c';                                   % index of parameter
                                                        % value of centers
        c = repmat(c', 1, t);
        s = Q(:,i,j)';                                  % center copies into matrix
        s = repmat(s, k,1);                             % sample points
                                                        % sample points copies into matrix
        R = (s-c).^2;
        Z(p,:) = exp(g/d^2 * R);                        % distance between sample and center
                                                        % exponential RBF
        a = a+k;
    end
end
```

Listing 3: fuzzyCluster

```
function [U C E] = fuzzyCluster(X, k, m, t, seed)
% function [U C E] = fuzzyCluster(X, k, m, t)
%  X, the data to be clustered
%  k, the number of clusters
%  m, a fuzziness coefficient
%  t, tolerance
%
% Returns U, the membership of each point in each cluster
% C, the cluster centers
% E, error as the Xie-Beni index

%% initialization
    [n d] = size(X);
    if nargin < 5
        rand('state',sum(100*clock));               % generate random memberships seed if
        U = rand(n, k);                             %   none specified
    else
        U = seed;
    end

%% Precalculations
    f = 1/(m-1);
    Xs = permute(repmat(X, [1 1 k]), [3 2 1]);      % X data replicated k times and rotated
    o = ones(n,d);
    U1 = inf;

%% Converge to Solve
    while any(abs(U1-U) > t)
        U1 = U;

    %% calculate center points
        B = (U.^m)';                                % quick way to calc centers
        C = (B*X) ./ (B*o);

    %% calculate membership in each cluster       % very quick way to calc memberships
        Cs = repmat(C, [1 1 n]);
        R  = sum((Xs-Cs).^2, 2).^f;                 % find distances (sqrt reduced into f),
        R  = permute(R, [1 3 2]);
        Rv = 1./R;                                  % copies of 1/ (x(i)-c(r))
        R  = R';                                    % copies of x(i)-c(j), rotated for mult.
        for i=1:n
            U(i,:) = sum(Rv(:,i)*R(i,:));           % sum over r collapsed into
        end                                         %    matrix multiplication
        U = 1./U;
    end                                             % find inverse

%% Calculate error statistics                     % the Xie-Beni index.  Distances
    F = sum((Xs-Cs).^2,2);                          %   calc'ed in the same way as above.
    F = permute(F, [3 1 2]);
    F = sum(sum((U.^m) .* F));
    v = dist(C);
    E = F/(n*min(v(v>0)))^2;
```

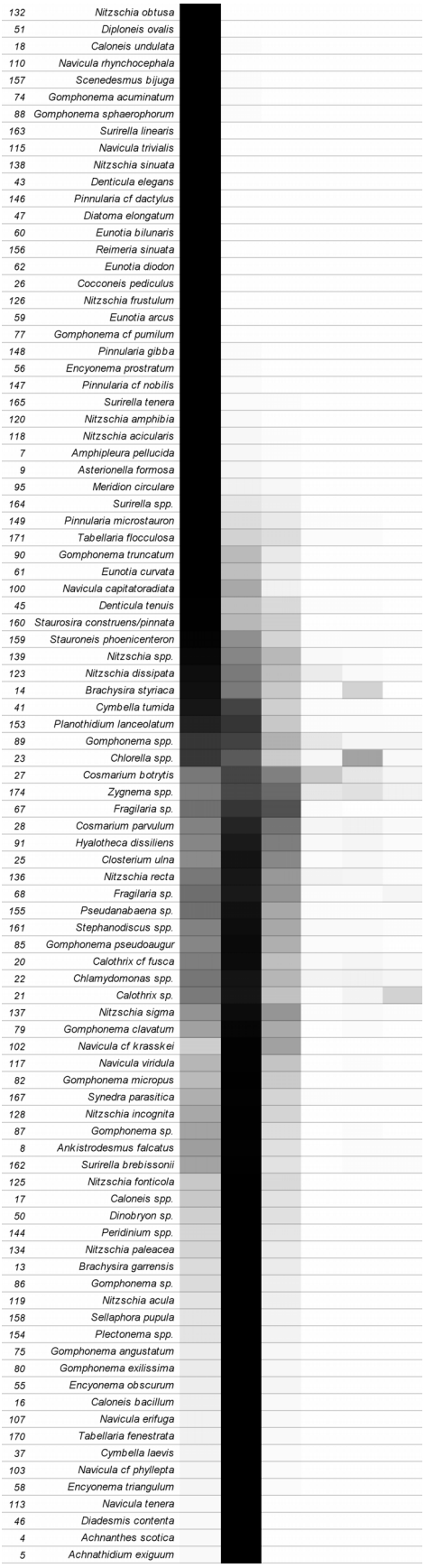
Listing 4: kNearN

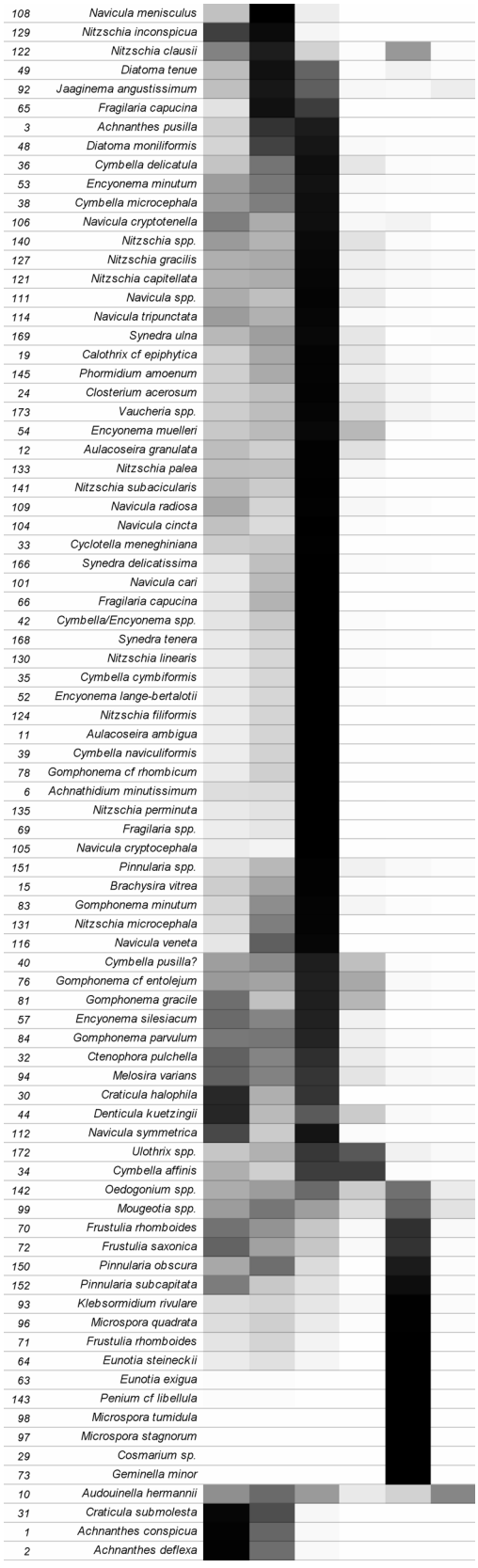
```
function C = kNearN(A, t, K)
% C = kNearN(A, t, K)
% A is the data to be clustered
% t is the smallest distance two different clusters may be apart
% K is the number of neighbors to check when aggregating clusters.

%% Initialize
    n = length(A);                                      % initial clusters, 1 in each
    C = 1:n;                                            % allow for quick zero-initialization
    Z = zeros(n);

%% Create Distance matrix, D
    D = Z;                                              % quick replication by adding 0
    copies = Z(1,:);
    for i=1:n                                           % difference squared matrix
        D(:,i) = sum((A-A(i+copies,:)).^2,2);
    end                                                 % find Euclidian dist
    D = sqrt(D);

%% Find the neighbors and add them to their group.     % vector of cluster numbers
    clusters = [];                                      % vector of points checked
    checked = Z(1,:);
    L = (D<t);
    for i=1:n                                           % check if not clustered
        if ~any(C(i) == clusters)                       % build clusters recursively
            [C checked] = FindNeighbors(i, L, C, checked); % add to cluster list
            clusters = [clusters C(i)];
        end
    end

%% Combine groups                                     % sort distance matrix
    [Y Q] = sort(D, 2, 'ascend');
    C1 = Z(1,:);
    while any(C1-C)
        C1 = C;
        for i=1:n                                       % Get cluster of K nearest to i
            X = C(Q(i,1:K));
            W = Z(:,1);                                 % for each neighbor
            for k=X                                     % increment its counter
                W(k) = W(k) + 1;
            end                                         % find the cluster of the most neighbors
            [m u] = max(W);
                                                        % if most neighbors in different cluster
            if C(i) ~= u                                % set all members of current cluster
                x = find(C == C(i));                    %   to k-nearest neighbor's cluster
                C(x) = u;
            end
        end
    end

%% Recursive Function FindNeighbors
function  [C checked] = FindNeighbors(i, L, C, checked)
    M = find(L(i,:));                                   % set all neighbors to my cluster
    C(M) = C(i);                                        % add me to checked
    checked(i) = 1;
    for j=M                                             % if neighbor not yet checked
        if ~checked(j)                                  %   check him
            [C checked] = FindNeighbors(j, L, C, checked);
        end
    end
```

Appendix 2. Maps

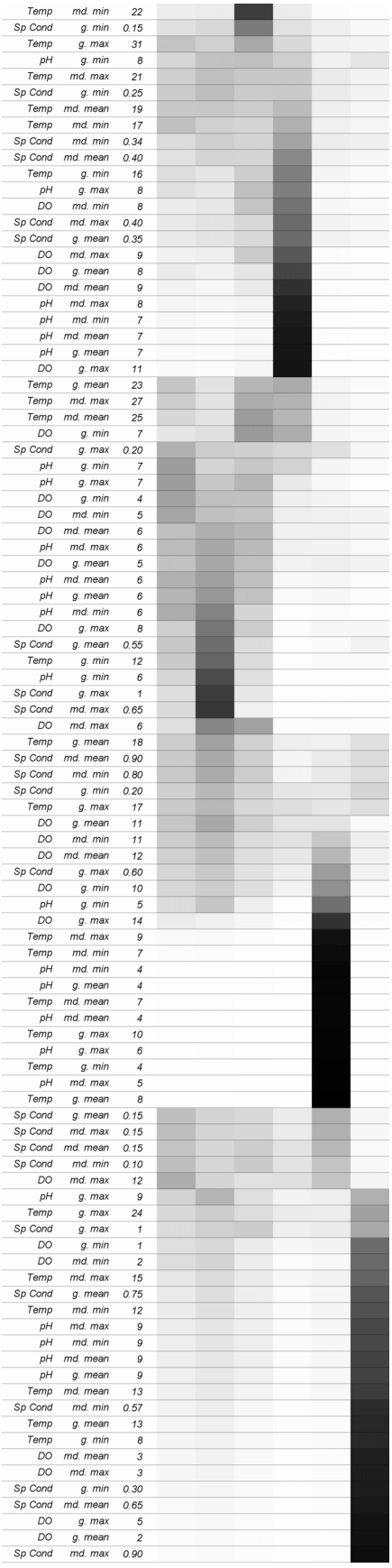
Appendix 3: Fuzzy C-Means Clustering Results

Error! Reference source not found.

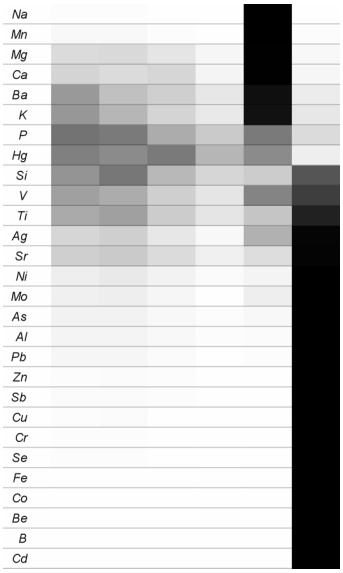




Error! Reference source not found.



Error! Reference source not found.



Appendix 4: k-Nearest Neighbors Clustering Results

Cyan (1)

Cosmarium sp.	Ba	Temp	g. min	4
Eunotia exigua	Ca		g. mean	8
Eunotia steineckii	K		g. max	10
Frustulia rhomboides	Mg		m. d. min	7
Frustulia rhomboides	Mn		m. d. mean	7
Frustulia saxonica	Na		m. d. max	9
		Sp Cond		
Geminella minor			g. mean	0.15
Klebsormidium rivulare			m. d. min	0.10
Microspora quadrata			m. d. mean	0.15
Microspora stagnorum			m. d. max	0.15
Microspora tumidula		DO	g. min	10
Mougeotia spp.			g. max	14
Penium cf libellula		pH	g. min	5
Pinnularia obscura			g. mean	4
Pinnularia subcapitata			g. max	6
			m. d. min	4
			m. d. mean	4
			m. d. max	5

Black (2)

Achnanthes scotica	Si	Temp	g. min	12
Achnathidium exiguum			g. mean	18
Ankistrodesmus falcatus			g. max	17
		Sp Cond		
Caloneis bacillum			g. min	0.20
Caloneis spp.			g. min	0.25
Calothrix cf fusca			g. mean	0.55
Calothrix sp.			g. max	0.60
Chlamydomonas spp.			g. max	1
Cosmarium botrytis			m. d. min	0.34
Cymbella laevis			m. d. min	0.80
Diadesmis contenta			m. d. mean	0.40
Diatoma tenue			m. d. mean	0.90
Encyonema triangulum			m. d. max	0.65
Fragilaria sp.		DO	g. mean	11
Gomphonema angustatum			g. max	8
Gomphonema exilissima			m. d. min	11
Gomphonema pseudoaugur			m. d. mean	12
Gomphonema sp.		pH	g. min	8
Gomphonema sp.			g. max	7
Jaaginema angustissimum			g. max	9
Navicula cf krasskei				
Navicula cf phyllepta				
Navicula erifuga				
Navicula menisculus				
Navicula tenera				

Nitzschia clausii
Nitzschia incognita
Nitzschia paleacea
Nitzschia recta
Planothidium lanceolatum
Plectonema spp.
Pseudanabaena sp.
Sellaphora pupula
Stephanodiscus spp.
Surirella brebissonii
Synedra parasitica

Red (3)

Achnanthes conspicua	Hg	Temp	g. min	16
Achnanthes deflexa	P		g. mean	23
Achnathidium minutissimum			g. max	31
Amphipleura pellucida			m. d. min	17
Asterionella formosa			m. d. mean	19
Aulacoseira granulata			m. d. max	21
Brachysira vitrea		Sp Cond	g. max	0.20
Caloneis undulata		DO	g. mean	8
Cocconeis pediculus			g. max	11
Craticula halophila			m. d. min	5
Craticula submolesta			m. d. mean	9
Ctenophora pulchella			m. d. max	9
Cyclotella meneghiniana			m. d. max	12
Cymbella affinis		pH	g. min	7
Cymbella microcephala			g. mean	7
Cymbella pusilla?			g. max	8
Cymbella tumida			m. d. min	7
Cymbella/Encyonema spp.			m. d. mean	7
Denticula elegans			m. d. max	8
Denticula kuetzingii				
Denticula tenuis				
Diatoma elongatum				
Diatoma moniliformis				
Diploneis ovalis				
Encyonema muelleri				
Encyonema prostratum				
Encyonema silesiacum				
Eunotia arcus				
Eunotia bilunaris				
Eunotia curvata				
Eunotia diodon				
Fragilaria sp.				
Fragilaria spp.				
Gomphonema acuminatum				
Gomphonema cf entolejum				
Gomphonema cf pumilum				
Gomphonema gracile				

Gomphonema minutum
Gomphonema parvulum
Gomphonema sphaerophorum
Gomphonema spp.
Gomphonema truncatum
Melosira varians
Meridion circulare
Navicula capitatoradiata
Navicula rhynchocephala
Navicula spp.
Navicula trivialis
Nitzschia acicularis
Nitzschia amphibia
Nitzschia dissipata
Nitzschia frustulum
Nitzschia inconspicua
Nitzschia obtusa
Nitzschia palea
Nitzschia sinuata
Nitzschia spp.
Nitzschia spp.
Pinnularia cf dactylus
Pinnularia cf nobilis
Pinnularia gibba
Pinnularia microstauron
Reimeria sinuata
Scenedesmus bijuga
Stauroneis phoenicenteron
Staurosira construens/pinnata
Surirella linearis
Surirella spp.
Surirella tenera
Synedra ulna
Tabellaria flocculosa

Green (4)

Achnanthes pusilla	Temp	m. d. min	22
Aulacoseira ambigua		m. d. mean	25
Calothrix cf epiphytica		m. d. max	27
	Sp		
Closterium acerosum	Cond	g. min	0.15
Cymbella cymbiformis		g. mean	0.35
Cymbella delicatula		m. d. max	0.40
Cymbella naviculiformis	DO	g. min	7
Encyonema lange-bertalotii		m. d. min	8

Encyonema minutum
Fragilaria capucina
Fragilaria capucina
Gomphonema cf rhombicum
Navicula cari
Navicula cincta
Navicula cryptocephala
Navicula cryptotenella

Navicula radiosa
Navicula symmetrica
Navicula tripunctata
Nitzschia capitellata
Nitzschia filiformis
Nitzschia gracilis
Nitzschia linearis
Nitzschia microcephala
Nitzschia perminuta
Nitzschia subacicularis
Oedogonium spp.
Phormidium amoenum
Pinnularia spp.
Synedra delicatissima
Synedra tenera
Ulothrix spp.
Vaucheria spp.

Blue (5)

Audouinella hermannii	DO	g. min	4
Brachysira garrensis		g. mean	5
Brachysira styriaca		m. d. mean	6
Chlorella spp.		m. d. max	6
Closterium ulna	pH	g. min	6
Cosmarium parvulum		g. mean	6
Dinobryon sp.		m. d. min	6
Encyonema obscurum		m. d. mean	6
Gomphonema clavatum		m. d. max	6
Gomphonema micropus			
Hyalotheca dissiliens			
Navicula veneta			
Navicula viridula			
Nitzschia acula			
Nitzschia fonticola			
Nitzschia sigma			
Peridinium spp.			
Tabellaria fenestrata			
Zygnema spp.			

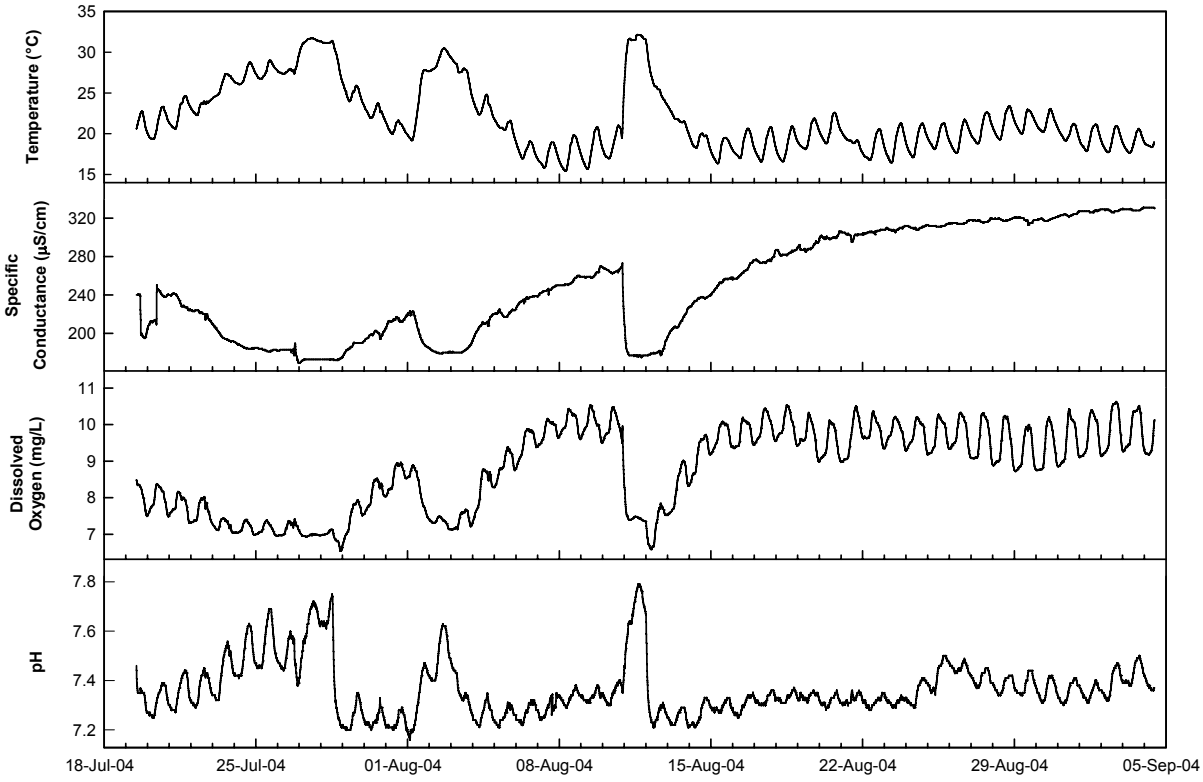
Magenta (6)

Ag	Temp	g. min	8
Al		g. mean	13
As		g. max	24
B		m. d. min	12
Be		m. d. mean	13
Cd		m. d. max	15
Co	Sp Cond	g. min	0.30
Cr		g. mean	0.75
Cu		g. max	1.40
Fe		m. d. min	0.57

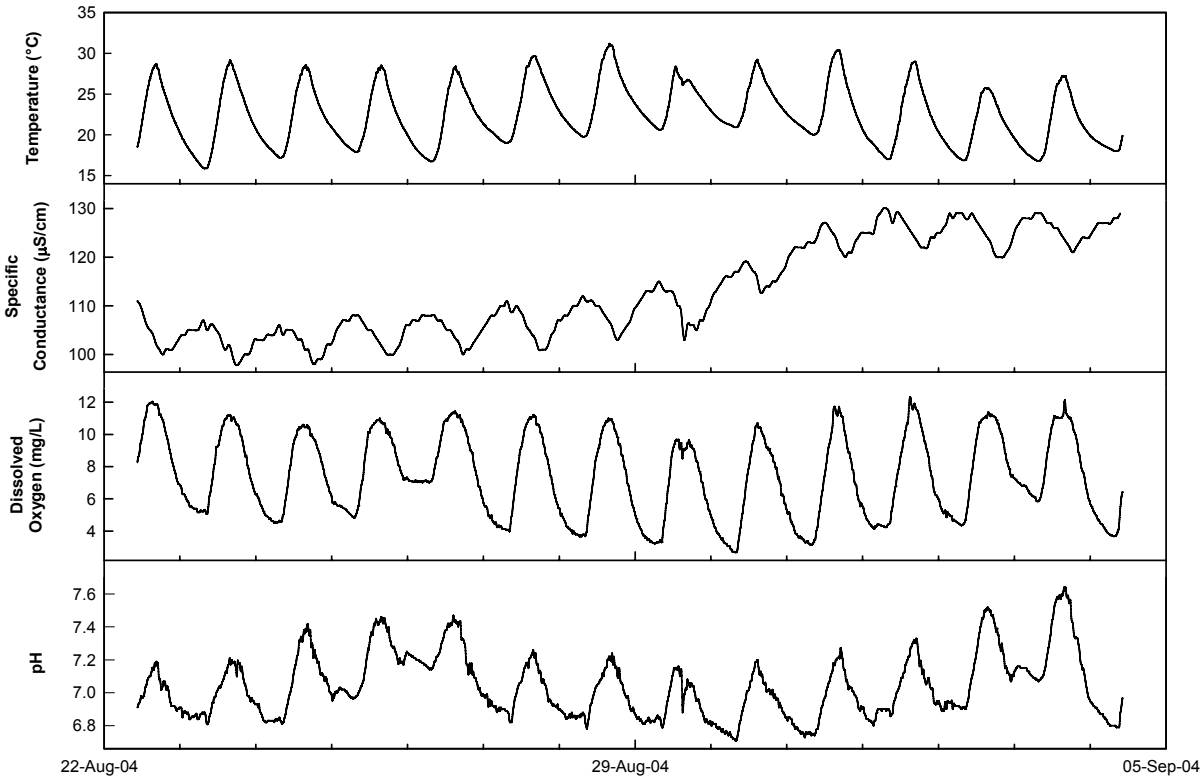
		m. d. mean	0.65
Mo		m. d. max	0.90
Ni	DO	g. min	1
Pb		g. mean	2
Sb		g. max	5
Se		m. d. min	2
Sr		m. d. mean	3
Ti		m. d. max	3
V	pH	g. mean	9
Zn		m. d. min	9
		m. d. mean	9
		m. d. max	9

Appendix 5: Water Chemistry by Site

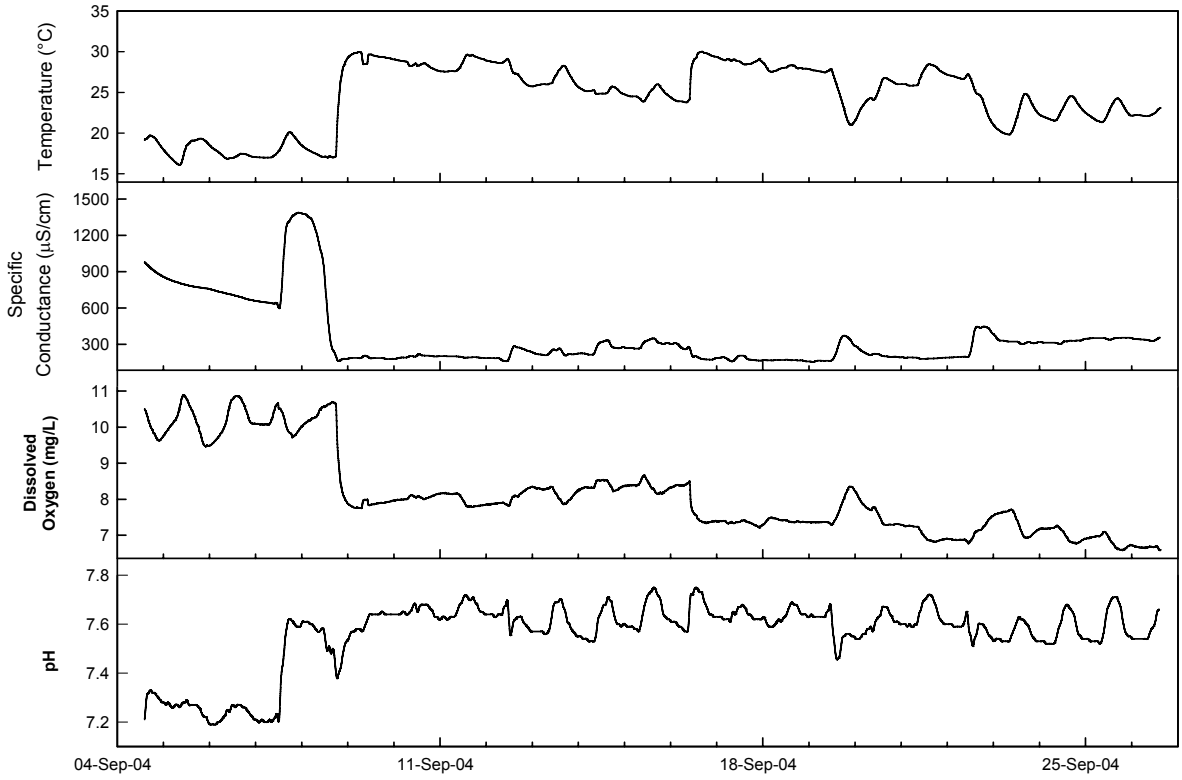
A: 0A Summer 2004



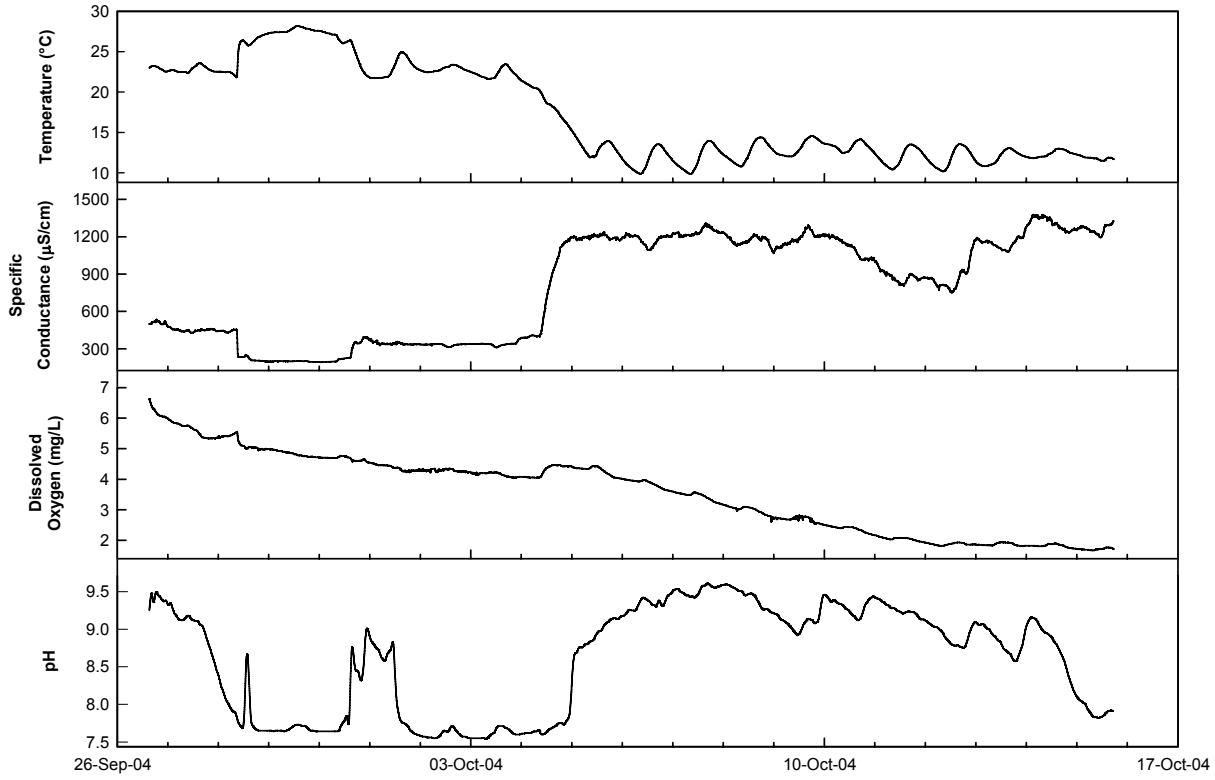
B. 0B Summer 2004



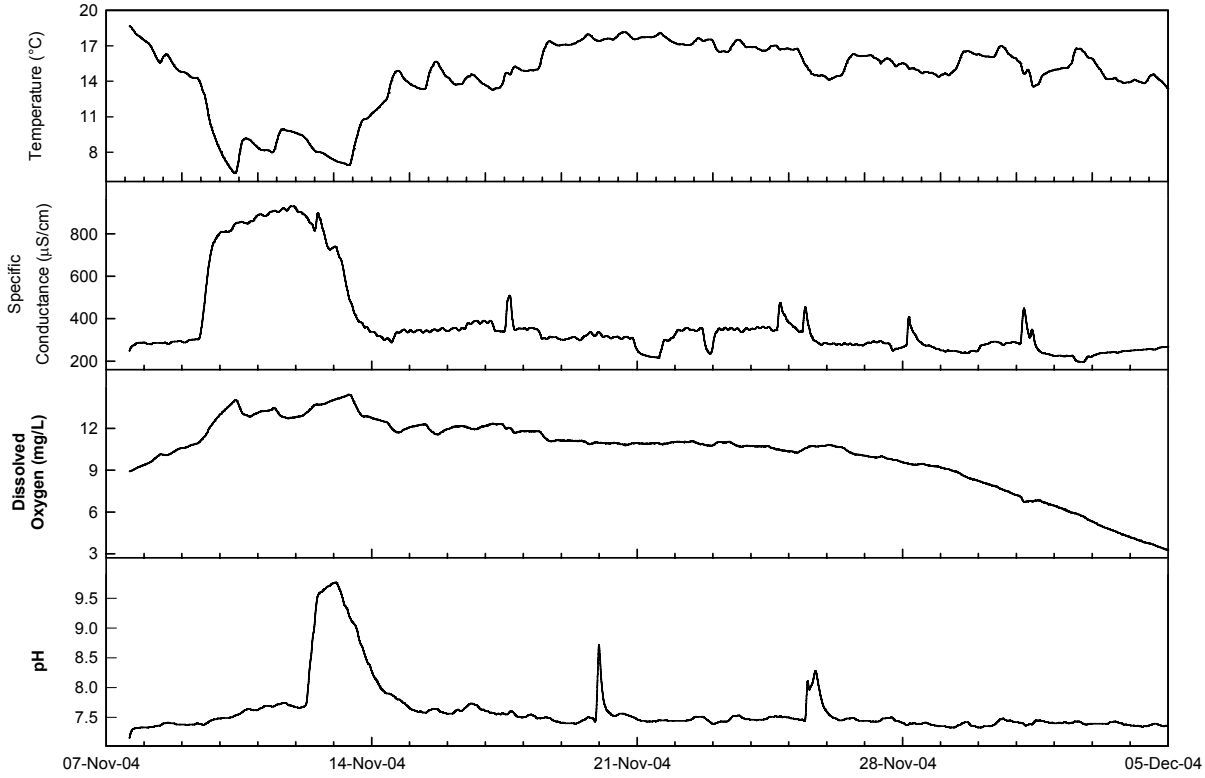
C. 4M1 Summer 2004



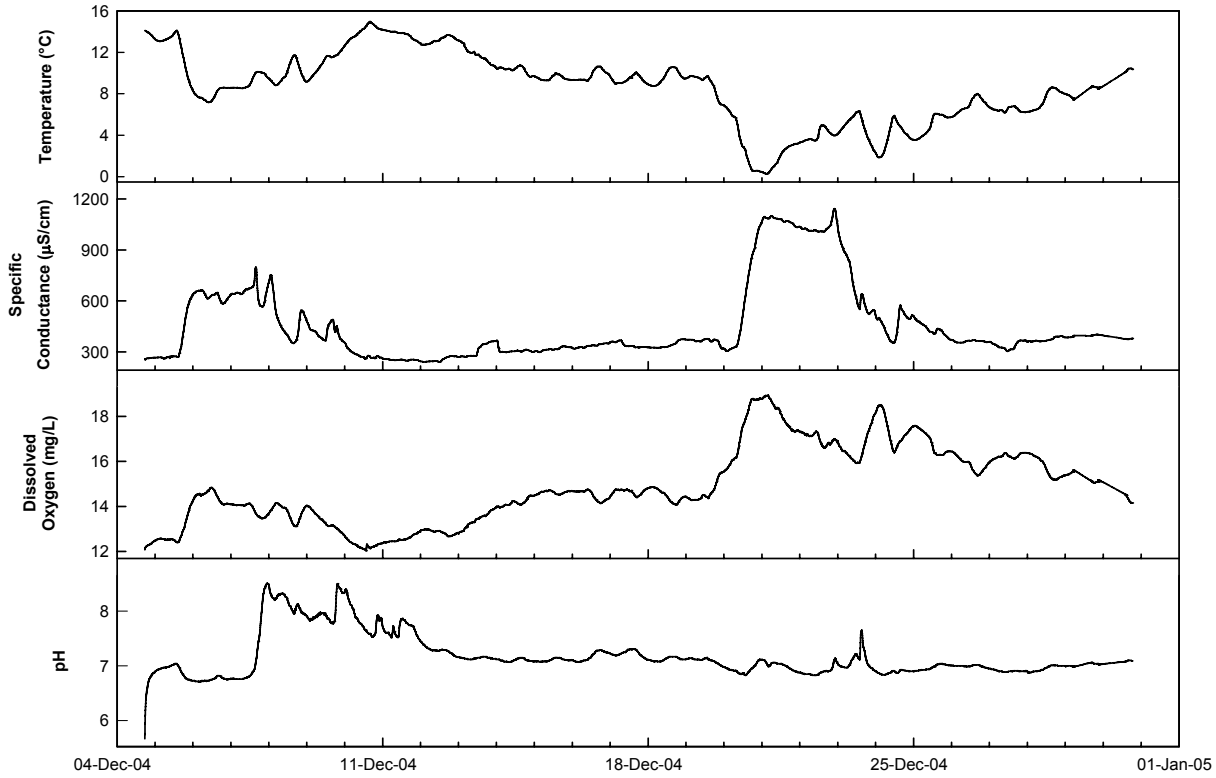
D. 4M2 Fall 2004



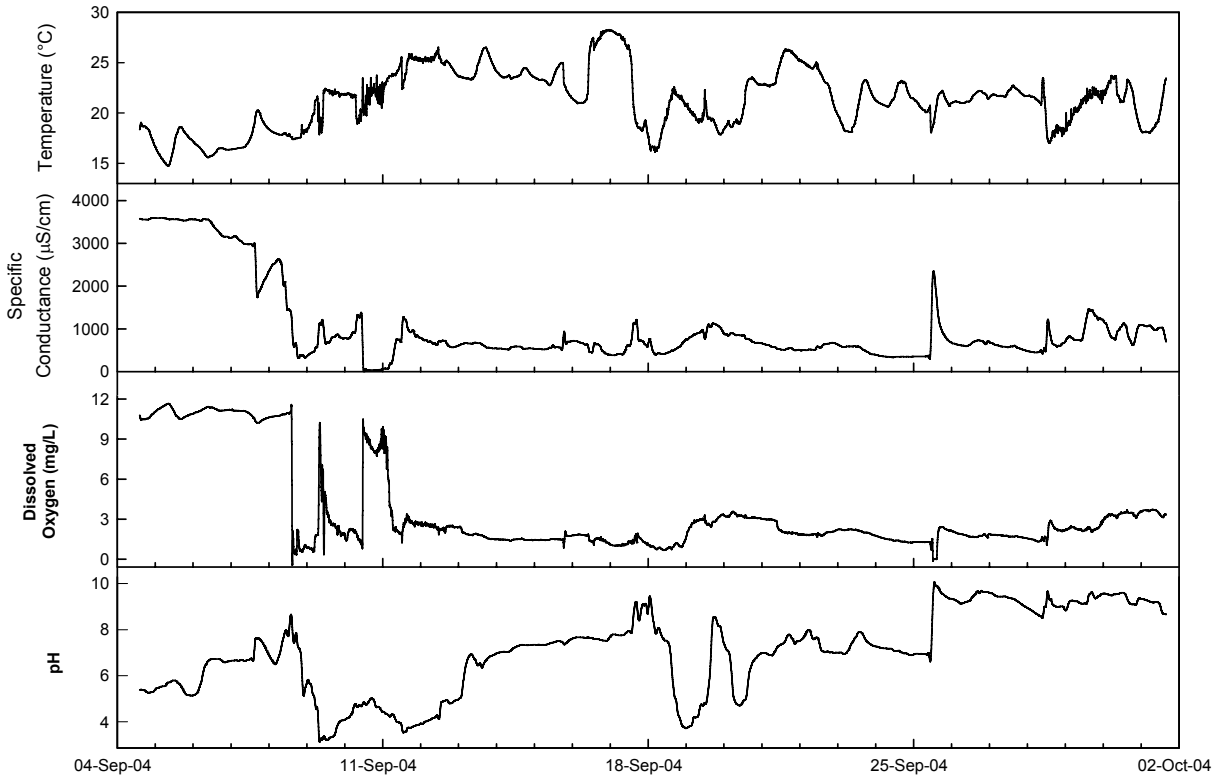
G. LR1 Winter 2004



H. LR2 Winter 2004



E. 4MR Fall 2004



F. LRR Winter 2004