HOW CHEMICAL DIFFERENCES IN DISSOLVED ORGANIC MATTER RELATE TO VEGETATION

By

Sarah Jean Seelen

RECOMMENDED: ______________________________

____

____
Advisory Committee Chair

____
Chair, Department of Civil and Environmental Engineering

APPROVED: ______________________________
Dean, College of Engineering and Mines

____
Dean of the Graduate School

____
Date

HOW CHEMICAL DIFFERENCES IN DISSOLVED ORGANIC MATTER RELATE TO VEGETATION

A

THESIS

Presented to the Faculty
of the University of Alaska Fairbanks

in Partial Fulfillment of the Requirements
for the Degree of

MASTER OF SCIENCE

By

Sarah Jean Seelen, B.A.

Fairbanks, Alaska

December 2004

ABSTRACT:

The purpose of this study was to better understand the link between dissolved organic matter (DOM) in soil leachate and different vegetation attributes. Soil cores were collected from the Caribou Poker Creeks Research Watershed (CPCRW) and subjected to a laboratory leaching procedure. The leachates were then subjected to a number of analytical tests, including pyrolysis-gas chromatography/mass spectrometry (py-GC/MS). Py-GC/MS is a “molecular fingerprinting” technique that was used to help determine similarities and differences in organic matter leached from soils with different vegetation attributes.

Numerous statistical tests were performed including Student-t, analysis of variance, principal components analysis, and partial least squares regression (PLS). Results from Student-t tests indicated that local vegetation plays an important role in the character of the DOM in soil leachate. Additionally, a principal components test revealed relationships between soil leachates and vegetation attributes. A prediction model was created using PLS to predict components of leachate DOM based on vegetation attributes. This model, while in its early development, was able to predict 70% of the total molecular fingerprint of leachate DOM based on cover vegetation.

Table of Contents

List of Figures

List of Tables

List of Appendices

Acknowledgements

Environmental Protection Agency - Alaska Training and Technical Assistance Center (ATTAC)

National Science Foundation - Experimental Program to Stimulate Competitive Research (EPSCoR)

United States Geological Survey (USGS)

CHAPTER 1: INTRODUCTION

1.1 IMPORTANCE TO BETTER UNDERSTANDING ORGANIC MATTER

Groundwater aquifers are replenished through surface recharge. Certain recharge areas in a watershed may have soils with higher hydraulic conductivity than others or may, for other reasons, be particularly vulnerable to the influx of non-point source pollutants. Dissolved organic matter (DOM) is believed to serve as a vector for heavy metal transport in groundwater (Koopal et al., 2001; Moore, T.R., 1997). The properties of DOM are a function of the original parent litter (e.g., leaves, roots) as well as the physical, chemical and biological transformations that occur in the soil (Lehtonen et al., 2001; Joly et al., 2001; Schulten and Gleixner 1999; Page et al., 2002; Suominen et al., 2003). Whether or not DOM can serve to mobilize non-point source pollutants in a recharge area depends on the properties of the DOM, the pollutant, and the soil/water chemistry.

1.2 OBJECTIVES, HYPOTHESIS AND DESCRIPTION OF STUDY

The overall objective of this study was to better understand the link between cover vegetation and dissolved organic matter (DOM) in soil leachates. The hypothesis for this research project is that the chemical nature of DOM can be statistically related to local vegetation. Armed with these two pieces of information, as well as information on soil properties, one could identify certain recharge areas in a watershed that may be particularly vulnerable to subsurface mobilization of non-point source pollutants. Appendix A includes a literature review of the role DOM plays in the mobilization of cadmium (Cd^{2+}), a heavy metal cation.

Soil cores collected from a boreal forest in interior Alaska were subjected to a laboratory leaching procedure. DOM was collected and analyzed using pyrolysis-gas chromatography/mass spectrometry (py-GC/MS). Py-GC/MS is a “molecular fingerprinting” technique that was used to identify similarities and differences in DOM leached from areas with different cover vegetation.

With the aid of statistical software, this project considered the many relationships between the molecular fingerprint of DOM from soil leachates and cover vegetation.

It was determined that the chemical nature of DOM in soil leachates can be statistically related to local vegetation.

CHAPTER 2: BACKGROUND AND LITERATURE REVIEW OF ORGANIC MATTER

Considerable attention has been given in recent years to organic matter contained in the soil solution, both from the standpoint of plant nutrition and contaminant control. For instance, organic matter can help alleviate metal ion toxicities through chelation or serve as a carrier of xenobiotics (Lehtonen et al., 2001; Stevenson, F.J. 1994).

Soil organic matter (SOM) refers to the sum-total of all carbon-containing substances in soil which includes, but is not limited to: litter, dead organic debris, water-soluble organic matter (humus), microbial biomass, stabilized organic matter, and anthropogenic inputs (Stevenson, F.J. 1994, Lichtfouse et al., 1998; Schulten, 1999). It influences plant growth through its effects on the physical, chemical, and biological properties of soils (Schulten, 1999). SOM exists as polyfunctional macromolecules (Lehtonen et al., 2001; Davies et al., 2001; Joly et al., 2000; Schulten and Gleixner, 1999; Schulten and Schnitzer, 1992). Some properties of organic matter in soils are of direct concern when considering complexation potentials with heavy metals. They include: water retention which can be up to 20 times its weight in water; combination with clay minerals which stabilizes structure and increases permeability;

chelation, forming stable complexes with copper (Cu^{2+}), manganese (Mn^{2+}), zinc (Zn^{2+}), and other polyvalent cations; enhancement of availability of micronutrients to higher plants; cation exchange capacity (CEC), 20 – 70 % of CEC of many soils is caused by organic matter (Stevenson, F.J. 1994).

The amount and type of organic matter is governed by the soil-forming factors of time, parent material, topography, vegetation, and climate (Kracht and Gleixner, 2000). Local environmental conditions such as chemical composition of the parent material, climate, acidity, biological activity, and the genetically different forest humus types, develop and are characterized by different chemical properties (Hempfling and Schulten, 1990).

Water collects organic matter as it passes through the canopy, soil litter, and subsurface soil. As the groundwater ages, the organic matter is subjected to greater degrees of transformation (White et al., 2002). Dissolved organic matter (DOM) is modified from SOM by chemical, physical, and biological processes (White et al., 2002). The water-soluble part obtained by filtration (<0.45 μm) is defined as DOM (Schulten 1999). DOM is released from vegetation and soil organic matter and influences almost every process in the soil and aquatic environments (Moore, T.R., 1997). It may reflect the major vegetation present (Hongve et al., 2000). However, climatic and microbial processes are important factors as these can also influence the character of DOM in soils (Page et al., 2002). According to Hongve et al. (2000), leaching from litter on the soil surface is often the primary source of DOM. However, even minor vegetation components can make a significant contribution to DOM when released as a highly soluble, stable, or recalcitrant macromolecule subjected to degradation during transport through the catchment (Page et al, 2002).

Many analytical techniques can be applied toward the analysis of organic matter. Some techniques used in the analysis of organic matter include nuclear magnetic resonance (C^{13}-NMR) spectroscopy; hydrogen, nitrogen, and carbon isotopes; and Fourier transform-infra red (FT-IR) spectrum. However, applied methods of analytical pyrolysis are well suited for investigations of structural and molecular properties of humic substances (Schulten 1999; Hempling and Schulten, 1990; White, et al. 2002). Py-GC/MS is a powerful tool for fingerprinting and characterizing organic matter, in particular, DOM. The principal results from pyrolysis are the characterization, differentiation, quantification, and structural identification of the pyrolysis products (Schulten and Gleixner, 1999). Parent plant materials such as wood, roots, and leaves have also been studied by py-GC/MS (Hempfling and Schulten, 1990).

Some advantages of py-GC/MS are sensitivity, specificity, and speed (Schulten 1999). Additionally, it is highly reproducible when compared with other methods (Davies et al., 2001). Despite the advantages, py-GC/MS is not without limitations. Different experimental conditions may reveal compounds important that may otherwise be true (Schulten and Schnitzer, 1992). Py-GC/MS produces hundreds of degradation by-products, which may undergo secondary reactions thus causing variation in the yields of the by-products (Joly et al., 2000; Page et al., 2002). The largest limitation within the analytical pyrolysis method exists within the mass spectrometry aspect. Because analytical pyrolysis produces thermal degradation of by-products of DOM, the mass spectra of these products may be too complicated to correctly identify DOM compounds with ease. Additionally, the mass spectra typically contain branched alkanes and alkenes as well as alkylated benzenes, which are generally unspecific, and no definite identification can be made (Davies et al. 2001).

Due to the ubiquitous nature of DOM, it is imperative to better understand its relevance. Two primary concerns of the role of DOM are in water treatment systems and heavy metal transport. DOM in drinking water is aesthetically unacceptable and can cause the formation of disinfection by-products in the presence of chlorine, typically used in disinfection (Schulten, 1993; Lehtonen et al., 2001; Page et al., 2002). The presence of DOM in waters to be treated with membrane filtration may also cause a considerable amount of fouling, thereby reducing membrane life (Naar et al., 2001). The interactions between heavy metals (e.g., Cd^{2+}) are discussed further in Appendix A. The interactions between DOM and heavy metals are important and to better understand them, the gap between plant precursors and DOM must be closed (Cronan and Aiken, 1985; Schulten and Gleixner, 1999). The combination of py-GC/MS and statistical analysis can better close the gap between vegetation and potential for membrane fouling and heavy metal transport.

CHAPTER 3: METHODOLOGY

3.1 SITE DESCRIPTION

Samples were collected from the Caribou Poker Creeks Research Watershed (CPCRW) and the Bonanza Creek Research Watershed. The CPCRW is approximately 100 km^2 and is reserved for research studies. However, gold mining and logging were conducted in CPCRW until the early 1900s.

3.2 SAMPLE COLLECTION

Samples were collected in August in 2001, 2002, and 2003. Most soil samples were collected in a 3-inch inside diameter steel core tube. The cone end of the core tube was fitted with cutting teeth and the other connected to a 2 hp power head. The depth of the core was limited by the depth of bedrock. Some samples could only be collected as grab samples. For the grab samples, a shovel was used to collect a 1 cubic foot sample.

The specific collection sites were selected based on vegetation. Sites were selected in an attempt to include all dominant types of vegetation found in a boreal forest. This included: birch, aspen, white and black spruce trees; shrub birch and arctic rose shrubs; and feather moss, sphagnum, lichen, and grass ground cover. After collection, samples were placed in a cooler at 4°C until they were subjected to a laboratory leaching procedure.

3.3 SOIL SAMPLE LEACHING PROCESS

Soil leachate was collected from each core using a leaching procedure modified after Wagai and Sollins (2002). The leaching process began by wet sieving the entire soil sample in 7.5 L of water stored in a collection vessel. Three sieves were used during this step, sieve numbers 16, 60, and 200 (screen openings: 1.18 mm, 0.25 mm, and 0.075 mm). The sieves were continuously moved up and down in the water to keep the sediment moving and to fully immerse the sample in the water. Sieving was conducted to disrupt the aggregates and to ensure all particles were in contact with the water. After a sufficient amount of time passed so that the aggregates were broken down, the sieves were removed along with the larger stones and plant debris (such as leaves, roots, and branches) so that only the leachate remained in the collection vessel.

Three liters of leachate were removed from the soil-water solution and subjected to centrifugation. This step was necessary to eliminate particles smaller than sieve #200 and larger than 0.45 µm, the particle size cutoff for DOM. The leachate was centrifuged for 30 minutes at 3900 rpm. The centrifuged leachate was then vacuum filtered over a 0.45µm glass micro-fiber filter to obtain the dissolved leachate sample. A total of 2 L of dissolved leachate was rotovapped under vacuum at 40°C to prevent loss of organic matter to volatilization until approximately 20 mL remained. The evaporating flask was rinsed out twice to obtain maximum amount of dissolved organic matter. This liquid was then placed in a drying oven at 60°C to fully evaporate the water. The dried DOM was collected and placed in a quartz tube for py-GC/MS.

3.4 ANALYTICAL PYROLYSIS

Py-GC/MS was conducted with a CDS Model 2500 pyrolyzer and autosampler in tandem with a GC/MS. During pyrolysis, the sample was heated from a starting temperature of 25 °C to 700 °C in 0.1 seconds and held at a constant 700 °C for 9.9 seconds. The pyrolysis reactor was mounted to an HP 5890 Series II GC, with a Supelco SPB 35 (35% Ph Me silicon) column, 60 m x 0.25 m x 0.25 μm. The GC interface temperature was set at 235 °C. The GC temperature program was 40 °C held for 30 minutes, 1 °C/min increase for 80 min, 20 °C/min increase for 50 minutes, 10 °C/min increase for 10 minutes and then held for 10 minutes. The GC was plumbed directly to an HP 5971A Series Mass Selective Detector on electron impact (EI) mode. The MS scanned mass units 45 to 650. All mass spectra were compared to the NBS54K spectral library. Helium served as a carrier gas at a flow rate of 0.5 cm^3/minute. Each sample was injected with a split ratio of 1:50.

3.5 VEGETATION ATTRIBUTES

The dominant vegetation attributes were assigned to each core location by field inspection. The values for each vegetation attribute was given a (+)1 for presence and a (-)1 for the absence of the vegetation attribute. A total of nine vegetation attributes were considered: trees (birch, aspen, spruce, n=3), shrubs (arctic rose, shrub birch, n=2) and ground cover (feather moss, sphagnum moss, lichen, grass, n=4). Each site was labeled with only one dominant tree, shrub, or ground cover attribute and if there was not a dominant species, it was given a (-)1. Table 1 contains each soil leachate and its assigned vegetation attributes. Two vegetation attributes were present only once in the leachates, arctic rose and grass. All other vegetation attributes were present in multiple soil leachates.

Table 1: Vegetation Attributes for all Soil Leachates.

	Tree			Shrub		Ground Cover			
Soil Leachate ID	birch	aspen	spruce	shrub birch	arctic rose	feather moss	sphagnum moss	lichen	grass
AP	-	-	-	-	+	+	-	-	-
Haystack	-	-	-	+	-	+	-	-	-
HR	-	-	-	+	-	-	+	-	-
LG	-	-	-	-	-	-	-	+	-
P2	-	-	+	+	-	-	+	-	-
P3	-	-	+	+	-	-	+	-	-
SHAY1	+	-	-	-	-	+	-	-	-
SHAY2	+	-	-	-	-	-	+	-	-
UBS	-	-	+	-	-	+	-	-	-
HWLB	+	-	-	-	-	-	-	-	-
CSCP	-	-	-	+	-	+	-	-	-
ASPEN	-	+	-	-	-	+	-	-	-
BIRCH	+	-	-	-	-	-	-	-	+
SPRUCE	-	-	+	-	-	-	+	-	-
SPRUCE2	-	-	+	-	-	-	+	-	-
ASPEN2	-	+	-	-	-	+	-	-	-
n	4	2	5	5	1	7	6	2	1

A (+) indicates the presence of vegetation attribute and a (-) indicates an absence of vegetation attribute.

3.6 MOLECULAR FINGERPRINT COMPOUNDS

Py-GC/MS of DOM produces a complex chromatogram, or “pyrogram,” containing hundreds of peaks that represent individual pyrolysis products. The pyrolysis products may have been organic molecules originally present in the soil or they may be thermal breakdown products of large macromolecules. It has been found in previous research that only a small portion of the pyrolysis products are needed to draw differences between individual and groups of soil and water samples (Joly et al, 2000; White and Beyer, 1999; White et al, 2004). A total of 16 pyrolysis products were selected from each pyrogram to identify similarities or differences in DOM from the 16 soil leachate samples. HP Chemstation software was used to quantify the relative abundance of each of the 16 pyrolysis products. The pyrolysis products were compared on a relative abundance basis and were not individually quantified on a per soil mass basis. Table 2 contains a list of the 16 pyrolysis products. The relative abundance of the 16 compounds in a given sample was referred to as the samples molecular fingerprint. The likely origin of the pyrolysis products is also listed in Table 2.

Table 2: Molecular Fingerprint Compounds and Likely Origin.

Compound	Major Ions	Likely Origin
dimethyl benzene	91	Aromatic Hydrocarbon Precursors
furfural	96	Primary Polysaccharides
methyl cyclopentenone	67	Unknown
trimethyl benzene	105	Aromatic Hydrocarbon Precursors
benzaldehyde	106	Unknown
benzofuran	118	Unknown
benzonitrile	103	Unknown
phenyl ethanone	105	Unknown
methyl benzonitrile	117	Unknown
phenol	94	Protein, Lignin
methylated indene	129, 144	Aromatic Hydrocarbon Precursors
tetradecene	57, 71	Unknown
methoxy phenol	124	Lignin
dimethoxyphenol	154	Lignin
napthalene	128	Unknown
2-methyl napthalene	142	Unknown

CHAPTER 4: STATISTICAL ANALYSIS

4.1 GENERAL OVERVIEW OF STATISTICAL ANALYSES PERFORMED

A combination of univariate and multivariate statistics was used to properly identify and describe vegetation influences on the DOM obtained through leaching soil samples. Three statistical programs were used in the analysis portion of the research project, Microsoft® Excel, Minitab™ and The Unscrambler™. Microsoft® Excel and Minitab™ were used for the univariate portion and the multivariate portion was computed with The Unscrambler™.

When evaluating environmental samples, there is no universally accepted approach to determine statistically significant differences or similarities. For this reason, the statistical analyses of environmental data have become the trial and error of collecting, analyzing, and interpreting data with each step leading to the next step. Additionally, statistical tests do not necessarily consist of a series of rules and standards with which to evaluate data. They are merely recommendations by which one can analyze and evaluate statistical differences or similarities. This becomes especially true in environmental samples when the number of data points may be small compared to the recommended number of data points needed for a specific statistical test. It is also perfectly acceptable for the findings from one test to indicate that another test needs to be performed as is commonly the case when evaluating data with ANOVA, for example.

Multivariate statistics offers the researcher the ability to compare multiple independent variables against multiple dependent variables. In this research project, one of the objectives was to identify relationships between vegetation attributes and pyrolysis products or molecular fingerprint compounds.

4.2 DESCRIPTIVE STATISTICS

Each molecular fingerprint compound was evaluated to determine its basic statistical information. The descriptive statistics for each molecular fingerprint were calculated using the relative percentage of all soil leachates, regardless of its vegetation attribute. This basic statistical information was used to better understand the data set and determine whether had a normal distribution. Table 3 includes the following basic statistical information for each molecular fingerprint compound: mean, standard error, median, standard deviation, sample variance, kurtosis, skewness, range, and the confidence interval at 95%. Additionally, histograms and probability plots were calculated for each molecular fingerprint compound.

The arithmetic average of the values for each of the molecular fingerprint compounds is the mean. This value is equal to the sum of the values divided by the number of observations. The variance of a set of observations is the average of the squares of the deviations of the observation of their mean. The standard deviation is used rather than the variance in describing a data set for two reasons. The first of which, the variance is in a different unit of measurement than the observations and makes it difficult to apply a qualitative value to the numerical value. The standard deviation is the square root of the variance. This value is a natural measure of spread for observations about its mean. The standard error for each molecular fingerprint compound is the estimate of the error (or uncertainty) involved in estimating the mean of the compound. This does not refer to the variability involved in measuring the molecular fingerprint compound. The standard error is often used when the uncertainty of the estimate of the mean is of concern.

Table 3: Descriptive Statistics for all Molecular Fingerprint Compounds.

Molecular Fingerprint Compound	Mean	Standard Error	Median	Standard Deviation	Sample Variance	Kurtosis	Skewness	Range	Confidence Level (95.0%)
dimethyl benzene	17.89	1.12	16.38	4.19	17.58	0.40	0.69	15.66	2.42
furfural	13.32	1.46	13.34	5.47	29.96	-0.49	0.49	18.70	3.16
methyl cyclopentenone	7.40	0.69	8.27	2.59	6.70	-1.30	-0.16	7.88	1.49
trimethyl benzene	4.88	0.58	4.47	2.17	4.72	7.39	2.50	8.56	1.25
benzaldehyde	1.70	0.18	1.78	0.69	0.48	-1.55	0.08	2.05	0.40
benzofuran	1.29	0.19	1.18	0.69	0.48	7.84	2.53	2.85	0.40
benzonitrile	3.07	0.35	2.84	1.32	1.75	2.35	1.45	4.84	0.76
phenyl ethanone	2.64	0.30	2.60	1.11	1.24	-1.44	0.22	3.36	0.64
methyl benzonitrile	0.72	0.07	0.73	0.26	0.07	-0.28	-0.11	0.93	0.15
phenol	36.50	2.06	34.97	7.71	59.39	-0.12	0.94	24.20	4.45
methylated indene	1.24	0.20	1.16	0.74	0.54	0.28	0.87	2.42	0.43
tetradecene	1.38	0.23	1.14	0.85	0.73	4.64	1.80	3.39	0.49
methoxy phenol	8.01	1.13	6.56	4.23	17.87	0.05	0.85	14.49	2.44
dimethoxy phenol	1.84	0.61	0.88	2.27	5.14	1.27	1.65	6.52	1.31

Two descriptive statistics used in part to determine the degree of normality of the data set are kurtosis and skewness. Kurtosis is measure of the data distribution with respect to its flatness or peakedness as compared to a normal distribution. A negative value characterizes a flat distribution and a positive value characterizes a peaked distribution. An acceptable kurtosis values is around 3.0.

Skewness is another statistic used to determine how close the specific data set is to a normally distributed data set. Ideally, more than 50 data points would be used in the calculation of skewness. Typical values for a moderately skewed data set are between 0.5 to 1. While for a normalized skewness, the range is between 0 to 0.5. Skewness is very sensitive to the number of data points.

The range is the minimum value within the distribution for the specific molecular fingerprint compound subtracted from the maximum value. The final descriptive statistic calculated for the molecular fingerprint compounds is the confidence interval at 95%. This indicates the amount of error added and subtracted to the compound mean for a confidence interval of 95%. The value added and subtracted to the molecular fingerprint compound is different for each compound.

While some molecular fingerprint compounds had values above the acceptable limit for skewness and kurtosis, this indicated that differences existed among the sample sites. Histograms offer a visual representation of the range of the data for each molecular fingerprint compound for each sample site. Figures B-1 – B-14 in Appendix B contain histograms for each of the molecular fingerprint compounds remaining after the test for outliers within the molecular fingerprint compounds. Figures C-15 –C-28 in Appendix C contain the probability plots for each of the molecular fingerprint compounds. A probability plot is another visual representation of whether a data set is considered to be normally distributed. The closer the data points are to the trend line with slope equal to 1, the more likely the data set is normally distributed. Another test used to determine normality in small data sets is the W-test. This test was developed by Shapiro and Wilk (1965) and is an effective method for testing whether a data set has been drawn from an underlying normal distribution. The test involves calculating the W-statistic and comparing the value with tabulated $W_{critical}$ values for varying statistical significance (Gilbert, 1987). The data set is said to be normally distributed if the $W_{critical}$ is greater than the W-statistic calculated. This study did not use the W-test.

4.3 Q-TEST

The Q-test is a test for outliers. The $Q_{calculated}$ is equal to the gap divided by the range within the data set (Harris, 1999; Dean and Dixon, 1951; Rorabacher, 1991). The gap is the difference between the largest and second largest value (or the smallest and second smallest value). The range is the total spread of the data. The questionable point should be removed if $Q_{calculated}$ is greater than $Q_{critical}$ for the number of observations. Of course data points considered to be outliers must be removed with caution. The analytical process of py-GS/MS involves a very small amount of DOM (mg) to be placed into a very small diameter quartz tube (mm). While great care was taken to ensure a properly homogenous DOM sample, heterogeneity exists in all natural samples. Due to the nature of the analytical analysis and the possibility of heterogeneity within DOM samples, outliers were confidently removed through the Q-test statistical test. The outlier test results will be discussed further in Chapter 5.

4.4 ANALYSIS OF VARIANCE

A one-way analysis of variance (ANOVA) test is a method used to compare multiple samples with a single variable. This method was useful for this study in that it allowed the comparison, for example, between all trees for one molecular fingerprint compound. Microsoft® Excel was used to compute the statistical test. The one-way ANOVA tested the null hypothesis, where all sample means are equal, against the alternative hypothesis, where all sample means are not equal.

The one-way ANOVA tests introduced a number of statistically significant relationships within the vegetation attribute group comparisons. This would typically be followed up with a post-hoc test to determine where the actual differences exist. However, due to the unequal number of samples in each vegetation attribute group the most common post-hoc test, the Tukey test, could not be used and the Scheffe test, considered to be the next best solution for a parametric data set with unequal samples in each group, was considered to be overly conservative resulting in too many false positives. It was determined that a Student-t test would give, within a reasonable amount of certainty, a statistically significant value for differences between two vegetation attribute groups with differing vegetation attributes.

4.5 STUDENT-t

The Student-t test compares the means of two sets of data and determines whether they are statistically different. Microsoft® Excel computed the Student-t test and for this project, the Student-t test applied was two-tailed assuming unequal variances. A strong statistically significant difference was accepted when the p-value was less than 0.050 ($p<0.050$). A moderate statistically significant difference was accepted when the p-value was less than 0.200 ($p<0.200$). The p-value of 0.050 represents, with 95 % confidence, that the means of the two data sets are significantly different.

4.6 PRINCIPAL COMPONENT ANALYSIS (PCA)

The statistical program used to compute principal component analysis (PCA) for this research project was The Unscrambler™. PCA was a statistical tool used to plot all vegetation attributes and all molecular fingerprint compounds within a 2-D plot. The data matrix for this research project was 16 x 25. This included 16 molecular fingerprint compounds, 9 vegetation attributes and 16 leachate samples. Table E-1 located in Appendix E contains the complete data matrix used for PCA.

The purpose of all multivariate data analyses is to reduce the data in order to discover some 'hidden meaning' which allows the researcher to identify patterns or explain phenomenon previously unidentified. PCA computes relationships for with x-independent variables.

PCA operates by computing and organizing the explained and error variance associated with each data point into principal components. Principal components, after computation, can be written in a multiple regression equation, such as: Y = 0.1 (Variable 1) – 0.3 (Variable 2) + 0.45 (Variable 3) and so on (Bryant and Yarnold, 1995). The first principal component typically contains the maximum amount of the total variance that is explained. If, for example, there were 10 variables and the first principal component explained 60 % of the variance, then this new variable would account for the same overall amount of the total variance as did the 6 original variables, assuming each variable would represent 10 % of the variance. This would essentially reduce the number of variables from 10 to 4. As the number of principal components increases the explained variance asymptotically approaches 100 %. It is important to determine a reasonable number of principal components to compute. Researchers typically specify the number of principal components to be analyzed equal to a set amount of total variance has been explained. This research project computed 6 principal components.

The Unscrambler™ also produces three main plots meant to assist in deciphering the relationships among vegetation attributes and molecular fingerprint compounds. The first, the scree plot or residual plot, indicates the cumulative variance associated with each principal component. A scree plot is meant to be used as a preliminary step to determine the number of principal components needed for the data set. A loading plot places the vegetation influences and the molecular fingerprint compounds on an x-y scatterplot. The loading plot gives (+) and (-) significances to the vegetation attributes and molecular fingerprint compounds.. The final plot, the score plot, represents where the soil leachate core sites plot with respect to the loading plot.

4.7 PARTIAL LEAST SQUARES

The Unscrambler™ statistical program has the power to calculate both the x-variables (independent) and y-variables (dependent). This research project used the computational analysis of PLS1. This acronym stands for partial least squares regression 1-y-variable. For example, PLS1 analyzes all sample site soil leachates with vegetation attributes for one molecular fingerprint compound. This analysis produces a number of graphs. Two of which are similar to PCA, loading and score plots, and a coefficient plot or predicted vs. measured plots. The loading plot indicates where the vegetation attributes plot in relation to a specific molecular fingerprint compound and the score plot indicates where the sample site soil leachates plot relative to the first and second principal components. The coefficient plot illustrates the measured y-variable compared to the predicted y-variable. Additionally, the coefficients are tabulated so that a linear combination of coefficients for each vegetation attribute can be combined to obtain a prediction model for that specific molecular fingerprint compound. For example, if the molecular fingerprint compound is dimethyl benzene, a possible linear combination may be: % dimethyl benzene = 0.02(birch) + 0.55 (aspen) – 5(spruce) and so on for all vegetation attributes.

4.8 LIMITATION OF STATISTICAL PREDICTION

Limitations within statistical analyses exist. This study includes only a limited number of observations from each soil leachate site with a limited number of vegetation attributes. Boreal forests have diverse vegetation which is certainly not completely characterized by this study. The aim of this study was to determine statistical differences among dominant vegetation attributes. By identifying the dominant vegetation attributes within the research watershed, statistically significant differences were determined.

The largest limitation in the study was the small data set. Small data sets hinder the ability to make definitive conclusions about the population as a whole. However, a number of statistical tests were applied to the data set to determine whether the data set departed from a hypothesized normal distribution.   Due to the variety of the statistical tests applied it can only be assumed that for an infinitely large data set it will be normally distributed. While 3 of the 16 molecular fingerprint compounds were found to be non-normal based upon their kurtosis values, it would be expected that if the number of observations would be increased, the kurtosis values would come into the acceptable range.

CHAPTER 5: RESULTS AND DISCUSSION

5.1 GENERAL RESULTS

The combination of univariate and multivariate statistics illustrated that cover vegetation has an impact on the DOM in soil leachates. A number of molecular fingerprint compounds were statistically significant in two or more different analyses. This indicates their importance in the prediction of vegetation attributes for specific leachates.

The relative percentages of molecular fingerprint compounds for all soil leachates were plotted in Figure 1. Differences between each of the soil leachates can be seen visually. The soil leachates were then grouped according to their vegetation attributes (trees, shrubs and ground cover) and analyzed by both univariate and multivariate statistical analysis.

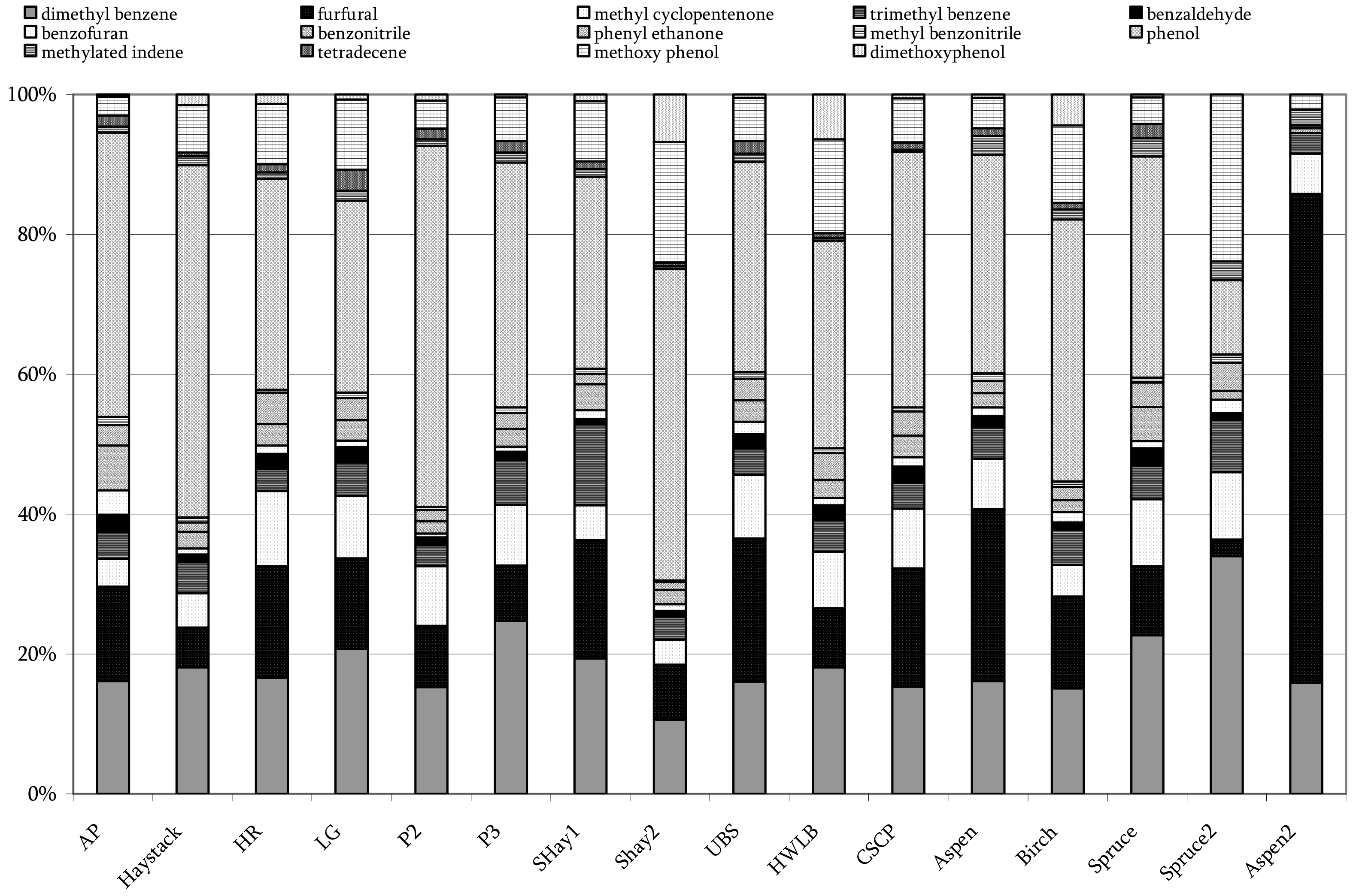

Figure 1: Relative Percentages for Each Molecular Fingerprint Compound for Each Sample Site.

As previously mentioned, each leachate was labeled with vegetation attributes. Each leachate was assigned one tree, one shrub, and one ground cover vegetation attribute. If the sample site for the leachate contained no dominant tree, shrub or ground cover, the leachate was assigned 'none' for that attribute and given a (-)1 value. For example, soil leachate from sample site P3 was assigned spruce, shrub birch, and sphagnum moss as vegetation attributes. A complete list of assigned vegetation attributes of all soil leachates can be found in Table 1. Appendix E, Table E-1, contains all relative percentages of the molecular fingerprint compounds for all soil leachates.

5.2 CREATING THE MOLECULAR FINGERPRINT

The pyrolysis products produced provide a molecular fingerprint which can be used to characterize soil leachates. This method of using pyrolysis products as a fingerprint is well documented (Joly et al., 2000; Page et al., 2002; White and Beyer, 1999; White et al., 2004). However, the analysis for this research project differed from previous methods in two ways. The first difference was that only 16 compounds were considered for the molecular fingerprint while some researchers used many more compounds (Davies et al., 2001; Hempfling and Schulten, 1990). It has been shown that some pyrolysates of humic acid can contain more than 250 pyrolysis products (Joly et al., 2000). A smaller molecular fingerprint was chosen to simplify the overall characterization of soil leachates and it has been shown that only a small portion of the pyrolysis products are needed to draw differences between individual and groups of soil and water samples (Joly et al., 2000; Page et al., 2002; White and Beyer, 1999; White et al., 2004). Within the smaller molecular fingerprint chosen for this study, the most abundant functional groups commonly found in organic matter were included. The functional groups in this molecular fingerprint are phenolics, benzenes, indenes, chains, and nitriles. As shown in Figure 2: Proposed Chemical Structure for Humic Acid (Schwarzenbach et al., 2003), the major functional groups from humic acid were included within the molecular fingerprint.

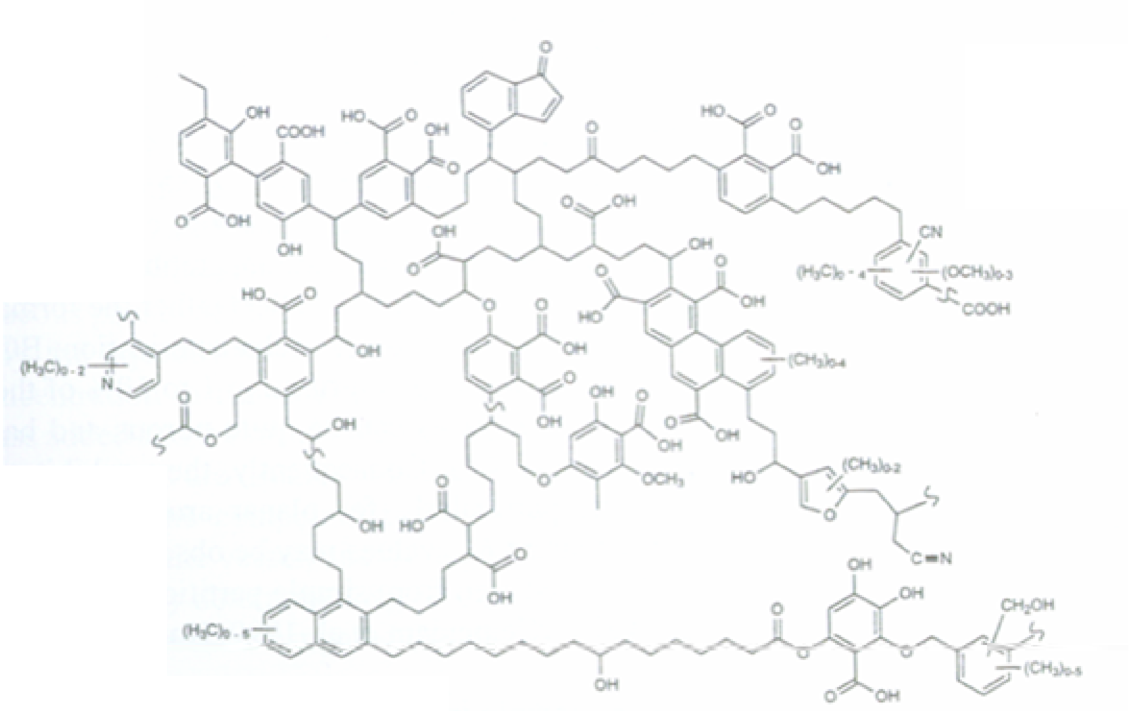

Figure 2: Proposed Chemical Structure for Humic Acid.

The second difference is how the relative abundance of each molecular fingerprint compound was determined. Samples were compared based on the relative abundance of a subset of individual molecular fingerprint compounds. In this study, only the most abundant ion for each molecular fingerprint compound was integrated. Extracting major ions created a smoother baseline and allowed for separation of overlapping compounds in the fingerprints (White et al., 2002; Tinoco et al., 2002) Other studies use a specific peak at a certain retention time or mass to charge ratio (*m/z*) (Joly et al., 2000). This limits the ability to which one is able to confidently identify the pyrolysis products. If a particular sample has a very high or very low concentration of organic matter, the retention time or *m/z* ratio may differ slightly, altering the outcome of the overall relative percentages for the pyrolysis products.

5.3 STATISTICAL DISCUSSIONS

The distribution for each molecular fingerprint compound was found to be normally distributed with moderate confidence. Table 3 contains the mean values, standard error, median, standard deviation, sample variance, kurtosis, skewness, range and confidence interval at 95% for all molecular fingerprint compounds. Each molecular fingerprint compound had 14 points. This is below the 50 data points recommended by McBean and Rovers (1998) to confidently compute the comparison of experimental distribution for each molecular fingerprint compound versus the normal distribution.

Based on histograms (Appendix B, Figures B-1 – B-14) and normal probability graphs (Appendix C, Figures C-15 – C-29), overall the molecular fingerprint compounds were normally distributed. It has been noted by McBean and Rovers (1998), that while a distribution is symmetrical it may not be normally distributed. As highlighted by dark box lines in Table 3, 3 of 14 compounds had kurtosis values above the recommended value of 3.0. Values above 3.0 indicate the distribution of the variable is more peaked than a normal distribution. A more peaked distribution can be seen in the histograms of those compounds with high kurtosis values, trimethyl benzene, benzofuran, and tetradecene.

Of the 14 molecular fingerprint compounds, 9 compounds were moderately skewed. As shown in Appendix B, Figures B-1 –B-14, those compounds with an obvious left or right skew tendency have an extreme skewed distribution. Skewness is only one indicator of a data set's tendency towards normal distribution. It can be said that through the combination of histograms, skewness, and kurtosis variables for all molecular fingerprint compounds, this data can be classified to have a moderately normal distribution.

5.3.1 IDENTIFYING VARIABILITY WITHIN SOIL LEACHATES

One soil leachate was chosen randomly to determine the variability in the compounds chosen for the molecular fingerprint and of the analytical analysis, py-GC/MS. This was done using the statistical analysis of Q-test and a computation of the 95 % confidence interval.

Soil leachate CSCP was analyzed 10 times by py-GC/MS. The 10 replicate sample runs were taken from the same DOM from a single leached core. Relative percentages were computed for each molecular fingerprint compound and the Q-test was performed. Table 4 contains the results from this analysis. The range and gap for each molecular fingerprint compound was computed. Those values within the dark boxes are outside the acceptable limits for the Q-test for 10 observations with the $Q_{critical}$ at 90 % confidence for 10 observations at 0.41 (Dean and Dixon, 1951; Rorabacher, 1991). Two soil leachate samples from the 10 replicate samples were removed from the molecular fingerprint because they failed the Q-test. Table 5 contains the $Q_{critical}$ values for varying number of observations (Harris, 1999).

Table 4: Q-Test: Values for Rejection of Data ($Q_{critical} = 0.41$).

Replicate Soil Leachate ID	dimethyl benzene	furfural	methyl cyclopentenone	trimethyl benzene	benzaldehyde	benzofuran	benzonitrile	phenyl ethanone	methyl benzonitrile	napthalene	phenol	2 methyl napthalene	methylated indene	tetradecene	methoxy phenol	dimethoxy phenol
CS01	15.19	14.93	8.10	3.56	2.35	1.34	3.13	3.53	0.52	1.98	35.61	2.77	0.31	1.00	5.19	0.50
CS03	15.62	13.55	7.77	3.65	2.20	1.42	3.21	3.69	0.55	2.43	36.10	3.36	0.31	1.02	4.61	0.51
CS04	14.04	16.87	8.43	3.33	2.13	1.24	2.94	3.27	0.49	1.59	35.73	2.35	0.28	1.11	5.65	0.57
CS05	14.81	15.07	8.23	3.56	2.23	1.35	2.85	3.39	0.51	1.87	34.56	2.78	0.31	1.01	6.95	0.53
CS06	13.63	17.95	8.58	3.23	2.32	1.26	2.73	3.05	0.46	1.45	35.35	2.15	0.28	0.95	6.05	0.56
CS07	15.20	16.57	7.80	3.34	2.16	1.42	2.94	3.25	0.53	2.41	33.35	3.42	0.29	1.00	5.87	0.46
CS08	15.06	15.64	8.08	3.52	2.24	1.40	2.96	3.29	0.53	2.06	34.09	3.01	0.30	1.02	6.32	0.49
CS09	13.38	18.44	8.48	3.19	2.33	1.27	2.59	2.92	0.46	1.41	33.99	2.13	0.27	1.13	7.42	0.60
CS10	14.43	**24.44**	**9.31**	3.50	**3.09**	1.38	3.26	3.26	0.43	1.31	22.29	1.93	0.28	1.07	**9.39**	0.59
CSCP	16.80	14.50	8.30	3.89	2.44	1.35	3.56	4.04	**1.00**	2.13	**30.22**	3.03	0.33	1.03	6.92	0.46

Range	3.42322305	10.89343	1.54	0.70	0.96	0.17	0.97	1.12	0.57	1.12	13.80	1.49	0.06	0.18	4.78	0.14
Gap (high)	1.17648362	6.001434	0.73	0.24	0.66	0.02	0.30	0.35	0.45	0.02	0.36	0.06	0.01	0.03	1.97	0.01
Gap (low)	0.25557008	0.950407	0.02	0.05	0.03	0.01	0.14	0.13	0.03	0.10	7.93	0.20	0.01	0.05	0.58	0.03
Qcalc (high)	0.34	**0.55**	**0.48**	0.34	**0.68**	0.10	0.31	0.31	**0.79**	0.02	0.03	0.04	0.24	0.14	**0.41**	0.05
Qcalc (low)	0.07	0.09	0.01	0.07	0.03	0.08	0.14	0.12	0.04	0.09	**0.57**	0.13	0.20	0.26	0.12	0.18

Table 5: Values of Q for Rejection of Data.

$Q_{critical}$ (90% confidence)	Number of Observations
0.76	4
0.64	5
0.56	6
0.51	7
0.47	8
0.44	9
0.41	10

Those CSCP samples that passed the Q-test were used to determine a final index of compounds. This was completed by calculating the 95 % confidence interval and then the percent error associated with each index compound. The percent error rejection value was 10 %. As shown in Table 6, with dark boxes, only two compounds were removed from the rough index compound list, naphthalene and 2-methyl naphthalene. The remaining 14 index compounds were at or under the 10 % percent error. As previously shown, the confidence interval takes into account the standard deviation for the average of the sample for which it was calculated. Due to the standard deviation being so large relative to its average, naphthalene and 2-methyl naphthalene had a considerably large percent error, 17% and 15%, respectively.

Table 6: Replicate CSCP Molecular Fingerprint.

Molecular Fingerprint Compound (Raw)	Average Major Ion Percentages	Standard Deviation	Confidence Interval 95 %	Percent Error (%)
dimethyl benzene	14.62	0.82	0.57	3
furfural	16.13	1.64	1.14	6
methyl cyclopentenone	8.18	0.30	0.21	2
trimethyl benzene	3.42	0.17	0.12	3
benzaldehyde	2.25	0.08	0.06	2
benzofuran	1.34	0.07	0.05	3
benzonitrile	2.92	0.20	0.14	4
phenyl ethanone	3.30	0.25	0.17	4
methyl benzonitrile	0.50	0.03	0.02	4
napthalene	1.90	0.40	0.28	12
phenol	34.85	0.99	0.68	2
2 methyl napthalene	2.75	0.51	0.35	11
methylated indene	0.29	0.02	0.01	3
tetradecene	1.03	0.06	0.04	3
methoxy phenol	6.01	0.91	0.63	9
dimethoxy phenol	0.53	0.05	0.03	5

Molecular Fingerprint Compound (adjusted)	Average Major Ion Percentages	Standard Deviation	Confidence Interval 95 %	Percent Error (%)
dimethyl benzene	15.34	1.00	0.69	5
furfural	16.90	1.60	1.11	7
methyl cyclopentenone	8.58	0.24	0.16	2
trimethyl benzene	3.59	0.21	0.14	4
benzaldehyde	2.35	0.08	0.05	2
benzofuran	1.40	0.09	0.06	4
benzonitrile	3.06	0.23	0.16	5
phenyl ethanone	3.46	0.28	0.20	6
methyl benzonitrile	0.53	0.04	0.03	5
phenol	36.55	1.05	0.73	2
methylated indene	0.31	0.02	0.01	4
tetradecene	1.08	0.06	0.04	4
methoxy phenol	6.29	0.92	0.64	10
dimethoxy phenol	0.55	0.04	0.03	6

After the removal of the naphthalene and 2-methyl naphthalene, all compounds were below 10 % error. With 95 % confidence, the variability within the fingerprint compounds was acceptable. Figure 3 shows the mean relative percentages for each molecular fingerprint compound with 95 % confidence error bars from the replicate CSCP soil leachate runs.

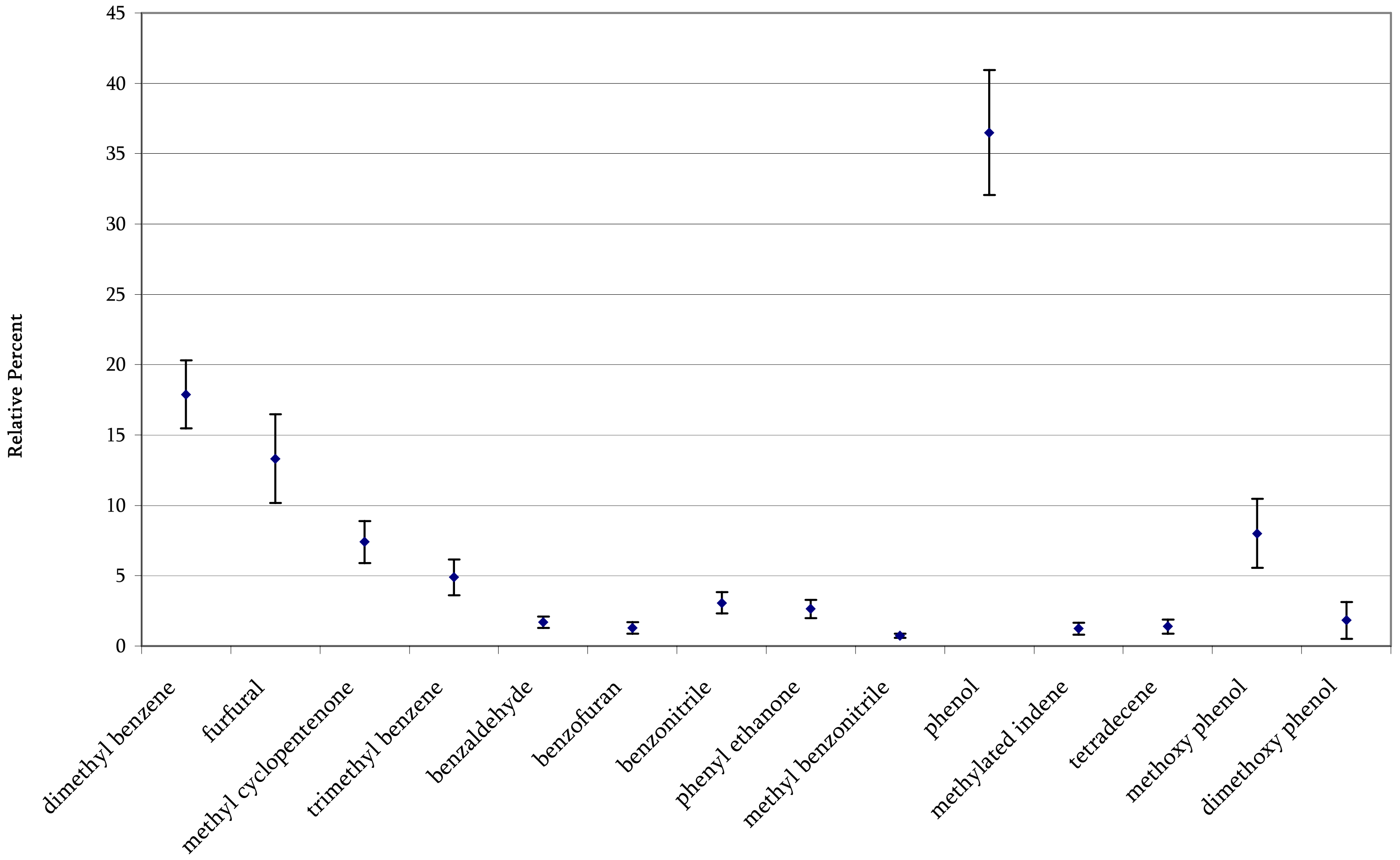

Figure 3: Mean Values of Molecular Fingerprint Compounds for Mean Values of Duplicate CSCP.

Once the variability in the molecular fingerprint compounds and the analytical technique was controlled, the Q-test was applied to the remainder of the soil leachates. The DOM from each soil leachate was run in replicates of four in order for the Q-test for outliers to be applied. Table E-2, in Appendix E contains the Q-test results for all soil leachates and their molecular fingerprint compounds with those molecular fingerprint compounds which failed the Q-test in a dark box. The $Q_{critical}$ for four observations is $Q_{critical} = 0.76$ (Dean and Dixon, 1951; Rorabacher, 1991). The final data set used in all statistical analysis is the averaged values of the molecular fingerprint compounds remaining after outliers were removed through the Q-test from each core site. Figure 1 plots the final relative percentages of each molecular fingerprint compound for each soil leachate. Table E-1, Appendix E, contains the final data set used for statistical analysis.

5.4 RESULTS FROM ANOVA

One-way analysis of variance (ANOVA) is a statistical tool used in this research project to compare the mean of several vegetation attributes (i.e. trees) for each molecular fingerprint compound. ANOVA provides a test to determine whether to accept or reject the hypothesis that all of the vegetation attributes are equal. For example, the ANOVA test was used to determine whether a statistical difference occurred within the comparison of all trees (aspen, birch, and spruce). To qualify as statically significant, the computed F needs to be equal or larger than $F_{critical}$. Table 7 contains all ANOVA results for trees, shrubs, and ground cover.

Table 7: ANOVA $F_{critical}$, F, and p-values for Molecular Compounds for Each Vegetation Group.

	Trees ($F_{critical}$=5.143)		Shrubs ($F_{critical}$=7.708)		Ground Cover ($F_{critical}$=3.585)	
Compounds	F	p-value	F	p-value	F	p-value
dimethyl benzene	0.924	0.447	0.185	0.689	4.131	0.034
furfural	2.215	0.190	0.200	0.678	1.508	0.267
methyl cyclopentenone	7.897	0.021	3.474	0.136	2.818	0.088
trimethyl benzene	0.271	0.772	0.024	0.884	0.510	0.684
benzaldehyde	0.663	0.549	1.584	0.277	0.858	0.491
benzofuran	0.118	0.891	45.863	0.002	1.384	0.299
benzonitrile	0.117	0.891	37.211	0.004	0.752	0.548
phenyl ethanone	0.941	0.441	0.033	0.865	0.839	0.500
methyl benzonitrile	1.348	0.329	11.951	0.026	1.927	0.184
phenol	0.092	0.913	0.000	0.992	0.287	0.834
methylated indene	3.018	0.115	0.037	0.858	0.259	0.853
tetradecene	0.687	0.540	0.614	0.477	3.585	0.051
methoxy phenol	0.556	0.600	4.292	0.107	2.174	0.149
dimethoxyphenol	5.241	0.045	1.379	0.305	4.191	0.033

5.4.1 ANOVA: TREES

Out of 14 ANOVA tests for the determination of different mean values from the vegetation attribute of trees, only 2 were statistically significant at p-values below

$p<0.050$. Methyl cyclopentenone ($F=7.897$, $p<0.021$) and dimethoxy phenol ($F=5.241$, $p<0.045$) had F statistic values above the $F_{critical}$ ($F_{critical}=5.143$). Two other molecular fingerprint compounds, furfural ($F=2.215$, $p<0.190$) and methylated indene ($F=3.018$, $p<0.115$), had moderate statistical significance at p-values below $p<0.200$.

5.4.2 ANOVA: SHRUBS

Because there were only two shrub attributes, shrub birch and arctic rose, it was more appropriate to perform individual student–t tests. Discussion and analysis of the statistical significant molecular fingerprint compounds for shrubs will be in section 5.5.2: Student-t Test: Shrubs.

5.4.3 ANOVA: GROUND COVER

There were four different vegetation attributes in the ground cover group: feather moss, sphagnum moss, lichen, and grass. Table 7 contains $F_{critical}$ and p-values for ground cover one-way ANOVA analysis. At $F_{critical}=3.585$, there were 6 molecular fingerprint compounds statistically significant at p-values greater than $p<0.200$.

A strong statistical significance at p-value, $p<0.05$ was observed for 2 compounds, dimethyl benzene ($F=4.131$, $p<0.034$) and dimethoxy phenol ($F=4.191$, $p<0.033$). Four compounds were observed at a moderate statistical significance at $p<0.200$. They were methyl cyclopentenone ($F=2.818$, $p<0.088$), methyl benzonitrile ($F=1.927$, $p<0.184$), tetradecene ($F=3.585$, $p<0.051$) and methoxy phenol ($F=2.174$, $p<0.149$).

5.5 STUDENT-t TESTS

Student-t tests were preformed on all molecular fingerprint compounds and all pair-wise vegetation attributes. Table 8 contains all pair-wise student-t test p-values for molecular fingerprint compounds and vegetation attributes. For example, vegetation attribute birch was compared to vegetation attribute aspen for all 14 molecular fingerprint compounds. This was done instead of conducting a post-hoc test after the determination of statistical significance within the ANOVA results previously presented. The disadvantage of performing a two-tailed student-t test is the possibility of committing at least one type I error. A type I error identifies a false statistical significance when it does not exist. The result of this is the possibility of identifying molecular fingerprint compounds as statistically significant when, in fact, they are not. Student-t tests were preferred over post-hoc tests primarily because of the unequal number of points within each pair-wise comparison, that is, there were only two soil leachates with vegetation attribute lichen compared to 4 soil leachates with sphagnum moss as vegetation attribute. It will be shown that this decision was not unreasonable and the results for pair-wise student-t tests did not contradict ANOVA results for statistical significant molecular fingerprint compounds.

Table 8: Student-t Test p-values for Soil Leachates.

Molecular Fingerprint Compounds	Birch to Aspen	Birch to Spruce	Aspen to Spruce	Shrub birch to Arctic Rose	Feather moss to Sphagnum	Feather moss to Lichen	Feather moss to Grass	Sphagnum moss to Lichen	Sphagnum moss to Grass	Lichen to Grass
dimethyl benzene	0.956	0.241	0.585	0.689	0.648	**0.0004**	**0.0429**	**0.140**	0.301	**0.033**
furfural	**0.074**	0.787	0.191	0.678	**0.081**	0.453	0.281	0.559	0.887	0.769
methyl cyclopentenone	0.444	**0.007**	**0.035**	**0.136**	0.263	0.084	0.191	0.440	**0.101**	**0.056**
trimethyl benzene	0.710	0.532	0.894	0.884	0.476	**0.071**	0.651	**0.125**	0.990	0.152
benzaldehyde	0.575	0.303	0.907	0.277	0.668	0.663	0.174	0.517	0.309	0.327
benzofuran	0.770	0.645	0.923	**0.002**	**0.108**	0.355	0.338	0.796	0.225	0.327
benzonitrile	0.683	0.803	0.731	**0.004**	0.526	0.796	0.223	0.808	0.386	0.566
phenyl ethanone	0.804	0.285	0.319	0.865	0.729	0.335	0.267	0.620	0.338	0.228
methyl benzonitrile	0.191	0.337	0.479	**0.026**	**0.055**	0.915	0.245	**0.116**	0.948	0.427
phenol	0.699	0.697	0.998	0.992	0.638	0.638	0.474	0.619	0.750	0.233
methylated indene	**0.045**	**0.118**	0.450	0.858	0.915	0.915	0.601	0.579	0.676	0.350
tetradecene	0.385	0.286	0.814	0.477	0.513	0.513	0.282	**0.119**	0.205	0.206
methoxy phenol	**0.136**	0.569	0.666	**0.107**	0.389	0.389	**0.008**	0.750	0.225	0.408
dimethoxy phenol	0.256	**0.020**	0.944	0.305	0.301	0.301	**0.0003**	0.557	0.148	**0.056**

5.5.1 STUDENT-t TEST: TREES

Student-t tests were preformed to compare the means of vegetation attributes birch, aspen and spruce. The p-values for each pair-wise comparison are in Table 8. Three student-t tests were computed (birch to aspen, birch to spruce and aspen to spruce) and, in at least one mean comparison, 3 molecular fingerprint compounds had a strong statistically significant p-value at $p<0.050$. Molecular fingerprint compound methyl cyclopentenone was statistically significant in birch ($n = 4$) to spruce ($n = 5$) ($p<0.007$) and aspen ($n = 2$) to spruce ($p<0.035$). Methyl cyclopentenone was the only molecular fingerprint compound to be statistically significant in two comparisons. Methylated indene has a strong statistically significant p-value in the mean comparison of birch to aspen ($p<0.045$) and was moderately significant in the mean comparison of birch to spruce ($p<0.118$). The last molecular fingerprint compound with a strong statistical significance was dimethoxy phenol for the mean comparison of birch to spruce ($p<0.020$).

A moderate statistical significance of $p<0.200$ was present in furfural ($p<0.074$, birch to aspen; $p<0.191$, aspen to spruce), methyl benzonitrile ($p<0.191$, birch to aspen) and methoxy phenol ($p<0.136$, birch to aspen).

5.5.1 STUDENT-t TEST: SHRUBS

Table 8 contains the results for the student-t test for shrubs. Overall, the mean comparison of shrub birch (n = 5) to arctic rose (n = 1) for all molecular fingerprint compounds was moderately significant with 5 of the 14 compounds resulting in p-values less than $p<0.200$. Strong statistical significance was present in 3 of the 5 compounds ($p<0.05$). Molecular fingerprint compounds benzofuran ($p<0.002$), benzonitrile ($p<0.004$) and methyl benzonitrile ($p<0.026$) exhibited strong statistical significance while molecular fingerprint compounds methyl cyclopentenone ($p<0.136$) and methoxy phenol ($p<0.107$) exhibited moderate statistical significance. Because molecular fingerprint compounds benzofuran, benzonitrile and methyl benzonitrile had $p<0.05$, it can be said that when comparing shrub birch to arctic rose the comparison is significant at 95 % for these molecular fingerprint compounds.

Vegetation attribute arctic rose was present in only one soil leachate sample. This appears to have had a strong impact on the student-t test resulting in the relatively high number of molecular fingerprint compounds exhibiting at least a moderate statistical significance ($p<0.20$). One reason the difference between the means of shrub birch to arctic rose was so significance is because shrub birch was a vegetation attribution for 5 leachate samples while arctic rose was attributed to one leachate. Unfortunately, there is no way to determine whether this data point is extreme in any way or if this soil leachate is a good example of a leachate under the influence of arctic rose. A solution would be to collect and analyze more soil leachates under the influence of arctic rose to determine its importance to the resulting molecular fingerprint.

5.5.2 STUDENT-t TEST: GROUND COVER

The vegetation group for ground cover presented the greatest opportunity for comparisons with the most number of mean comparisons within all the vegetation attributes. Student-t tests were preformed to compare the means of ground cover vegetation attributes. There were six pair-wise mean comparisons in total and of these comparisons 11 of the 14 molecular fingerprint compounds exhibited at least a moderate statistical significant p-value ($p<0.20$). This information is presented in Table 8.

Molecular fingerprint compound dimethyl benzene was the most statistically significant overall within the ground cover comparisons. This molecular fingerprint compound had $p<0.050$ in 3 of 6 pair wise mean comparisons, feather moss to lichen ($p<0.0004$), feather moss to grass ($p<0.043$) and lichen to grass ($p<0.033$). Because p-values for these comparisons were below $p<0.050$, it can be said that these comparisons are significant with at least 95 % confidence. The comparison of lichen to sphagnum moss produced a moderately statistically significant value of $p<0.14$. This relationship between lichen and sphagnum moss with respect to molecular fingerprint compound dimethyl benzene can be reported to be statistically different at 86% confidence.

Strong statistical significance was also observed between leachates with feather moss to leachates with grass vegetation attributes in molecular fingerprints methoxy phenol ($p<0.008$) and dimethoxy phenol ($p<0.0003$). Since, in this case, $p<0.01$, it can be said with 99 % there exists a statistical significance between leachates from vegetation attribute feather moss and leachates with vegetation attribute grass.

When comparing leachates with vegetation attributes feather moss ($n = 7$) to leachates with vegetation attribute sphagnum moss ($n = 6$), three molecular fingerprint compounds were observed to have a moderate statistical significance at $p<0.20$: furfural ($p<0.081$), benzofuran ($p<0.108$) and methyl benzonitrile ($p<0.055$). All other molecular fingerprint compounds were considered to not have statistical significance.

The comparison of leachates with vegetation attribute lichen to leachates with vegetation attribute grass resulted in 3 molecular fingerprint compounds to be at least moderately significant ($p<0.20$). As previously mentioned, molecular fingerprint compound dimethyl benzene was statistically significant at $p<0.033$. Since p-value is less than $p<0.05$, it can be said that for dimethyl benzene there is statistical significance at least 95 % confidence. Methyl cyclopentenone ($p<0.056$) and dimethoxy phenol were moderately significant. No other molecular fingerprint compounds were statistically significant.

5.6 PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal component analysis (PCA) was used to compare all vegetation attributes at once to all molecular fingerprint compounds in a 2-D plot. This ideally would introduce relationships between the vegetation attributes and molecular fingerprint compounds.

PCA was first computed for all vegetation attributes and molecular fingerprint compounds to determine general trends within the interrelated soil leachates. The data matrix was maximized to produce the largest amount of explained variance within the soil leachates and their relationships with vegetation and molecular fingerprint compounds. For this reason, three soil leachates were removed from the data matrix. Two samples, SPR2 and ASP2, were removed because they were collected from a different sample site (Bonanza Creek Research Watershed) than the other soil leachates, which were collected from CPCRW. These samples contributed little to the overall picture of similarities and differences within the data matrix.

The third soil leachate removed, AP, was also removed due to sample site location. While AP was collected from CPCRW, its localized site location was considerably unique. This particular soil leachate was from a soil sample taken from a site that is under the influence of a pingo. This phenomenon introduced an ecologically plausible situation in a permafrost region whereby through the creation of the pingo, the soil moisture changes over the period of years from relatively moist to relatively dry. This changing moisture content provokes the vegetation to also change which in turn changes the organic matter. While currently the soil site had dominant vegetation of shrub birch and arctic rose it is possible that it was under entirely different dominant vegetation only a short time ago. It is for this reason that AP was removed from this multivariate analysis.

The loading plot containing the remaining soil leachates is found in Figure 4. Only 8 vegetation attributes are present since AP contained the ninth (arctic rose). The explained variance for this PCA in the x-direction is 51% and in the y-direction 34 % and six principal components were analyzed overall. Table 9 lists the positive and negative relationships for principal component 1 (PC1) and principal component 2 (PC2 for molecular fingerprint compounds and vegetation attributes.

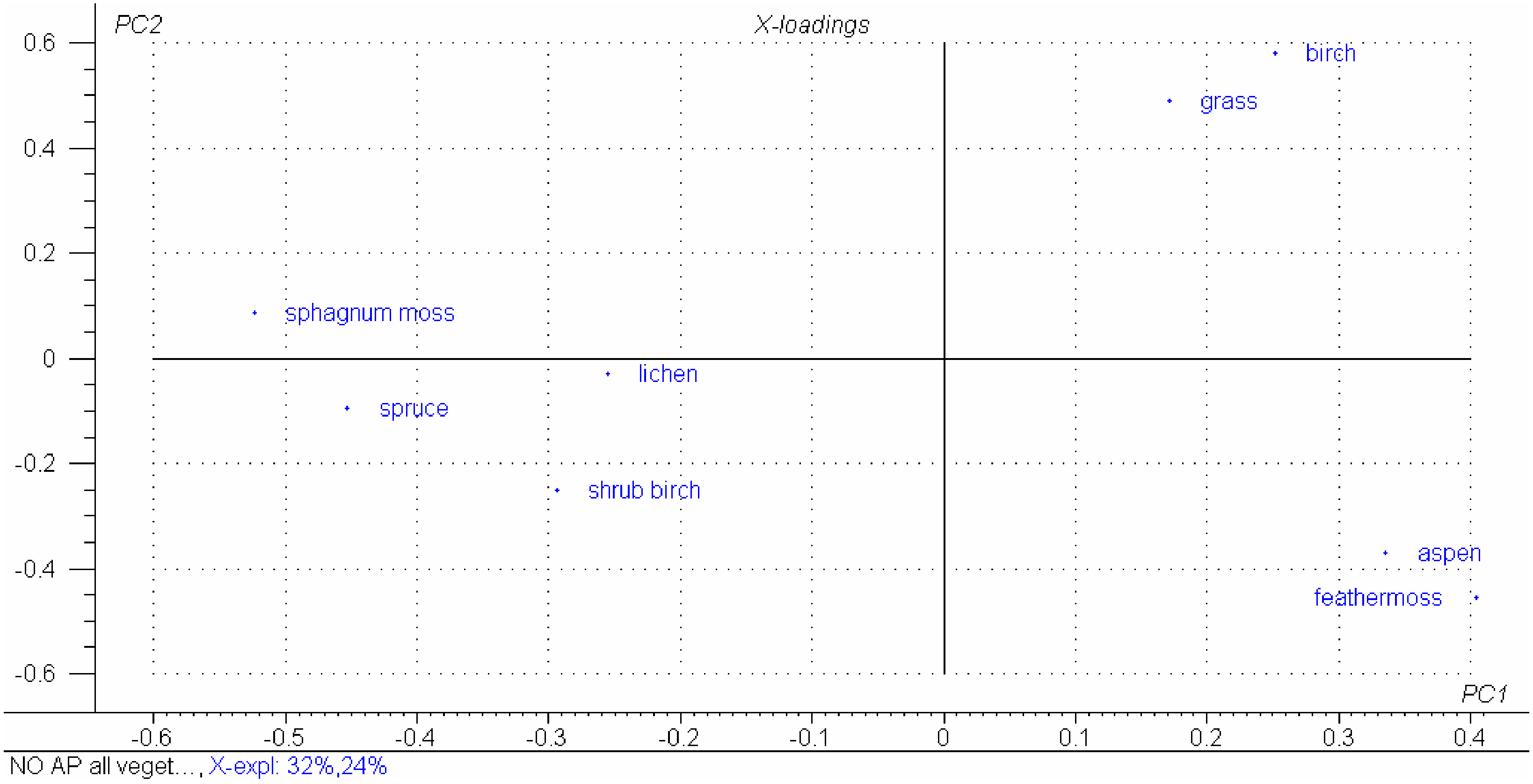

Figure 4: Principal Component Analysis Loading Plot: All Vegetation Attributes; AP Removed.

With an explained variance in the x-direction of 51 %, this PCA is relatively robust in the x-direction, which in turn places more weight on PC1 than on PC2 since only 34 % variance is explained in the y-direction. Vegetation attributes birch and grass and molecular fingerprint compounds benzofuran, methoxy phenol, and dimethoxy phenol was positively related. Vegetation attributes spruce, lichen, shrub birch and sphagnum moss and molecular fingerprint compounds methyl cyclopentenone, tetradecene, and dimethyl benzene were negatively related in PC1. These relationships are similar to those found during the computation of student-t statistical significance. Within two-tail unequal variance student-t tests for vegetation attribute group trees, birch and spruce were found to have strong statistical significance for molecular fingerprint compound methyl cyclopentenone ($p<0.007$) and dimethoxy phenol ($p<0.020$). This data suggests that the presence of molecular fingerprint compounds methyl cyclopentenone is related to spruce while molecular fingerprint dimethoxy phenol is inversely related to spruce and directly related to birch. This is relationship is confirmed by Tinoco et al (2002), who determined that pyrolytic products from humic acids derived from an oak forest (deciduous) were different from those humic acids derived from a pine forest (coniferous) based upon their relative methoxy and dimethoxy phenol percentages.

Table 9: Summary of PCA.

All Vegetation Attributes		
PC	(-) significant factors	(+) significant factors
PC1	spruce, shrub birch, lichen, sphagnum moss	birch, grass, benzofuran, methoxy phenol, dimethoxy phenol
PC2	birch, sphagnum moss, phenol, methoxy phenol,dimethoxy phenol,	aspen, feather moss, methyl benzonitrile, furfural, methylated indene, benzofuran

Only Trees		
PC	(-) significant factors	(+) significant factors
PC1	birch, diemethoxy phenol	spruce, methyl cyclopentenone
PC2	aspen, furfural	spruce, methoxy phenol, phenyl ethanone, dimethyl benzene

Only Shrubs		
PC	(-) significant factors	(+) significant factors
PC1	arctic rose, benzofuran, benzonitrile, methyl benzonitrile	shrub birch, methyl cylcopentenone
PC2	arctic rose, shrub birch, phenol	methylated indene, methoxy phenol

Only Ground cover		
PC	(-) significant factors	(+) significant factors
PC1	feather moss, benzofuran	lichen, sphagnum moss, tetradecene, methyl cyclopentenone, methoxy phenol
PC2	sphagnum moss, grass, methoxy phenol, dimethoxy phenol	lichen, feather moss, tetradecene, furfural

5.6.1 PRINCIPAL COMPONENT ANALYSIS: TREES

The loading plot containing the PCA for all molecular fingerprint compounds and only the vegetation attribute trees is depicted in Figure 5. One soil leachate sample, ASP2, was removed due to its contribution to limiting the explained variance. Soil leachate ASP2 was obtained from Bonanza Creek Research Watershed. Once ASP2 was removed, the x-direction explained 84 % of the total x-variance and in the y-direction only 19 % of the total y-variance was explained. Table 9 lists the positive and negative relationships for PC1 and PC2. Because of the limited amount of explained variance in the y-direction, the relationships presented in the y-direction are negligible in relation to the relationships presented in the x-direction.

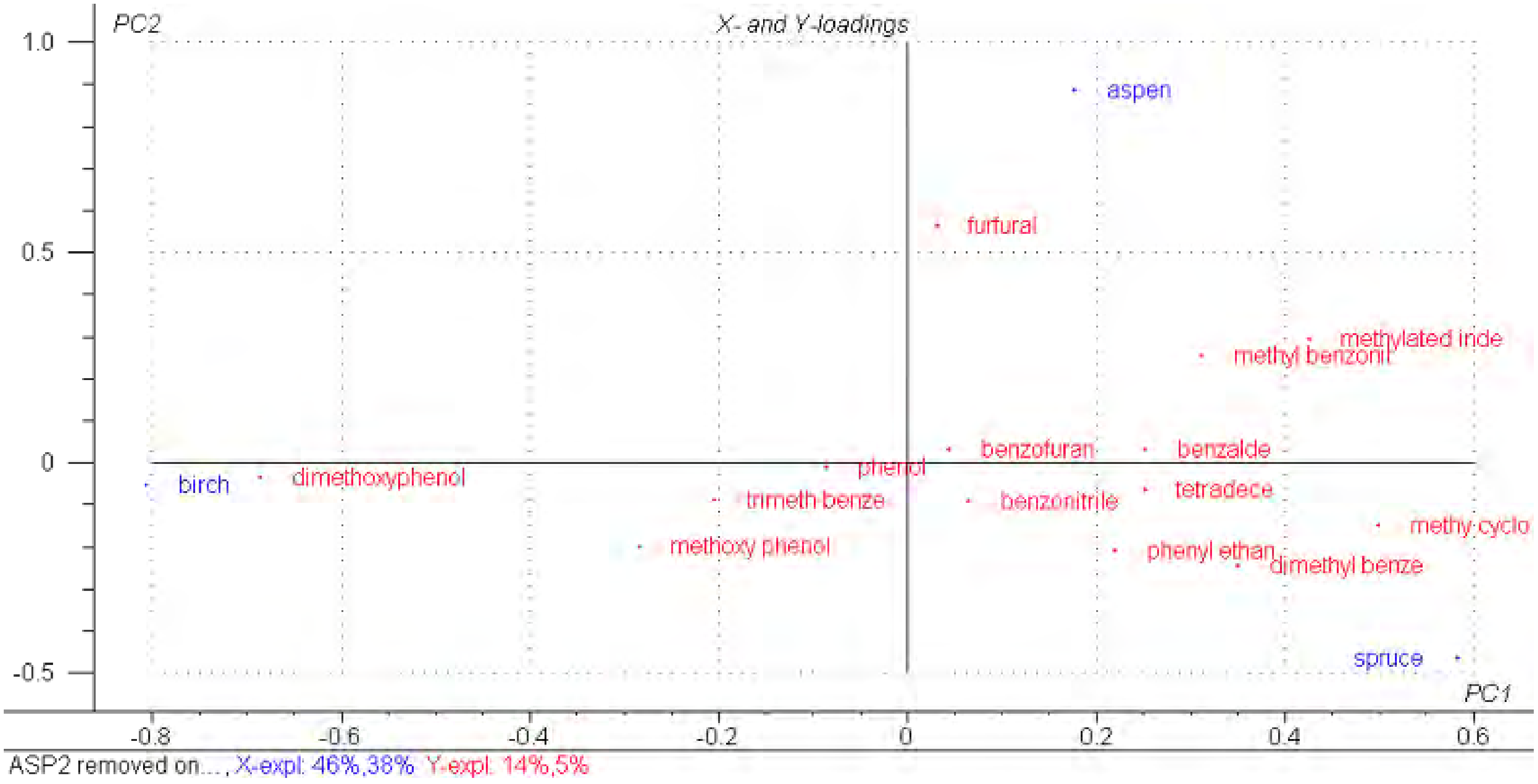

Figure 5: Principal Component Analysis Loading Plot Vegetation Attribute – Trees; AP Removed.

5.6.2 PRINCIPAL COMPONENT ANALYSIS: GROUND COVER

The loading plot containing the PCA for all molecular fingerprint compounds and only vegetation attribute ground cover is in Figure 6. The PCA for all molecular fingerprint compounds and only vegetation attributes from ground cover explained the least amount of variance in the y-direction at only 15 %. The total variance in the x-direction explained was 70 %, which is considerably higher than in the y-direction. Just as in the analysis of PCA-Trees, the relationships in PC2 are negligible compared to those in PC1.

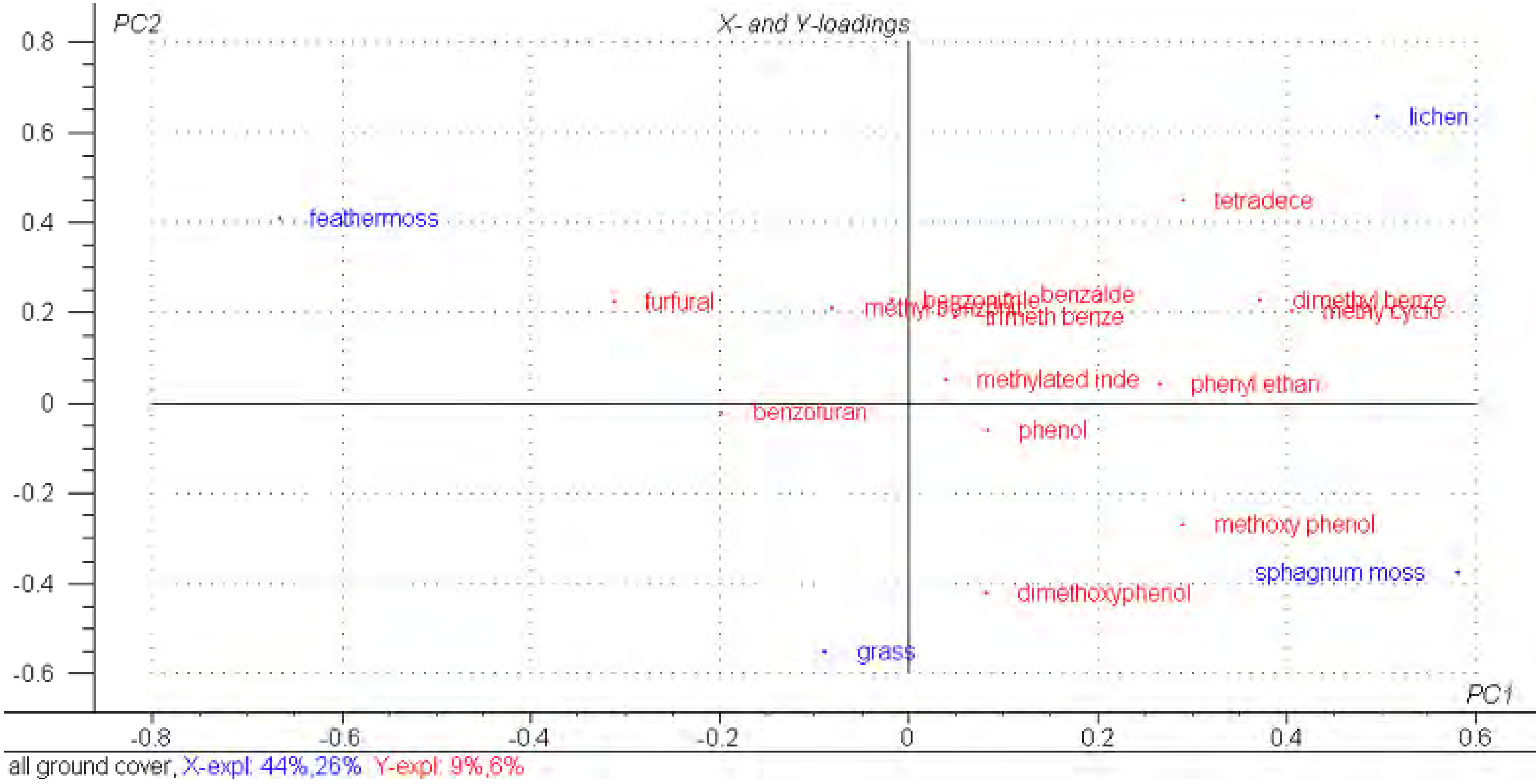

Figure 6: Principal Component Analysis Loading Plot: Vegetation Attribute –Ground Cover; AP Removed.

5.7 PARTIAL LEAST SQUARES

The second multivariate statistical analysis, partial least squares (PLS1), was used to create a prediction model. The statistical software The Unscrambler™ was used. Each molecular fingerprint compound was compared individually against all vegetation influences. By reducing the number of molecular fingerprint compounds from 14 to one, distinctions between not only molecular fingerprint compounds and vegetation could be more obvious, but relationships among the vegetation could also be more apparent.

From these analyses, a linear combination was constructed and used to predict a relative percentage of that specific molecular fingerprint compound. Table 10 contains all positive and negative significance vegetation for each molecular fingerprint compound. Each molecular fingerprint compound elicits multiple positive and negative significances and no one molecular fingerprint is constantly a positive or negative significance. Appendix D includes all PLS1 loading plots and predicted vs. measured graph for all molecular fingerprint compounds (Figures D-29 – D-65).

Table 10: Summary of PLS1 Models.

Y-Variable	(-) significant factors	(+) significant factors
dimethyl benzene	birch, feather moss	spruce, lichen, sphanum moss
furfural	spruce, sphagnum moss	feather moss
methyl cyclopentenone	aspen, birch, feather moss	lichen, sphagnum moss
trimethyl benzene	aspen, feather moss	spruce, lichen, sphagnum moss
benzaldehyde	aspen, birch, grass	spruce, shrub birch, arctic rose, lichen
benzofuran	shrub birch, sphagnum moss	arctic rose, feather moss
benzonitrile	spruce, sphagnum moss	birch, shrub birch, feather moss
phenyl ethanone	aspen, feather moss	spruce, shrub birch, sphagnum moss, lichen
methyl benzonitrile	birch, shrub birch, sphagnum moss	aspen, arctic rose, feather moss
phenol	aspen, birch	spruce, shrub birch, sphagnum moss
methylated indene	birch, shrub birch, grass	spruce, aspen
tetradecene	aspen, birch, feather moss	spruce, shrub birch, sphagnum moss, lichen
methoxy phenol	feather moss	birch, spruce, grass
dimethoxy phenol	aspen, spruce, shrub birch, feather moss	birch, grass

5.7.1 PREDICTION

An underlying goal of this study is to be able to identify regions in risk of the potential transport of non-point source pollutants due to DOM. Partial least squares regression (PLS1) is one technique used to create a prediction of future values based upon experimental values. In this study the experimental values are the molecular fingerprint compounds obtained from the soil leachates with varying vegetation attributes. The known values are the vegetation attributes. Each molecular fingerprint compound, for an example site, was analyzed separately with the input of only the vegetation attributes. From this analysis a predicted vs. measured graph was produced with coefficients for each of the vegetation attributes. These coefficients were combined to create a model in which the molecular fingerprint compounds are the ‘predictors’ and the vegetation attributes are ‘known’.   For example, Figure 7 illustrates the x- and y-loadings for methyl cyclopentenone and all vegetation attributes.

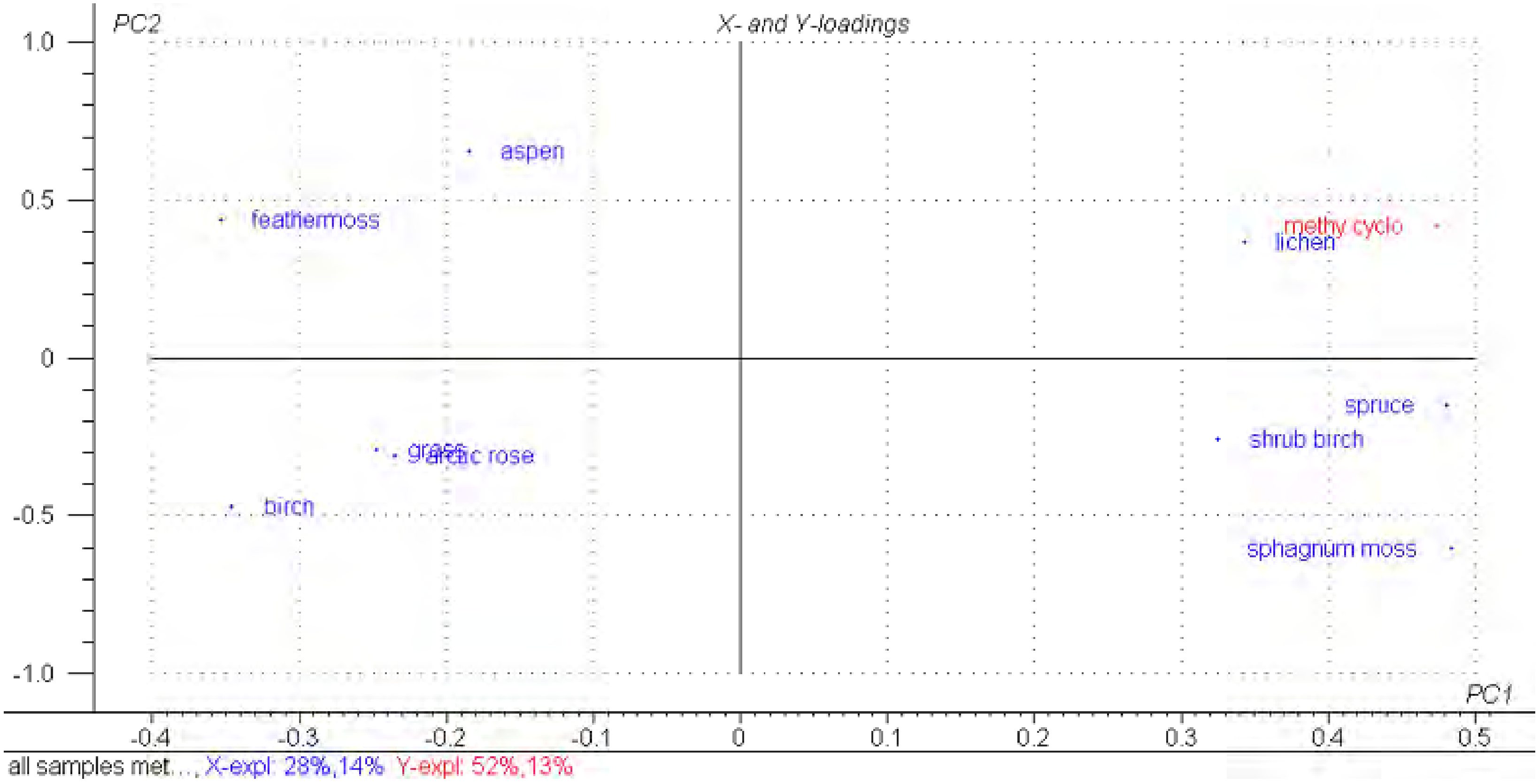

Figure 7: Partial Least Squares: Loading Plot for Methyl Cyclopentenone.

The total explained variance in the x-direction was 42 % and the total explained variance in the y-direction was 65 %. Table 10 lists the positive (+) and negative (-) significance for each molecular fingerprint. Positive significance for methyl cyclopentenone is lichen, spruce and shrub birch and the negative significance are birch, aspen and feather moss. Table 11 lists the coefficients and constants for all molecular fingerprint compounds. Larger coefficients are more influential on the final predicted relative percentage than smaller coefficients. If coefficients are negligible, the vegetation attribute for that coefficient can be removed from the prediction. For illustrative purposes, no vegetation attributes and their respective coefficients were removed during this analysis. Table 11 contains all coefficients to provide an illustration of the varied relationships between compounds and vegetation. Figure 9 contains the plot of the measured relative percentage of molecular fingerprint compound methyl cyclopentenone versus the predicted relative percentage of methyl cyclopentenone.

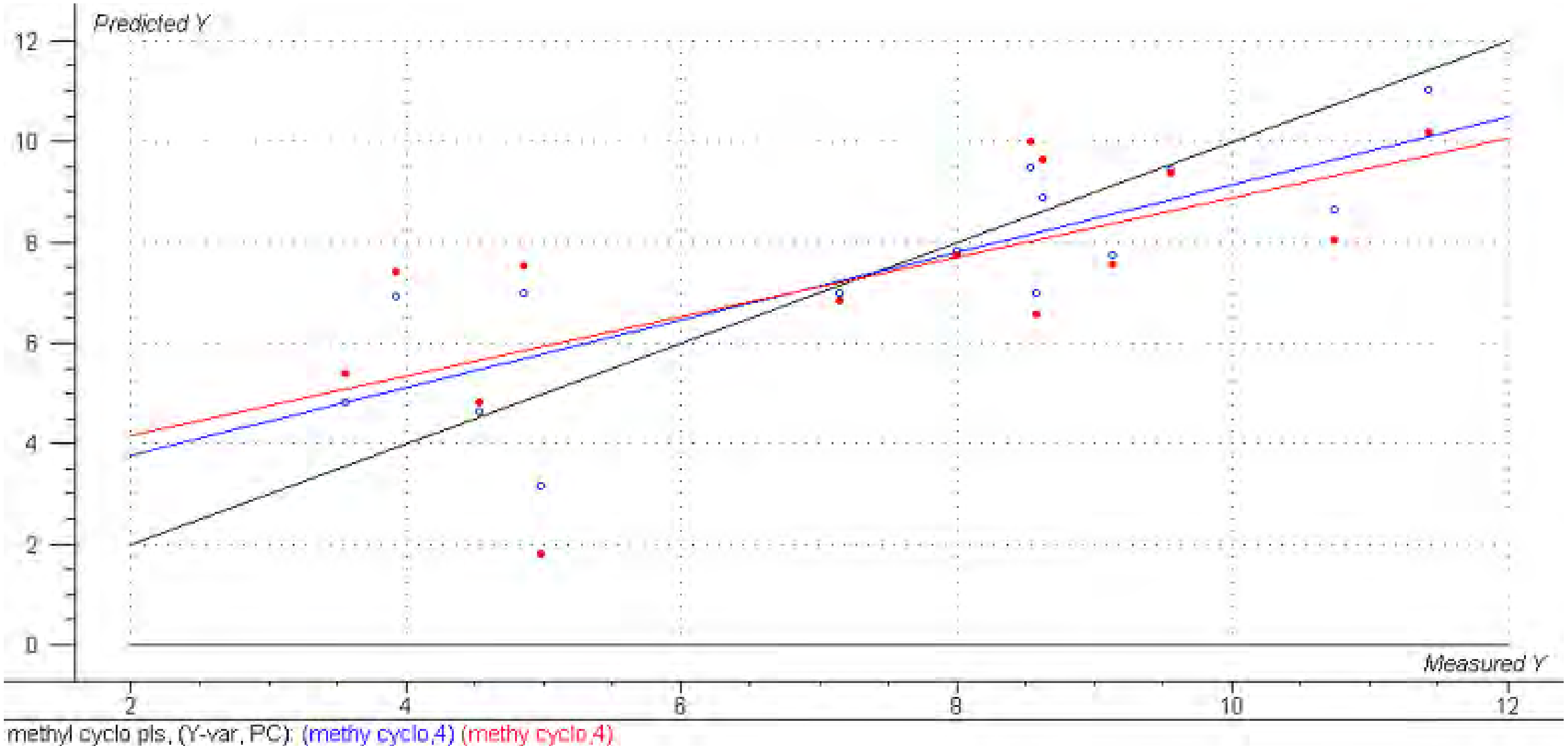

Figure 8: Partial Least Squares: Measured vs. Predicted for Methyl Cyclopentenone.

Table 11: Weighted Coefficients from PLS1.

	Tree			Shrub		Ground Cover			
Molecular Fingerprint Compounds	birch	aspen	spruce	shrub birch	arctic rose	feather moss	sphagnum moss	lichen	grass
dimethyl benzene	-0.249	-0.08752	0.274	-0.32	-0.05556	-0.149	-0.01511	0.383	-0.105
furfural	-0.008917	0	-0.134	-0.04963	0.04772	0.214	-0.247	-0.01787	0.03779
methyl cyclopentenone	-0.825	-0.367	0.03215	-0.23	-0.486	-0.612	-0.437	-0.0248	-0.276
trimethyl benzene	-0.105	-0.106	0.229	-0.298	-0.0265	-0.19	-0.06974	0.487	0.144
benzaldehyde	-0.582	-0.691	-0.232	-0.298	0.07218	-0.472	-0.466	0.05043	-0.175
benzofuran	-0.218	-0.417	0.06739	-0.323	0.639	0.0802	-0.08286	-0.11	0.16
benzonitrile	0.007594	-0.358	-0.32	0.04603	0	0.249	-0.334	0.453	-0.406
phenyl ethanone	-0.806	-0.708	-0.289	-0.425	-0.202	-0.753	-0.479	-0.195	-0.323
methyl benzonitrile	-0.243	0.0177	0.158	-0.336	0.349	0.131	-0.237	0.159	0.09887
phenol	-0.1	-0.138	0.03447	0.288	0.104	-0.03871	0.143	-0.0571	0.0355
methylated indene	-0.364	0.496	0.277	-0.482	-0.191	-0.05891	0.07114	0.0932	0.159
tetradecene	-0.239	-0.24	0.02189	-0.147	0.146	-0.114	-0.22	0.608	-0.02128
methoxy phenol	0.262	-0.218	-0.101	-0.31	-0.182	-0.277	0.11	-0.007342	-0.04147
dimethoxyphenol	0.549	-0.146	-0.291	-0.02441	-0.09948	-0.324	0.08106	-0.118	-0.0178

All soil leachates are plotted with their respective methyl cyclopentenone relative percentages on the graph in blue coloring. The predicted relative percentages for each sample soil leachate are listed in red type. The black trend line represents the ideal solution where the measured values and the predicted values form a slope of one. The blue trend line represents the actual trend line and the red tread line represents the predicted trend line. The coefficients are extracted from the predicted tread line. A linear equation in the form:

$$\text{Relative percentage compound} = \text{constant} + \Sigma\ (\text{coefficients} * \text{vegetation attribute})$$

The vegetation attribute is either a (+)1 or a (-)1 depending on the presence (+) or absence (-) of the specific vegetation. The constants are derived from The Unscramble™ coefficient listing within the PLS1 output. Chapter 6 contains an application of this prediction model.

5.8 INFLUENCE OF SPRUCE ON LEACHATE

Two molecular fingerprint compounds within this study were used to compare the influence deciduous and coniferous trees have on the organic matter beneath their canopies. According to Tinoco et al. (2002) while studying the effects of soil perturbation, the relative percentage of molecular fingerprint compound dimethoxy phenol is negligible with primarily deciduous tree influences upon the leachates while leachates obtained from coniferous tree influences produce a molecular fingerprint with higher percentages of dimethoxy phenol. Figure 10 plots soil leachates according to their tree vegetation attributes as the relative percentage of methoxy and dimethoxy phenol.

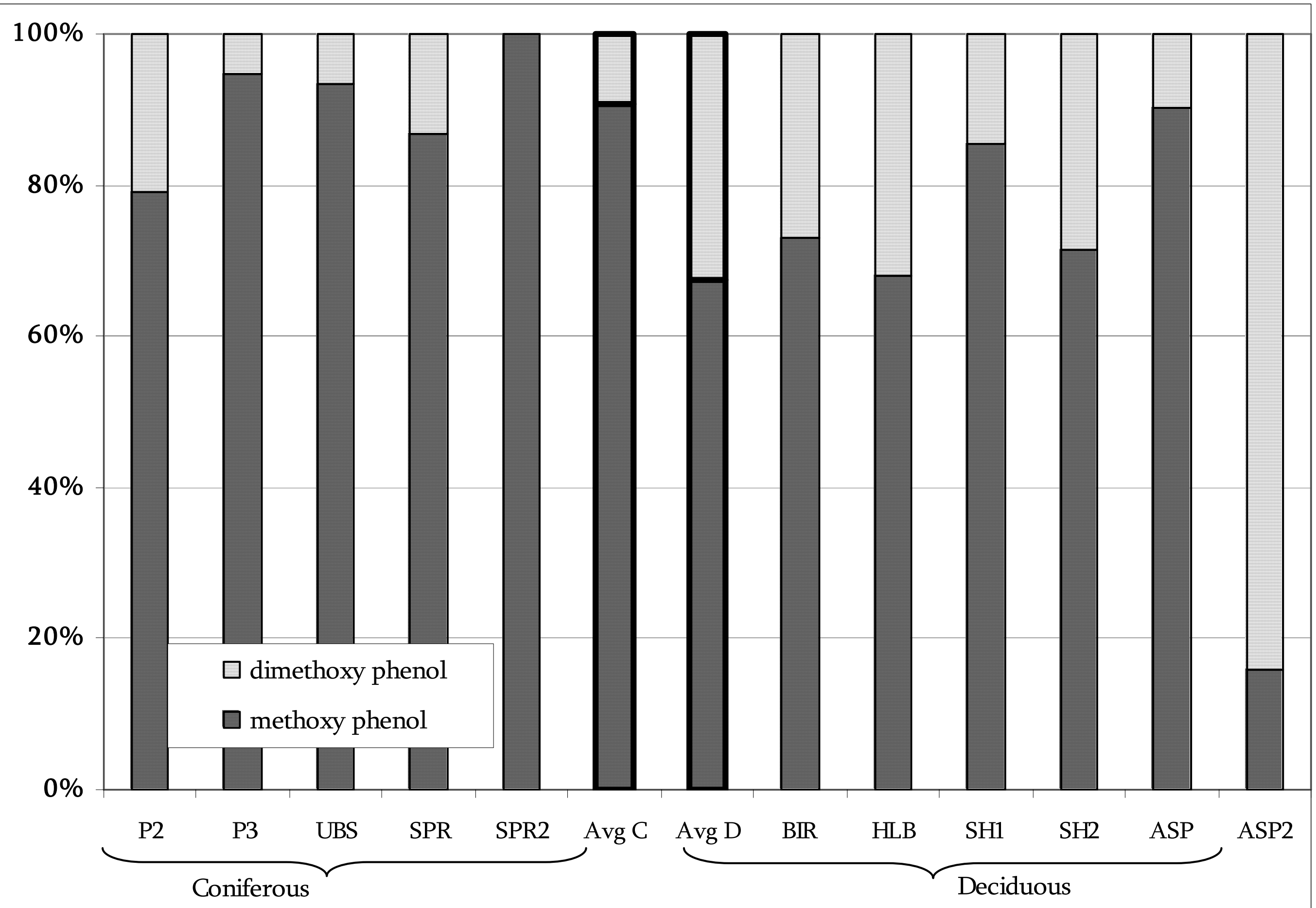

Figure 9: Relative Percentages for Methoxy Phenol and Dimethoxy Phenol.

*Avg C and Avg D are the averages of leachates with coniferous and deciduous, respectively

Soil leachates P2, P3, UBS, SPR, and SPR2 are influenced by spruce (coniferous) while soil leachates BIR, HLB, SH1, SH2, ASP, and ASP2 are influenced by aspen or birch (deciduous). The averaged values of the coniferous (Avg C) and the averaged values of the deciduous (Avg D) are outlined in heavy black. When all soil leachates used in this study are plotted according to their relative percentages of these two molecular fingerprint compounds, a moderate visual difference is observed. However, a student-t test for mean comparison with unequal variance between relative percentages of methoxy phenol compared to relative percentages of dimethoxy phenol resulted in a strong statistical significance ($p < 6.91 \times 10^{-7}$). The difference is more visually pronounced when the relative percentages for methoxy and dimethoxy phenol are averaged into a 'coniferous' and 'deciduous' value (plotted as Avg C and Avg D, respectively, Figure 10).

Table 12: Relative Percentage of Methoxy Phenol and Dimethoxy Phenol

	methoxy phenol	dimethoxy phenol
P2*	79.21	20.79
P3*	94.65	5.35
UBS*	93.29	6.71
SPR*	86.71	13.29
SPR2*	100.00	0.00
Avg C*	90.77	9.23
Avg D^	67.36	32.64
BIR^	73.02	26.98
HLB^	68.10	31.90
SH1^	85.48	14.52
SH2^	71.44	28.56
ASP^	90.25	9.75
ASP2^	15.85	84.15
mean	78.17	21.83

* indicates confiferous, ^ indicates decidous influences

Tinoco et al. (2002) suggests samples with the influence of vegetation attribute coniferous would contain zero to a very small relative percentage of dimethoxy phenol compared to a high relative percentage of methoxy phenol. It is important to note that these average values include soil leachates taken from soil cores from CPCRW and Bonanza Creek Research Watershed. While these research watersheds are relatively close regionally (Interior Alaska), they are quite different locally. This difference can be seen by comparing the soil leachates ASP and ASP2 and SPR and SPR2. As suggested by the soil leachate names, soil leachates ASP and ASP2 both had vegetation attribute aspen and soil leachates SPR and SPR2 both had vegetation attribute spruce. Figure 11 contains only these four leachates' relative percentages of methoxy and dimethoxy phenol.

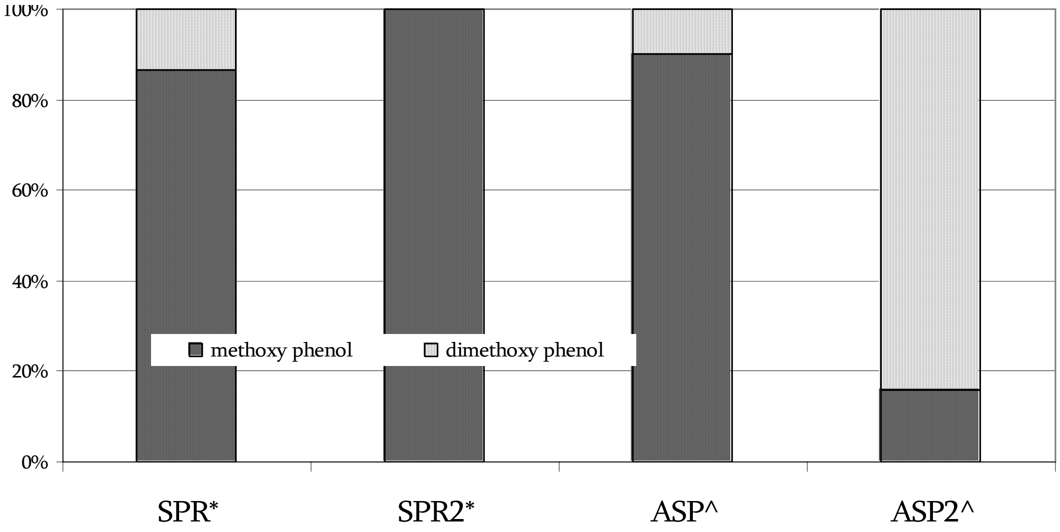

Figure 10: Relative Percentages for Methoxy Phenol and Dimethoxy Phenol for Soil Leachates: SPR, SPR2, ASP, and ASP2.

*coniferous, ^deciduous

Samples from CPCRW, soil leachates SPR and ASP, are quite similar to each other while samples from Bonanza Creek Research Watershed, soil leachates SPR2 and ASP2, display the relationship explained by Tinoco et al. (2002). Soil leachate ASP2 contains a large abundance of dimethoxy phenol relative to methoxy phenol and soil leachate SPR2 contains only methoxy phenol, the indicator molecular fingerprint compound for coniferous vegetation attributes. One explanation as to why leachates from CPCRW do not display such a significant difference within these two molecular fingerprint compounds could be due to limited number of soil leachates analyzed from the watershed. Only one sample was analyzed for each vegetation attribute (aspen and spruce). Also, CPCRW undergoes fires quite regularly with the latest fire, impacting the site where soil core used for leaching were removed, occurring during 1963 (Yoshikawa et al., 2003). Fire is a short-term influence on the vegetation in a boreal forest. Few boreal forest stands reach an age of 150-200 years and the areas typically burn every 50 – 150 years, on average (Johnson et al. 1995). Due to the natural perturbation of surface litter and surface soils after fire, vegetation may be continuously changing in CPCRW. Also, due to the slow cycling of organic matter in cold regions, the presence of spruce influence on soil leachates may still be noticeable within the molecular fingerprint. The typical succession of vegetation after a fire is related to the soil moisture. Characteristically the vegetation returns in the order of: grasses and shrubs to birch and aspen to spruce (Yoshikawa et al., 2003). This underlying influence of spruce in all soil leachates obtained from CPCRW may limit the model predictions with respect to trees and could help explain limited statistical differences seen within comparisons of tree vegetation attributes.

CHAPTER 6: APPLICATION

6.1 USING THE PLS1 PREDICTION MODEL

While dissolved organic matter may undergo numerous chemical, physical and biological transformations along its transport path, an important first step in understanding the impact of these transformations is to understand what is entering the cycle by the leaching of water through different vegetation. The application of all previous statistical analyses is the important first step.

DOM plays an important role in the transport of heavy metals in the environment. Additionally DOM abundance in drinking water is itself of concern. There are three applications of the PLS1 prediction model and are supported by student-t tests and PCA previously developed in this research project. In the first application, one knows the vegetation at a field site and predicts the relative percentages for each of the molecular fingerprint compounds. If information can be developed that links molecular fingerprint compounds to heavy metals with affinity to bind with specific molecular fingerprint compounds, knowing the relative percentage for each molecular fingerprint compound could aid in the determination of sites with the potential to mobilize heavy metals in soil leachates. The prediction model could be used in the opposite direction as well. In this case, on the knowledge of the relative percentages of the molecular fingerprint compounds in soil leachate, one could know the vegetation at the site where water entered the subsurface. The final application is in the water treatment industry. A previous study by Naar et al. (2001) studied the effects of DOM removal by nanofilters. The prediction model could help to better understand if a certain type of DOM entering the treatment system could be a fouling hazard to filters.

6.1.1 HEAVY METAL CONTAMINANT TRANSPORT

The relationship between heavy metals such as cadmium (Cd), lead (Pb) and zinc (Zn) and DOM is well understood. What is not known is the impact of vegetation on DOM and whether DOM from specific vegetation sources will have more potential to mobilize the heavy metals.

Based on the multiple linear equations obtained through PLS1, a sample site with known vegetation could be entered into the model and the relative abundance of molecular fingerprint compounds could be determined. An example of this application could be the following: Input vegetation attributes spruce and lichen as (+)1 and all other vegetation attributes as (-)1 into Table 13. Then sum the predicted relative percentages of all molecular fingerprint compounds. This is the predicted overall molecular fingerprint for this soil leachate under vegetation attributes spruce and lichen. This prediction model explained 71 % of the total molecular fingerprint while only 29 % was not explained. For a rough estimate, this result is satisfactory. According to Autier and White (2004), furfural seemed to have a high sorption capacity for Cd^{2+}. If, for example, this predictive model for vegetation attributes spruce and lichen, indicates a higher than expected relative percentage of furfural, then the sample site may be more prone to mobilization of cadmium or similar heavy metals.

Table 13: Predictive Illustration for Spruce and Lichen Vegetation Attributes.

	Tree			Shrub		Ground Cover						
Molecular Fingerprint Compounds	birch	aspen	spruce	shrub birch	arctic rose	feather moss	sphagnum moss	lichen	grass	Standard Deviation	Constant	Predicted %
dimethyl benzene	-1	-1	1	-1	-1	-1	-1	1	-1	4.19	17.69	24.56
furfural	-1	-1	1	-1	-1	-1	-1	1	-1	5.47	2.39	1.59
methyl cyclopentenone	-1	-1	1	-1	-1	-1	-1	1	-1	2.59	0.48	8.86
trimethyl benzene	-1	-1	1	-1	-1	-1	-1	1	-1	2.17	3.83	6.80
benzaldehyde	-1	-1	1	-1	-1	-1	-1	1	-1	0.69	0.40	2.08
benzofuran	-1	-1	1	-1	-1	-1	-1	1	-1	0.69	2.26	2.34
benzonitrile	-1	-1	1	-1	-1	-1	-1	1	-1	1.32	2.38	3.60
phenyl ethanone	-1	-1	1	-1	-1	-1	-1	1	-1	1.11	-0.68	2.90
methyl benzonitrile	-1	-1	1	-1	-1	-1	-1	1	-1	0.26	3.28	3.42
phenol	-1	-1	1	-1	-1	-1	-1	1	-1	7.71	4.75	2.31
methylated indene	-1	-1	1	-1	-1	-1	-1	1	-1	0.74	2.05	2.60
tetradecene	-1	-1	1	-1	-1	-1	-1	1	-1	0.85	1.68	2.93
methoxy phenol	-1	-1	1	-1	-1	-1	-1	1	-1	4.23	4.78	7.10
dimethoxyphenol	-1	-1	1	-1	-1	-1	-1	1	-1	2.27	0.40	-0.57

This model must be viewed as only a preliminary evaluation of the impact a vegetation attribute has on the molecular fingerprint. Soil leachates were obtained from a laboratory leaching experiment and the vegetation attributes were based only on dominant vegetation. These two steps account for large assumptions about the impact vegetation has on the molecular fingerprint. Leachates in a boreal forest do not flow only perpendicular from the surface. Lateral flow of precipitation after a rain event or base flow overlying permafrost or water from emerging springs will interact with numerous vegetation attributes and organic matter. Additionally, as previously mentioned, a boreal forest contains numerous vegetation attributes. This model only accounts for those determined to be dominant at a localized site.

The adsorption of cadmium to DOM is discussed in depth in Appendix A. The complexation between DOM and Cd^{2+} is controlled primarily by two factors. The first is pH. Lumsdon, D.G. (2004) determined organic soils adsorbed much more Cd^{2+} at about pH 4 when compared to the two B horizons which adsorbed little at this pH. The second factor favoring the complexation of DOM and Cd^{2+} are functional groups, including the carboxylic groups, COOH and the phenolic (Ar-OH) groups in DOM. The most abundant molecular fingerprint compound found in all soil leachates was phenol. On average, the total phenolic contribution (phenol, methoxy phenol and dimethoxy phenol) to the molecular fingerprint was 40%. This finding agrees with Page et al. (2002), who states that compounds such as phenols may provide a more significant input to DOM than a major part of plant biomass, such as lignin, that degrades more slowly. The overall high relative percentages of phenol did not lead to statistically significant relationships between molecular fingerprint compounds and vegetation attributes but its presence in the soil leachates may indicate a more elevated potential for complexation with Cd^{2+}.

6.2 IDENTIFYING PRECURSORS FOR MEMBRANE-FOULING

Based on research conducted by Naar et al. (2001) it was found that certain groups of molecular fingerprint compounds were more likely to be retained on a nanofilter. Figures 13-17 indicate the relative percentages alkyl benzenes, phenol, methylated indene, methyl cyclopentenone, and nitriles (modified from Naar et al. 2001), in addition to the averaged values of the different vegetation attributes from the soil leachates identified in this study.

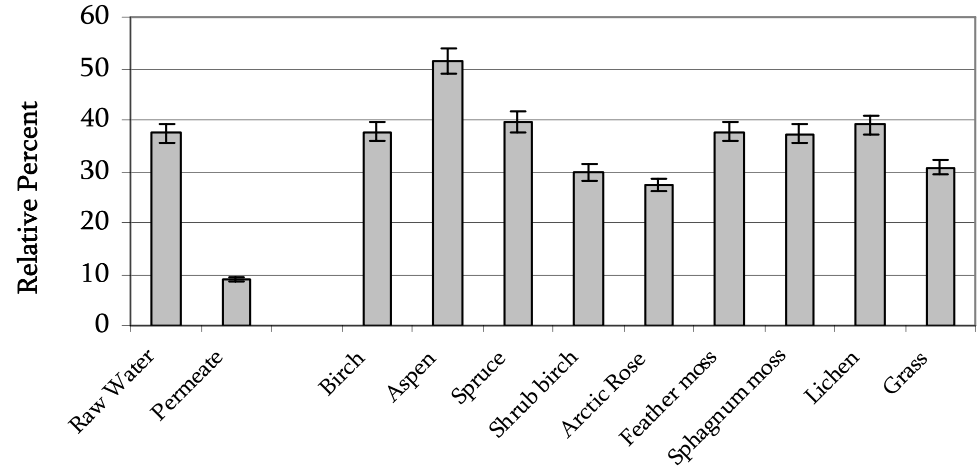

Figure 11: Relative Percentage of Alkyl Benzenes.

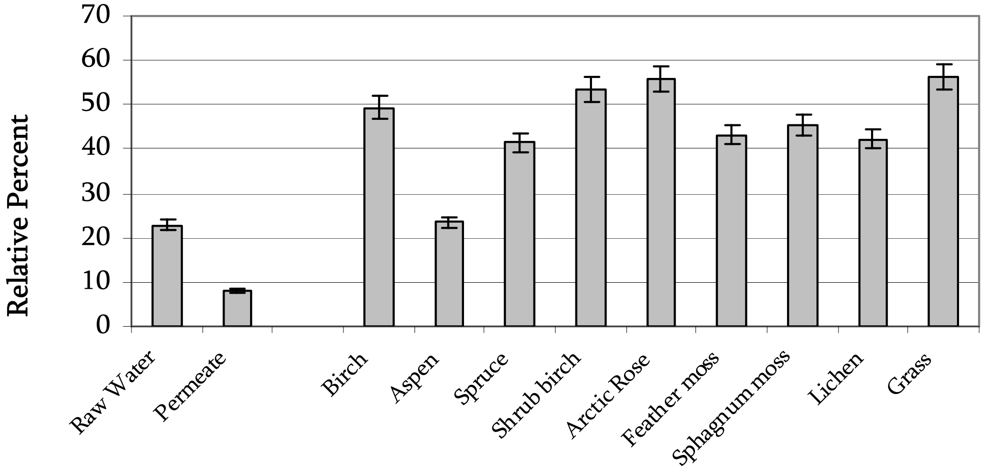

Figure 12: Relative Percentage of Phenol.

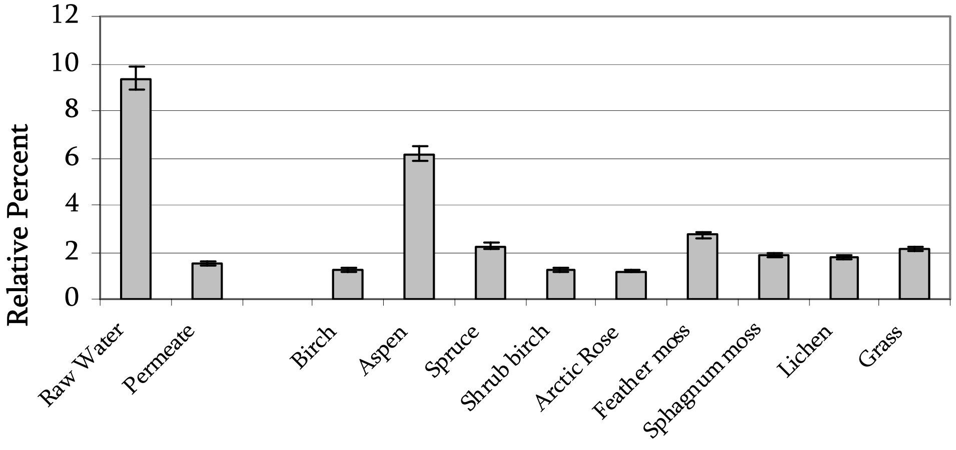

Figure 13: Relative Percentage of Methylated Indene.

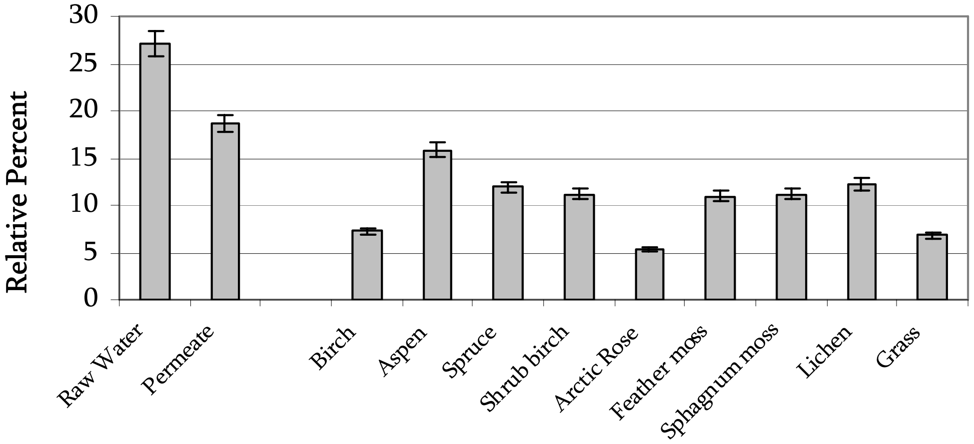

Figure 14: Relative Percentage of Methyl Cyclopentenone.

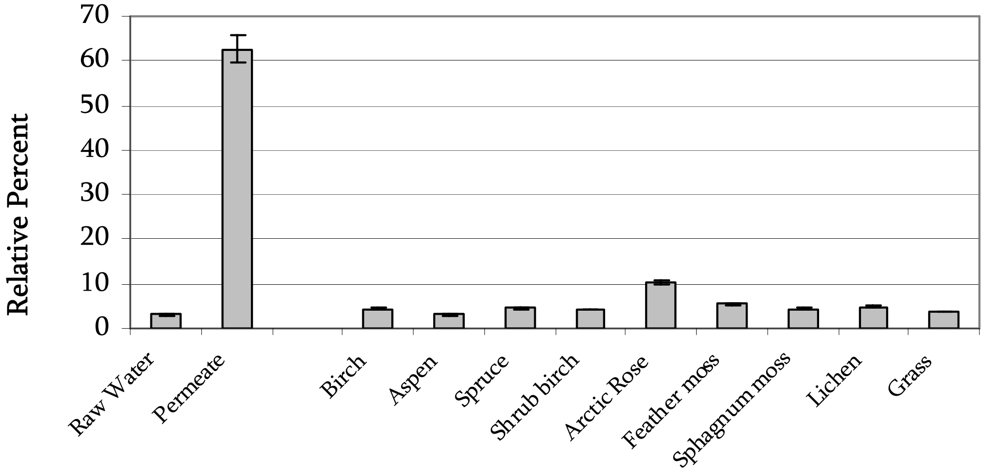

Figure 16: Relative Percentage of Nitriles.

The values for raw water refer to water entering the membrane and the values for permeate refer to the water exiting the membrane. An error of 5% was placed on each relative percentage. Two molecular fingerprint groups, methyl cyclopentenone and nitriles, had relative percentages in the raw water below that in the permeate. This means molecular fingerprint compounds methyl cyclopentenone and nitriles were not retained on the nanofilter and most likely do not contribute to membrane fouling. However, molecular fingerprint compounds alkyl benzene, phenol, methylated indene (for most vegetation attributes) had relative percentages from the leachates that were higher than the relative percent in the permeate. This suggests that they were retained on the filter and contribute to membrane fouling. Within molecular fingerprint compound methylated indene (Figure 15) important distinctions between vegetation attributes was discovered. Vegetation attributes aspen, spruce, feather moss and grass were within 95% confidence greater than the relative percentage in the permeate. Also, by referring to the student-t results for methylated indene (Table 8) and the tabulated (+) and (-) significance from PCA and PLS1 (Tables 9 and 10), it is shown that a statistical significance is present for birch to aspen ($p<0.045$) and because PCA and PLS1 related aspen to feather moss, may suggest the source of methylated indene may be related to aspen and/or feather moss. This further suggests that DOM derived from aspen and/or feather moss vegetation attributes may be more susceptible to membrane fouling than other molecular fingerprint compounds.

CHAPTER 7: CONCLUSIONS

7.1 CONCLUSIONS ABOUT THE INFLUENCE OF TREES ON LEACHATE

The influence of trees on the molecular fingerprint is statistically significant and throughout the statistical analyses of the data matrix, coniferous and deciduous trees could be differentiated based upon their molecular fingerprints. Three molecular fingerprint compounds, methyl cyclopentenone (birch to spruce, $p<0.007$; aspen to spruce, $p<0.035$), methylated indene (birch to aspen, $p<0.045$) and dimethoxy phenol (birch to spruce, $p<0.020$) appeared to be more statistically connected to trees than to the other vegetation attributes. This connection could indicate a closer relationship between the molecular fingerprint compounds and the specific tree attributes. For instance, methyl cyclopentenone appears to be directly related to spruce based on the strong statistical p-values and PCA loading plots (Table 8 and Figure 5). Also, there appears to be an underlying influence of spruce upon the leachates obtained from CPCRW from to the relationship between methoxy phenol and dimethoxy phenol.

7.2 CONCLUSIONS ABOUT THE INFLUENCE OF SHRUBS ON LEACHATE

Statistical differences between molecular fingerprints of leachates from shrub birch and arctic rose were confirmed through ANOVA, student-t, PCA and PLS1. Three molecular fingerprint compounds, benzofuran, benzonitrile and methyl benzonitrile, were determined to be statistically significant at 95 % confidence ($p<0.05$). These three compounds were also significant in PCA. Shrub birch was negatively related and arctic rose was positively related to the three molecular fingerprint compounds. Additionally, methyl cyclopentenone was moderately significant ($p<0.136$) and was positively related to shrub birch.

However, to make the data matrix more robust, more soil leachates containing arctic rose and perhaps including more shrubs as vegetation attributes would increase the overall weight of the importance of shrubs on the molecular fingerprint.

7.3 CONCLUSIONS ABOUT THE INFLUENCE OF GROUND COVER ON LEACHATE

Ground cover plays an important role in the influence on soil leachates. Strong statistical significance was discovered for all four vegetation attributes. Additionally, through the combination of PLS1 for each molecular fingerprint compound and student-t p-values, more weight can be placed on certain molecular fingerprint compounds than on others. Molecular fingerprint compounds methoxy and dimethoxy phenol were statistically significant ($p<0.008$, $p<0.0003$, respectively) in the comparison between feather moss and lichen. Feather moss was inversely related to both compounds while grass was directly related to both compounds in PLS1. Molecular fingerprint compound dimethyl benzene was statistically significant between vegetation attributes lichen and grass in the student-t test. PLS1 indicates a positive relationship between vegetation attribute lichen and molecular fingerprint compound dimethyl benzene while vegetation attribute grass is not significant in PLS1 or PCA. This suggests dimethyl benzene is related to vegetation attribute lichen.

7.4 CONCLUSIONS ABOUT POSSIBLE APPLICATIONS OF STATISTICAL PREDICTION MODEL

While the prediction model is still in development, the applied applications are promising. The vegetation attributes included within the model are common throughout boreal forests. The next stage for this prediction model is to collect more soil leachates with specific vegetation attributes. These leachates would then be subjected to the same analytical analysis and the same molecular fingerprint compounds selected as this study. The relative percentages of the molecular fingerprint compounds would be used as inputs into the prediction model. The actual relative percentages of the molecular fingerprint compounds from the pyrograms would be compared with the relative percentages the model predicted. This would be a true test of the robustness of the model.

7.5 OVERALL CONCLUSIONS ABOUT INFLUENCE OF VEGETATION ON SOIL LEACHATES

Through this study, it was determined that the chemical nature of DOM in soil leachates can be statistically related to local vegetation. Through the various statistically analyses, it was confirmed with statistical significance that certain molecular fingerprint compounds are related to specific vegetation and alternatively, the presence of some molecular fingerprint compounds are inversely related to specific vegetation.

Statistically significant differences between the chemical nature of DOM from soil leachates with varying vegetation attributes confirm the hypothesis of the study. These differences also confirm what is experienced in the natural environment of a boreal forest. While all leachates were obtained from a boreal forest in Interior Alaska, the boreal forest can be delineated into smaller sub-regions based upon the underlying geology and the differences in soil type and drainage. The differences in soil type and drainage account primarily for the assortment of vegetation attributes in a boreal forest. Spruce trees typically are found in poorly drained sites while deciduous trees are found on more well-drained sites (Johnson et al. 1995). Since trees can out compete other forms of plants in a boreal forest because they can live longer, grow taller, and can tolerate a poor nutrient supply, general trends can be established locally through formation a plant community (Johnson et al. 1995). For example, a plant community consisting of dominant vegetation attributes of aspen and feather moss would be reasonable as well as a plant community consisting of dominant vegetation attributes spruce and lichen. The specific plant community has the largest impact on localized soil leachates. Numerous studies have confirmed that differences exist when comparing DOM between birch and conifers (Suominen et al. 2003; Quideau et al. 2001; Kaiser et al., 2001; Smolander et al. 2002; Hongve et al. 2000). Due the varying soil chemical requirements within the different plant communities, the organic matter will vary and therefore the DOM will also vary. Smolander et al. (2002) found the carbon to nitrogen ratios varied largely when comparing birch stands to pine stands. Among the many reasons differences can be statistically determined between coniferous and deciduous plant communities, is the timing and rate at which the two communities grow and drop leaves or needles. Since conifers can photosynthesize at lower temperatures than deciduous trees, they will have a longer growing season. Deciduous trees shed their leaves in the autumn of the year. A thick forest floor mat is formed with some leaves degraded in the autumn months and some left for degradation over the winter months and spring of the year (Hongve et al. 2000). Conifer litter is much more recalcitrant than deciduous litter. Additionally the difference between the litters may be related to the special pattern of needle senescence and shedding of needles (Hongve et al. 2000).

With the better understanding of the vegetation influences of specific molecular fingerprint compounds believed to serve as vectors for contaminant transport, better decisions regarding non-point source determination within watersheds can be made. In addition, the better understanding of vegetation influences on leachates that can potentially cause membrane fouling during filtration of water can potentially save money and time for operators.

References:

Autier, V. and White, D. 2004. Examination of cadmium sorption characteristics for a boreal soil near Fairbanks, Alaska. Journal of Hazardous Materials. **106B**: 149-155.

Berk, K.N., and Carey, P. Data Analysis with Microsoft® Excel, Updated for Windows XP. Canada: Brooks/Cole, 2004.

Bohn, H.L., McNeal, B.L., and O'Connor, G.A. Soil Chemistry, 2nd edition. NY: John Wiley & Sons, Inc, 1985.

Bryant, F.B. and Yarnold, P.R. "Principal Components Analysis and Exploratory and Confirmatory Factor Analysis." Reading and Understanding Multivariate Statistics. Ed. Laurence G. Grimm and P.R. Yarnold. Washington, D.C., American Psychological Association, 2003. 99-136.

Buffle, J. Complexation Reactions in Aquatic Systems: An analytical approach. Ellis Horwood Limited West Sussex, England. 1990.

Cronan, C.S. and Aiken, G.R. 1985. Chemistry and transport of soluble humic substances in forested watershed of the Adirondack Park, New York. Geochimica et cosmochimica acta. **49**: 1697-1705.

Davies, A.N., Kuckuk, R., Hill, W., Nolte, J., and Burba, P. 2001. A comparison of various pyrolysis experiments for the analysis of reference humic substances. Journal of Analytical and Applied Pyrolysis. **60**: 145-157.

Dean, R.B. and Dixon, W.J. 1951. Simplified statistics for small number of observations. Analytical Chemistry. **23**: 636-638.

Ebensen, K.H. Multivariate Data Analysis – in practice. CAMO Process, Woodbridge, N.J. 2002.

Gilbert, R.O. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold, New York. 1987.

Gray, C.W., McLaren, R.G., Roberts, A.H.C., and Condron, L.M. 1999. Solubility, sorption and desorption of native and added cadmium in relation to properties of soils in New Zealand. European Journal of Soil Science. **50**: 127-137.

Hempfling, R. and Schulten, H.-R. 1990. Chemical characterization of the organic matter in forest soils by Curie point pyrolysis-GC/MS and pyrolysis-field ionization mass spectrometry. Organic Geochemistry. **15**: 131-145.

Hongve, D., van Hees, A.W., and Lundstrom, U.S. 2000. Dissolved components in precipitation water percolated through forest litter. European Journal of Soil Science. **51**: 667-677.

Johnson, D., Kershaw, L., MacKinnon, A., and Pojar, J. Plants of the Western Boreal Forest and Aspen Parkland. Lone Pine Publishing and the Canadian Forest Service, Canada. 1995.

Joly, Helen, A., Li, H., and Belzile, N. 2000. Pyrolysis gas chromatography-mass spectrometry of humic substances extracted from Canadian lake sediments. Canadian Journal of Chemistry. **78**: 51-63.

Kaiser, K., Guggenbergerger, G., Haumaier, L., and Zech, W. 2001. Seasonal variations in the chemical composition of dissolved organic matter in organic forest floor layer leachates of old-growth Scots pine (*Pinus sylvestris* L.) and European beech (*Faus sylvatica* L.) stands in northeastern Bavaria, Germany. Biogeochemistry. **55**: 103-143.

Koopal, L. K., van Riemsdijk, W.H., and Kinniburgh, D.G. 2001. Humic matter and contaminants. General aspects and modeling metal ion binding. Pure Applied Chemistry. **73**: 2005-2016.

Kracht, O. and Gleixner, G. 2000. Isotope analysis of pyrolysis products from Spagnum peat and dissolved organic matter from bog water. Organic Geochemistry. **31**: 645-654.

Lehtonen, K, Hannien, K., and Ketola, M. 2001. Structurally bound lipids in peat humic acids. Organic Geochemistry. **32**: 33-43.

Lichtfouse, E., Chenu, C., Baudin, F., Leblond, C., da Silva, M., Behar, F., Derenne, S., Largeau, C., Wehrung, P., and Albrecht, P. 1998. A novel pathway of soil organic formation by selective preservation of resistant straight-chain biopolymers: chemical and isotope evidence. Organic Geochemistry. **28**: 411-415.

Lumsdon, D.G. 2004. Partitioning of organic carbon, aluminum and cadmium between solid and solution in soils: application of a mineral-humic particle additivity model. European Journal of Soil Science

Martensson, A.M., Aulin, C., Wahlberg, O., and Agren S. 1999. Effect of humic substances on the mobility of toxic metals in a mature landfill. Waste Management Resource **17**: 296-304.

McBean, E.A. and F.A. Rovers. <u>Statistical Procedures for Analysis of Environmental Monitoring Data and Risk Assesment.</u> Prentice Hall: Upper Saddle River, NJ, 1998.

Moore, T. R. “Dissolved organic carbon: Sources, sinks, and fluxes and role in the soil carbon cycle.” <u>Soil Processes and the Carbon Cycle</u>, Ed.: Rattan Lal, John M. Kimble, Ronald F. Follett, and B. A. Stewart. Boca Raton, FL, CRC Press LLC. 1997.

Naar, J., McDonald, N, Garland, D.S., White, D., and Woolard, C. 2001. Nanofilter removal of disinfection by-product precursors. Proceedings of the 6th Research and Development Conference on Rural Sanitation, Anchorage, Alaska. pp. 9-20.

Page, D.W., van Leeuwen, J.A., Spark, K.M., and Mulcahy, D.E. 2002. Pyrolysis characterization of plant, humus and soil extracts from Australian catchments. Journal of Analytical and Applied Pyrolysis. **65**: 269-285.

Parker, J.L., Fernandez, I.J., Rustad, L.E., and Norton, S.A. 2002. Soil organic matter fractions in experimental forested watersheds. Water, Air, and Soil Pollution. **138**: 101-121.

Quideau, S.A., Chadwick, O.A., Trumbore, S.E., Johnson-Maynard, J.L., Graham, R.C., and Anderson, M.A. 2001. Vegetation control on soil organic matter dynamics. Organic Geochemistry. **32**: 247-252.

Rorabacher, D.B. 1991. Statistical treatment for rejection of deviant values: critical values of Dixon's "Q" parameter and related subrange ratios at the 95% confidence interval. Analytical Chemistry. **63**: 139-146.

Schulten, H.-R. 1999. Analytical pyrolysis and computational chemistry of aquatic humic substances and dissolved organic matter. Journal of Analytical and Applied Pyrolysis. **49:** 385-415.

Schulten, H.-R., and Gleixner, G. 1999. Analytical pyrolysis of humic substances and dissolved organic matter in aquatic systems: structure and origin. Water Resources. **33**: 2489-2498.

Schulten, H.-R. and Schnitzer, M. 1992. Structural studies on soil humic acids by Curie-point pyrolysis-gas chromatography/mass spectrometry. Soil Science. **153**: 205-224.

Schwarzenbach, R.P., Gschwend, P.M., and Imboden, D.M. Environmental Organic Chemsitry, 2nd edition, John Wiley & Sons, 2003, Hoboken, New Jersey.

Shapiro, S.S. and Wilk, M.B. 1965. An analysis of variance test for normality (complete samples). Biometrika. **52**: 591-611.

Smeulders, D. E., Wilson, M.A., and Kamali Kannangara, G.S. 2001. Host-guest interactions in humic materials. Organic Geochemistry. **32**: 1357-1371.

Smolander, A. and Kitunen, V. 2002. Soil microbial activities and characteristics of dissolved organic C and N in relation to tree species. Soil Biology & Biochemistry. **34**: 651-660.

Stevenson, F.J. <u>Humus Chemistry: Genesis, Composition, Reactions</u> 2nd Edition, John Wiley and Sons, New York. 1994

Suominen, K., Kitunen, V., and Smolander A. 2003. Characteristics of dissolved organic matter and phenolic compound in forest soils under silver birch (*Betula pendula*), Norway spruce (*Picea abies*) and Scots pine (*Pinus sylvestris*). European Journal of Soil Science. **54**: 287-293.

Tinoco, P., Almendros, G., and Gonzalez-Vila, F.J. 2002. Impact of the vegetation on the lignin pyrolytic signature of soil humic acids from Mediterranean soils. Journal of Analytical and Applied Pyrolysis. **64**: 407-420.

Voegelin, A. and Kretzschmar, R. 2003. Modeling sorption and mobility of cadmium and zinc in soils with scaled exchange coefficients. European Journal of Soil Science. **54**: 387-400.

Wagai, R. and Sollins, P. 2002. Biodegradation and regeneration of water-soluble carbon in a forest soil: leaching column study. Biology and Fertility in Soils. **35**: 18-26.

Weng, L., Temminghoff, E.J.M., Lofts, S., Tipping, E., and van Riemsdijk, W.H., 2002. Complexation with dissolved organic matter and solubility control of heavy metals in a sandy soil. Environmental Science and Technology. **36**: 4804-4810.

White, D.M., and Beyer, L. 1999. Pyrolysis-GC/MS and GC/FID of three Antarctic soils. Journal of Analytical and Applied Pyrolysis. **50**: 63-76.

White, D.M., Yoshikawa, K., and Garland, D. S. 2002. Use of dissolved organic matter to support hydrologic investigations in a permafrost-dominated watershed. Cold Regions Science and Technology. **35**: 27-33.

Yin, Y, Impellitteri, C.A., You, S., and Allen, H.E. 2002. The importance of organic matter distribution and extract soil:solution ratio on the desorption of heavy metal from soils. Science of the Total Environment. **287**: 107-119.

Yoshikawa, K., Bolton, W.R., Romanovkey, V.E., Fukuda, M., and Hinzman, L. (2003). Impacts of wildfire on the permafrost in the boreal forests of Interior Alaska. Journal of Geophysical Research. **108**: FFR 4-1 – 4-14.

Appendix A: Dissolved Organic Matter and Cadmium Adsorption

Study of dissolved organic matter (DOM) is difficult because of its complicated nature. DOM is ubiquitous in all environmental systems and is primarily derived from soils and peat. The molecular structure of DOM is still under an examination. Figure 2 is a recent structural design of a portion of DOM, humic acid (Schwarzenbach et al., 2003). To create this DOM structure, pyrolysis-gas chromatograph/mass spectrometer (py-GS/MS), C^{13}-NMR, and FT-IR among many other experimental techniques were used to examine the molecular structure. As previously mentioned, DOM is a combination of the physical, chemical and biological environment under which the organic matter was derived from. DOM is operationally defined as the natural organic material that passes a glass fiber filter with a pore size <0.45 μm. DOM is a source of cation exchange capacity within soil (Bohn et al., 1985; Schwarzenbach et al., 2003). Additionally, it provides pH-buffering capacity and is a large carbon source (Bohn et al., 1985). In addition, DOM has the ability to *ad*sorb and *ab*sorb ions. The overall surface charge of DOM is typically negatively charged at ambient pH due to carboxylic functional groups. More discussion regarding functional groups will be later in this review.

In recent years, much attention has been drawn to the impact of non-point source heavy metal pollution, within systems with high amounts of DOM, will have on the mobility of the heavy metals. This literature review will highlight the potential mobility of cadmium ($^{2+}$) within a natural system and the partitioning that is available.

Studies of DOM and solution systems including soils, sediments and suspending particles in water bodies is incredibly complicated due to the large degree of heterogeneity of the systems. The solids within the systems are composed of a conglomeration of aggregates of several homologous components including clays, oxides, DOM and microorganisms which are also contributors to the overall characteristics of the system (Buffle 1990).

A factor influencing humic solubility (DOM solubility) includes the humic surface charge (Lumsdon, D.G. 2004). The mechanism for the carboxylic functional group deprotonation is:

$$\text{R-COOH} = \text{R-COO}^- + \text{H}^+$$

where R is most likely an aromatic group. Following deprotonation, the molecule assumes a stretch configuration due to the repulsion of the charged groups (Stevenson, F.J. 1994). Another factor influencing the solubility of DOM is the hydrophobicity that occurs when the molecule is protonated. With the addition of metal ions, the charge is reduced, salt is formed and the molecule collapses (Stevenson, F.J. 1994). Also precipitation increases through the increased oxygen-containing functional groups as the molecule becomes neutralized, the reduction of charge on the humic molecule (Stevenson, F.J. 1994).

Cadmium enters the soil through multiple sources. They include the deposition of sewage sludge from wastewater facilities, agricultural usages as a fertilizer, mine tailings, emissions from industrial facilities, atmospheric deposition and dumps and land fills from which cadmium is released as a leachate (Voegelini et al., 2003; Martensson et al., 1999).

The risks associated with soluble cadmium are typical for all dissolved heavy metals. Cadmium is toxic to plants, animals and humans. It also possesses the ability to be phytoavailable and there are a number of plants available that are used in phytoremediation for cadmium contamination. However, due to non point source contamination or atmospheric deposition, animals grazing on these plants will also uptake cadmium, which then could bioaccumulate within their bodies or be extracted through their milk (Grey et al., 1999).

Cadmium released as a leachate is incredibly hazardous. Because aerobic conditions cause an increase in acidity of the soil/metal-solution system, metals previously bound within the leachate become disassociated. It has been suggested that when previously anaerobic leachates become aerobic, metal bound to organic matter becomes released (Martensson et al., 2003). Therefore, the production of chelating agents, which enhance the mobility of metals with the material and thus will increase mobility of metals which could affect long-term decisions regarding the control of contaminants (Martensson, et al., 2003).

One factor influencing cadmium solubility is pH. Common solubility constants for cadmium range from 2.3×10^{-14} to 7.2×10^{-15}. The dissociation reaction for Cadmium hydroxide is:

$$Cd(OH)_2\ (s) \ {<}{-}{-}{-}{-}{-}{>}\ Cd^{2+}\ (aq) + 2\ OH^-\ (aq)$$

Therefore, as the pH increases, the amount OH^- increases.

There are many factors affecting the molecular adsorption of compounds. They include primarily the chemical character, shape and configuration of the compound. Also the acidity or basicity of the environment, the water solubility, the charge distribution, the polarity of the compound, the size, the organic matter content of the systems, the pH, and even the kind and type clay minerals present will influence the potential for adsorption of a compound (Bohn 1985; Schwarzenbach et al., 2003).

In the natural environment, soluble cadmium mainly complexes with: hydrous ferric oxide and humic substances (DOM). There is a strong affinity for Cd^{2+} to complex with (Fe(OH) and a weak affinity to complex with ($Fe(OH)_2$) (Lunsdon, D.G. 2004).

Adsorption is difficult to quantify and will vary from soil to soil. Adsorption isotherms are commonly used to determine the relationship of solute adsorption by solids at constant temperature and pressure. They show the amount of adsorbate (solute) sorbed as a function of its equilibrium concentration (Bohn (1985)). A variety of isotherms are possible depending on the affinity of the adsorbent (solid) for the adsorbate (solute) (Bohn (1985)).

There are a number of forms a metal ion may take on in a soil. They include water soluble, as the free cation, as organic and inorganic complexes, exchangeable, bound to carbonates or specifically adsorbed, bound to Fe and Mn oxides, organically complexed or held in primary minerals. All these forms are dependent upon the type of metal cation, the pH, the kind and amount of clay (inorganic sorbants) and the organic matter content of the soil. A suggested relationship for exchangeable and soluble copper (Cu^{2+}) can roughly be applied to soluble cadmium (Cd^{2+}) (Stevenson, F.J. 1994).

exchangeable and soluble Cu ↔ specifically adsorbed Cu ↔ organically bound Cu

The adsorption of natural organic matter to mineral surfaces involves the direct exchange at the clay-organic interface (Stevenson, F.J. 1994). The soil humic partitioning is affected by a bimodal distribution of acidic sites. The sites with weak proton affinity are carboxylic and those with stronger proton affinity are phenolic. The pKa value for the carboxylic group is ~5.0 and for the phenolic group is ~9.0.

The water content within the system will affect the way the humic acid (HA) responds to the mineral oxide surfaces. Buffle (1990) compiled a number of experiments outlining the relationship between the variations in water content, specific surface area and pore volume. He found that inner water within the open pore space is not expelled by the adsorption of HA. This suggests that HA does not penetrate between layers of mineral oxides.

When HA is adsorbed on the surface this leads to the elimination of their hydration water. At the same time, occupation of the smallest pores by HA produces a decrease in the corresponding pore volume and specific surface area. Buffle (1990) also found that when the content of HA is high, there is a further increase in the water content, due to hydration of the HA and a decrease in the volume of larger pores which become partially clogged.

When DOM complexes with cadmium it can form both soluble and insoluble complexes with metal ions. DOM with low molecular weight acts as a transporter of metal ions and those with high molecular weights act as a sink for polyvalent cations (Stevenson, F.J. 1994). Metal complexes with humic acid are soluble at low metal-humic acid ratios (few combined molecules in the chain) (Stevenson, F.J. 1994).

The role of pH (and soil type) as an effective regulator of sorption between DOM and Cd^{2+} has been extensively studied (Koopal et al., 2001; Lumsdon, D.G. 2004; Martensson et al., 1999). It has been experimentally shown that organic soils adsorbed much more Cd at about pH 4 in contrast to the two B horizons which adsorbed little at this pH (Lumsdon, D.G. 2004). Additionally, during an experiment where Cd sorption of 15 soils was studied, it was found that the SOM normalized partitioning coefficient of Cd correlated well with SOM content (r^2=0.92) (Yin et al., 2002).

The primary group responsible for this control is the COOH group. As previously mentioned the mechanism is governed by the deprotonation of COOH:

$$R\text{-}COOH = R\text{-}COO^{-} + H^{+}$$

As the pH increases, the surface potential increases and thus competition from protons decreases. The reduction in competition from protons allows for metal binding with the functional groups (Yin et al., 2002). One researcher described the mechanism as that DOM behaves like a weak-acid polyelectrolyte in which the ionization of COOH groups are controlled by pH which then affects their ability to bind to metal ions (Lumsdon, D.G. 2004).

On average for most soils, pH of above and about 4.5 leads to hydrolysis of the metal ion with the formation of oxide hydrates (Stevenson, F.J. 1994). Hydrolysis reactions involve the formation of monomeric species and polymers of the metal ions.

Once the complex between DOM and cadmium is formed, it shields the bound cations from hydrolysis and precipitation reactions. This complexation then reduces the overall availability of the soluble cadmium within the soil system.

Those soil systems with more DOM than mineral surfaces will most likely have more sorbed cadmium. Additionally, the overall cation toxicities inherent in Cd^{2+} decrease in more acidic soils (Bohn et al., 1985). This further supports the suggestion that the Cd^{2+} becomes bound with the DOM and is no longer a free ion in solution. Lumsdon, D.G. (2004) found that as the surface area increased the affinity of DOM for Cd^{2+} also increased.

When the system is dominated by DOM complexed cadmium species, the pH effects on the solid-solution partitioning will be less obvious than when the metal is present as the free ion (Weng et al., 2002). Also, when DOM is complexed with the metal, the solubility of DOM is increased a considerable degree. Weng et al. (2002) found the concentration of dissolved metals to increase by more than 2 orders of magnitude when complexed with DOM.

Sorption may occur between oxides and DOM or between DOM and cadmium or other combinations. One such combination is one in which DOM binds with the oxide surface and cadmium binds to the DOM or the cadmium may bind with the oxide and then the DOM binds with the cadmium. This forms an oxide-DOM-cadmium complex. Stevenson, F. J. (1994) determined that polyvalent cations might serve as linkages between humic substances and clay minerals, thereby affecting the physical properties of soils.

Some proposed mechanisms for this type of sorption are: the formation of chain-like structures through metal-ion bridges, the formation of hydroxyl complexes of the metal ion at high pH values, or the attachment to clay particles at dioxide surfaces, such as through metal-ion linkages. The bulk of the organic matter in most is soils bound to clay minerals, probably through linkages with Fe, Al, and other polyvalent cations (Stevenson, F.J. 1985). When DOM, cadmium and metal oxides combine in this way, this ligand will likely influence the adsorptive properties of the adsorption site associated with the other available mineral surfaces (Lumsdon, D.G. 2004). Grey (1999) found that there was no significance between Fe and Al oxide because it was suggested that they were already masked by organic matter coatings.

There are a number of different models that rely on theoretical and experimental data and are utilized to predict the sorption capacity of the solute and sorbent. The linear additivity model attempts to account for several independently action sorbent phases specific to each application of the model, Langmuir and Freundlich models are commonly applied.

Within natural organic matter there is a theoretical maximum of the amount of free metal ions that can bind to DOM. This is approximately equal to the content of acidic functional groups, primarily COOH (Stevenson, F.J. 1994). It is also dependent upon pH, kind and amount of acidic functional groups, ionic strength and the molecular weight of the molecules involved.

Often experimental values are obtained in the form of Kd values which is the distribution coefficient expressing the ratio of sorbed Cd to solution Cd at equilibrium. Kd values range due of the type and strength of complexes being formed. Numerous studies compiled in Buffle (1990) support a range between less than 0.1 second to greater than 72 hours for half-dissociation times. Labile complexes dissociate the most rapidly, but those more strongly bound in environmental conditions have pH greater than 7 and the metal concentration to humic acid less than 10^{-2} mmole /g of carbon.

The most commonly applied model is the linear additivity model. This model is based on the assumption that the different adsorbing phases behave independently, and there is no interaction between them that may alter their reactivity (Lumsdon, D.G. 2004). This model is used to predict metal sorption to soils from the composition of their solid phases (Voegelin et al., 2003). A valuable aspect to this model is its ability to be adapted to specific combinations of available information. Constants are applied to combat limited information and curve fitting. The limitation is that it requires the quantification of all soil components. This includes the reactive surface areas, which is very tedious and often only estimated (Voegelin et al., 2003).

Langmuir and Freundlich models are limited applications due to their inability to account for charge balances within the system. The Freundlich model also assumes that the decrease in energy of adsorption with increasing surface coverage is due to surface heterogeneity (Bohn 1985).

Statistical correlations and multivariable approaches are becoming more common as statistical programs become more easy to use. The linear regression approach is a multivariate statistical model. Results are evaluated based on large correlations and small standard errors used as criteria for best fit (Gray et al., 1999). Linear regression allows for the input of the many soil characteristic variables within soil complexation experiments. The largest limitation is that it is a statistical program, which allows for subjective analysis. Additionally, there is no across the board criterion for variable correlations. Grey (1999) determined that organic matter is critical in controlling metal solubility; however, experiments in soil often fail to reveal strong correlations between organic matter and metal sorption capacity. Statistically, organic matter could only explain 24% (of 100%) of variation of sorbed Cd in the soils. The other variables determined to be important were organic matter content and cadmium concentration in the system.

The cationic exchange capacity approach is an outdated approach since the model only includes theoretical values for the input parameters. The input parameters include the surface complexation constants and the total number of adsorption sites. Another limitation is that it ignores the concentration of DOC and the possibility of complexation between Cd and DOC (Lumsdon, D.G. 2004).

Complexation between cadmium and DOM is complicated and the discipline has made large strides in the last decade, however much work still needs to be done. Primarily, theoretical and experimentally based modelers need to combine efforts to complete a reliable prediction model for the complexation of DOM molecules. Two models are getting more attention because of their innovative approaches.

The multi-site and multilayer adsorption applies the experimentally encountered coating oxides with DOM. The model assumes the multiple sites exhibit their own binding strength and adsorption maximum (Bohn et al., 1985). This model agrees with the conceptual idea that the type of sorption occurring may suggest that multiple mechanisms of adsorption are also occurring on similar sites.

The second model is the nonideal competitive adsorption model (NICA). This model includes parameters for heterogeneous substrates and a variety of affinities for the multiple sites encountered within complexation reactions. This model is coupled with electrostatic interactions, encompassing not only intrinsic affinities for certain sites but also electrostatic interactions. Overall, this model is a thermodynamically consistent competitive binding model for heterogeneous systems and a component-specific binding stoichiometry (Koopal et al., 2001).

Natural organic matter is a known transport vessel for soluble cadmium through soil systems. Prediction models incorporate a variety of theoretical and experimental parameters to predict the movement of cadmium through the soil systems. Complexation with oxides or DOM depends primarily on three major variables: the pH, organic matter content and overall cadmium concentration. With the complexation chemistry better understood, the next step is to better understand the link between vegetation and its potential to leach compounds complexed with Cd^{2+}.

Appendix B: Histograms

Figures B-1 – B-14 contain histograms for each of fourteen molecular fingerprint compounds. They were computed with the statistical software package Minitab™.

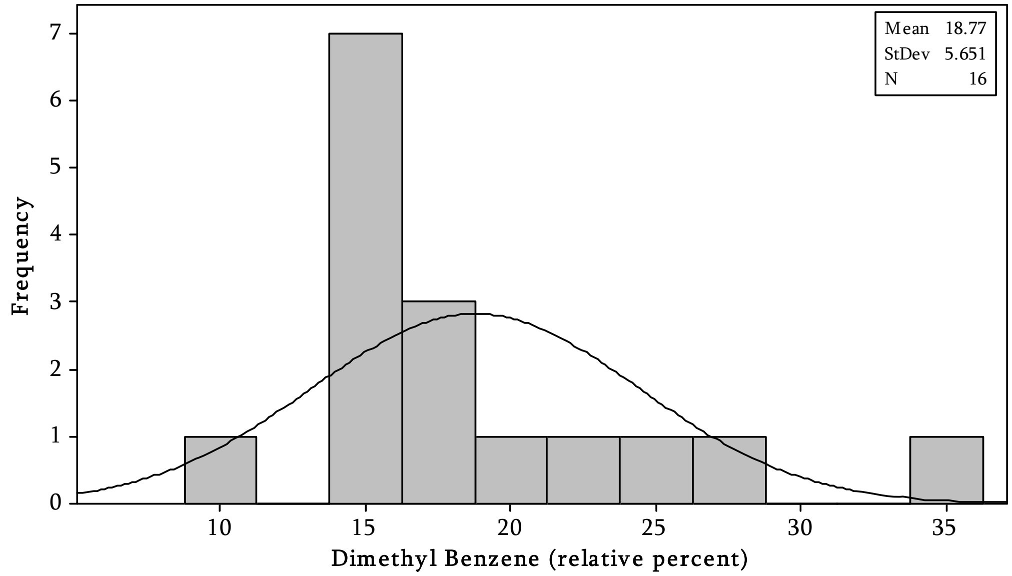

Figure B-1: Histogram (with Normal Curve) of Dimethyl Benzene.

Figure B-1: Histogram (with Normal Curve) of Dimethyl Benzene. This histogram with 16 observations shows the distribution of the relative percentages of molecular fingerprint compound dimethyl benzene to be moderately skewed to the left with a mean of 18.77 and a standard deviation of 5.651.

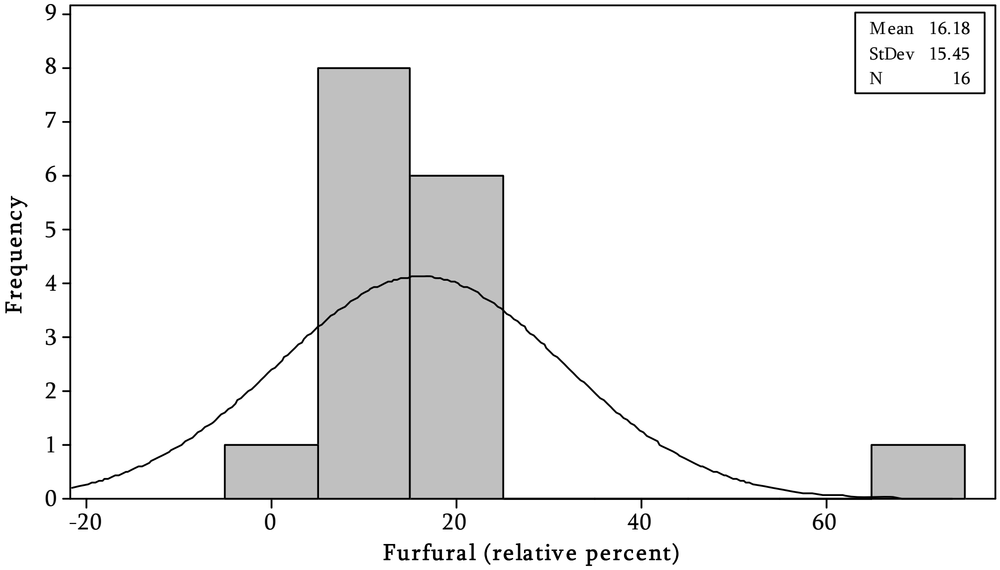

Figure B-2: Histogram (with Normal Curve) of Furfural.

Figure B-2: Histogram (with Normal Curve) of Furfural. This histogram with 16 observations shows the relative percentages of molecular fingerprint compound furfural to be moderately skewed to the left with a mean of 16.18 and a standard deviation of 15.45.

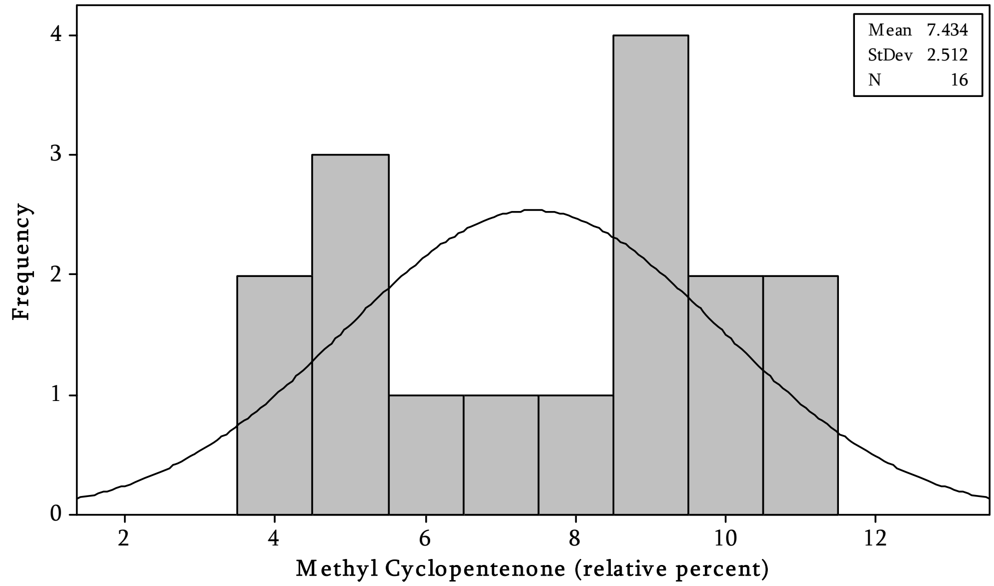

Figure B-3: Histogram (with Normal Curve) of Methyl Cyclopentenone.

Figure B-3: Histogram (with Normal Curve) of Methyl Cyclopentenone. This histogram with 16 observations shows the relative percentage of molecular fingerprint compound methyl cyclopentenone to have a bimodal distribution with a mean of 7.434 and a standard deviation of 2.512.

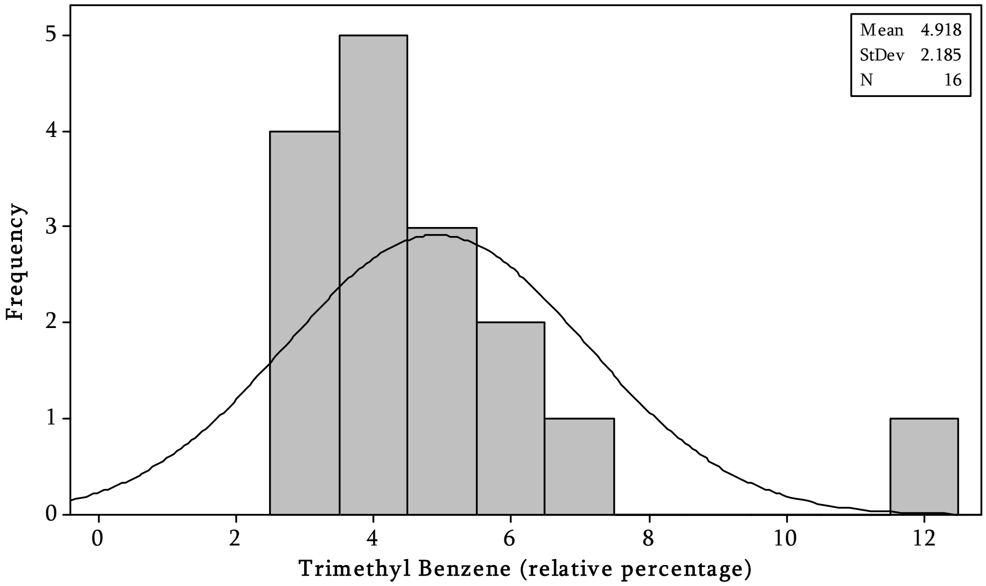

Figure B-4: Histogram (with Normal Curve) of Trimethyl Benzene.

Figure B-4: Histogram (with Normal Curve) of Trimethyl Benzene. This histogram with 16 observations shows the relative percentage of molecular fingerprint compound trimethyl benzene and the distribution to be left skewed with a mean of 4.918 and a standard deviation 2.185.

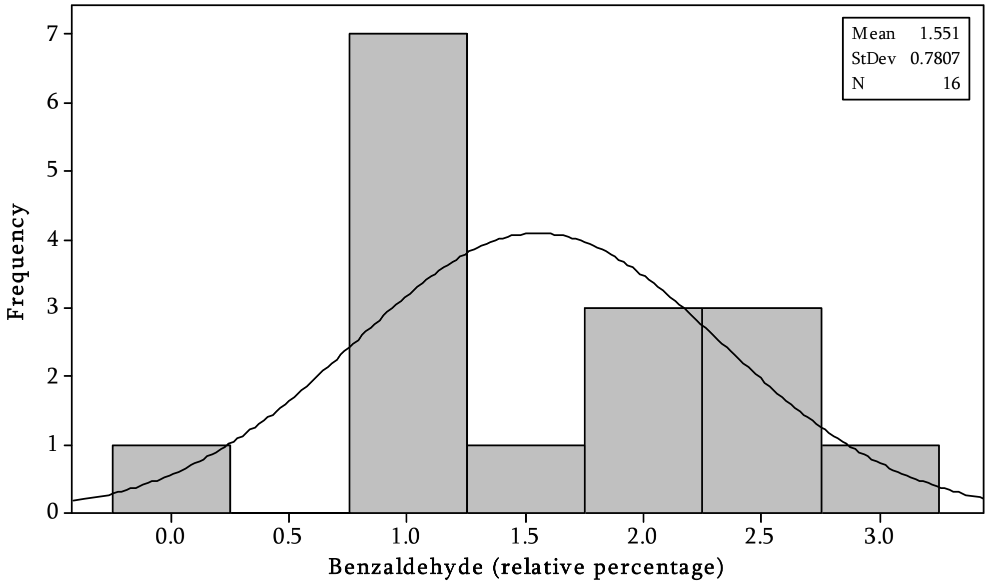

Figure B-5: Histogram (with Normal Curve) of Benzaldehyde.

Figure B-5: Histogram (with Normal Curve) of Benzaldehyde. This histogram with 16 observations shows the distribution of the relative percentages of molecular fingerprint compound benzaldehyde with a mean of 1.551 and a standard deviation of 0.7807.

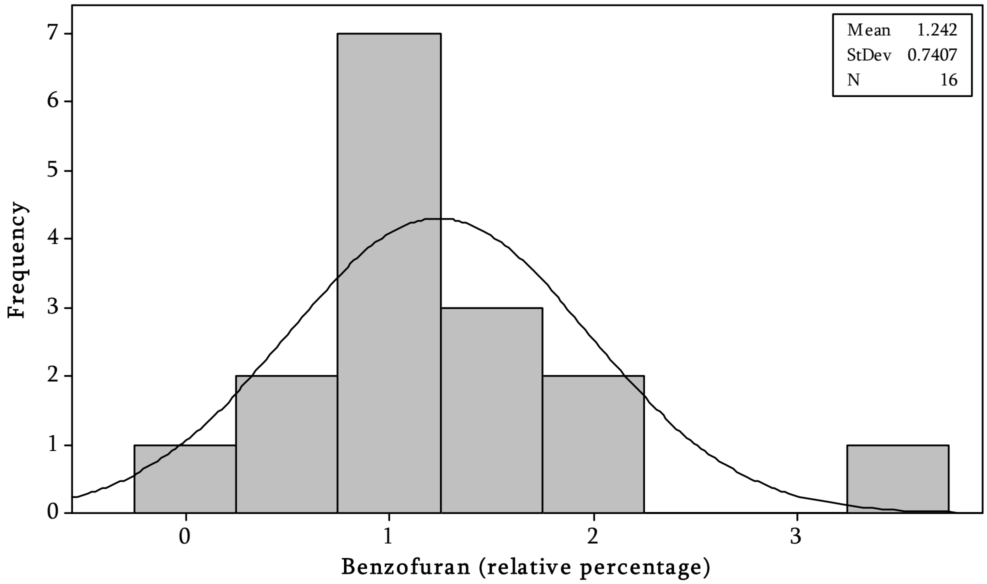

Figure B-6: Histogram (with Normal Curve) of Benzofuran.

Figure B-6: Histogram (with Normal Curve) of Benzofuran. This histogram with 16 observations shows the distribution of the relative percentages of molecular fingerprint compound benzofuran with a mean of 1.242 and a standard deviation of 0.7407.

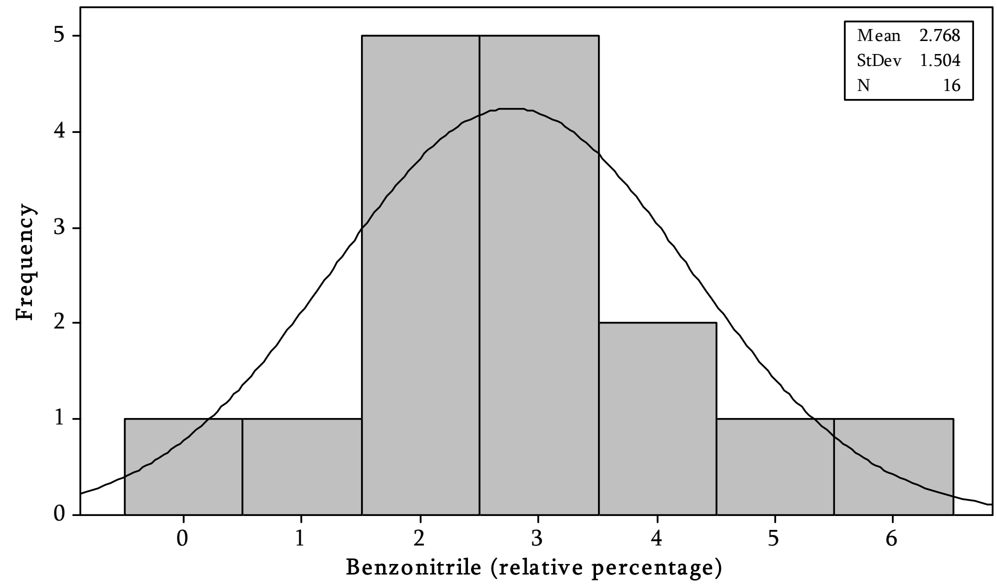

Figure B-7: Histogram (with Normal Curve) of Benzonitrile.

Figure B-7: Histogram (with Normal Curve) of Benzonitrile. This histogram with 16 observations shows the normal distribution of the relative percentages of molecular fingerprint compound benzonitrile with a mean of 2.768 and a standard deviation of 1.504.

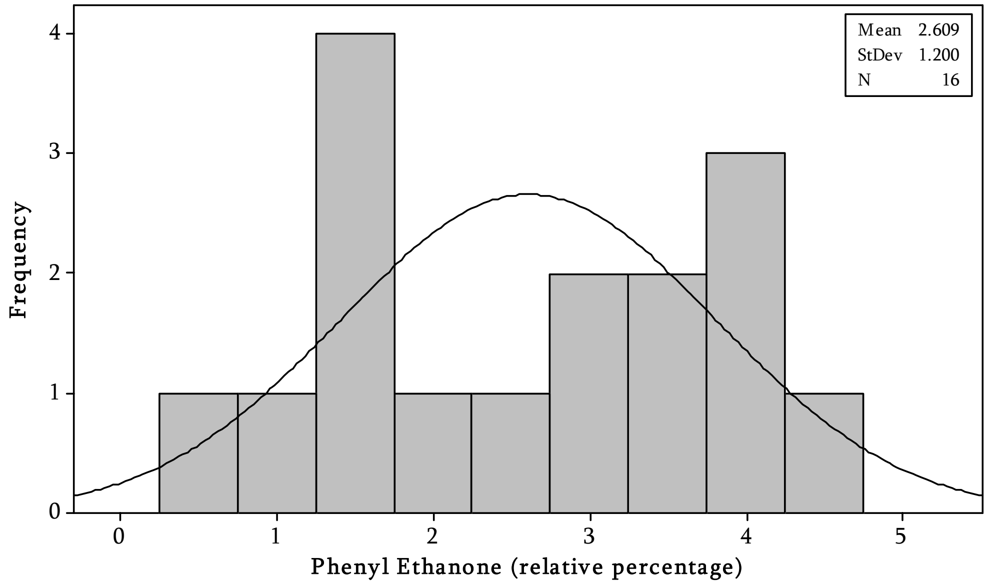

Figure B-8: Histogram (with Normal Curve) of Phenyl Ethanone.

Figure B-8: Histogram (with Normal Curve) of Phenyl Ethanone. This histogram with 16 observations shows the distribution of the relative percentages of molecular fingerprint compound phenyl ethanone with a mean of 2.609 and a standard deviation of 1.200.

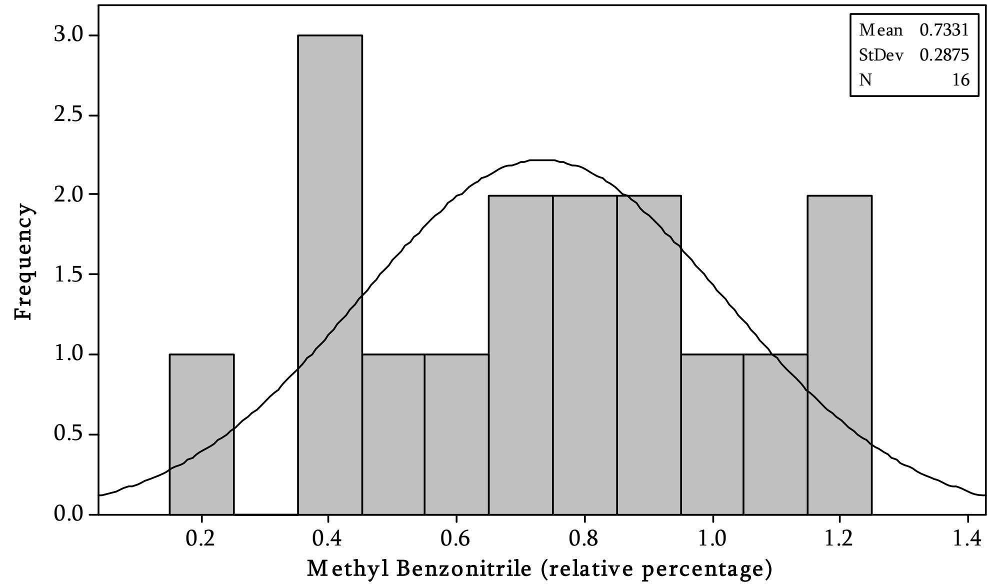

Figure B-9: Histogram (with Normal Curve) of Methyl Benzonitrile.

Figure B-9: Histogram (with Normal Curve) of Methyl Benzonitrile. This histogram with 16 observations shows the distribution of the relative percentages of molecular fingerprint compound methyl benzonitrile with a mean of 0.7331 and a standard deviation of 0.2875.

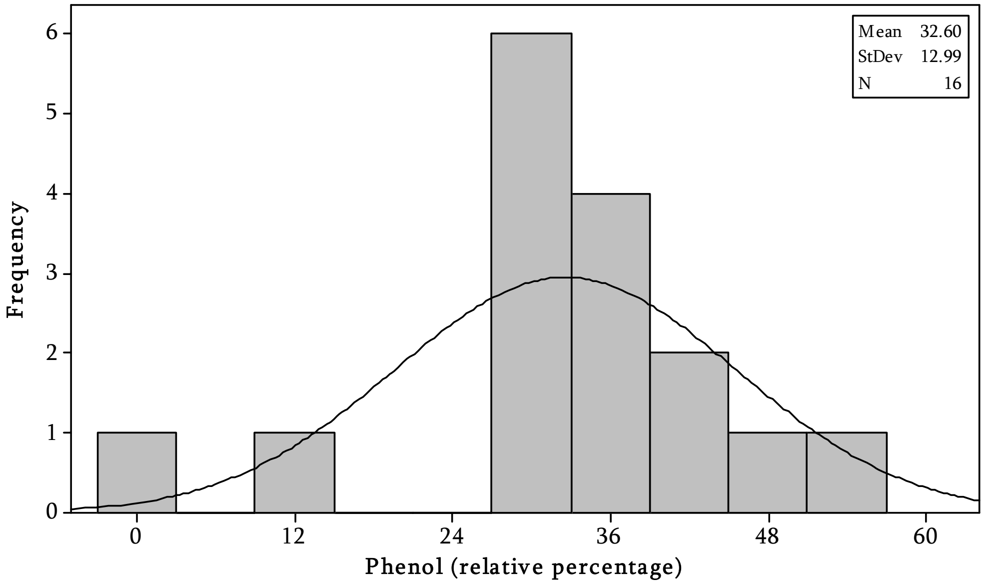

Figure B-10: Histogram (with Normal Curve) of Phenol.

Figure B-10: Histogram (with Normal Curve) of Phenol. This histogram with 16 observations shows the distribution of the relative percentages of molecular fingerprint compound phenol with a mean of 32.60 and a standard deviation of 12.99.

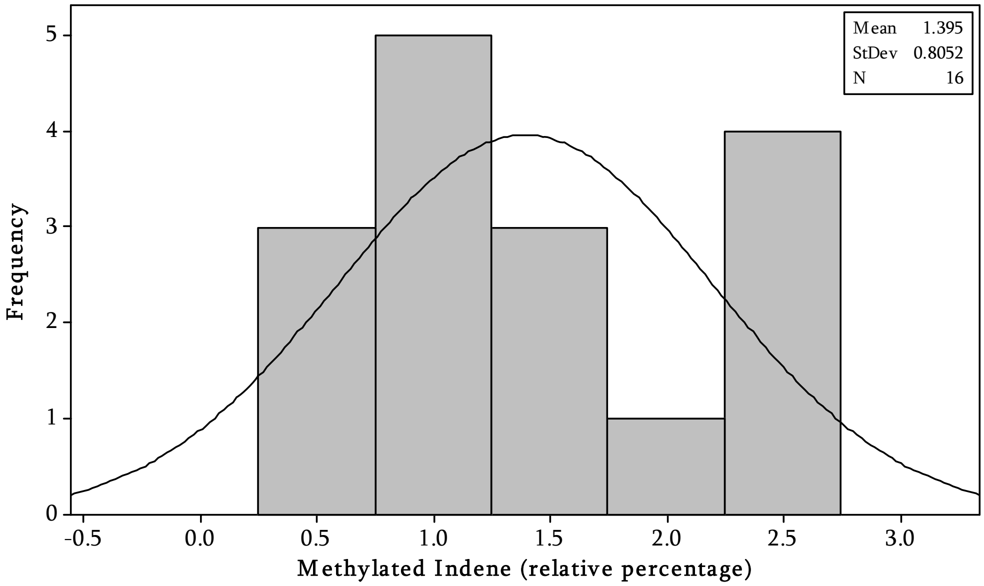

Figure B-11: Histogram (with Normal Curve) of Methylated Indene.

Figure B-11: Histogram (with Normal Curve) of Methylated Indene. This histogram with 16 observations shows the distribution of the relative percentages of molecular fingerprint compound methylated indene with a mean of 1.395 and a standard deviation of 0.8052.

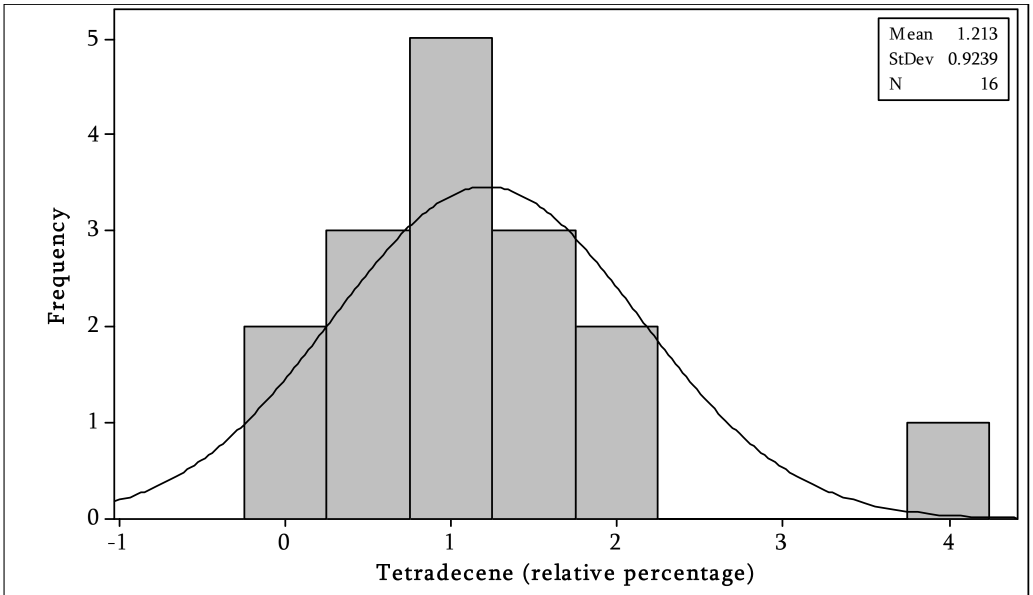

Figure B-12: Histogram (with Normal Curve) of Tetradecene.

Figure B-12: Histogram (with Normal Curve) of Tetradecene. This histogram with 16 observations shows the distribution of the relative percentages of molecular fingerprint compound tetradecene with a mean of 1.213 and a standard deviation of 0.9239.

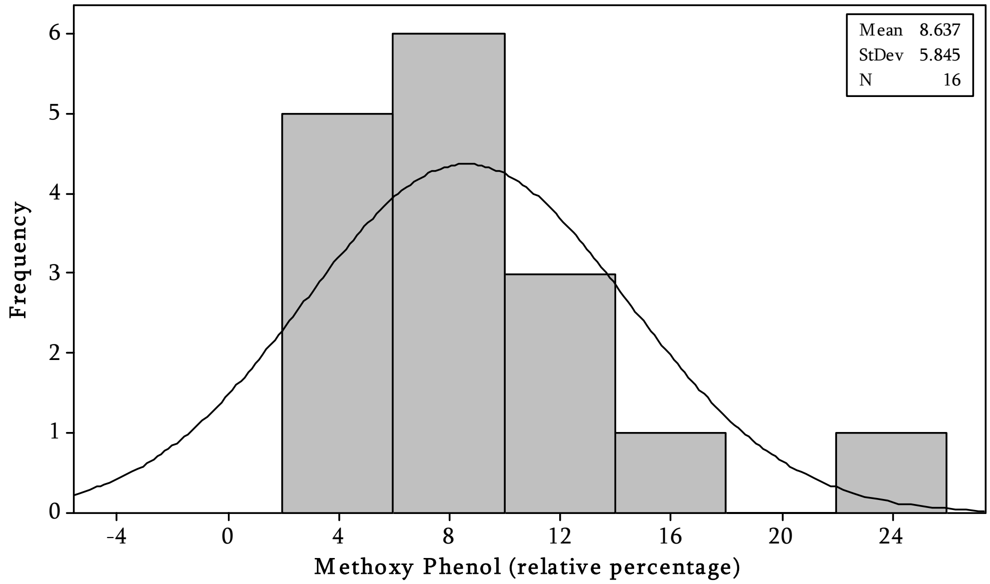

Figure B-13: Histogram (with Normal Curve) of Methoxy Phenol.

Figure B-13: Histogram (with Normal Curve) of Methoxy Phenol. This histogram with 16 observations shows the distribution of the relative percentages of molecular fingerprint compound methoxy phenol with a mean of 8.637 and a standard deviation of 5.845.

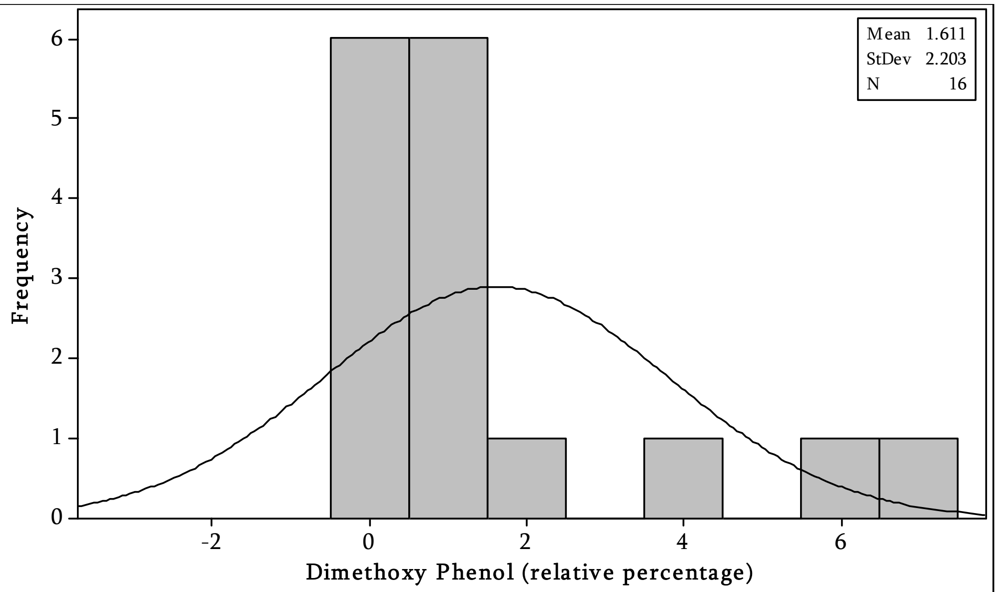

Figure B-14: Histogram (with Normal Curve) of Dimethoxy Phenol.

Figure B-14: Histogram (with Normal Curve) of Dimethoxy Phenol. This histogram with 16 observations shows the distribution of the relative percentages of molecular fingerprint compound dimethoxy phenol with a mean of 1.611 and a standard deviation of 2.203.

Appendix C: Probability Plots

Figures C-15 – C-28 contain the probability plots for each of fourteen molecular fingerprint compounds. They were computed with the statistical software package Minitab™.

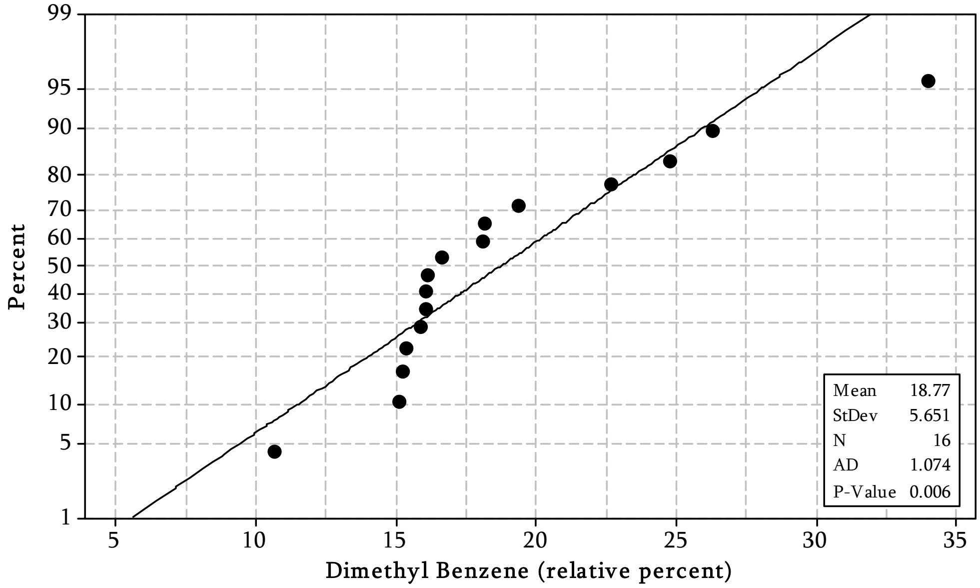

Figure C-15: Probability Plot of Dimethyl Benzene.

Figure C-15: Probability Plot of Dimethyl Benzene. This probability plot shows the 16 observations for molecular fingerprint compound dimethyl benzene and is considered to be positively skewed.

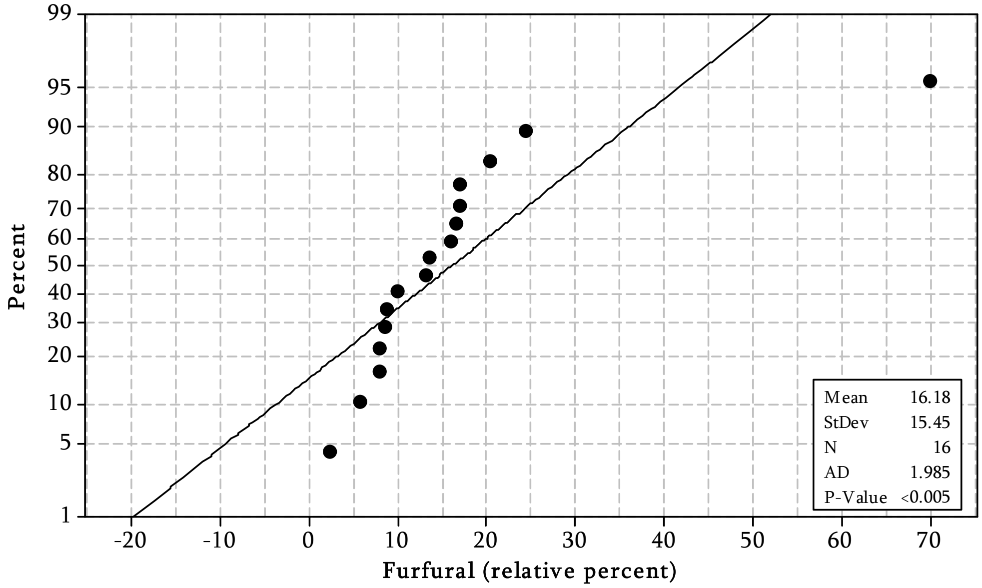

Figure C-16: Probability Plot of Furfural.

Figure C-16: Probability Plot of Furfural. This probability plot shows the 16 observations for molecular fingerprint compound furfural and is considered to be positively skewed.

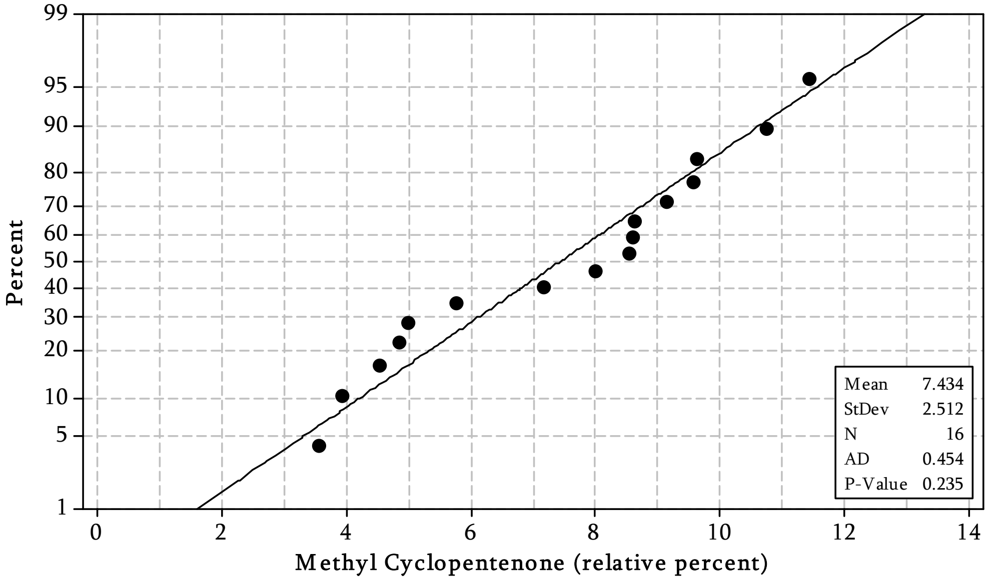

Figure C-17: Probability Plot of Methyl Cyclopentenone.

Figure C-17: Probability Plot of Methyl Cyclopentenone. This probability plot shows the 16 observations for molecular fingerprint compound methyl cyclopentenone and is considered to have normal distribution.

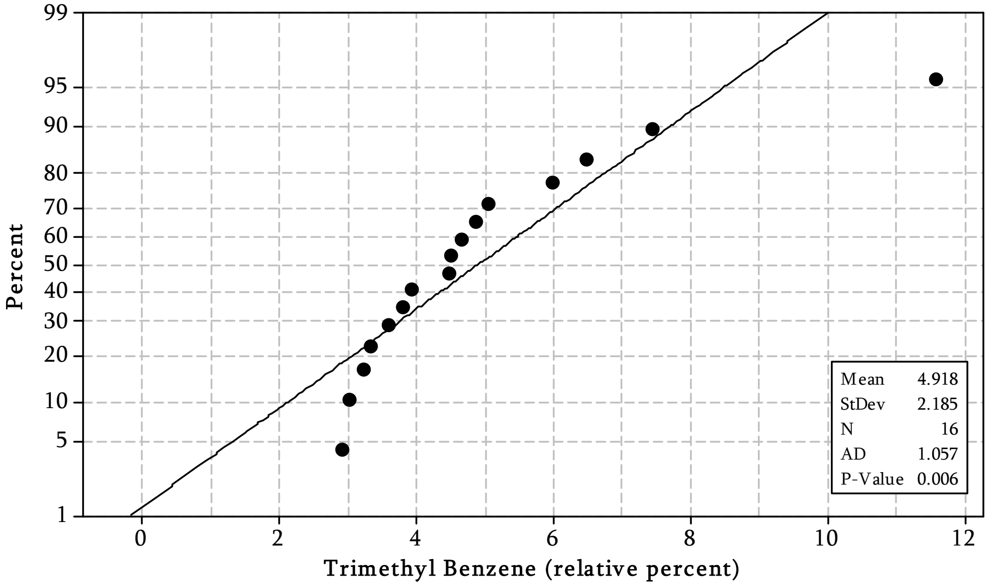

Figure C-18: Probability Plot of Trimethyl Benzene.

Figure C-18: Probability Plot of Trimethyl Benzene. This probability plot shows the 16 observations for molecular fingerprint compound trimethyl benzene.

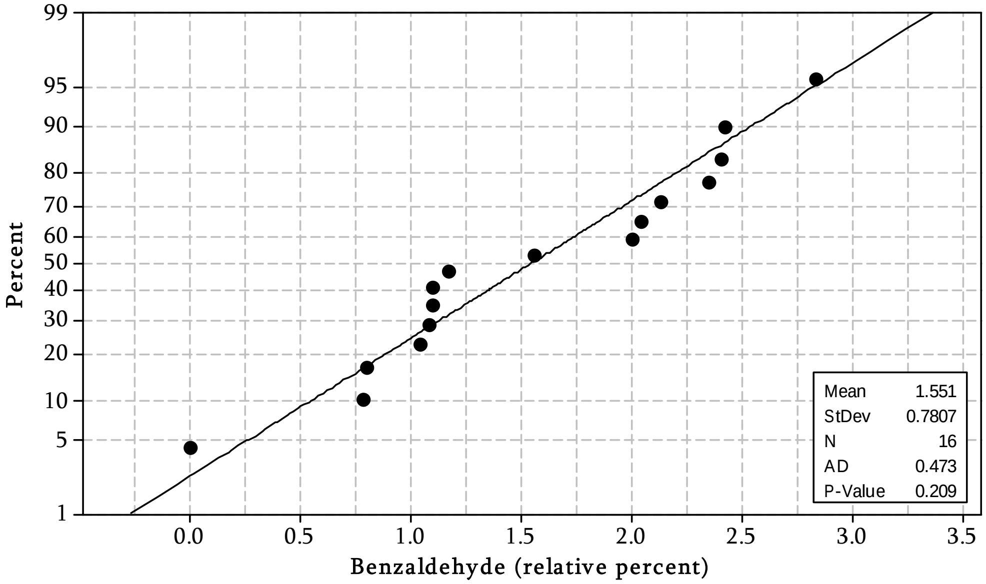

Figure C-19: Probability Plot of Benzaldehyde.

Figure C-19: Probability Plot of Benzaldehyde. This probability plot shows the 16 observations for molecular fingerprint compound benzaldehyde and is considered to be normally distributed.

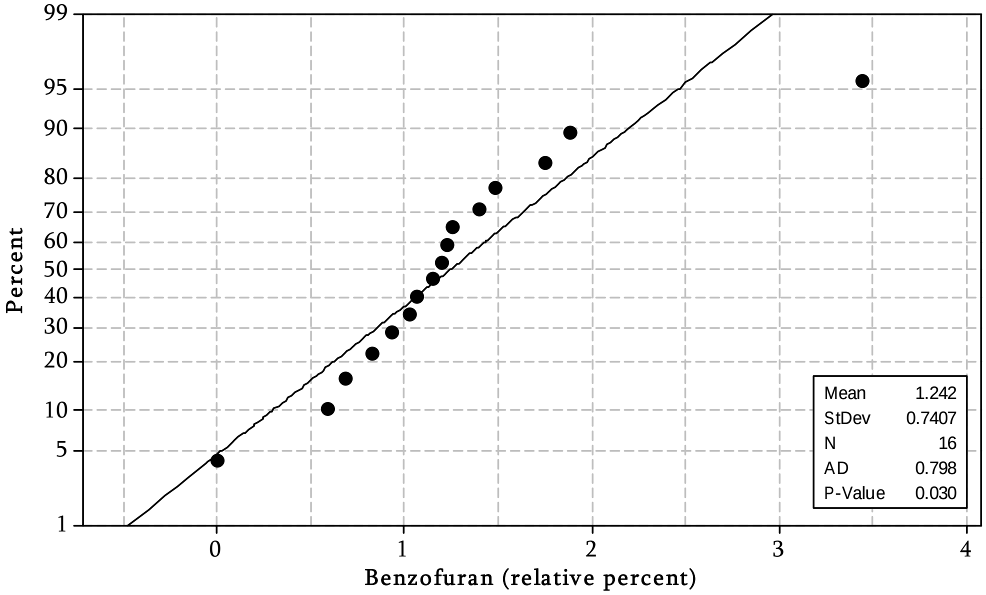

Figure C-20: Probability Plot of Benzofuran.

Figure C-20: Probability Plot of Benzofuran. This probability plot shows the 16 observations for molecular fingerprint compound benzofuran.

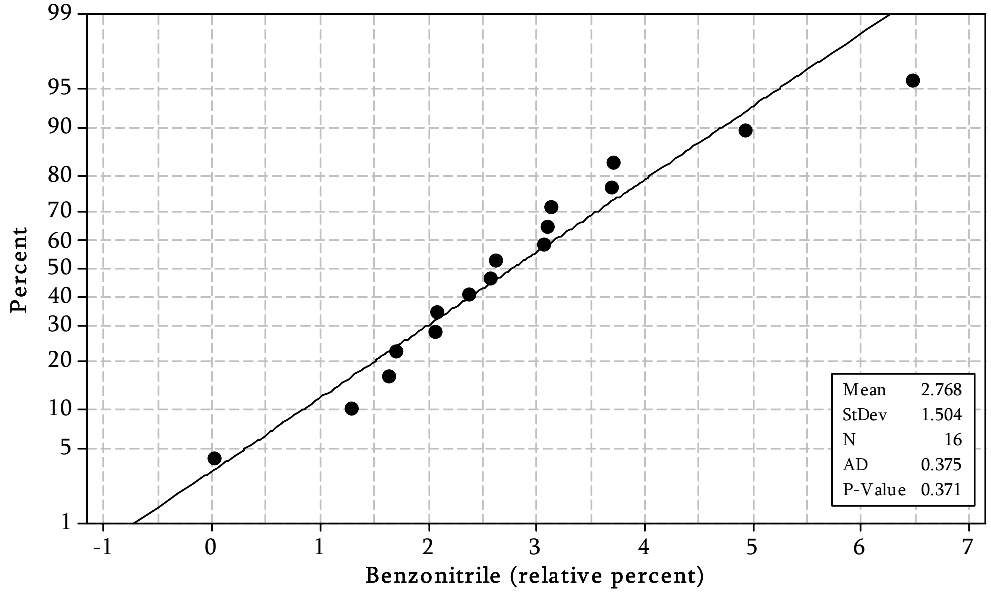

Figure C-21: Probability Plot of Benzonitrile.

Figure C-21: Probability Plot of Benzonitrile. This probability plot shows the 16 observations for molecular fingerprint compound benzonitrile and is considered to be normally distributed.

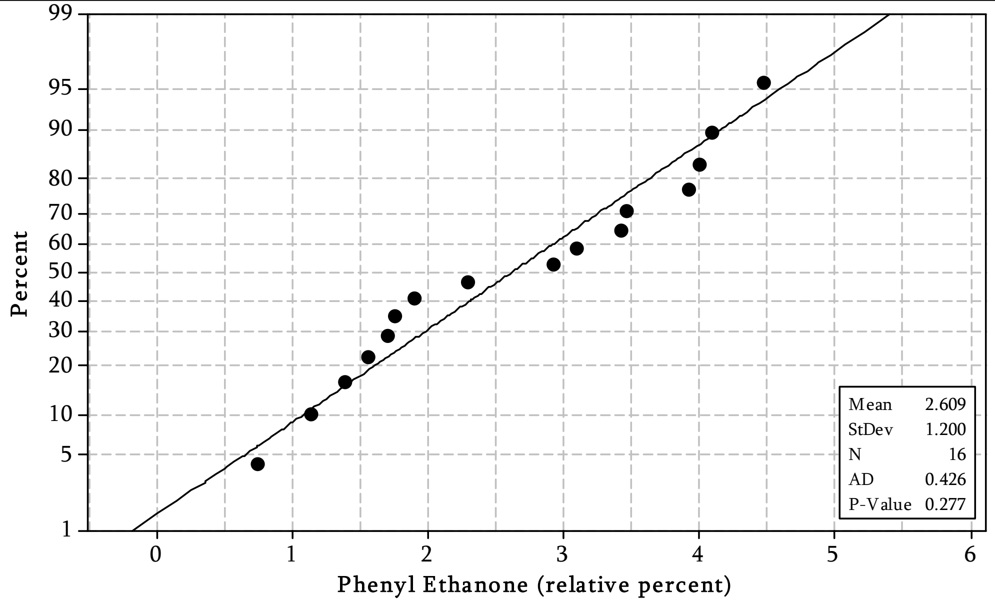

Figure C-22: Probability Plot of Phenyl Ethanone.

Figure C-22: Probability Plot of Phenyl Ethanone. This probability plot shows the 16 observations for molecular fingerprint compound phenyl ethanone.

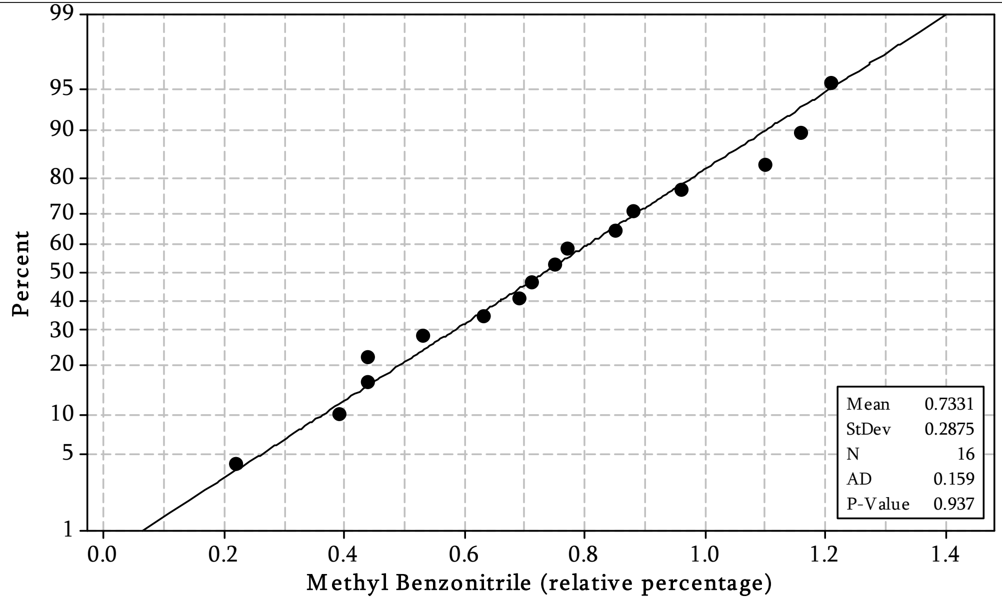

Figure C-23: Probability Plot of Methyl Benzonitrile.

Figure C-23: Probability Plot of Methyl Benzonitrile. This probability plot shows the 16 observations for molecular fingerprint compound methyl benzonitrile and is considered to have normal distribution.

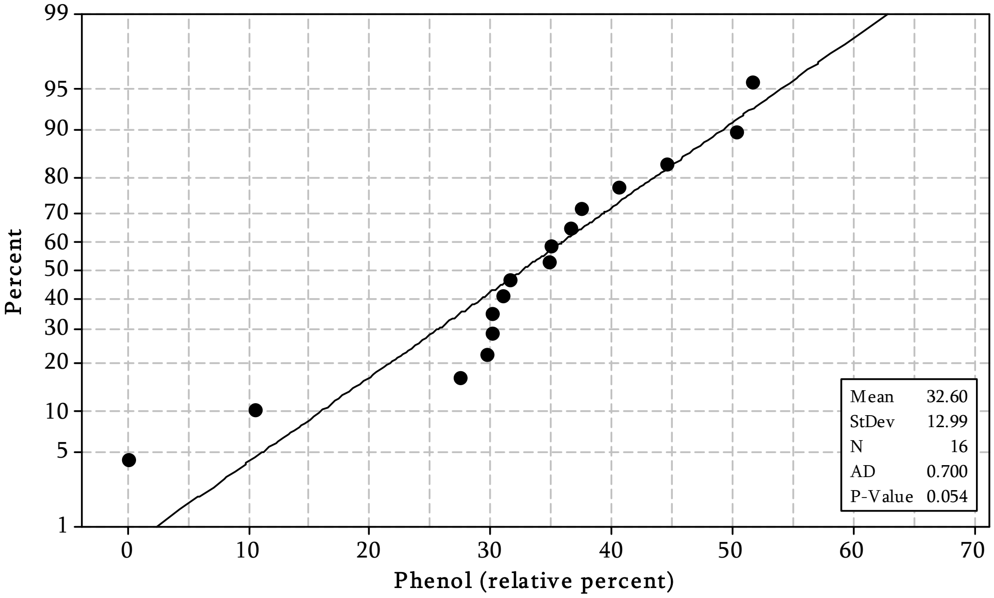

Figure C-24: Probability Plot of Phenol.

Figure C-24: Probability Plot of Phenol. This probability plot shows the 16 observations for molecular fingerprint compound phenol.

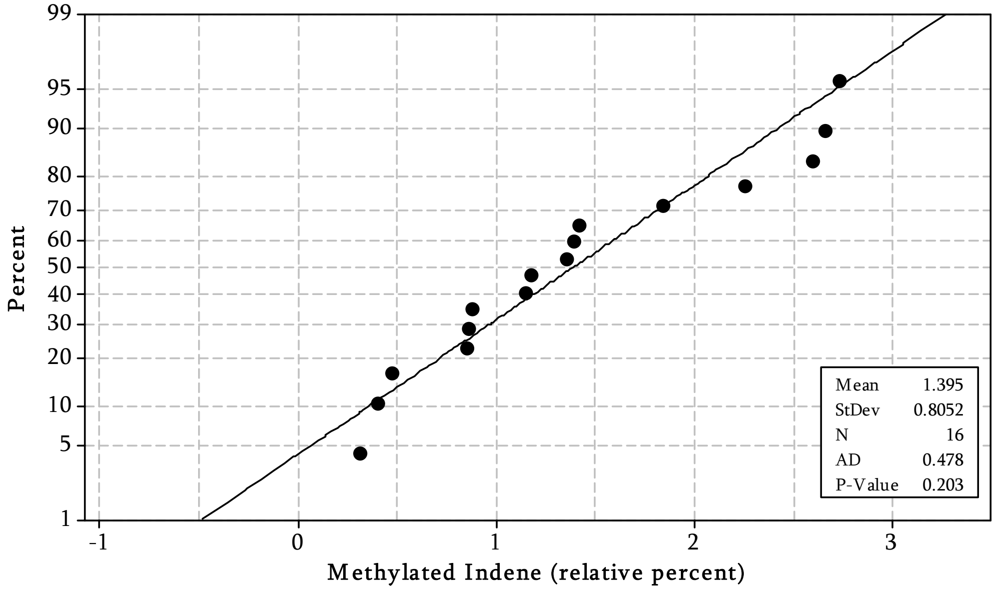

Figure C-25: Probability Plot of Methylated Indene.

Figure C-25: Probability Plot of Methylated Indene. This probability plot shows the 16 observations for molecular fingerprint compound methylated indene.

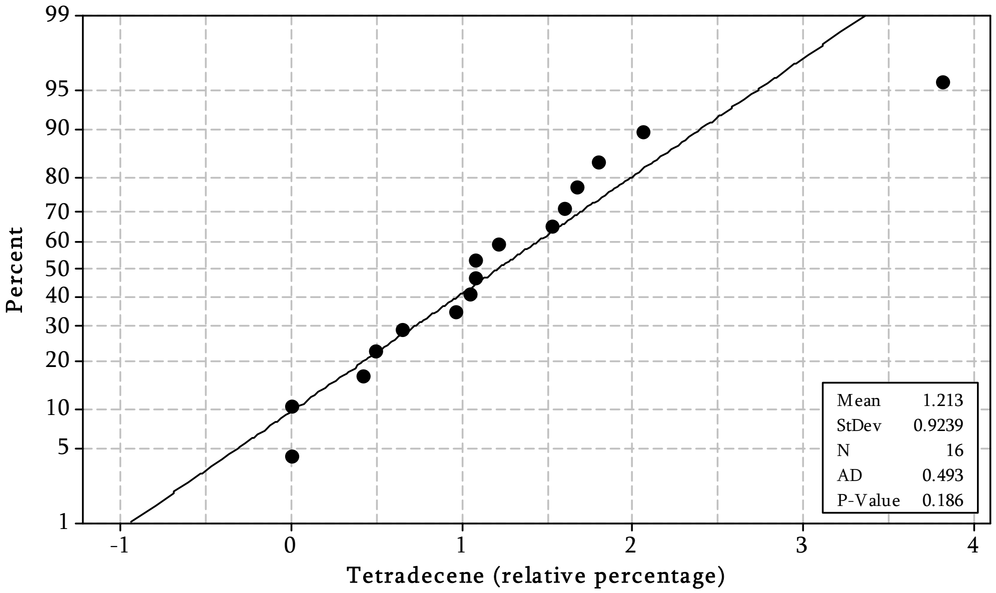

Figure C-26: Probability Plot of Tetradecene.

Figure C-26: Probability Plot of Tetradecene. This probability plot shows the 16 observations for molecular fingerprint compound tetradecene and is considered to have normal distribution.

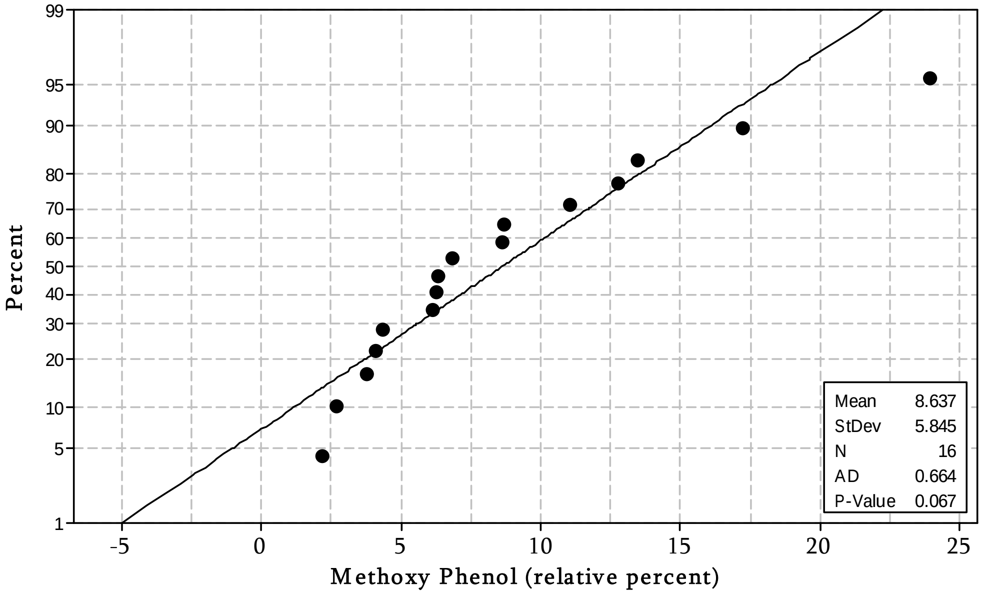

Figure C-27: Probability Plot of Methoxy Phenol.

Figure C-27: Probability Plot of Methoxy Phenol. This probability plot shows the 16 observations for molecular fingerprint compound methoxy phenol.

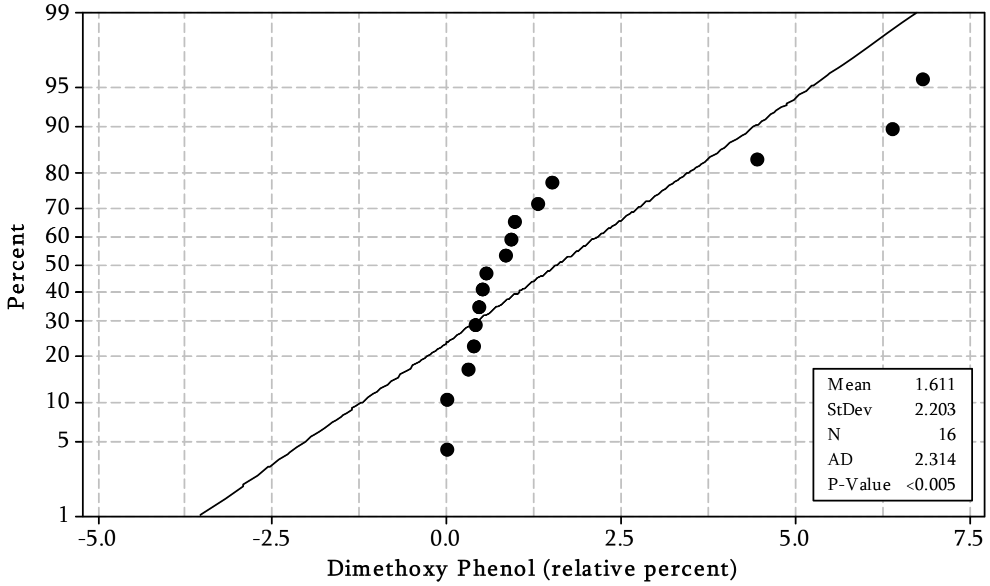

Figure C-28: Probability Plot of Dimethoxy Phenol.

Figure C-28: Probability Plot of Dimethoxy Phenol. This probability plot shows the 16 observations for molecular fingerprint compound dimethoxy phenol and is positively skewed.

Appendix D: Partial Least Squares Regression Plots

Figures D-29 – D-70 contains the partial least square regression (PLS1) loading plot and predicted vs. measured plots for all fourteen molecular fingerprint compounds.

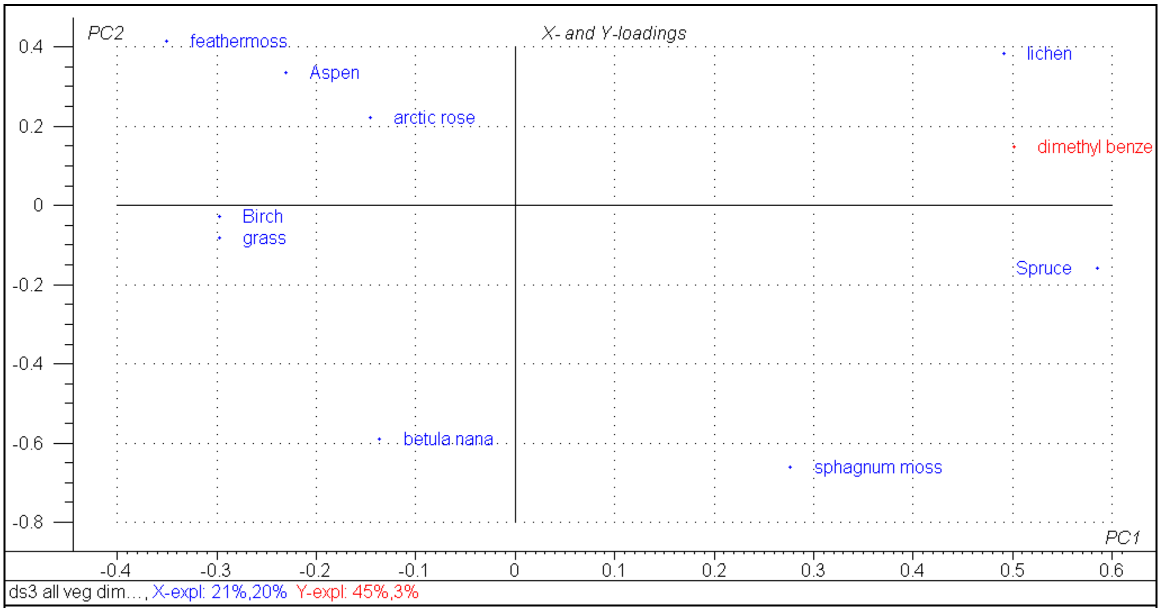

Figure D-29: PLS1 Loading Plot for Dimethyl Benzene.

Figure D-29: PLS1 Loading Plot for Dimethyl Benzene. The input for this plot included all vegetation attributes and dimethyl benzene as only the molecular fingerprint compound. This loading plot explains 41 % of the total variance in the x-direction and 48 % of the total variance in the y-direction.

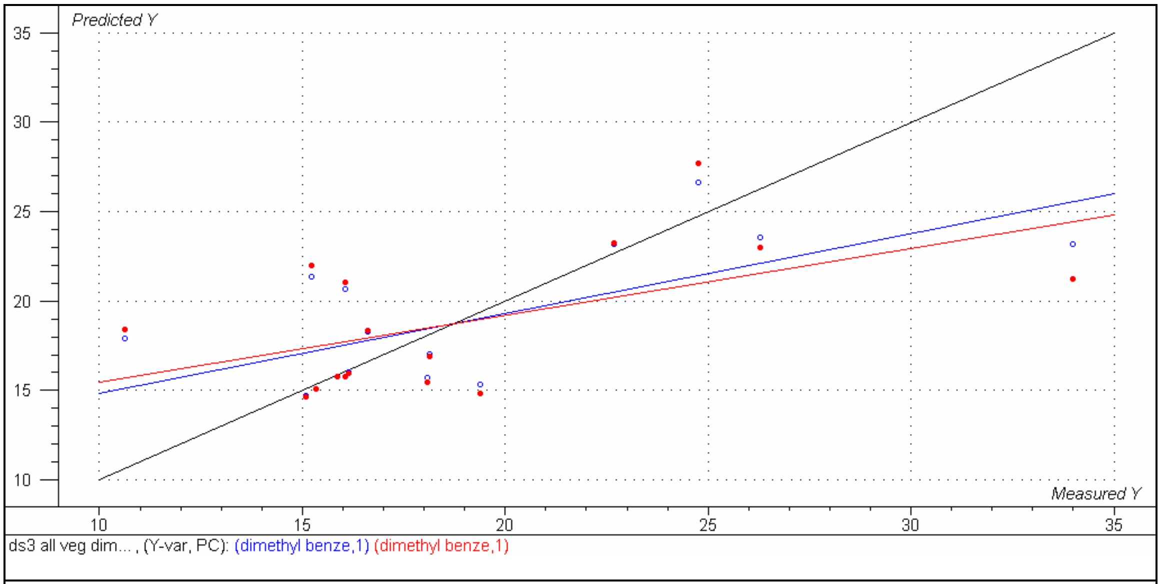

Figure D-30: PLS1 Predicted vs. Measured for Dimethyl Benzene.

Figure D-30: PLS1 Predicted vs. Measured for Dimethyl Benzene. The x-axis contains the measured relative percentages for molecular fingerprint compound dimethyl benzene for all soil leachates and the y-axis contains the predicted relative percentages for molecular fingerprint compound dimethyl benzene. The black trend line represents the target line. The blue trend line represents the measured relative percentages (r^2= 0.668) and the red trend line represents the predicted relative percentages (r^2= 0.546).

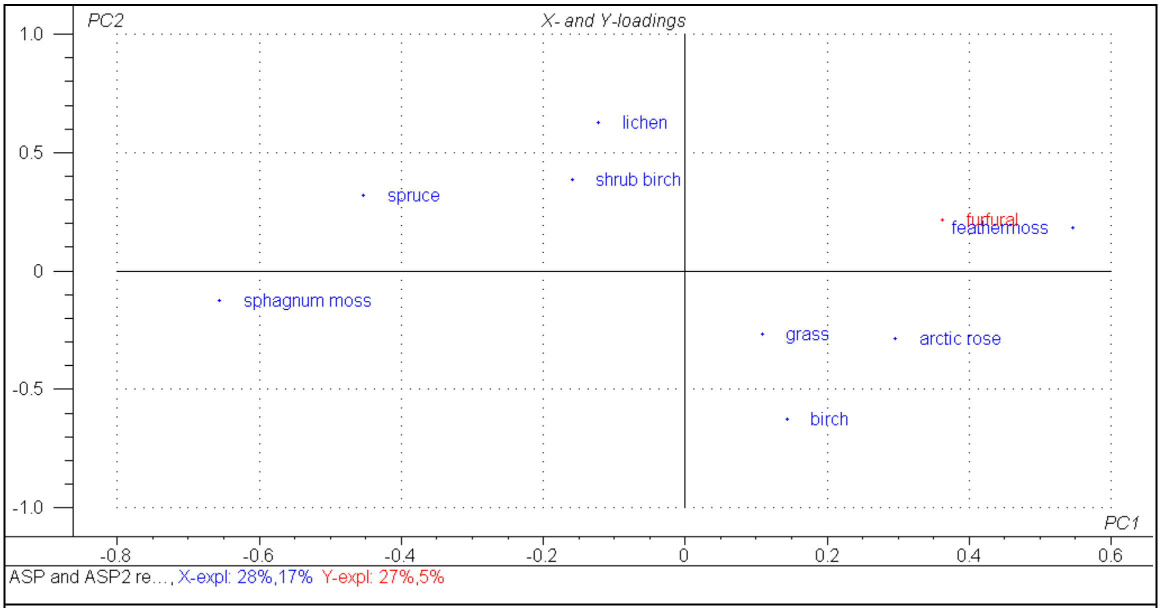

Figure D-31: PLS1 Loading Plot for Furfural.

Figure D-31: PLS1 Loading Plot for Furfural. The input for this plot included all vegetation attributes and furfural as only the molecular fingerprint compound. This loading plot explains 45 % of the total variance in the x-direction and 32 % of the total variance in the y-direction.

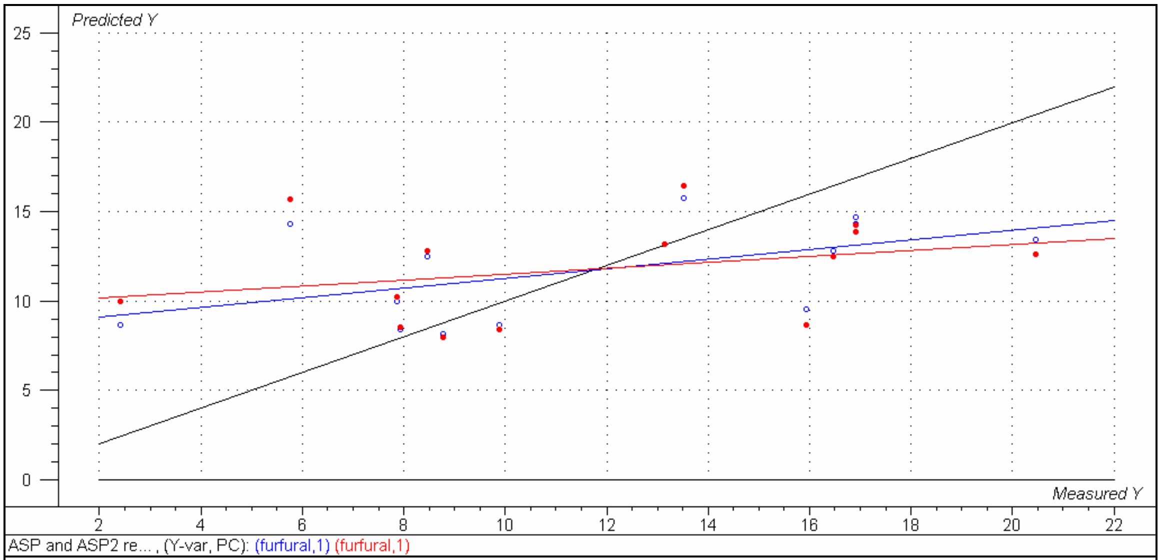

Figure D-32: PLS1 Predicted vs. Measured for Furfural.

Figure D-32: PLS1 Predicted vs. Measured for Furfural. The x-axis contains the measured relative percentages for molecular fingerprint compound furfural for all soil leachates and the y-axis contains the predicted relative percentages for molecular fingerprint compound furfural. The black trend line represents the target line. The blue trend line represents the measured relative percentages (r^2= 0.521) and the red trend line represents the predicted relative percentages (r^2= 0.304).

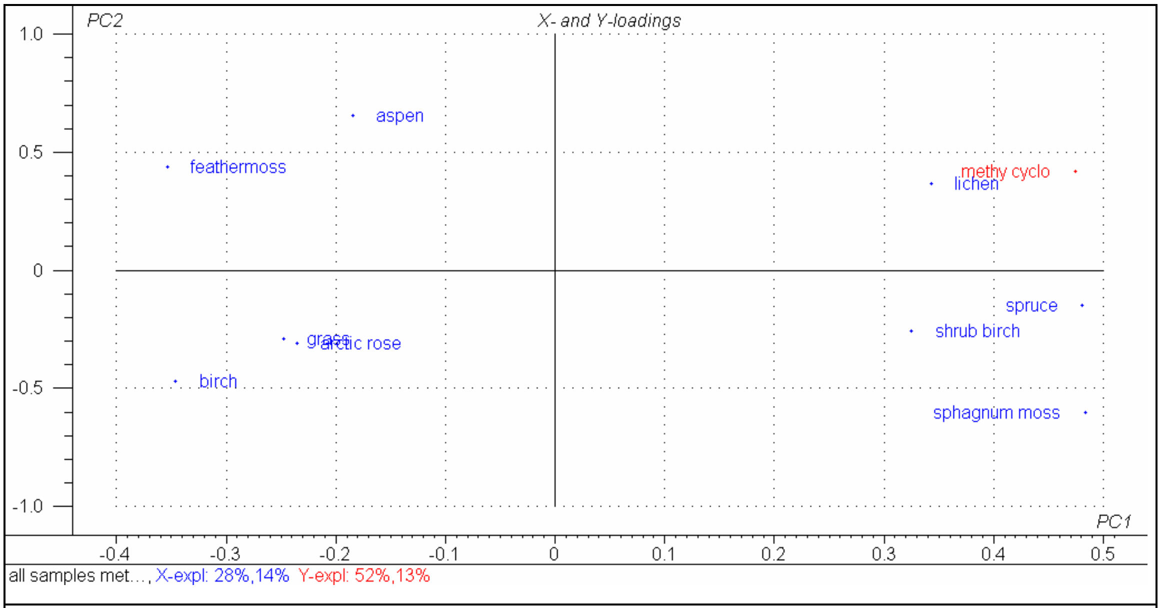

Figure D-33: PLS1 Loading Plot for Methyl Cyclopentenone.

Figure D-33: PLS1 Loading Plot for Methyl Cyclopentenone. The input for this plot included all vegetation attributes and methyl cyclopentenone as only the molecular fingerprint compound. This loading plot explains 42 % of the total variance in the x-direction and 65 % of the total variance in the y-direction.

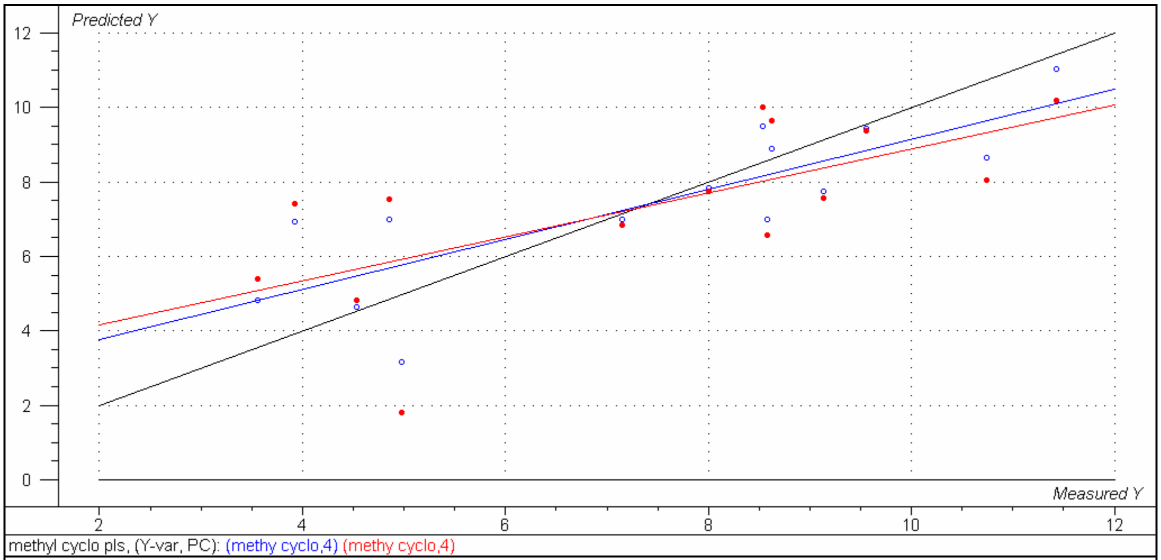

Figure D-34: PLS1 Predicted vs. Measured for Methyl Cyclopentenone.

Figure D-34: PLS1 Predicted vs. Measured for Methyl Cyclopentenone. The x-axis contains the measured relative percentages for molecular fingerprint compound methyl cyclopentenone for all soil leachates and the y-axis contains the predicted relative percentages for molecular fingerprint compound methyl cyclopentenone. The black trend line represents the target line. The blue trend line represents the measured relative percentages (r^2= 0.820) and the red trend line represents the predicted relative percentages (r^2= 0.672).

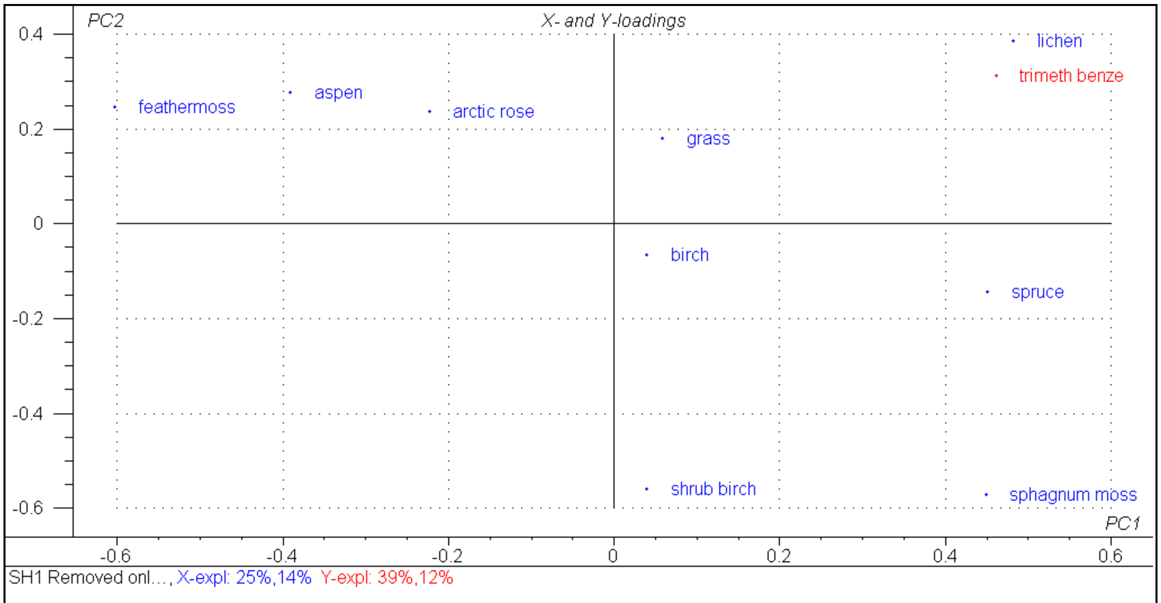

Figure D-35: PLS1 Loading Plot for Trimethyl Benzene.

Figure D-35: PLS1 Loading Plot for Trimethyl Benzene. The input for this plot included all vegetation attributes and trimethyl benzene as only the molecular fingerprint compound. This loading plot explains 39 % of the total variance in the x-direction and 51 % of the total variance in the y-direction.

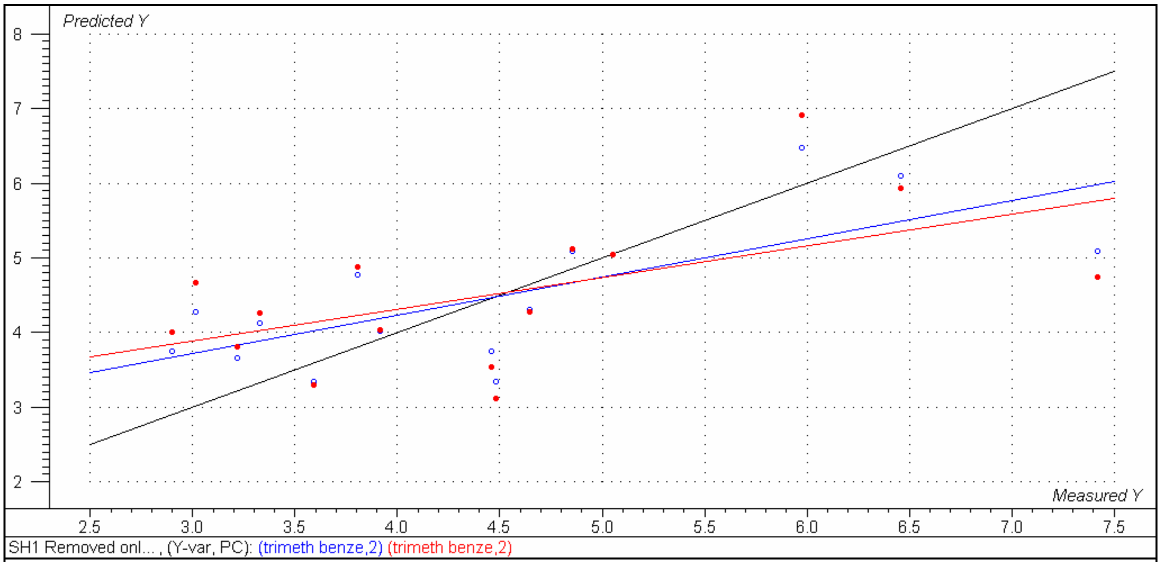

Figure D-36: PLS1 Predicted vs. Measured for Trimethyl Benzene.

Figure D-36: PLS1 Predicted vs. Measured for Trimethyl Benzene. The x-axis contains the measured relative percentages for molecular fingerprint compound trimethyl benzene for all soil leachates and the y-axis contains the predicted relative percentages for molecular fingerprint compound trimethyl benzene. The black trend line represents the target line. The blue trend line represents the measured relative percentages (r^2= 0.715) and the red trend line represents the predicted relative percentages (r^2= 0.561).

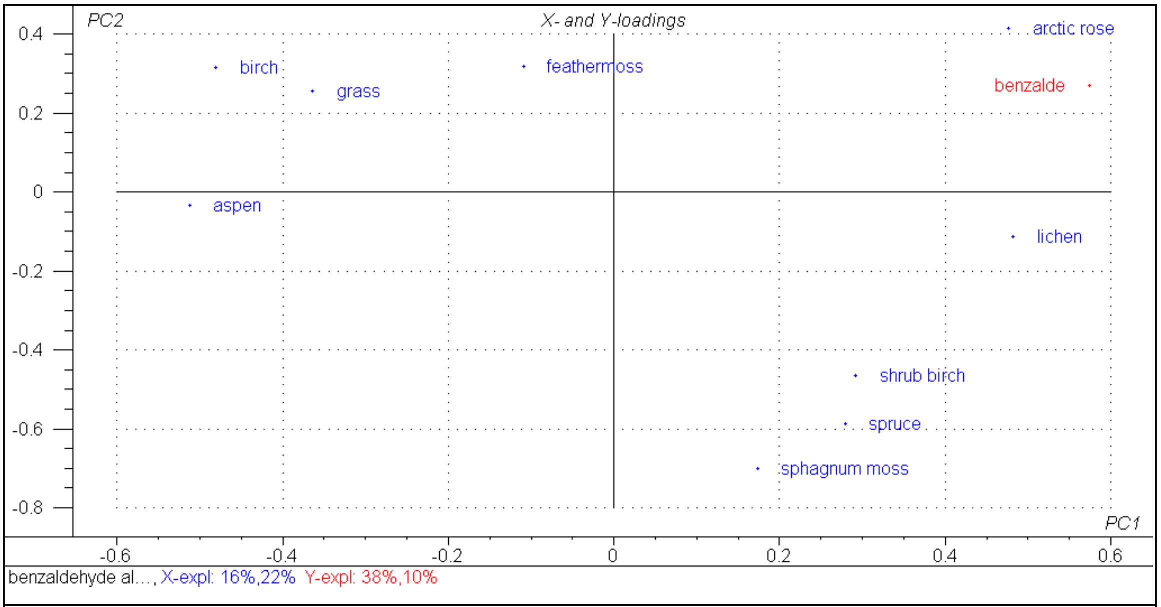

Figure D-37: PLS1 Loading Plot for Benzaldehyde.

Figure D-37: PLS1 Loading Plot for Benzaldehyde. The input for this plot included all vegetation attributes and benzaldehyde as only the molecular fingerprint compound. This loading plot explains 38 % of the total variance in the x-direction and 48 % of the total variance in the y-direction.

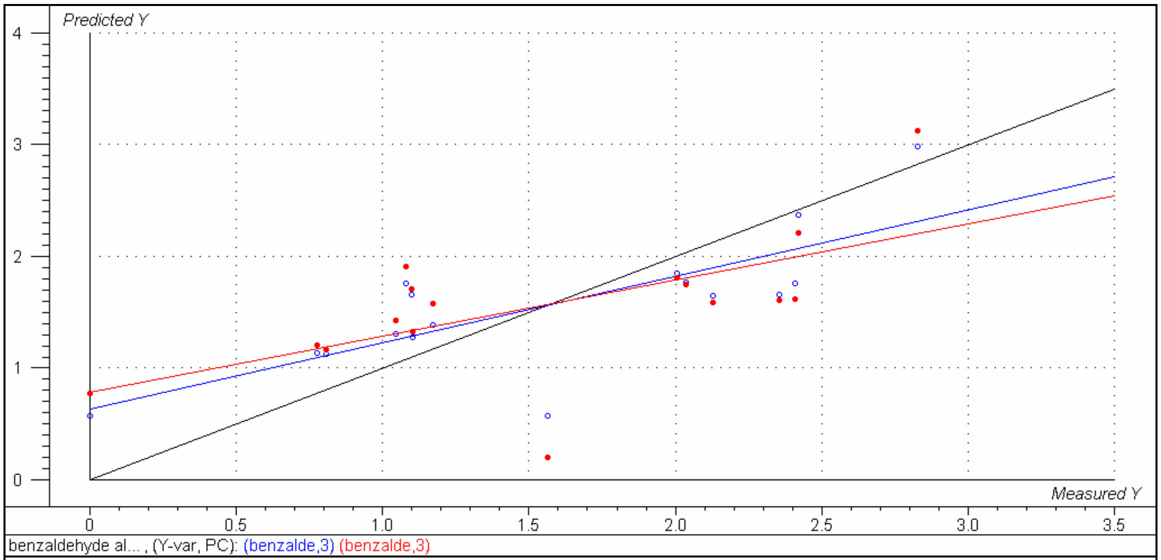

Figure D-38: PLS1 Predicted vs. Measured for Benzaldehyde.

Figure D-38: PLS1 Predicted vs. Measured for Benzaldehyde. The x-axis contains the measured relative percentages for molecular fingerprint compound benzaldehyde for all soil leachates and the y-axis contains the predicted relative percentages for molecular fingerprint compound benzaldehyde. The black trend line represents the target line. The blue trend line represents the measured relative percentages (r^2= 0.771) and the red trend line represents the predicted relative percentages (r^2= 0.622).

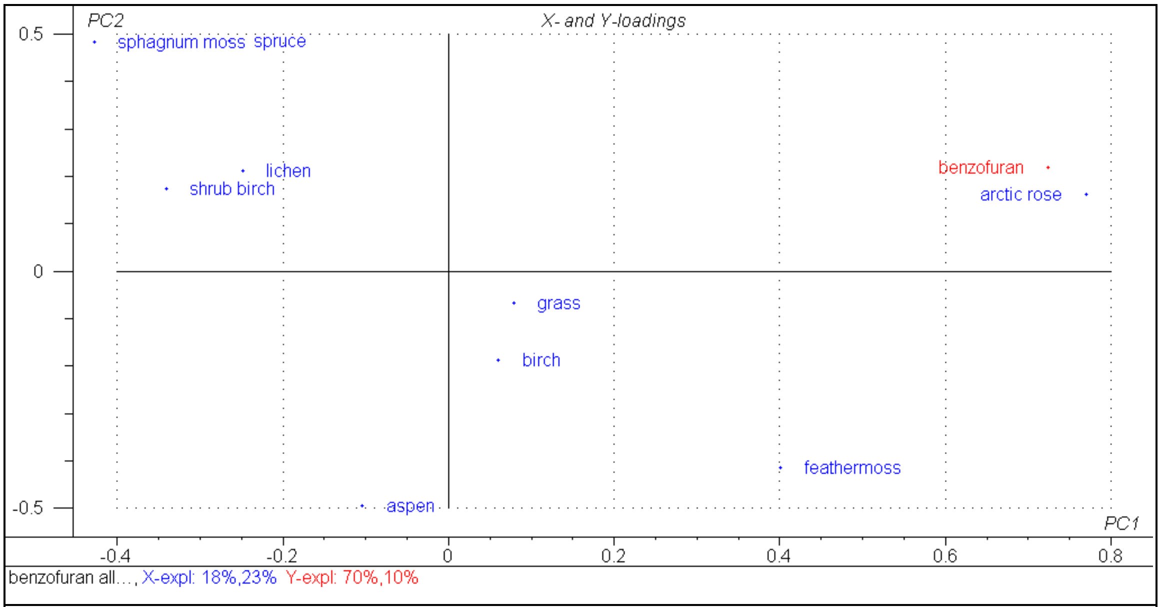

Figure D-39: PLS1 Loading Plot for Benzofuran.

Figure D-39: PLS1 Loading Plot for Benzofuran. The input for this plot included all vegetation attributes and benzofuran as only the molecular fingerprint compound. This loading plot explains 41 % of the total variance in the x-direction and 80 % of the total variance in the y-direction.

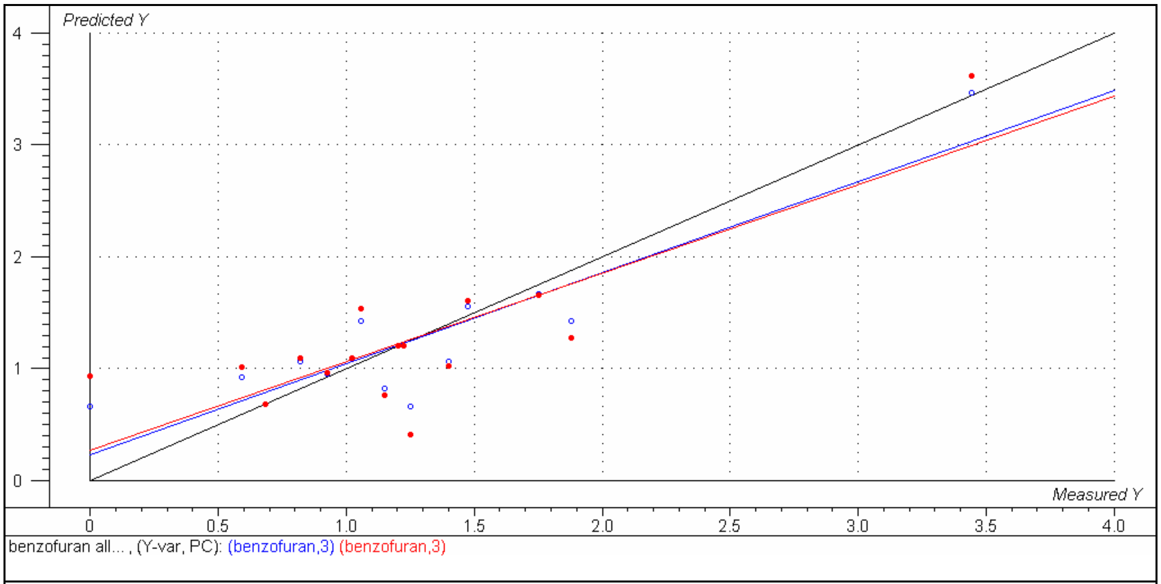

Figure D-40: PLS1 Predicted vs. Measured for Benzofuran.

Figure D-40: PLS1 Predicted vs. Measured for Benzofuran. The x-axis contains the measured relative percentages for molecular fingerprint compound benzofuran for all soil leachates and the y-axis contains the predicted relative percentages for molecular fingerprint compound benzofuran. The black trend line represents the target line. The blue trend line represents the measured relative percentages (r^2= 0.902) and the red trend line represents the predicted relative percentages (r^2= 0.826).

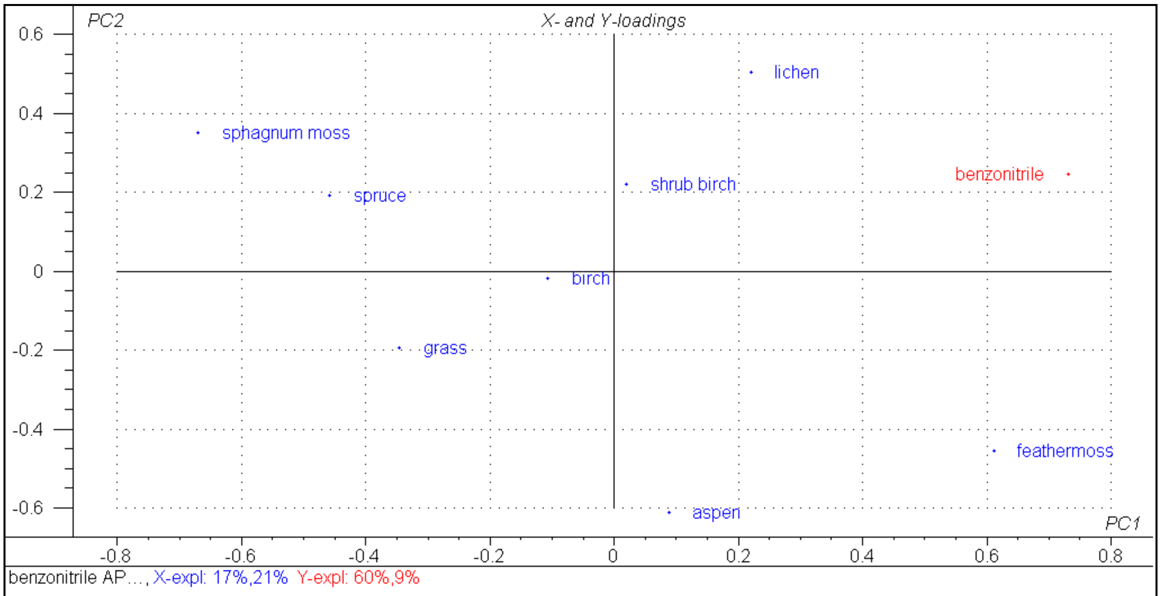

Figure D-41: PLS1 Loading Plot for Benzonitrile.

Figure D-41: PLS1 Loading Plot for Benzonitrile. The input for this plot included all vegetation attributes and benzonitrile as only the molecular fingerprint compound. This loading plot explains 38 % of the total variance in the x-direction and 69 % of the total variance in the y-direction.

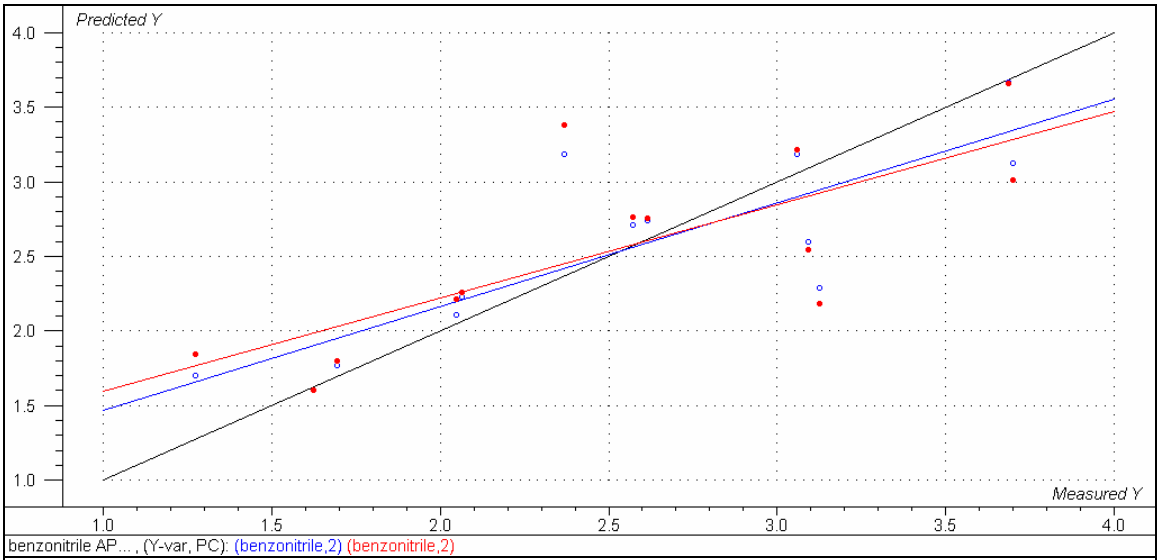

Figure D-42: PLS1 Predicted vs. Measured for Benzonitrile.

Figure D-42: PLS1 Predicted vs. Measured for Benzonitrile. The x-axis contains the measured relative percentages for molecular fingerprint compound benzonitrile for all soil leachates and the y-axis contains the predicted relative percentages for molecular fingerprint compound benzonitrile. The black trend line represents the target line. The blue trend line represents the measured relative percentages (r^2= 0.833) and the red trend line represents the predicted relative percentages (r^2= 0.754).

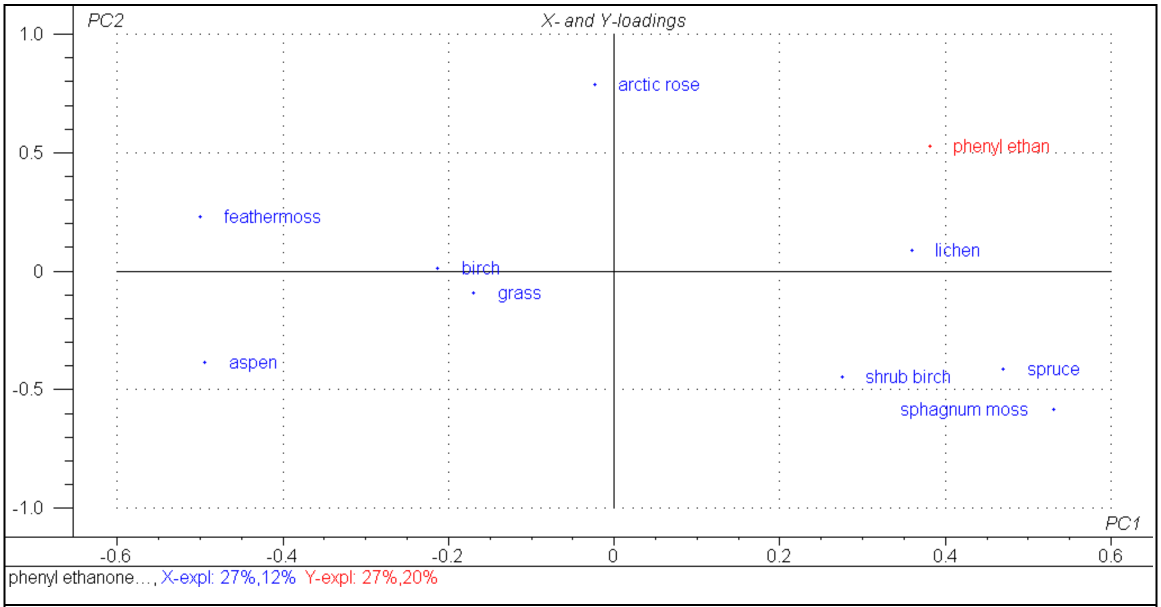

Figure D-43: PLS1 Loading Plot for Phenyl Ethanone.

Figure D-43: PLS1 Loading Plot for Phenyl Ethanone. The input for this plot included all vegetation attributes and phenyl ethanone as only the molecular fingerprint compound. This loading plot explains 39 % of the total variance in the x-direction and 47 % of the total variance in the y-direction.

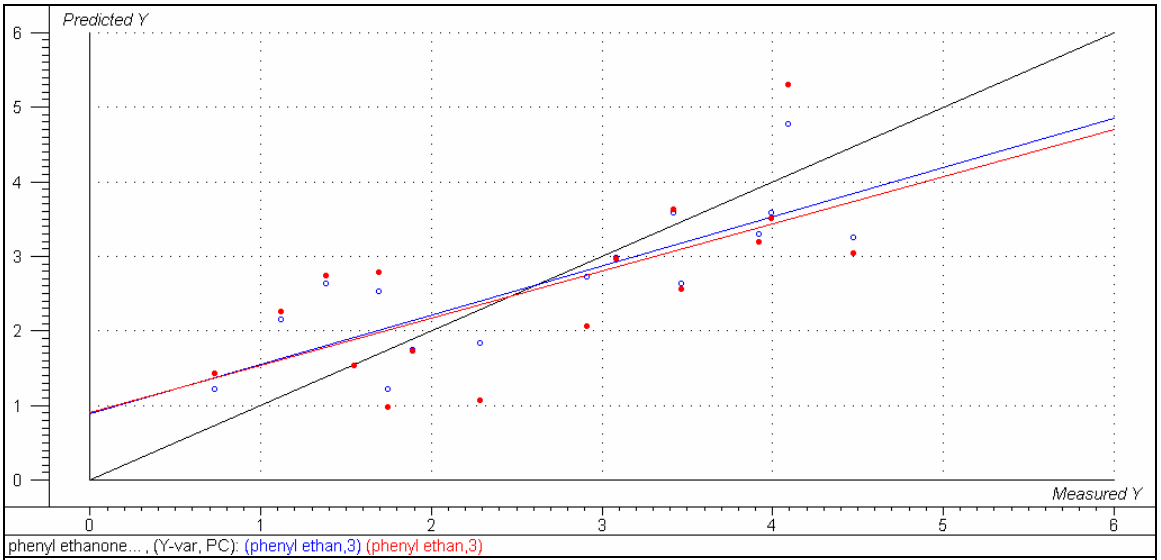

Figure D-44: PLS1 Predicted vs. Measured for Phenyl Ethanone.

Figure D-44: PLS1 Predicted vs. Measured for Phenyl Ethanone. The x-axis contains the measured relative percentages for molecular fingerprint compound phenyl ethanone for all soil leachates and the y-axis contains the predicted relative percentages for molecular fingerprint compound phenyl ethanone. The black trend line represents the target line. The blue trend line represents the measured relative percentages (r^2= 0.812) and the red trend line represents the predicted relative percentages (r^2= 0.684).

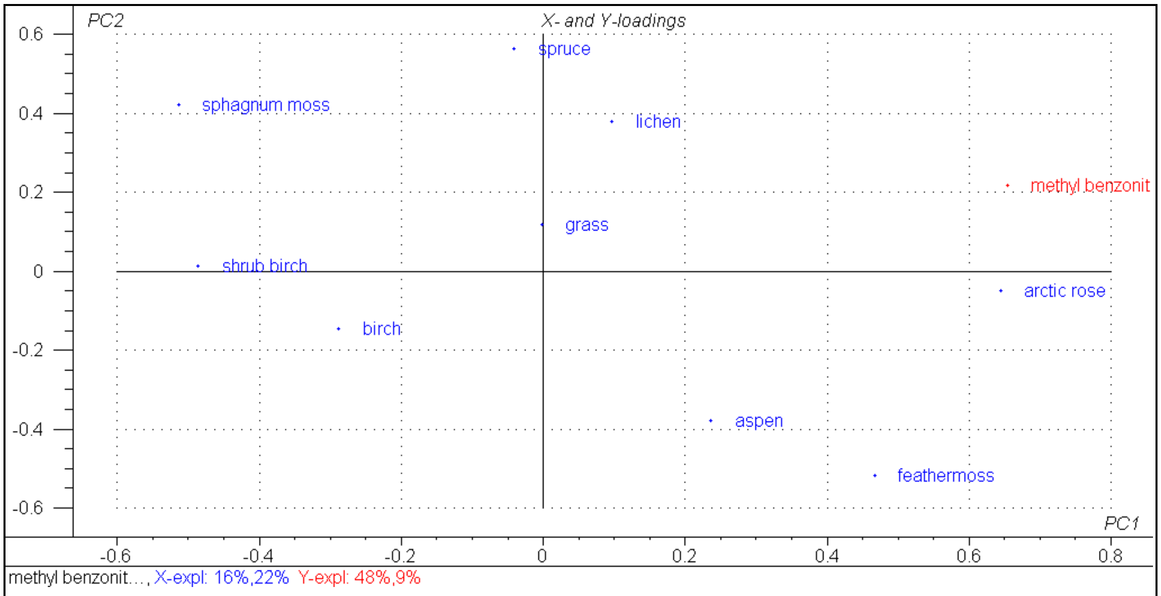

Figure D-45: PLS1 Loading Plot for Methyl Benzonitrile.

Figure D-45: PLS1 Loading Plot for Methyl Benzonitrile. The input for this plot included all vegetation attributes and methyl benzonitrile as only the molecular fingerprint compound. This loading plot explains 38 % of the total variance in the x-direction and 57 % of the total variance in the y-direction.

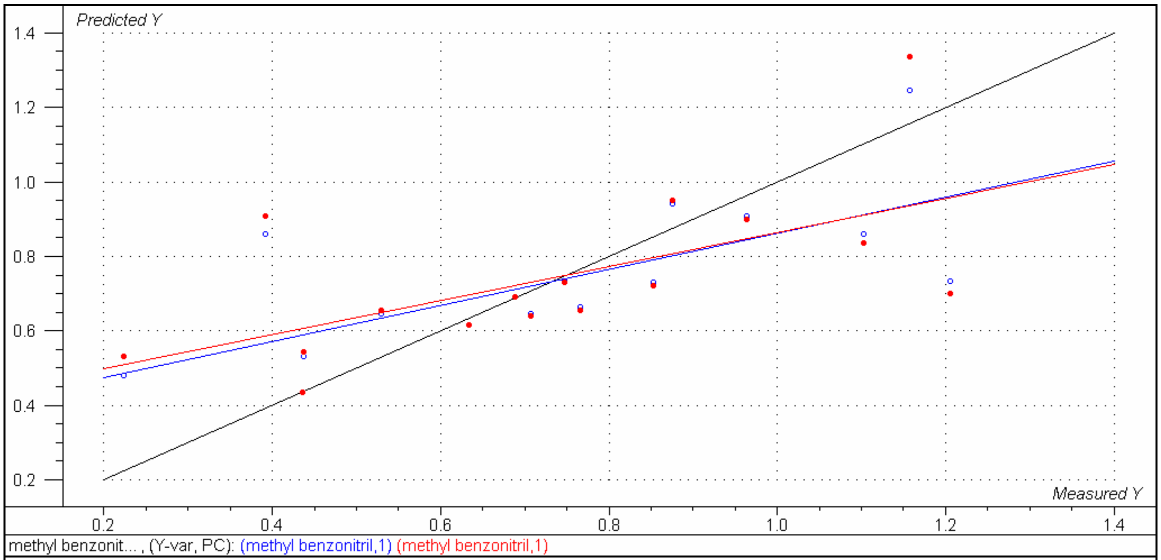

Figure D-46: PLS1 Predicted vs. Measured for Methyl Benzonitrile.

Figure D-46: PLS1 Predicted vs. Measured for Methyl Benzonitrile. The x-axis contains the measured relative percentages for molecular fingerprint compound methyl benzonitrile for all soil leachates and the y-axis contains the predicted relative percentages for molecular fingerprint compound methyl benzonitrile. The black trend line represents the target line. The blue trend line represents the measured relative percentages (r^2= 0.696) and the red trend line represents the predicted relative percentages (r^2= 0.614).

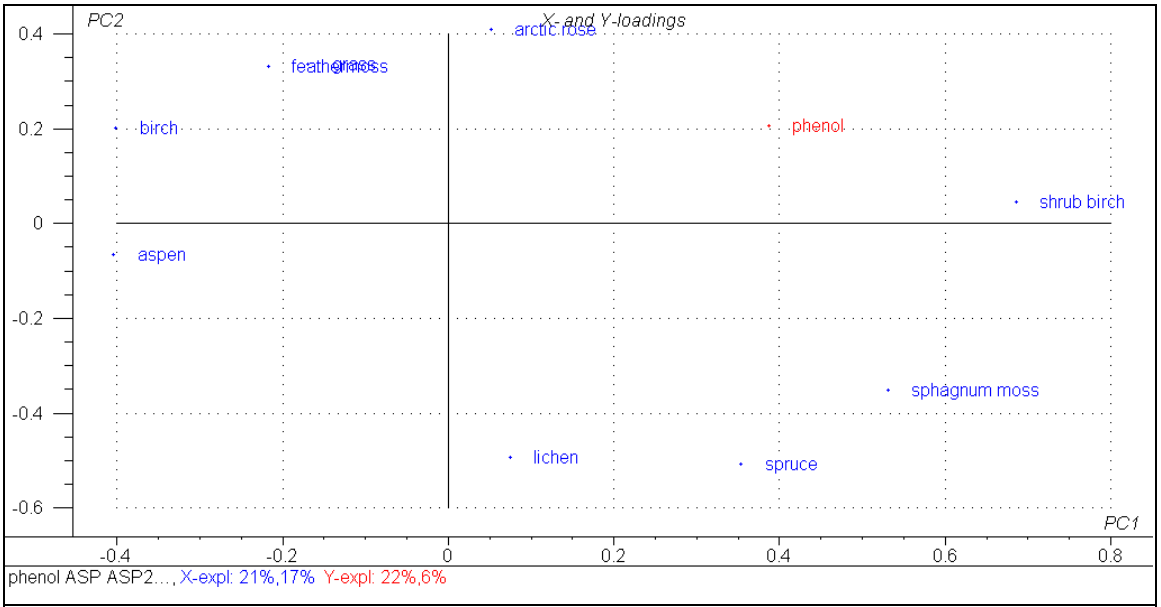

Figure D-47: PLS1 Loading Plot for Phenol.

Figure D-47: PLS1 Loading Plot for Phenol. The input for this plot included all vegetation attributes and phenol as only the molecular fingerprint compound. This loading plot explains 38 % of the total variance in the x-direction and 28 % of the total variance in the y-direction.

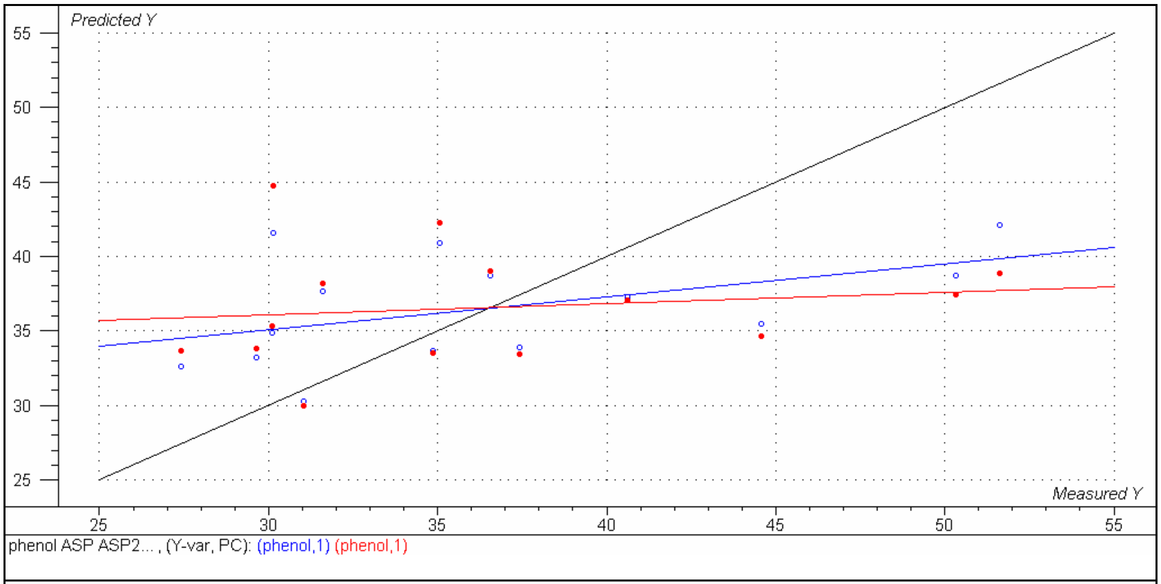

Figure D-48: PLS1 Predicted vs. Measured for Phenol.

Figure D-48: PLS1 Predicted vs. Measured for Phenol. The x-axis contains the measured relative percentages for molecular fingerprint compound phenol for all soil leachates and the y-axis contains the predicted relative percentages for molecular fingerprint compound phenol. The black trend line represents the target line. The blue trend line represents the measured relative percentages (r^2= 0.470) and the red trend line represents the predicted relative percentages (r^2= 0.470).

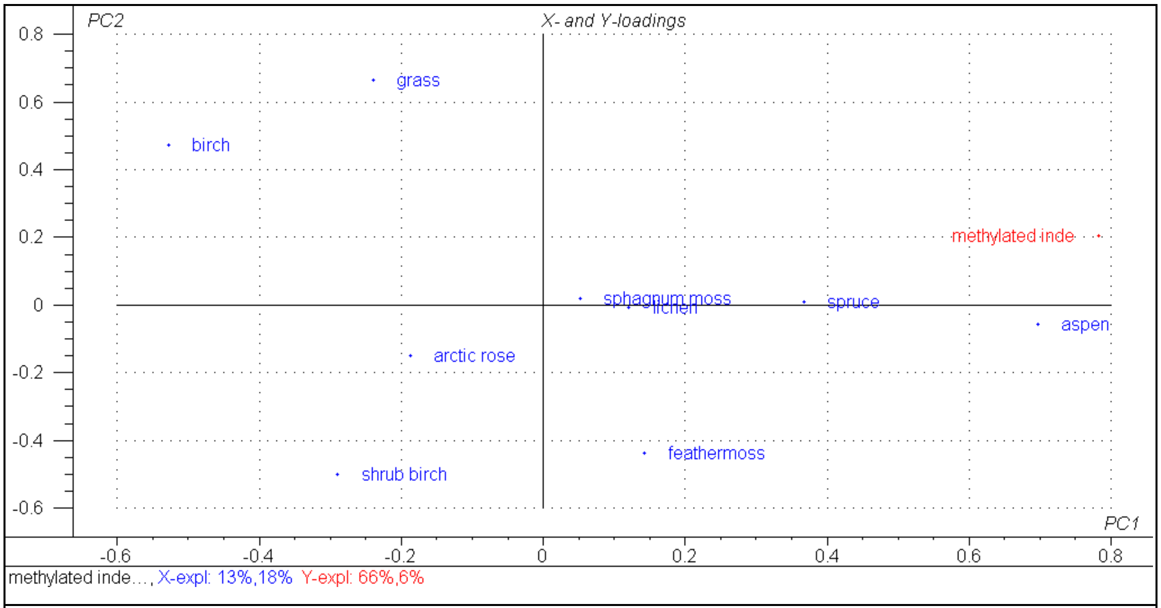

Figure D-49: PLS1 Loading Plot for Methylated Indene.

Figure D-49: PLS1 Loading Plot for Methylated Indene. The input for this plot included all vegetation attributes and methylated indene as only the molecular fingerprint compound. This loading plot explains 31 % of the total variance in the x-direction and 72 % of the total variance in the y-direction.

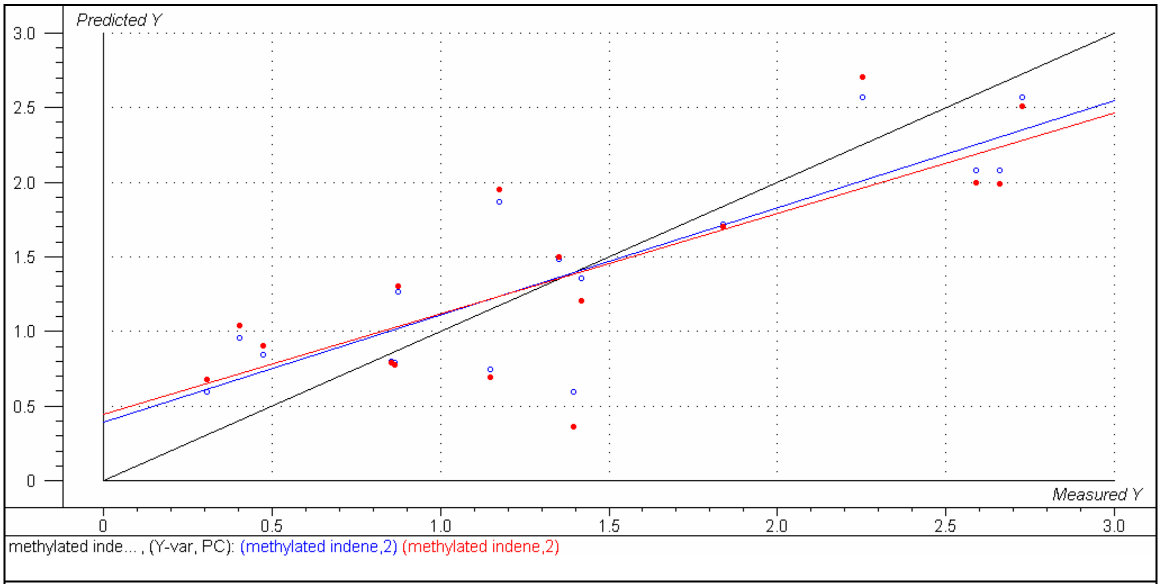

Figure D-50: PLS1 Predicted vs. Measured for Methylated Indene.

Figure D-50: PLS1 Predicted vs. Measured for Methylated Indene. The x-axis contains the measured relative percentages for molecular fingerprint compound methylated indene for all soil leachates and the y-axis contains the predicted relative percentages for molecular fingerprint compound methylated indene. The black trend line represents the target line. The blue trend line represents the measured relative percentages (r^2= 0.470) and the red trend line represents the predicted relative percentages (r^2= 0.470).

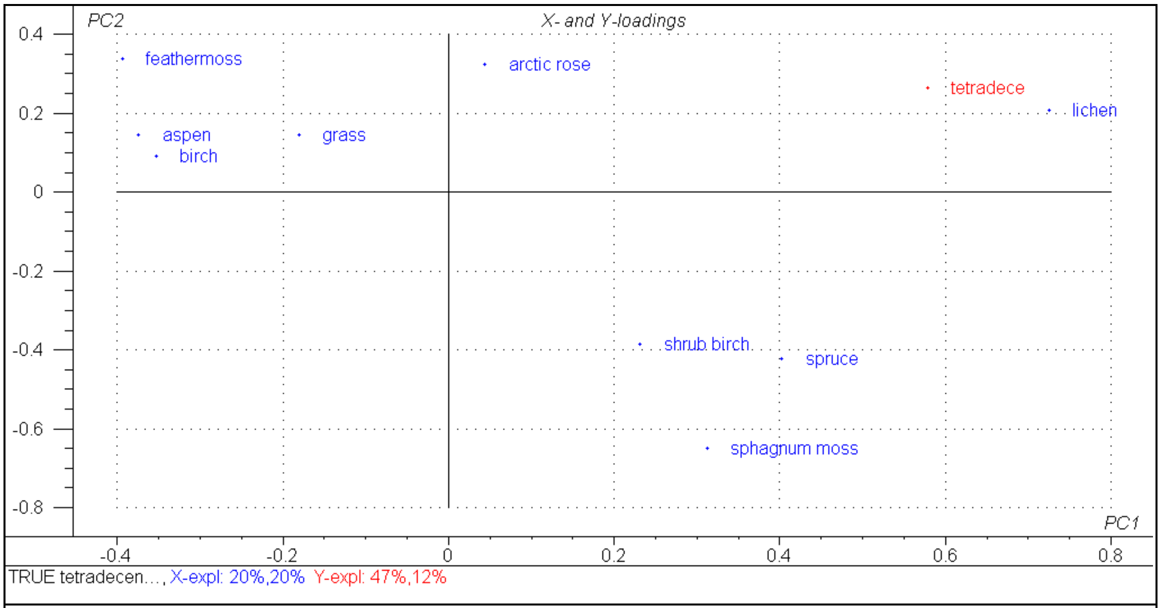

Figure D-51: PLS1 Loading Plot for Tetradecene.

Figure D-51: PLS1 Loading Plot for Tetradecene. The input for this plot included all vegetation attributes and tetradecene as only the molecular fingerprint compound. This loading plot explains 40 % of the total variance in the x-direction and 59 % of the total variance in the y-direction.

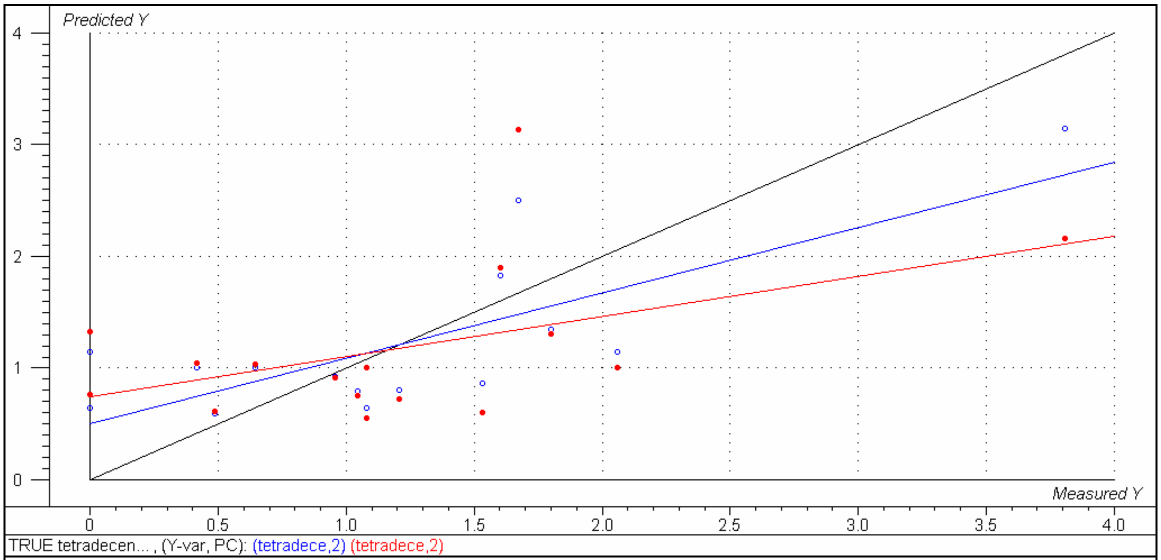

Figure D-52: PLS1 Predicted vs. Measured for Tetradecene.

Figure D-52: PLS1 Predicted vs. Measured for Tetradecene. The x-axis contains the measured relative percentages for molecular fingerprint compound tetradecene for all soil leachates and the y-axis contains the predicted relative percentages for molecular fingerprint compound tetradecene. The black trend line represents the target line. The blue trend line represents the measured relative percentages (r^2= 0.765) and the red trend line represents the predicted relative percentages (r^2= 0.484).

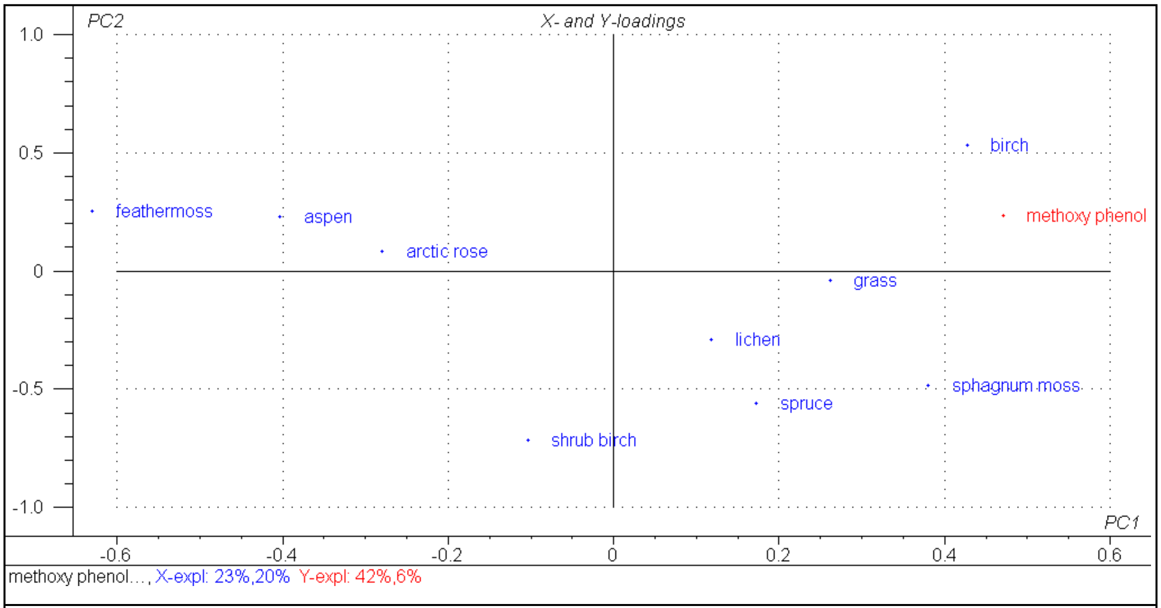

Figure D-53: PLS1 Loading Plot for Methoxy Phenol.

Figure D-53: PLS1 Loading Plot for Methoxy Phenol. The input for this plot included all vegetation attributes and methoxy phenol as only the molecular fingerprint compound. This loading plot explains 43 % of the total variance in the x-direction and 48 % of the total variance in the y-direction.

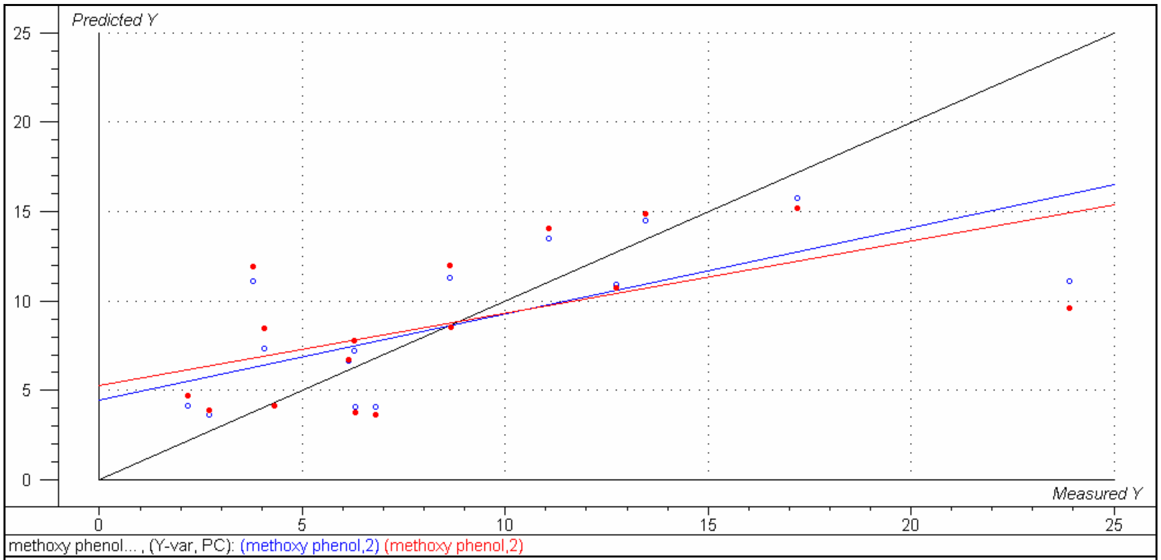

Figure D-54: PLS1 Predicted vs. Measured for Methoxy Phenol.

Figure D-54: PLS1 Predicted vs. Measured for Methoxy Phenol. The x-axis contains the measured relative percentages for molecular fingerprint compound methoxy phenol for all soil leachates and the y-axis contains the predicted relative percentages for molecular fingerprint compound methoxy phenol. The black trend line represents the target line. The blue trend line represents the measured relative percentages (r^2= 0.694) and the red trend line represents the predicted relative percentages (r^2= 0.581).

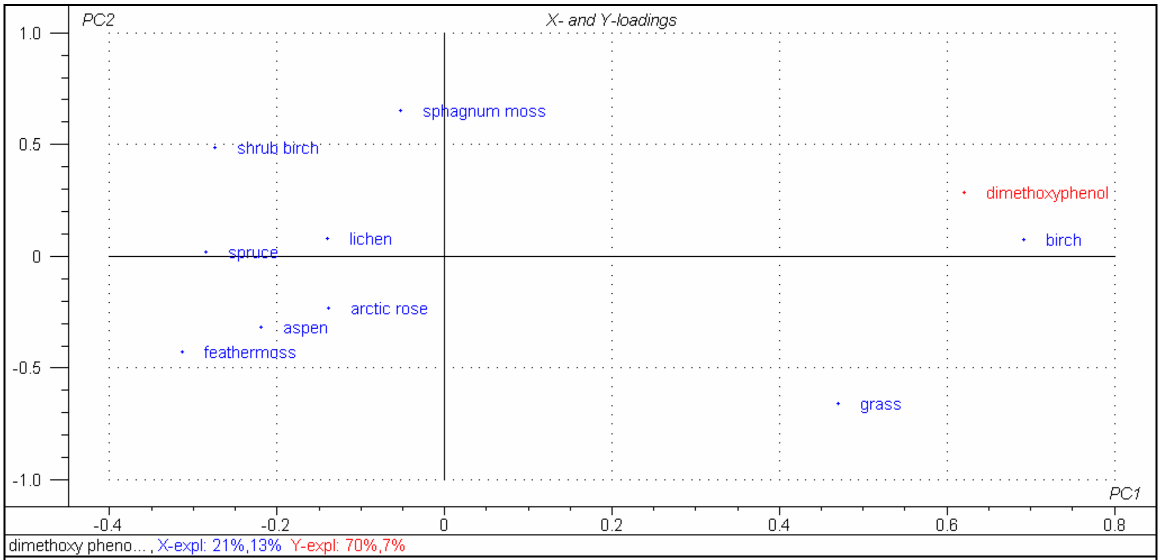

Figure D-55: PLS1 Loading Plot for Dimethoxy Phenol.

Figure D-55: PLS1 Loading Plot for Dimethoxy Phenol. The input for this plot included all vegetation attributes and dimethoxy phenol as only the molecular fingerprint compound. This loading plot explains 34 % of the total variance in the x-direction and 77 % of the total variance in the y-direction.

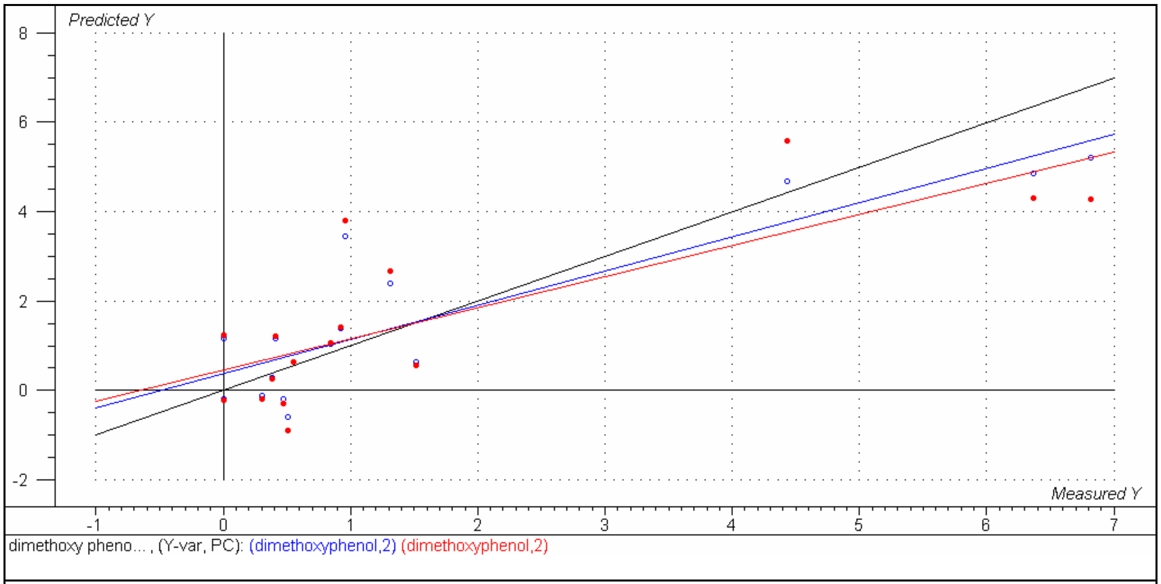

Figure D-56: PLS1 Predicted vs. Measured for Dimethoxy Phenol.

Figure D-56: PLS1 Predicted vs. Measured for Dimethoxy Phenol. The x-axis contains the measured relative percentages for molecular fingerprint compound dimethoxy phenol for all soil leachates and the y-axis contains the predicted relative percentages for molecular fingerprint compound dimethoxy phenol. The black trend line represents the target line. The blue trend line represents the measured relative percentages (r^2= 0.874) and the red trend line represents the predicted relative percentages (r^2= 0.787).

Appendix E: Tables

Table E-1: Final Data Matrix for Soil Leachates Including Relative Percentages for all Molecular Fingerprint Compounds and their Vegetation Attributes.

	dimethyl benzene	furfural	methyl cyclopentenone	trimethyl benzene	benzaldehyde	benzofuran	benzonitrile	phenyl ethanone	methyl benzonitrile	phenol	methylated indene	tetradecene	methoxy phenol	dimethoxy phenol
AP	16.14	13.52	3.93	3.91	2.42	3.44	6.46	2.91	1.16	40.62	0.86	1.60	2.71	0.30
Haystack	18.07	5.76	4.85	4.48	1.10	0.82	2.37	1.38	0.71	50.33	1.39	0.42	6.82	1.51
HR	16.62	15.93	10.74	3.22	2.13	1.15	3.13	4.47	0.44	30.16	0.85	1.21	8.66	1.31
LG	26.29	16.46	11.43	5.97	2.83	1.20	3.69	4.09	0.96	34.87	1.84	3.81	12.72	0.93
P2	15.23	8.77	8.53	3.01	1.10	0.59	1.69	1.69	0.44	51.62	0.88	1.53	4.07	0.84
P3	24.75	7.93	8.63	6.46	1.17	0.68	2.57	2.29	0.77	35.06	1.35	1.67	6.28	0.38
SHAY1	19.38	16.92	4.98	11.57	0.78	1.22	3.70	1.55	0.69	27.42	1.15	1.05	8.63	0.96
SHAY2	10.63	7.87	3.55	3.33	0.80	0.93	2.06	1.12	0.22	44.57	0.40	0.49	17.20	6.82
UBS	16.05	20.46	9.13	3.80	2.04	1.75	3.09	3.08	0.88	30.10	1.17	1.80	6.14	0.50
HWLB	18.13	8.46	8.00	4.65	2.00	1.02	2.61	3.92	0.63	29.63	0.47	0.65	13.44	6.37
CSCP	15.34	16.90	8.58	3.59	2.35	1.40	3.06	3.46	0.53	36.55	0.31	1.08	6.29	0.55
ASPEN	16.07	24.46	7.15	4.46	1.56	1.25	2.05	1.74	1.10	31.04	2.73	1.08	4.31	0.47
BIRCH	15.08	13.15	4.53	5.05	1.04	1.48	1.62	1.89	0.85	37.44	1.42	0.96	11.06	4.43
SPRUCE	22.68	9.89	9.55	4.86	2.41	1.06	4.92	3.42	0.75	31.62	2.59	2.06	3.79	0.41
SPRUCE2	33.98	2.41	9.61	7.42	1.08	1.88	1.27	4.00	1.21	10.59	2.66	0.00	23.89	0.00
ASPEN2	15.87	69.93	5.75	2.90	0.00	0.00	0.00	0.73	0.39	0.00	2.25	0.00	2.17	0.00

	Tree			Shrub		Ground Cover			
	birch	aspen	spruce	shrub birch	arctic rose	feather moss	sphagnum moss	lichen	grass
AP	-1	-1	-1	-1	1	1	-1	-1	-1
Haystack	-1	-1	-1	1	-1	1	-1	-1	-1
HR	-1	-1	-1	1	-1	-1	1	-1	-1
LG	-1	-1	-1	-1	-1	-1	-1	1	-1
P2	-1	-1	1	1	-1	-1	1	-1	-1
P3	-1	-1	1	1	-1	-1	1	-1	-1
SHAY1	1	-1	-1	-1	-1	1	-1	-1	-1
SHAY2	1	-1	-1	-1	-1	-1	1	-1	-1
UBS	-1	-1	1	-1	-1	1	-1	-1	-1
HWLB	1	-1	-1	-1	-1	-1	-1	-1	-1
CSCP	-1	-1	-1	1	-1	1	-1	-1	-1
ASPEN	-1	1	-1	-1	-1	1	-1	-1	-1
BIRCH	1	-1	-1	-1	-1	-1	-1	-1	1
SPRUCE	-1	-1	1	-1	-1	-1	1	-1	-1
SPRUCE2	-1	-1	1	-1	-1	-1	1	-1	-1
ASPEN2	-1	1	-1	-1	-1	1	-1	-1	-1

Table E-2: Q-test for Soil Leachates.

	dimethyl benzene	furfural	methyl cyclopentenone	trimethyl benzene	benzaldehyde	benzofuran	benzonitrile	phenyl ethanone	methyl benzonitrile	phenol	methylated indene	tetradecene
APA	11.64	7.71	2.90	2.38	18.20	1.70	3.78	2.12	0.98	44.21	0.46	3.92
APB	17.84	11.73	4.27	3.80	2.92	2.96	6.68	3.51	0.34	42.77	0.19	3.01
APC	23.93	13.94	5.52	4.79	3.80	3.84	8.75	4.58	0.44	27.85	1.50	1.06
APD	15.35	11.62	4.82	4.01	3.33	3.13	7.30	4.02	1.82	42.87	0.81	0.92
Range	12.29	6.23	2.62	2.41	15.29	2.14	4.97	2.46	1.48	16.36	1.31	3.00
Gap$_{high}$	6.10	2.21	0.70	0.78	14.41	0.72	1.45	0.56	0.84	1.34	0.68	0.92
Gap$_{low}$	3.71	3.91	1.37	1.42	0.41	1.25	2.90	1.39	0.11	14.92	0.27	0.14
Qcalc$_{high}$	0.50	0.35	0.27	0.32	0.94	0.34	0.29	0.23	0.57	0.08	0.52	0.31
Qcalc$_{low}$	0.30	0.63	0.52	0.59	0.03	0.59	0.58	0.57	0.07	0.91	0.21	0.05

	dimethyl benzene	furfural	methyl cyclopentenone	trimethyl benzene	benzaldehyde	benzofuran	benzonitrile	phenyl ethanone	methyl benzonitrile	phenol	methylated indene	tetradecene
2APA	18.18	20.64	3.92	4.01	1.42	5.04	6.27	1.52	1.32	34.29	0.85	0.82
2APB	17.79	19.59	3.30	5.72	1.39	5.09	6.46	1.63	1.44	37.41	0.81	0.62
2APC	16.73	18.33	3.09	4.47	1.29	4.54	6.03	1.53	1.42	40.07	1.65	1.04
2APD	18.34	17.98	3.35	4.29	1.31	4.78	6.90	1.63	1.56	38.04	0.78	1.01
Range	1.62	2.67	0.83	1.71	0.13	0.55	0.87	0.11	0.24	5.78	0.87	0.41
Gap$_{high}$	0.17	1.05	0.58	1.26	0.02	0.05	0.44	0.10	0.11	2.03	0.80	0.03
Gap$_{low}$	1.06	0.35	0.20	0.28	0.02	0.24	0.24	0.01	0.11	3.12	0.03	0.20
Qcalc$_{high}$	0.10	0.39	0.70	0.73	0.17	0.09	0.51	0.91	0.48	0.35	0.91	0.07
Qcalc$_{low}$	0.66	0.13	0.24	0.16	0.16	0.43	0.28	0.09	0.44	0.54	0.04	0.49

	benzene	furfural	methyl cyclopentenone	benzene	benzaldehyde	benzofuran	benzonitrile	ethanone	benzonitrile	phenol	indene	tetradecene
H2A	22.36	6.55	5.50	5.24	1.17	0.94	2.72	2.04	1.14	50.51	1.52	0.32
H2B	37.31	10.30	15.02	10.74	2.12	1.66	4.80	4.13	1.35	10.30	1.19	1.08
H2C	18.61	5.96	5.21	4.81	1.21	0.86	2.51	0.40	0.60	57.76	1.57	0.51
H2D	18.12	6.33	5.17	4.61	1.21	0.89	2.52	2.07	0.57	56.50	1.48	0.53
Range	19.19	4.34	9.85	6.13	0.95	0.80	2.29	3.72	0.78	47.46	0.37	0.76
Gap$_{high}$	14.95	3.75	9.52	5.50	0.91	0.72	2.08	2.06	0.21	1.26	0.05	0.55
Gap$_{low}$	0.48	0.38	0.04	0.20	0.04	0.03	0.02	1.63	0.04	40.21	0.28	0.19
Qcalc$_{high}$	0.78	0.86	0.97	0.90	0.96	0.90	0.91	0.55	0.27	0.03	0.13	0.72
Qcalc$_{low}$	0.03	0.09	0.00	0.03	0.04	0.03	0.01	0.44	0.05	0.85	0.76	0.25

Table E-2: Q-test for Soil Leachates (cont'd).

	dimethyl benzene	furfural	methyl cyclopentenone	trimethyl benzene	benzaldehyde	benzofuran	benzonitrile	phenyl ethanone	methyl benzonitrile	phenol	methylated indene	tetradecene
HRA	20.32	6.79	5.53	6.31	4.40	0.62	1.47	11.66	0.44	42.02	0.11	0.33
HRB	12.28	21.27	10.43	2.74	2.46	1.25	2.72	3.74	0.33	41.81	0.20	0.78
HRC	15.10	19.70	11.03	3.24	2.46	1.45	3.33	4.65	0.42	36.77	0.87	1.00
HRD	27.48	12.41	14.19	4.69	2.19	1.14	4.32	6.42	0.69	22.54	1.72	2.20

Range	15.20	14.47	8.66	3.57	2.21	0.82	2.85	7.92	0.36	19.48	1.61	1.87
Gap$_{high}$	7.17	1.57	4.90	1.62	1.94	0.20	0.99	5.23	0.25	0.21	0.86	1.20
Gap$_{low}$	2.82	5.62	4.90	0.50	0.27	0.51	1.25	0.91	0.09	14.23	0.09	0.45
Qcalc$_{high}$	0.47	0.11	0.57	0.45	$\boxed{0.88}$	0.24	0.35	0.66	0.68	0.01	0.53	0.64
Qcalc$_{low}$	0.19	0.39	0.57	0.14	0.12	0.62	0.44	0.11	0.24	0.73	0.06	0.24

	dimethyl benzene	furfural	methyl cyclopentenone	trimethyl benzene	benzaldehyde	benzofuran	benzonitrile	phenyl ethanone	methyl benzonitrile	phenol	methylated indene	tetradecene
LGA	25.38	16.93	11.21	5.69	2.66	0.00	3.08	3.78	1.40	26.79	0.24	2.83
LGB	13.85	12.05	7.13	33.21	2.05	1.00	2.31	2.18	0.40	23.65	0.23	1.94
LGC	19.24	17.30	10.86	4.71	2.68	1.33	2.86	3.17	0.55	32.49	1.78	3.02
LGD	24.12	11.67	9.03	5.34	2.30	1.49	3.55	3.75	0.71	31.93	2.28	3.83

Range	11.53	5.63	4.08	28.49	0.63	1.49	1.24	1.60	1.01	8.85	2.05	1.89
Gap$_{high}$	1.26	0.37	0.35	27.52	0.02	0.16	0.48	0.03	0.69	0.56	0.50	0.82
Gap$_{low}$	5.39	0.38	1.90	0.62	0.24	1.00	0.55	1.00	0.15	3.14	0.01	0.89
Qcalc$_{high}$	0.11	0.06	0.09	$\boxed{0.97}$	0.03	0.11	0.38	0.02	0.69	0.06	0.24	0.43
Qcalc$_{low}$	0.47	0.07	0.46	0.02	0.39	0.67	0.44	0.62	0.15	0.36	0.00	0.47

	dimethyl benzene	furfural	methyl cyclopentenone	trimethyl benzene	benzaldehyde	benzofuran	benzonitrile	phenyl ethanone	methyl benzonitrile	phenol	methylated indene	tetradecene
P2A	21.65	11.43	9.58	4.41	1.37	0.00	0.00	2.02	3.23	44.75	0.33	1.22
P2B	13.19	9.27	8.02	2.97	1.28	0.66	1.66	1.73	0.41	59.46	0.26	1.10
P2C	21.98	12.05	9.64	4.26	1.56	0.80	2.33	2.18	0.65	40.79	1.56	2.20
P2D	12.89	6.40	9.24	2.29	0.65	0.40	1.36	1.44	0.32	62.57	0.92	1.52

Range	9.09	5.65	1.62	2.12	0.91	0.80	2.33	0.73	2.90	21.77	1.30	1.10
Gap$_{high}$	0.32	0.62	0.06	0.15	0.18	0.14	0.67	0.15	2.57	3.11	0.64	0.68
Gap$_{low}$	0.30	2.87	1.22	0.68	0.63	0.40	1.36	0.29	0.08	3.96	0.07	0.12
Qcalc$_{high}$	0.04	0.11	0.04	0.07	0.20	0.17	0.29	0.21	$\boxed{0.89}$	0.14	0.49	0.62
Qcalc$_{low}$	0.03	0.51	0.75	0.32	0.69	0.50	0.58	0.39	0.03	0.18	0.05	0.11

Table E-2: Q-test for Soil Leachates (cont'd).

	dimethyl benzene	furfural	methyl cyclopentenone	trimethyl benzene	benzaldehyde	benzofuran	benzonitrile	phenyl ethanone	methyl benzonitrile	phenol	methylated indene	tetradecene
P3A	31.99	10.73	10.70	0.84	1.45	0.00	3.39	2.85	1.74	34.85	0.35	1.10
P3B	26.68	9.43	9.08	7.25	1.41	0.83	3.04	2.60	0.84	36.75	0.53	1.55
P3C	23.89	8.67	8.60	6.39	1.31	0.73	2.68	2.42	0.72	41.81	1.46	1.31
P3D	28.99	7.43	10.05	7.14	1.06	0.63	2.56	2.34	0.89	34.10	2.30	2.50
Range	8.10	3.30	2.10	6.40	0.40	0.83	0.83	0.51	1.02	7.71	1.95	1.40
Gap_{high}	3.00	1.29	0.65	0.11	0.05	0.10	0.35	0.25	0.85	5.06	0.84	0.95
Gap_{low}	2.78	1.24	0.48	5.55	0.25	0.63	0.11	0.08	0.12	0.75	0.18	0.21
$Qcalc_{high}$	0.37	0.39	0.31	0.02	0.12	0.12	0.42	0.49	$\boxed{0.83}$	0.66	0.43	0.68
$Qcalc_{low}$	0.34	0.38	0.23	$\boxed{0.87}$	0.63	$\boxed{0.76}$	0.14	0.15	0.12	0.10	0.09	0.15

	dimethyl benzene	furfural	methyl cyclopentenone	trimethyl benzene	benzaldehyde	benzofuran	benzonitrile	phenyl ethanone	methyl benzonitrile	phenol	methylated indene	tetradecene
SH1A	18.94	19.00	6.23	12.53	0.74	1.52	3.86	1.90	0.00	34.77	0.37	0.14
SH1B	26.64	22.16	7.48	0.00	1.05	1.90	5.36	2.34	0.98	12.93	1.60	1.12
SH1C	20.03	18.39	5.92	13.15	1.06	1.51	4.27	1.88	0.81	30.68	1.27	1.03
SH1D	22.84	19.03	5.11	12.45	0.65	1.20	3.91	1.55	0.71	29.98	1.27	1.29
Range	7.69	3.76	2.38	13.15	0.41	0.69	1.51	0.80	0.98	21.84	1.23	1.15
Gap_{high}	3.80	3.12	1.25	0.62	0.02	0.37	1.09	0.44	0.17	4.09	0.33	0.17
Gap_{low}	1.08	0.61	0.82	12.45	0.08	0.30	0.06	0.33	0.71	17.05	0.91	0.88
$Qcalc_{high}$	0.49	0.83	0.53	0.05	0.04	0.54	0.72	0.56	0.17	0.19	0.27	0.15
$Qcalc_{low}$	0.14	0.16	0.34	$\boxed{0.95}$	0.20	0.44	0.04	0.41	0.72	$\boxed{0.78}$	0.73	$\boxed{0.77}$

	dimethyl benzene	furfural	methyl cyclopentenone	trimethyl benzene	benzaldehyde	benzofuran	benzonitrile	phenyl ethanone	methyl benzonitrile	phenol	methylated indene	tetradecene
SH2A	14.23	10.54	4.76	4.41	1.06	1.21	2.61	1.45	0.00	59.30	0.18	0.26
SH2B	13.61	9.37	4.31	4.29	1.01	1.19	2.73	1.44	0.44	60.22	0.61	0.78
SH2C	14.15	11.18	4.97	4.45	1.11	1.25	2.80	1.51	0.45	56.44	0.80	0.89
SH2D	28.42	16.24	6.73	6.67	1.44	1.66	3.52	1.13	0.75	30.55	1.67	1.22
Range	14.82	6.86	2.41	2.38	0.43	0.47	0.91	0.38	0.75	29.67	1.49	0.96
Gap_{high}	14.20	5.06	1.76	2.22	0.32	0.40	0.79	0.06	0.30	0.92	0.87	0.33
Gap_{low}	0.54	1.17	0.44	0.12	0.05	0.02	0.12	0.31	0.44	25.88	0.43	0.52
$Qcalc_{high}$	$\boxed{0.96}$	0.74	0.73	$\boxed{0.93}$	0.75	$\boxed{0.86}$	$\boxed{0.87}$	0.16	0.40	0.03	0.58	0.35
$Qcalc_{low}$	0.04	0.17	0.18	0.05	0.11	0.04	0.13	$\boxed{0.82}$	0.58	$\boxed{0.87}$	0.29	0.54

Table E-2: Q-test for Soil Leachates (cont'd).

	dimethyl benzene	furfural	methyl cyclopentenone	trimethyl benzene	benzaldehyde	benzofuran	benzonitrile	phenyl ethanone	methyl benzonitrile	phenol	methylated indene	tetradecene
UBSA	16.08	23.31	7.61	4.20	2.35	2.30	4.21	3.51	0.97	33.99	0.34	1.14
UBSB	14.89	19.45	6.17	3.70	2.24	2.19	4.41	3.48	0.62	39.98	1.17	1.71
UBSC	16.38	22.07	8.03	3.90	2.12	2.08	3.82	3.24	0.87	34.33	1.19	1.98
UBSD	16.52	20.94	8.06	3.76	2.12	2.08	3.98	3.32	0.88	34.74	1.35	2.26

Range	1.63	3.86	1.89	0.51	0.23	0.21	0.59	0.27	0.35	5.99	1.01	1.12
Gap$_{high}$	0.14	1.24	0.03	0.31	0.12	0.10	0.20	0.03	0.09	5.24	0.16	0.27
Gap$_{low}$	1.19	1.49	1.44	0.06	0.12	0.11	0.16	0.08	0.25	0.34	0.83	0.57
Qcalc$_{high}$	0.08	0.32	0.02	0.61	0.50	0.49	0.34	0.10	0.26	0.87	0.16	0.24
Qcalc$_{low}$	0.73	0.39	0.76	0.12	0.52	0.51	0.27	0.30	0.72	0.06	0.82	0.51

	dimethyl benzene	furfural	methyl cyclopentenone	trimethyl benzene	benzaldehyde	benzofuran	benzonitrile	phenyl ethanone	methyl benzonitrile	phenol	methylated indene	tetradecene
2UBSA	18.08	19.37	8.96	4.76	2.31	1.91	3.07	3.59	0.62	36.19	0.39	0.74
2UBSB	17.93	18.19	9.01	4.65	2.28	1.89	3.04	0.00	0.60	36.40	1.29	1.21
2UBSC	17.94	22.23	11.54	4.24	2.18	1.64	2.64	3.29	0.99	30.37	1.23	1.71
2UBSD	18.07	22.51	11.88	4.45	2.32	1.65	2.68	3.37	1.02	29.06	1.27	1.72

Range	0.15	4.31	2.92	0.52	0.15	0.27	0.43	3.59	0.43	7.35	0.90	0.98
Gap$_{high}$	0.01	0.28	0.34	0.11	0.01	0.02	0.03	0.22	0.03	0.21	0.02	0.01
Gap$_{low}$	0.01	1.18	0.06	0.20	0.11	0.01	0.04	3.29	0.02	1.32	0.83	0.47
Qcalc$_{high}$	0.09	0.07	0.12	0.21	0.05	0.08	0.07	0.06	0.07	0.03	0.02	0.01
Qcalc$_{low}$	0.05	0.27	0.02	0.40	0.73	0.04	0.10	0.92	0.05	0.18	0.93	0.48

	dimethyl benzene	furfural	methyl cyclopentenone	trimethyl benzene	benzaldehyde	benzofuran	benzonitrile	phenyl ethanone	methyl benzonitrile	phenol	methylated indene	tetradecene
HLBA	16.66	5.35	5.39	38.44	1.64	0.84	2.07	3.09	0.82	24.97	0.28	0.46
HLBB	26.31	9.27	10.47	6.75	2.86	1.50	3.37	5.98	0.84	30.83	0.55	1.27
HLBC	23.68	7.80	8.42	5.87	2.42	1.30	3.10	5.26	0.72	39.97	0.47	0.99
HLBD	68.38	0.00	15.72	11.94	0.00	0.00	0.00	0.00	0.00	0.00	3.96	0.00

Range	51.72	9.27	10.32	32.56	2.86	1.50	3.37	5.98	0.84	39.97	3.68	1.27
Gap$_{high}$	42.07	1.47	5.24	26.49	0.44	0.20	0.27	0.72	0.02	9.14	3.41	0.28
Gap$_{low}$	7.02	5.35	3.02	0.88	1.64	0.84	2.07	3.09	0.72	24.97	0.27	0.46
Qcalc$_{high}$	0.81	0.16	0.51	0.81	0.15	0.13	0.08	0.12	0.02	0.23	0.93	0.22
Qcalc$_{low}$	0.14	0.58	0.29	0.03	0.57	0.56	0.61	0.52	0.86	0.62	0.07	0.37

Table E-2: Q-test for Soil Leachates (cont'd).

	dimethyl benzene	furfural	methyl cyclopentenone	trimethyl benzene	benzaldehyde	benzofuran	benzonitrile	phenyl ethanone	methyl benzonitrile	phenol	methylated indene	tetradecene
2HLBA	21.23	11.87	10.29	5.47	2.38	1.19	3.27	4.34	0.81	37.89	0.65	0.63
2HLBB	21.32	11.87	10.19	5.44	2.37	1.18	3.32	4.38	0.79	38.01	0.65	0.48
2HLBC	20.38	12.15	10.64	5.41	2.47	1.20	3.25	4.45	0.79	37.98	0.65	0.65
2HLBD	45.95	5.32	16.17	8.82	0.00	0.00	2.89	3.15	1.66	8.03	6.93	1.07

Range	25.57	6.83	5.98	3.41	2.47	1.20	0.42	1.30	0.88	29.98	6.28	0.59
Gap_{high}	24.63	0.28	5.53	3.35	0.09	0.01	0.05	0.07	0.86	0.03	6.28	0.43
Gap_{low}	0.85	6.55	0.10	0.03	2.37	1.18	0.35	1.23	0.01	29.85	0.00	0.15
$Qcalc_{high}$	0.96	0.04	0.92	0.98	0.04	0.01	0.11	0.05	0.98	0.00	1.00	0.72
$Qcalc_{low}$	0.03	0.96	0.02	0.01	0.96	0.98	0.83	0.95	0.01	1.00	0.00	0.25

	dimethyl benzene	furfural	methyl cyclopentenone	trimethyl benzene	benzaldehyde	benzofuran	benzonitrile	phenyl ethanone	methyl benzonitrile	phenol	methylated indene	tetradecene
BIRA	16.35	14.72	5.05	5.81	1.30	1.73	1.91	2.15	0.99	47.25	1.79	0.96
BIRB	19.36	16.42	5.67	6.15	1.17	1.77	1.94	2.33	1.03	41.30	1.56	1.30
BIRC	18.12	19.60	5.51	0.58	1.36	1.78	1.95	2.24	0.28	45.65	1.76	1.18
BIRD	20.43	26.08	7.09	5.20	1.07	2.02	2.04	2.33	1.02	28.24	2.74	1.74

Range	4.09	11.37	2.04	5.57	0.29	0.29	0.13	0.19	0.74	19.00	1.17	0.77
Gap_{high}	1.08	6.48	1.41	0.35	0.06	0.23	0.09	0.00	0.01	1.60	0.95	0.43
Gap_{low}	1.77	1.70	0.46	4.62	0.10	0.04	0.03	0.09	0.71	13.05	0.19	0.21
$Qcalc_{high}$	0.26	0.57	0.69	0.06	0.19	0.81	0.69	0.00	0.01	0.08	0.81	0.56
$Qcalc_{low}$	0.43	0.15	0.23	0.83	0.34	0.13	0.24	0.48	0.95	0.69	0.16	0.27

	dimethyl benzene	furfural	methyl cyclopentenone	trimethyl benzene	benzaldehyde	benzofuran	benzonitrile	phenyl ethanone	methyl benzonitrile	phenol	methylated indene	tetradecene
SPRA	25.61	10.91	10.84	5.65	2.97	1.76	5.71	4.02	0.89	27.38	2.17	2.08
SPRB	24.55	10.89	10.26	5.20	2.56	0.85	5.31	3.61	0.78	28.75	4.34	2.90
SPRC	20.94	9.20	8.86	4.39	2.03	0.72	4.41	3.11	0.67	42.58	1.61	1.47
SPRD	54.82	5.70	20.85	9.42	0.00	0.00	9.22	0.00	0.00	0.00	0.00	0.00

Range	33.87	5.22	12.00	5.03	2.97	1.76	4.81	4.02	0.89	42.58	4.34	2.90
Gap_{high}	29.21	0.02	10.01	3.77	0.41	0.92	3.51	0.41	0.11	13.83	2.17	0.82
Gap_{low}	3.61	3.50	1.40	0.81	2.03	0.72	0.90	3.11	0.67	27.38	1.61	1.47
$Qcalc_{high}$	0.86	0.00	0.83	0.75	0.14	0.52	0.73	0.10	0.12	0.32	0.50	0.28
$Qcalc_{low}$	0.11	0.67	0.12	0.16	0.68	0.41	0.19	0.77	0.76	0.64	0.37	0.51