Evaluation of DNR Aquatic Vegetation Surveys: Data Summaries and Comparative Analysis

Carol Reschke, George E. Host and Lucinda B. Johnson
University of Minnesota Duluth
Natural Resources Research Institute
5013 Miller Trunk Highway
Duluth, MN 55811

March 16, 2006 update

Submitted to:

Ms. Donna Perleberg, Project Officer
Minnesota Department of Natural Resources

NRRI/TR-2006/32

In fulfillment of:

Minnesota Department of Natural Resources
CFMS Contract # A61156

Table of Contents

Background ……………………………………………………….…………..p. 1

Research Rationale and Objectives ……………………………………………p. 2

Methods ……...……………………………..…………………….………..p. 3
    Data Manipulations ....……………………..…………….……………p. 4
    Spatial Analysis Methods …….……..……………………………..p. 9
    Community Analysis Methods .. ………………………………………p. 10
    Data Integration ..………………………………………………………p. 11

Results and Discussion ...…………………………………………………p. 12
    Exploratory Data Analyses ……………………………………………p. 12
    Community Analyses ………………………………………………p. 17
    Comparison of Plant Community Types (PC Types) …………………………p. 21
    Aquatic Plant Community Groups ………………………………….……p. 21
    Aquatic Plant Communities and Schupp Lake Types ………………………p. 26
    Aquatic Vegetation Response to Disturbance ………………………………p. 29
    Spatial Trends in Aquatic Vegetation Community Types …………………..p. 30
    Vegetation Response to Landscape and Local Variables ..…………………p. 30
    Comparing Sampling Methods: Species Richness …………………………p. 37
    Exotic Species as Indicators of Disturbance ...………………………………p. 40
    Data Integration … …………………………………………………………p. 41

Recommendations ………………………………………………………………p. 42

References ……………………………………………………………………..p. 45

TABLES

Table 1: Dataset characteristics ……………………………………………….…p. 8
Table 2: Number of lakes, land use within lake buffers, and Aquatic Vegetation diversity metrics, summarized by Omernik Level III Ecoregions..……..…p. 13
Table 3: Number of lakes, land use within lake buffers, and Aquatic Vegetation Diversity metrics, summarized by DNR ECS Sections ………………..p. 13
Table 4: Average species richness (S') by Omernik Ecoregion for each dataset .….p. 15
Table 5: Average species richness (S') by DNR ECS Section for each dataset …...p. 16
Table 6: Aquatic Plant Community Groups, and the plant community types included within those groups . ………………………………………….…p. 25
Table 7: Distribution of ALLKS aquatic plant community types by Schupp Lake Class ………………………………………………………………...…p. 28
Table 8: Pearson and Kendall Correlations between Environmental Variables and Ordination Axes for the FISH dataset ……………………………………..p. 31

Table 9: Pearson and Kendall Correlations between Plant Taxa and Ordination Axes for the FISH dataset……………………………………………………..p. 32
Table 10: Pearson and Kendall Correlations between Environmental Variables and Ordination Axes for the ECO dataset…………………………………..p. 33
Table 11: Pearson and Kendall Correlations between Plant Taxa and Ordination Axes for the ECO dataset……………………………………………………..p. 34
Table 12: Pearson and Kendall Correlations between Environmental Variables and Ordination Axes for the ALLKS dataset ………………………………….p. 35
Table 13: Pearson and Kendall Correlations between Plant Taxa and Ordination Axes for the ALLKS dataset ..…………………………………………….p. 36
Table 14: Numbers of lakes for each combination of sampling effort, for the 120 lakes sampled by two or more methods ……………………………………..p. 37
Table 15: Species richness for lakes sampled with two or more methods …………p. 38
Table 16: Exotics in Aquatic Vegetation Datasets ..……………………………..p. 40

FIGURES

Figure 1: Frequency distribution of Schupp Lake Classes . ………………………..p. 27

APPENDICES

(* indicates appendices that are not printed; instead they are attached as separate spreadsheet files. See page 67 for file names for these appendices.)

Appendix A: **Data Editing and Manipulations…**……………………..…………p. 47

Appendix B: **Derived Data Calculations…**………………………………….…..p. 56

Appendix C: **Final List of Plant Taxa…**………………….………………..……..p. 59

Appendix D: ***Taxa presence/absence in each lake by dataset**.

Appendix E: ***Composition of each of the 41 plant community types**.

Appendix F.1: ***Composition of Plant Community Types in the FISH dataset**.

Appendix F.2: ***Composition of Plant Community Types in the ECO dataset**.

Appendix F.3: ***Composition of Plant Community Types in the WJL dataset**.

Appendix F.4: ***Composition of Plant Community Types in the WPI dataset**.

Appendix F.5: ***Composition of Plant Community Types in the ALLKS dataset**.

Appendix F.6: **Composition of Aquatic Plant Community Groups.**

Appendix G: **Species Richness (S') for each Lake by Dataset**

Appendix H.1: **ECO Ordination Graphs.**

Appendix H.2: **FISH Ordination Graphs.**

Appendix H.3: **WJL Ordination Graphs.**

Appendix H.4: **WPI Ordination Graphs.**

Appendix H.5: **ALLKS Ordination Graphs and PCTYPES Ordination Graph.**

Appendix **Map Plates**

Background

The composition of aquatic vegetation in lakes provides an index of ecological integrity, and potentially a diagnostic tool to identify causes of disturbance or impairment. Aquatic vegetation in Minnesota lakes is monitored by several programs within the Minnesota Department of Natural Resources (DNR), in the Divisions of Fish and Wildlife, and Ecological Services. Protocols for vegetation sampling among the programs differ, however, based on programmatic objectives.

As of 2003**,** approximately 2407 lakewide vegetation surveys in 2176 lakes have been conducted in Minnesota using either a semi-quantitative "modified belt-transect" method (Anon. 1993) or a quantitative point-intercept method with sample points placed along transects. DNR Fisheries staff conducted lakewide surveys using the modified belt-transect method on over 1500 lakes since 1992. The Shallow Lakes Program in the Division of Fish and Wildlife has collected data on 570 lakes since the 1940's using two types of point samples. The first is called the Jessen-Lound method, it uses 4 rake tosses off the corners of a boat at each sample point (DNR Wildlife – Jessen and Lound 1962). The second method used by the Shallow Lakes program is the Point Intercept method, which employs a single rake toss at each point. The Ecological Services Division's Aquatic Plant Ecology program has conducted surveys in 60 lakes since 1999 using the point intercept method. The data compiled from these programs are summarized in Table 1. DNR's Minnesota County Biological Survey Program (MCBS, also in the Division of Ecological Services) also has collected detailed species list data from portions of about 1225 lakes, primarily in northern Minnesota.

Newman (1998) and Middleton (1998) reviewed data collected by DNR Wildlife and Perleberg (2001a) reviewed data collected by DNR Fisheries. These reviews primarily provide recommendations for revised sampling approaches as opposed to plant community comparisons. Perleberg (2001b) compared the "modified belt-transect method" used by DNR Fisheries (Anon. 1993), the transect method used by DNR Wildlife (Jessen and Lound 1962), and a point-intercept method (Madsen 1999) and suggested that while the species list information was useful for comparison, the frequency and abundance data collected by these different methods were not directly comparable. Perleberg (2003) also summarized species list information and calculated diversity indices from the Fisheries and Wildlife data. Further work in this area is needed, however, for the following reasons:

a) Additional species lists are now available, including about 1225 from northern Minnesota lakes where MCBS has conducted rare species searches.
b) Perleberg (2003) looked at the species list data and did not compare plant community data with environmental data. Nichols (1999), however, did conduct such an analysis of Wisconsin aquatic plant data.
c) "Community type profiles" can be created from existing data, regardless of whether databases can be integrated.

Based on the reviews by Newman (1998) and Perleberg (2001), and because GPS is now readily available to field staff, Wildlife and Ecological Services determined there is a need and ability to collect more quantitative data on at least a subset of lakes. The method selected is a point-intercept survey with sample sites placed at evenly spaced distances along a grid pattern across an entire lake (Madsen 1999).

Research Rationale and Objectives

With increasing concerns about the status of aquatic vegetation beds resulting from the increased rate of shore land development, there is a need to:

- summarize and communicate results from existing data;
- develop statistical approaches for analyzing data collected using different sampling protocols;
- develop a strategy for identifying data gaps to guide future sampling efforts.

To fulfill these goals, we were asked to assemble and synthesize data across the various monitoring programs. The project had the following specific objectives:

1. **Assemble the following data sets in a common format and database:**
   a. DNR Fisheries Lake Survey Program, semi-quantitative modified belt-transect data with estimates of abundance that cannot be directly compared to “b” below.
   b. DNR Wildlife Shallow Lakes Program, a long-term program that has collected data using two distinct methods including:
      - Jessen-Lound, with abundance ranks of species in lakes
      - Point-intercept, with species lists and frequency data.
   c. DNR Ecological Services, a database of point-intercept data with species lists and frequency data.
   d. DNR Ecological Services, Minnesota County Biological Survey (MCBS), a database with detailed species lists primarily covering Northern MN lakes.

   A key outcome of this objective was an assessment of whether data could be analyzed as a single comprehensive data set, or whether they needed to be treated independently.

2. **Compile available water quality, clarity, and chemistry data for a subset of lakes; compare with plant community data using correlation and regression techniques; compile trends for existing variables.**

3. **Create basic data summaries by relevant geographic regions and lake types.**
   a. Examine differences in species richness, diversity, and composition across MN lakes.
   b. Describe the large-scale patterns in aquatic plant communities across Minnesota. Important levels of geographic stratification include Omernik ecoregions and DNR Ecological Classification System (ECS) section**,** Moyle’s three plant zones, Schupp’s lake classification.
   c. Determine whether physical and/or chemical characteristics of select waterbodies are useful for modeling/describing aquatic plant communities.

4. **If possible, conduct a separate analysis of the data from lakes sampled with the point-intercept data approach**.
   Determine whether this dataset is sufficient for statewide analyses of plant communities and if not, provide recommendations to address deficiencies.

5. **Assemble list of potential indicators**
   Indicator variables such as % exotics can be summarized. Presence or dominance of exotic species of concern (e.g., coverage of Eurasian water milfoil; curly-leafed pondweed) can be displayed as an overlay on ordinations to determine correlations with vegetation patterns.

6. **Use spatial data to assess level of disturbance around lakes, and determine relationships between level of disturbance and vegetation pattern**
   A number of publicly available data sets can be use to quantify levels of disturbance or impairment. Among these are land use/ land cover, population density, road density, and locations of point source discharges.

7. **Evaluate the extent to which the level of disturbance explains the remaining variance in the various large-scale patterns identified in 2**. How important are local conditions in influencing the expression of the large-scale pattern? What are the strongest patterns/associations? What disturbance variables are most valuable in explaining differences among plant communities?

8. Assess trends in indicators and community composition across the composite (normalized) data set. Identify set(s) of conditions where invasive, non-native plants are most prevalent.

9. Based on discussions of data quality with DNR staff, test regional data normalization techniques (e.g., Wiley et al. 2003) to compare trends across quantitative data sets generated using different sampling methods.

10. Test stressor-response relationships using spatial data from (4), and indicators from (7).

11. For the strongest patterns/associations described above (both at the intra-lake and inter-lake scales), develop data visualization and delivery tools for agency personnel and the public.

12. Examine the geographic distribution of data and recommend future sampling locations relative to ecoregions, lake types, sensitive regions, etc.

Due to the very large time task of assembling the data sets and conducting an extensive quality control procedure, not all objectives were fully addressed. This report concentrates on Objectives 1-.7.

Methods

The original project objectives called for assessment of five aquatic vegetation datasets; however, the MN County Biological Survey (MCBS) data were not available for analysis, so the current study focuses on four datasets: Fisheries (FISH), Ecological Services Aquatic Ecology

Program (ECO), Wildlife Jessen-Lound (WJL), and Wildlife Point-Intercept (WPI) (Table 1). Extensive quality assurance and data editing were required, especially for the FISH and WPI datasets, before the data could be assembled for analysis. The vegetation data from each dataset were copied into four separate Access Databases. Environmental and GIS-derived data were compiled into a single database for all 2042 lakes in this study. We have assembled four vegetation databases and one environmental database that contain the raw data and queries used to edit data (Objectives 1, 2**)**. These databases were used to produce data matrices used in quantitative analyses. Databases were assembled with the goal of permitting species data to be summarized by: 1) presence/absence, 2) relative frequency, 3) average or maximum abundance, and 4) importance value (an algorithm that combines relative frequency and abundance).

Data Manipulations

Data editing, database manipulations, and derived data calculations are summarized below for each dataset and described fully in Appendices A and B.

1. Fisheries Dataset (FISH)

Raw data were edited to enable species abundance data from multiple transects in a lake that were stored in multiple data records to be summarized into a single data record for each lake (see Appendix A.1 for a detailed description of the editing procedures). A new database was created with the original Division of Water (DOW) lake number, vegetation code, and sample date fields, columns for each transect sampled, and sorting codes. Data for each lake were then summarized by transects and species to yield a single value for each species in a lake. Data from 1669 lake surveys conducted from 1992 through 2003 were compiled, and then these were sorted by lake number and sample date. The most recent year's data for each lake were selected for the analyses, and earlier samples removed (for lakes with multiple sample years). Lakes with fewer than 9 transects sampled were also eliminated from the analysis dataset due to their small sample size. The final dataset included vegetation data on 1500 lakes.

After the raw data were compiled into a matrix of species by lakes, the resulting species list was reviewed. There were 220 species present in the original data set. Species nomenclature was reviewed and verified using the USDA PLANTS online database as a reference (http://plants.usda.gov/). The species list was then reviewed in consideration of ease of field identification, since no voucher specimens were collected for this dataset. Species that are difficult to identify in the field were combined into a taxon representing a genus or group of similar-looking species. These revised taxa were then assigned new computer codes we called "NRRILUMP" codes. Species with multiple synonyms were combined into the currently accepted name. This reduced the dataset to 136 taxa in 1500 lakes (see Appendices A.1 and C).

Species or taxa that are rare in the dataset can introduce statistical "noise" that can make analytical results confusing; therefore, species or taxa rarely observed were removed from the dataset. Species were sorted by raw frequency (total number of lakes in which the species was observed present). Species present in fewer than 16 lakes (1.07 % of the lakes) were eliminated from the classification and ordination analyses (see Appendix A.1 for further details). The final dataset included 103 taxa in 1500 lakes. The following derived variables were calculated: 1) presence/absence; 2) relative frequency, 3) average abundance, and 4) importance value (see Appendix B.1 for details).

2. Ecological Services Dataset (ECO)

The Ecological Services "Aquatic Vegetation Database" includes data sampled using point-intercept sampling methods in 60 lakes (Table 1). This database is well organized and nearly error-free. Sampling was conducted from 2000-2004. Tables of lakes by species (raw frequency) were easily produced using the crosstab function in Access; tables were then exported to Excel for editing (see Appendix A.2). The database includes 66 surveys in 60 lakes, mostly located in north-central Minnesota. There are a total of 87 species or taxa recorded in the database. Six lakes were sampled twice; earlier samples were removed from the dataset. To allow comparison with other datasets, the 87 species were assigned NRRILUMP codes to combine taxa that are difficult to distinguish in the field. This was necessary even though we trust the species identifications in the ECO dataset, because we want to compare between datasets. When species were lumped, the dataset consisted of 60 lakes and 73 taxa.

3. Wildlife Jessen-Lound Dataset (WJL)

The Shallow Lakes database, created by the DNR Division of Wildlife, includes vegetation data collected using two sampling methods: Jessen-Lound and point-intercept. These data are supposed to be in separate tables in the Wildlife Lakes Survey database (the file name is "Wildlake2000"). Data from Jessen-Lound samples are stored in the table called "Data Vegetation" in Wildlake2000. A total of 507 aquatic vegetation samples were collected by DNR Wildlife staff from 1946 to 1994 using the Jessen-Lound method (Appendix A.3).

In Jessen-Lound sampling, the sample area consists of four sub-samples, taken at each corner of the boat. Each sub-sample consists of a visual survey and single rake toss which covers an area approximately one square meter. Some of the very old surveys from the 1940's contained only species lists. At most sample sites, a species abundance rank was assigned based on the number of sub-samples in which a species is observed. (See Appendix B.3 for further details.) Of 7138 species samples in the WJL dataset, 182 had no abundance rating, therefore, an abundance of 0 was assumed. The data in the "Data Vegetation" table include only a single abundance rating for each species in each lake. In a summary data matrix provided by Donna Perleberg, the number of points sampled per lake is reported as 10 to 40. So someone must have summarized original field data from elsewhere to boil it down to one abundance number for each species in each lake. It would be better to have the original data so we could calculate both relative frequency and average abundance. But without the original data, the best we can do is use these single abundance numbers (one abundance per species per lake) and assume that whoever summarized the data did it correctly. Since we don't know how these single abundance values were compiled from the original data, it is safer to use these to create a presence/absence dataset.

The abundance data were compiled using a crosstab query in the Access database to create a matrix of lakes by species, showing the abundance value recorded for each species per lake. This matrix was then used to create a presence/absence dataset, converting all values >0 to 1. Some lakes were sampled in multiple years; only the most recent data were included in our analysis (with 3 exceptions explained in Appendix B.3.) This presence/absence dataset was then modified by applying a "Beals smoothing" algorithm, also known as the sociological favorability index. Beals smoothing replaces each cell in the data matrix with a probability of each taxon occurring in a sample unit, based on the joint occurrences of the target taxon with the other taxa that are actually present in the sample unit. This algorithm helps solve a problem with a statistical distribution of the data known as the "zero truncation problem", in which there are a large number of zeroes in the dataset. This algorithm replaces presence/absence data with quantitative values that represent the "favorability" of each sample for each plant taxon, regardless of whether the taxon was actually present in that sample (McCune and Grace, 2002). The final dataset consisted of 425 lakes and 83 taxa (Table 1).

4. Wildlife Point-Intercept Dataset (WPI)

The Shallow Lakes database, created by the DNR Division of Wildlife, includes vegetation data collected using two sampling methods: Jessen-Lound and Point-Intercept. These data are supposed to be in separate tables in the Wildlife Lakes Survey database (the file name is "Wildlake2000"). Data from Point-Intercept samples are stored in the table called "Data Sample Stations Vegetation" in Wildlake2000; however a copy of the entire WJL dataset is also included in the same table. The WJL data are stored with the "DSSV-station" field entered as "999". Data were sorted on the field "DSSV-station", and all records with "999" in the "DSSV-station"field were removed. This left a total of 165 lake samples that were recorded using the point-intercept sampling method (Table 1). Sampling was conducted from 1988 to 2003. Five lakes were sampled in multiple years; only the most recent data were included in the analysis. In point-intercept sampling, the sample area consists of a visual survey and single rake toss which covers an area of approximately one square meter. The point intercept sampling method was refined in more recent years to include an evenly spaced grid of sample points across a lake, where empty sample points are also recorded. Some of the older samples do not record "empty" points where no aquatic vegetation was found.

The NRRILUMP codes were added to the dataset, then the raw frequency or number of observations of each plant taxon was tabulated by using a crosstab query in Access (Appendix B.4). The query was set up so that lakes were in rows, plant taxa were in columns, and the cells in the table calculated a count of all the DSSV stations (sample points) where each species or taxon was observed. This yielded a table of 165 lakes and 76 plant taxa. The table was then sorted by the NRRILUMP codes and data examined for how to combine the records of taxa that needed to be lumped (see Appendix A.4 for details). Raw frequency values were combined as appropriate, leaving a table of 165 lakes and 64 taxa.

Some lakes were sampled with very few points, others with hundreds of points. In addition, some lakes were sampled in multiple years; only the most recent data were included in our analysis. Some lakes were empty after rare taxa and wetland edge taxa were removed, so three more lakes were deleted. The dataset then consisted of 157 lakes and 64 taxa. Twelve lakes with fewer than 7 points sampled were deleted for their small sample size. The WPI dataset then consisted of 145 lakes and 64 taxa. When this dataset was analyzed with NMS, it was not possible to find a stable solution (McCune and Medford 1999). So an outlier analysis was run on the 145 lakes dataset, and seven outlier lakes were identified that were greater than 2.5 standard deviations from the mean. These seven lakes were removed and one additional taxon was then removed because it was no longer present in the dataset. The final dataset used in the analyses consisted of 138 lakes and 63 taxa (see Appendix A.4 for more details).

5. Combined Point-Intercept Dataset (ECOWPI)

The relative frequency data from the two datasets sampled with the point-intercept method (ECO and WPI) were combined into one dataset that consisted of 205 lakes and 83 taxa. This combined dataset was analyzed using the same procedures used for those datasets individually: a flexible beta classification, and a series of NMS ordinations.

6. All Lakes from Four Datasets Combined (ALLKS)

Each of the four datasets was either compiled as presence/absence data, or converted to presence/absence data using a power transformation of x^0 in PC-ORD. The four datasets were then combined into one composite presence/absence dataset that consisted of 1994 lakes and 131 taxa. These presence/absence values were modified using the “Beals smoothing” algorithm used previously with the WJL data. After running one set of ordinations using the Beals smoothing data, it was clear that there were some lakes that were statistical outliers in the dataset. These outliers cause the rest of the lakes to be clumped together in the ordination, making it difficult to interpret. An outlier analysis was run in PC-ORD and ten outlier samples were identified that were more than 4 standard deviations from the mean of the combined dataset. These outliers were removed and the remaining 1984 lakes and 130 taxa were included in the analysis.

7. Plant Community Types Dataset (PC Types)

The plant community types (PC Types) identified in the analyses of five datasets (ECO, FISH, WJL, WPI, and ALLKS) were used to prepare a dataset that would allow comparison of the types recognized by the different datasets. This is an alternative approach to looking at patterns across all four datasets. This dataset was prepared by converting the data in each dataset separately to presence/absence data, and sorting the presence/absence data in each dataset by the PC Type that the lake was assigned in the dataset’s flexible beta classification (or the modified version for WJL). Then a summary composition of each PC Type was prepared based on relative frequency of each taxon across all the lakes in a single PC Type. This produced, for each PC Type, a relative frequency of each taxon in each dataset. Since these relative frequency values were all derived from presence/absence data, they are comparable between datasets. A summary dataset of the composition of all 41 PC Types was prepared, it had 41 PC Types and 137 taxa (relative frequency values across multiple lakes in the type).

This PC Type dataset was analyzed using the same classification and ordination procedures described above for each of the original four datasets.

Table 1. Dataset characteristics. Dataset abbreviation is shown in parenthesis.

Source	Initial # Lake Surveys	Final # Lakes (repeated years and small sample sizes removed)	Initial # Taxa	Final # Taxa	Survey Period	Survey method	Original Data type	Derived Data
Fisheries (FISH)	1669 (in 1510 lakes)	1500	220	103	1992-2003	Transects (6-52 per lake)	Qualitative abundance: None (0) Rare (1) Common (3) Abundant (5)	Average abundance; Relative frequency; Importance value
Ecological Services (ECO)	66 (in 60 lakes)	60	87	73	2000-2004	Point Intercept (Single rake-toss) (79 – 2129 per lake)	Presence/ absence	Relative frequency
Shallow Lakes (WJL)	507 (in 446 lakes)	425	127	83	1946-1994	Jessen-Lound (Four rake-tosses) (~10-40 per lake)	Semi-quantitative: Abundance None (0) 1 toss (1) 2 tosses(2) 3 tosses (3) 4 tosses (4) 4 tosses, full (5)	Presence/absence; Beals Smoothing value
Shallow Lakes (WPI)	165 (in 160 lakes)	138	75	63	1988-2003	Point Intercept (Single rake-toss) (1-305 per lake)	Presence/ absence	Relative frequency

Spatial Analysis Methods

GIS database development

One of the key products of this study was an integrated spatial database that stores lake polygons from the different sampling regimes, along with ancillary data both within the lake and from the surrounding landscape. This composite coverage was particularly useful for Objectives 2 and 6, which assessed the degree to which aquatic vegetation corresponds to ecological regions of the state, and the response of aquatic vegetation to disturbance and surrounding land use.

Ecoregional distributions of aquatic vegetation

There are currently two major ecological classifications used in the state. The Ecological Classification System (ECS) created and maintained by DNR is widely used in forest planning and wildlife management. This classification divides the state into 4 Provinces, 11 Sections, and 25 Subsections. The alternative classification is the Omernik Level III Ecoregion classification, which recognizes seven ecoregions. [The Omernik classification is used by the MN Pollution Control Agency.] We used the composite lakes database to determine if there were correlations between the composition of aquatic communities in lakes and ecological land units.

Quantifying disturbance in the lake environment

Objectives 6 and 7 were to quantify the degree of anthropogenic disturbance in the immediate environment of the lakes and assess the impact on aquatic vegetation. We created 1000 m (1 km) buffers around each lake based on the composite lake polygon GIS coverage. We then summarized several spatial data layers useful for quantifying anthropogenic stress within each lake buffer. The data sets employed here have previously been used to identify reference conditions and environmental stress gradients for coastal wetlands and high energy shorelines of the Great Lakes (Host et al. 2005). The National Land Cover Dataset (NLCD) is a Landsat Thematic Mapper-based classification of land-use/land cover at a 30 m pixel resolution. Data were summarized to a modified Anderson Level II land use classification and proportional land use within the 1000 m buffer was calculated for each class. Using NLCD data we also calculated two summary land cover classes that consisted of a combination of NLCD classed to illustrate the distinction of natural lands versus developed or agriculture lands. Roads from the USGS TIGER database were clipped to the lake buffers, and road densities summarized on a length/unit area (km/km^2) basis. Population density from US Census block data were similarly summarized within buffers. Point source pollution data were derived the EPA BASINS data catalogue, which contains locations of discharge permits within the EPA's National Pollutant Discharge Elimination System (NPDES). These were tabulated as a point source density – the number of point sources per unit area within the buffer region.

We used 2004 lake water quality data compiled by the Minnesota Pollution Control Agency (MPCA), from their website: http://www.pca.state.mn.us/water/lakequality.html#reports for the following environmental parameters for each lake:

1. Mean Secchi disk depth (m)
2. Alkalinity (ppm)
3. Total Phosphorus (ppb)
4. Mean Chlorophyll a (ppb)
5. Trophic State Indices (based on Total P, Chlorophyll A, and Secchi disk)
6. Color (Platinum Cobalt units) 9

Community Analysis Methods

Classification

Several classification methods were tested with the FISH dataset (the largest of the four). The cluster analysis program in PC-ORD performs eight variants on the general class of cluster analyses that are hierarchical, agglomerative, and polythetic. In this context, "hierarchical" means that large clusters are composed of smaller clusters. "Agglomerative" means that the analysis proceeds by joining clusters rather than by dividing clusters. “Polythetic" means that many attributes of the items are used to decide the optimum way to combine or divide clusters (McCune and Mefford 1999).

The best classification results for the FISH dataset were produced by the Flexible Beta Method, using Relative Sorensen distance measure, and beta set to -0.25, and the dendrogram scaled normally. The same classification routine was then used with the other three datasets. The classification produces a large dendrogram (hierarchical tree diagram) showing relationships between samples and groups of samples, and it can assign group numbers to each sample at whatever level of the classification dendrogram is requested. We chose to print classification group numbers for a range of group levels, generally from 5 to 12 groups. Our assumption is that there are somewhere between 5 to 12 lake community types that could be usefully recognized in each dataset based on the aquatic vegetation in each lake. Each of the four datasets was analyzed using PC-ORD software (McCune and Mefford 1999).

Ordination and Overlays

Each dataset was analyzed using ordination methods in PC-ORD software (McCune and Medford 1999). Ordination is an ordering of samples along axes according to similarity of composition. Ordination produces a series of graphs displaying the samples as points arranged according to their positions along two axes at a time. The strongest compositional and environmental gradients are usually found in relation to the first axis, then the second axis, and so on. Samples close together in the graph have similar composition (of aquatic plants in this study), and samples far apart in the graph have very different composition. Several ordination methods were tested.

We used the Non-metric Multidimensional Scaling (NMS) ordination method because this method is well-suited to data that are non-normal or are on arbitrary, discontinuous, or otherwise questionable scales (McCune and Mefford 1999). NMS consists of an iterative search for a ranking and placement of samples on a specified number of axes that minimizes the stress of the selected configuration (McCune and Mefford 1999). “Stress” is a measure of departure from monotonicity in the relationship between the distances in the original dataset and the distances in the ordination space. Each dataset was analyzed using one of two methods. The first method was running the NMS program in “Autopilot” mode set to “medium” level of speed and thoroughness. The distance measure used was Relative Sorensen, and the number of axes chosen for calculation was 3. The starting configuration was random. The program in autopilot mode does 15 runs with real data and 30 runs with randomized data, with a maximum of 200 iterations. A second option was to initially run a Bray-Curtis ordination, with Relative Sorensen distance measure, variance-regression endpoint selection, Euclidean projection geometry, and Euclidean calculation of residuals. Then the axis coordinates from this Bray-Curtis ordination were used as a starting configuration for the NMS ordination (instead of a random start). The NMS was run with Relative Sorenson distance measure, 3 or 4 axes were calculated, and the program usually was set for 10 runs with real data and 20 runs with randomized data, with a maximum of 100 iterations. The first NMS run is evaluated for the stress and stability of the solution. In most cases a second NMS ordination was run using the number of axes and starting configuration recommended in the results from the first NMS ordination. Each dataset was analyzed until the most stable solutions with relatively low stress were found.

The ordination results are displayed in a graph of two axes at a time. Each symbol in the graph represents one lake, and the position of each lake represents its position in the aquatic plant compositional gradient represented by the axes shown. Patterns of abundance of individual species in the ordination graph can be displayed in overlays showing the relative importance or relative frequency of a species in the diagram. Environmental characteristics such as ecoregion location, surrounding landscape setting, and water chemistry can be displayed as overlays on the graph by color and size coding the symbols for each lake. These overlays are a useful tool for evaluating the importance of environmental variables as they influence vegetation composition. Overlays can also show characteristics of the aquatic vegetation community such as species richness, and the classification group to which each lake was assigned.

For each dataset, the overlays of a series of Flexible Beta classification groups (ranging from 6 to 9 groups) were reviewed to determine the level of classification (i.e., the number of groups) that produced the clearest set of sample groups in the ordination graph. That level was then chosen to describe lake aquatic plant community types for each dataset.

Data Integration

A fundamental task of the project was to assess the degree to which the above data sets could be combined, and to conduct summaries of individual and combined data sets (Objective 1). This consisted of: 1) editing vegetation datasets for consistent taxonomy, 2) summarizing taxonomic data, and 3) calculating derived data (e.g., species richness, diversity), 4) creating a composite database.

A complete list of taxa included in the four datasets was compiled (Appendix C). This list was reviewed to determine which taxa represented synonyms, and which taxa should be combined because they are difficult to accurately distinguish in the field. Using this list as a guideline, the individual datasets were reviewed to correct any duplicate species entries using taxonomic synonyms, and combining some taxa as we agreed upon. Species in each original dataset were changed either to newer nomenclature, combined into broader groups of similar-looking species, or deleted because they were considered wetland herbs or shrubs that don’t occur in open lake waters, or were too rare in the dataset to be included in the final analyses. This lumping of plant taxa is essential for being able to compare between datasets. A composite database of taxa by lake also was created (ALLKS). In addition, we created a composite table containing the derived variables (i.e., species richness and diversity) for all lakes and all databases (EnvGIS).

Results and Discussion

Exploratory Data Analyses

Exploratory data analyses included basic summaries of species/taxa counts, as well as species diversity metrics. Surveys were also summarized according to the geographic regions and lake type analyses described in the previous section; the geographic analyses presented in the following two sections are based on the FISH data set, which is the largest data set in terms of numbers of species and numbers of lakes sampled; this is followed by a comparative analysis of all data sets.

Analysis by Omernik Level III Ecoregions (FISH data set) (Objective 3)

Lakes are not uniformly distributed across the Omernik Level III Ecoregions (Map Plate 1). The great majority of lakes occur in the Northern Lakes and Forest or the North Central Hardwood Forest, which in the larger FISH data set contain 1303 and 937 lake samples, respectively (Table 2). The degree of human impact to lakes is strongly influenced by the type of land use adjacent to the lake. To assess this, we quantified land use within a 1km buffer surrounding each lake. There were strong differences in land use between these two lake-dominated ecoregions; lakes of the North Central Hardwood Forest had an average of 49% agricultural land use within the buffer, compared with 18% in forested land cover (Table 2). The opposite pattern occurred in the Northern Lakes And Forests Ecoregion, where lake buffers contained 56% forested land, and only 11% in agricultural land uses.

Metrics of species diversity, as expressed by species richness and the Shannon index, were higher in the Northern Lakes and Forests region (Table 2). In terms of floristic composition, it is important to note that these two ecoregions had opposite signs on the first NMS axis (based on the FISH data set), indicating that there is a major difference in the composition of aquatic vegetation in lakes between these two ecoregions.

There were 135 and 100 lake samples in the Western Cornbelt and Northern Glaciated plains, respectively. Both of these were strongly dominated by agricultural land use (76 and 80%, respectively) and had the lowest species richness and Shannon diversity levels (Table 3). The remaining three ecoregions each had 15 or fewer and no interpretations are made on these data.

Analysis by DNR ECS Sections (FISH data set) (Objective 3)

The numbers of lakes were distributed more equitably among ecological Sections (Map Plate 2), with five sections containing between 100 and 800 lake samples (Table 3). As in the Omernik classification, there were strong differences in immediate land use within the lake buffer among sections. The Northern Superior Uplands were the most heavily forested (70%) followed by the Drift and Lake Plains (51%). The North Central Glaciated Plateau and Minnesota and NE Iowa Moraines were highly agricultural. Again there was a strong contrast in species richness and diversity between the heavily forested Northern Superior Uplands and the agricultural North Central Glaciated plains, although more intermediate conditions were found with the section-scale classification.

Table 2. Number of lakes, land use within lake buffers, and Aquatic Vegetation diversity metrics, summarized by Omernik Level III Ecoregions. Nonmetric Multidimensional Scaling (NMS) methods were used to derive ordination scores for aquatic vegetation communities. Mean NMS axis 1 scores from 1500 lakes sampled by FISH method are shown, with species richness (S) and Shannon species diversity (H')) (both calculated from the trimmed species data set).

Omernik Ecoregion	# of lakes	% Agric.	% Forest	% Wetland	S	H'	mean NMS1
Red River Valley	15	0.74	0.09	0.10	8.53	1.59	-0.36
Northern Minnesota Wetlands	13	0.17	0.34	0.40	13.69	2.15	0.10
Northern Lakes and Forests	1303	0.11	0.56	0.22	19.65	2.38	0.25
North Central Hardwood Forest	937	0.49	0.18	0.16	16.06	2.09	-0.36
Driftless Area	5	0.24	0.24	0.16	9.60	1.65	-0.46
Western Cornbelt Plains	135	0.76	0.06	0.07	6.60	1.17	0.08
Northern Glaciated Plains	100	0.80	0.04	0.09	7.65	1.21	0.18

Table 3. Number of lakes, land use within lake buffers, and Aquatic Vegetation diversity metrics, summarized by DNR ECS Sections. Nonmetric Multidimensional Scaling (NMS) methods were used to derive ordination scores for aquatic vegetation communities. Mean NMS axis 1 scores from 1500 lakes sampled by FISH method are shown, with species richness (S) and Shannon species diversity (H') (both calculated from the trimmed species data set).

DNR ECS Sections	# of lakes	% Agric.	% Forest	% Wetland	S	H'	mean NMS1
N. Minnesota & Ontario Peatlands	5	0.03	0.30	0.57	12.20	1.90	0.28
Northern Superior Uplands	358	0.01	0.70	0.19	17.78	2.01	0.72
N. Minnesota Drift & Lake Plains	810	0.14	0.51	0.23	19.89	2.51	-0.15
Southern Superior Uplands	3	0.50	0.23	0.24	20.33	2.22	0.61
Western Superior Uplands	213	0.28	0.38	0.27	20.85	2.49	-0.01
Lake Agassiz, Aspen Parklands	4	0.54	0.14	0.14	11.25	1.75	-0.33
Minnesota & NE Iowa Morainal	792	0.48	0.18	0.15	16.05	2.07	-0.37
Paleozoic Plateau	7	0.38	0.23	0.12	8.43	1.53	-0.14
Red River Valley	14	0.77	0.08	0.10	8.36	1.43	-0.20
North Central Glaciated Plains	302	0.74	0.07	0.10	8.92	1.39	-0.02

Species Richness Analysis by Omernik Level III Ecoregions for all datasets

As in the FISH dataset analysis, the majority of lakes for all four datasets are in the Northern Lakes and Forests and North Central Hardwood Forest, which contain 1085 and 808 lake samples, respectively (Table 4). Overall, the Northern Lakes and Forests and Northern Central Hardwood Forest have the highest species richness, while the Western Cornbelt and Northern Glaciated Plains have the lowest. The FISH and ECO datasets detected more species than the WJL and WPI datasets, with average species richness comparable to the average species richness for all datasets combined all the ecoregions. For example, the ECO and FISH datasets have an average species richness 10 greater than the WJL and WPI datasets in the Northern Lakes and Forests. In the Western Cornbelt Plains, however, average species richness is nearly the same for the FISH, WJL, and WPI datasets (with no lakes sampled by ECO).

Species Richness Analysis by DNR ECS Sections for all datasets

The lakes are more evenly distributed across the ECS sections than the Omernik Ecoregions, as seen in the FISH analysis, with 5 sections containing between 150 and 700 lake samples (Table 5). In general, the Northern Minnesota Drift and Lake Plains and Western Superior Uplands have the highest species richness and the North Central Glaciated Plains have the lowest species richness. In the ECO and FISH datasets, the Northern Superior Uplands section also has high species richness and the Minnesota and NE Iowa Morainal section has lower species richness. For these last two ECS Sections, the opposite is true in the WJL and WPI datasets (with low species richness in the Northern Superior Uplands and higher species richness in the Minnesota and NE Iowa Morainal section). In 3 of the sections (Northern Minnesota Drift and Lake Plains, Northern Superior Uplands, and Western Superior Uplands) the ECO and FISH datasets have an average species richness of 10 or more species greater than the WJL or WPI datasets. In the other 2 sections with high sample size (North Central Glaciated Plains and Minnesota and NE Iowa Morainal) the average species richness is closer across datasets, however the ECO and FISH datasets are still higher than the WJL and WPI datasets. As seen in the analysis by Omernik Level III Ecoregions, the FISH and ECO datasets have average species richness comparable or greater than the average composite species richness (based on all datasets) across all the DNR sections.

In summary, the ECO and FISH sampling detected more overall species richness than WPI and WJL sampling; species richness values in ECO and FISH are comparable to combining the 4 datasets.

Table 4: Average species richness (S') by Omernik Ecoregion for each dataset. Note: In the ECO dataset Leech Lake is divided into 6 bays which are combined for the composite ALLKS dataset.

	ECO		FISH		WJL		WPI		ALLKS	
Omernik Ecoregion	S'	# of lakes	S'	# of lakes	S'	# of lakes	S'	# of lakes	S'	# of lakes
Red River Valley		0	12	7	5	2	6	5	9.5	13
Northern Minnesota Wetlands		0	18.4	5	10	3	14.3	3	15.4	10
Northern Lakes and Forests	23.7	31	23.1	776	13.9	252	13.4	26	21.2	1029
North Central Hardwood Forest	16.2	28	18.5	569	12.4	148	10.3	63	17.2	750
Driftless Area		0	11.7	3	7	1		0	10.5	4
Western Cornbelt Plains		0	6.9	84	7.4	11	6.2	28	6.9	116
Northern Glaciated Plains	7	1	10.5	56	4	8	5.6	20	9.1	79
total number of lakes		60		1500		425		145		2001

Table 5: Average species richness (S') by DNR ECS Section for each dataset. Note: In the ECO dataset Leech Lake is divided into 6 bays which are combined for the composite ALLKS dataset.

	ECO		FISH		WJL		WPI		ALLKS	
DNR ECS Section	S'	# of lakes	S'	# of lakes	S'	# of lakes	S'	# of lakes	S'	# of lakes
Lake Agassiz, Aspen Parklands		0	18.5	2	4	1		0	13.7	3
Minnesota & NE Iowa Morainal	16.3	24	18.2	493	12.4	118	10.9	55	17.1	644
N. Minnesota & Ontario Peatlands		0	16	3	8	1		0	14	4
N. Minnesota Drift & Lake Plains	23.7	20	26.6	359	13.9	221	14.4	20	22.2	578
North Central Glaciated Plains	13.5	4	10.6	179	7.4	26	6.6	60	9.8	247
Northern Superior Uplands	22	6	18.3	339	7.7	6	8	1	18.3	345
Paleozoic Plateau		0	9.4	5	7	1		0	9	6
Red River Valley		0	12.4	7	5	1	4.2	5	8.7	13
Southern Superior Uplands		0	21.3	3		0		0	21.3	3
Western Superior Uplands	23.8	6	26.4	110	14.4	50	10.3	4	23.3	158
total number of lakes		60		1500			425		145	2001

Community Analyses (Objective 5)

Both classification and ordination analyses were performed on species compositional data for each dataset. Classifications identify natural "assemblages" of taxa, and produce dendrograms that display the hierarchical combination of similar lakes to form the groups of lakes used to define plant assemblages or plant community types. Ordinations provide a ranking of lakes based on gradients in their plant species composition, and using these rankings, they display lakes as points arrayed along two or three axes in graphs. Lakes with similar plant composition are shown as symbols close together in the ordination graphs, and lakes with very different plant composition have symbols that are far apart in the graphs (see Appendices H.1 – H.5).

Ordination graphs use color-coded symbols to represent the assemblages or plant community types for each lake that were identified in the flexible beta classification analysis. Other environmental variables or community metrics can be overlaid on the ordination by varying the size of the symbols, so that larger symbols represent higher environmental values, or higher community metrics (e.g. taxa richness). By overlaying environmental or other descriptive data on the ordination of lakes, relationships between species composition and environmental factors can be assessed. By overlaying values of relative frequency or importance values of individual species or taxa on the ordination graph, patterns of species composition can be interpreted. Indicator species can be identified as species that occur in only one plant community type; and taxa characteristic of a plant community, but not necessarily restricted to one type, can also be easily observed in species overlays. The number of meaningful or useful assemblages defined by the flexible beta classification was determined for each dataset by examining the patterns of the plant community types in relation to overlays of plant taxa values and values of environmental data. Another way to display relationships between quantitative environmental variables (such as percent of 1 km buffer that is forested, or taxa richness) and ordination axis scores is displayed in a joint plot. A joint plot is a set of radiating lines drawn through the ordination graph; the angle and length of the line tell the direction and strength of the relationship between the variable and the ordination axes. One joint plot displays lines for all the variables that meet the criterion of having an r^2 correlation of greater than 0.2. The joint plot for each dataset helps with interpretation of the environmental gradients correlated to the vegetation gradients represented by the ordination axes. A summary of the results for each dataset is provided below.

1. Fisheries Dataset (FISH)

The best ordination of the large FISH dataset was obtained by first running a Bray-Curtis ordination, and then running a NMS ordination using the Bray-Curtis coordinates as a starting configuration. The results were an ordination of 1500 lakes and 103 taxa (importance values), with a final stress of 20.91, and final instability of 0.004, with 21 iterations run. The stress value is relatively high compared to guidelines for a good ordination (McCune and Grace, 2002), except that stress increases with the size of the dataset. Given that this is a large dataset, and the ordination showed fairly clear correlations with environmental variables, we accepted this solution. It may be possible to get a clearer ordination by removing some outlier lakes, or subdividing the dataset into smaller groups of lakes.

The flexible beta classification identified up to 12 types of lakes based on the FISH importance value data. After examining overlays of these groups, taxa importance values, and environmental variables, the 9 type level was selected for identifying plant community types. The composition of these nine plant community types are summarized in Appendix F.1. A map of the statewide distribution of lakes in the FISH dataset, and their plant community types (PC Type) from the classification is shown in Map Plate 13.

The joint plot (Appendix H.2, Figure FISH-1) of the FISH ordination shows that the aquatic plant vegetation gradient represented by the first NMS axis was most strongly correlated with alkalinity, and the second NMS axis was strongly correlated with surrounding land use, total Phosphorus, mean Chlorophyll a, Taxa richness, and maximum depth of the lakes. The lower part of the second axis represents higher diversity lakes in a forested setting; these tend to be higher elevation sites, and deeper lakes. The upper part of the second axis represents lower diversity lakes in agricultural or developed settings, with higher total Phosphorus, and higher mean Chlorophyll a.

2. Ecological Services Dataset (ECO)

The best ordination of the ECO dataset was run in autopilot mode set to medium speed and thoroughness. The results were an ordination of 60 lakes and 73 taxa (relative frequency values), with a final stress of 14.38, and final instability of 0.0001, with 79 iterations run. This is a very stable solution, and it showed clear correlations to several landscape setting and water quality variables.

The flexible beta classification identified up to 9 types of lakes based on the ECO relative frequency data. After examining overlays of these groups, taxa relative frequency values, and environmental variables, the 6 type level was selected for identifying plant community types. The composition of these six plant community types are summarized in Appendix F.2. A map of the statewide distribution of lakes in the ECO dataset, and their plant community types (PC Type) from the classification is shown in Map Plate 12.

The joint plot (Appendix H.1, Figure ECO-1) of the ECO ordination shows that the aquatic plant vegetation gradient represented by the first NMS axis was correlated with lake Alkalinity, and the second NMS axis was strongly correlated with quite a few landscape setting and water chemistry variables. The lower part of Axis 2 represents higher diversity lakes in forested settings; these lakes have higher elevations and greater maximum depths than lakes in the upper part of Axis 2. The upper part of the second axis represents lower diversity lakes with greater proportions of agriculture lands in the 1 km buffer area, and the water of these lakes has higher total Phosphorus, higher mean Chlorophyll a, and a higher Trophic State Index based on Chlorophyll a values.

3. Wildlife Jessen-Lound Dataset (WJL)

The best ordination of the WJL dataset was first run in autopilot mode set to medium speed and thoroughness, and then the optimal configuration from that run was used as starting coordinates for the final run. The results were an ordination of 425 lakes and 83 taxa (Beals smoothing favorability values), with a final stress of 8.72, and final instability of 0.00450, with

18 iterations run. This is a very low stress solution, and it shows clear correlations to several landscape setting variables.

The flexible beta classification identified up to 12 types of lakes based on the WJL Beals smoothing favorability values. After examining overlays of these 12 groups, taxa favorability values, and environmental variables, the 9 type level was selected for identifying plant community types. The composition of these nine plant community types is summarized in Appendix F.3. A map of the statewide distribution of lakes in the WJL dataset, and their plant community types (PC Type) from the classification is shown in Map Plate 14.

The joint plot (Appendix H.3, Figure WJL-1) of the WJL ordination shows that the aquatic plant vegetation gradient represented by the second NMS axis was strongly correlated with a gradient of landscape settings. The lower part of Axis 2 represents lakes in forested or primarily natural landscape settings; while the upper part of Axis 2 represents lakes in primarily agriculture or developed landscape settings. There weren't significant correlations with water chemistry variables, probably because there was very little water chemistry data available for many of the WJL lakes.

4. Wildlife Point-Intercept Dataset (WPI)

The best ordination of the WPI dataset was first run in autopilot mode set to medium speed and thoroughness, and then the optimal configuration from that run was used as starting coordinates for the final run. The results were an ordination of 138 lakes and 63 taxa (relative frequency values), with a final stress of 19.61, and final instability of 0.005, with 200 iterations run. This was the most stable solution found after trying many variations, and it is still less stable than the solutions for some of the other datasets.

The flexible beta classification identified up to 11 groups of lakes based on the taxa relative frequency values. After examining overlays of these 11 groups, taxa relative frequency values, and environmental variables, a modified version of the 7 type level from the flexible beta classification was selected for identifying plant community types. The 7 flexible beta types were slightly modified, since two types were intermingled in the ordination, and these two types did not get lumped at the 6 group level; so those intermingled types were combined into a single type based on their array in the ordination, leaving 6 plant community types as our final result. The composition of these six plant community types is summarized in Appendix F.4. A map of the statewide distribution of lakes in the WPI dataset, and their plant community types (PC Type) from the classification is shown in Map Plate 15.

The joint plot (Appendix H.4, Figure WPI-1) of the WPI ordination shows that the aquatic plant vegetation gradient represented by the first NMS axis was strongly correlated with a gradient of landscape settings. The left end of Axis 1 represents lower diversity lakes in primarily agriculture or developed landscape settings, and the right end of Axis 1 represents higher diversity lakes in predominantly natural landscapes, with high proportions of the surrounding landscape in the buffer that is forested or wetlands.

5. Combined Point-Intercept Dataset (ECOWPI) (Objective 4)

The two datasets sampled using the point-intercept method were combined into one dataset of 205 lakes and 83 taxa (relative frequency values) in order to address this project's Objective 4: to determine if this combined dataset is sufficient for statewide analyses of aquatic plant communities. We were unable to obtain a stable, low-stress solution with meaningful environmental correlations for this dataset, despite numerous runs of both NMS and Bray-Curtis ordinations. Therefore we are not reporting ordination and classification results for this combined dataset.

One possible solution would be to convert the ECOWPI relative frequency data to presence/absence data, modify the presence/absence data with the Beals smoothing algorithm, and run analyses of the Beals smoothing favorability dataset. Since we were able to do this same procedure for a combination of all four datasets, we did not proceed further with this approach to analysis of the combined ECOWPI dataset.

6. All Lakes from Four Datasets Combined (ALLKS) (Objective 1)

All four original datasets (ECO, FISH, WJL, and WPI) were converted to presence/absence values and combined into one large dataset of 1984 lakes and 130 plant taxa (Beals smoothing values), in order to address the question in this project's Objective 1: to produce an assessment of whether the data can be analyzed in a single comprehensive data set. The best ordination of the ALLKS dataset was produced by first running a Bray-Curtis ordination, and then using the axis coordinates from the Bray-Curtis ordination as a starting point for the NMS runs. Initially this procedure was run on the full dataset of 1994 lakes, but the NMS solutions were unstable and could not be used. An outlier analysis in PC-ORD identified 10 lakes that were extreme outliers in the ALLKS dataset. These 10 outliers were removed, leaving a dataset of 1984 lakes and 130 taxa. The results were an ordination of 1984 lakes and 130 taxa (Beals smoothing favorability values), with a final stress of 6.06864, and final instability of 0.00318, with 33 iterations run. This is the lowest stress NMS solution of all the datasets, and second most stable after the ECO dataset. This ordination has strong correlations with the first axis. There weren't significant correlations with water chemistry variables, probably because there were very few water chemistry data available for many of the lakes. Ordination graphs with overlays of landscape, environment, and aquatic plant taxa favorability values are provided in Appendix H.5.

The flexible beta classification of ALLKS identified up to 12 groups of lakes based on the taxa Beals smoothing favorability values. After examining overlays of these 12 groups, taxa favorability values, and environmental variables, the 11 type level from the flexible beta classification was selected for identifying plant community types. The composition of these eleven plant community types is summarized in Appendix F. 5. A map of the statewide distribution of lakes in the ALLKS dataset, and their plant community types (PC Type) from the classification are shown in Map Plate 16.

The joint plot (Appendix H.5, Figure ALLKS-1) of the ALLKS ordination shows that the aquatic plant vegetation gradient represented by the first NMS axis was strongly correlated with a gradient of landscape settings. The left end of Axis 1 represents lakes in predominantly natural landscapes, with high proportions of the surrounding landscape in the buffer that is forested;

these lakes tend to be at higher elevations also. The right end of Axis 1 represents lakes in primarily agriculture or developed landscape settings.

Comparison of Plant Community Types (PC Types)

There were 41 plant community types identified in the analyses of five datasets (FISH, ECO, WJL, WPI, and ALLKS). The species values in these datasets consisted of different types of variables: importance values for FISH, relative frequencies for ECO and WPI, and presence/absence data modified by the Beals smoothing algorithm for WJL and ALLKS. In order to compare the composition of plant community types in these five datasets, the presence/absence dataset was used to calculate the relative frequency of each species across the set of lakes in each of the 41 recognized plant community types. A table of plant taxa by PC Types was compiled that included the relative frequency of each species in each PC Type (Appendix E). These relative frequencies were calculated by adding the number of lakes in a PC Type where a species was present, and dividing that by the number of lakes in the PC Type. The summary dataset was then called “Combined PCtypes”; it consisted of the 41 plant community types and 137 plant taxa. In order to evaluate the similarity of plant community types identified from the different datasets, the plant community types were run through the same classification and ordination procedures used on each of the lakes by species datasets. This time the symbols in the ordination represent each of the 41 plant community types, and we called the assemblages identified “plant community groups” (see Appendix H.5, Figure ALLKS-19). These groups represent a higher level in an hierarchical community classification, and they make it easier to make comparisons between datasets.

Aquatic Plant Community Groups

Eight plant community groups were identified as a result of the flex beta classification and NMS ordinations of the “Combined PCtypes” dataset. These eight plant community groups identify broad assemblages of aquatic plants characteristic of Minnesota lakes. Finer scale differences in whole lake plant community types are represented by the 41 plant community types; however some of these types may represent replicates of the same lakewide plant communities. They vary due to differences in sampling dates, differences in sampling methods, or differences in aquatic plant identification skills of the staff gathering the data. The summary of plant community types is presented below and in Appendix F.6 by aquatic plant community groups (the broader scale). Table 6 shows which PC Types are included in each of the 8 groups. The composition of each of the 41 plant community types can be examined in the tables in Appendix F.1 – F.5. Even finer scale differences can be recognized within lakes, if sampling is stratified within lakes. The only example of such finer scale stratification of sampling in this study is in the ECO data for Leech Lake. Six bays of Leech Lake were sampled separately. In the analysis of ECO data these six bays were assigned to two different plant community types; but in the analysis of 1984 lakes statewide, Leech Lake as a whole was assigned to just one plant community type.

Group 1: Water lily - cattail - muskgrass - sedge alkaline marsh

The water lily – cattail – muskgrass – sedge alkaline marsh group includes shallow water types with a mixture of floating-leaved, emergent, and submersed aquatic plants. The full composition of each included type is summarized in Appendix E. The taxa with the highest relative frequencies in the included PC Types are yellow waterlilies (*Nuphar* spp.), broad-leaved cattail (*Typha latifolia*), white waterlilies (*Nymphaea* spp.), muskgrasses (*Chara* spp.), and sedges (*Carex* spp.). These lakes tend to have alkaline water chemistry. This group includes the largest number of lakes, and ten PC Types recognized in the five datasets. It includes ALLKS types 1, 5, 6, and 15, FISH types 1 and 16, and WJL types 1, 2, 4, and 29 (Table 6; and Appendix F.6).

Group 2: Sedge - yellow water lily - iris rich marsh

The sedge – yellow water lily – iris rich marsh group includes shallow water types with a mixture of sedges (*Carex* spp.) and yellow water lilies (*Nuphar* spp.) as the most frequent taxa. The full composition of each included type is summarized in Appendix E. This group includes a large number of lakes, and four PC Types recognized in the five datasets. It includes ALLKS types 2 and 34, and FISH types 3 and 59 (Table 6; and Appendix F.6).

Group 7: Sago pondweed - cattail - coontail - duckweed marsh

The sago pondweed – cattail – coontail – duckweed marsh group includes shallow water types with a fairly high species diversity, including sago pondweed (*Stuckenia pectinata*), broad-leaved cattail (*Typha latifolia*), coontail (*Ceratophyllum demersum*), lesser duckweed, algae, and hardstem bulrush (*Schoenoplectus acutus*) as the most frequent taxa, occurring in more than 50% of lakes. PC Types in this group often include exotic species, with curly-leaf pondweed (*Potamogeton crispus*) most frequently reported. The full composition of each included type is summarized in Appendix E. This group includes a large number of lakes, and nine PC Types recognized in the five datasets. It includes ALLKS types 46, 194, and 218, and FISH types 40, 42, 66, 117, and 134, and WJL type 37 (Table 6; and Appendix F.6).

Group 8: Coontail - pondweed - muskgrass aquatic bed

The coontail – pondweed – muskgrass aquatic bed group includes deep water aquatic bed types of submerged aquatic plants, including sago pondweed (*Stuckenia pectinata*), coontail (*Ceratophyllum demersum*), lesser duckweed (*Lemna minor*), flat-stem pondweed (*Potamogeton zosteriformis*), muskgrasses (*Chara* spp.), narrow-leaf pondweeds (*Potamogeton* spp.). The full composition of each included type is summarized in Appendix E. This group includes a large number of lakes, and eleven PC Types recognized in the five datasets. It includes ALLKS type 69, ECO types 4 and 31, WJL types 46, 47, and 57, and WPI types 2, 3, 10, 26, and 104 (Table 6; and Appendix F.6).

Group 11: Disturbed pondweed - coontail - northern milfoil aquatic bed

The disturbed pondweed – coontail – northern milfoil aquatic bed group includes deep water aquatic bed types with low diversity that tends to occur in disturbed landscapes. The most frequent native taxa are sago pondweed (*Stuckenia pectinata*), coontail (*Ceratophyllum demersum*), northern milfoil (*Myriophyllum sibiricum*) and muskgrasses (*Chara* spp.). This group frequently has exotic species as an important component of the vegetation, with curly-leaf pondweed (*Potamogeton crispus*) the most common, along with Eurasian milfoil (*Myriophyllum spicatum*) at lower frequency. The full composition of each included type is summarized in Appendix E. This is a small group of 15 lakes in two PC Types: ALLKS type 814 and WJL type 128 (Table 6; and Appendix F.6).

Group 12: Muskgrass - bulrush - pondweed alkaline aquatic bed

The muskgrass – bulrush – pondweed alkaline aquatic bed group includes aquatic bed types that are in alkaline lakes and tend to have high diversity. The taxa with the highest frequencies in this group are muskgrasses (*Chara* spp.), bulrushes (*Scirpus* spp.), large-leaf pondweed (*Potamogeton amplifolius*), flat-stem pondweed (*Potamogeton zosteriformis*), coontail (*Ceratophyllum demersum*), narrow-leaf pondweed with floating leaves (*Potamogeton* spp.), clasping-leaf pondweed (*Potamogeton richardsonii*), and waterweeds (*Elodea* spp.). The full composition of each included type is summarized in Appendix E. This group includes 43 lakes, in three PC Types, that were only present in the ECO dataset: ECO types 1, 2, and 6 (Table 6; and Appendix F.6).

Group 15: Water shield - hornwort - spikerush lake

The water shield – hornwort – spikerush lake group is a very small but distinct group of two lakes with composition that suggests boggy wetlands or low alkalinity ponds. The most frequent taxa are water shield (*Brasenia schreberi*), hornwort (*Ceratophyllum echinatum*), spikerushes (*Eleocharis* spp.), small-leaved milfoil (*Myriophyllum farwellii*), stoneworts (*Nitella* spp.), yellow and white water lilies (*Nuphar* spp., *Nymphaea* spp.), Nuttall's pondweed (*Potamogeton epihyrus*), floating-leaf pondweed (*Potamogeton natans*), narrow-leaf pondweeds with floating leaves (*Potamogeton* spp.), floating-leaved burreeds (*Sparganium* spp.), and common bladderwort (*Utricularia macrorhiza*). This group includes two lakes in ECO type 5 (Table 6; and Appendix F.6). There may well be other lakes of this type in other datasets, however they may have come out differently in the analyses because some of the characteristic species are difficult to identify in the field, and so the composition based on common plants would not be so distinct from other groups.

Group 39: Sago pondweed - hardstem bulrush - widgeon grass marsh

The sago pondweed – hardstem bulrush – widgeon grass marsh group is another very small but distinct group of three lakes. The taxa with the highest frequencies are sago pondweed (*Stuckenia pectinatus*), hardstem bulrush (*Schoenoplectus acutus*), widgeon grass (*Ruppia occidentalis*), sedges (*Carex* spp.), muskgrasses (*Chara* spp.), bushy pondweed (*Najas flexilis*), horned pondweed (*Zannichellia palustris*), and several pondweeds (*Potamogeton* spp.). This group includes three lakes in WPI type 20 (Table 6; and Appendix F.6).

Classification Comments

The plant community groups and types are not strongly correlated to ECS provinces, sections, subsections, nor Omernik ecoregions); however some regional patterns are apparent in the statewide maps for each dataset. Although some aquatic plant communities are more likely to occur in one or a few ECS sections than elsewhere, aquatic plant communities tend to cross these ecoregional boundaries, and are much more influenced by surrounding landscape features and water chemistry. There are some correlations with ECS LTA's and Schupp classes, but they are weak and inconsistent. It might be worth looking at large scale watersheds to see if they are correlated (re: Red River, James Bay, Lake Superior, Mississippi River). We haven't yet been able to examine those large-scale watershed patterns.

Some aquatic plant community types are not adequately represented in the datasets, in particular bog ponds and acidic lakes. These types are probably much better represented in the MCBS aquatic vegetation data. Probably any lake type that is not favorable for production of game fish or waterfowl are under-represented in these datasets. Some rare aquatic types, such as saline ponds or coastal-plain disjunct ponds seem to be missing, or they aren't popping out as distinct types because of small sample size or inadequate identification of uncommon plant species. These rare types might be better represented with the addition of MCBS data.

Inconsistencies in how a lake is classified by different datasets may be due to a variety of factors: sampling year (big differences could be due to big changes such as development or pollution), sampling date (June vs July vs Sept), plant identification skills of field staff collecting data, and sample size/sampling method in a lake (e.g. FISH transects cover a lot more area than point-intercept samples).

Finer scale patterns and whole lake community types can probably be detected in the larger datasets by subdividing them at upper levels of the classification (e.g., 2 to 4 groups) and then running similar analyses on the subsets of the data. Some community metrics (e.g. taxa richness) had clear multivariate patterns, such as a binomial distribution, that do not yield significant correlation statistics because the correlation is positive in the first half of the axis and negative in the second half. Dividing the data into two or three parts would allow these correlations to be clarified.

Combining the two point-intercept datasets (ECO and WPI) does not adequately represent the range of aquatic plant communities that are included in the FISH and WJL datasets. A couple community groups were not identified at all in either ECO or WPI dataset. There is also something about the WPI dataset that makes it difficult to analyze with the same techniques that worked well on the FISH and ECO datasets. The combined ECOWPI dataset was even more problematic (with greater stress and instability in the NMS solutions) than WPI alone. We don't know what is different about the WPI dataset; it might have to do with accuracy of plant identification (a guess based on the low numbers of exotics such as *Potamogeton crispus* and *Myriophyllum spicatum*), or inconsistency between field workers. Or maybe the data have an odd distribution pattern that needs to be transformed in some way to produce better results.

Presence/absence data, especially in large quantities (over 400 samples) can produce a useful plant community classification with clear environmental correlations, especially when the Beals smoothing favorability index is applied to the data. Finer scale patterns may be easier to detect with quantitative data on smaller datasets (as in the ECO dataset), but large-scale, statewide patterns may be easiest to detect with large datasets of presence/absence data. Accuracy in species identification is probably more important than which sampling method is used, in terms of detecting plant community patterns and producing meaningful results. For example, narrow-leaf pondweeds, *Carex* spp., and some aquatic mosses are known by botanists familiar with the species to be good indicators of fine scale environmental gradients such as pH and alkalinity. However, these species are challenging to identify, and usually need to be collected for laboratory identification by most field workers who are not experts in the taxonomy of these groups. Collecting voucher specimens of these particular plant groups, and investing in getting the specimens correctly identified, could make a big difference in recognizing and refining plant community patterns. Since the MCBS aquatic vegetation data are vouchered and carefully identified for at least the pondweeds and sedges, the MCBS dataset could be used to assess the difference in results produced by datasets with difficult species lumped together into broader taxa versus datasets with accurate species identifications (species not lumped).

Table 6. Aquatic Plant Community Groups, and the plant community types included within those groups. Groups were identified across five data sets based on classification of the 41 plant community types derived from individual analyses of the five data sets. (See text for details.) The plant community types represent subsets of the Aquatic Plant Community Groups.

Aquatic Plant Community Group	Aquatic Plant Community Group Names	Included Plant Community Types				
		All Lakes	FISH	WJL	WPI	ECO
Group 1	Water lily - cattail - muskgrass - sedge alkaline marsh	ALLK001 ALLK005 ALLK006 ALLK015	FISH001 FISH016	WJL001 WJL002 WJL004 WJL029		
Group 2	Sedge - yellow water lily - iris rich marsh	ALLK002 ALLK034	FISH003 FISH059			
Group 7	Sago pondweed - cattail - coontail - duckweed marsh	ALLK046 ALLK194 ALLK218	FISH040 FISH042 FISH066 FISH117 FISH134	WJL037		
Group 8	Coontail - pondweed - muskgrass aquatic bed	ALLK069		WJL046 WJL047 WJL057	WPI002 WPI003 WPI010 WPI026 WPI104	ECOpc04 ECOpc31
Group 11	Disturbed pondweed - coontail - northern milfoil aquatic bed	ALLK814		WJL128		
Group 12	Muskgrass - bulrush - pondweed alkaline aquatic bed					ECOpc01 ECOpc02 ECOpc06
Group 15	Water shield - hornwort - spikerush aquatic bed					ECOpc05
Group 39	Sago pondweed - hardstem bulrush - widgeon grass marsh				WPI020	

Aquatic Plant Communities and Schupp Lake Types (Objective 3)

Schupp (1992) developed an ecological classification of Minnesota lakes; the classification was derived from three types of variables, related to lake size, depth, and fertility. Forty-four lake classes were identified, falling onto four primary groups. The first major division separated lakes of northeastern Minnesota (St. Louis, Lake and Cook counties) from the rest of the state. Lakes of NE MN are derived from scouring of pre-Cambrian rock, contrasted with the lakes formed within glacial deposits outside of this region. Lake alkalinity is the major discriminating variable: NE MN lakes are soft water lakes, with alkalinity values typically < 40. The 2nd major division separates lakes based on a threshold of 80% littoral zone – those with > 80% littoral zone are likely those that frequently experience winter-kill.

The composite set of sample lakes (ALLKS dataset) was non-uniformly distributed among Schupp classes, with a greater representation of hard-water lakes compared with those of NE MN (Figure 1). Two lake classes in particular were strongly represented in the data set – classes 24 and 43 each comprised > 120 lakes.

There was a strong correlation between the ALLKS plant community types (derived from the composite dataset) and aggregated Schupp classes (Table 7). There was a clear discrimination in aquatic vegetation communities with respect to alkalinity. Two PC Types characterized the soft-water lakes: ALLKS #2 (*Nuphar-Nymphaea* marsh) and ALLKS #34 (*Carex-Dulichium* marsh). These were mostly restricted to Schupp classes 1-19, which occur in NE MN, but they also occurred in Schupp classes 20 and 21, which are low-alkalinity lakes outside the NE area, as well as Schupp classes 29, 37 and 39, all of which have relatively low (<100 ppm) alkalinities for their region. Schupp class 32 is anomalous, a higher-alkalinity lake characterized by soft-water aquatic plant communities.

Three other aquatic vegetation communities were strongly representative of and restricted to hard-water lakes: ALLKS #1 (*Potamogeton zosteriformis-Chara* marsh), ALLKS #6 *Potamogeton zosteriformis-Nuphar* marsh) and ALLKS #46 (*Potamogeton pectinatus-Ceratophyllum* marsh). These communities only rarely occurred in NE MN, but were very common outside this region. ALLKS #194 (*Potamogeton pectinatus-Scirpus* marsh) was similarly restricted to hard-water lakes, but was much less common.

One vegetation community, ALLKS #5 (*Nuphar-Nymphaea* marsh) was more ubiquitous, occurring in both soft and hard water lakes. Overall, there was little evidence of discrimination with respect to aquatic plant communities between lakes that winter-kill and those that do not.

Schupp class 43, one of the two classes with more than 120 samples, had strong representations of several relatively uncommon groups. Of the 33 lakes represented by ALLKS #218 (*Potamogeton pectinatus-Potamogeton crispus* aquatic bed), 25 occurred in Schupp class 43. To a lesser degree ALLKS #69 (*Ceratophyllum-Potamogeton* aquatic bed) and ALLKS #194 (*Ceratophyllum-Potamogeton* aquatic bed) also occurred with relatively high frequency in this group.

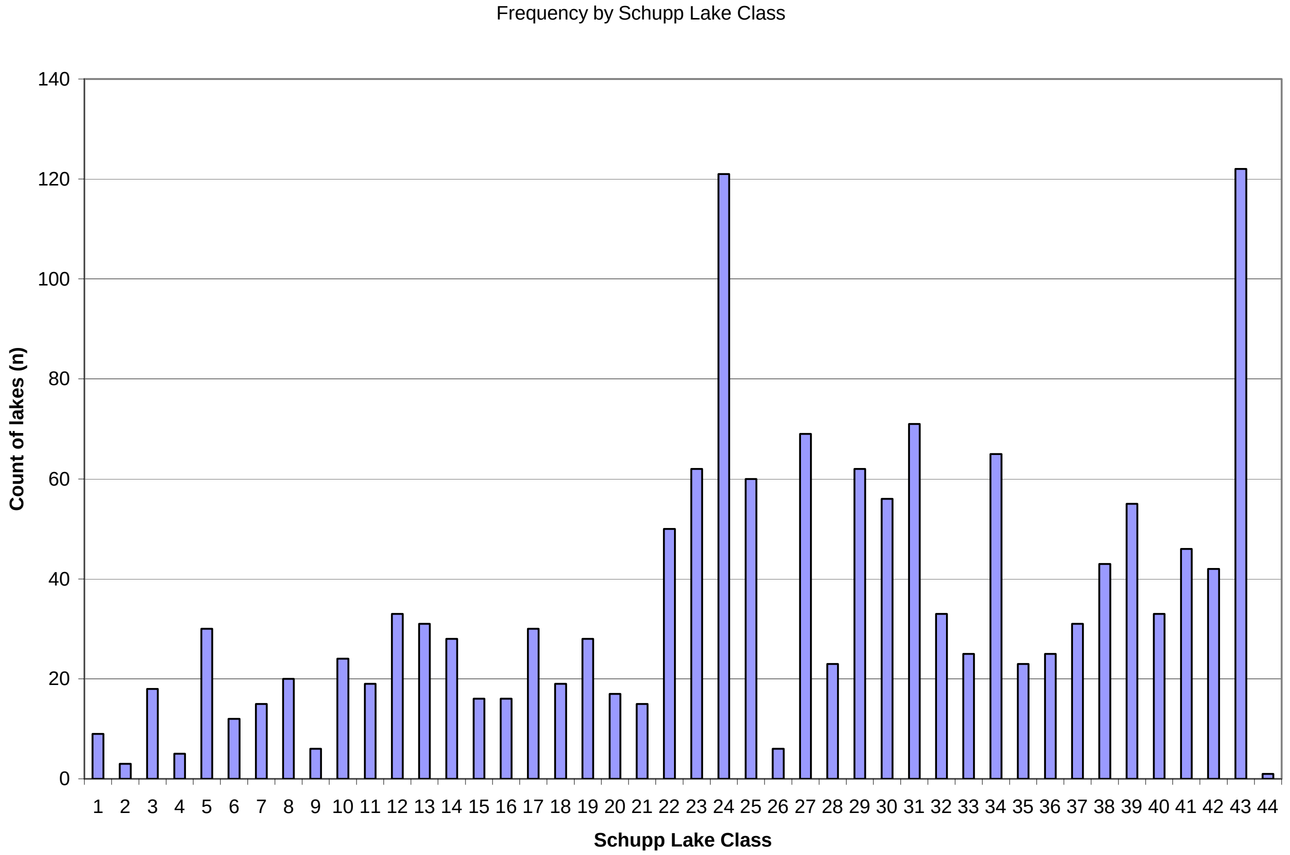


Figure 1. Frequency distribution of Schupp Lake Classes.

Table 7. Distribution of ALLKS aquatic plant community types by Schupp Lake Class. See Appendix E for a description of plant community types.

Aquatic Community Classes - All Lakes

Schupp Lake Class	1	2	5	6	15	34	46	69	194	218	814	Grand Total
1		8				1						9
2		2	1									3
3		12				5						17
4		2				3						5
5		20	7	1		2						30
6		11				1						12
7		10	4		1							15
8		15	1			3						19
9		4				2						6
10		16				8						24
11		14	1	2		1	1					19
12		26	1			4						31
13		19	2		1	9						31
14		15	1			10						26
15		10	1			5						16
16	1	13	1	1								16
17		24				6						30
18		11	1		1	4						17
19	1	16	7	3					1			28
20	2	10	3			2						17
21	1	8	4		2							15
22	34		1	12			2		1			50
23	30	3	10	17			2					62
24	28			15			63	10	1	1	2	120
25	30		6	18	1		4		1			60
27	38		4	20			7					69
28	13	1		7	1		1					23
29	24	7	11	17	1		2					62
30	7		4	12	1		21	5	6			56
31	48		1	16			5		1			71
32	4	12	9	5	1	2						33
33	6		1	9	1		6	1	1			25
34	35			9			15	1	5			65
35	8	2	4	5			3		1			23
36	5	4	3	5	2	2	3		1			25
37	2	11	6	4	4	3	1					31
38	15		1	5			16	3	1	1	1	43
39	13	6	11	15	2		7		1			55
40	1			3			10	12	5	2		33
41	13			4			17	5	5	2		46
42	9		3	3	1		17		7	2		42
43	4			6	1		56	11	15	25	2	120
44							1					1
Grand Total	372	312	110	214	21	73	260	48	53	33	5	1501

Aquatic Vegetation Response to Disturbance (Objectives 6, 7)

There were significant ecological gradients on the primary axes of variation in the aquatic vegetation that appeared to be related to a combination of geographic and anthropogenic factors (e.g. the NMS analysis of FISH dataset in Appendix H.1). With respect to geographic gradients, there was spatial patterning in the composition of aquatic variation with respect to both of the ecological land classifications. The Omernik Level III classification is relatively coarse, with the two forested Level III Ecoregions (Northern Central Hardwood Forest & Northern Lakes and Forests) accounting for 90% of the lakes. The DNR ECS had a better distribution of lakes among Sections (Map Plate 2). While there is considerable overlap among the 100's of lakes included in this analysis, there appeared to be strong associations between floristic communities and Ecological Sections. Some of these associations are biogeographic – there are strong gradients of climate and physiography across the state. Since one intent of ecological classification is to capture this variability, it provides a means to quantify factors that define species distribution limits. Axis 1 of the NMS analysis of the FISH dataset (Appendix H.2, Figure FISH-1) appears to separate floristic communities of the MN and NE Iowa Moraines from the Northern Superior Uplands (Map Plate 13). Axis 2 similarly separates the two plain-dominated Sections: North Central Glaciated Plains from the Drift and Lake Plains.

The other dominant gradient in the landscape results from human use of the landscape. There is a strong gradient of agricultural land uses from the southwest to northeast parts of the state. The distribution of agriculture itself is constrained by the climatic and physiographic factors described in the ECS classification, so it is difficult to discriminate floristic differences resulting from the impacts of agriculture (e.g. increased nutrient and sediment inputs) with biogeographic effects (but see Richards et al. 1996). Nonetheless, an overlay of the amount of agriculture within the 1 km lake buffer indicates that were floristic gradients observed on both NMS Axes 1 and 2, related to the amount of agriculture in the N Central Glaciated Plains and the MN and NE Iowa's Moraines respectively (Map Plate 3). The converse of this plot, based on the proportion of forested land in the lake buffer, separates the lake-rich Northern Superior Uplands at the upper end of Axis 1. A similar effect occurs on Axis 3, which places lakes of the heavily forested Drift and Lake Plains and Northern Superior Uplands at the upper end of the 3rd NMS axis (Map Plate 4).

Joint plots which overlay environmental variables with significant correlations to the vegetation data on top of the lake ordinations show the direction and magnitude of correlation between species composition and environment (e.g. Figure FISH-1). Maps of the statewide distribution of some of these environmental variables are shown in Map Plates 3-6. The effects occur across both axes one and two. The amounts of agriculture and forest land in the lake buffer work in opposite directions, with agriculture having a negative correlation with Axis 1 and a positive correlation with Axis 2 (Figure FISH-4). Two variables related to topography were also flagged as important: the median and minimum elevation were positively correlated with axis 1 (Figure FISH-6), and were positively correlated with the amount of forested uplands (Figure FISH-2). In assessing Axis 3, the Schupp Lake Typology Class was identified as a key weighting factor, correlated with the amount of agriculture and developed land (Table 8).

Spatial Trends in Aquatic Vegetation Community Types (Objectives 3, 6, 7)

We found some correlation between aquatic vegetation types and geographic location of the lake, but the community types cross ecoregional boundaries at ECS province, section and subsection levels. So the pattern of species distribution is more complex than simply a response to geographic location within the state. Stronger correlations were found with the percent of forested uplands or wetlands (natural lands) versus agriculture or developed lands within the 1 km wide buffer of each lake.

Vegetation Response to Landscape and Local Variables (Objective 7)

FISH**:** Flex-beta groups were strongly associated with the following taxa and environmental variables: *Carex*, *Sparganium fluctuans, Iris versicolor,* and *Dulichium arundinaceum* were negatively correlated with alkalinity and landscape variable including forested and natural land, while *Ceratophyllum demersum, Potamogeton zosteriformis,* and *Myriophyllum sibiricum* were positively correlated with alkalinity, road density, agricultural and developed land. Taxa richness was negatively correlated with total phosphorus, and positively correlated with natural land covers (Tables 8, 9).

ECO: Vegetation groups in the Ecological Services database were more strongly correlated with the physical and landscape variables than the FISH data set, in all likelihood because of the smaller sample size which may have spanned a stronger gradient of lakes (Tables 10, 11). *Brasenia schreberi*, *Potamogeton epihydrus*, *Nymphaea* sp. *(*includes *N. leibergii, N. odorata, N. tuberosa) Sparganium* spp. *(S. angustifolium, S. fluctuans, S minimum) Myriophyllum farwellii*, *Ceratophyllum echinatum* and *Drepanocladus/Fontinalis* were all strongly correlated with lower alkalinity lakes with lower secchi and shallower depths. *Chara* spp. and *Potamogeton crispus* were both correlated with the higher alkalinity lakes with higher secchi and maximum depths. These two species had opposite correlations with the second axis. *Chara* was positively correlated with natural lands and deeper lakes, while *P. crispus* was strongly positively correlated with agricultural and developed land uses. *Potamogeton crispus, P. pectinatus, Spirodela polyrhiza, Ceratophyllum demersum*, and *Myriophyllum spicatum* were all positively correlated with total phosphorus, chlorophyll a levels and developed land with higher proportions of agriculture. *Najas flexilis* was negatively correlated with all of the above variables, and positively correlated with natural land cover. Taxa richness was negatively correlated with agricultural and developed land use, chlorophyll a and total phosphorus (Tables 10, 11).

ALLKS: The composite dataset had many more taxa with strong correlations with the ordination axes, compared to the FISH dataset (Tables 12, 13). In addition, the environmental data were more strongly correlated with the ordination axes than were the FISH dataset. Comparable correlations were observed with the ECO dataset. The explanation for this is not obvious, given the high degree of overlap with FISH dataset. However, different techniques other than Beals smoothing for presence/absence data are warranted. In general the correlations were stronger with respect to the landscape variables versus those from the lakes themselves (Tables 8, 10, 12).

Table 8. Pearson and Kendall Correlations between Environmental Variables and Ordination Axes for the FISH dataset. Variables in bold have higher correlations with the ordination axes. (N = 1500).

	Axis 1			Axis 2			Axis 3		
Env Var.	r	r^2	tau	r	r^2	tau	r	r^2	tau
ECS LTAs	**-0.483**	**0.234**	**-0.305**	**0.475**	0.226	0.308	**-0.535**	0.286	-0.374
Dev Ag Lnd	**-0.422**	**0.178**	**-0.319**	**0.582**	0.339	0.376	**-0.544**	0.296	-0.359
SchLclas	**-0.379**	0.144	-0.195	**0.354**	0.125	0.273	**-0.381**	0.145	-0.252
RoadDens	**-0.376**	0.141	-0.399	**0.330**	0.109	0.207	-0.122	0.015	-0.162
Agricult	**-0.323**	0.104	-0.259	**0.460**	0.212	0.280	**-0.562**	0.316	-0.360
Alkalinity	-0.297	0.089	-0.170	0.236	0.056	0.119	-0.300	0.09	-0.163
Taxa Richness	-0.220	0.049	-0.126	**-0.438**	0.192	-0.286	**0.325**	0.106	0.222
meanSecc	-0.211	0.045	-0.200	-0.274	0.075	-0.149	0.024	0.001	-0.011
maxDepFt	-0.206	0.042	-0.191	-0.207	0.043	-0.119	-0.010	0	-0.011
maxDepM	-0.206	0.042	-0.191	-0.207	0.043	-0.119	-0.010	0	-0.011
Median Elev	**0.504**	**0.254**	**0.309**	**-0.397**	0.158	-0.296	0.158	0.025	0.091
MinElev	**0.473**	**0.224**	**0.297**	**-0.406**	0.165	-0.296	0.145	0.021	0.087
ForestUp	**0.468**	**0.219**	**0.282**	**-0.547**	0.300	-0.378	**0.514**	0.264	0.355
NatLand	**0.422**	**0.178**	**0.319**	**-0.582**	0.339	-0.376	**0.544**	0.296	0.359
STDelev	0.290	0.084	0.109	-0.101	0.010	-0.083	0.139	0.019	0.076

Interpretation:	Axis 1 is positively correlated with: Median elevation, minimum elevation, forested uplands and natural lands; Axis 1 is negatively correlated with: ECS LTAs, Developed land, and to a lesser extent: Schupp Class, Road density, and Agriculture Axis 2 is positively correlated with: ECS LTAs, Developed Land, Agriculture; Axis 2 is negatively correlated with: Minimum elevation, Forested Uplands, and Natural Lands, and taxa richness Axis 3 is postively correlated with: Forest Uplands and Natural Land Axis 3 is negatively correlated with: ECS LTAs, Developed and Ag land, and Agriculture

Table 9. Pearson and Kendall Correlations between Plant Taxa and Ordination Axes for the FISH dataset. Taxa in bold have high correlations with the ordination axes. N=1500

	Axis 1			Axis 2			Axis 3		
Taxa Code	r	r^2	tau	r	r^2	tau	r	r^2	tau
CERADEME	**-0.473**	0.224	-0.492	0.197	0.039	0.168	0.020	0	-0.036
POTAZOST	**-0.383**	0.146	-0.391	-0.134	0.018	-0.144	0.122	0.015	0.080
MYRISIBI	**-0.305**	0.093	-0.306	-0.129	0.017	-0.166	0.046	0.002	0.041
POTARICH	**-0.300**	0.090	-0.332	-0.163	0.026	-0.204	0.001	0	-0.024
CHARASP	-0.295	0.087	-0.278	**-0.382**	0.146	-0.355	-0.224	0.05	-0.180
NAJAFLEX	-0.267	0.071	-0.275	**-0.313**	0.098	-0.292	0.111	0.012	0.095
POTAPECT	-0.264	0.070	-0.356	0.164	0.027	0.178	**-0.355**	0.126	-0.378
ELODEA	-0.247	0.061	-0.339	0.125	0.016	0.013	0.153	0.023	0.108
CAREX	**0.423**	0.179	0.338	-0.275	0.075	-0.256	0.174	0.030	0.248
SPARGAFL	**0.334**	0.111	0.367	-0.111	0.012	-0.194	0.247	0.061	0.333
IRISVERS	**0.307**	0.094	0.290	-0.067	0.005	-0.158	0.150	0.023	0.284
DULIARUN	**0.302**	0.091	0.388	-0.105	0.011	-0.181	0.194	0.038	0.289
EQUISETU	0.279	0.078	0.282	-0.098	0.010	-0.223	0.152	0.023	0.262
BRASSCHR	0.211	0.044	0.158	-0.086	0.007	-0.136	0.264	0.070	0.289

Table 10. Pearson and Kendall Correlations between Environmental Variables and Ordination Axes for the ECO dataset. Variables in bold have high correlations with the ordination axes. N = 60

	Axis 1			Axis 2			Axis 3		
Veg Code	r	r^2	tau	r	r^2	tau	r	r^2	tau
Forest Up	**-0.336**	0.113	-0.216	**-0.538**	0.289	-0.34	-0.217	0.047	-0.133
Median El	-0.268	0.072	-0.171	**-0.454**	0.206	-0.402	-0.221	0.049	-0.008
Min Elev	-0.243	0.059	-0.151	**-0.461**	0.213	-0.387	-0.206	0.043	-0.005
Nat Land	-0.224	0.050	-0.153	**-0.666**	0.443	-0.469	-0.156	0.024	-0.104
Taxa Richn	-0.209	0.044	-0.146	**-0.509**	0.259	-0.342	-0.129	0.017	-0.135
Alkalinity	**0.495**	0.245	0.361	0.212	0.045	0.138	0.056	0.003	0.043
Secchi TSI	**0.397**	0.158	0.314	0.290	0.084	0.287	-0.231	0.053	-0.152
Max DepM	**0.327**	0.107	0.244	**-0.475**	0.226	-0.327	0.186	0.035	0.092
meanChl_a	**0.300**	0.090	0.243	**0.581**	0.338	0.348	-0.257	0.066	-0.243
TP _TSI	0.295	0.087	0.238	**0.373**	0.139	0.316	**-0.322**	0.104	-0.217
Chl_a TSI	0.289	0.083	0.241	**0.427**	0.183	0.349	**-0.356**	0.127	-0.241
Agricult	0.264	0.069	0.218	**0.581**	0.338	0.391	0.078	0.006	0.044
Water	0.233	0.054	0.180	-0.265	0.07	-0.232	**0.309**	0.096	0.226
DevAgLnd	0.224	0.050	0.153	**0.666**	0.443	0.469	0.156	0.024	0.104
ECS LTAs	0.159	0.025	0.078	**0.558**	0.311	0.353	0.139	0.019	0.113
totPppb	0.158	0.025	0.238	**0.521**	0.271	0.316	-0.248	0.062	-0.214
RoadDens	0.133	0.018	0.120	**0.422**	0.178	0.331	0.148	0.022	0.049

Interpretation:

Axis 1 is positively correlated with: Alkalinity, Secchi Trophic State Index, Maximum depth
Axis 1 is negatively correlated with: Forested uplands
Axis 2 is positively correlated with: mean Chlorophyll a, Agriculture, Developed land, ECS LTAs, Total Phosphorus and Road density
Axis 2 is negatively correlated with: Forested uplands, median elevation, minimum elevation, Natural lands, Taxa richness.
Axis 3 is negatively correlated with Total Phosphorus and Chlorophyll a

Table 11. Pearson and Kendall correlations between Plant Taxa and Ordination Axes for the ECO dataset. Taxa in bold have high correlations with the ordination axes. N = 60.

	Axis 1			Axis 2			Axis 3		
	r	r^2	tau	r	r^2	tau	r	r^2	tau
BRASSCHR	**-0.560**	0.314	-0.359	-0.297	0.088	-0.234	-0.127	0.016	-0.174
NYMPHAEA	**-0.479**	0.230	-0.386	0.111	0.012	0.076	0.264	0.070	0.143
POTAEPIH	**-0.464**	0.216	-0.296	-0.287	0.082	-0.270	-0.044	0.002	-0.072
MYRIFARW	**-0.459**	0.211	-0.245	**-0.327**	0.107	-0.285	0.015	0	-0.050
SPARGAFL	**-0.457**	0.209	-0.295	-0.288	0.083	-0.292	-0.031	0.001	-0.110
CERAECHI	**-0.405**	0.164	-0.257	-0.178	0.032	-0.222	0.132	0.017	0.033
DREPFONT	**-0.396**	0.156	-0.166	-0.186	0.035	-0.197	0.104	0.011	-0.157
POLYAMPH	**-0.389**	0.151	-0.187	-0.167	0.028	-0.143	0.117	0.014	0.011
POTAZOST	**-0.366**	0.134	-0.291	0.154	0.024	0.163	0.212	0.045	0.044
ACORCALA	**-0.355**	0.126	-0.125	-0.151	0.023	-0.090	0.065	0.004	-0.024
SCIRSUBT	**-0.331**	0.109	-0.234	-0.272	0.074	-0.249	-0.214	0.046	-0.205
ISOETES	**-0.307**	0.095	-0.201	-0.214	0.046	-0.160	-0.130	0.017	-0.201
POTAROBB	-0.295	0.087	-0.272	-0.267	0.071	-0.154	**-0.471**	0.222	-0.375
CERADEME	-0.252	0.063	-0.204	**0.427**	0.182	0.332	0.293	0.086	0.163
SPIRPOLY	-0.228	0.052	-0.167	**0.307**	0.094	0.207	-0.150	0.022	-0.062
MEGABECK	-0.172	0.030	-0.146	-0.200	0.040	-0.185	**-0.594**	0.352	-0.343
LEMNMINO	-0.169	0.029	-0.090	**0.374**	0.140	0.296	-0.014	0	-0.010
UTRIVULG	-0.160	0.026	-0.061	**-0.390**	0.152	-0.395	**0.316**	0.100	0.234
POTAPRAE	-0.139	0.019	-0.078	-0.110	0.012	-0.117	**-0.492**	0.242	-0.384
ELEOCHAR	-0.133	0.018	-0.067	-0.258	0.067	-0.328	**-0.306**	0.094	-0.177
POTAPECT	-0.129	0.017	0.039	**0.338**	0.114	0.252	0.216	0.047	0.353
POTANARR	-0.123	0.015	-0.114	-0.138	0.019	-0.074	**-0.633**	0.400	-0.383
POTAGRAM	-0.095	0.009	-0.122	-0.228	0.052	-0.285	**-0.363**	0.131	-0.256
POTAALPI	-0.078	0.006	-0.096	-0.090	0.008	-0.090	**-0.393**	0.154	-0.183
POTACRIS	**0.593**	0.352	0.463	**0.637**	0.406	0.524	-0.230	0.053	-0.201
MYRISPIC	0.016	0	-0.056	**0.398**	0.159	0.249	0.289	0.084	0.176

Table 12. Pearson and Kendall Correlations between Environmental Variables and Ordination Axes for the ALLKS dataset. The first axis is most highly (negatively) correlated with forested upland and natural land cover classes, and the Std Dev of Elevation. That axis is positively correlated with Developed and Agricultural land, and to a lesser extent with alkalinity and the Schupp lake classes. (N= 1984)

:	Axis 1			Axis 2				Axis 1			Axis 2		
Env Var	r	r^2	tau	r	r^2	tau	Env Var	r	r^2	tau	r	r^2	tau
Forest	**-0.759**	0.576	-0.584	-0.203	0.041	-0.125	**Dev + Ag**	**0.739**	0.546	0.598	0.225	0.051	0.126
Nat Land	**-0.739**	0.546	-0.598	-0.225	0.051	-0.126	**Agriculture**	**0.646**	0.417	0.518	0.204	0.042	0.101
Min Elev	**-0.538**	0.289	-0.353	-0.263	0.069	-0.2	**Alkalinity**	**0.350**	0.122	0.184	0.056	0.003	0.036
STD Elev	**-0.327**	0.107	-0.142	-0.068	0.005	-0.057	**Schupp Class**	**0.323**	0.105	0.24	0.121	0.015	0.099
Taxa Richness	-0.270	0.073	-0.258	-0.003	0	0.039	Developed	0.273	0.075	0.323	0.065	0.004	0.093
Wetlands	-0.224	0.050	-0.205	-0.059	0.003	-0.014	Total P (ppb)	0.267	0.071	0.162	0.171	0.029	0.031
Barren	-0.221	0.049	-0.284	-0.041	0.002	-0.096	Pop density	0.267	0.071	0.325	0.071	0.005	0.091
Open Upland	-0.174	0.030	-0.32	0.016	0	-0.086	Mean Chl a	0.238	0.057	0.205	0.15	0.023	0.026
Mean Secchi	-0.046	0.002	0.018	-0.167	0.028	-0.087	%Water	0.011	0	-	-	0.014	-
Color (PCu)	-0.025	0.001	0.102	0.041	0.002	0.042							
Depth max	-0.018	0	0.003	-0.126	0.016	-0.073							

Table 13. Pearson and Kendall Correlations between Plant Taxa and Ordination Axes for the ALLKS dataset. Taxa in bold represent those that are highly correlated with the first ordination axis. Taxa are ordered by their correlations with the first ordination axis. Negative correlations with Axis 1 represent an association with Forested Uplands and % Natural Land within 1 km of the lake. To a lesser extent this axis is also correlated with the Standard deviation of elevation, representing hilly landscapes. Strong positive correlations with Axis 1 represent a strong association with agricultural or developed land, and to a lesser extent with higher minimum elevation, higher alkalinity, and Schupp Class. (N= 1984)

	Axis 1			Axis 2		
Taxa Code	r	r^2	tau	r	r^2	tau
SPARGAFL	**-0.986**	0.972	-0.931	-0.070	0.005	-0.035
DULIARUN	**-0.985**	0.971	-0.929	0.001	0	0.022
EQUISETU	**-0.969**	0.939	-0.855	-0.107	0.011	-0.045
POTAEPIH	**-0.964**	0.929	-0.892	-0.064	0.004	-0.031
POTEPALU	**-0.964**	0.928	-0.884	0.010	0	0.024
IRISVERS	**-0.957**	0.915	-0.824	0.049	0.002	0.068
ERIOSEPT	**-0.944**	0.891	-0.904	-0.121	0.015	-0.058
ISOETES	**-0.941**	0.885	-0.863	-0.148	0.022	-0.076
CAREX	**-0.939**	0.882	-0.804	0.172	0.03	0.13
SCIRSUBT	**-0.909**	0.826	-0.863	-0.144	0.021	-0.088
NITELLA	**-0.908**	0.824	-0.752	-0.185	0.034	-0.108
BRASSCHR	**-0.906**	0.820	-0.819	-0.099	0.01	-0.062
CALLPALU	**-0.901**	0.812	-0.809	0.095	0.009	0.076
POTAROBB	**-0.896**	0.804	-0.778	-0.170	0.029	-0.096
POTASPIR	**-0.891**	0.793	-0.840	-0.068	0.005	-0.04
NUPHAR	**-0.886**	0.785	-0.795	-0.246	0.061	-0.128
ELEOSMAL	**-0.871**	0.758	-0.746	0.137	0.019	0.099
CALLITRI	**-0.869**	0.755	-0.849	-0.043	0.002	-0.012
UTRIMINO	**-0.862**	0.744	-0.809	-0.034	0.001	-0.017
POTAAMPL	**-0.852**	0.726	-0.714	**-0.367**	0.135	-0.225
LOBEDORT	**-0.849**	0.721	-0.838	-0.190	0.036	-0.114
MENYTRIF	**-0.833**	0.693	-0.850	-0.065	0.004	-0.036
SCIRCYPE	**-0.818**	0.669	-0.693	**0.328**	0.108	0.233
MEGABECK	**-0.814**	0.662	-0.671	-0.164	0.027	-0.085
LYCOAMER	**-0.803**	0.645	-0.718	0.243	0.059	0.163
DECOVERT	**-0.792**	0.627	-0.821	-0.149	0.022	-0.052
SAGITTAR	**-0.771**	0.595	-0.597	**-0.301**	0.091	-0.165
SIUMSUAV	**-0.769**	0.592	-0.674	0.178	0.032	0.147
POTAPECT	**0.988**	0.976	0.912	0.020	0	0.001
CERADEME	**0.966**	0.932	0.839	-0.028	0.001	-0.042
POTACRIS	**0.877**	0.770	0.784	0.159	0.025	0.099
LEMNMINO	**0.851**	0.724	0.659	**0.415**	0.172	0.294
LEMNTRIS	**0.837**	0.700	0.586	0.113	0.013	0.092
ELODEA	**0.808**	0.653	0.542	-0.105	0.011	-0.076
POTAZOST	**0.784**	0.615	0.510	-0.246	0.061	-0.226
HETERANT	**0.769**	0.592	0.571	0.064	0.004	0.062
WOLFFIA	**0.757**	0.572	0.699	0.236	0.056	0.176
SPIRPOLY	**0.752**	0.565	0.534	**0.411**	0.169	0.300
ALGAE	**0.739**	0.546	0.551	0.204	0.042	0.117
SCIRACUT	**0.673**	0.453	0.395	-0.131	0.017	-0.157
TYPHANGU	**0.657**	0.432	0.589	**0.430**	0.185	0.271
POTARICH	**0.650**	0.423	0.393	**-0.420**	0.177	-0.324
POTASTRI	**0.643**	0.413	0.545	0.041	0.002	0.019
MYRISPIC	**0.615**	0.379	0.741	0.089	0.008	0.054
UTRIVULG	**-0.751**	0.564	-0.591	-0.172	0.030	-0.090
MYRIALTE	**-0.739**	0.547	-0.780	-0.222	0.049	-0.139
ELATINE	**-0.726**	0.527	-0.766	-0.036	0.001	-0.011
UTRIINTE	**-0.724**	0.524	-0.651	-0.277	0.077	-0.177
LYSIMACH	**-0.720**	0.519	-0.735	0.130	0.017	0.110
MYRIFARW	**-0.710**	0.504	-0.767	-0.210	0.044	-0.159
POTAGRAM	**-0.697**	0.485	-0.561	**-0.368**	0.135	-0.227
PONTCORD	**-0.696**	0.485	-0.754	-0.009	0	0.004
POTAOBTU	**-0.692**	0.478	-0.777	-0.040	0.002	-0.024
NYMPHAEA	**-0.620**	0.385	-0.522	-0.202	0.041	-0.094
SUBUAQUA	**-0.602**	0.362	-0.735	-0.027	0.001	-0.030

Comparing Sampling Methods: Species Richness

In the composite dataset (ALLKS), 1895 lakes were sampled in a single dataset, 113 in two datasets, and 7 in three datasets (Table 6). These multiple samplings allow a comparison of the number and composition of species recorded. With some exceptions, the FISH dataset, collected by modified belt-transects, consistently recorded a greater number of species than the ECO, WJL or WPI datasets (Table 7). This is likely due to the greater area of the lake sampled by the fish transects, compared with the point estimates used in the remaining three methods. Both the FISH and ECO datasets recorded 150 to 230% more species compared with the WPI and WJL datasets**.** Mean richness in the FISH dataset was 24.5, compared with 19.5 for ECO and 12.8 and 10.4 for the WJL and WPI methods, respectively (Table 7**).** There was little overlap between the ECO and WJL/WPI datasets on individual lakes, so direct comparisons among these are not appropriate.

Table 14. Numbers of lakes for each combination of sampling effort, for the 120 lakes sampled in two or more datasets.

	FISH	ECO	WJL	WPI
FISH				
ECO	49			
WJL	36	3		
WPI	17	2	33	

The WJL and WPI datasets were typically in close agreement. The WJL dataset recorded an overall average of 2.4 more species per lake compared with WPI. In a direct comparison of lakes sampled by both methods, there was an average difference of 1 species of WJL over WPI (Table 15).

A few caveats must be considered in this comparison – data were frequently collected in different years, and it is likely that some differences are due to multi-year variation in species composition. Also, there is a considerable difference in the precision of the abundance measures among the methods, from the very qualitative measures recorded with the FISH transect method (rare, common, abundant), or the more quantitative occurrence by number of rake tosses used in the WJL method, to the quantitative point intercept methods used to sample ECO and WPI datasets. The ordination and classification methods described above rely heavily on abundance measures, and the community metrics calculated here may be more robust when derived from point-intercept data (ECO and WPI). Lastly, for the purpose of quantifying species richness, it appears that the ECO and FISH data sets result in similar taxa richness values to the composite data set, despite having sampled a very different number of lakes. With respect to detecting effects of disturbance, more analyses are needed to evaluate conclusively the effectiveness of the ECO versus the FISH datasets. Although much stronger correlations with environmental variables were apparent in the ECO dataset compared to FISH, the composite ALLKS dataset also resulted in high correlations.

Table 15. Species richness for lakes sampled with two or more methods. Maximum values are highlighted. Table also shows paired comparisons between different sampling methods

DOW LKNUM	ECO	FISH	WJL	WPI	Average Richness	Number of sampling methods	FISH-ECO DIFF	FISH-WJL DIFF	FISH-WPI DIFF	WJP-WPI DIFF
69037100	22	42	5		23.0	3	20	37		
41004500		18	3	3	8.0	3		15	15	0
66004800		13	5	7	8.3	3		8	6	-2
8002600		9	7	7	7.7	3		2	2	0
41008900	7	9		3	6.3	3	2		6	
66005000		8	3	4	5.0	3		5	4	-1
11001600	17		19	14	16.7	3				5
18003400	36	56			46.0	2	20			
31019000	26	47			36.5	2	21			
18008800	25	46			35.5	2	21			
31019100	28	46			37.0	2	18			
49001600	29	44			36.5	2	15			
73009200		43	27		35.0	2		16		
1004000		43	33		38.0	2		10		
49007900	32	42			37.0	2	10			
77008900	23	41			32.0	2	18			
1013600		40	20		30.0	2		20		
15001600	25	39			32.0	2	14			
34015800		38		24	31.0	2			14	
1021200		37	16		26.5	2		21		
1008900	32	37			34.5	2	5			
18037400	29	37			33.0	2	8			
31092100	25	37			31.0	2	12			
40005100	27	36			31.5	2	9			
77002300	24	36			30.0	2	12			
73012800		35	16		25.5	2		19		
49002400	20	35			27.5	2	15			
73022600		33	5		19.0	2		28		
18038700		33	20		26.5	2		13		
49013300		33	24		28.5	2		9		
15001000	23	33			28.0	2	10			
56038300		33		17	25.0	2			16	
1009600	28	32			30.0	2	4			
77006600		31	14		22.5	2		17		
73012300		31	21		26.0	2		10		
36001800		30	9		19.5	2		21		
1009900		30	12		21.0	2		18		
58007800		29	10		19.5	2		19		
40006300	22	29			25.5	2	7			
36002100		28	11		19.5	2		17		
33000900		28	14		21.0	2		14		
3010700	28	28			28.0	2	0			
38040600	23	28			25.5	2	5			
40003900	20	28			24.0	2	8			
11020100		26	17		21.5	2		9		
34025100	17	25			21.0	2	8			
18021200	26	24			25.0	2	-2			
34022400	13	24			18.5	2	11			
76014100		24		5	14.5	2			19	
3010200		23		19	21.0	2			4	
33001500	14	23			18.5	2	9			
47002600	10	23			16.5	2	13			
9001000		22	3		12.5	2		19		
86028200		22	15		18.5	2		7		
2006500		22		15	18.5	2			7	
47011900	17	22			19.5	2	5			
56076800	30	22			26.0	2	-8			
71001600	10	22			16.0	2	12			
76003300		22		9	15.5	2			13	
29011700		21	13		17.0	2		8		
1007700		21	17		19.0	2		4		
29004500		21	20		20.5	2		1		
73019600		21	22		21.5	2		-1		
21010100	21	20			20.5	2	-1			

Table 15. Species richness for lakes sampled with two or more methods. Maximum values are highlighted. Table also shows paired comparisons between different sampling methods

DOW LKNUM	ECO	FISH	WJL	WPI	Average Richness	Number of sampling methods	FISH-ECO DIFF	FISH-WJL DIFF	FISH-WPI DIFF	WJP-WPI DIFF
29025000	23	20			21.5	2	-3			
69074200	21	20			20.5	2	-1			
73009700	33	20			26.5	2	-13			
21010500	21	19			20.0	2	-2			
38039300	12	19			15.5	2	7			
34024600		18		7	12.5	2			11	
21009500	23	17			20.0	2	-6			
86005300	15	17			16.0	2	2			
11020000		16	6		11.0	2		10		
27006700	16	16			16.0	2	0			
30010000		15	10		12.5	2		5		
29021700		15	14		14.5	2		1		
18011000	8	15			11.5	2	7			
70001000	9	15			12.0	2	6			
71014700	5	15			10.0	2	10			
10008900		13	3		8.0	2		10		
56061300		13		13	13.0	2			0	
66005200	8	13			10.5	2	5			
86011400	5	13			9.0	2	8			
70009800		12	7		9.5	2		5		
6000100		12		5	8.5	2			7	
71008200	12	12			12.0	2	0			
30009600		11	12		11.5	2		-1		
10004400	10	11			10.5	2	1			
47003200	7	10			8.5	2	3			
66004400		9		6	7.5	2			3	
41002200		8		2	5.0	2			6	
62000600	16	8			12.0	2	-8			
71014500	5	8			6.5	2	3			
42002000		7	2		4.5	2		5		
24002800		2		2	2.0	2			0	
73027300			4	13	8.5	2				-9
76005700			4	8	6.0	2				-4
86025000			4	9	6.5	2				-5
73028500			5	8	6.5	2				-3
51006200			6	3	4.5	2				3
61003400			6	7	6.5	2				-1
26029400			7	7	7.0	2				0
15005900			8	15	11.5	2				-7
18003200			9	11	10.0	2				-2
52001000			9	7	8.0	2				2
52003400			10	18	14.0	2				-8
18038600	18		11		14.5	2				
4019800			12	17	14.5	2				-5
73025500			13	10	11.5	2				3
11007300			14	10	12.0	2				4
3057700			16	7	11.5	2				9
18010700			16	14	15.0	2				2
49002500			16	18	17.0	2				-2
80002700			16	15	15.5	2				1
80001800			17	11	14.0	2				6
80002800			18	12	15.0	2				6
80002200			20	17	18.5	2				3
1009200			23	10	16.5	2				13
49002600			24	16	20.0	2				8
18017500			26	12	19.0	2				14
Average	19.53	24.47	12.82	10.40			6.53	11.79	7.82	1.03

Exotic Species as Indicators of Disturbance (Objective 5)

There are four exotic species documented in the four original datasets: flowering rush (*Butomus umbellatus*), purple loosestrife (*Lythrum salicaria*), Eurasian milfoil (*Myriophyllum spicatum*), and curly-leaf pondweed (*Potamogeton crispus*). The presence of these species and the first year they were sampled are presented in Table 16.

Table 16. Exotics in Aquatic Vegetation Datasets

	Percent of lakes in datasets with exotic species present				
Exotic aquatic plants	ECO	FISH	WJL	WPI	grand total
Flowering Rush (*Butomus umbellatus*)	0.6	0.4	0.0	0.0	0.1
Purple Loosestrife (*Lythrum salicaria*)	0.0	5.0	0.0	0.3	1.2
Eurasian Milfoil (*Myriophyllum spicatum*)	18.2	3.7	0.1	0.0	2.6
Curly-leaf Pondweed (*Potamogeton crispus*)	33.7	22.2	1.4	3.4	9.5
	First year exotics documented in datasets				
	ECO	FISH	WJL	WPI	earliest year
Flowering Rush (*Butomus umbellatus*)	2004	1999			1999
Purple Loosestrife (*Lythrum salicaria*)		1993		2002	1993
Eurasian Milfoil (*Myriophyllum spicatum*)	2001	1993	1991		1991
Curly-leaf Pondweed (*Potamogeton crispus*)	2001	1993	1951	2002	1951

1. Flowering rush (*Butomus umbellatus*) is a narrow-leaved, grass-like plant that can have emergent, floating, or submersed leaves. It has a flowering stalk with pink flowers. It is a native of Europe and temperate Asia, and occurs in 14 northern US states from Idaho to Vermont and Connecticut. Flowering rush is considered an invasive plant, and is listed as noxious or invasive in 3 states (USDA PLANTS online database: http://plants.usda.gov). It first appeared in FISH dataset in 1999, and it is rare in the ECO and FISH datasets, appearing in less than 1% of the lakes in each. It was not recorded in either WJL or WPI datasets.

2. Purple loosestrife (*Lythrum salicaria*) is a showy flowering herbaceous plant that grows in shallow water and wetlands. It was introduced from Eurasia as a garden plant in upstate NY. It is now widespread in the U.S., found in 42 states. It is invasive and listed as a noxious or prohibited plant in 32 states (USDA PLANTS online database: http://plants.usda.gov). Purple loosestrife first appeared in the FISH dataset in 1993, and it is present in 5% of the lakes in the FISH dataset. It was first recorded in the WPI dataset in 2002, where it is rare, present in less than 1% of the lakes. Purple loosestrife was not recorded in either ECO or WJL datasets.

3. Eurasian milfoil (*Myriophyllum spicatum*) is a submersed aquatic plant with finely divided leaves that is native to Europe and Asia. It has spread into 38 states from Florida to Alaska. It is invasive and listed as noxious or invasive in 16 states (USDA PLANTS online database: http://plants.usda.gov). There are seven other species of milfoil that are native in the upper midwest, and Eurasian milfoil closely resembles northern water milfoil (*Myriophyllum sibiricum*), a native species. Eurasian milfoil first appeared in the WJL dataset in 1991, where it is rarely recorded – present in fewer than 1% of lakes. Eurasian milfoil was first reported in the

FISH dataset in 1993, and it is present in nearly 4% of FISH lakes. It was first recorded in the ECO dataset in 2001, and it is present in nearly 18% of the lakes sampled. Eurasian milfoil was not recorded in the WPI dataset. Ordination overlays showing distribution of Eurasian milfoil are shown in Figures ECO-14, FISH-14, and ALLKS-12.

4. Curly-leaf pondweed (*Potamogeton crispus*) is a submersed aquatic plant with relatively large, curly leaves that often grows up to the surface of the lake and is easily visible by mid-June. It looks similar to a native species, clasping-leaf pondweed (*Potamogeton richardsonii*). It is native to Europe, and the first U.S. collection was in Delaware in the mid-1800's. The first record in Wisconsin was in 1905. It has spread throughout the continental U.S., present in 46 states. It is invasive and listed as noxious or invasive in 4 states (USDA PLANTS online database: http://plants.usda.gov). Curly-leaf pondweed was first recorded in the WJL dataset in 1951, and it is present in all four datasets. Curly-leaf pondweed is common in the ECO and FISH datasets, present in nearly 34% of ECO lakes and 22% of FISH lakes. It is rarely recorded in WJL dataset, in 1.4% of lakes; and it is uncommon in WPI lakes, present in 3.4% of lakes. Ordination overlays showing distribution of curly-leaf pondweed are shown in Figures ECO-16, FISH-16, WJL-16, WPI-12, and ALLKS-14.

The presence of these exotic species, especially Eurasian milfoil and curly-leaf pondweed, are correlated with more disturbed landscapes, with a greater proportion of the surrounding buffer area made up of agricultural or developed lands. Eurasian milfoil and curly-leaf pondweed tend to occur in lakes with higher mean Chlorophyll-a, and higher total phosphorus. Therefore, their presence is a good indicator of a disturbed landscape.

The low numbers of both Eurasian milfoil and curly-leaf pondweed in the WJL and WPI datasets seem unusual compared to the FISH and ECO datasets. Since both these exotics are similar in appearance to native species that are common in our lakes, it is possible that both these species were overlooked by field staff conducting the WJL and WPI samples. The fact that the highest percentage of lakes with each of these species was recorded in the ECO dataset could well reflect the skills of the ECO sampling crews in identifying these species, rather than a bias towards sampling weedy lakes.

Data Integration (Objectives 1, 3)

As noted in Table 1, the datasets vary based on the number of lake sampled and methods used, with some degree of overlap among the various sampling methods. We assessed the degree to which data collected by the various methods could be combined for larger scale analyses. The lowest common denominator is presence/absence data, derived from the species list collected at the lake. Some sampling methods provide qualitative or quantitative information on species abundance and/or frequency; from these data it is possible to calculate relative frequency, and average or maximum abundance, as well as community descriptors such as species richness, Shannon diversity, or multivariate community metrics.

Confounding of the WJL and WPI data in the "Data Sample Stations Vegetation" table in the Wildlife Shallow Lakes database was discovered late in the course of this project while preparing comparisons between datasets. The WPI data needed to be completely edited and analyzed again after the WPI data were separated from the WJL data. At least we found that the

WPI data were suitable for analysis, after being separated from the WJL data. In contrast, the WJL data in this table are not the original field data. Instead the original data from 10 to 40 sample points in each lake has been summarized with one abundance value for each plant taxon in each lake. How these data were compiled is not at all clear. Without the original field data by sample point, it is not possible to evaluate how these abundance values were summarized: are they an average of the points sampled, or a maximum value, or the first one entered? Without more information on how the abundances were compiled, it seems safer to analyze presence/absence data rather than the questionable abundance data in the WJL dataset.

Given that these four datasets represent three different sampling methods with three widely varying amounts of area sampled, we find that the only appropriate common denominator for pooling data from all four datasets is presence/absence data. Although relative frequency values can be calculated from FISH, ECO, and WPI datasets, the sample area is much larger in the FISH belt-transects than in the point-intercept. Relative frequency values are sensitive to sample area size, so they cannot be reasonably compared between the FISH dataset and the two point-intercept datasets (ECO and WPI). Only the point-intercept data from the ECO, WPI, and WJL datasets could be pooled in a single dataset for analysis.

An alternative approach to comparing datasets is to compare the results of the community analysis procedures. Each dataset was classified using the Flexible Beta cluster analysis routine, and a range of groupings identified in each dataset. These groupings were reviewed in PCORD to evaluate how well they represent patterns of plant relative frequency or importance values, as well as how well they reflect geographic patterns of environmental variables such as water quality and chemistry. The numbers of lake types or groups that seem most meaningful were identified, and the composition of these broader lake types were compared across datasets.

Recommendations

Following are some recommendations regarding the management of lake monitoring data in DNR databases, and some thoughts on additional work that could be done with analysis of DNR aquatic vegetation datasets.

We found that there were large disparities in data quality across the four data sets. The ECO dataset was in excellent shape and was easy to summarize. The ECO database was also well-documented, with text files that explain data entry and a data dictionary or metadata information explaining the meaning of each field in the database. The ECO dataset was the smallest and most recent of the four we used, and both the small size and recent design may have contributed to the clean organization of data, and the thoroughness of documentation. This database could serve as a model for other programs to shoot for in maintaining and updating their databases.

The FISH dataset was the largest of the four we used. The FISH dataset had numerous problems that required extensive data editing and review, and in some cases interpretation. The FISH database was fairly well-documented, therefore, it was not difficult to interpret the data tables and data fields. However, the structure of the data made the data summary process very tedious. The VEGEFIL and VEGEDEP tables are examples of poorly designed data tables. Instead of assigning unique variable codes for each field sample (transect number), the same variable names are used for several different transect numbers, and the only way to interpret which transect was sampled was through the use of a second field. For example, as explained in Appendix A, a lake could have three records for a species (e.g. a water lily) in a sample called Transect 2; although one of them (with beginning transect field 1) is actually from transect 2, another one is actually from Transect 12, and the third is actually from Transect 22. The only way to distinguish these three “Transect 2” entries was to refer to the beginning transect (1, 11, or 21). There’s no reason to use two fields to define a variable name, when it can easily be represented in a single field. In order to sort out the data, the transect numbers were reassigned from 5 sets of 1 through 10 to one set of 1 through 52.

Another problem with the FISH dataset was the inconsistent use of blanks and zeroes (coded as “N” in the raw data, for not found). Sometimes a transect that was sampled had “N” entered for a species if the species was not found in that transect, and sometimes the field was left blank. There was no single field found that contained a tally of the number of transects sampled (the sample size) in each lake. Given that the sampling design allowed a wide variation in the number of samples recorded at each lake, it seems logical to keep track of the sample size. Most univariate or multivariate statistical tests require the use of sample size in the calculations. Comparisons between lakes are difficult unless the data are relativized by dividing summary statistics by the sample size. Keeping track of the sample size should be a basic part of any data collection and management. Since there was no field found indicating the number of transects sampled in each lake, the sample size had to be interpreted from the data. This was further complicated because of the inconsistent use of blanks and zeroes (“N”s in the raw data). The solution was to laboriously review the spreadsheets for two distinct patterns of blanks and zeroes in a table of the raw data with lake number and species codes in rows, and transect numbers in columns. The cells in the table contained the raw abundance codes (N, R, C, and A) converted to numbers (0, 1, 3, and 5). This large table (with 31,776 rows and 55 columns) was scanned to look for patterns of blank cells and filled cells. Large blocks of columns of blank cells including all the rows used to record data in one lake were found, and these were interpreted as transects not sampled. If a block of cells including all the rows for one lake had a scattered pattern of filled and blank cells, then the blank cells within that block were interpreted as zeros (transect sampled but species not found). The scanning, interpretation, and editing/correcting process took weeks. It is sometimes difficult to find financial support for data management; however, the usefulness of the data is strongly related to the quality of database design, consistency and accuracy of data entry, and quality control to make sure that data standards are met. Therefore we highly recommend that such an investment be made in this important effort.

The Wildlife Lakes database had a completely different set of problems from the FISH database. This database is the oldest database, including samples going back to the 1940’s. Aquatic vegetation data were collected using two sampling methods, and therefore it includes two aquatic vegetation datasets: WJL and WPI. A logical database design would keep the data sampled by different methods in different data tables. We were told that Jessen-Lound samples were in the “Data Vegetation” table, that Point-Intercept data were in the “Data Sample Stations Vegetation” table, and that the additional survey information (lake number, survey date) was in a third table that included surveys using both sampling methods. After editing and analyzing both datasets, we found that one of the vegetation tables actually included both sets of data. This is both a database design and documentation problem. The names of the tables did not distinguish the difference between the datasets, and there was no documentation regarding the table contents and the field contents. These were explained in phone calls from the database manager; however, the information provided was either incorrect or misunderstood. We discovered that the data from both methods were mixed together in the “Data Sample Stations Vegetation” table by comparing lakes sampled by WJL method to lakes sampled by WPI method. We found that every single WJL lake was also in the WPI dataset, and this seemed suspicious. We reviewed the original data tables and found identical data for about 385 project numbers. We also found that the older data in the WJL dataset must have been summarized in some way, because the data from individual sample points was not found, and each species had a single abundance value for each lake. How these abundance values were determined is unknown, and so the accuracy of the data is highly questionable. We felt obliged to use only presence/absence data with this dataset, because of the uncertainty of how the abundance data were compiled for each lake.

In summary, two of the three DNR programs that routinely gather data on aquatic vegetation (Fisheries, and Wildlife Shallow Lakes) require extensive database design and management efforts to make these data more useable. Financial support for data management and database design is an essential part of maintaining useful and retrievable data. Financial support and commitment to database design and management is just as important as support for the field monitoring efforts. Database design needs to take into consideration how the data will be used for a variety of purposes, and how data can be routinely updated and maintained.

References

Anonymous. 1993. Manual of instructions for lake survey. Minnesota Dept. of Natural Resources, Section of Fisheries, Lake Survey Program, St. Paul, MN.

Host, G. E., J. Schuldt, J. J. Ciborowski, L. B. Johnson, T. Hollenhorst, and C. Richards. 2005. Use of GIS and remotely-sensed data for *a priori* identification of reference areas for Great Lakes coastal ecosystems. International J. of Remote Sensing 26: *In Press.*

Jessen, J. R. and R. Lound. 1962. An evaluation of a survey technique for submerged aquatic plants. Minnesota Dept. of Conservation Division of Game and Fish. Section of Research and Planning Fish and Wildlife Survey Unit. Game Investigational Report No. 6, St. Paul, MN.

Madsen, J. D. 1999. Point intercept and line intercept methods for aquatic plant management. APCRP Technical Notes Collection (TN APCRP-M1-02). U.S. Army Engineer Research and Development Center, Vicksburg, MN www.wes.army.mil.el.aqua.

McCune, B. and J. B. Grace. 2002. Analysis of Ecological Communities. MjM Sofware Design, Gleneden Beach, Oregon, USA.

McCune, B. and M. J. Mefford. 1999. PC-ORD. Multivariate Analysis of Ecological Data, Version 4. MjM Software Design, Gleneden Beach, Oregon, USA.

Middleton, D. A. J. 1998. Statistical aspects of macrophyte assessment in Game Lakes. Final report submitted to Dr. Mark Hanson, Minnesota Dept. of Natural Resources, Division of Wildlife, Wetland Wildlife Populations & Research Group, Bemidji.

Newman, R. M. 1998. Assessing macrophytes in Minnesota’s game lakes. Final report submitted to Dr. Mark Hanson, Minnesota Dept. of Natural Resources, Division of Wildlife, Wetland Wildlife Populations & Research Group, Bemidji.

Nichols, S. A. 1999. Floristic quality assessment of Wisconsin lake plant communities with example applications. J. Lake and Reserv. Mgmt. 15(2):133-141.

Nichols, S. A. 1999. Distribution and habitat descriptions of Wisconsin Lake plants. WI Geological and Natural History Survey Bulletin 96.

Perleberg, D. 2001a. Evaluation of DNR Fisheries Lake Vegetation Survey Method. Minnesota Dept. of Natural Resources, Ecological Services Division. 1601 Minnesota Dr., Brainerd, MN 56401.

Perleberg, D. 2001b. Estimating species abundance and distribution: a comparison of three quantitative survey methods for lakewide assessment of submerged macrophytes. 1601 Minnesota Dr., Brainerd, MN 56401.

Perleberg, D. 2003. Floristic quality assessment of Minnesota’s aquatic macrophyte communities (draft report). 1601 Minnesota Dr., Brainerd, MN 56401.

Richards, C., L. B. Johnson, and G. E. Host. 1996. Landscape-scale influences on stream habitats and biota. Can. J. Fish. Aquat. Sci. 53:295-311.

Appendix A: Data Editing and Manipulations

1. Fisheries Data Set (FISH)

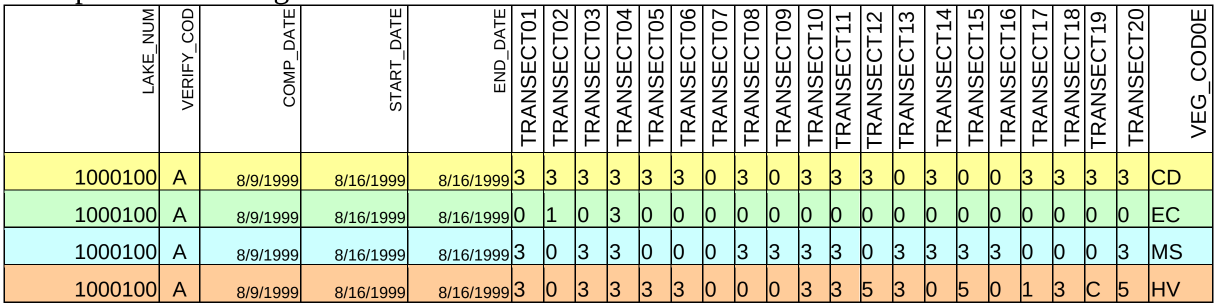
Fisheries data on aquatic vegetation of over 1500 lakes were provided within a much larger Access database that also included fish populations, water quality, and other environmental and geographical features of the lakes. Several database tables were integrated into a new file in order to edit the aquatic vegetation data for analysis with PCORD. The main table is called Vegefil.dbf, and has 76,274 records of aquatic vegetation presence and abundance classes from transects in lakes. The table includes DOW lake numbers, survey dates, and abundance codes for aquatic plants recorded in a variable number of transects, based on lake size. We retained all lakes represented by 9 or more transects.

The original data were written so that only 10 transects fit on one record or row in the data table. Therefore, if at least 41 transects were sampled there could be up to 5 rows (records) of data in the table. The data were structured so that the identity of each lake was recorded in two fields (lake number and sub-basin number) in each record. For example, a lake number in the data table consisted of a five or six character DOW lake number such as “10033” and a two character sub-basin code such as “00”. The identity of each transect was also recorded in two fields (beginning transect number and transect order after that beginning transect number).   For example, the 35th transect sampled would be identified in a record by a combination of the beginning transect number “31” (meaning that the ten transects recorded in that row were transects 31 through 40), and the transect number would be recorded as “5”.   For lakes in which 50 transects were sampled, data for one species was recorded in 5 records (rows of data); each row had a different beginning transect number (1, 11, 21, 31, and 41), but they had identical variables for the transect order number (1 – 10). Most lakes were sampled only one year, but some lakes were sampled several times, in different years. Therefore, the same lake might have two or three sets of up to 5 records for each species, one set for each year the vegetation was sampled.

Example of original vegefil records for a lake with 20 transects:

DOW_NUM	SUB_BASIN	VERIFY_COD	COMP_DATE	START_DATE	END_DATE	TRANSNUM	TRANSECT1	TRANSECT2	TRANSECT3	TRANSECT4	TRANSECT5	TRANSECT6	TRANSECT7	TRANSECT8	TRANSECT9	TRANSECT10	VEG_CODE
10001	00	A	8/9/1999	8/16/1999	8/16/1999	1	C	C	C	C	C	C		C		C	CD
10001	00	A	8/9/1999	8/16/1999	8/16/1999	11	C	C	N	C	N	N	C	C	C	C	CD
10001	00	A	8/9/1999	8/16/1999	8/16/1999	1	N	R	N	C	N	N	N	N	N	N	EC
10001	00	A	8/9/1999	8/16/1999	8/16/1999	11	N	N	N	N	N	N	N	N	N	N	EC
10001	00	A	8/9/1999	8/16/1999	8/16/1999	1	C	N	C	C	N	N	N	C	C	C	MS
10001	00	A	8/9/1999	8/16/1999	8/16/1999	11	C	N	C	C	C	C	N	N	N	C	MS
10001	00	A	8/9/1999	8/16/1999	8/16/1999	1	C	N	C	C	C	C	N	N	N	C	HV
10001	00	A	8/9/1999	8/16/1999	8/16/1999	11	C	A	C	N	A	N	R	C	C	A	HV

Example of edited vegefil records for a lake with 20 transects



LAKE_NUM	VERIFY_COD	COMP_DATE	START_DATE	END_DATE	TRANSECT01	TRANSECT02	TRANSECT03	TRANSECT04	TRANSECT05	TRANSECT06	TRANSECT07	TRANSECT08	TRANSECT09	TRANSECT10	TRANSECT11	TRANSECT12	TRANSECT13	TRANSECT14	TRANSECT15	TRANSECT16	TRANSECT17	TRANSECT18	TRANSECT19	TRANSECT20	VEG_COD0E
1000100	A	8/9/1999	8/16/1999	8/16/1999	3	3	3	3	3	3	0	3	0	3	3	3	0	3	0	0	3	3	3	3	CD
1000100	A	8/9/1999	8/16/1999	8/16/1999	0	1	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	EC
1000100	A	8/9/1999	8/16/1999	8/16/1999	3	0	3	3	0	0	0	3	3	3	3	0	3	3	3	3	0	0	0	3	MS
1000100	A	8/9/1999	8/16/1999	8/16/1999	3	0	3	3	3	3	0	0	0	3	3	5	3	0	5	0	1	3	C	5	HV

The first task in data editing was to create single field codes for lake numbers and for transect numbers. For the purpose of rearranging data, a combined code was created for each record that included lake number, plant species code, and year sampled. These codes would allow species abundance data to be sorted and rearranged into a matrix with 50 columns, one for each transect, with a unique code for each transect. This task was accomplished in several steps using both Access database queries, and Excel spreadsheets. The steps for rearranging the data table were as follows:

1) The Vegefil table was sorted in Access by “TRANSNUM”, which is the beginning transect number in each set of 10 transects. The Vegefil table was then divided into five tables, one for each beginning transect number. These five tables were then exported to Excel spreadsheets, and those spreadsheets were given file names reflecting the data subset: VegefilTransnum01, VegefilTransnum11, VegefilTransnum21, VegefilTransnum31, and VegefilTransnum41. Each of these spreadsheets was edited to renumber and rename the transects using sortable field names: TRSECT01 through TRSECT10 in the first spreadsheet, TRSECT11 through TRSECT20 in the second spreadsheet, etc.

2) In each of the 5 spreadsheets, leading zeros were added to the 5 character DOW numbers, so that all DOW numbers would consist of 6 characters to allow efficient sorting on that field. The fields “Dow_Num” and “Sub-Basin” were then combined into a new eight character field called “LAKE_NUM” using the *Concatenate* function in Excel.

3) The fields for “LAKE_NUM” and “VEG_CODE” were combined into a single field that I called “LAKENVEG” using the *Concatenate* function in Excel.

4) A new field for the year in which sampling was created by taking the right four characters from the “End_Date” field in the original database, using the *Right* function in Excel. I called this new field “ENDYEAR”.

5) The fields for “LAKENVEG” and “ENDYEAR” were combined into a new field “LAKENVEGNYR” using the *Concatenate* function in Excel. This combined field was later used as a common field to join data across the five data tables.

6) Letter codes for abundance categories were converted to numerical codes. The four abundance categories in the original dataset were A = abundant, C = common, R = rare, and N = not found. These categories reflect descriptive estimates, and no specific definitions were provided to samplers for these classes. A numerical abundance value was developed by DNR Fisheries **(**Anon. 1993**)**which was basically a "weighted frequency" that combined the frequency a species occurred across transects in a lake with the abundance values for each transect. The "weights" in their abundance value calculations were used for the replacement values for the four classes as follows: A = 5, C = 3, R = 1, and N = 0. This conversion was accomplished with the *Find and Replace* function in Excel.

7) Missing zeros were filled in the 5 separate spreadsheets. This corrected an inconsistency in data entry. Most lakes had transects with N's entered for species not found, but some lakes, or in some years for a lake, the N's were not entered, and only the A's, C's and R's were entered, and the fields that should have been N's were left blank. Since the whole data set had varying numbers of transects for each lake, *it is important to distinguish blank cells that mean the transect wasn't sampled from blank cells that mean the transect was sampled but the species recorded in that line was not found*. These missing N's or zeros were located by visually scanning through the thousands of lines of data in Excel, and looking for patterns of blank cells scattered through blocks of cells with the 1, 3, and 5 values. If completely blank blocks of cells occurred to the right of the block of cells for a lake with scattered values recorded, then it was assumed that the completely blank block meant *not sampled*, and the scattered blanks were replaced with zeros.

8) The five tables were imported into a new Access database, and combined using a query that joined the 5 tables on the "LAKENVEGNYR" field, with the join property set to include all the data in the first 10 transects, and then only the data in the second set that matched the "LAKENVEGNYR" field in the first set; and likewise for the additional three tables. This query was used to create a table in Access, and then that table was then exported into a new Excel spreadsheet called "VegefilTR0111213141fix2". This new spreadsheet has the original lake number, vegetation code, and sample date fields, the combined fields (described above), and 50 columns, one for each transect sampled. The abundance class codes are displayed as numbers as described above.

9) The combined spreadsheet "VegefilTR0111213141fix2" was sorted by lake, year sampled, and species, and then visually scanned for missing zeros. This was necessary for lakes with more than 10 transects, where some species were not entered for the higher number transects (10-52). These missing zeros appeared as partially blank lines within blocks of cells for a lake and sample year that had data entered for other species. As before, blocks of completely blank cells to the right of blocks of cells with partially or fully completed data were treated as if they were not sampled. There were a few areas where some zeros were recorded in the otherwise blank blocks of cells. If these no species were recorded as present in a transect at the right side of a block of clearly sampled transects, and zeros were entered for just a few species that had occurred in the first one or two transects in that set of 10 transects (from the original data set), then those zeros were treated as *not sampled* and removed from the data set. If there were any cells

in a block of cells for a lake that had at least one species present, then all the other blank cells were assumed to be sampled but “not found” and zeros were filled into all the blank cells in the block of cells for that lake and year sampled.

10) The combined spreadsheet needs to be compared with what we call “reverse-joined” data, to search for species that were not recorded in transects 1-10, but were added to the lake in transects 11-50. These were omitted in the first join because that join listed only species present in transects 1-10. The reverse-join requires two queries per pair of files. The first query is between the end set of transects (41-50) and the first combined set (1-50). The joined dataset includes all the lake number, species codes, and sample date fields, and the transect data from transects 1-10 followed by 41-50. The 2nd join is set to the common field LAKENVEGNYR, but this time the join points the other way, so all records in transects 41-50 are selected, and only those records from transects 1-10 with matching LAKENVEGNYR field. The resulting table is visually scanned to look for blank fields in transects 1-10, and those rows are highlighted. The highlighted rows are compared to the combined file to see if they are included; so far those ones have been missing from the combined file, so those species records need to be added to the combined table. In order to combine them, a second query is created that joins the first query to the other three tables on the LAKENVEGNYR field, but this time the join arrows all point to the right, so all the records in the reverse join are selected, and only those records from the other three subset that match on the LAKENVEGNYR field. The rows with the blanks in the first ten columns are highlighted (in spreadsheet), and these are selected to copy and add to the combined table spreadsheet. This new amended-combined file was then compared with the other data subsets: reverse join between transects 31-40 and the amended-combined file, between transects 21-30 and the second amended-combined file, and between transects 11-20 and the third amended-combined file. This should result in adding all species sampled in the transects 11-50 but not recorded from the first 10 transects. The fourth and final amended-combined file should now include all the species in the original data set, now rearranged into a matrix of 50 transects (columns) by about 36,000 rows of abundance class data for a single species in one lake in one sample year.

The second task in data editing was to summarize data across all the transects in each lake for each species, to yield a single value for each species in each lake. This would allow the creation of a matrix of species by lakes. There are many options for how to summarize the data. The simplest is presence/absence. Relative frequency across all the sampled transects can also be calculated. An average abundance class across all the transects can be calculated. Median or mode of abundance classes across all the transects could be a more meaningful summary, but some species that were sampled but found in less than half the transects would end up as zeros in the summarized value. One more possibility is calculating an importance value based on an algorithm that combines relative frequency and abundance values. A suggestion for calculating an importance value is to multiply relative frequency by average abundance class across all the transects sampled for each lake in each year sampled. The problem with any calculated value is the variable sample size, (i.e., some lakes include only 10 transects, while some lakes have up to

52 transects). Relativized data can yield single values that are comparable, even though the sample size, distribution, and standard deviations vary quite a bit across lakes of different sample size.

The third task in data editing was to create spreadsheets of lakes by species, with one of the summary values: presence/absence, relative frequency, or average abundance class. This was accomplished by importing the combined and edited spreadsheet into Access, and creating a series of queries. The data in the edited spreadsheet was imported into the Access database named fisheries_aqveg5. The first query (e.g. "Qry Final Select AqVeg_presabs_for Crosstab") selects the data fields from the full data set that will be used in a cross-tab query. Next is a cross-tab query is created that rearranges the data into the species by lake matrix, with the summary value in each cell (e.g. "Qry Final AqVeg_presabs_Crosstab"). Then a make-table query is run to save the results in a table (e.g. table "FinalAqVeg_PresAbs2"). The table is opened and then all fields are exported to an Excel spreadsheet (e.g. "FinalAqVeg_PresAbs2.xls). Final edits for deleting rare species or problematic lakes (e.g. with small sample size) are made in this spreadsheet. Finally one of the tabs in this spreadsheet has the edited data arranged in PCORD-compatible format, and that individual sheet is exported to a Lotus123 *.wk1 file format in the subdirectory PCORD. These *.wk1 files are used as the starting point for running the classifications and ordinations in PCORD (e.g. "FAQVprab.wk1").

The fourth task is to select the most recent year's data for lakes that were sampled in multiple years. In consultation with project officer Donna Perleberg, we eventually decided to use the most recent year's data for lakes with multiple years' data. A few lakes had sample sizes smaller than 10 transects. After reviewing the data, we decided to keep a few lakes with 9 transects, and delete several others with fewer than 9 transects (10005200, 73008300, 73008600, 73008700, and 73008900). These lakes remain in the Access database.

The fifth task was to review species nomenclature, and combine duplicate names for the same species or taxon (Appendix C, and see also spreadsheet in the data packet to be delivered at a later date). There were 220 species present in the original data set. The species list was reviewed in consideration of ease of field identification; difficult to identify species were combined into a species group or genus. Species with multiple synonyms were combined into the currently accepted name (using the USDA PLANTS online database as a reference). For example the narrow-leaved species of *Potamogeton* are difficult to distinguish in the field, so all the narrow-leaved species were combined into one taxon with the code POTANARR for "*Potamogeton* spp. narrow-leaved". Certain wetland herbs and shrubs were deleted from the dataset; these are species that do not occur in open lake water, but only on the wetland fringe. These revisions were completed in the Excel spreadsheet file called "COMBOvegefil52" by assigning a new lumped taxon code, called "NRRILUMP" for any NRRICODEs to be lumped into the combined taxon. This revised spreadsheet, called "COMBOvegefil52_lumpedveg5" was then imported into an Access database. A query was designed to combine multiple rows or entries of the NRRILUMP field in the same lake by using the maximum abundance entered for any of the original taxa lumped into the NRRILUMP code. The revised table was then exported to Excel to calculate revised values of relative frequency, average abundance and importance value for each NRRILUMP taxon in each lake. The revised spreadsheet, called was then imported back into an Access table called "COMBOvegefil52_LUMP2", and queries described above were run on this table to create Crosstab tables of lakes x NRRILUMP taxa, with presence/absence, relative frequency, average abundance, and importance value. The lumped and revised spreadsheets resulted in tables of 1505 lakes x 136 taxa.

The sixth task was to remove species rare in the dataset, because they can introduce confusing "noise" for analytical purposes. Therefore, species or taxa rarely observed were removed from the data (Appendix C, and data packet to be delivered at a later date). Species were sorted by raw frequency (total number of lakes in which the species was present). Species present in fewer than 16 lakes (1.07% of the lakes) were eliminated from the classification and ordination analyses (see Appendix B.1 for further details). There was a natural break point in the raw frequency data near the 1% raw frequency level: several species present in 15 or fewer lakes, none present in 16, 17, 18 or 19 lakes, and many species present in 20 or more lakes. Some of the species eliminated because of low raw frequency were wetland or upland edge species that probably should not have been included in the sample anyway: *Leersia oryzoides*, *Scirpus palustris*, and *Lysimachia* are some examples. Other taxa in this group are mostly rare or easily overlooked species. Removing these rare species left a few lakes without any vegetation; these were removed from the dataset. The final presence/absence data set included 103 taxa in 1500 lakes (Appendix C).

2. Ecological Services Dataset (ECO)

The Ecological Services "Aquatic Vegetation Database" includes data sampled using point-intercept sampling methods in 66 lake surveys. Sampling was conducted from 2000 – 2004. Tables of lakes by species (raw frequency) were easily produced using the crosstab function in Access; tables were then exported to Excel for editing. Raw frequency of each species was calculated by using the "count" function in the crosstab query: this provides a number of sample points where the species was found in each lake. The database includes 66 lake surveys in 60 lakes, mostly located in north-central Minnesota. Six lakes were sampled twice; earlier samples were removed from the dataset. There are a total of 87 species or taxa recorded in the database. To allow comparison with other datasets, the 87 species were assigned NRRILUMP codes to combine taxa that are difficult to distinguish in the field. This was necessary even though we trust the species identifications in the ECO dataset, because we want to compare between datasets. When species were lumped, the dataset consisted of 60 lakes and 73 taxa. Presence/absence data were summarized for 60 lake surveys by converting raw frequency values > 0 to "1" for presence and "0" was entered for any blanks in the table.

3. Wildlife Jessen-Lound Dataset (WJL)

The Shallow Lakes database, created by the DNR Division of Wildlife, includes vegetation data collected using two sampling methods: Jessen-Lound and Point-Intercept. These data are in separate tables in the "Wildlake2000" database. The Jessen-Lound vegetation data are in a table called "Data Vegetation". A total of 507 aquatic vegetation samples were collected by DNR Wildlife staff from 1946 to 1996 using the Jessen-Lound method. Some of the lakes were sampled in multiple years; in those cases only the most recent year's data were included in our analysis (with 3 exceptions explained below). After removing the earlier samples of lakes sampled in multiple years, the dataset consisted of 425 lakes.

In Jessen-Lound sampling, the sample area consists of four sub-samples, one taken at each corner of the boat. Each sub-sample consists of a visual survey and single rake toss which covers an area approximately one square meter. Some of the very old surveys from the 1940's do not have individual sample station data, but only species lists.

At each sample site, a species abundance rank is assigned based on the number of sub-samples in which a species is observed, as follows:

species found at 0 of 4 rake tosses are assigned an abundance rank of 0;
species found at 1 of 4 rake tosses are assigned an abundance rank of 1;
species found at 2 of 4 rake tosses are assigned an abundance rank of 2;
species found at 3 of 4 rake tosses are assigned an abundance rank of 3;
species found at 4 of 4 rake tosses are assigned an abundance rank of 4; and
species found at 4 of 4 rake tosses AND rake is very full on each toss are assigned an abundance rank of 5.

Of 7138 samples in the WJL dataset, 182 had no abundance rating, so for those records, an abundance code of 0 was assumed. The data in the "Data Vegetation" table include only a single abundance rating for each species in each lake. In a summary data matrix provided by Donna Perleberg, the number of points sampled per lake is reported as 10 to 40. So someone must have summarized original field data from elsewhere to boil it down to one abundance number for each species in each lake. It would be better to have the original data so we could calculate both relative frequency and average abundance. But without the original data, the best we can do is use this single abundance number (one abundance per species per lake) and assume that whoever summarized the data did it correctly. Since we don't know how these single abundance values were compiled from the original data, it is safer to use these to create a presence/absence dataset.

The abundance data were compiled using a crosstab query in the Access database to create a matrix of lakes by species, showing the abundance value recorded for each species per lake. This matrix was then used to create a presence/absence dataset, converting all values >0 to 1. Some lakes were sampled in multiple years; only the most recent data were included in our analysis (with 3 exceptions explained in Appendix B.3.) This presence/absence dataset was then modified by applying a "Beals smoothing" algorithm, also known as the sociological favorability index. Beals smoothing replaces each cell in the data matrix with a probability of each taxon occurring in a sample unit, based on the joint occurrences of the target taxon with the other taxa that are actually present in the sample unit. This algorithm helps solve a problem with a statistical distribution of the data known as the "zero truncation problem", in which there are a large number of zeroes in the dataset. This algorithm replaces presence/absence data with quantitative values that represent the "favorability" of each sample for each plant taxon, regardless of whether the taxon was actually present in that sample (McCune and Grace, 2002). The final dataset consisted of 425 lakes and 83 taxa (Table 1).

4. Wildlife Point-Intercept Dataset (WPI)

The Shallow Lakes database, created by the DNR Division of Wildlife, includes vegetation data collected using two sampling methods: Jessen-Lound and Point-Intercept. These data are supposed to be in separate tables in the Wildlife Lakes Survey database (the file name is

"Wildlake2000"). Data from Point-Intercept samples are stored in the table called "Data Sample Stations Vegetation" in Wildlake2000; however a copy of the entire WJL dataset is also included in the same table. The WJL data are stored with the "DSSV-station" field entered as "999". Data were sorted on the field "DSSV-station", and all records with "999" in the "DSSV-station"field were removed. This left a total of 165 lake samples that were recorded using the point-intercept sampling method (Table 1). Sampling was conducted from 1988 to 2003. Five lakes were sampled in multiple years; only the most recent data were included in the analysis. In point-intercept sampling, the sample area consists of a visual survey and single rake toss which covers an area of approximately one square meter. The point intercept sampling method was refined in more recent years to include an evenly spaced grid of sample points across a lake, where empty sample points are also recorded. Some of the older samples do not record "empty" points where no aquatic vegetation was found.

The NRRILUMP codes were added to the dataset, then the raw frequency or number of observations of each plant taxon was tabulated by using a crosstab query in Access (Appendix B.4). The query was set up so that lakes were in rows, plant taxa were in columns, and the cells in the table calculated a count of all the DSSV stations (sample points) where each species or taxon was observed. This yielded a table of 165 lakes and 76 plant taxa. The table was then sorted by the NRRILUMP codes and data were examined for how to combine the records of taxa that needed to be lumped. Raw frequency values were combined by using the maximum value recorded for the component taxa as the value for the lumped taxon, leaving a table of 165 lakes and 64 lumped taxa.

Some lakes were sampled with very few points, others with hundreds of points. In future analyses the lakes with less than 9 total observations of any aquatic plant taxon will be omitted due to their small sample. Some lakes were sampled in multiple years; only the most recent data were included in our analysis. Some lakes were empty after rare taxa and wetland edge taxa were removed, so three more lakes were deleted. The final dataset consists of 157 lakes and 64 taxa.

After the WPI data were separated from the WJL data, a table of project numbers by vegecodes was created using the crosstab function in Access. Lake numbers and sample dates were added to this table by using a query in Access and the "Data Sample Stations" table as a source of the lake numbers and sample dates for each project number. This table of 165 lakes and 109 taxa was exported to an Excel spreadsheet. The list of vegecodes were copied into a new sheet and imported back into Access. A query was used to match the vegecodes with NRRILUMP codes and species names. The resulting table was exported back out to Excel and pasted into the table of 165 lakes by 109 taxa. Next the columns representing each taxon were sorted by their NRRILUMP codes so taxa to be lumped would appear side by side. The next step was to evaluate whether the raw frequencies of individual taxa that were to be lumped could be added together. This required review of the original data table to see if the different vegecodes to be lumped were recorded in separate sample points versus whether different to-be-lumped vegecodes were recorded at the same sample point. If the to-be-lumped taxa were recorded at different sample stations, then the raw frequencies from adjacent columns could simply be added together. If different to-be-lumped vegecodes were reported at the same sample station, the duplicate entries were highlighted and counted. The number of duplicate entries for a sample point was then subtracted from one of the soon-to-be-lumped taxa's raw frequency, so the adjacent columns could be added together without over-representing the lumped taxa. After this review, the columns of to-be-lumped taxa were added together to produce a raw frequency for the newly lumped taxon, and the columns for the individual taxa that had been lumped were deleted.

The sum of raw frequencies was calculated for each species across all the lakes. This list was copied to a new worksheet and sorted by sum of raw frequencies. Taxa that were rarest in the dataset, with only one observation or hit across all the lakes, were removed from the data set. Also several wetland edge herbs and shrubs were removed from the dataset; this resulted in a data table with 165 lakes and 64 taxa. Five lakes were sampled in multiple years, so the earlier samples were removed. Three lakes had no species left after the wetland edge species were removed. The reduced dataset was then 157 lakes and 64 plant taxa.

Some lakes were sampled with very few points, others with hundreds of points. Twelve lakes with fewer than 7 points sampled were deleted for their small sample size. The WPI dataset then consisted of 145 lakes and 64 taxa. When this dataset was analyzed with NMS, it was not possible to find a stable solution (McCune and Medford 1999). So an outlier analysis was run on the 145 lakes dataset, and seven outlier lakes were identified that were greater than 2.5 standard deviations from the mean. These seven lakes were removed and one additional taxon was then removed because it was no longer present in the dataset. The final dataset used in the analyses consisted of 138 lakes and 63 plant taxa.

5. All Lakes from Four Datasets Combined (ALLKS)

Each of the four datasets was either compiled as presence/absence data, or converted to presence/absence data using a power transformation of x^0 in PC-ORD. The four datasets were then combined into one composite presence/absence dataset that consisted of 1994 lakes and 131 taxa. These presence/absence values were modified using the "Beals smoothing" algorithm described above for the WJL data.

6. Plant Community Types Dataset (PC Types)

This dataset was prepared by converting the data in each dataset separately to presence/absence data, and sorting the presence/absence data in each dataset by the PC Type that the lake was assigned in the dataset's flexible beta classification (or the modified version for WJL). Then a summary composition of each PC Type was prepared based on relative frequency of each taxon across all the lakes in a single PC Type. This produced, for each PC Type, a relative frequency of each taxon in each dataset. Since these relative frequency values were all derived from presence/absence data, they are comparable between datasets.

Appendix B: Derived Data Calculations

1. Fisheries (FISH)

a. Relative Frequency:
Relative frequency of each of 103 taxa per lake was calculated using the formula:

$$\text{relative frequency} = \frac{\text{number of transects with plant taxon present in a lake}}{\text{total number of transects sampled in the same lake.}}$$

This was calculated in an Excel spreadsheet.

b. Average abundance:
Average abundance of each of 103 taxa per lake was calculated using the formula:

$$\frac{\text{sum of abundance estimate values for a plant taxon in a lake}}{\text{total number of transects sampled in a lake.}}$$

This was calculated in an Excel spreadsheet.

c. Importance value:
Importance value (IV) for each taxon per lake was calculated using the formula:

$$\text{IV} = \text{relative frequency} \times \text{average abundance.}$$

Importance values were calculated in an Excel spreadsheet using values generated by the above formulas. These importance values were used in the classification and ordination analyses.

d. Taxa Richness: Taxa Richness for each taxon was calculated in the vegetation spreadsheets after the wetland edge herb and shrub taxa were removed, but before the taxa rare in the dataset were removed for quantitative analyses. Taxa Richness is a count of how many different taxa are present in a lake. Therefore the Taxa Richness field may be slightly higher than the Shannon diversity index calculated using the reduced dataset that was used in quantitative analyses.

e. Average Relative Frequency:
This was calculated in the Flexible Beta plant community types composition table in order to compare types. The sum of all the relative frequencies for the group of lakes in each type of each plant taxon was calculated, and then this sum of relative frequencies was divided by the number of lakes in the group or type. This derived value helps sort out which taxa are the dominant taxa within each Flexible Beta type.

2. Ecological Services (ECO)

a. Relative Frequency

The total number of observations (hits) of all species was calculated for each lake in a spreadsheet. Then the relative frequency of each species was calculated for each lake using the following formula:

$$\text{relative frequency} = \frac{\text{number of hits of a single plant taxon in a lake (raw frequency)}}{\text{total number of hits of all plant taxa in the same lake.}}$$

These relative frequency values were used in the classification and ordination analyses.

b. Taxa Richness: Taxa Richness for each taxon was calculated in the vegetation spreadsheets after the wetland edge herb and shrub taxa were removed, but before the taxa rare in the dataset were removed for quantitative analyses. Taxa Richness is a count of how many different taxa are present in a lake. Therefore the Taxa Richness field may be slightly higher than the Shannon diversity index calculated using the reduced dataset that was used in quantitative analyses.

c. Average Relative Frequency:
This was calculated in the Flexible Beta plant community types composition table in order to compare types. The sum of all the relative frequencies for the group of lakes in each type of each plant taxon was calculated, and then this sum of relative frequencies was divided by the number of lakes in the group or type. This derived value helps sort out which taxa are the dominant taxa within each Flexible Beta type.

3. Shallow Lakes - Wildlife Jessen-Lound (WJL)

a. Presence/absence Data:

The raw abundance data, consisting of a matrix of lakes by species with the abundance value recorded for each species per lake. This matrix was then used to create a presence/absence dataset, converting all values >0 to 1. This presence/absence dataset was then modified by applying a “Beals smoothing” algorithm, also known as the sociological favorability index. Beals smoothing replaces each cell in the data matrix with a probability of each taxon occurring in a sample unit, based on the joint occurrences of the target taxon with the other taxa that are actually present in the sample unit. This algorithm helps solve a problem with a statistical distribution of the data known as the “zero truncation problem”, in which there are a large number of zeroes in the dataset. This algorithm replaces presence/absence data with quantitative values that represent the “favorability” of each sample for each plant taxon, regardless of whether the taxon was actually present in that sample (McCune and Grace, 2002).
These Beals smoothing values were used in the classification and ordination analyses.

b. Taxa Richness: Taxa Richness for each taxon was calculated in the vegetation spreadsheets after the wetland edge herb and shrub taxa were removed, but before the taxa rare in the dataset were removed for quantitative analyses. Taxa Richness is a count of how many different taxa are present in a lake. Therefore the Taxa Richness field may be slightly higher than the Shannon diversity index calculated using the reduced dataset that was used in quantitative analyses.

c. Average Relative Frequency:
This was calculated in the Flexible Beta plant community types composition table in order to compare types. The sum of all the relative frequencies for the group of lakes in each type of each plant taxon was calculated, and then this sum of relative frequencies was divided by the number of lakes in the group or type. This derived value helps sort out which taxa are the dominant taxa within each Flexible Beta type.

4. Shallow Lakes - Wildlife Point-Intercept (WPI)

a. Relative Frequency

The total number of observations of all species was calculated for each lake in a spreadsheet. Then the relative frequency of each species was calculated for each lake using the following formula:

$$\text{relative frequency} = \frac{\text{number of hits of a single plant taxon in a lake (raw frequency)}}{\text{total number of hits of all plant taxa in the same lake.}}$$

These relative frequency values were used in the classification and ordination analyses.

b. Taxa Richness: Taxa Richness for each taxon was calculated in the vegetation spreadsheets after the wetland edge herb and shrub taxa were removed, but before the taxa rare in the dataset were removed for quantitative analyses. Taxa Richness is a count of how many different taxa are present in a lake. Therefore the Taxa Richness field may be slightly higher than the Shannon diversity index calculated using the reduced dataset that was used in quantitative analyses.

c. Average Relative Frequency:
This was calculated in the Flexible Beta plant community types composition table in order to compare types. The sum of all the relative frequencies for the group of lakes in each type of each plant taxon was calculated, and then this sum of relative frequencies was divided by the number of lakes in the group or type. This derived value helps sort out which taxa are the dominant taxa within each Flexible Beta type.

5. All Lakes from Four Datasets Combined (ALLKS)

a. Taxa Richness: Taxa Richness for each taxon was calculated in the vegetation spreadsheets after the wetland edge herb and shrub taxa were removed, but before the taxa rare in the dataset were removed for quantitative analyses. Taxa Richness is a count of how many different taxa are present in a lake. Therefore the Taxa Richness field may be slightly higher than the Shannon diversity index calculated using the reduced dataset that was used in quantitative analyses.

6. Plant Community Types Dataset (PC Types)

a. Relative Frequency in each PC Type

The total number of lakes was calculated for each PC Type in a spreadsheet. Then the relative frequency of each species was calculated for each PC Type using the following formula:

$$\text{relative frequency in each PC Type} = \frac{\text{number of lakes within a PC Type with a single plant taxon present (raw frequency)}}{\text{total number of lakes in that PC type.}}$$

These relative frequency values were used in the classification and ordination analyses.

Appendix C: Final list of taxa, including wetland species and those removed from analyses due to rarity in a dataset.

Rare species (occurring in < 16 lakes) were removed from dataset prior to ordination analyses.

Notes: *a*) removed from analysis because it was a wetland edge species that did not grow in open water; *b*) removed from FISH analysis due to rarity; *c*) removed from WJL analysis due to rarity; *d*) removed from WPI analysis due to rarity. A spreadsheet with more detailed notations will be delivered with the data package.

Scientific Name	ECO	FISH	WJL	WPI	Lumped to – Scientific Name
Acorus calamus	X	X	X		*Acorus calamus (= americanus)*
Agrostis stolonifers (alba)		X	X		*Agrostis stolonifera (= alba)* [b,c]
Algae		X	X		*algae (includes blue-green, green, filamentous, and planktonic algae)*
Blue-green Algae		X	X		
Green algae			X		
Filamentous Algae		X	X	X	
Planktonic Algae		X	X		
Alisma gramineum		X			*Alisma sp. (includes A. gramineum, A. subcordatum, A. triviale)*
Alisma sp.		X			
Alisma triviale		X	X		
Alnus sp. [a]		X	X		
Alopecurus aequalis		X			*Alopecurus aequalis* [b]
Andromeda glaucophylla [a]		X			
Asclepias incarnata		X	X		*Asclepias incarnata*
Aster sp. [a]		X		X	
Beckmannia syzigachne		X	X		*Beckmannia syzigachne* [b,c]
Betula glandulifera [a]		X	X		
Bidens cernua		X	X		*Bidens sp. (includes B. cernua)*
Bidens sp.		X			
Boltonia asteroides		X			*Boltonia asteroids* [b]
Brasenia schreberi	X	X	X	X	*Brasenia schreberi*
Butomus umbellatus	X	X			*Butomus umbellatus (EXOTIC)* [b]
Calamagrostis canadensis		X	X		*Calamagrostis canadensis*
Calla palustris		X	X		*Calla palustris* [c]
Callitriche sp.		X			*Callitriche sp. (includes C. palustris)*
Callitriche palustris (=verna)		X			
Caltha palustris		X	X		*Caltha palustris* [c]
Carex aquatilis		X	X		*Carex sp. (includes C. aquatilis, C. comosa, C. lacustris, C. lasiocarpa)*
Carex comosa		X	X		
Carex lacustris		X			
Carex lasiocarpa		X			
Carex sp.	X				
Carex sp.			X		
Carex sp.	X	X	X	X	
Carex sp.			X		
Cyperaceae		X	X		
Ceratophyllum demersum	X	X	X	X	*Ceratophyllum demersum*
Ceratophyllum echinatum	X				*Ceratophyllum echinatum*
Chamaedaphne calyculata [a]		X	X		
Chara sp.	X	X	X	X	*Chara sp.*
Cicuta bulbifera		X	X		*Cicuta bulbifera* [c]
Cicuta maculata		X	X		*Cicuta maculate* [c]

Scientific Name	ECO	FISH	WJL	WPI	Lumped to – Scientific Name
Cornus sp. [a]		X			
Cyperus esculentus		X	X		*Cyperus sp. (includes C. esculentus)* [c]
Cyperus sp.		X	X		
Decodon verticillata		X	X		*Decodon verticillata* [c]
Drepanocladus sp.	X	X	X	X	*Drepanocladus; Fontinalis; etc.*
Drepanocladus; Fontinalis; etc.	X	X		X	
Fontinalis sp. or Drepanocladus sp.			X		
Dulichium arundinaceum		X	X	X	*Dulichium arundinaceum* [d]
Echinochloa muricata (=E. pungens) [a]		X	X		
Elatine minima		X			*Elatine sp.* [b]
Elatine sp.		X			
Eleocharis acicularis	X	X	X		*Eleocharis acicularis*
Eleocharis palustris (=smallii)		X	X		*Eleocharis palustris (= smallii)*
Eleocharis sp.	X	X	X	X	*Eleocharis sp. (includes E. palustris, E.*
Elodea canadensis	X	X	X	X	*Elodea sp. (includes E. canadensis, E. nuttalii)*
Elodea spp.	X	X		X	
Elymus virginicus [a]		X	X		
Equisetum fluviatile	X	X	X		*Equisetum sp. (includes E. fluviatile)*
Equisetum sp.	X	X			
Eragrostis spp. [a]		X			
Eriocaulon septangulare	X	X	X		*Eriocaulon septangulare*
Eriophorum sp.		X			*Eriophorum sp.* [b]
Eupatorium maculatum [a]		X	X		
Eupatoruim perfoliatum [a]		X	X		
Galium sp. [a]		X	X		
Glyceria borealis [a]		X	X		
Glyceria grandis [a]		X	X		
Glyceria sp. [a]		X			
Gramineae Family [a]		X	X	X	
Heteranthera dubia	X	X	X	X	*Heteranthera sp. (includes H. dubia)*
Heteranthera sp.		X	X		
Hippuris vulgaris	X	X	X	X	*Hippuris vulgaris*
Hordeum jubatum [a]		X	X		
Hypericum ellipticum [a]		X			
Hypericum sp. [a]		X			
Impatiens capensis [a]		X			
Impatiens sp. [a]		X	X		
Iris sp.	X	X			*Iris sp. (includes I. versicolor, I . virginica)*
Iris versicolor		X	X		
Isoetes echinospora		X			*Isoetes sp. (includes I. echinospora, I. macrospora, I. melanopoda)* [d]
Isoetes macrospora		X			
Isoetes sp.	X	X	X	X	
Juncus balticus		X	X		*Juncus sp. (includes J. balticus, J. effusus, J. filiformis, J. nodosus)*
Juncus effusus		X			
Juncus nodosus		X	X		
Juncus sp.		X	X		
Labiatae Family		X	X		*Labiatae Family (includes Mentha arvensis)* [b,c]
Mentha arvensis		X			
Philonotis sp.		X			
Ledum groenlandicum [a]		X	X		
Leersia oryzoides		X	X		*Leersia oryzoides* [b]
Lemna minor	X	X	X	X	*Lemna minor*
Lemna trisulca	X	X	X	X	*Lemna trisulca*
Lobelia cardinalis [a]		X			
Lobelia dortmanna		X	X		*Lobelia dortmanna* [c]
Lycopus americanus		X			*Lycopus americanus*
Lysimachia sp.		X			*Lysimachia sp. (includes L.. terrestris)* [b]

Scientific Name	**ECO**	**FISH**	**WJL**	**WPI**	**Lumped to – Scientific Name**
Lysimachia terrestris		X			
Lythrum salicaria		X		X	*Lythrum salicaria (EXOTIC)* [d]
Megalondonta/Bidens beckii	X	X	X	X	*Megalondonta/Bidens beckii*
Menyanthes trifoliata		X			*Menyanthes trifoliata*
Mimulus ringens		X			*Mimulus ringens* [b]
Myriophyllum alterniflorum		X	X		*Myriophyllum alterniflorum* [c]
Myriophyllum farwellii	X	X		X	*Myriophyllum farwellii* [b,d]
Myrica gale [a]		X			
Myriophyllum heterophyllum		X			*Myriophyllum heterophyllum* [b]
Myriophyllum sp.	X	X		X	*Myriophyllum sp.*
Myriophyllum sibiricum	X	X	X	X	*Myriophyllum sibiricum*
Myriophyllum spicatum	X	X	X		*Myriophyllum spicatum (EXOTIC)* [c]
Myriophyllum tenellum	X	X	X	X	*Myriophyllum tenellum*
Myriophyllum verticillatum	X	X			*Myriophyllum verticillatum* [b]
Najas flexilis	X	X	X	X	*Najas flexilis*
Najas gracillima		X	X		*Najas gracillima*
Najas guadalupensis		X	X		*Najas guadalupensis* [b,c]
Najas marina	X	X	X	X	*Najas marina*
Najas sp.	X	X		X	*Najas sp.*
Nelumbo lutea		X	X	X	*Nelumbo lutea* [b,c,d]
Nitella sp.	X	X	X	X	*Nitella sp.*
Nuphar sp.		X		X	*Nuphar sp. (includes N. lutea, N. microphylla, N. rubrodisca, N. variegata*
Nuphar lutea ssp. variegata (=N.	X	X	X	X	
Nuphar lutea ssp rubrodisca (=N.		X	X		
Nuphar lutea ssp. pumila (=N.		X		X	
Nymphaea sp.		X		X	*Nymphaea sp. (includes N. leibergii, N. odorata, N. tuberosa)*
Nymphaea leibergii		X	X		
Nymphaea odorata	X	X	X	X	
Nymphaea odorata ssp. tuberosa (=N.		X			
Phalaris arundinacea [a]		X	X		
Phragmites australis (communis)	X	X	X	X	*Phragmites australis (= communis)*
Polygonum amphibium	X	X	X		*Polygonum amphibium (= coccinium)*
Polygonum sp. [a]		X	X	X	*Polygonum sp.*
Polygonum lapathifolium [a]		X	X		
Pontederia cordata		X	X	X	*Pontederia cordata*
Potamogeton alpinus	X	X	X		*Potamogeton alpinus* [b,c]
Potamogeton amplifolius	X	X	X	X	*Potamogeton amplifolius*
Potamogeton sp.		X		X	*Potamogeton sp.*
Potamogeton crispus	X	X	X	X	*Potamogeton crispus (EXOTIC)*
Potamogeton diversifolius		X		X	*Potamogeton diversifolius* [b]
Potamogeton epihydrus	X	X	X	X	*Potamogeton epihydrus*
Stuckenia filiformis (=Potamogeton	X	X	X		*Potamogeton sp. (includes P. filiformis, P. foliosus, P. friesii, P. hillii, P. pusillus)*
Stuckenia filiformis ssp. occidentalis			X		
Potamogeton foliosus	X	X	X	X	
Potamogeton friesii	X	X	X	X	
Potamogeton hillii				X	
Potamogeton sp.		X			
Potamogeton sp. (narrow lvs)	X	X	X	X	
Potamogeton pusillus	X	X	X	X	
Potamogeton gramineus	X	X	X	X	*Potamogeton gramineus*
Potamogeton illinoensis	X	X	X	X	*Potamogeton Illinoensis*
Potamogeton natans	X	X	X	X	*Potamogeton natans*
Potamogeton nodosus		X	X	X	*Potamogeton nodosus*
Potamogeton obtusifolius		X	X		*Potamogeton obtusifolius* [c]
Stuckenia pectinata (=Potamogeton	X	X	X	X	*Stuckenia pectinata (=Potamogeton*
Potamogeton praelongus	X	X	X	X	*Potamogeton praelongus*
Potamogeton richardsonii	X	X	X	X	*Potamogeton richardsonii*
Potamogeton robbinsii	X	X	X	X	*Potamogeton robbinsii*

Scientific Name	ECO	FISH	WJL	WPI	Lumped to – Scientific Name
Potamogeton spirillus	X	X			*Potamogeton spirillus*
Potamogeton strictifolius		X	X		*Potamogeton strictifolius*
Potamogeton vaginatus		X	X		*Potamogeton vaginatus (=Stuckenia*
Potamogeton vaseyi		X			*Potamogeton vaseyi* [b]
Potamogeton zosteriformis	X	X	X	X	*Potamogeton zosteriformis*
Potentilla palustris		X	X	X	*Potentilla palustris* [d]
Ranunculus flabellaris		X			*Ranunculus flabellaris*
Ranunculus flammula	X	X			*Ranunculus flammula* [b]
Ranunculus gmelini		X			*Ranunculus gmelini* [b]
Ranunculus sp.	X	X	X	X	*Ranunculus sp. (includes R. longirostris, R. sceleratus)*
Ranunculus sceleratus		X			
Riccia fluitans	X	X	X		*Riccia fluitans* [b,c]
Rorippa spp.		X			*Rorippa sp. and related genera* [b]
Rumex maritimus		X			*Rumex sp. (includes R. maritimus, R. orbiculatus)*
Rumex orbiculatus		X	X		
Rumex sp.		X	X		
Ruppia occidentalis		X	X	X	*Ruppia occidentalis*
Sagittaria cristata		X	X		*Sagittaria cristata*
Sagittaria cuneata	X	X	X		*Sagittaria cuneata*
Sagittaria graminea		X			*Sagittaria graminea* [b]
Sagittaria latifolia		X	X	X	*Sagittaria latifolia*
Sagittaria rigida		X	X	X	*Sagittaria rigida*
Sagittaria sp.	X	X	X	X	*Sagittaria sp.*
Salix sp. [a]		X	X	X	
Schoenoplectus acutus (=Scirpus acutus)	X	X	X	X	*Schoenoplectus acutus (=Scirpus acutus)*
Scirpus atrovirens		X	X		*Scirpus atrovirens*
Scirpus cyperinus	X	X	X		*Scirpus cyperinus*
Schoenoplectus fluviatilis (=Scirpus		X	X	X	*Schoenoplectus fluviatilis (=Scirpus*
Scirpus heterochaetus [a]			X		
Schoenoplectus maritimus (=Scirpus		X	X		*Schoenoplectus maritimus (=Scirpus*
Schoenoplectus americanus (=Scirpus	X	X	X	X	*Schoenoplectus americanus (=Scirpus*
Scirpus sp.	X	X	X	X	*Scirpus sp.*
Schoenoplectus subterminalis (=Scirpus	X	X	X	X	*Schoenoplectus subterminalis (=Scirpus*
Schoenoplectus tabernaemontani	X	X	X		*Schoenoplectus tabernaemontani (=Scirpus*
Scolochloa festucacea		X	X		*Scolochloa festucacoa* [b]
Scutellaria sp.		X			*Scutellaria sp. (includes S. galericulata, S. lateriflora)* [c]
Scutellaria epilobiifolia (=galericulata)		X			
Scutellaria lateriflora		X	X		
Sium suave		X	X		*Sium suave*
Solidago sp. [a]		X	X		
Sparganium angustifolium		X	X	X	*Sparganium sp. (S. angustifolium, S. fluctuans, S minimum)*
Sparganium angustifolium or S. fluctuans	X			X	
Sparganium fluctuans		X		X	
Sparganium sp. (S. angustifolium, S.	X			X	
Sparganium minimum (=S. natans)		X			
Sparganium americanum	X				*Sparganium spp. (S. chlorocarpum, S. eurycarpum)*
Sparganium erectum ssp. stoloniferum		X	X		
Sparganium eurycarpum		X	X	X	
Sparganium spp. (S. chlorocarpum, S.	X	X	X	X	
Spartina pectinata [a]		X	X		
Sphagnum magellanicum		X			*Sphagnum sp. (includes S. magellanicum)* [b]
Sphagnum sp.		X			
Spiraea alba [a]		X			
Spirodela polyrhiza	X	X	X	X	*Spirodela polyrhiza*
Subularia aquatica		X			*Subularia aquatica* [b]
Typha sp.	X	X	X	X	*Typha sp.* [c]
Typha angustifolia		X	X	X	*Typha angustifolia*
Typha glauca		X		X	

Scientific Name	**ECO**	**FISH**	**WJL**	**WPI**	**Lumped to – Scientific Name**
Typha latifolia	X	X	X	X	*Typha latifolia*
Fern Group [a]		X	X	X	
Utricularia cornuta		X		X	*Utricularia cornuta* [b,d]
Utricularia sp.	X	X	X	X	*Utricularia sp.*
Utricularia gibba	X	X			*Utricularia gibba* [b]
Utricularia intermedia	X	X	X	X	*Utricularia intermedia*
Utricularia minor	X	X	X	X	*Utricularia minor*
Utricularia macrorhiza (=U. vulgaris)	X	X	X	X	*Utricularia macrorhiza (=U. vulgaris)*
Vaccinium sp.		X			*Vaccinium sp. (includes V. macrocarpon, V. oxycoccus)* [b,c]
Vaccinium macrocarpon		X	X		
Vaccinium oxycoccos		X			
Vallisneria americana	X	X	X	X	*Vallisneria americana*
Verbena sp. [a]		X			
Veronica americana			X		*Veronica sp. (includes V. americana)* [b,c]
Veronica sp.		X			
Wolffia columbiana		X	X		*Wolffia sp. (includes W. columbiana)* [c]
Wolffia sp.	X	X		X	
Zannichellia palustris	X	X	X	X	*Zannichellia palustrus* [c]
Zizania palustris and Z. aquatica	X	X	X	X	*Zizania palustris and Z. aquatica*

:*a*) removed from analysis because it was a wetland edge species that did not grow in open water;
b) removed from FISH analysis due to rarity;
c) removed from WJL analysis due to rarity;
d) removed from WPI analysis due to rarity.

A spreadsheet with more detailed notations will be delivered with the data package.

ADDITIONAL APPENDICES:

Appendix D. Taxa presence/absence in each lake by dataset. (This is included as a separate file, entitled **Appendix D Taxa Presence Absence In Each Lake By Dataset**)

Appendix E. Composition of each of the 41 plant community types. (This is attached as a separate file entitled: **Appendix G Composition of the 41 Plant Community Types**)

Appendix F.1 Composition of Plant Community Types in the FISH dataset. (This is attached as a separate file entitled: **Appendix F.1 Composition of the Plant Community Types by Dataset_FISH dataset**)

Appendix F.2. Composition of Plant Community Types in the ECO dataset. (This is attached as a separate file entitled: **Appendix F.2 Composition of the Plant Community Types by Dataset_ECO dataset**)

Appendix F.3. Composition of Plant Community Types in the WJL dataset. (This is attached as a separate file entitled: **Appendix F.3 Composition of the Plant Community Types by Dataset_WJL dataset**)

Appendix F.4. Composition of Plant Community Types in the WPI dataset. (This is attached as a separate file entitled**: Appendix F.4 Composition of the Plant Community Types by Dataset_WPI dataset**)

Appendix F.5. Composition of Plant Community Types in the ALLKS dataset. (This is attached as a separate file entitled: **Appendix F.5 Composition of the Plant Community Types by Dataset_ALLKS dataset**)

Appendix F.6: **Composition of Aquatic Plant Community Groups**

Appendix G. **Taxa Richness by Lake and Dataset**.

Appendix H.1: **ECO Ordination Graphs**

Appendix H.2: **FISH Ordination Graphs**

Appendix H.3: **WJL Ordination Graphs**

Appendix H.4: **WPI Ordination Graphs**

Appendix H.5: **ALLKS Ordination Graphs and PCTYPES Ordination Graph**

Appendix **Map Plates**

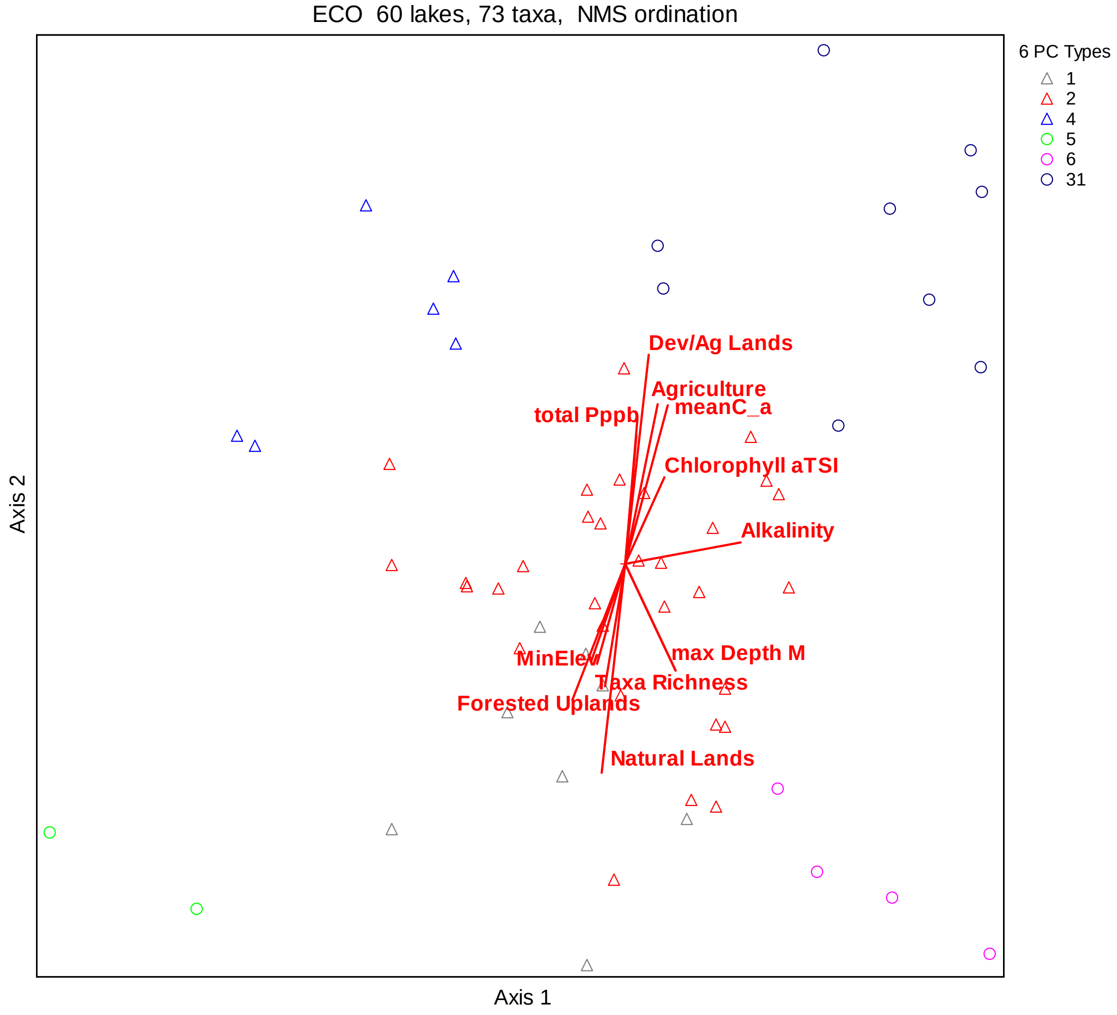


Figure ECO-1: NMS ordination of ECO dataset; joint plot showing lines of strongest correlation with quantitative environmental variables. The strongest correlations with the lower end of Axis 2 are with proportion of **Natural Lands** and **Forested Uplands** in the 1 km buffer around each lake, **Taxa Richness**, **Minimum Elevation**, and **Maximum Depth** of each lake. The strongest correlations with the upper end of Axis 2 are with proportion of **Agricultural Lands** and **Developed/Ag Lands** in the buffer, **Mean Chlorophyll a**, **Total Phosphorus**, and **Trophic State Index based on Chlorophyll a**. The strongest correlations with the right end of Axis 1 are with lake water **Alkalinity**. Symbol colors and shapes represent the 6 PC types recognized in analysis of the ECO dataset.

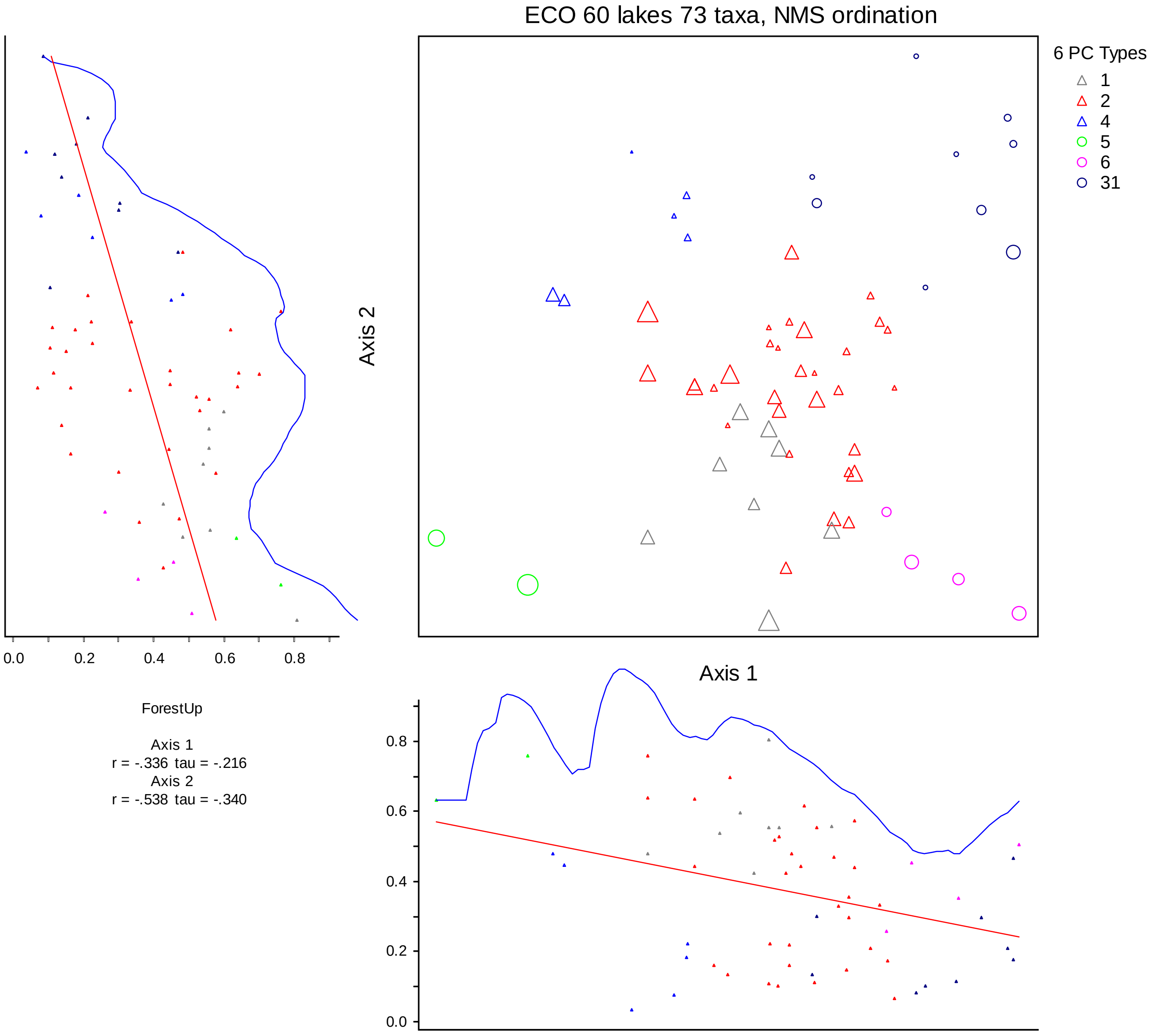


Figure ECO-2: NMS ordination of ECO dataset, overlay of proportion of 1 km buffer around each lake that is **Forested Uplands**. Larger symbols represent lakes in landscapes with higher proportions of forest in the buffer. Symbol colors and shapes represent the 6 PC types recognized in analysis of the ECO dataset.

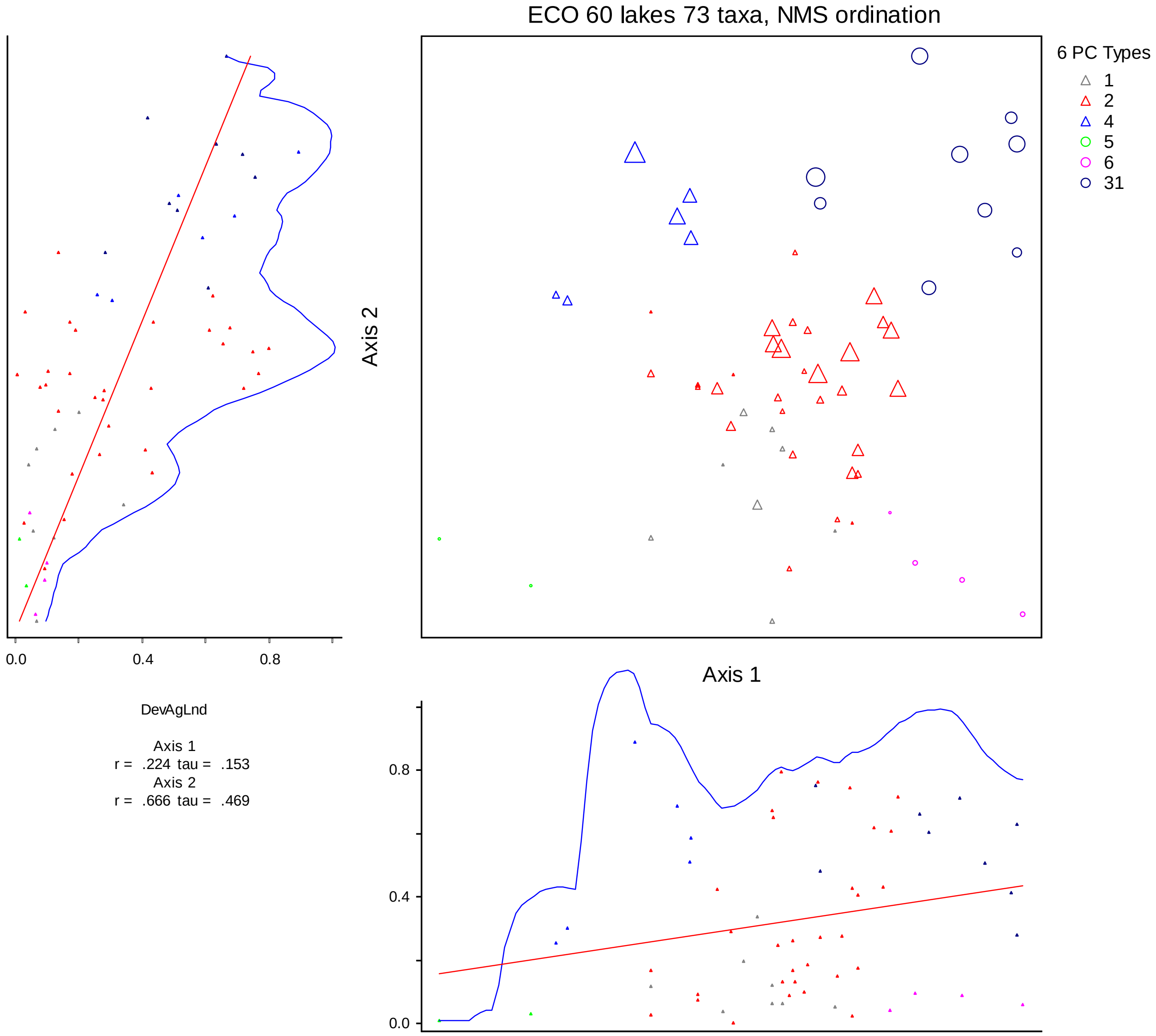


Figure ECO-3: NMS ordination of ECO dataset, overlay of proportion of 1 km buffer around each lake that is **Developed and Agriculture Lands**. Larger symbols represent lakes in landscapes with higher proportions of developed/ag lands in the buffer. Symbol colors and shapes represent the 6 PC types recognized in analysis of the ECO dataset.

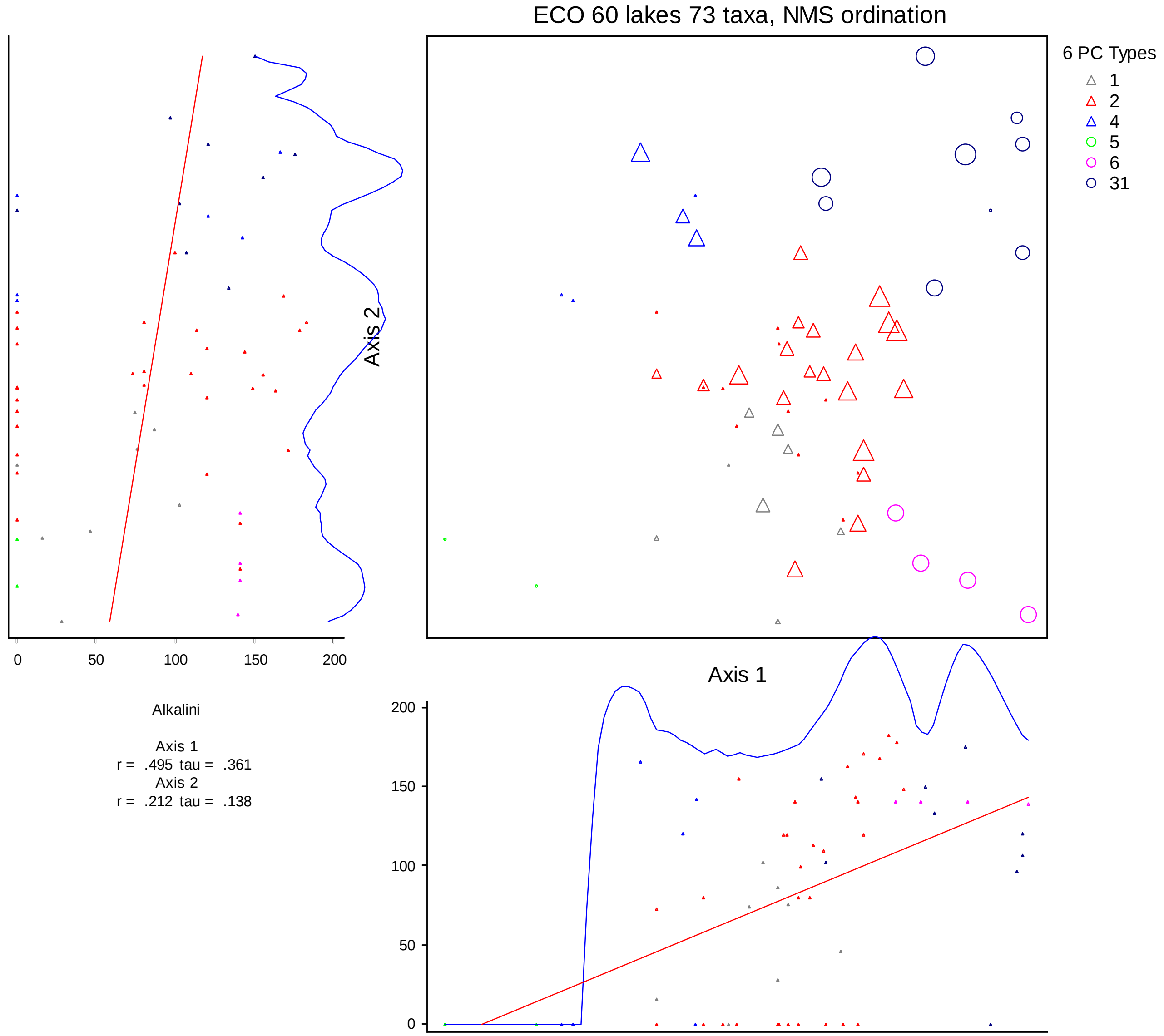


Figure ECO-4: NMS ordination of ECO dataset, overlay of **Alkalinity**. Larger symbols represent lakes with higher water alkalinity. Symbol colors and shapes represent the 6 PC types recognized in analysis of the ECO dataset.

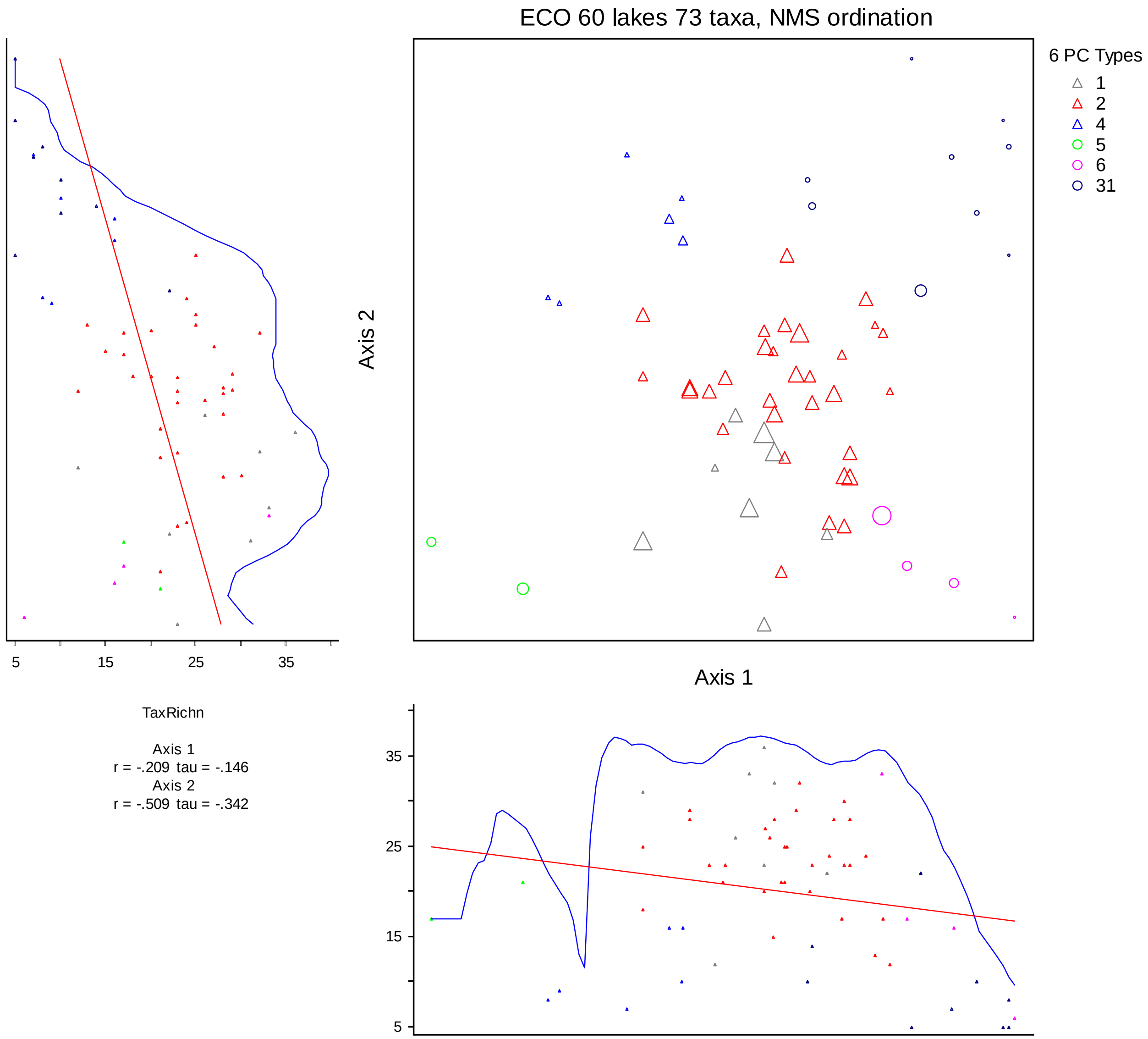


Figure ECO-5: NMS ordination of ECO dataset, overlay of **Taxa Richness**. Larger symbols represent lakes with greater numbers of taxa recorded for each lake. Symbol colors and shapes represent the 6 PC types recognized in analysis of the ECO dataset.

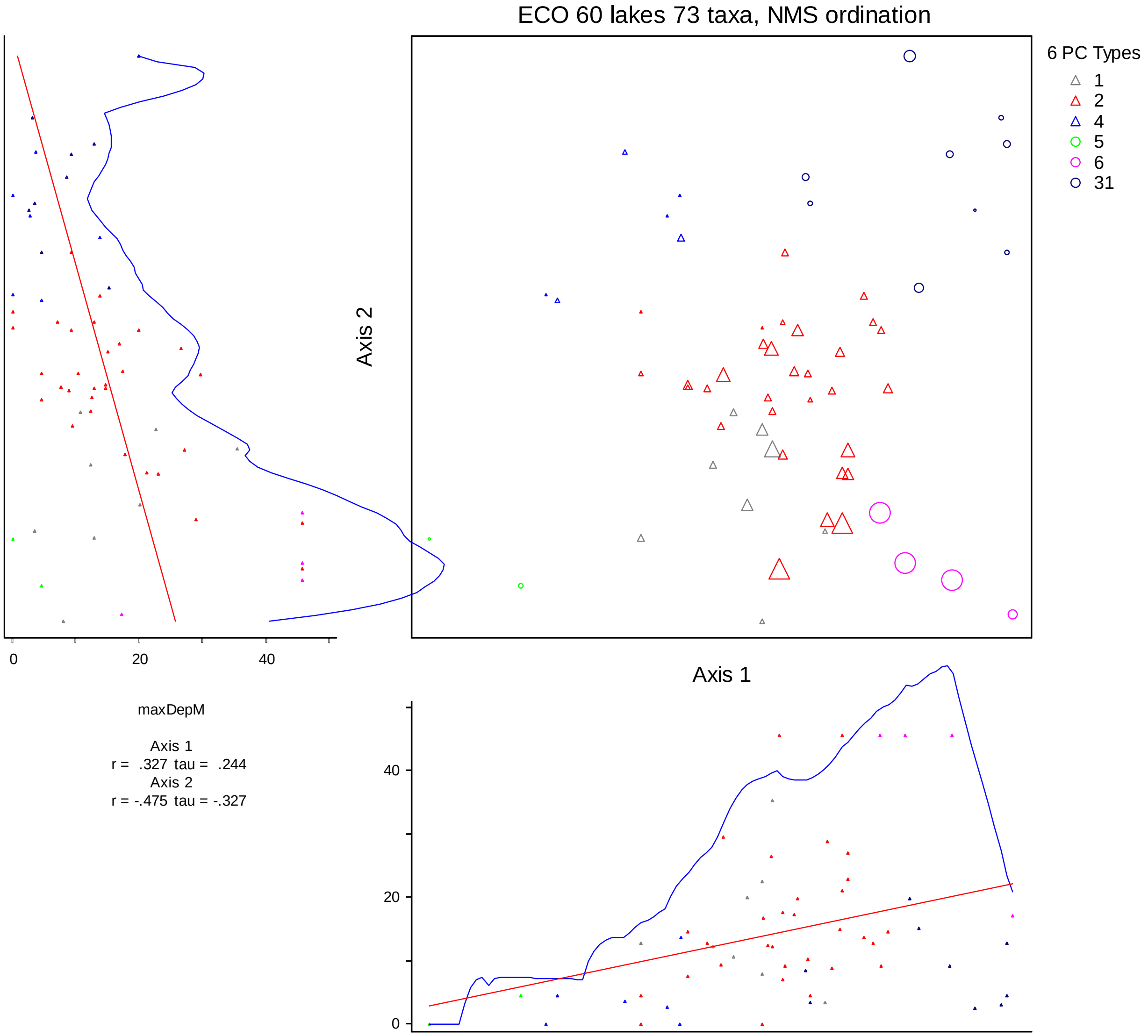


Figure ECO-6: NMS ordination of ECO dataset, overlay of **Maximum Depth in meters**. Larger symbols represent lakes with greater maximum depths recorded in the lake (unvegetated areas included). Symbol colors and shapes represent the 6 PC types recognized in analysis of the ECO dataset.

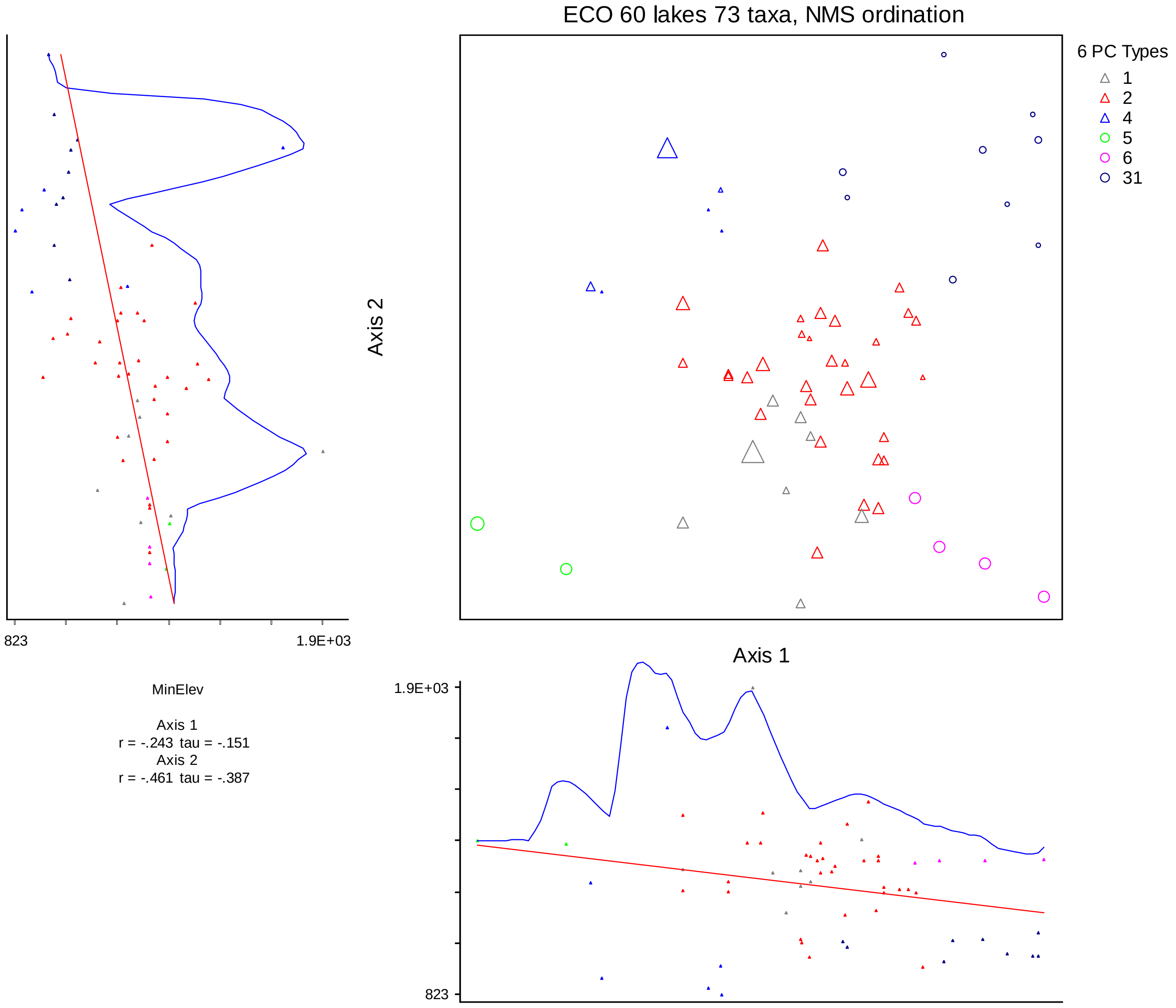


Figure ECO-7: NMS ordination of ECO dataset, overlay of **Minimum Elevation** of 1 km buffer around each lake (approximately equal to lake elevation). Larger symbols represent lakes with higher minimum elevations recorded for each lake's buffer. Symbol colors and shapes represent the 6 PC types recognized in analysis of the ECO dataset.

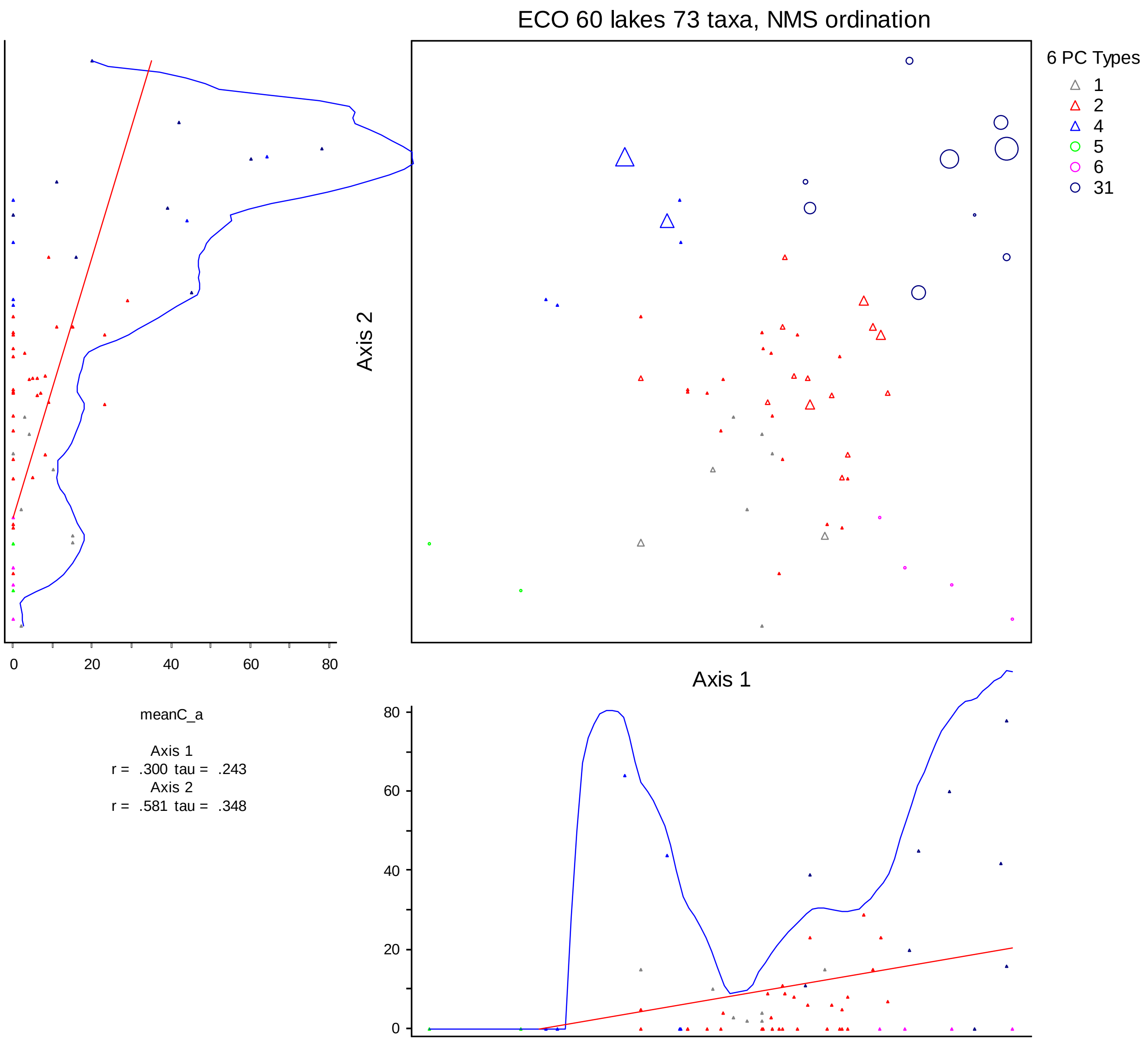


Figure ECO-8: NMS ordination of ECO dataset, overlay of **Mean Chlorophyll a**. Larger symbols represent lakes with higher mean Chlorophyll a levels. Symbol colors and shapes represent the 6 PC types recognized in analysis of the ECO dataset.

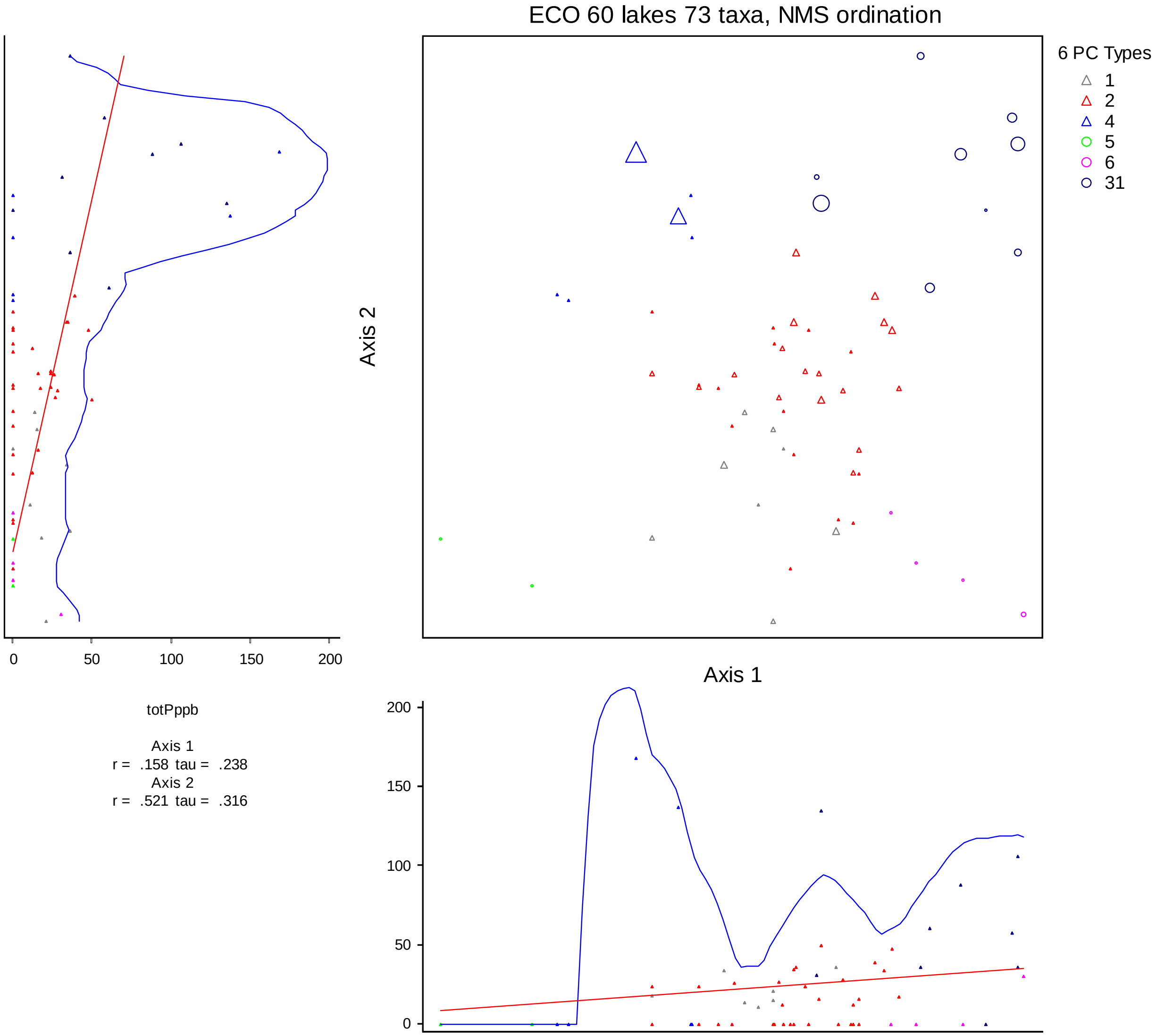


Figure ECO-9: NMS ordination of ECO dataset, overlay of **Total Phosphorus (ppb)**. Larger symbols represent lakes with higher total Phosphorus levels. Symbol colors and shapes represent the 6 PC types recognized in analysis of the ECO dataset.

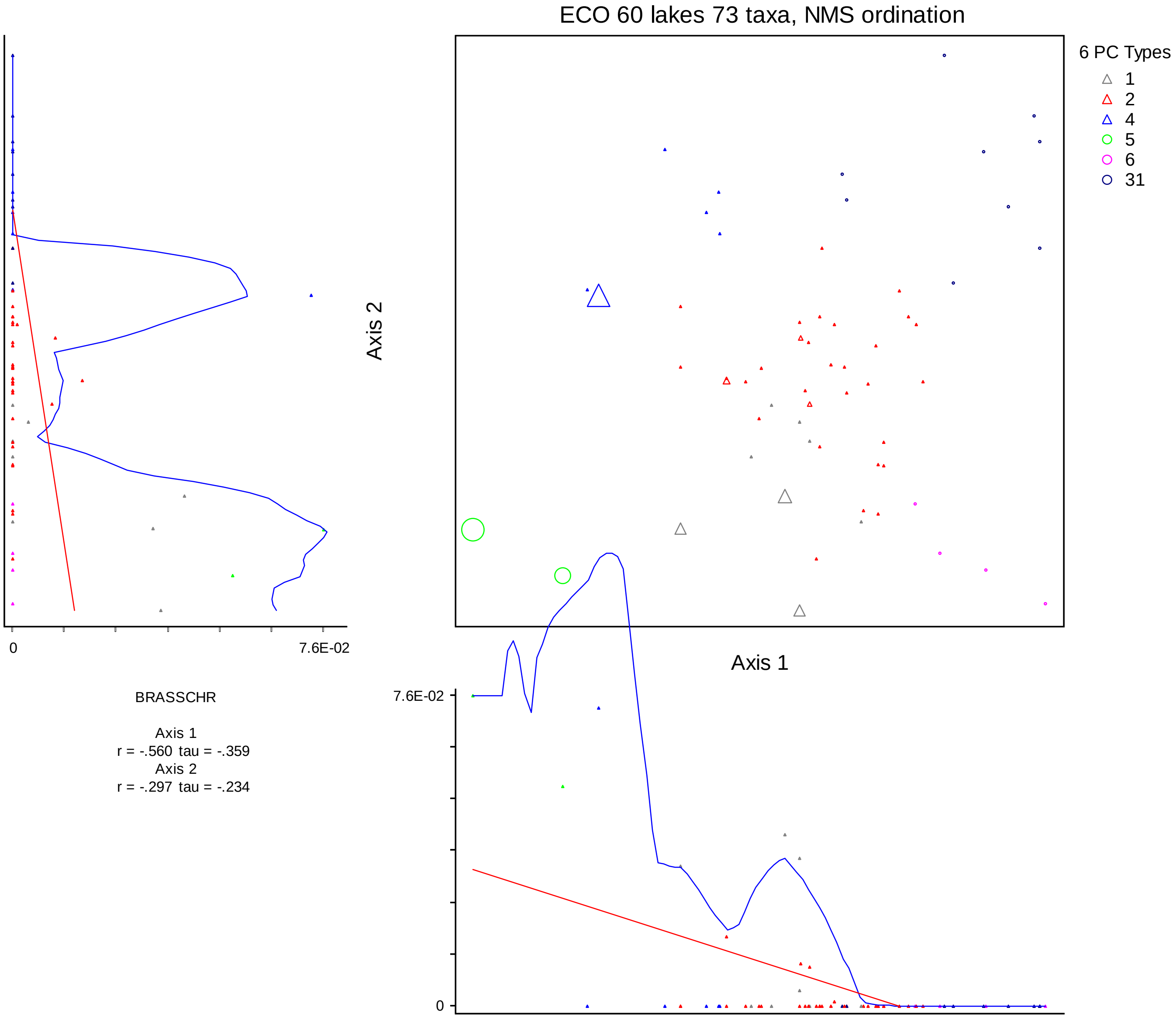


Figure ECO-10: NMS ordination of ECO dataset, overlay of favorability of lakes for **water shield** (*Brasenia schreberi*). Large symbols represent lakes with higher relative frequency values for the taxon. Symbol colors and shapes represent the 6 PC types recognized in analysis of the ECO dataset.

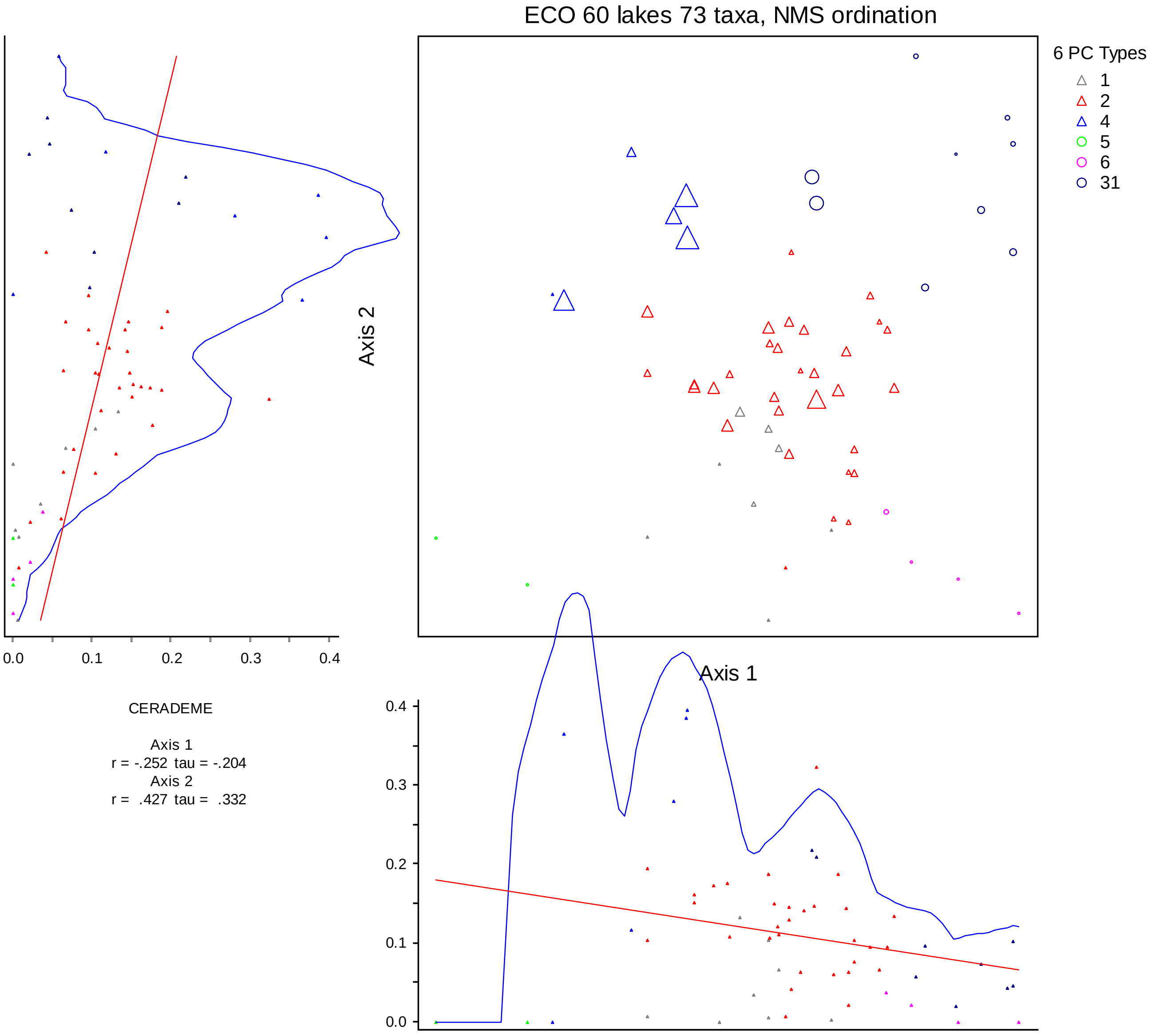


Figure ECO-11: NMS ordination of ECO dataset, overlay of favorability of lakes for **coontail** (*Ceratophyllum demersum*). Large symbols represent lakes with higher relative frequency values for the taxon. Symbol colors and shapes represent the 6 PC types recognized in analysis of the ECO dataset.

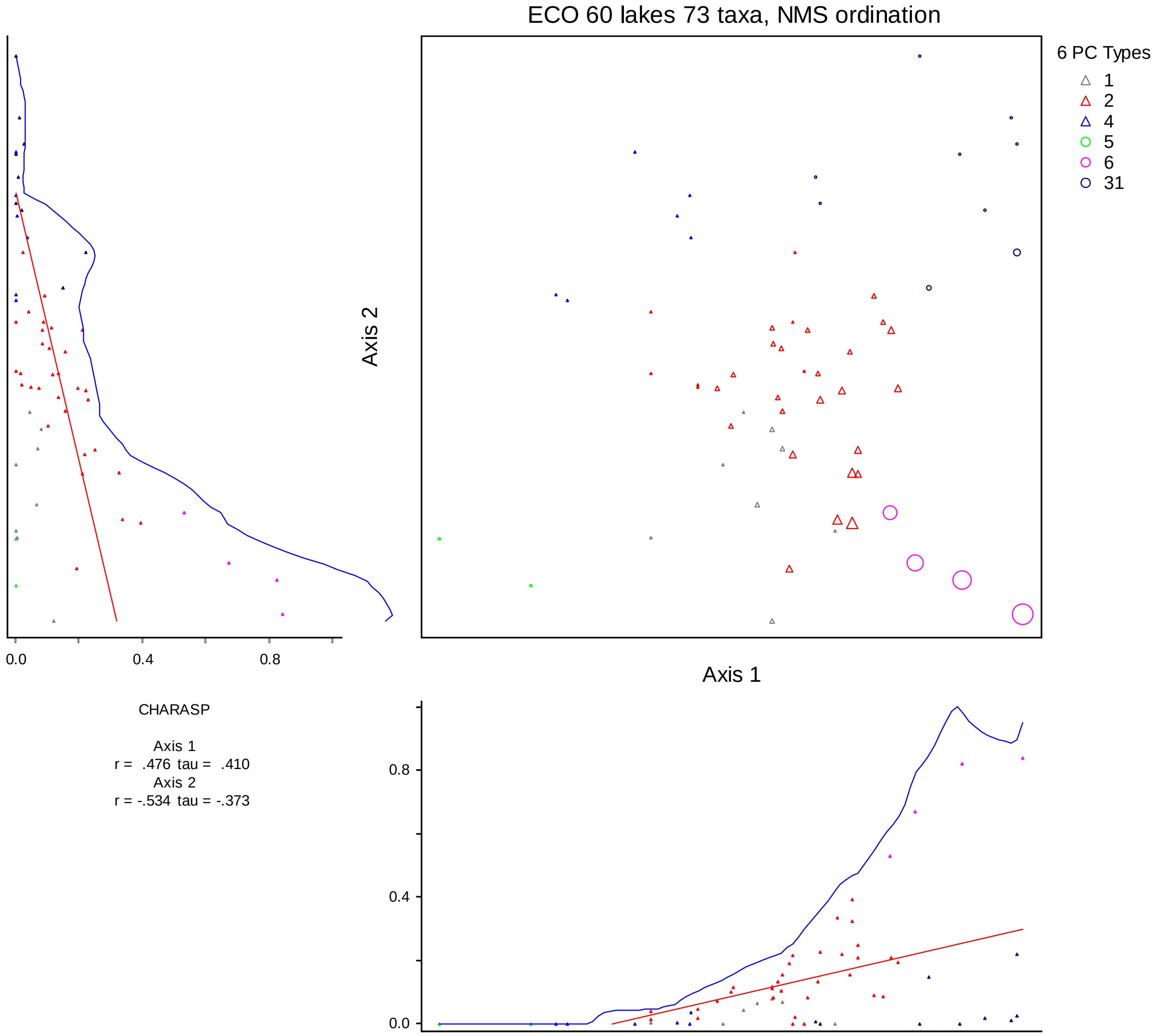


Figure ECO-12: NMS ordination of ECO dataset, overlay of favorability of lakes for **muskgrasses** (*Chara* spp.). Large symbols represent lakes with higher relative frequency values for the taxon. Symbol colors and shapes represent the 6 PC types recognized in analysis of the ECO dataset.

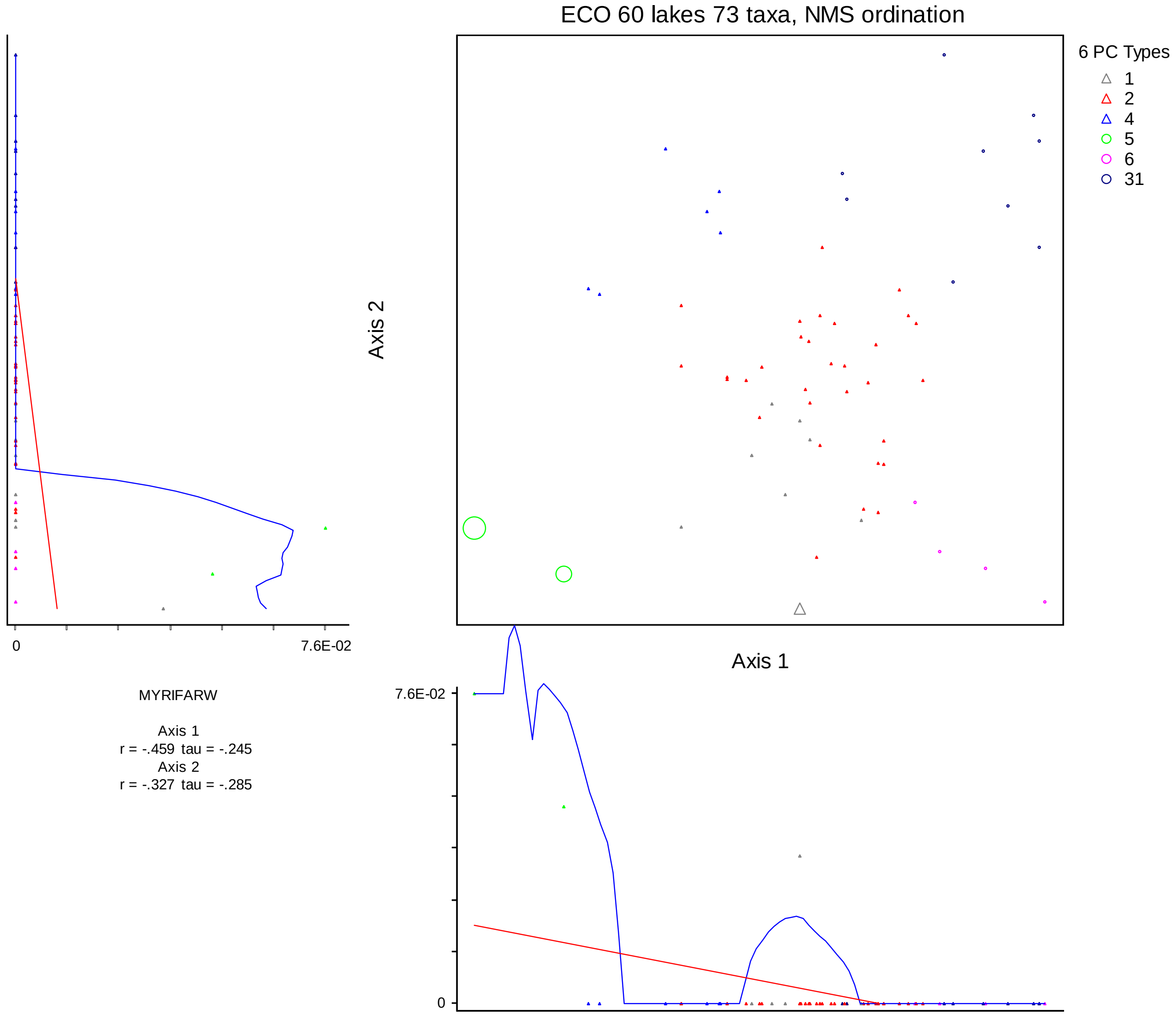


Figure ECO-13: NMS ordination of ECO dataset, overlay of favorability of lakes for **small-leaved milfoil** (*Myriophyllum farwellii*). Large symbols represent lakes with higher relative frequency values for the taxon. Symbol colors and shapes represent the 6 PC types recognized in analysis of the ECO dataset.

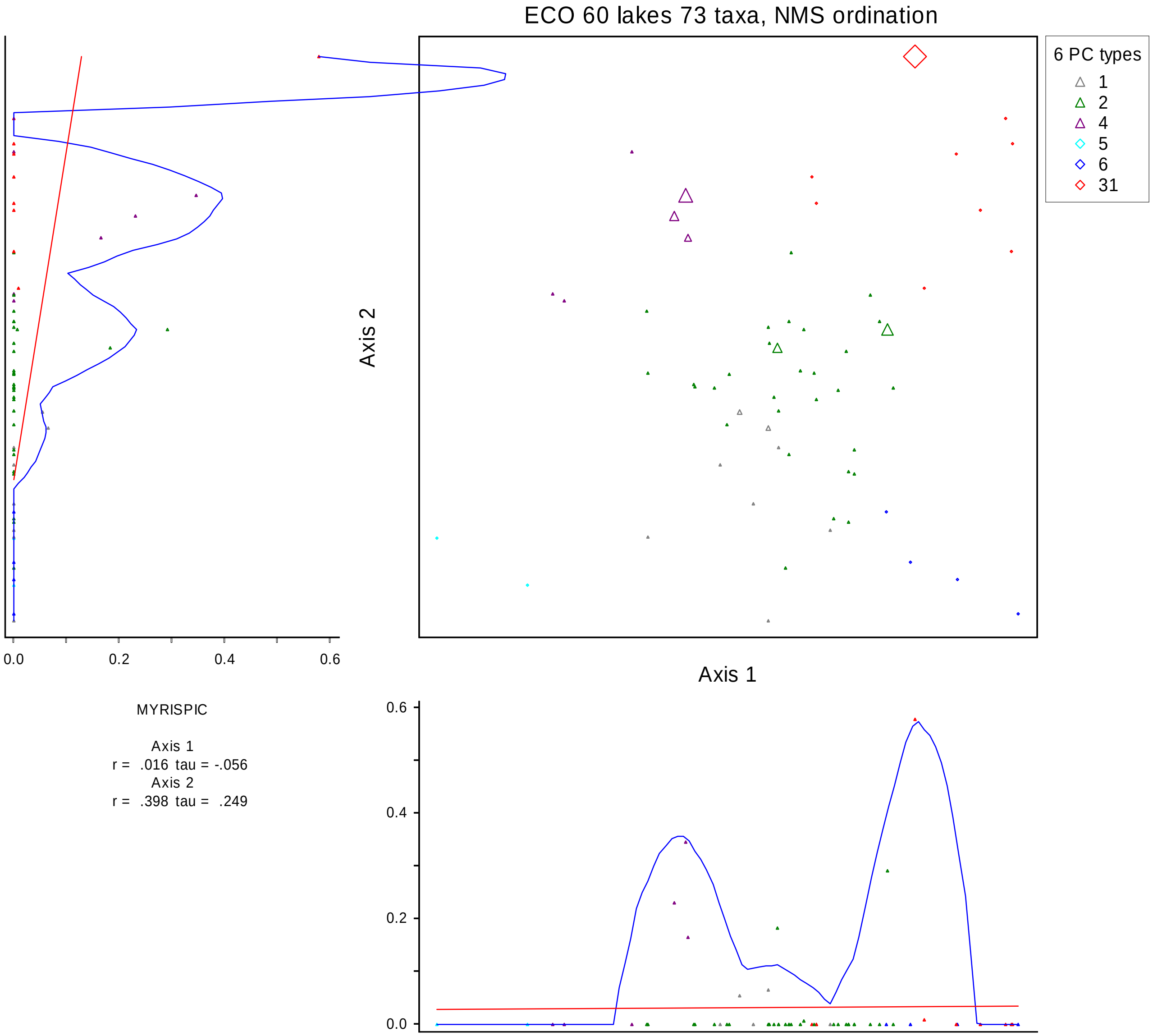


Figure ECO-14: NMS ordination of ECO dataset, overlay of favorability of lakes for exotic **Eurasian milfoil** (*Myriophyllum spicatum*). Large symbols represent lakes with higher relative frequency values for the taxon. Symbol colors and shapes represent the 6 PC types recognized in analysis of the ECO dataset.

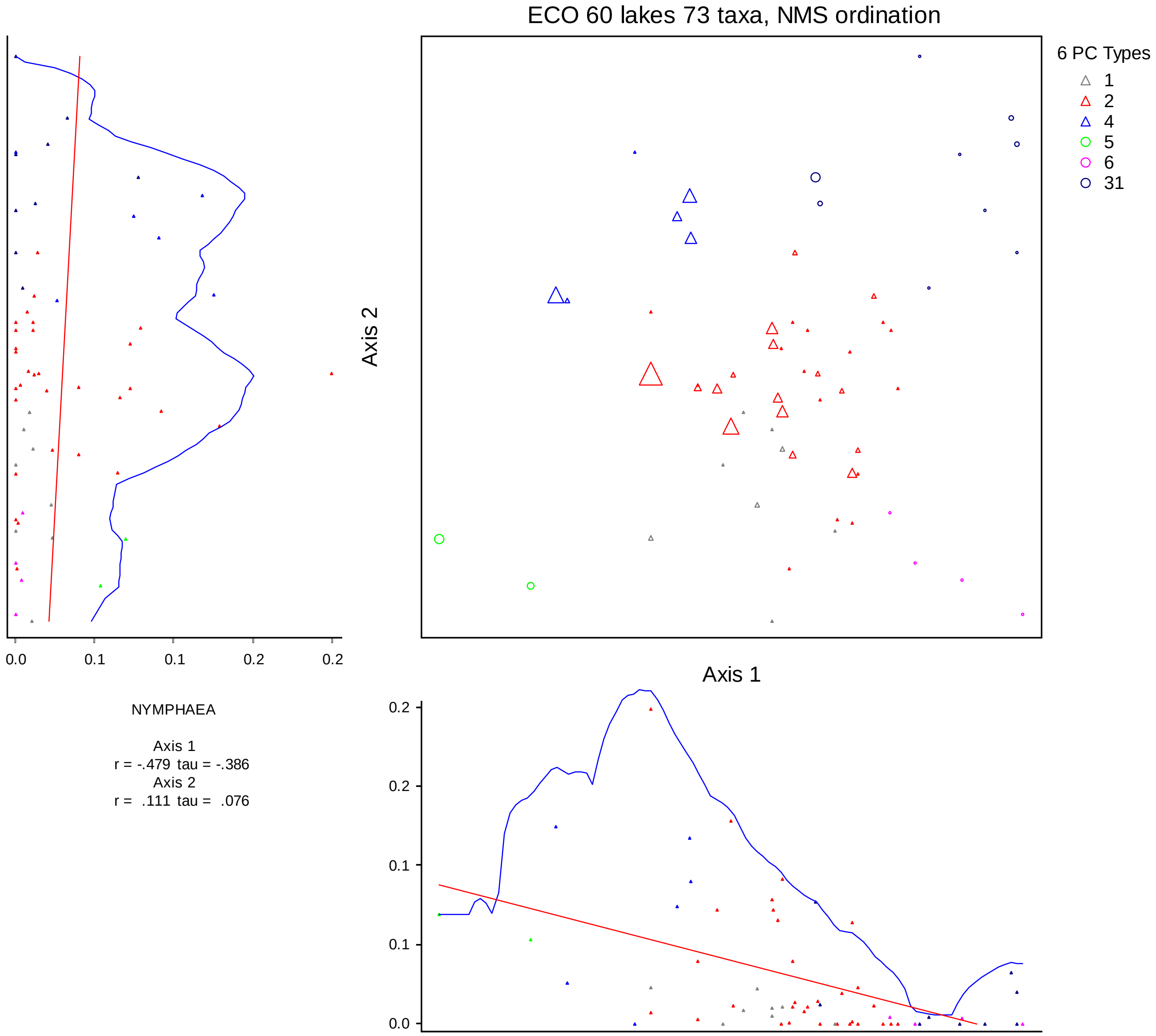


Figure ECO-15: NMS ordination of ECO dataset, overlay of favorability of lakes for **white water lilies** (*Nymphaea* spp.). Large symbols represent lakes with higher relative frequency values for the taxon. Symbol colors and shapes represent the 6 PC types recognized in analysis of the ECO dataset.

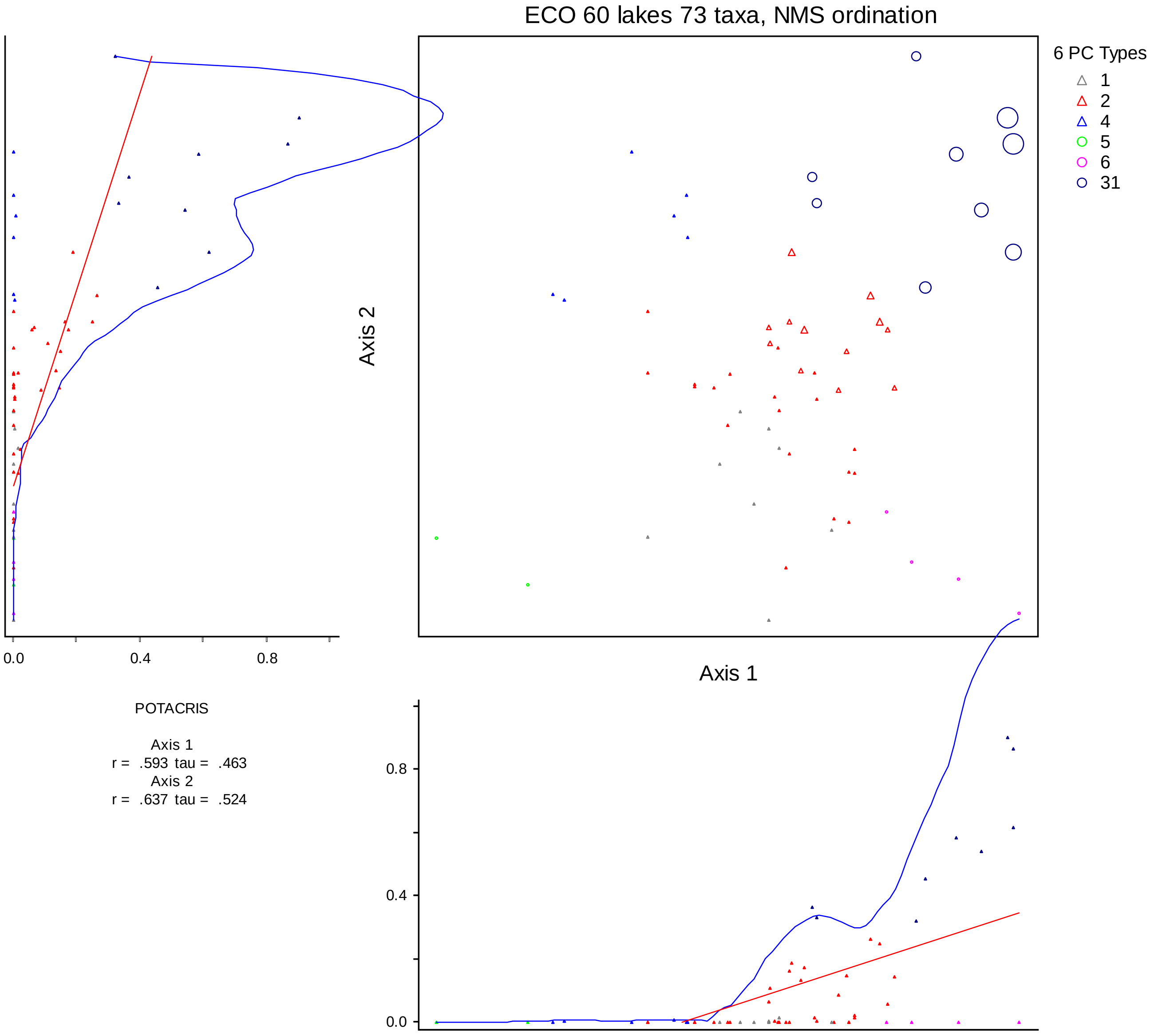


Figure ECO-16: NMS ordination of ECO dataset, overlay of favorability of lakes for exotic **curly-leaf pondweed** (*Potamogeton crispus*). Large symbols represent lakes with higher relative frequency values for the taxon. Symbol colors and shapes represent the 6 PC types recognized in analysis of the ECO dataset.

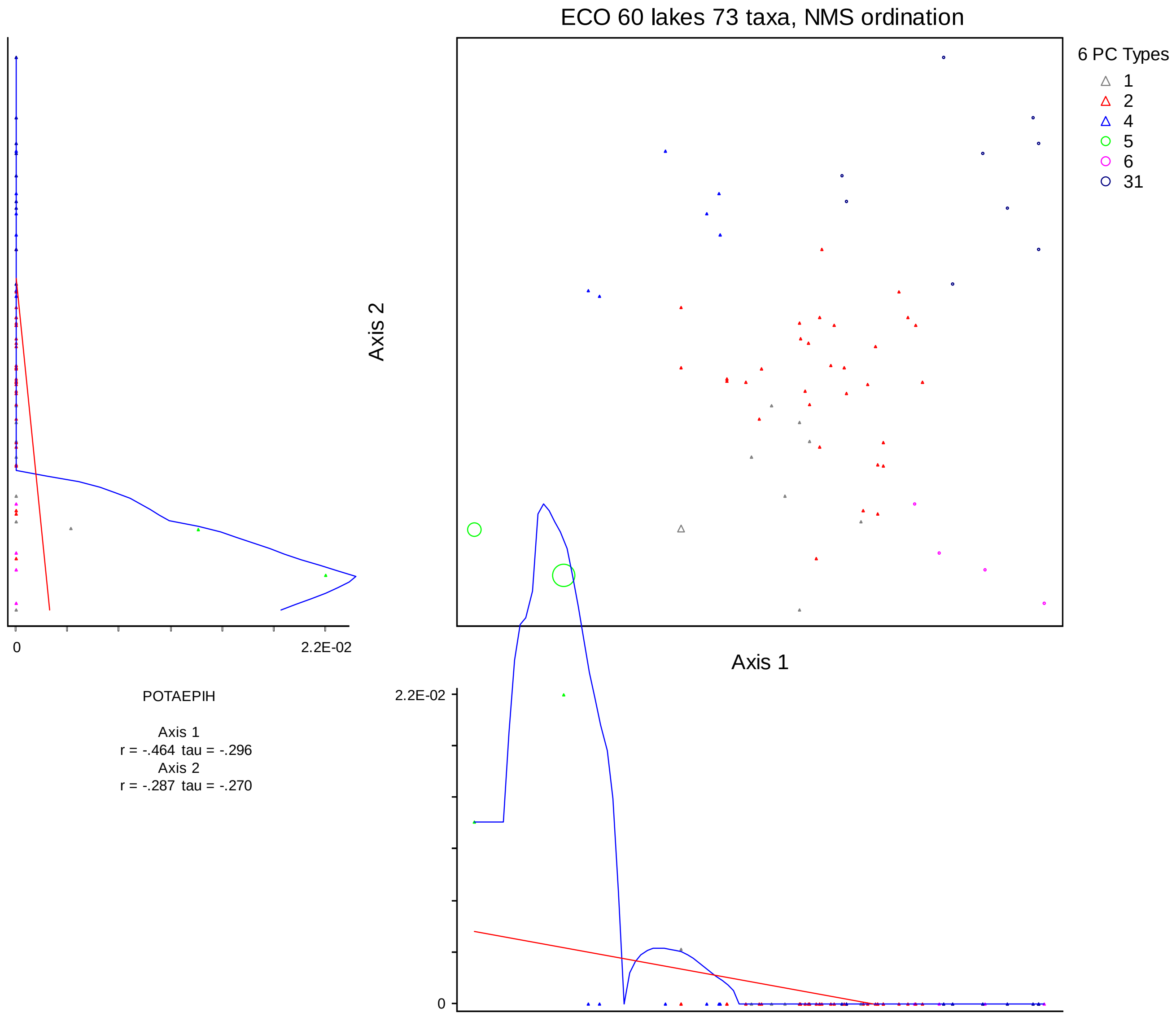


Figure ECO-17: NMS ordination of ECO dataset, overlay of favorability of lakes for **Nuttall's (ribbon-leaf) pondweed** (*Potamogeton epihydrus*). Large symbols represent lakes with higher relative frequency values for the taxon. Symbol colors and shapes represent the 6 PC types recognized in analysis of the ECO dataset.

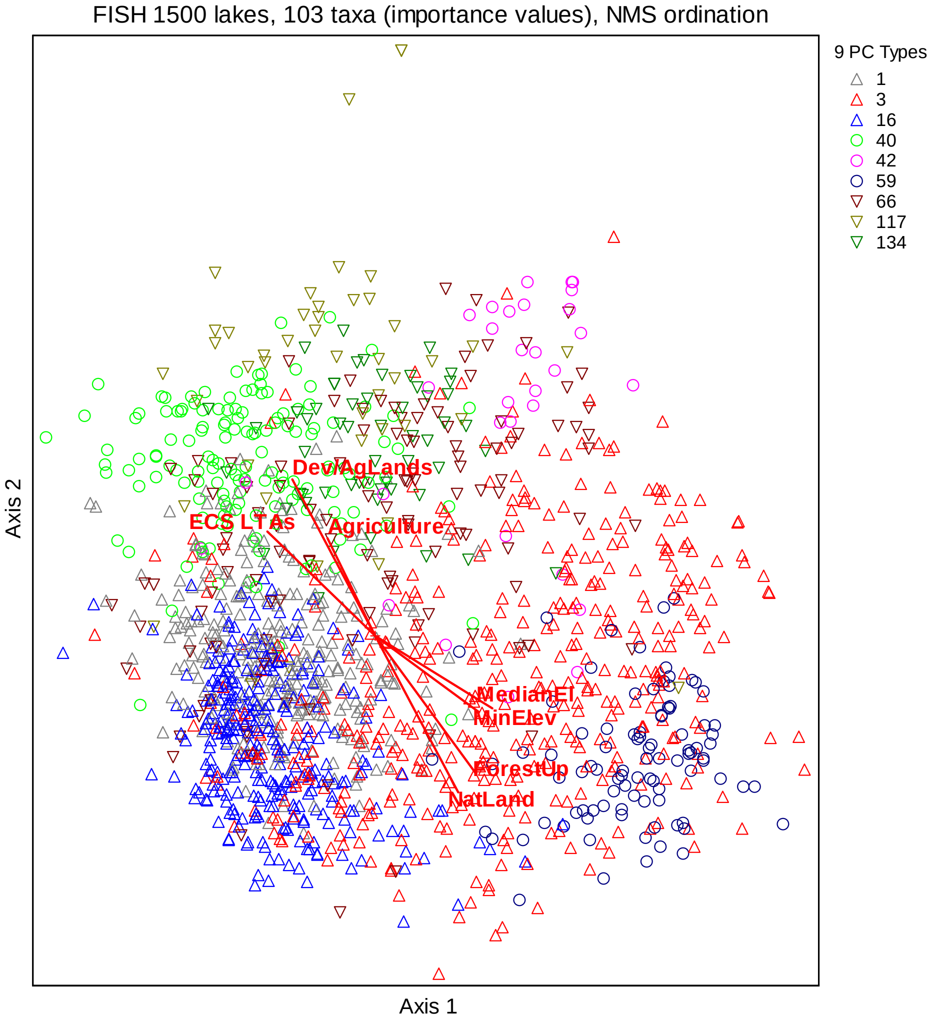


Figure FISH-1: NMS ordination of FISH dataset; joint plot showing lines of strongest correlation with quantitative environmental variables. The strongest correlations with the lower end of Axis 2 are with proportion of **Natural Lands** and **Forested Uplands** in the 1 km buffer around each lake, **Minimum Elevation**, and **Median Elevation** of the buffer around each lake (the Minimum Elevation of the buffer is approximately equal to the elevation of the lake). The strongest correlations with the upper end of Axis 2 are with proportion of **Agricultural Lands**, **Developed/Ag Lands, and ECS Land Type Associations (LTAs)** in the buffer. Symbol colors and shapes represent the 9 PC types recognized in analysis of the FISH dataset.

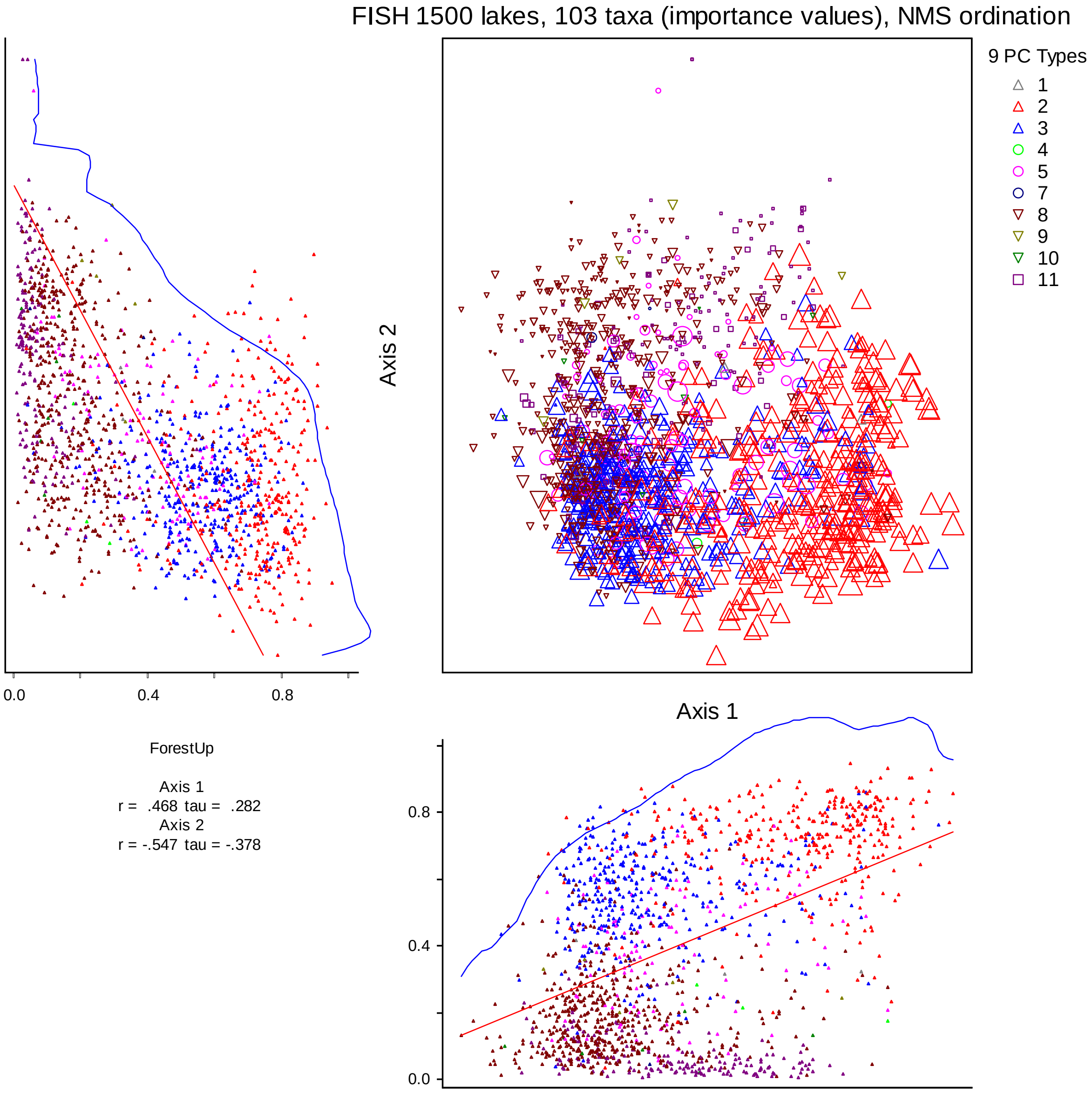


Figure FISH-2: NMS ordination of FISH dataset, overlay of proportion of 1 km buffer around each lake that is **Forested Uplands**. Larger symbols represent lakes in landscapes with higher proportions of forest in the buffer. Symbol colors and shapes represent the 9 PC types recognized in analysis of the FISH dataset.

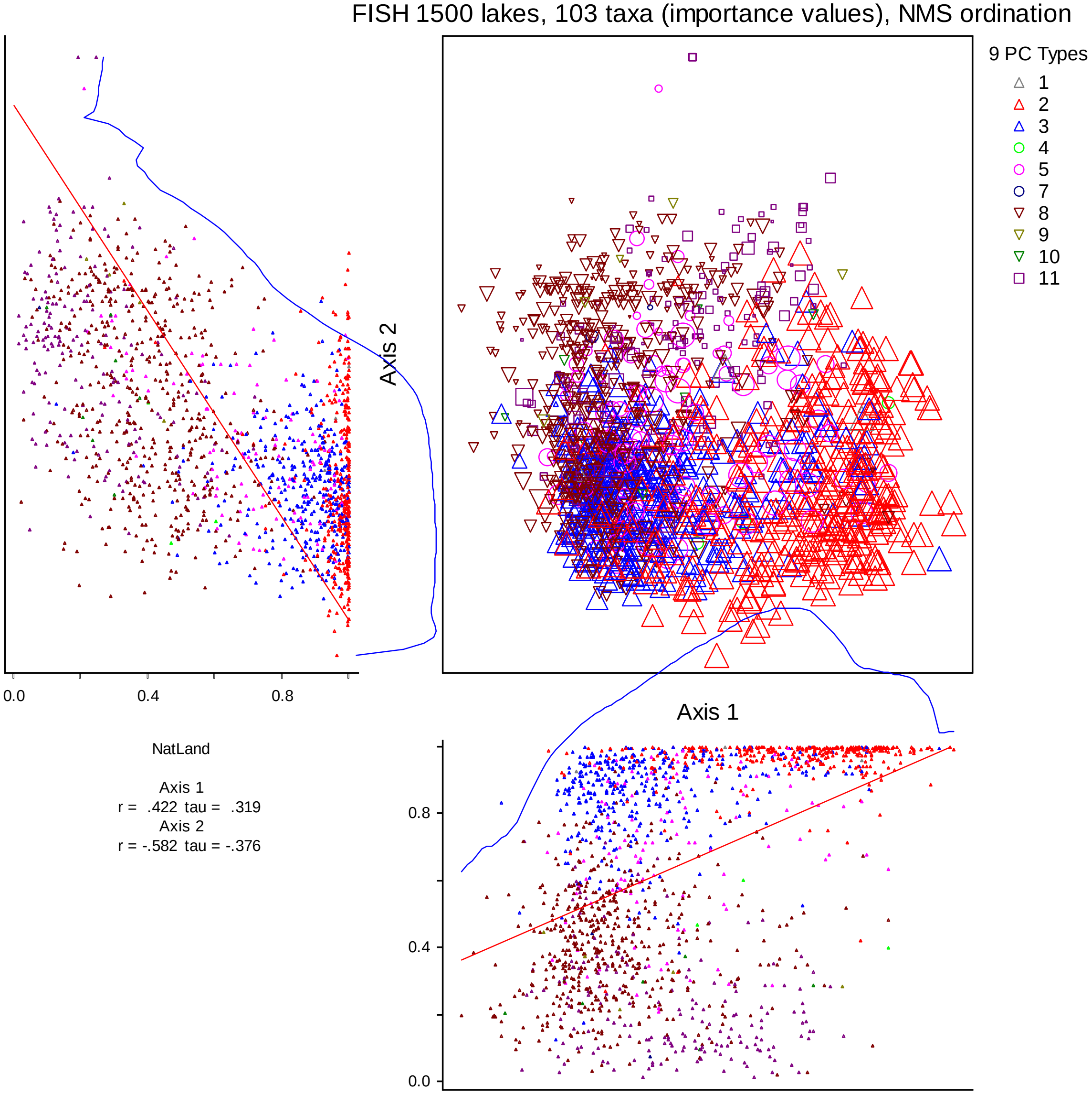


Figure FISH-3: NMS ordination of FISH dataset, overlay of proportion of 1 km buffer around each lake that is **Natural Lands**. Larger symbols represent lakes in landscapes with higher proportions of forest in the buffer. Symbol colors and shapes represent the 9 PC types recognized in analysis of the FISH dataset.

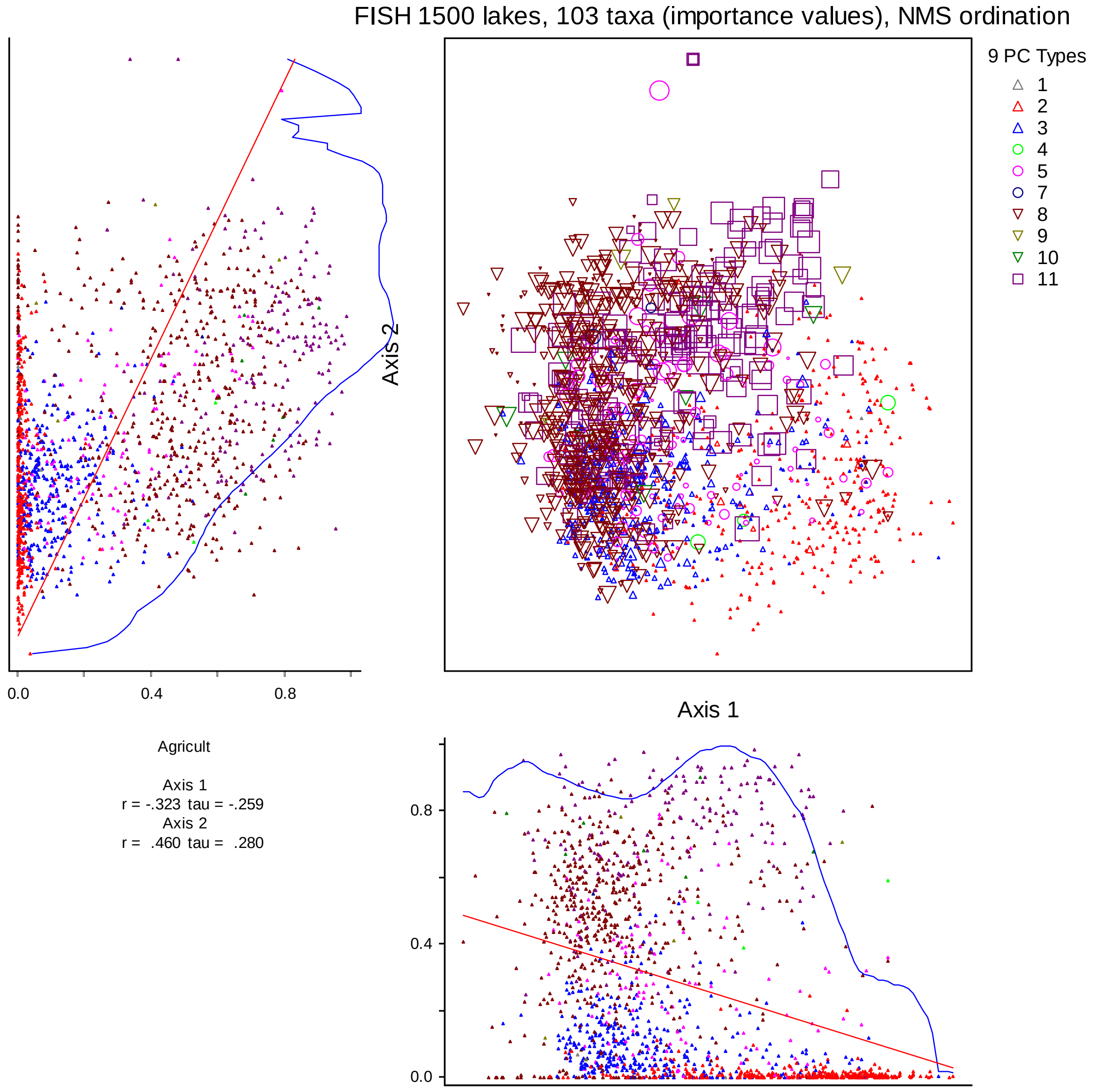


Figure FISH-4: NMS ordination of FISH dataset, overlay of proportion of 1 km buffer around each lake that is **Agriculture Lands**. Larger symbols represent lakes in landscapes with higher proportions of developed/ag lands in the buffer. Symbol colors and shapes represent the 9 PC types recognized in analysis of the FISH dataset.

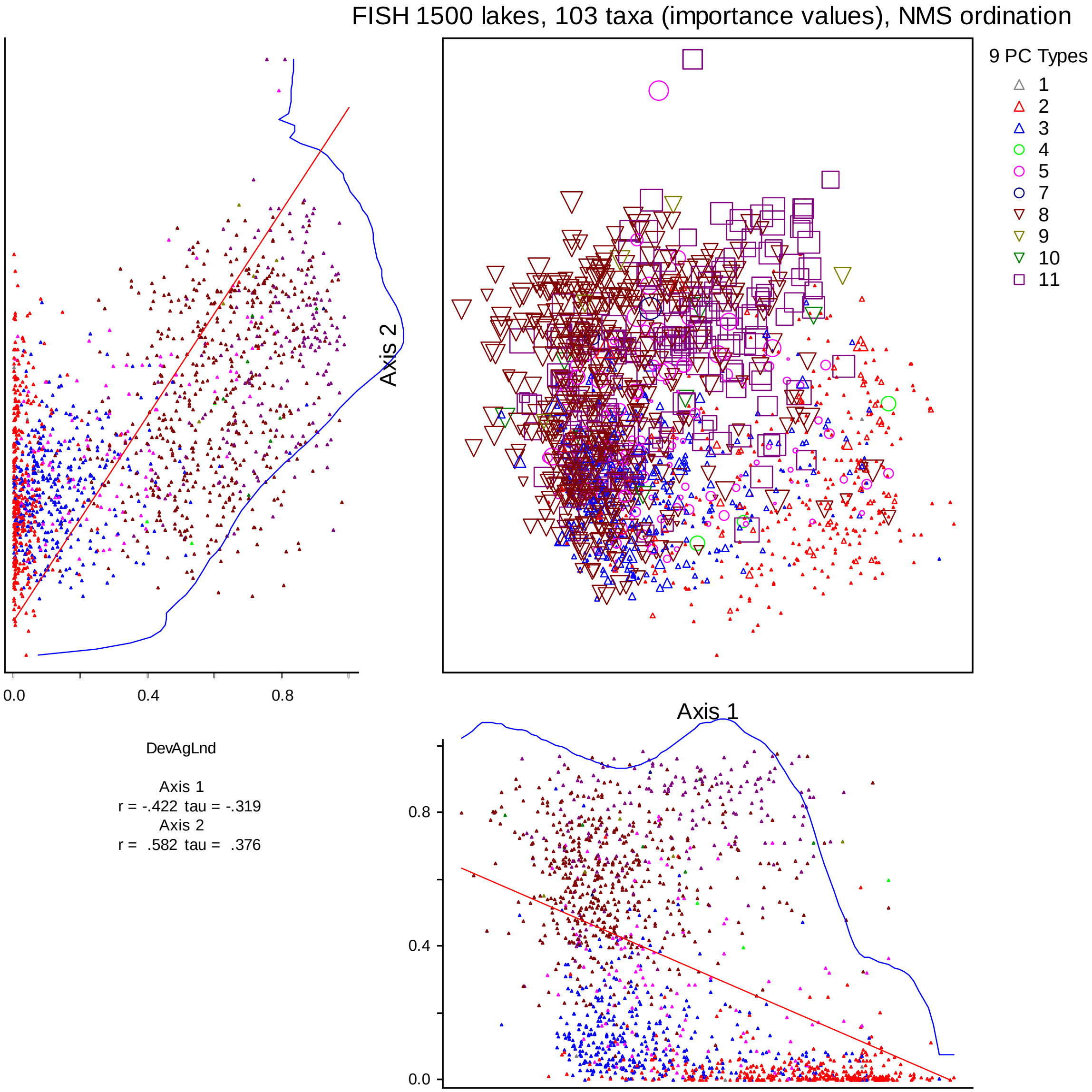


Figure FISH-5: NMS ordination of FISH dataset, overlay of proportion of 1 km buffer around each lake that is **Developed and Agriculture Lands**. Larger symbols represent lakes in landscapes with higher proportions of developed/ag lands in the buffer. Symbol colors and shapes represent the 9 PC types recognized in analysis of the FISH dataset.

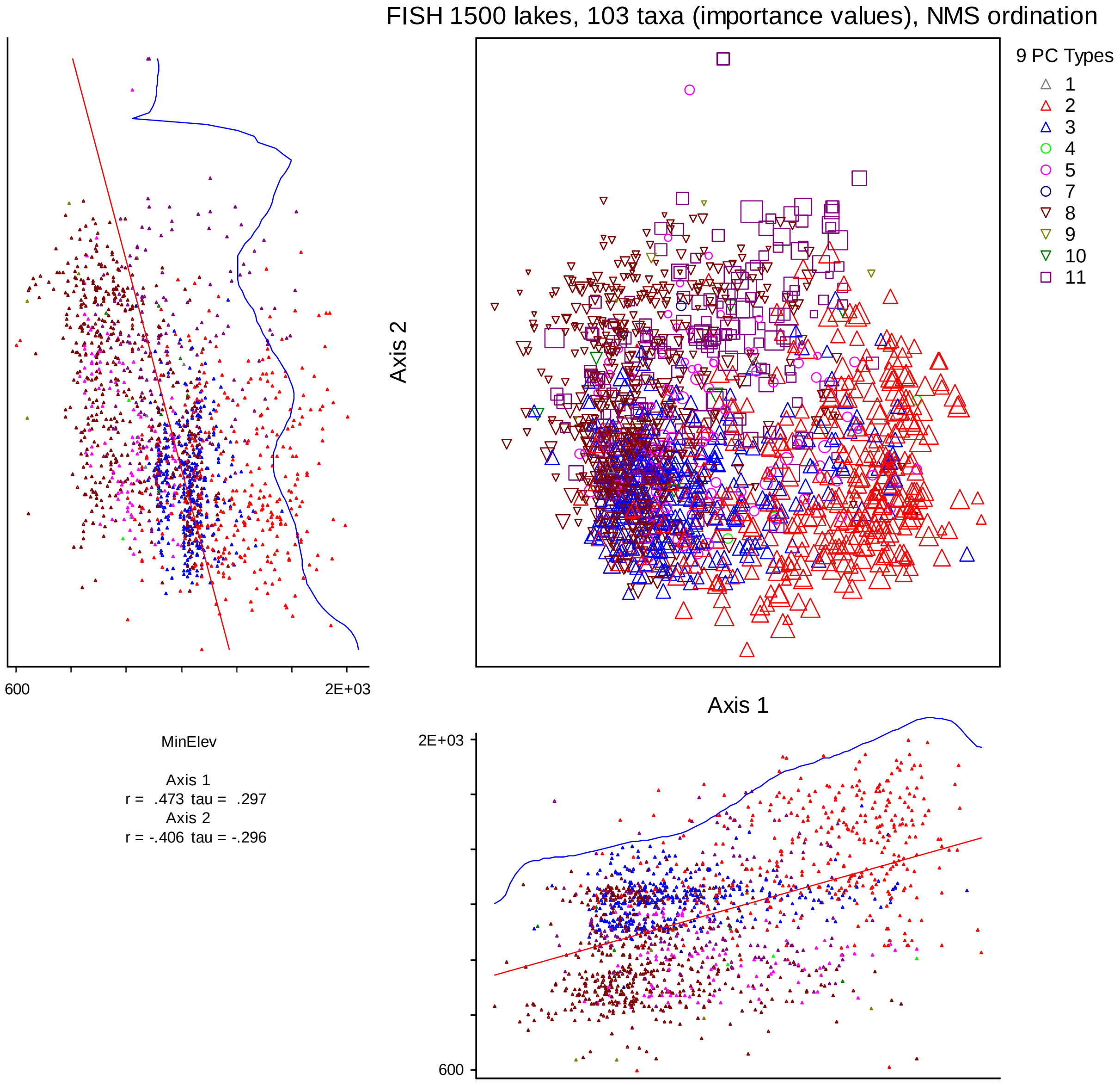


Figure FISH-6: NMS ordination of FISH dataset, overlay of **Minimum Elevation** of 1 km buffer around each lake (approximately equal to lake elevation). Larger symbols represent lakes with higher minimum elevations recorded for each lake's buffer. Symbol colors and shapes represent the 9 PC types recognized in analysis of the FISH dataset.

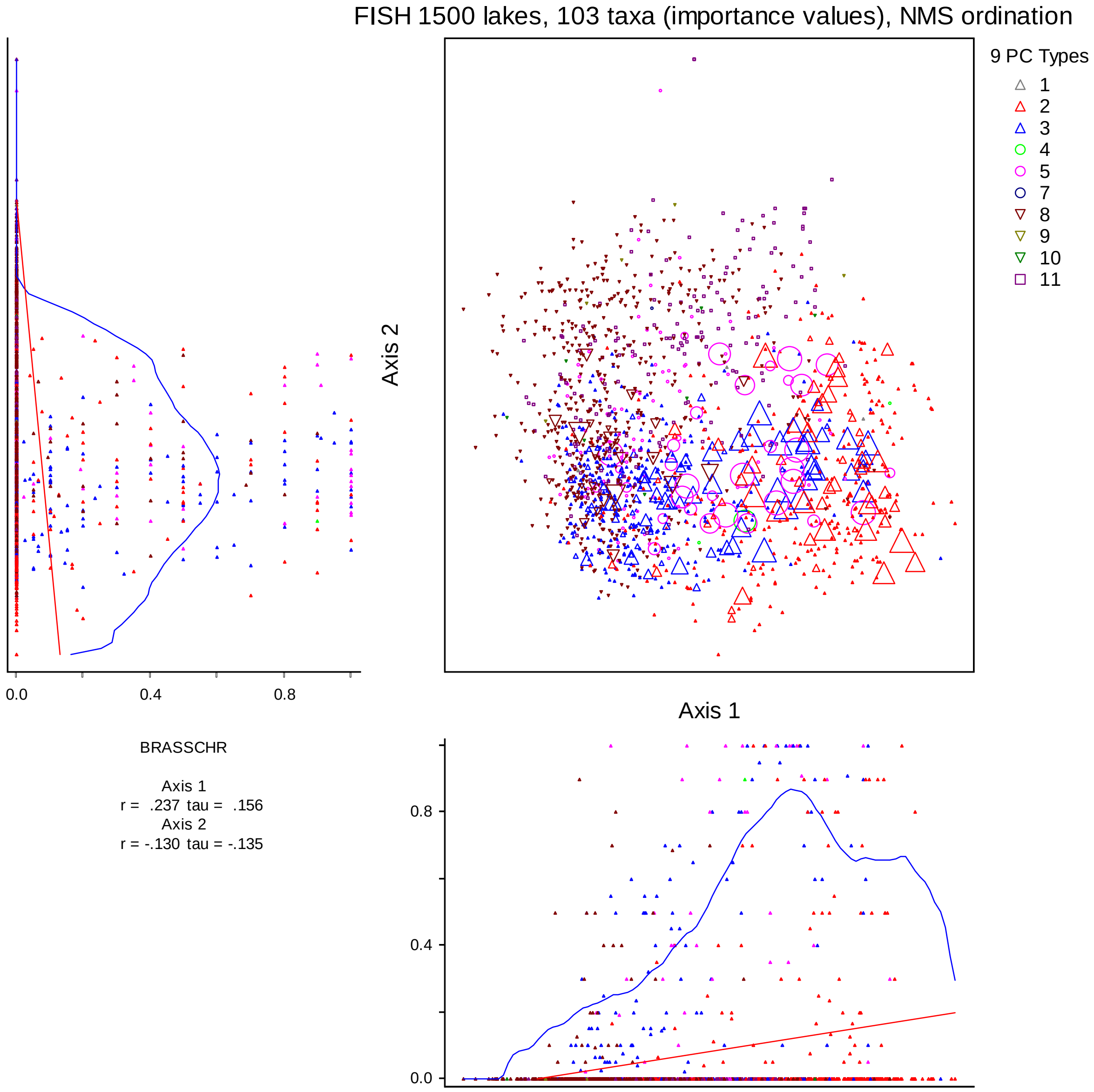


Figure FISH-7: NMS ordination of FISH dataset, overlay of favorability of lakes for **water shield** (*Brasenia schreberi*). Large symbols represent lakes with higher importance values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the FISH dataset.

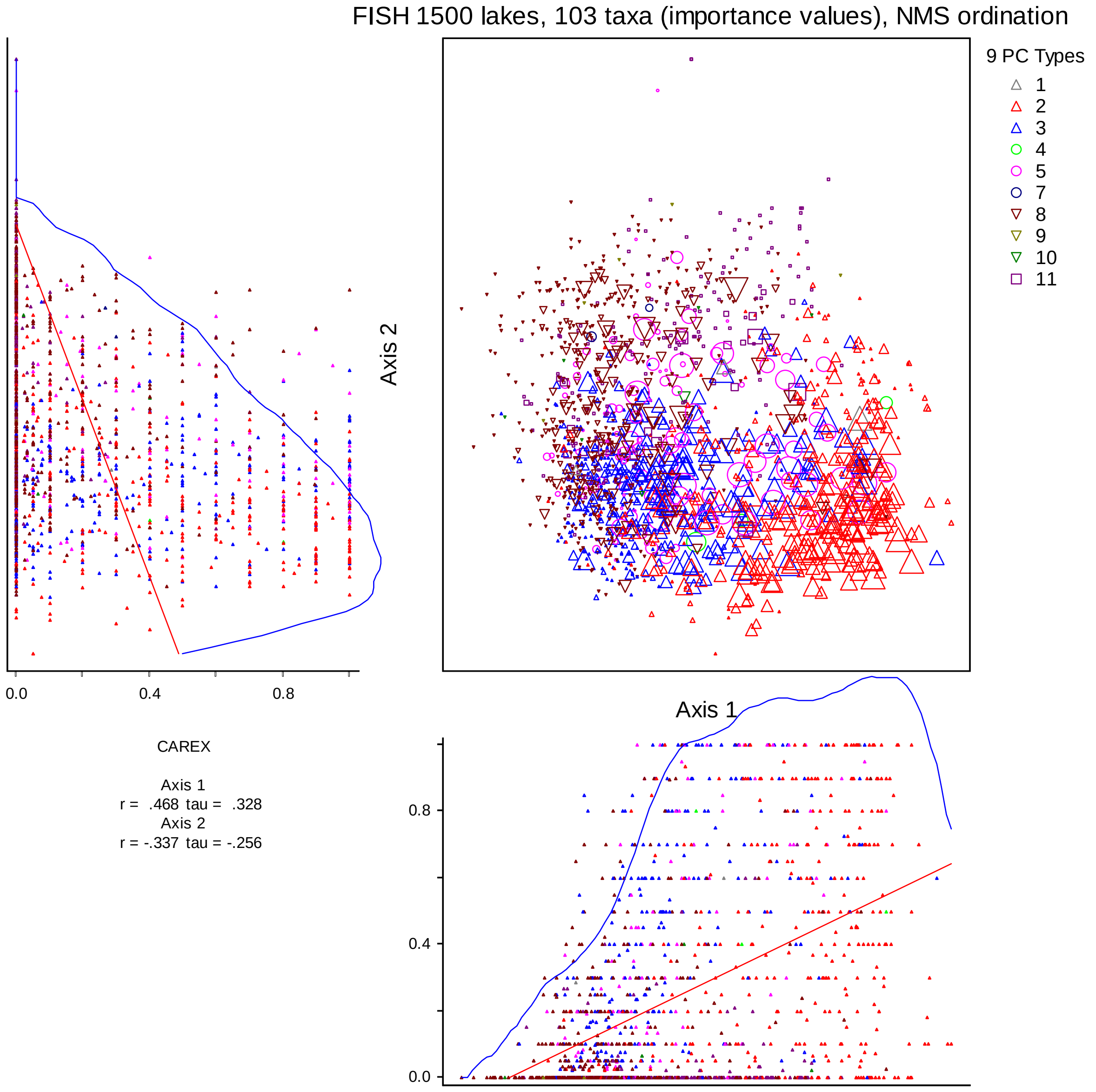


Figure FISH-8: NMS ordination of FISH dataset, overlay of favorability of lakes for **sedges** (*Carex* spp.). Large symbols represent lakes with higher importance values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the FISH dataset.

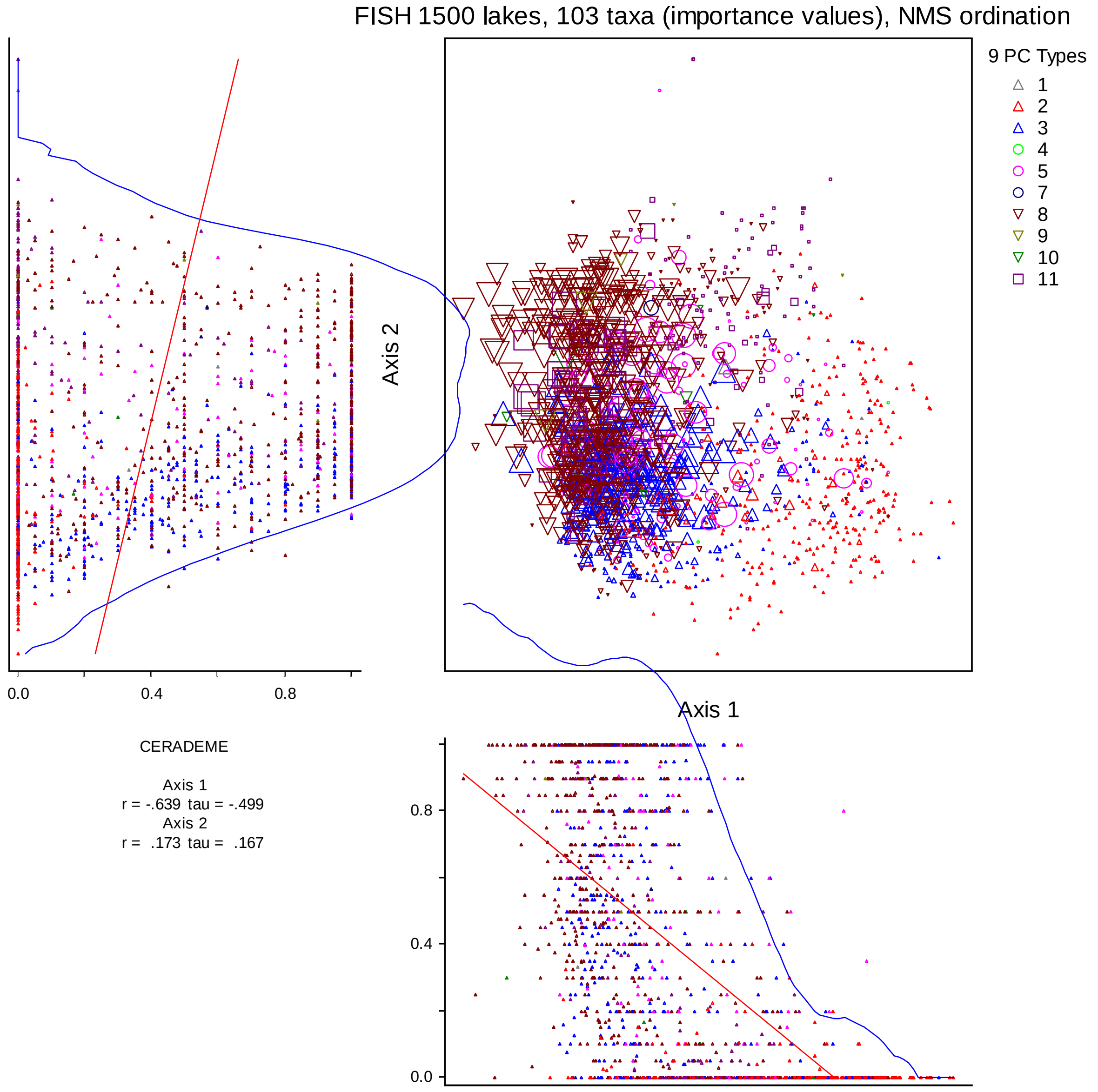


Figure FISH-9: NMS ordination of FISH dataset, overlay of favorability of lakes for **coontail** (*Ceratophyllum demersum*). Large symbols represent lakes with higher importance values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the FISH dataset.

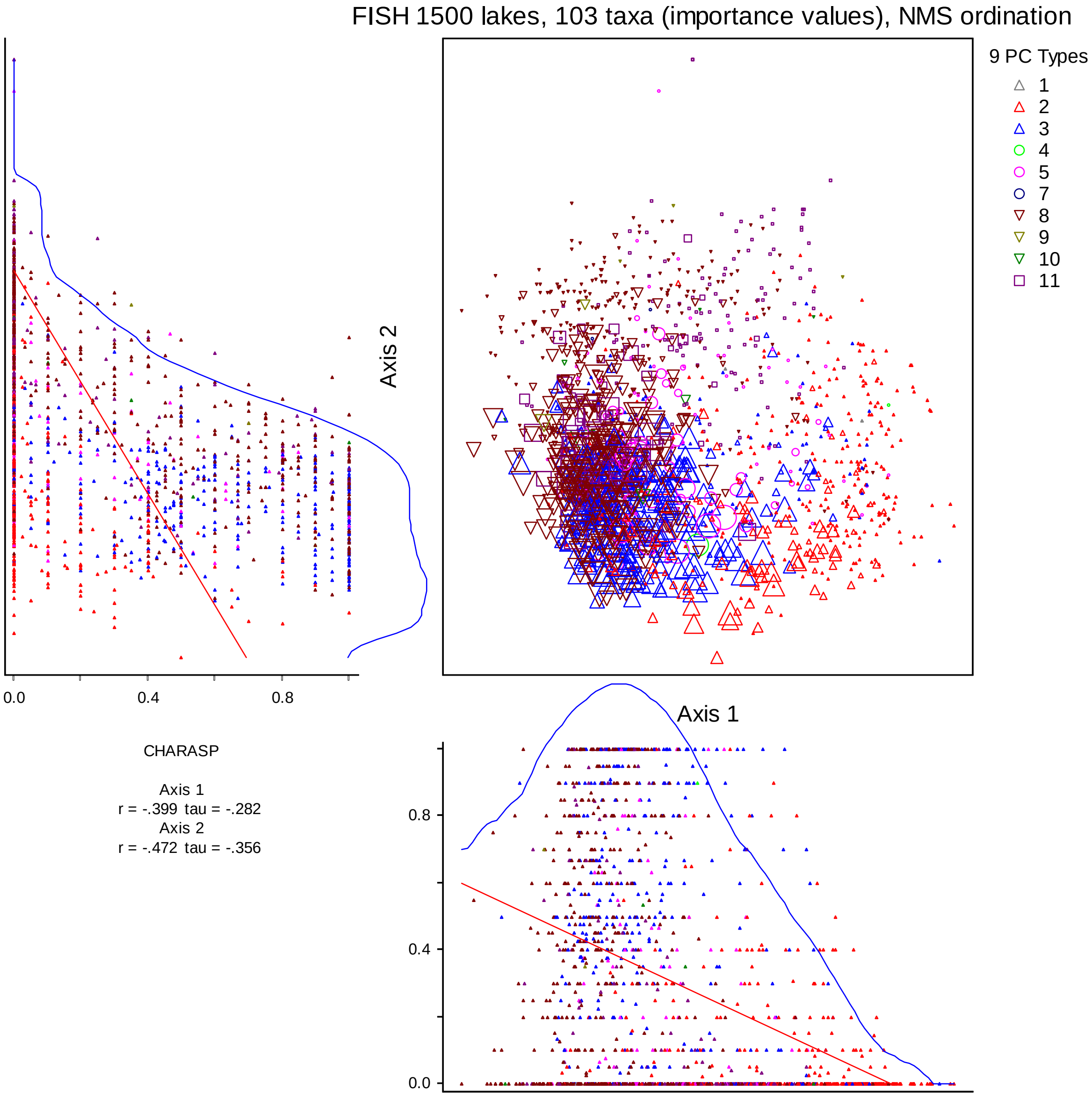


Figure FISH-10: NMS ordination of FISH dataset, overlay of favorability of lakes for **muskgrasses** (*Chara* spp.). Large symbols represent lakes with higher importance values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the FISH dataset.

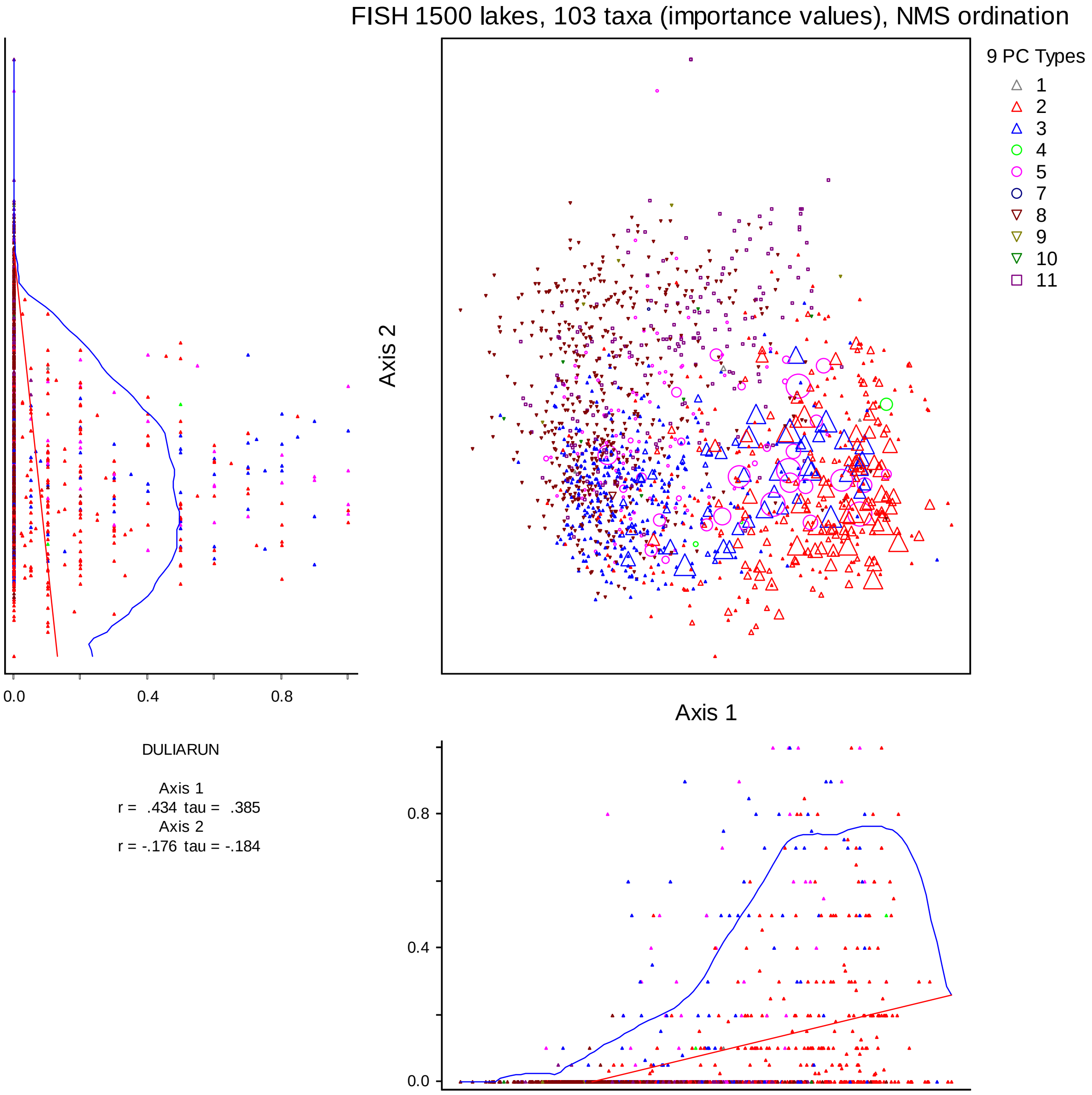


Figure FISH-11: NMS ordination of FISH dataset, overlay of favorability of lakes for **three-way sedge** (*Dulichium arundinaceum*). Large symbols represent lakes with higher importance values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the FISH dataset.

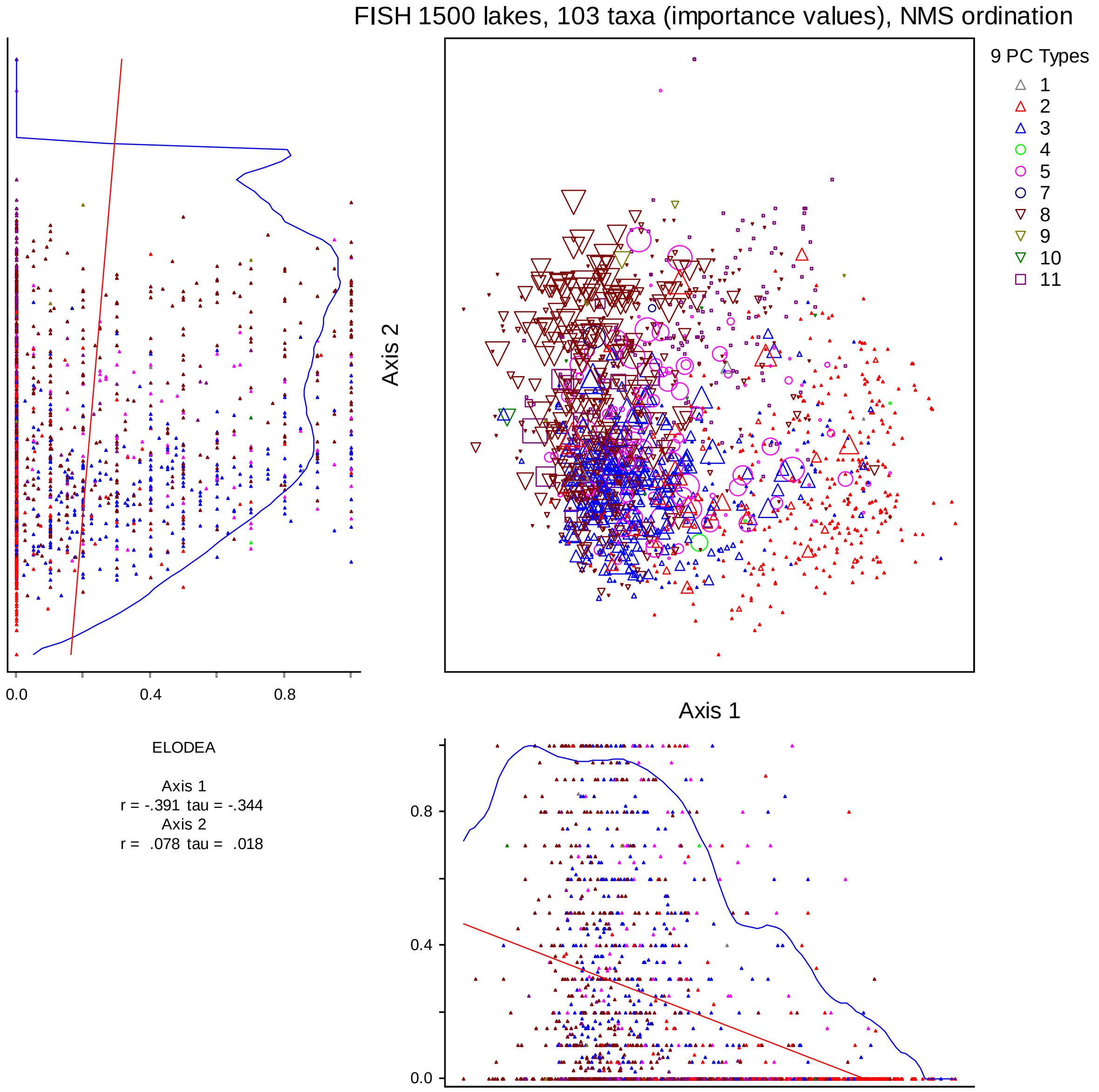


Figure FISH-12: NMS ordination of FISH dataset, overlay of favorability of lakes for **waterweeds** (*Elodea* spp.). Large symbols represent lakes with higher importance values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the FISH dataset.

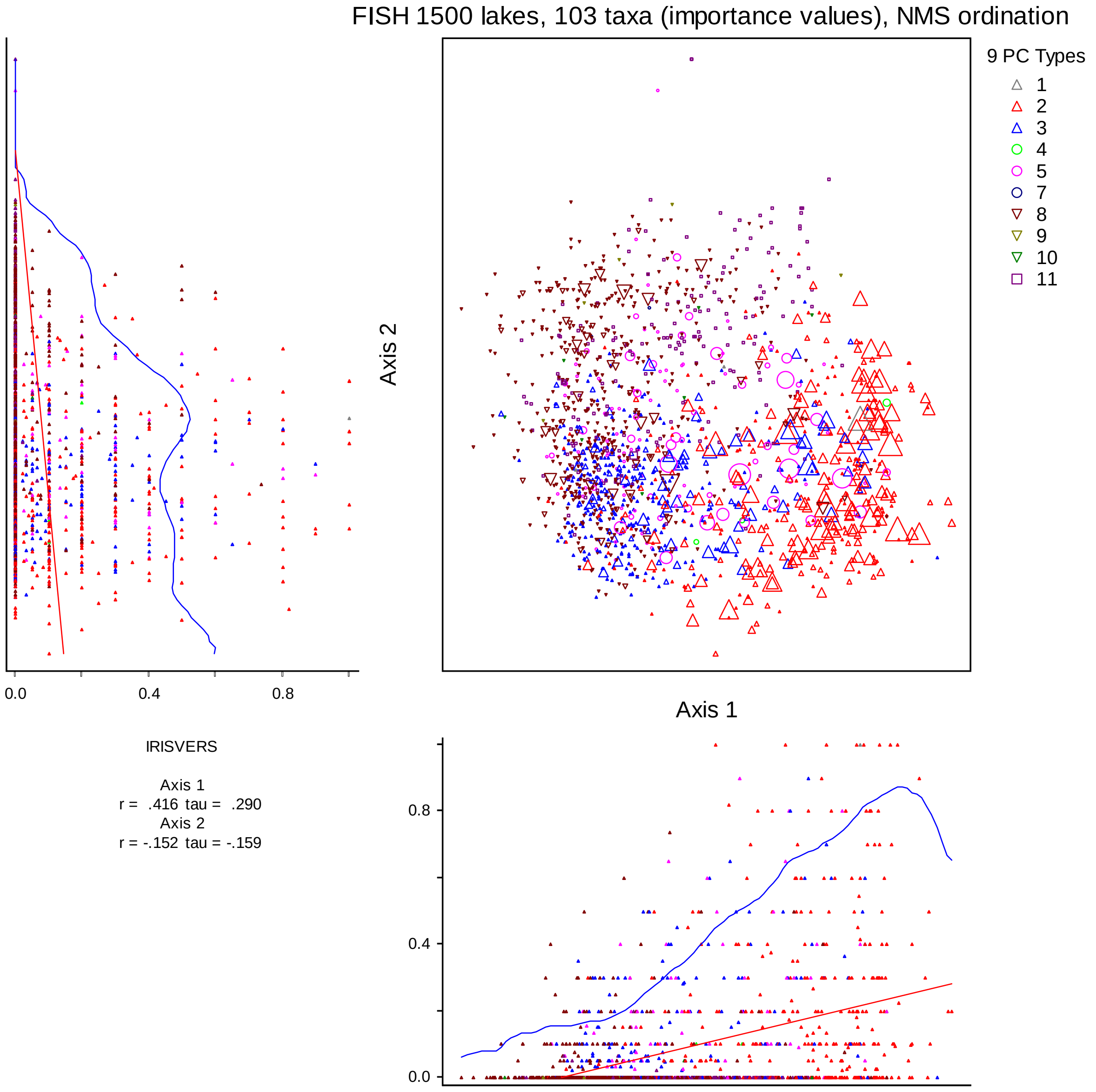


Figure FISH-13: NMS ordination of FISH dataset, overlay of favorability of lakes for **blue flag iris** (*Iris versicolor*). Large symbols represent lakes with higher importance values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the FISH dataset.

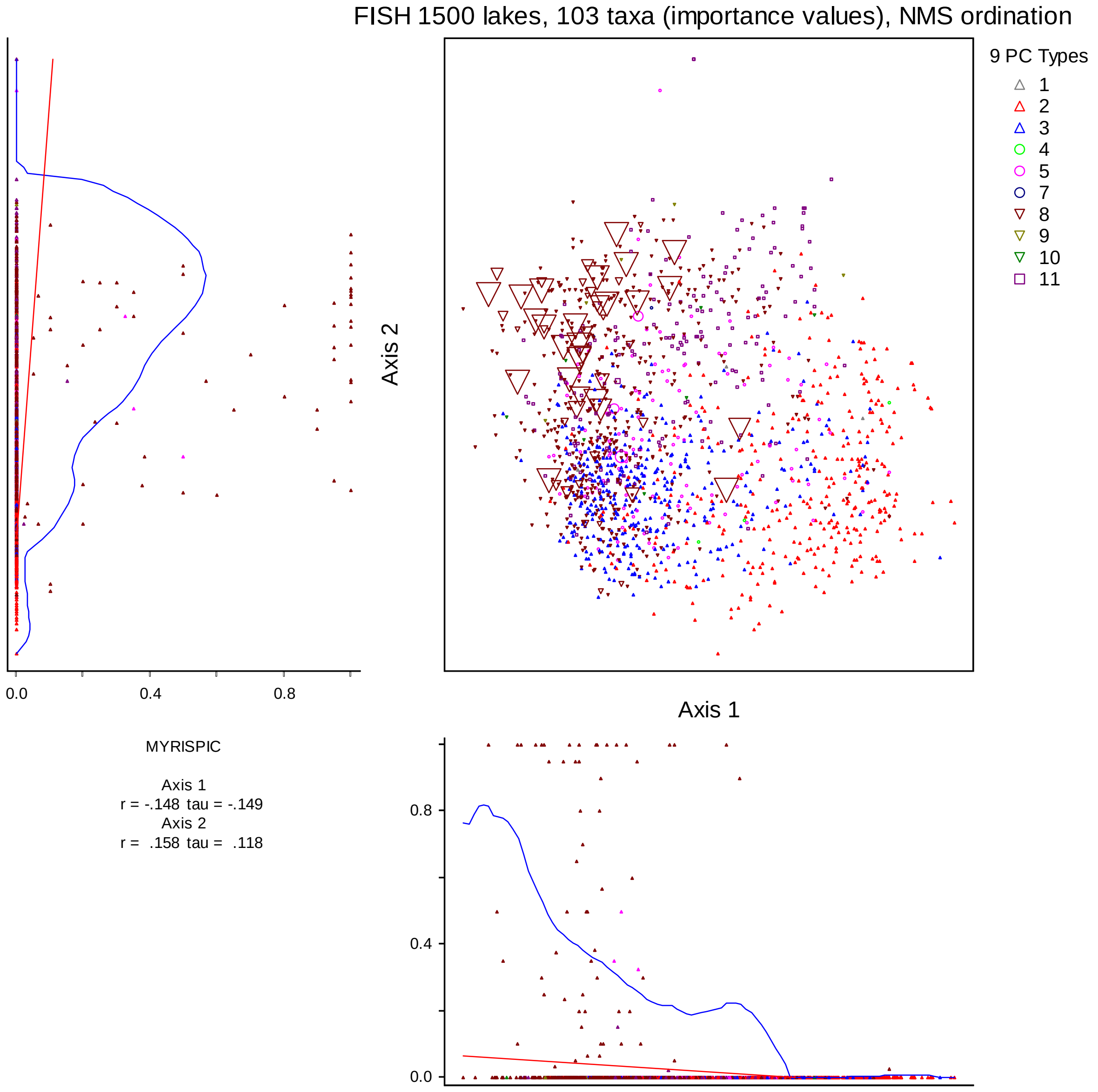


Figure FISH-14: NMS ordination of FISH dataset, overlay of favorability of lakes for exotic **Eurasian milfoil** (*Myriophyllum spicatum*). Large symbols represent lakes with higher importance values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the FISH dataset.

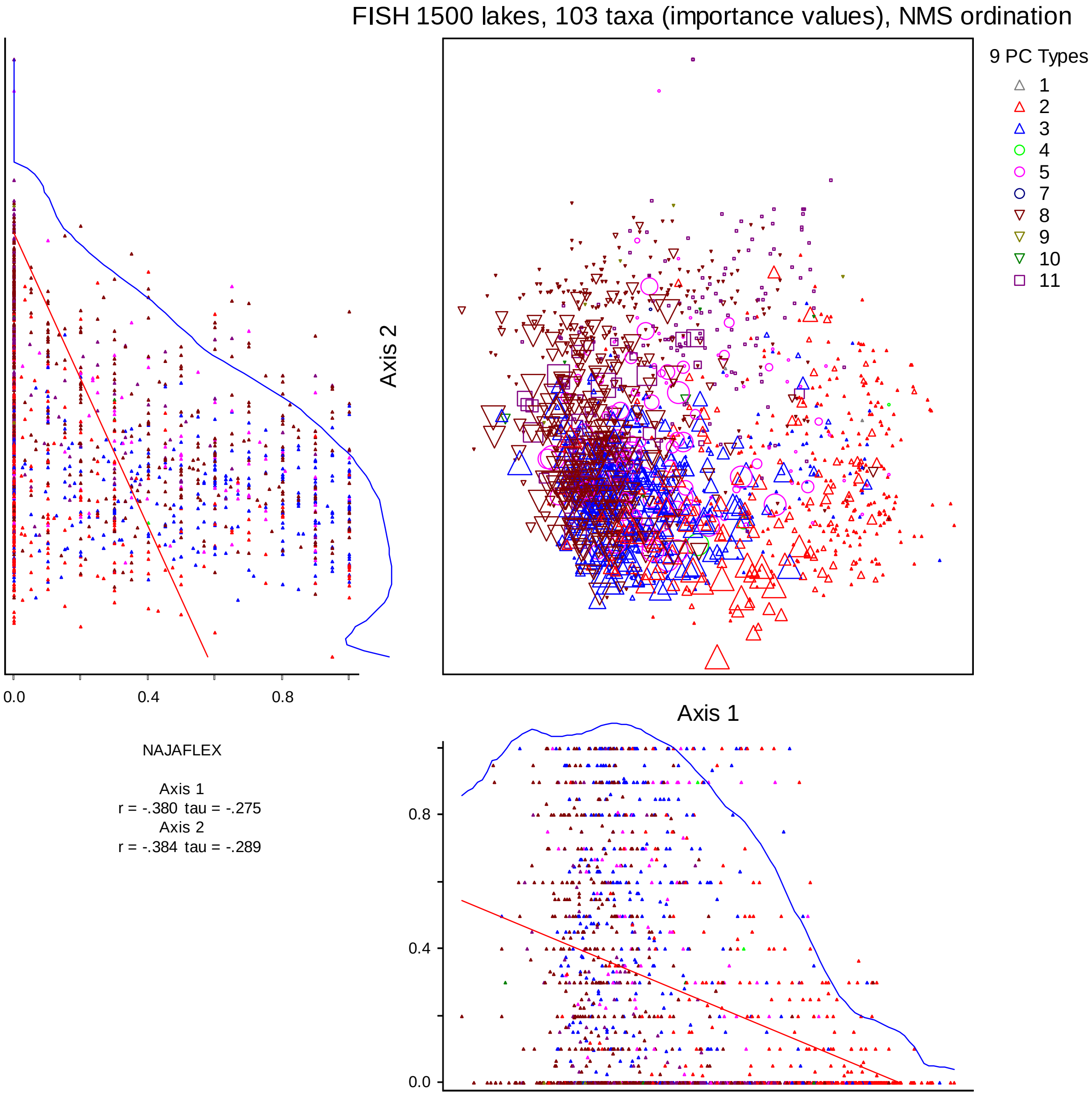


Figure FISH-15: NMS ordination of FISH dataset, overlay of favorability of lakes for **bushy pondweed** (*Najas flexilis*). Large symbols represent lakes with higher importance values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the FISH dataset.

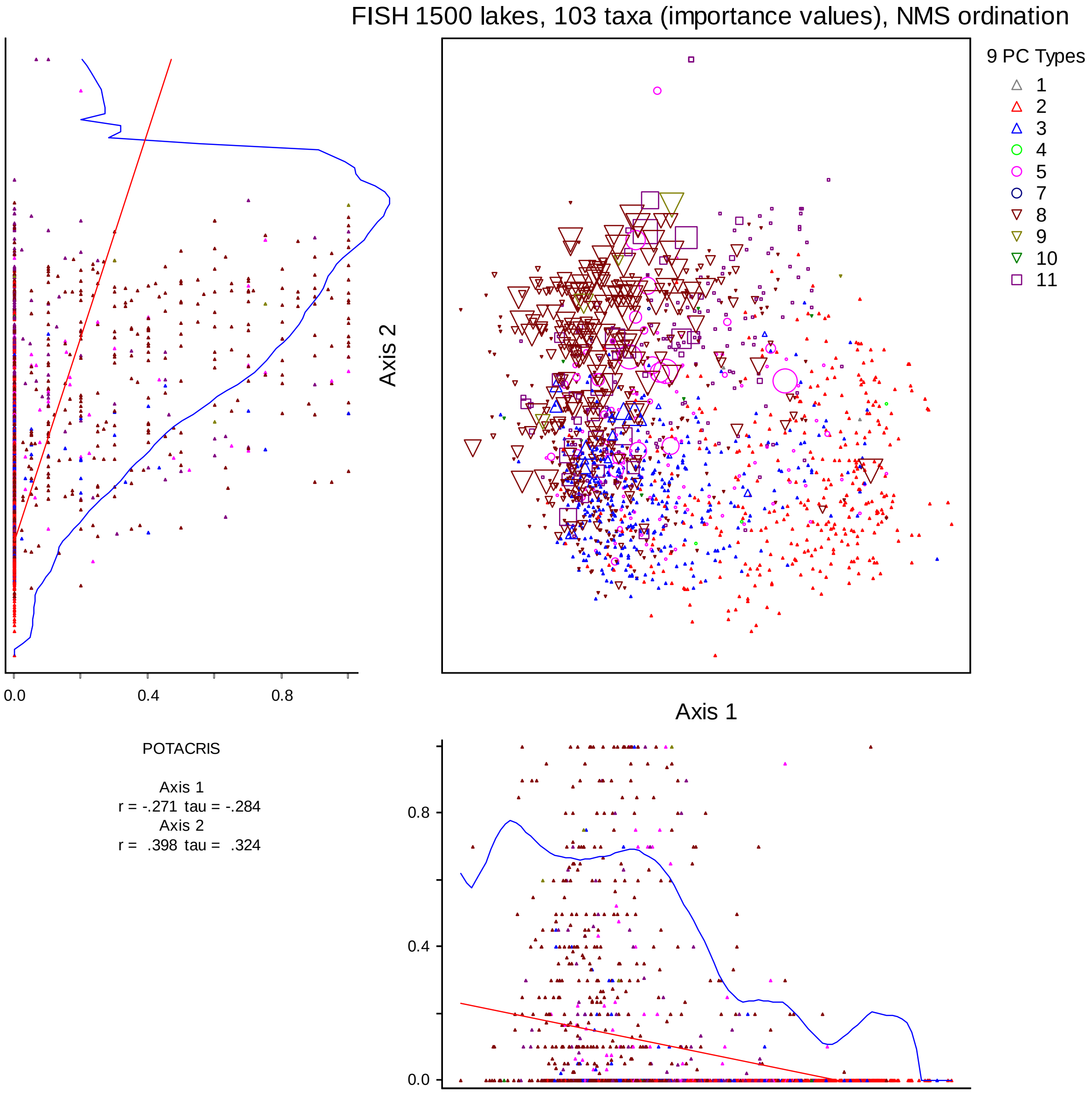


Figure FISH-16: NMS ordination of FISH dataset, overlay of favorability of lakes for exotic **curly-leaf pondweed** (*Potamogeton crispus*). Large symbols represent lakes with higher importance values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the FISH dataset.

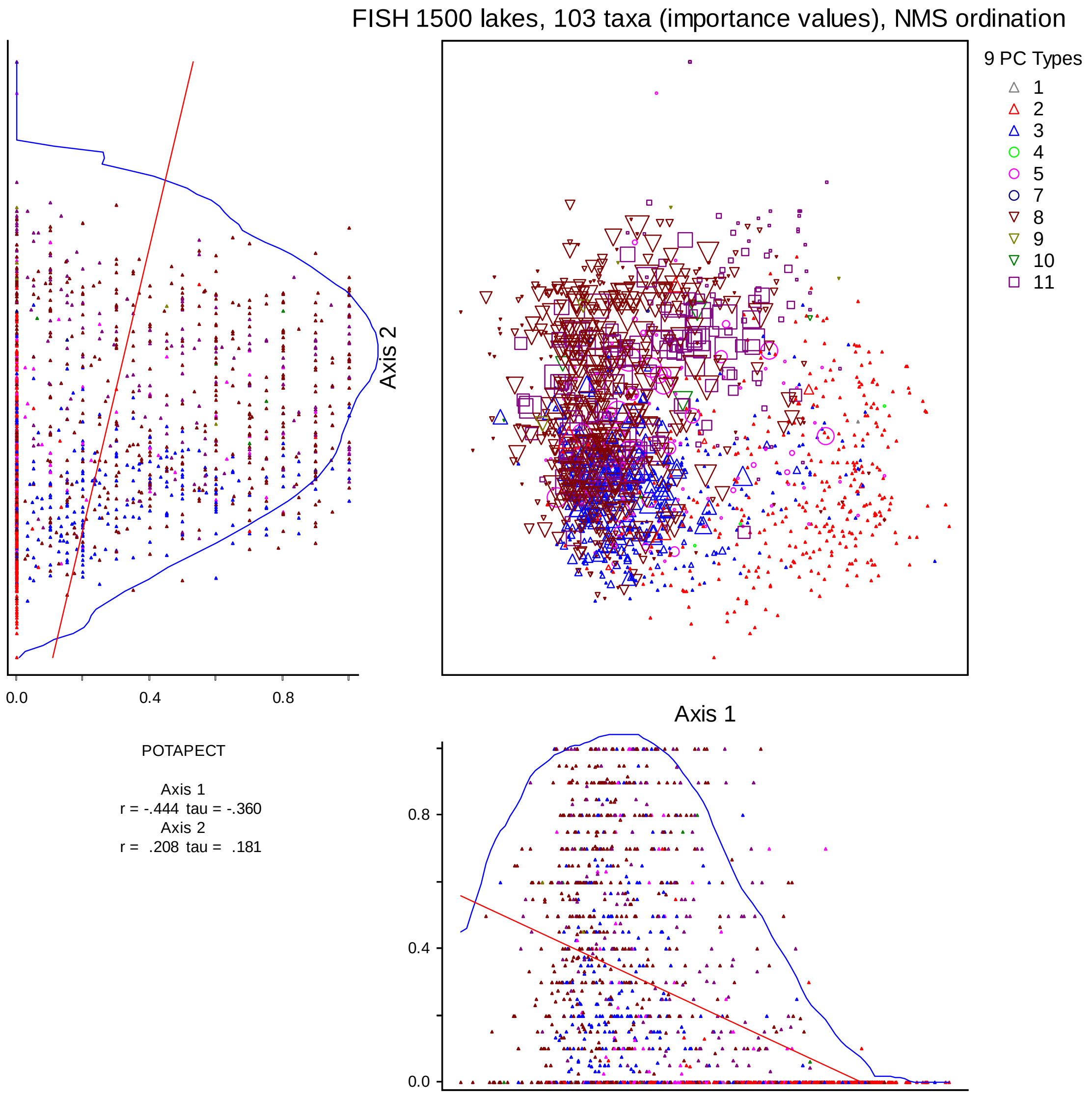


Figure FISH-17: NMS ordination of FISH dataset, overlay of favorability of lakes for **sago pondweed** (*Potamogeton pectinatus*). Large symbols represent lakes with higher importance values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the FISH dataset.

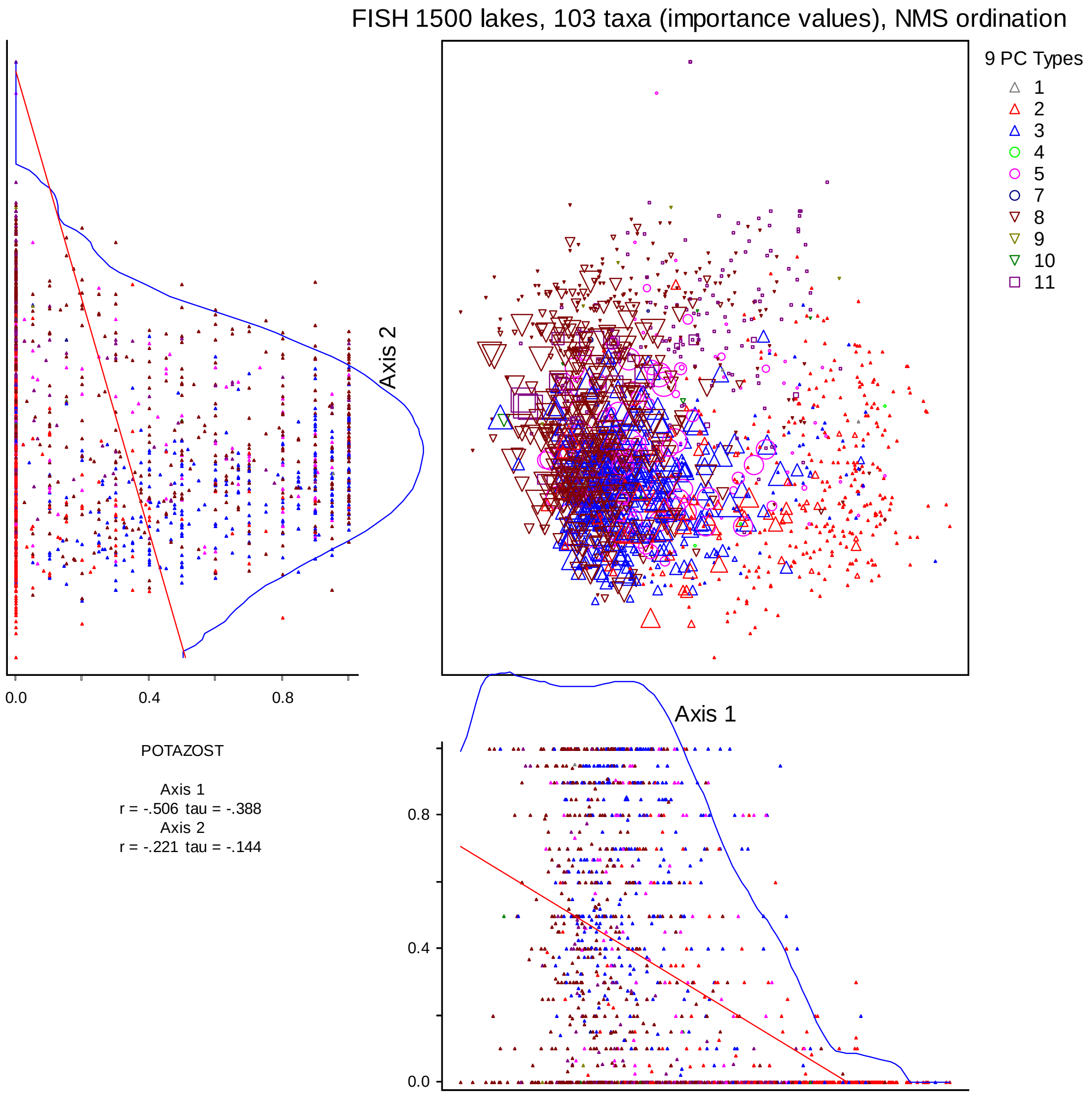


Figure FISH-18: NMS ordination of FISH dataset, overlay of favorability of lakes for **flat-stem pondweed** (*Potamogeton zosteriformis*). Large symbols represent lakes with higher importance values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the FISH dataset.

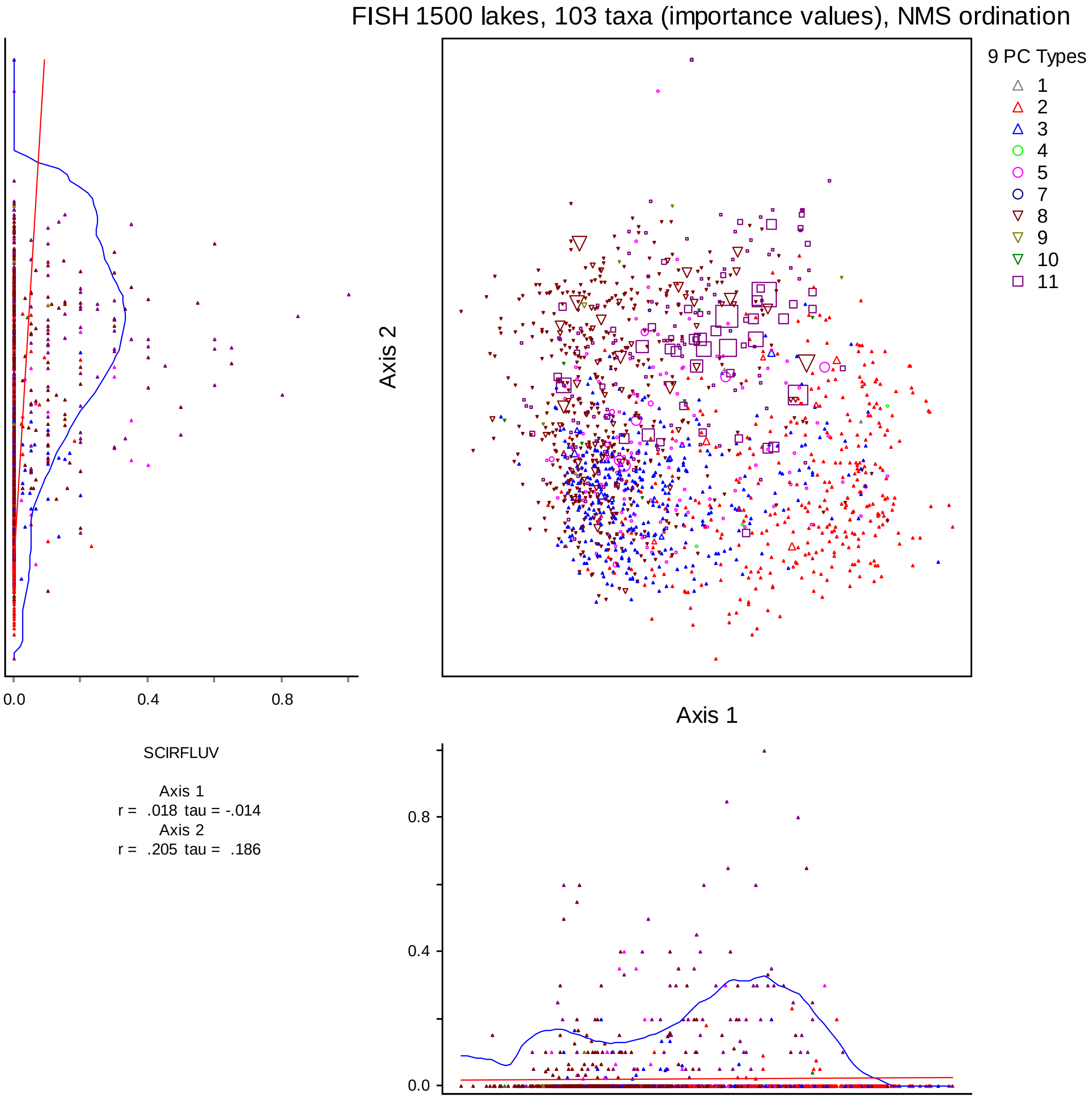


Figure FISH-19: NMS ordination of FISH dataset, overlay of favorability of lakes for **river bulrush** (*Scirpus fluviatilis*). Large symbols represent lakes with higher importance values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the FISH dataset.

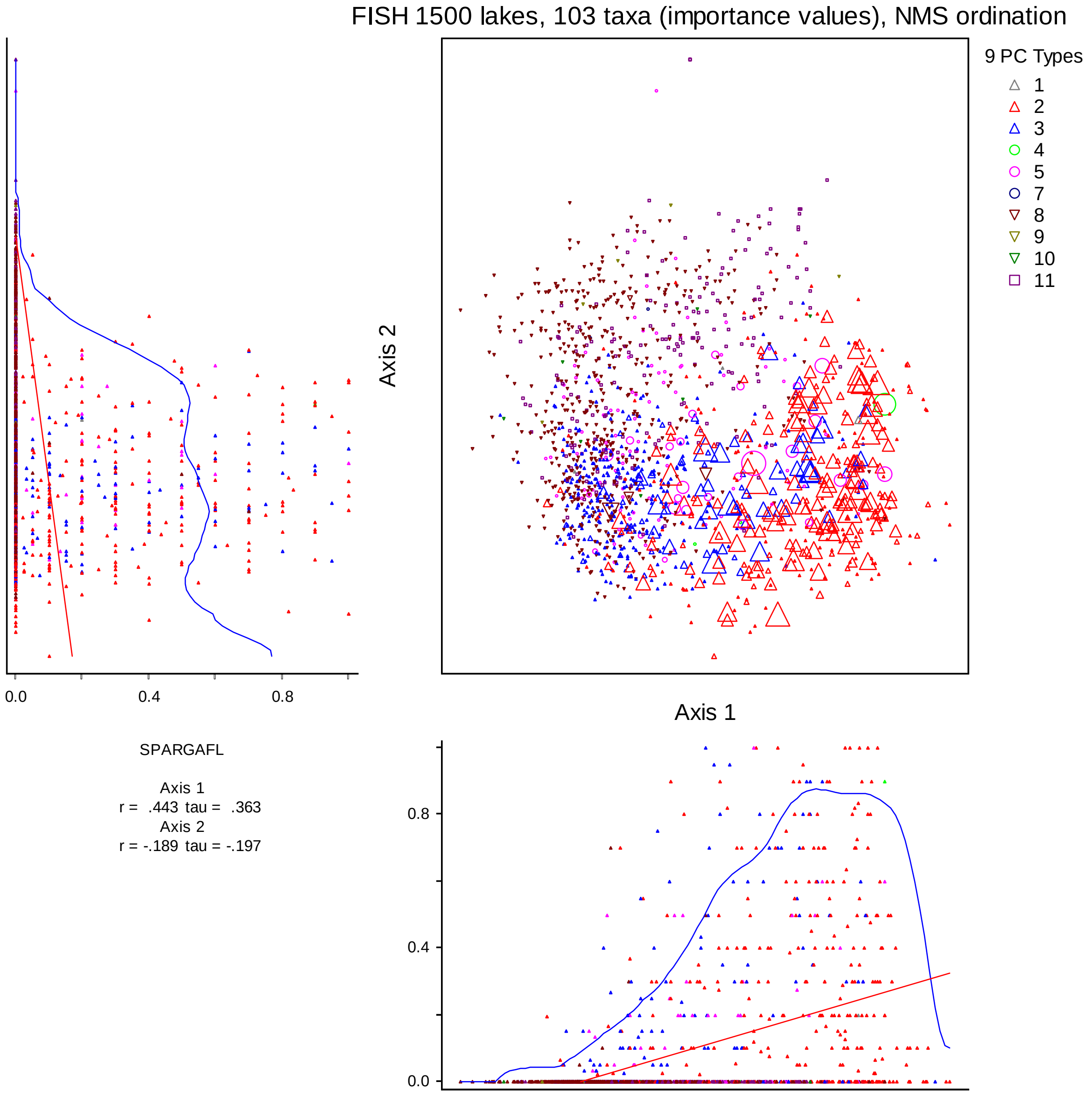


Figure FISH-20: NMS ordination of FISH dataset, overlay of favorability of lakes for **floating-leaved burreeds** (*Sparganium* spp.: *S. angustifolium*, *S. fluctuans*, *S minimum*). Large symbols represent lakes with higher importance values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the FISH dataset.

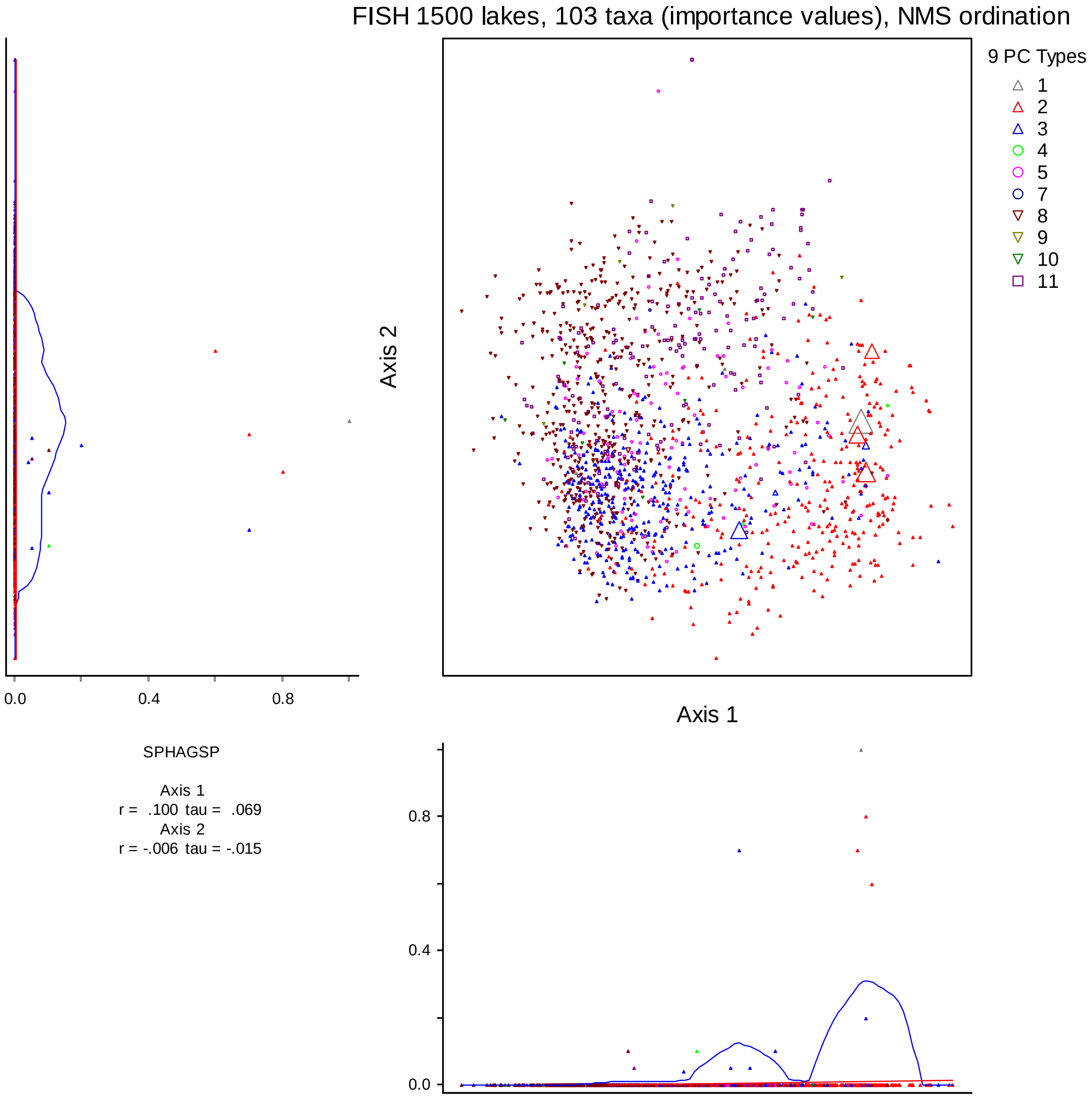


Figure FISH-21: NMS ordination of FISH dataset, overlay of favorability of lakes for **peat mosses** (*Sphagnum* spp.). Large symbols represent lakes with higher importance values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the FISH dataset.

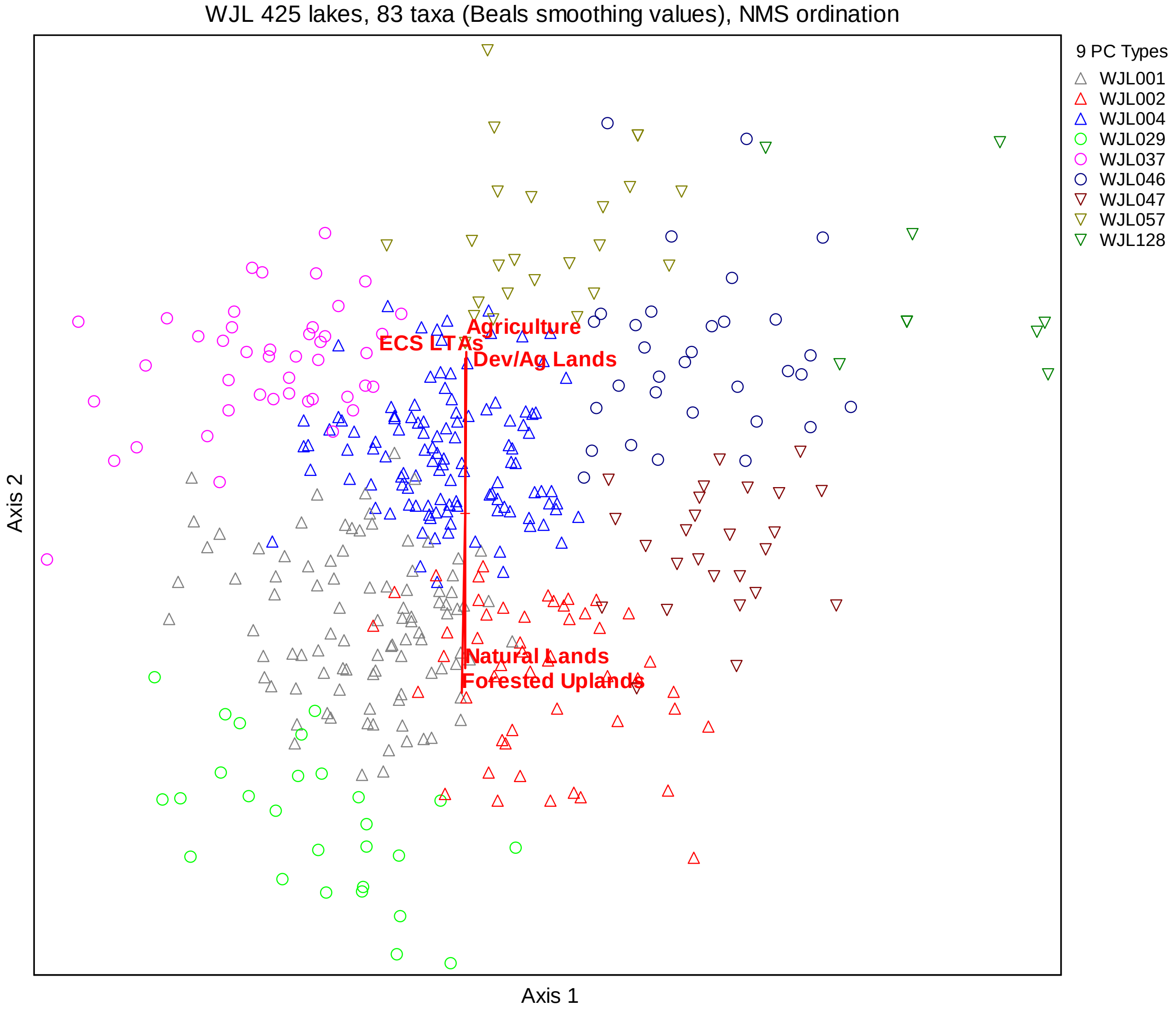


Figure WJL-1: NMS ordination of WJL dataset; joint plot showing lines of strongest correlation with quantitative environmental variables. The strongest correlations with the lower end of Axis 2 are with proportion of **Natural Lands** and **Forested Uplands** in the 1 km buffer around each lake. The strongest correlations with the upper end of Axis 2 are with proportion of **Agricultural Lands**, **Developed/Ag Lands, and ECS Land Type Associations (LTAs)** in the buffer. Symbol colors and shapes represent the 9 PC types recognized in analysis of the WJL dataset.

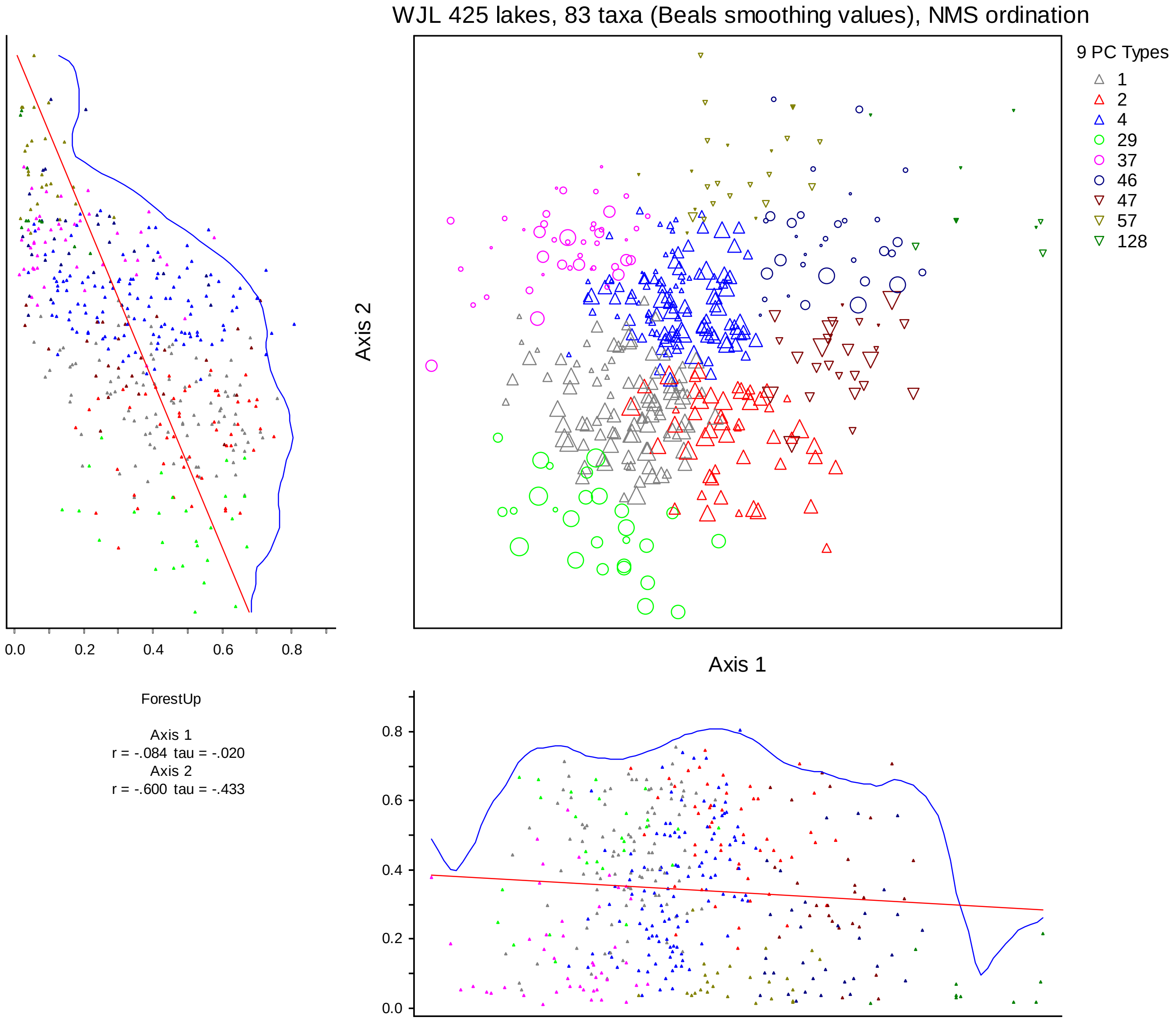


Figure WJL-2: NMS ordination of WJL dataset, overlay of proportion of 1 km buffer around each lake that is **Forested Uplands**. Larger symbols represent lakes in landscapes with higher proportions of upland forest in the buffer. Symbol colors and shapes represent the 9 PC types recognized in analysis of the WJL dataset.

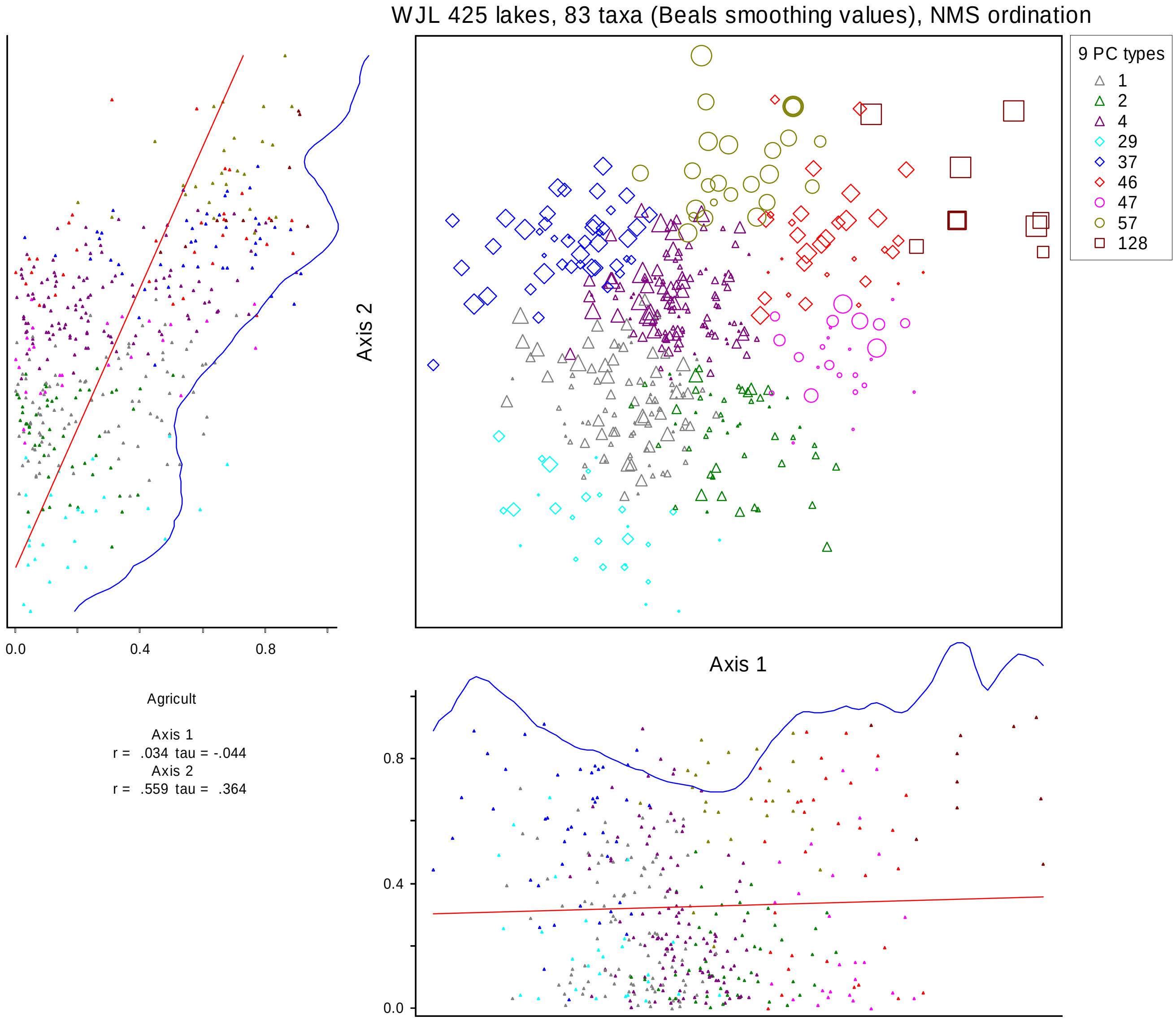


Figure WJL-3: NMS ordination of WJL dataset, overlay of proportion of 1 km buffer around each lake that is **Agriculture Lands**. Larger symbols represent lakes in landscapes with higher proportions of agriculture lands in the buffer. Symbol colors and shapes represent the 9 PC types recognized in analysis of the WJL dataset.

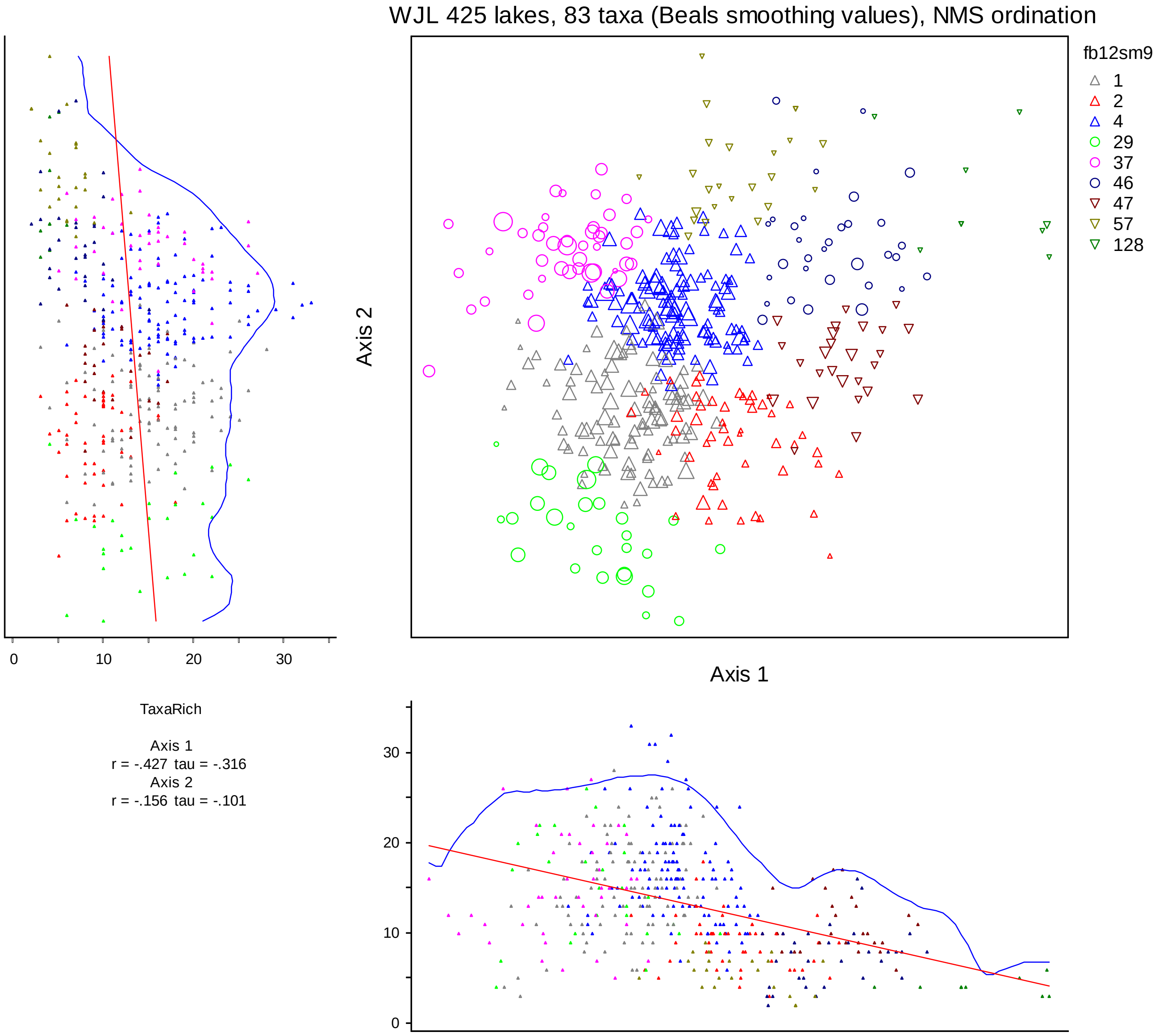


Figure WJL-4: NMS ordination of WJL dataset, overlay of **Taxa Richness** of each lake. Larger symbols represent lakes with higher numbers of taxa recorded for each lake. Symbol colors and shapes represent the 9 PC types recognized in analysis of the WJL dataset.

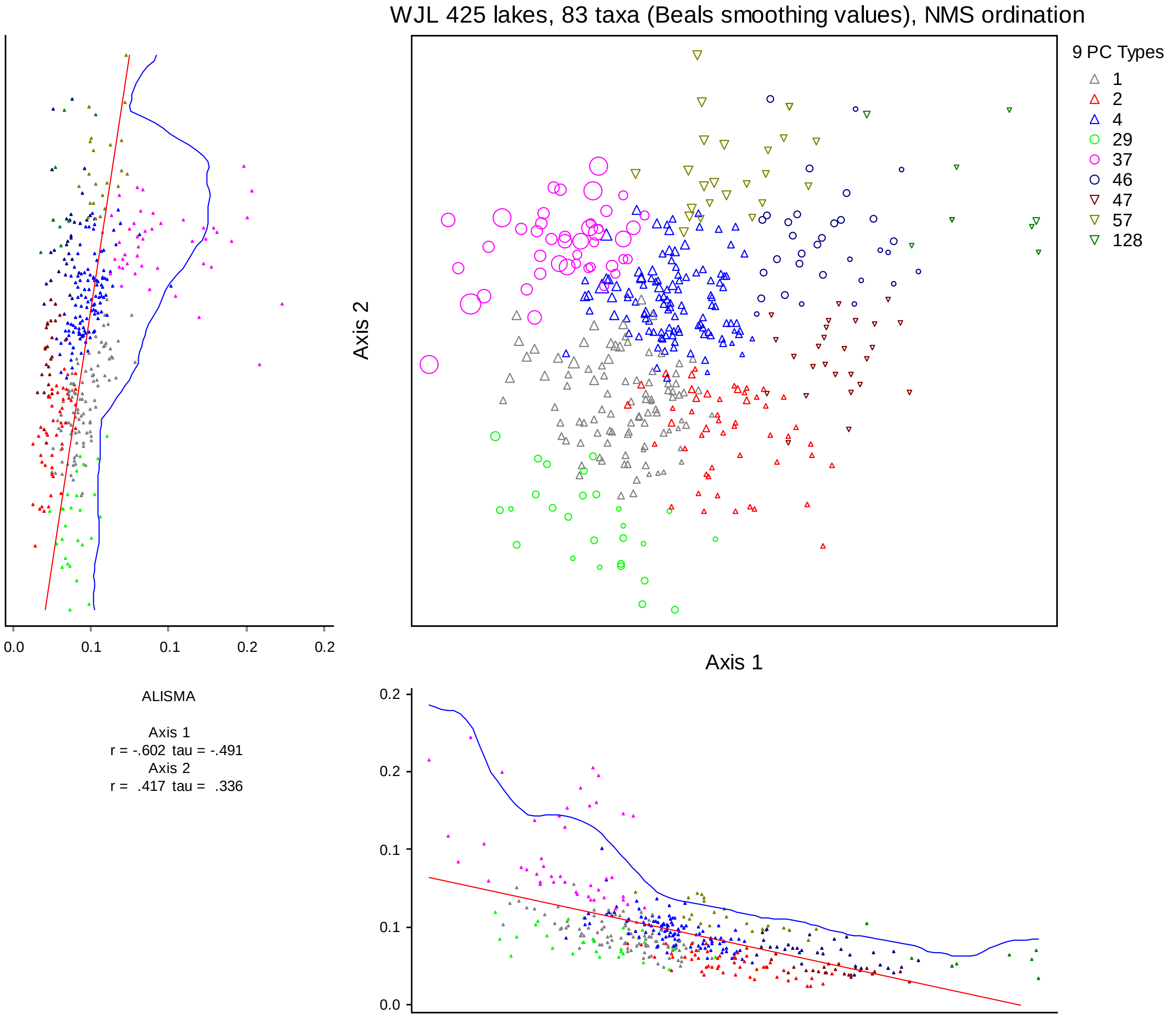


Figure WJL-5: NMS ordination of WJL dataset, overlay of favorability of lakes for **water plantains** (*Alisma* spp., includes *A. gramineum*, *A. triviale*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the WJL dataset.

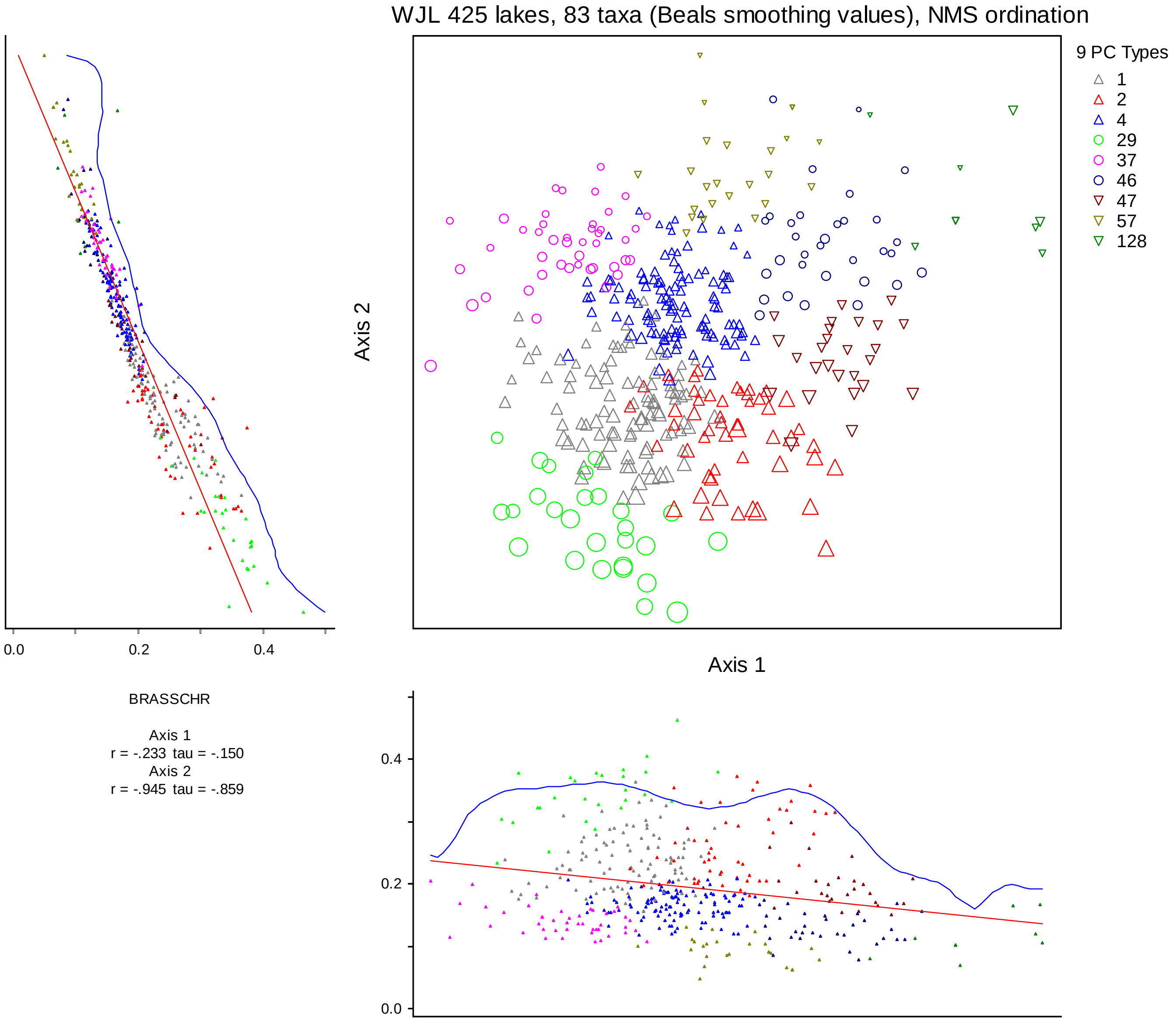


Figure WJL-6: NMS ordination of WJL dataset, overlay of favorability of lakes for **water shield** (*Brasenia schreberi*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the WJL dataset.

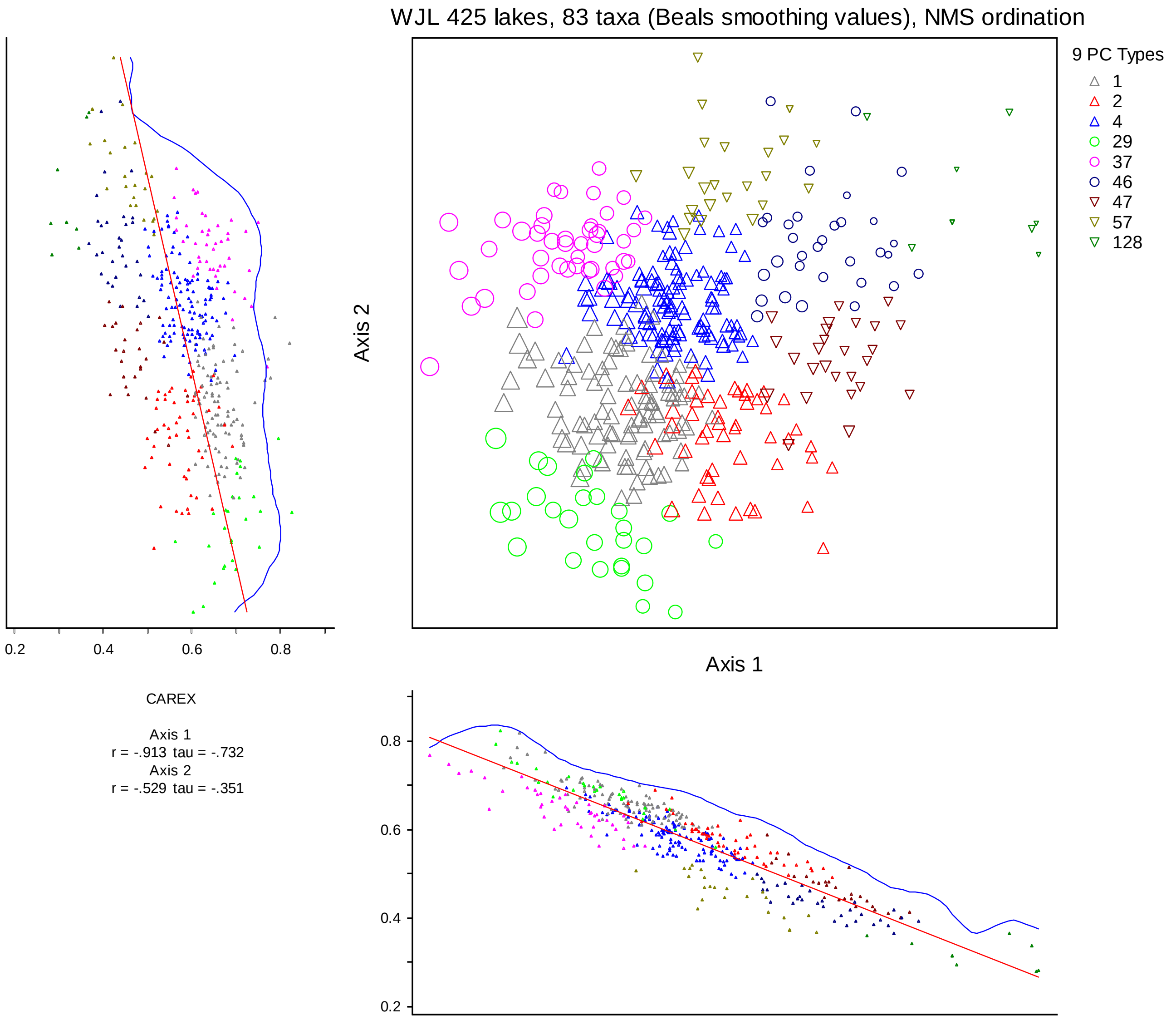


Figure WJL-7: NMS ordination of WJL dataset, overlay of favorability of lakes for **sedges** (*Carex* spp.). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the WJL dataset.

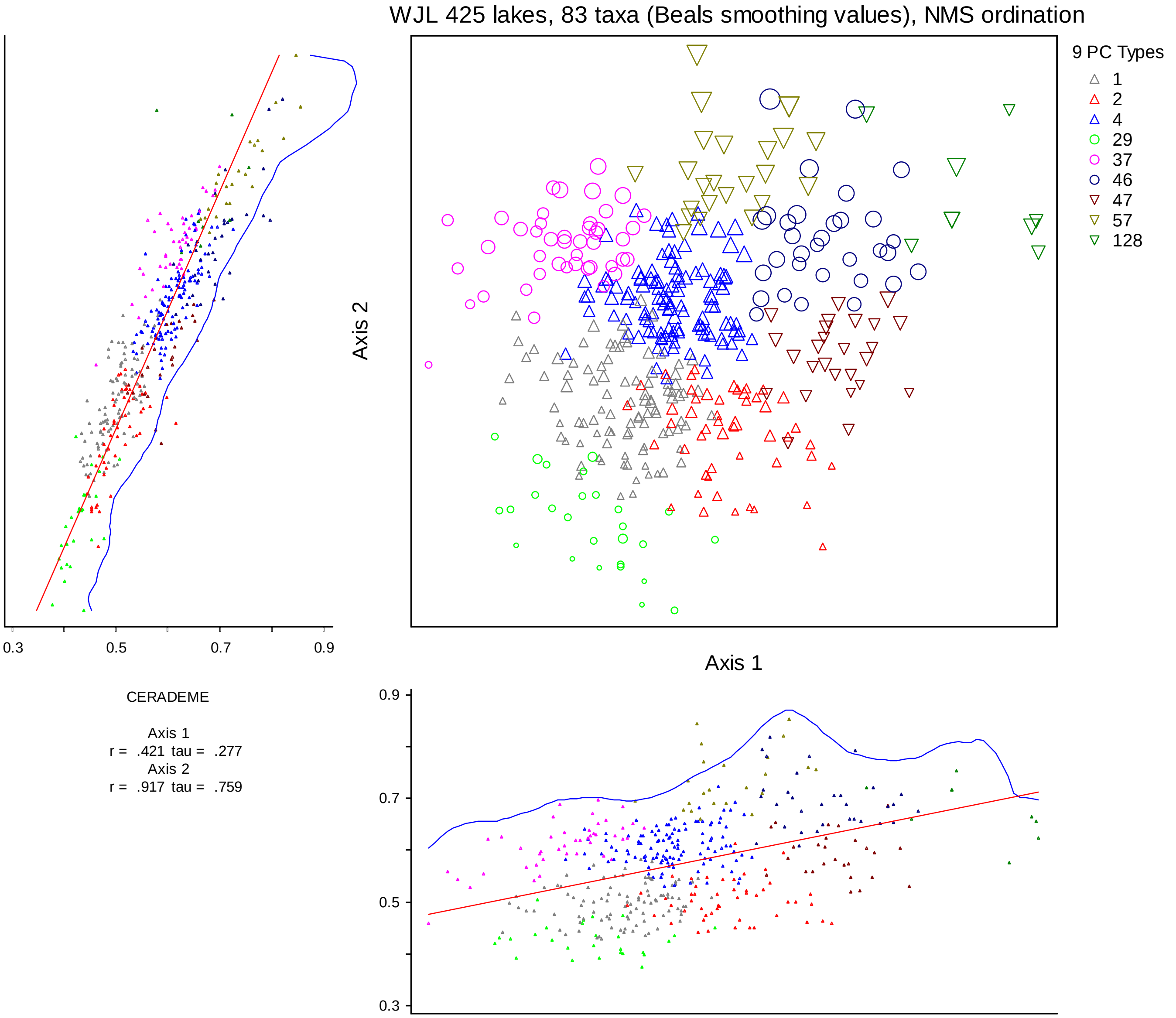


Figure WJL-8: NMS ordination of WJL dataset, overlay of favorability of lakes for **coontail** (*Ceratophyllum demersum*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the WJL dataset.

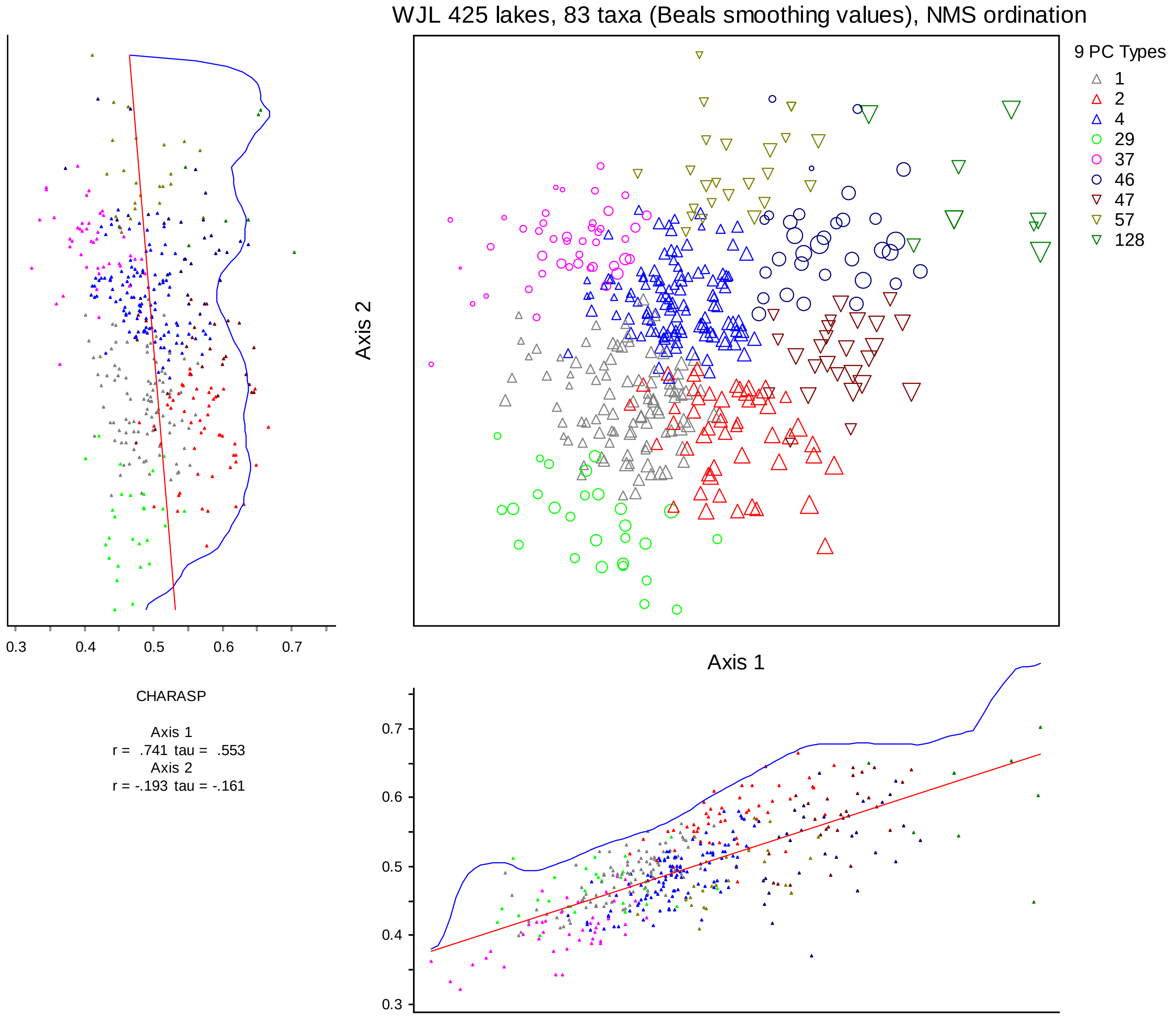


Figure WJL-9: NMS ordination of WJL dataset, overlay of favorability of lakes for **muskgrasses** (*Chara* spp.). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the WJL dataset.

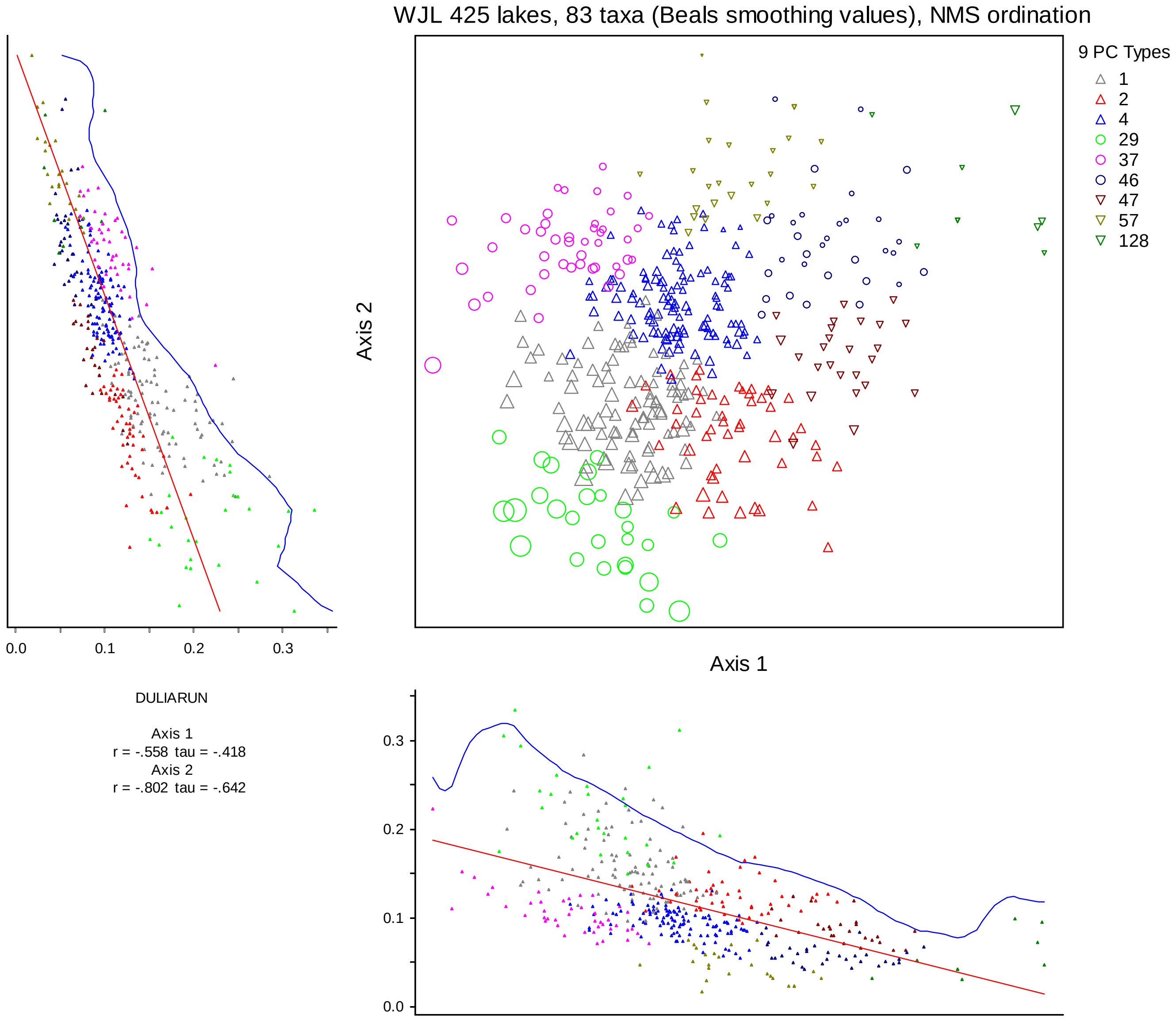


Figure WJL-10: NMS ordination of WJL dataset, overlay of favorability of lakes for **three-way sedge** (*Dulichium arundinaceum*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the WJL dataset.

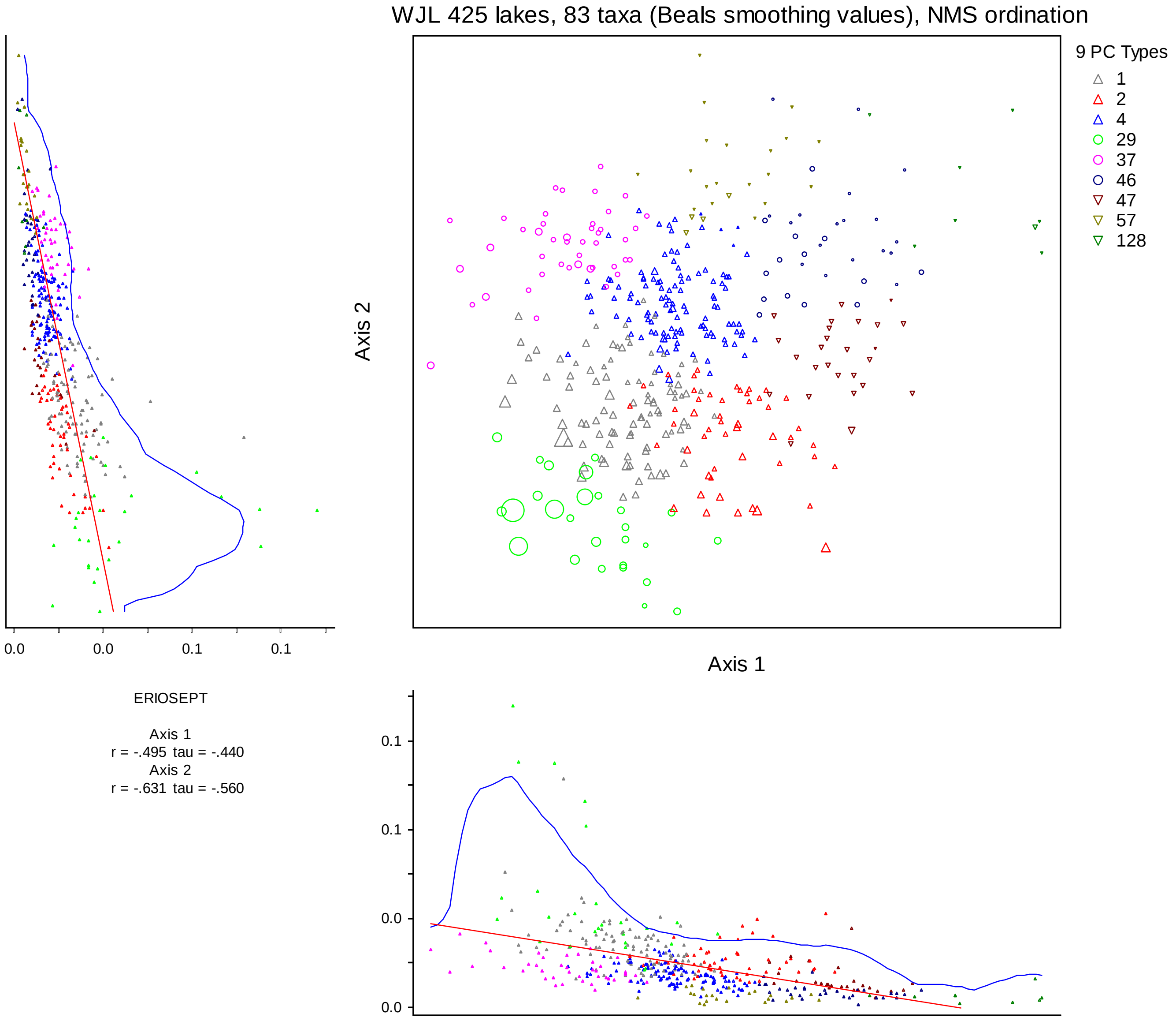


Figure WJL-11: NMS ordination of WJL dataset, overlay of favorability of lakes for **pipewort** (*Eroocaulon septangulare*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the WJL dataset.

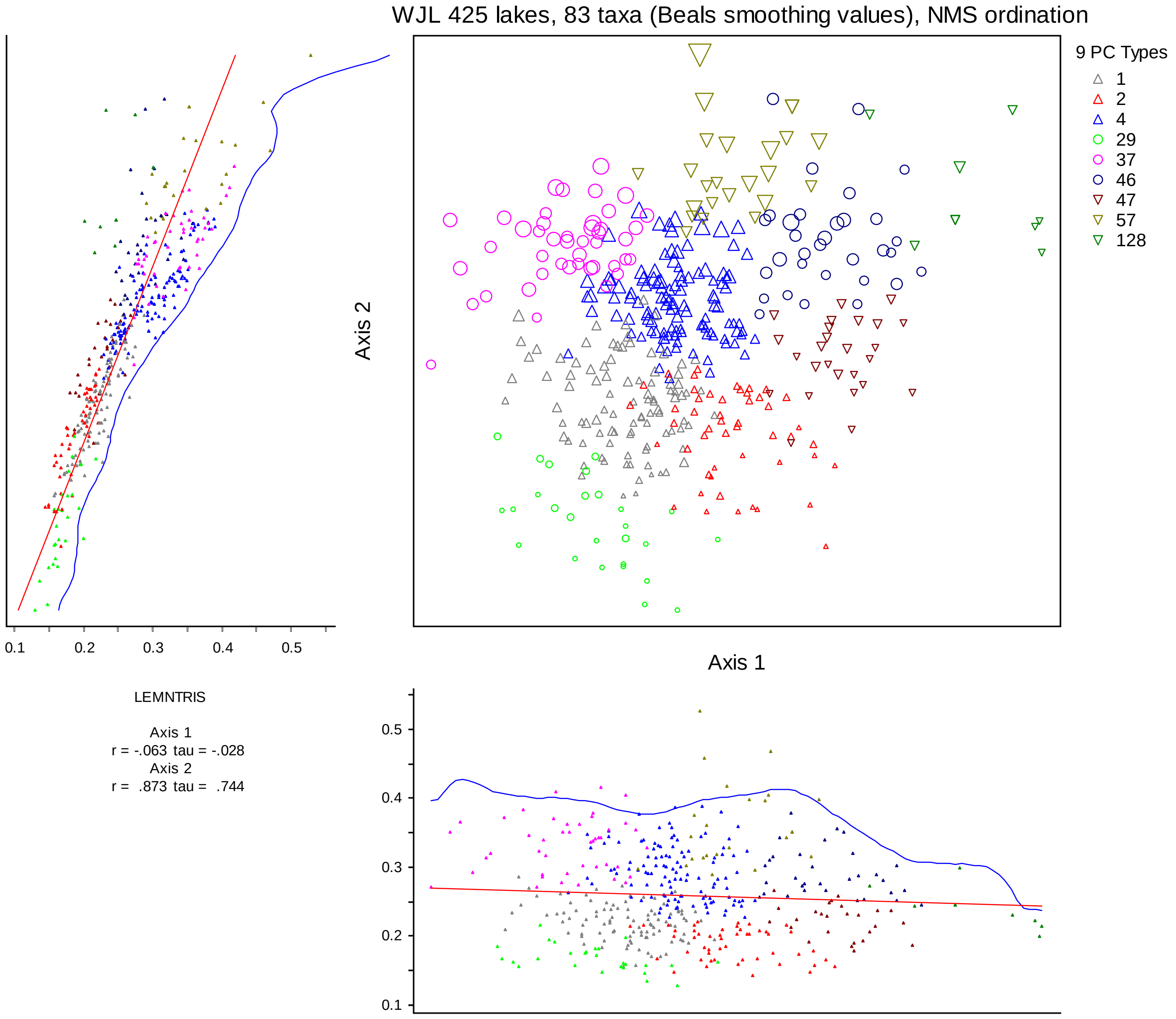


Figure WJL-12: NMS ordination of WJL dataset, overlay of favorability of lakes for **star duckweed** (*Lemna trisulca*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the WJL dataset.

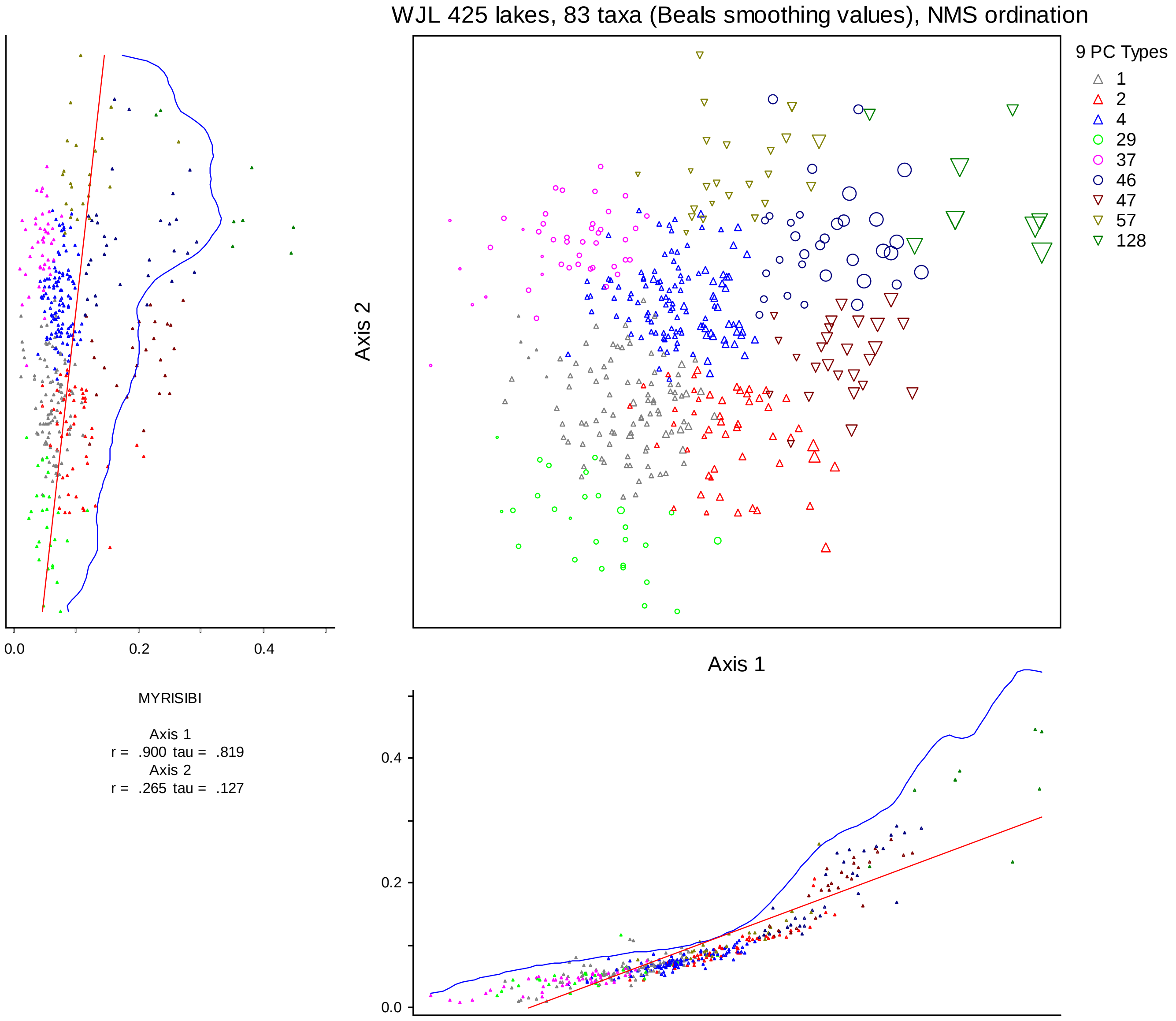


Figure WJL-13: NMS ordination of WJL dataset, overlay of favorability of lakes for exotic **northern milfoil** (*Myriophyllum sibiricum*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the WJL dataset.

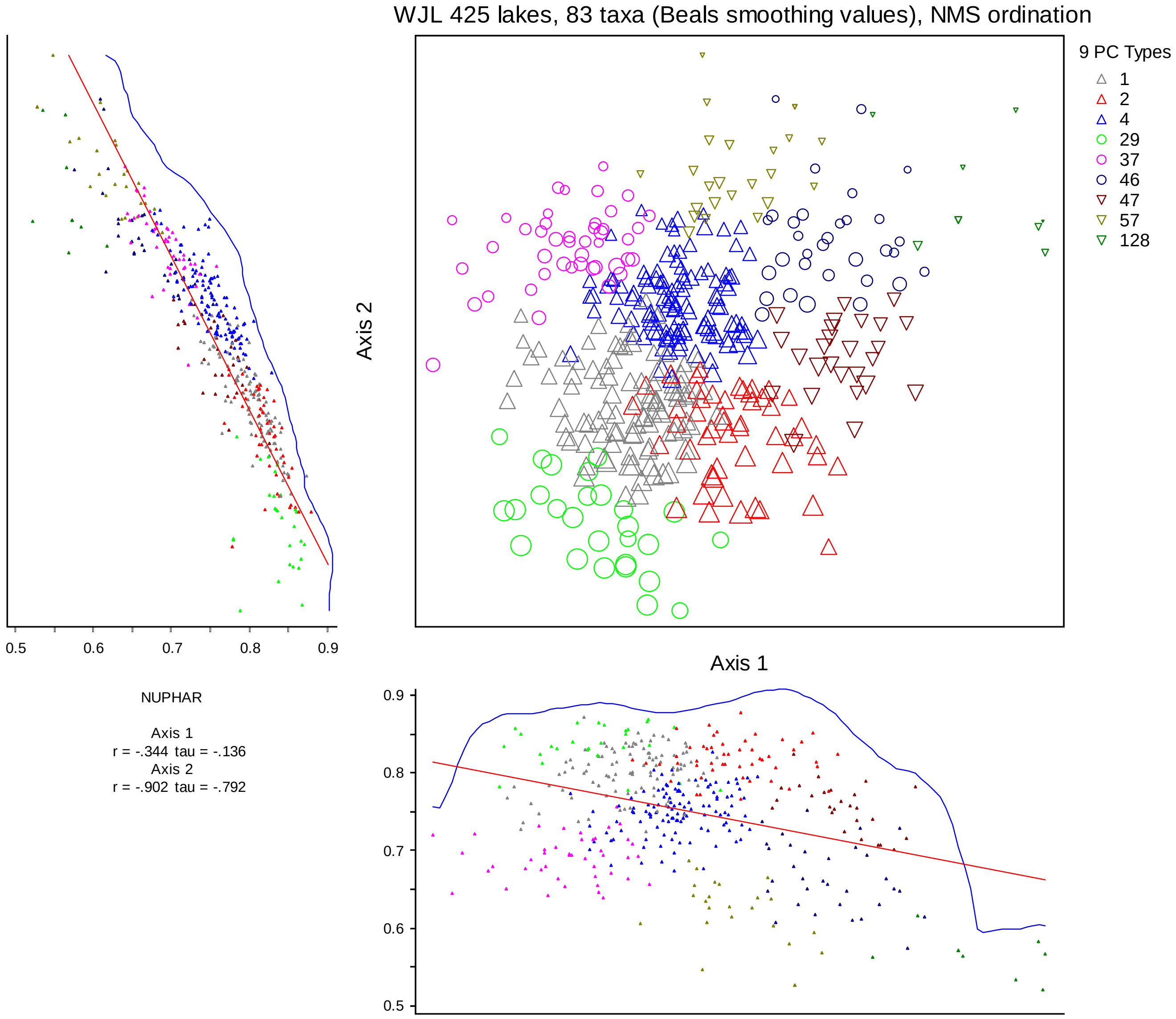


Figure WJL-14: NMS ordination of WJL dataset, overlay of favorability of lakes for **yellow waterlilies** (*Nuphar* spp., includes *N. lutea* ssp. *pumila*, ssp. *rubrodisca*, ssp. *variegata*, *N. microphylla*, *N. rubrodisca*, and *N. variegata*). Large symbols represent lakes with higher favorability for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the WJL dataset.

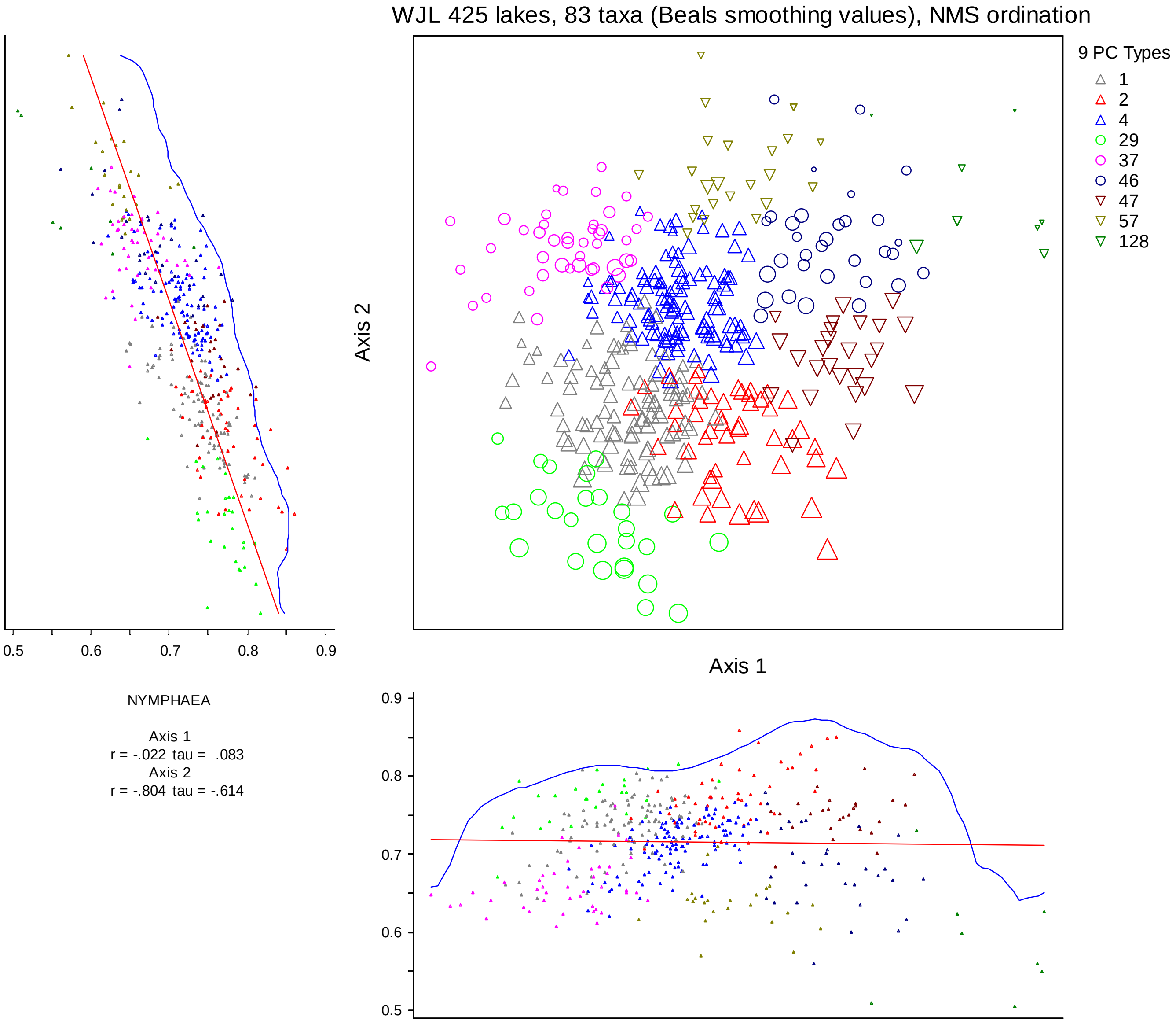


Figure WJL-15: NMS ordination of WJL dataset, overlay of favorability of lakes for **white waterlilies** (*Nymphaea* spp., includes *N. leibergii*, *N. odorata*, and *N. tuberosa*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the WJL dataset.

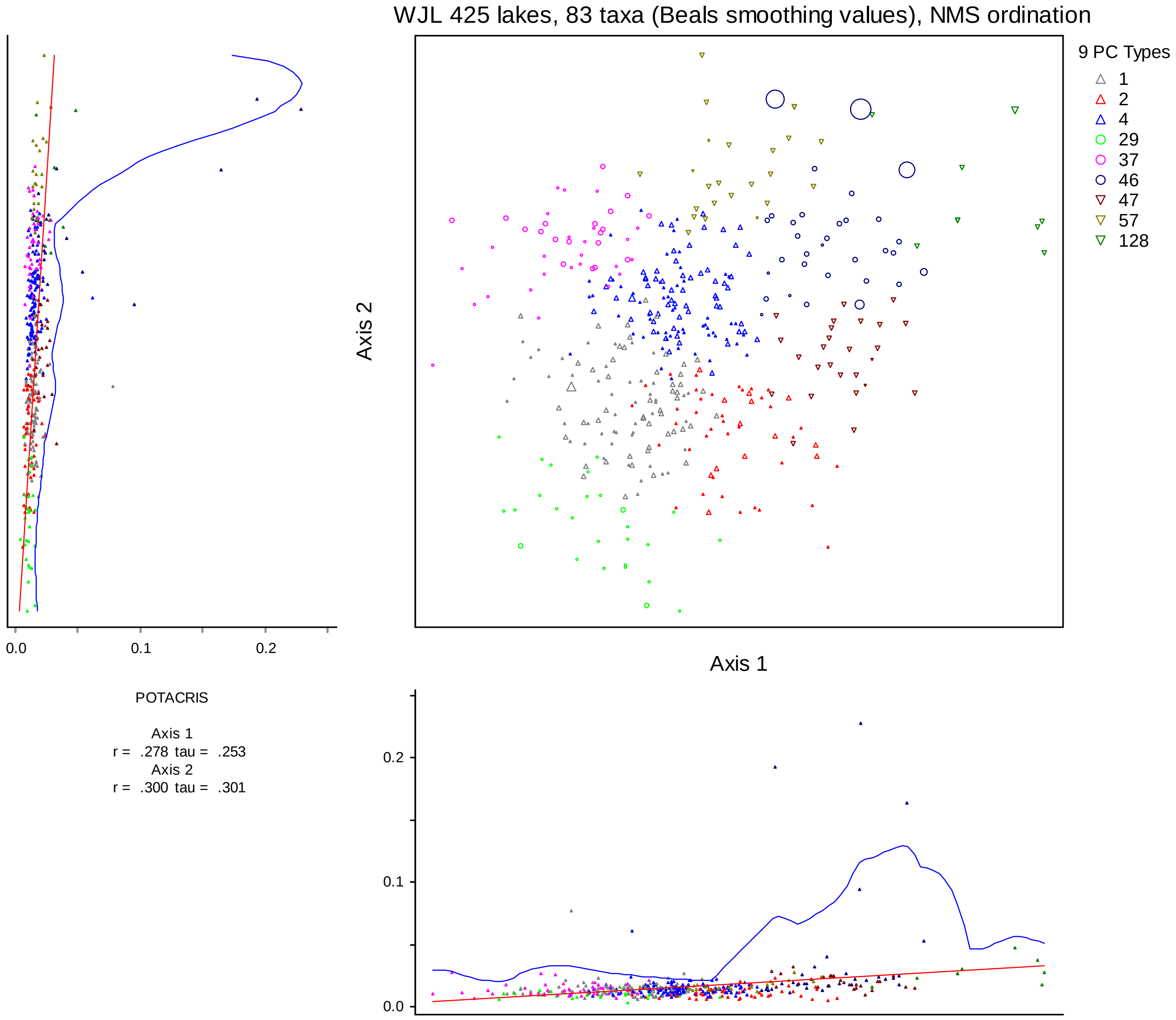


Figure WJL-16: NMS ordination of WJL dataset, overlay of favorability of lakes for exotic **curly-leaf pondweed** (*Potamogeton crispus*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the WJL dataset.

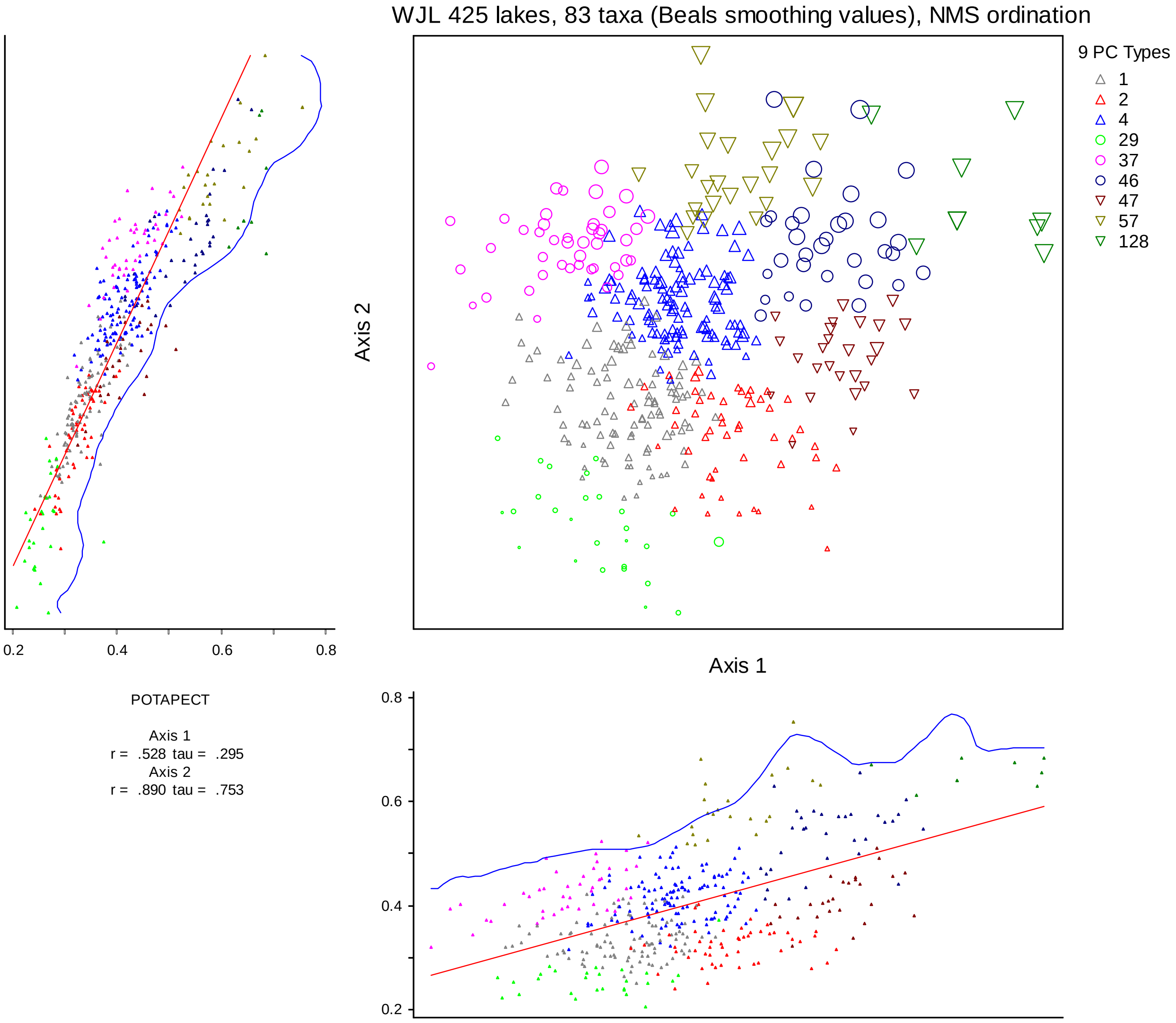


Figure WJL-17: NMS ordination of WJL dataset, overlay of favorability of lakes for **sago pondweed** (*Potamogeton pectinatus*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the WJL dataset.

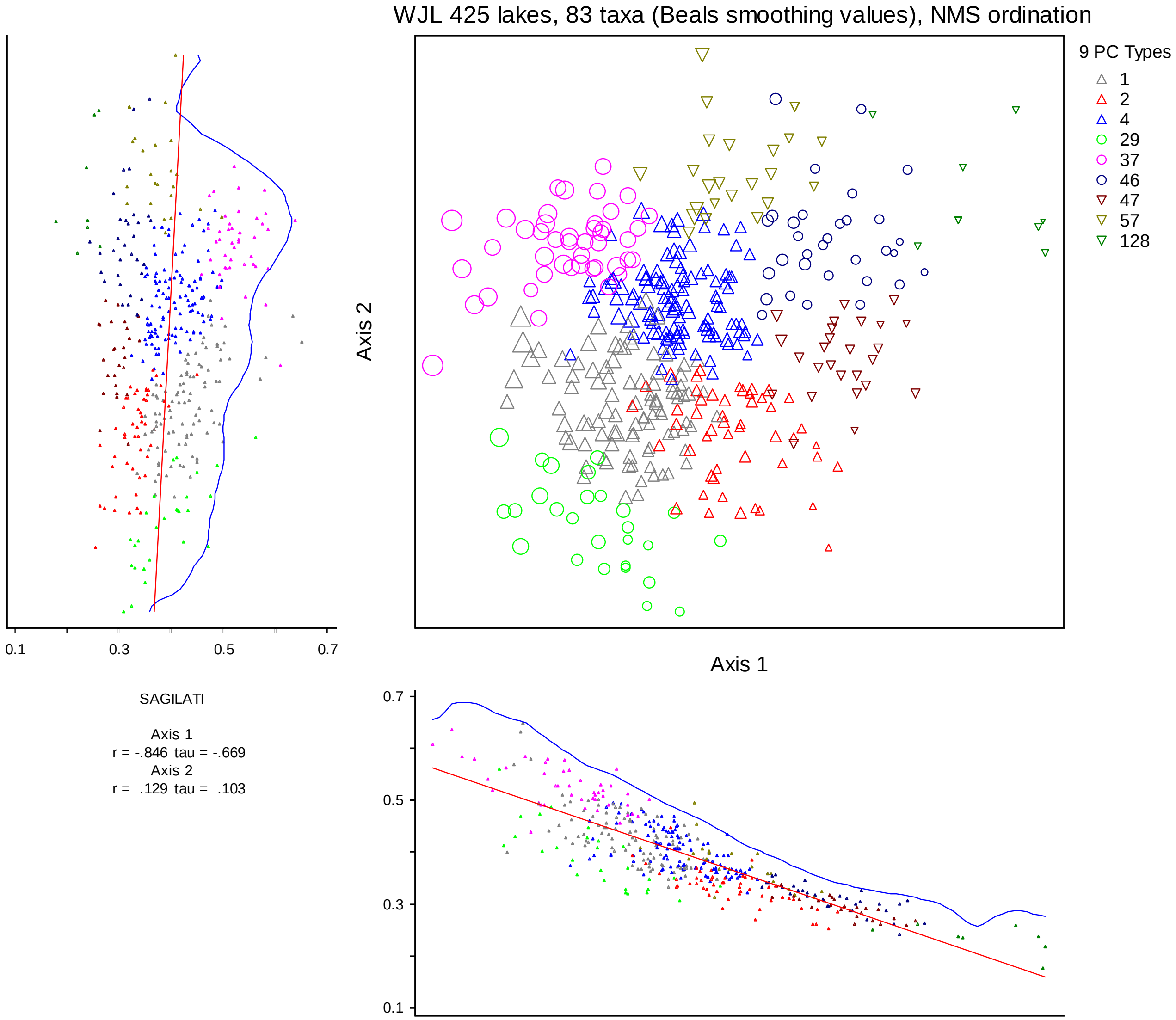


Figure WJL-18: NMS ordination of WJL dataset, overlay of favorability of lakes for **broad-leaved arrowhead** (*Sagittaria latifolia*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the WJL dataset.

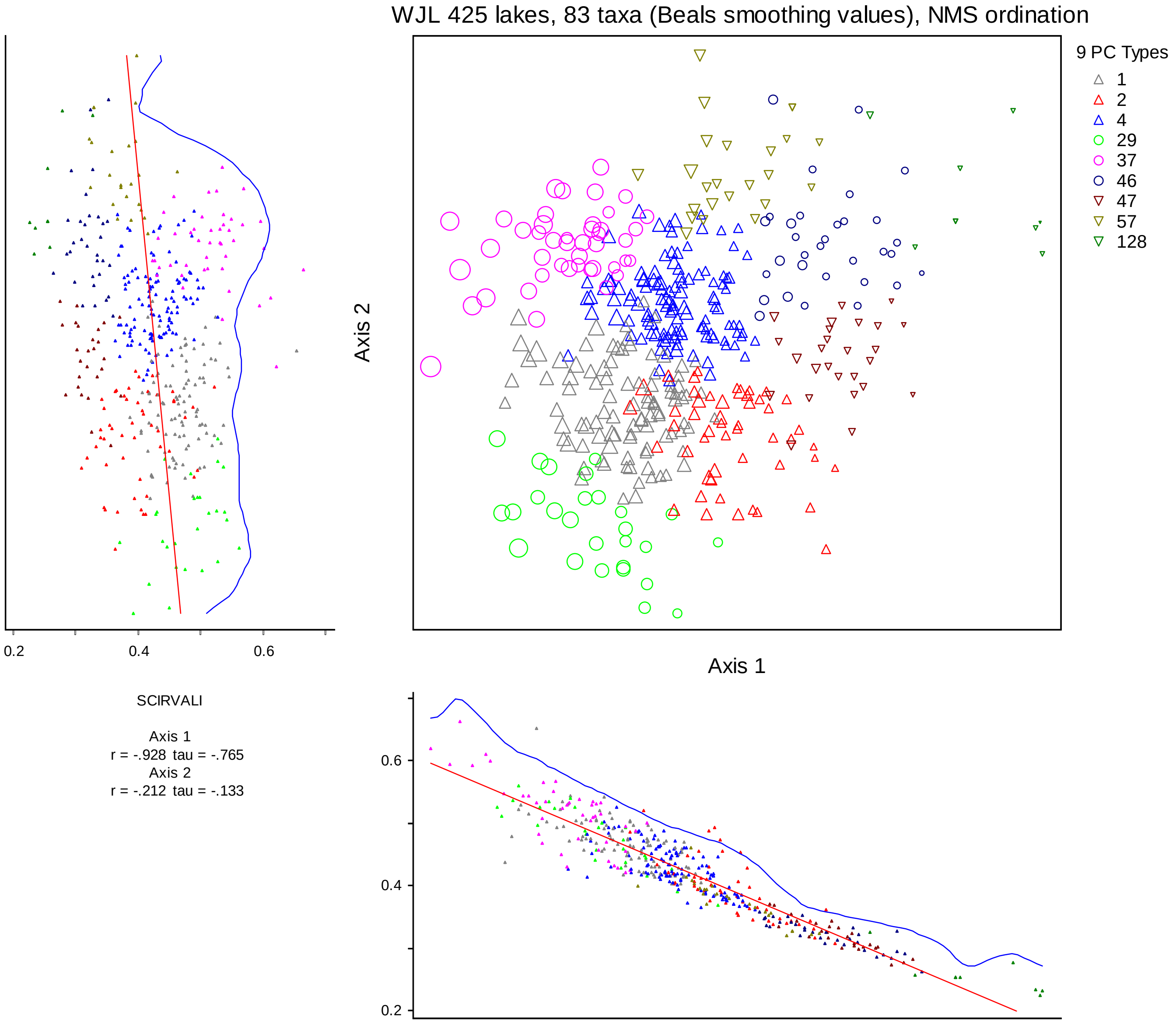


Figure WJL-19: NMS ordination of WJL dataset, overlay of favorability of lakes for **soft-stem bulrush** (*Scirpus validus* = *Schoenoplectus tabernaemontani*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the WJL dataset.

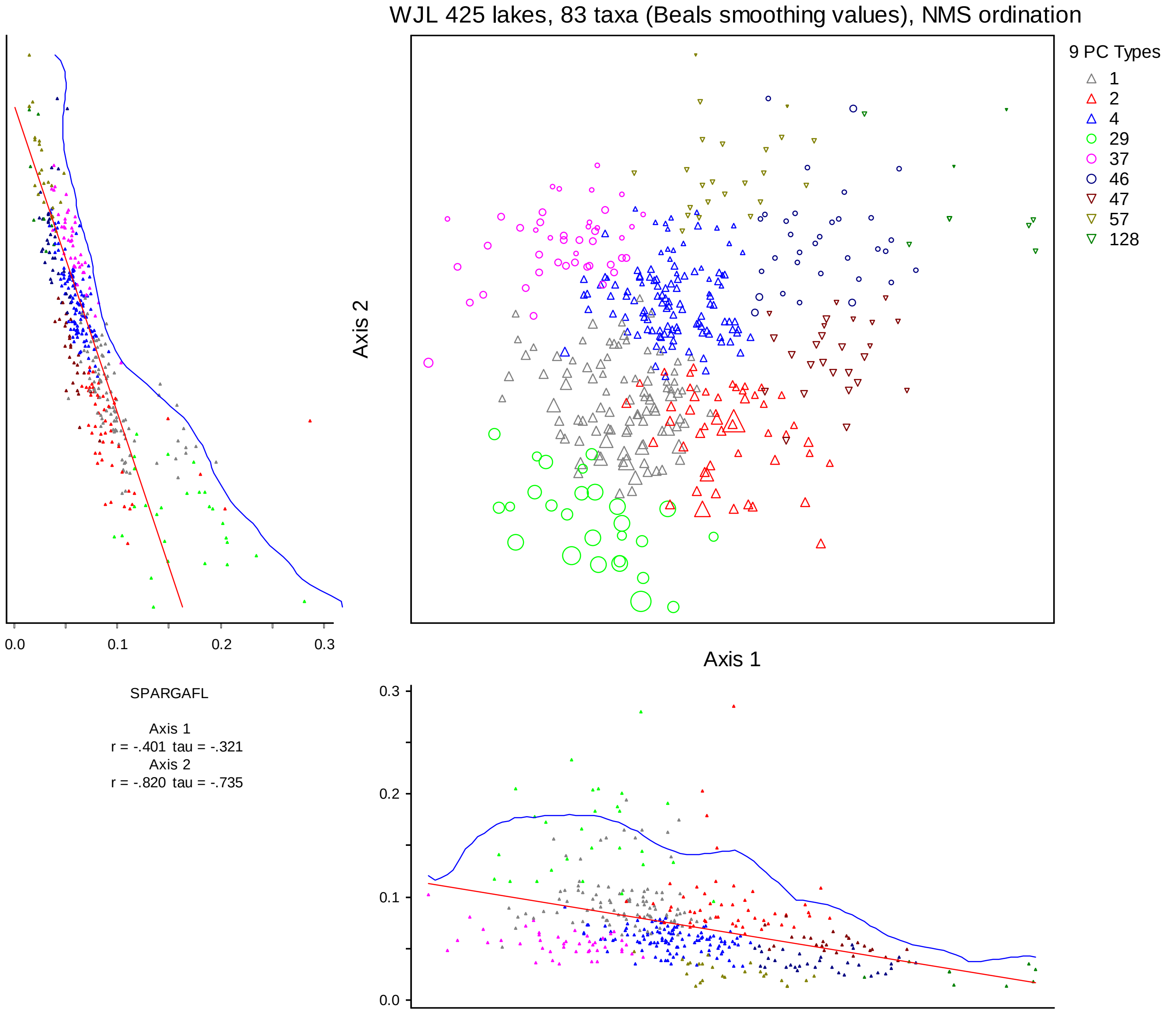


Figure WJL-20: NMS ordination of WJL dataset, overlay of favorability of lakes for **floating-leaved burreeds** (*Sparganium* spp., includes *S. angustifolium*, *S. fluctuans*, and *S. minimum*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the WJL dataset.

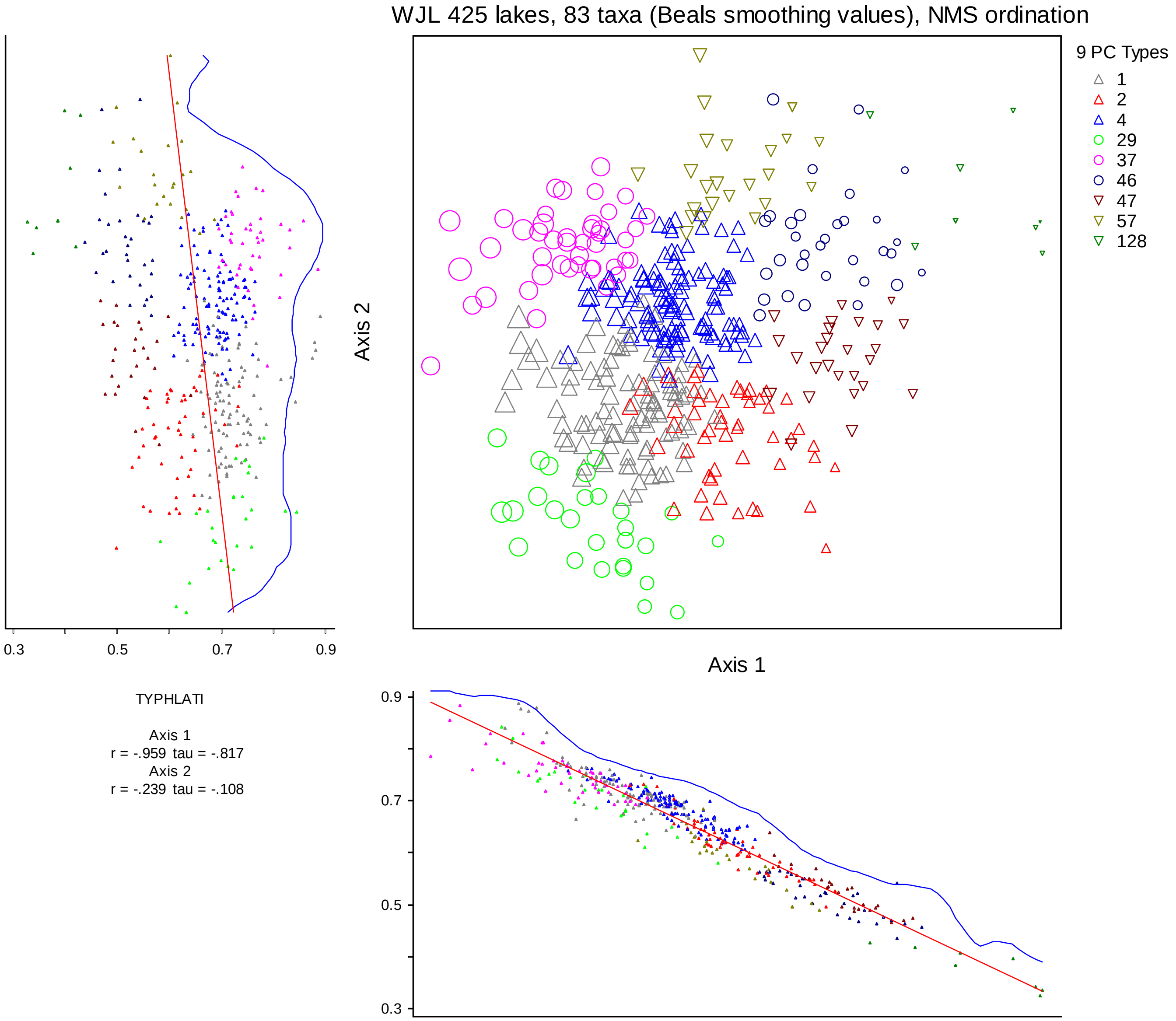


Figure WJL-21: NMS ordination of WJL dataset, overlay of favorability of lakes for **broad-leaved cattail** (*Typha latifolia*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent the 9 PC types recognized in analysis of the WJL dataset.

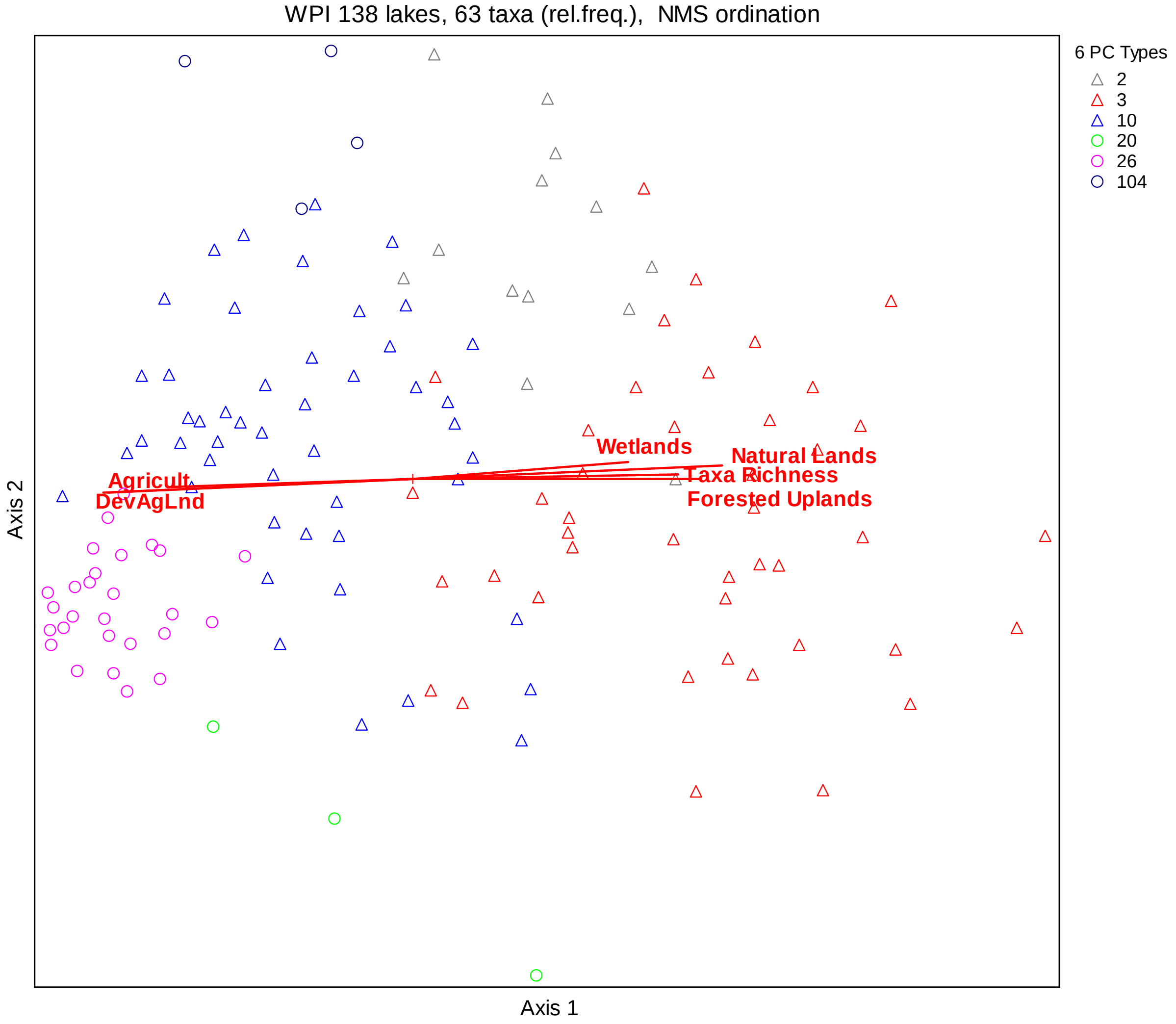


Figure WPI-1: NMS ordination of WPI dataset; joint plot showing lines of strongest correlation with quantitative environmental variables. The strongest correlations with the right end of Axis 1 are with proportion of **Natural Lands, Forested Uplands,** and **Wetlands** in the 1 km buffer around each lake, and **Taxa Richness** of each lake. The strongest correlations with the left end of Axis 1 are with proportion of **Agricultural Lands**, and **Developed/Ag Lands** in the buffer. Symbol colors and shapes represent the 6 PC types recognized in analysis of the WPI dataset.

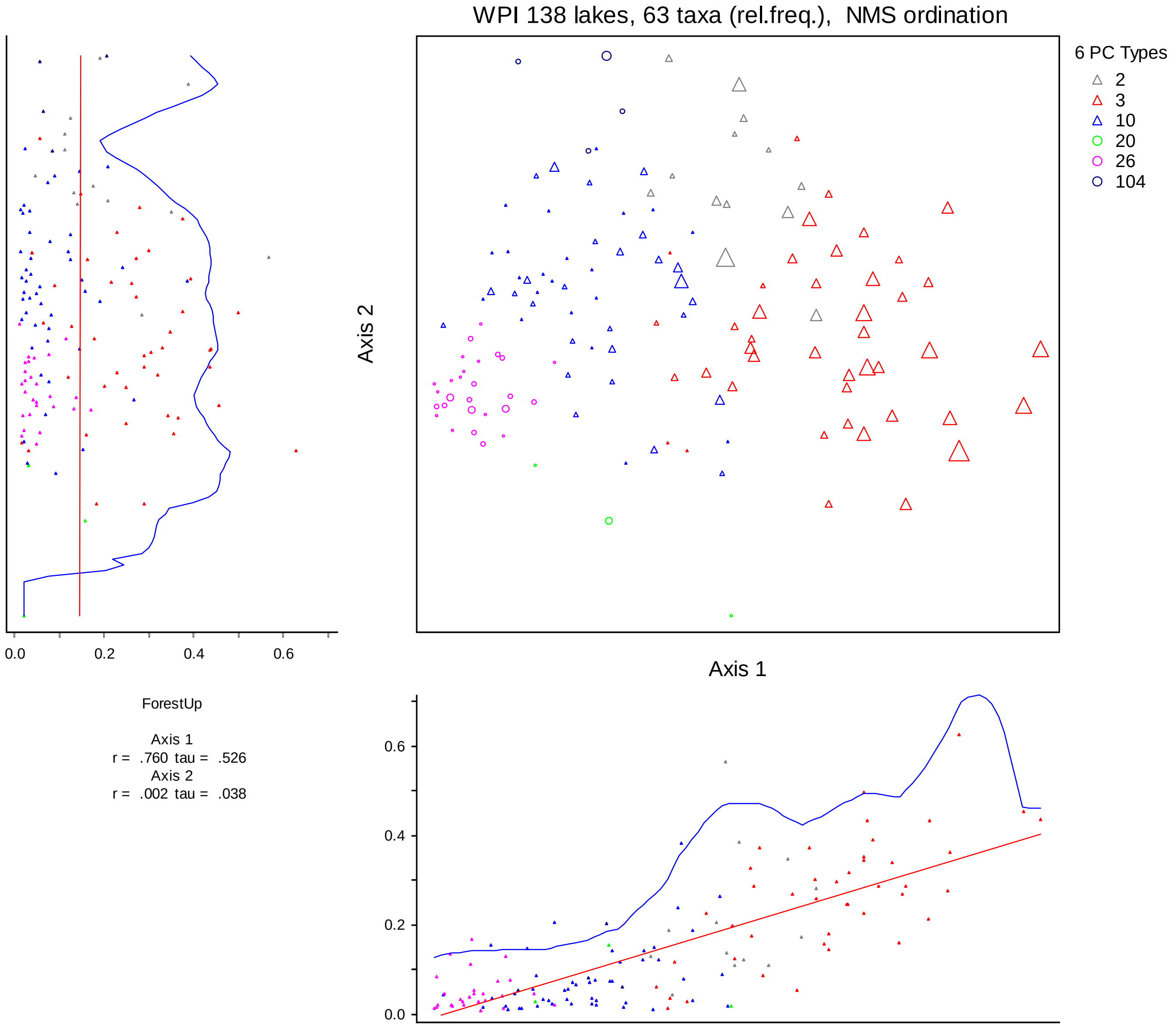


Figure WPI-2: NMS ordination of WPI dataset, overlay of proportion of 1 km buffer around each lake that is **Forested Uplands**. Larger symbols represent lakes in landscapes with higher proportions of forest in the buffer. Symbol colors and shapes represent the 6 PC types recognized in analysis of the WPI dataset.

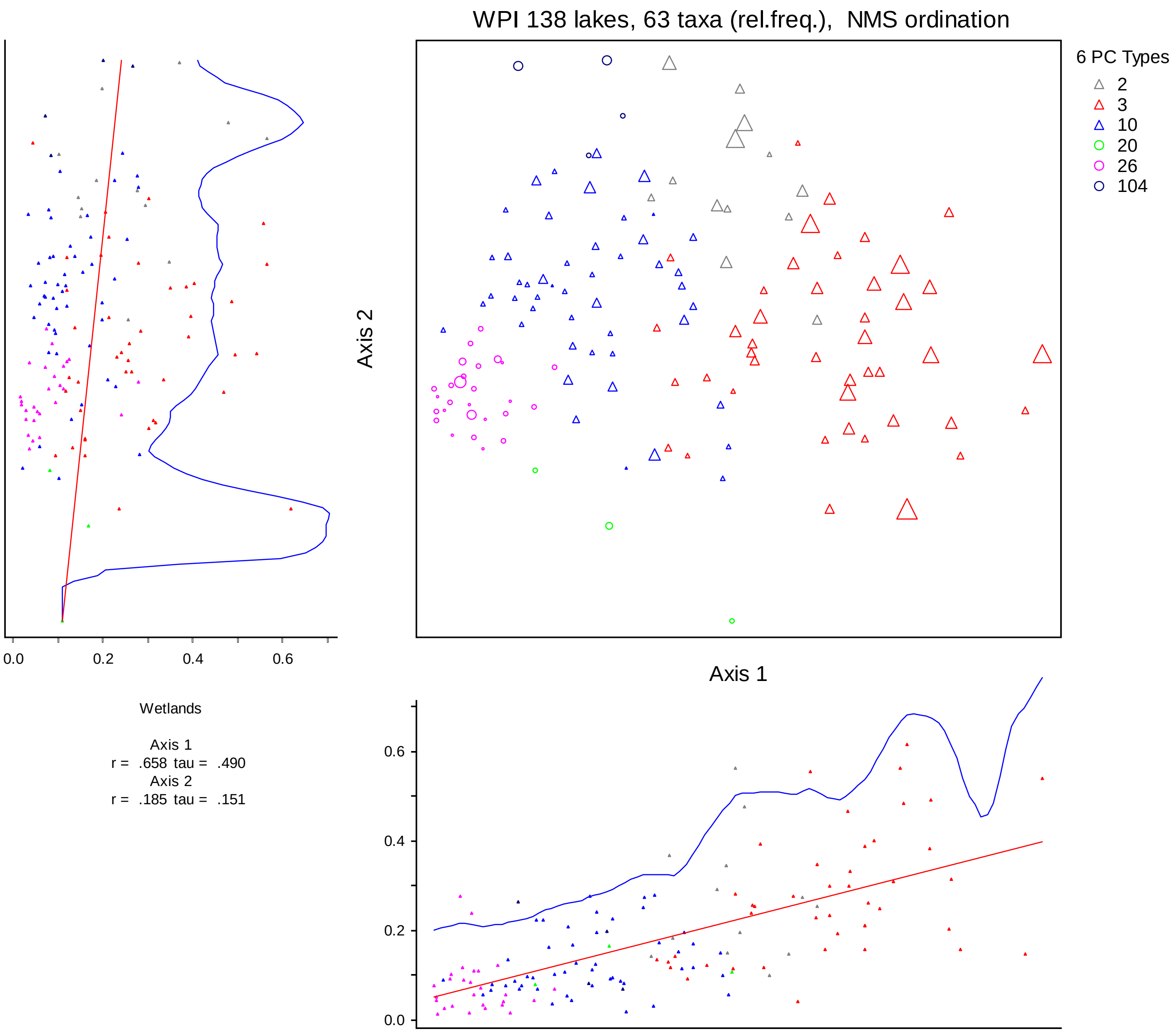


Figure WPI-3: NMS ordination of WPI dataset, overlay of proportion of 1 km buffer around each lake that is **Wetlands**. Larger symbols represent lakes in landscapes with higher proportions of developed/ag lands in the buffer. Symbol colors and shapes represent the 6 PC types recognized in analysis of the WPI dataset.

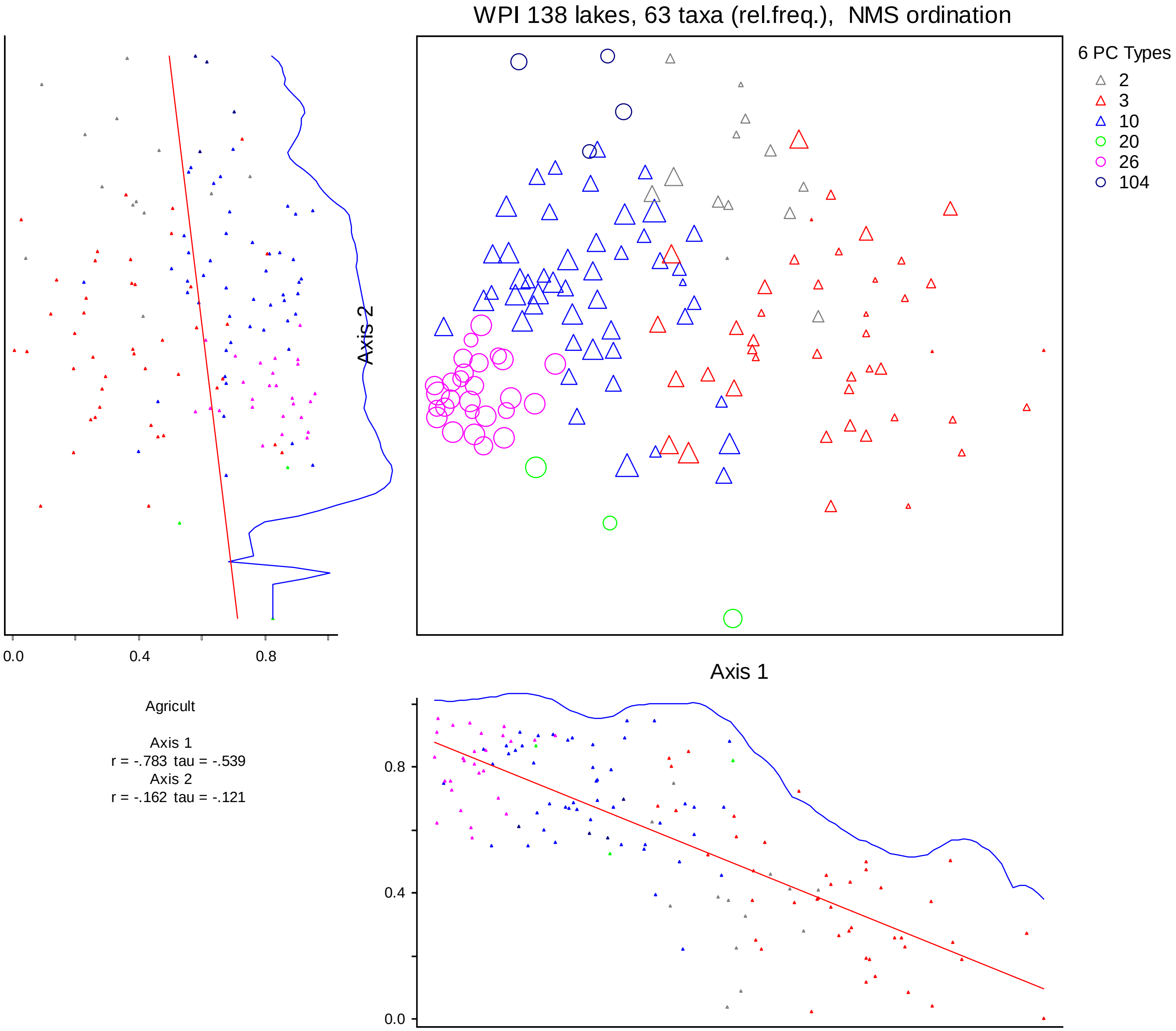


Figure WPI-4: NMS ordination of WPI dataset, overlay of proportion of 1 km buffer around each lake that is **Agriculture Lands**. Larger symbols represent lakes in landscapes with higher proportions of developed/ag lands in the buffer. Symbol colors and shapes represent the 6 PC types recognized in analysis of the WPI dataset.

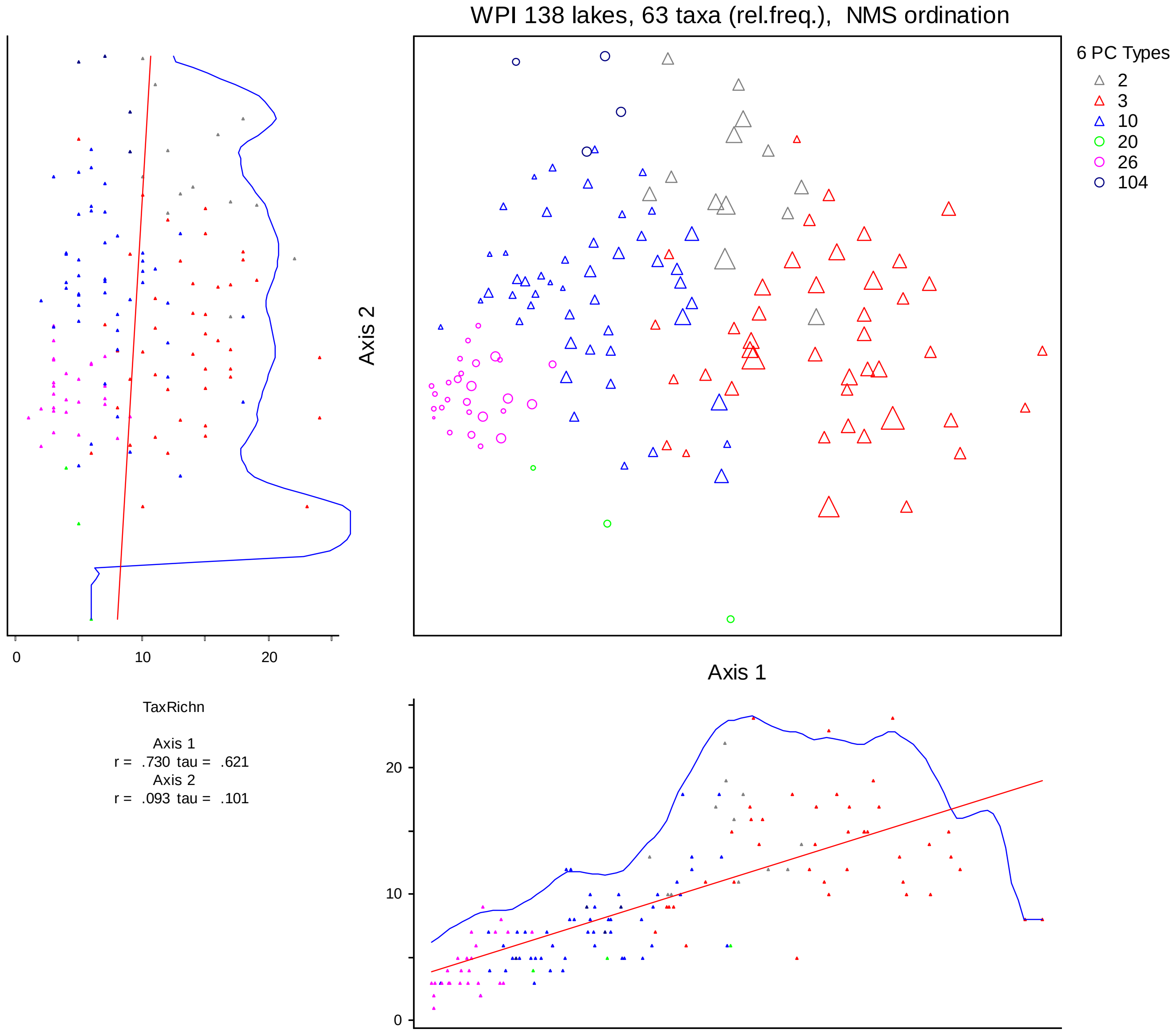


Figure WPI-5: NMS ordination of WPI dataset, overlay of **Taxa Richness** of each lake. Larger symbols represent lakes with higher numbers of taxa recorded for each lake. Symbol colors and shapes represent the 6 PC types recognized in analysis of the WPI dataset.

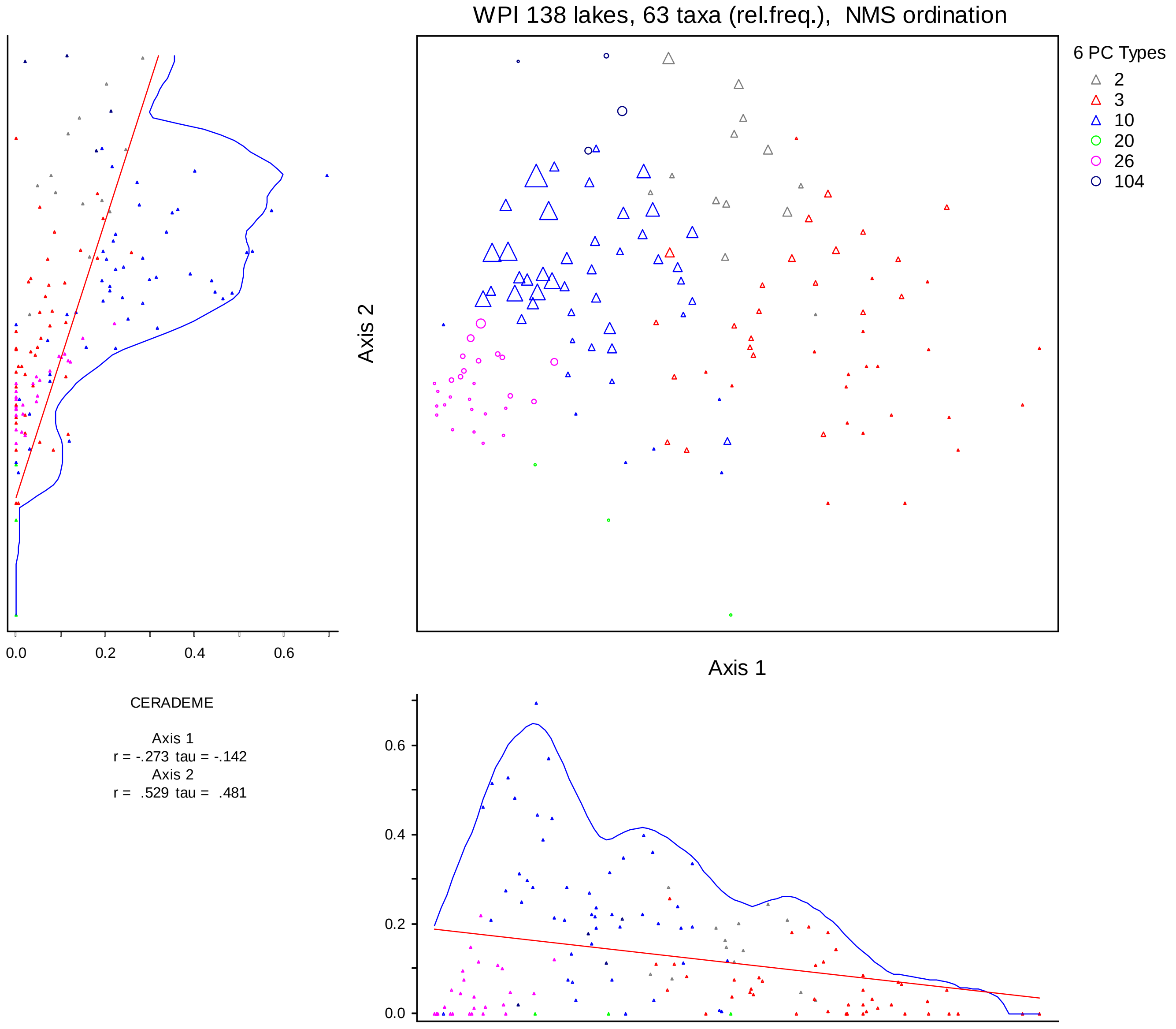


Figure WPI-6: NMS ordination of WPI dataset, overlay of favorability of lakes for **coontail** (*Ceratophyllum demersum*). Large symbols represent lakes with higher relative frequency values for the taxon. Symbol colors and shapes represent the 6 PC types recognized in analysis of the WPI dataset.

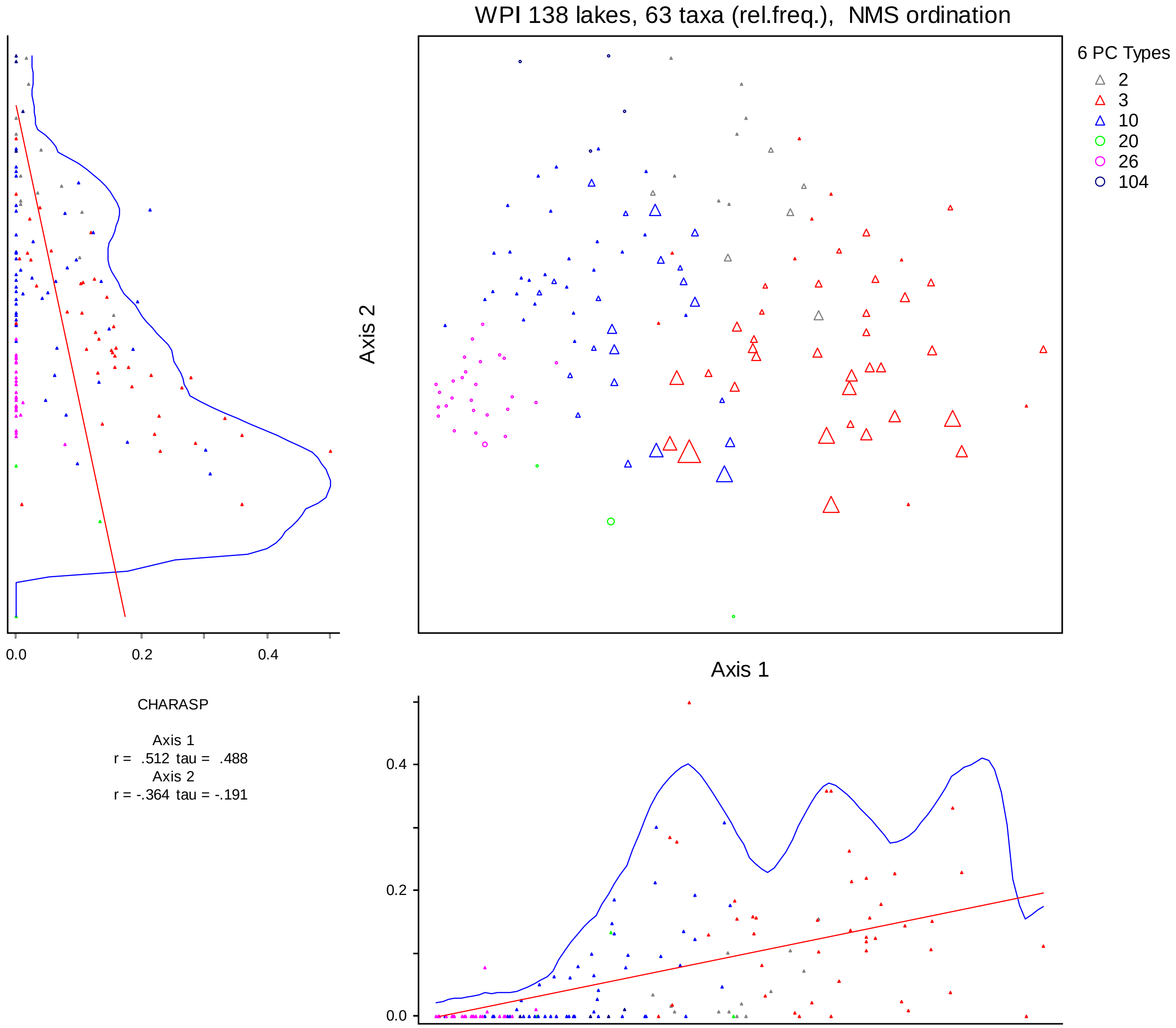


Figure WPI-7: NMS ordination of WPI dataset, overlay of favorability of lakes for **muskgrasses** (*Chara* spp.). Large symbols represent lakes with higher relative frequency values for the taxon. Symbol colors and shapes represent the 6 PC types recognized in analysis of the WPI dataset.

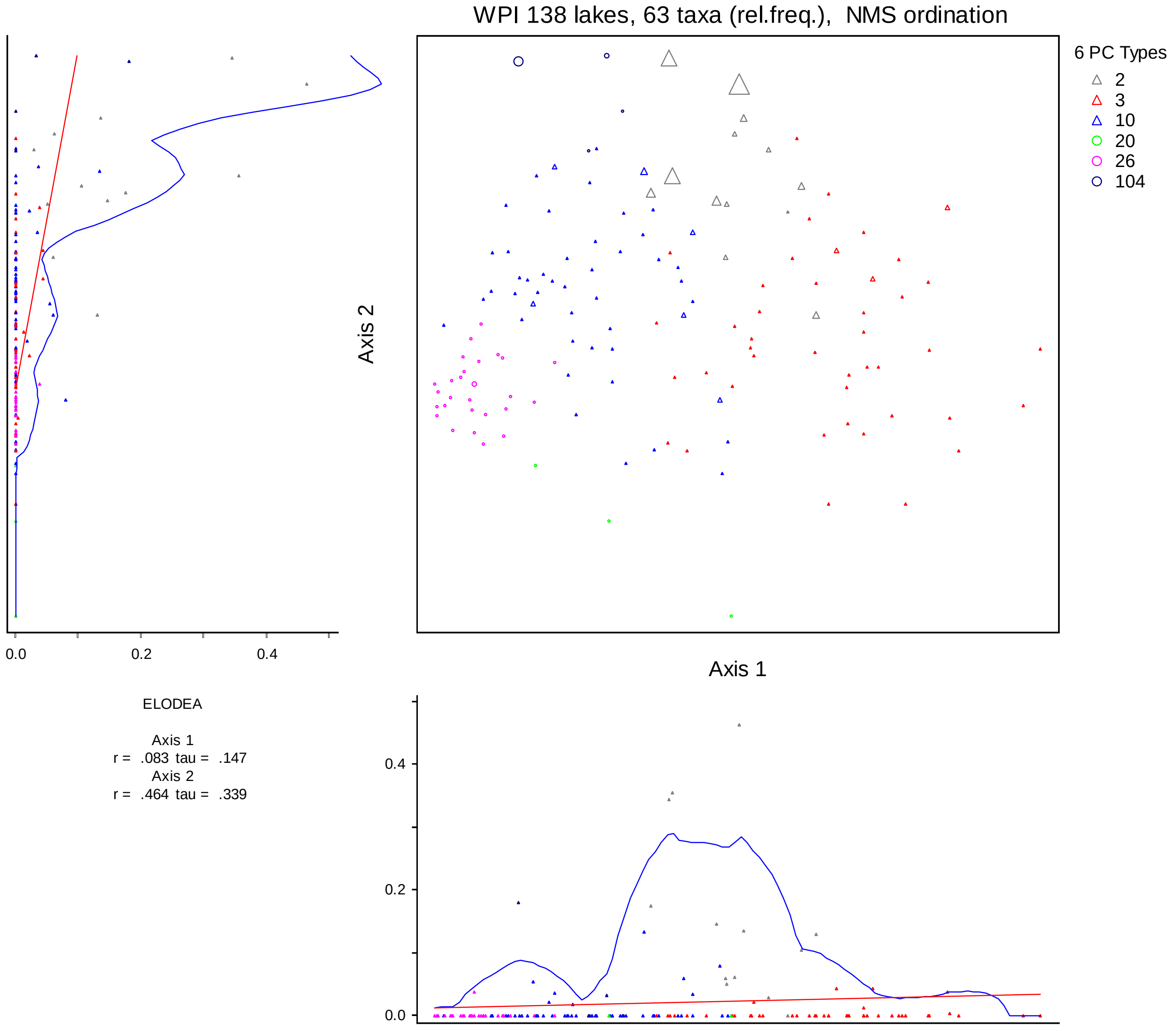


Figure WPI-8: NMS ordination of WPI dataset, overlay of favorability of lakes for **waterweeds** (*Elodea* spp., includes *E. canadensis* and *E. nuttallii*). Large symbols represent lakes with higher relative frequency values for the taxon. Symbol colors and shapes represent the 6 PC types recognized in analysis of the WPI dataset.

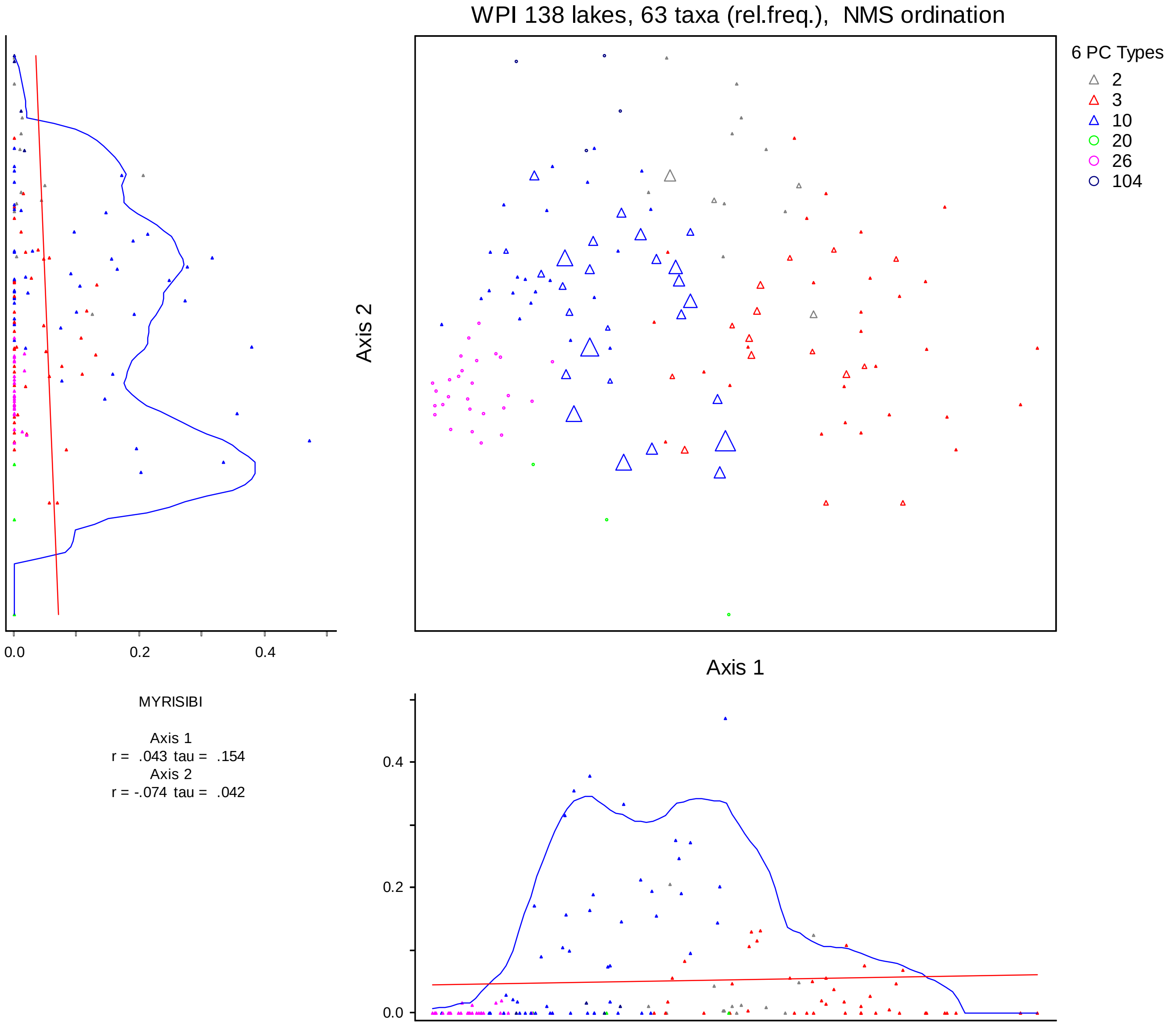


Figure WPI-9: NMS ordination of WPI dataset, overlay of favorability of lakes for **northern milfoil** (*Myriophyllum sibiricum*). Large symbols represent lakes with higher relative frequency values for the taxon. Symbol colors and shapes represent the 6 PC types recognized in analysis of the WPI dataset.

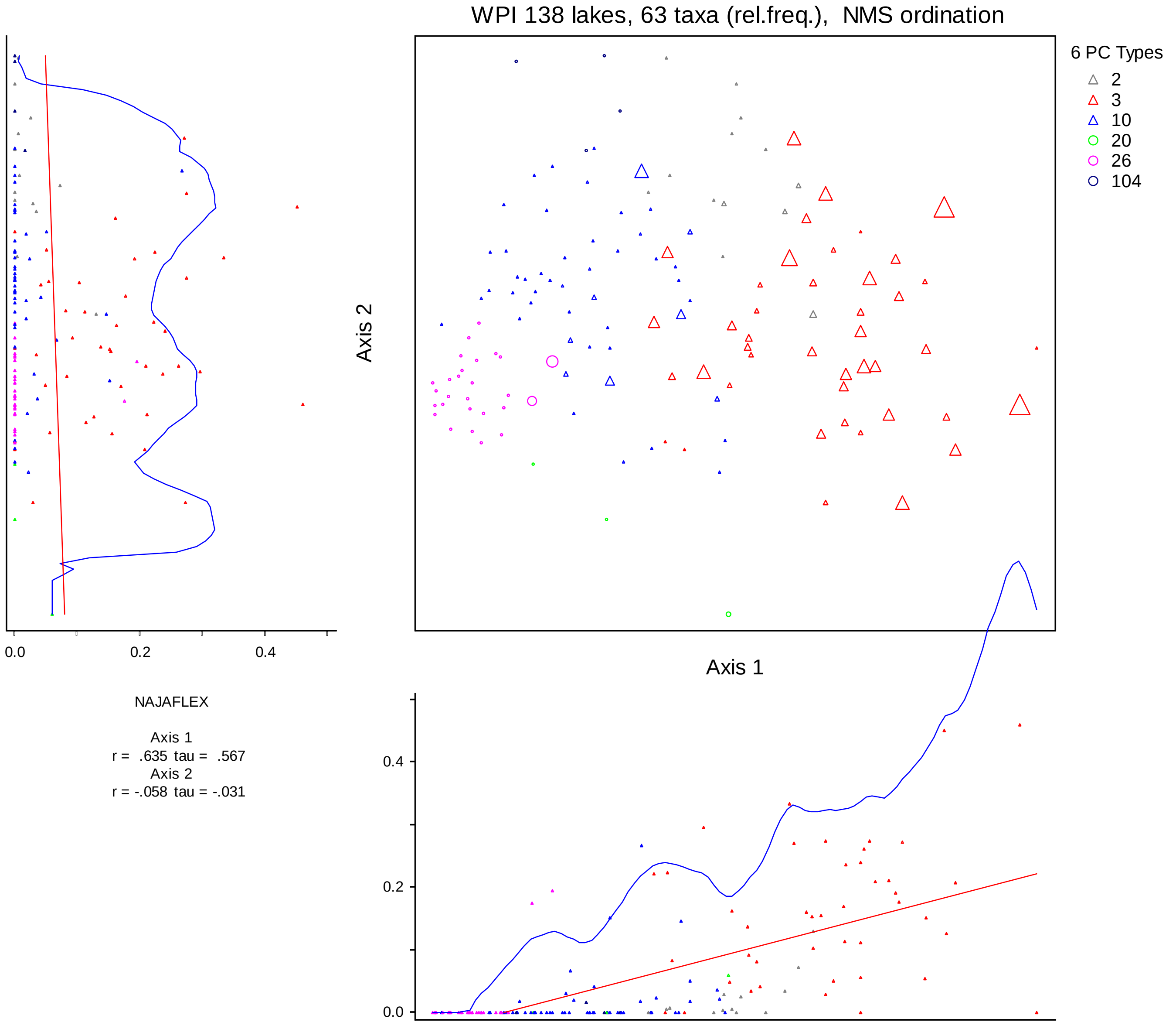


Figure WPI-10: NMS ordination of WPI dataset, overlay of favorability of lakes for **bushy pondweed** (*Najas flexilis*). Large symbols represent lakes with higher relative frequency values for the taxon. Symbol colors and shapes represent the 6 PC types recognized in analysis of the WPI dataset.

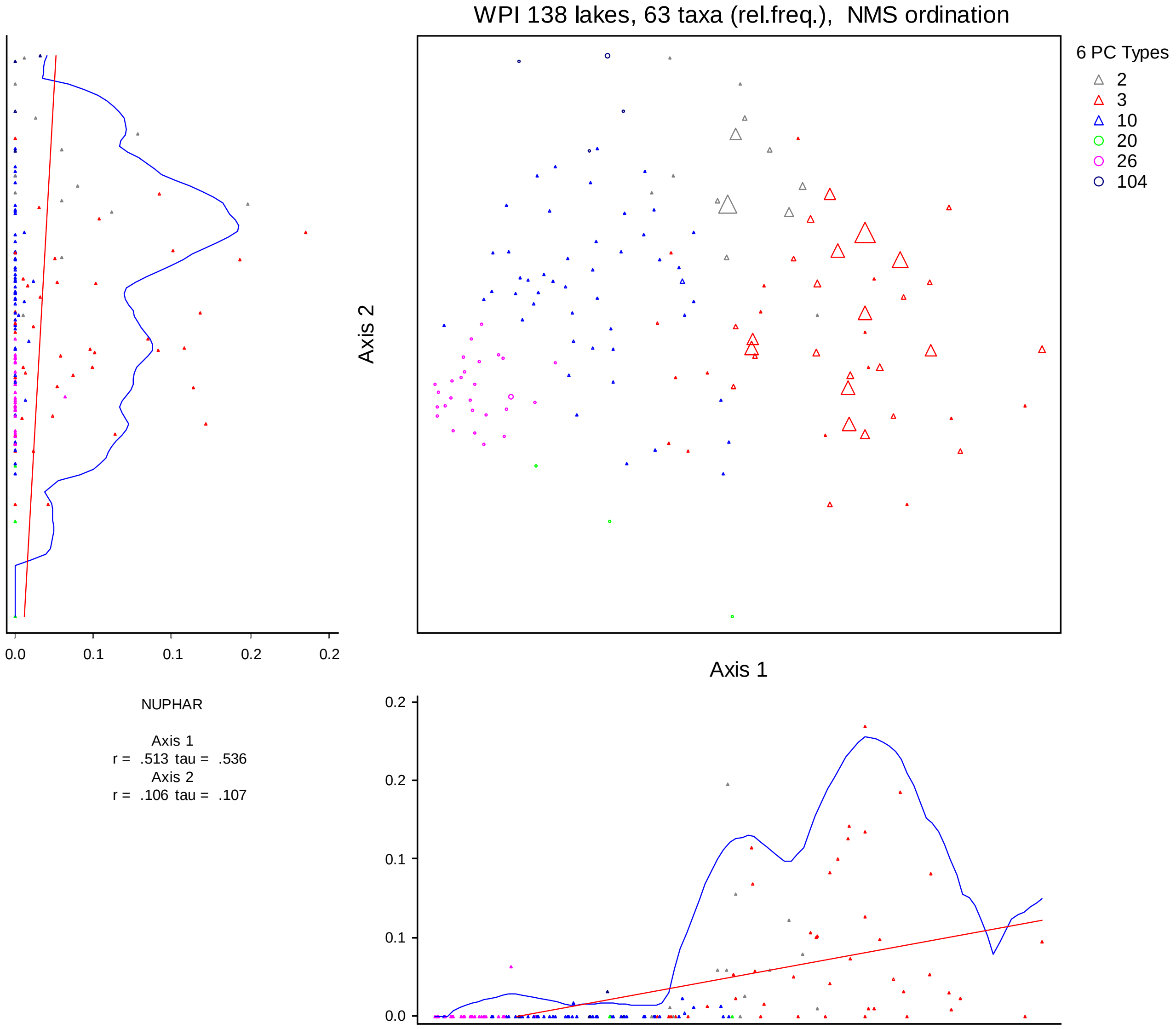


Figure WPI-11: NMS ordination of WPI dataset, overlay of favorability of lakes for **yellow waterlilies** (*Nuphar* spp., includes *N. lutea* ssp. *pumila*, ssp. *rubrodisca*, ssp. *variegata*, *N. microphylla*, *N. rubrodisca*, and *N. variegata*). Large symbols represent lakes with higher relative frequency values for the taxon. Symbol colors and shapes represent the 6 PC types recognized in analysis of the WPI dataset.

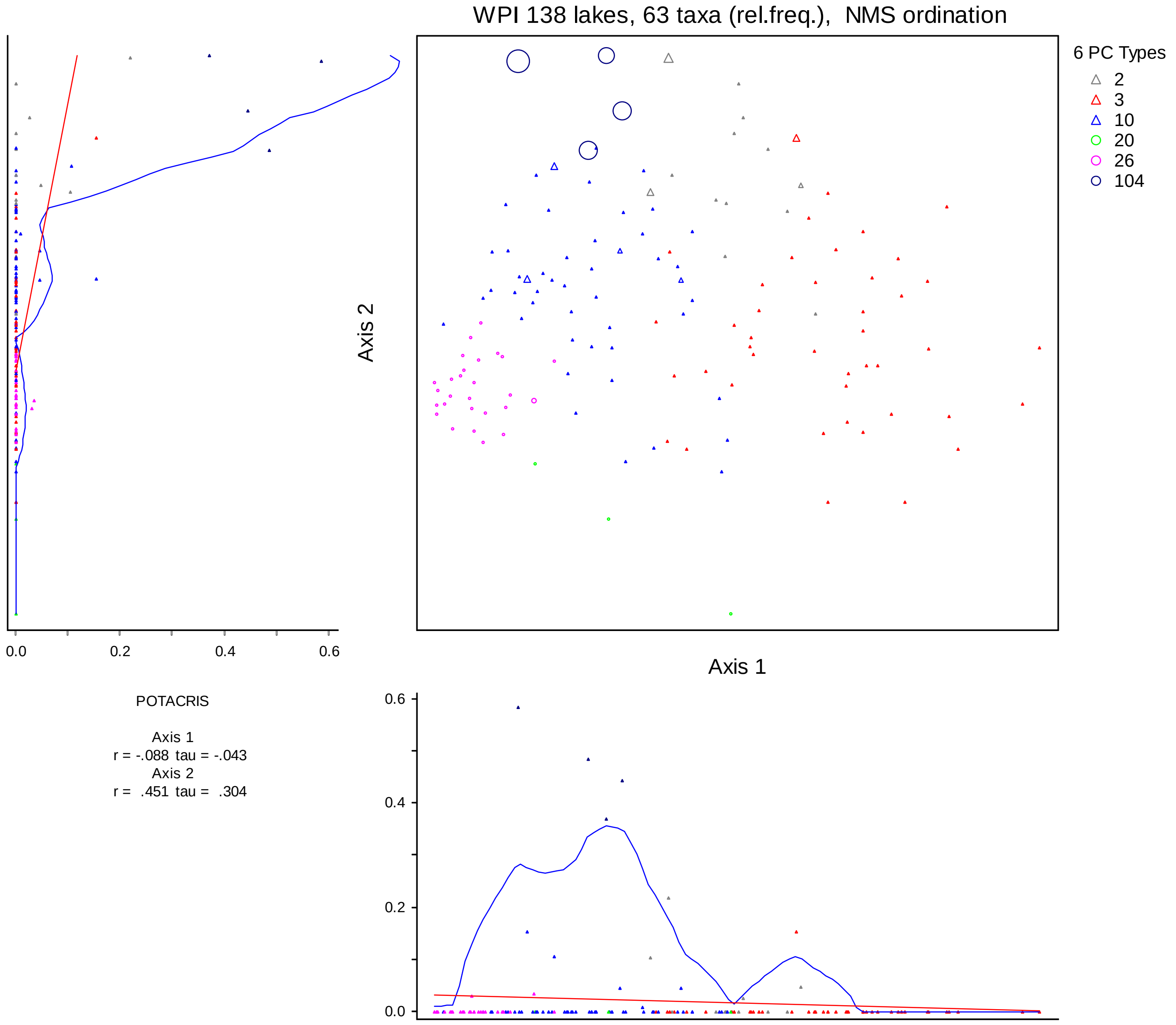


Figure WPI-12: NMS ordination of WPI dataset, overlay of favorability of lakes for exotic **curly-leaf pondweed** (*Potamogeton crispus*). Large symbols represent lakes with higher relative frequency values for the taxon. Symbol colors and shapes represent the 6 PC types recognized in analysis of the WPI dataset.

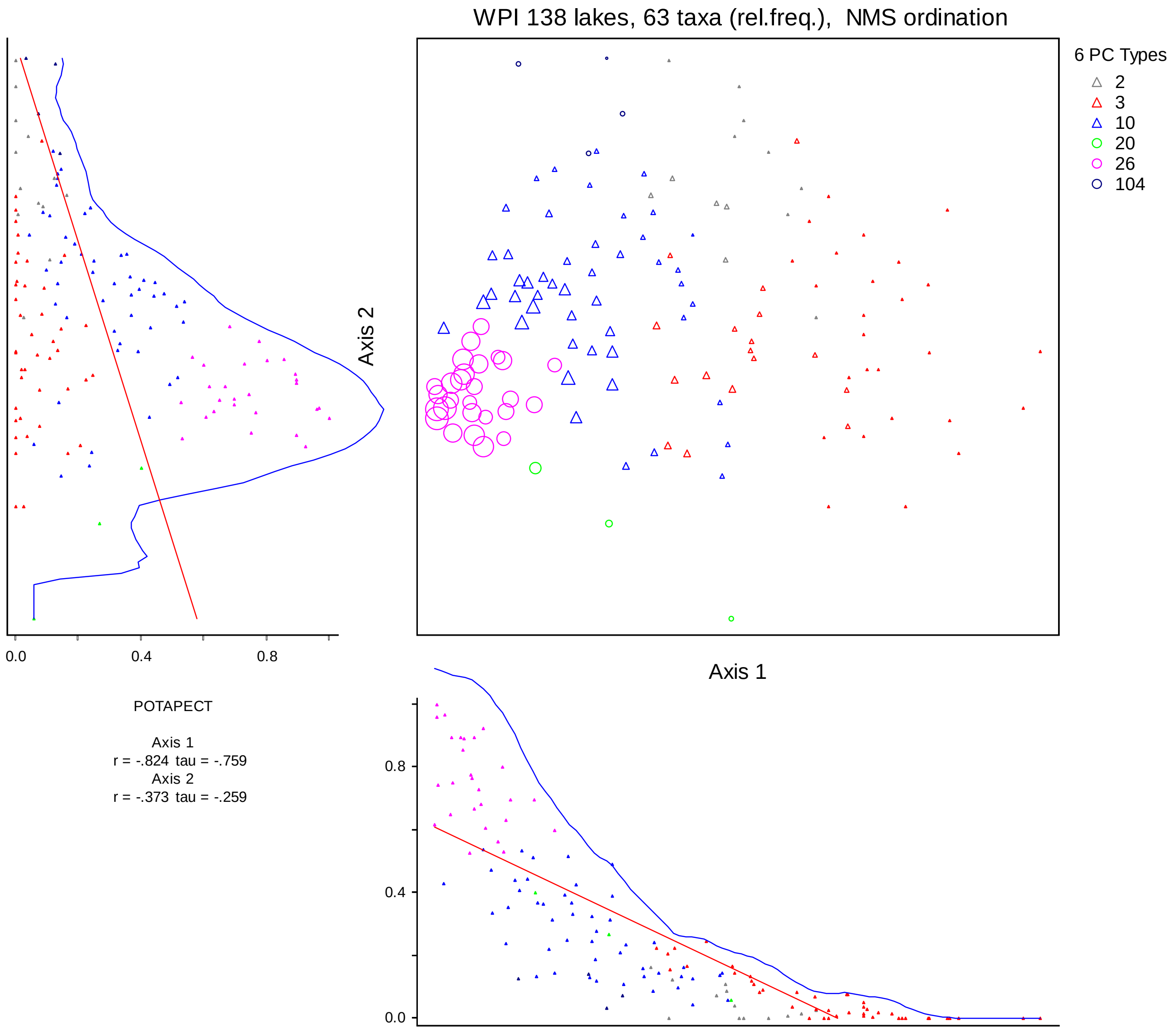


Figure WPI-13: NMS ordination of WPI dataset, overlay of favorability of lakes for **sago pondweed** (*Potamogeton pectinatus*). Large symbols represent lakes with higher relative frequency values for the taxon. Symbol colors and shapes represent the 6 PC types recognized in analysis of the WPI dataset.

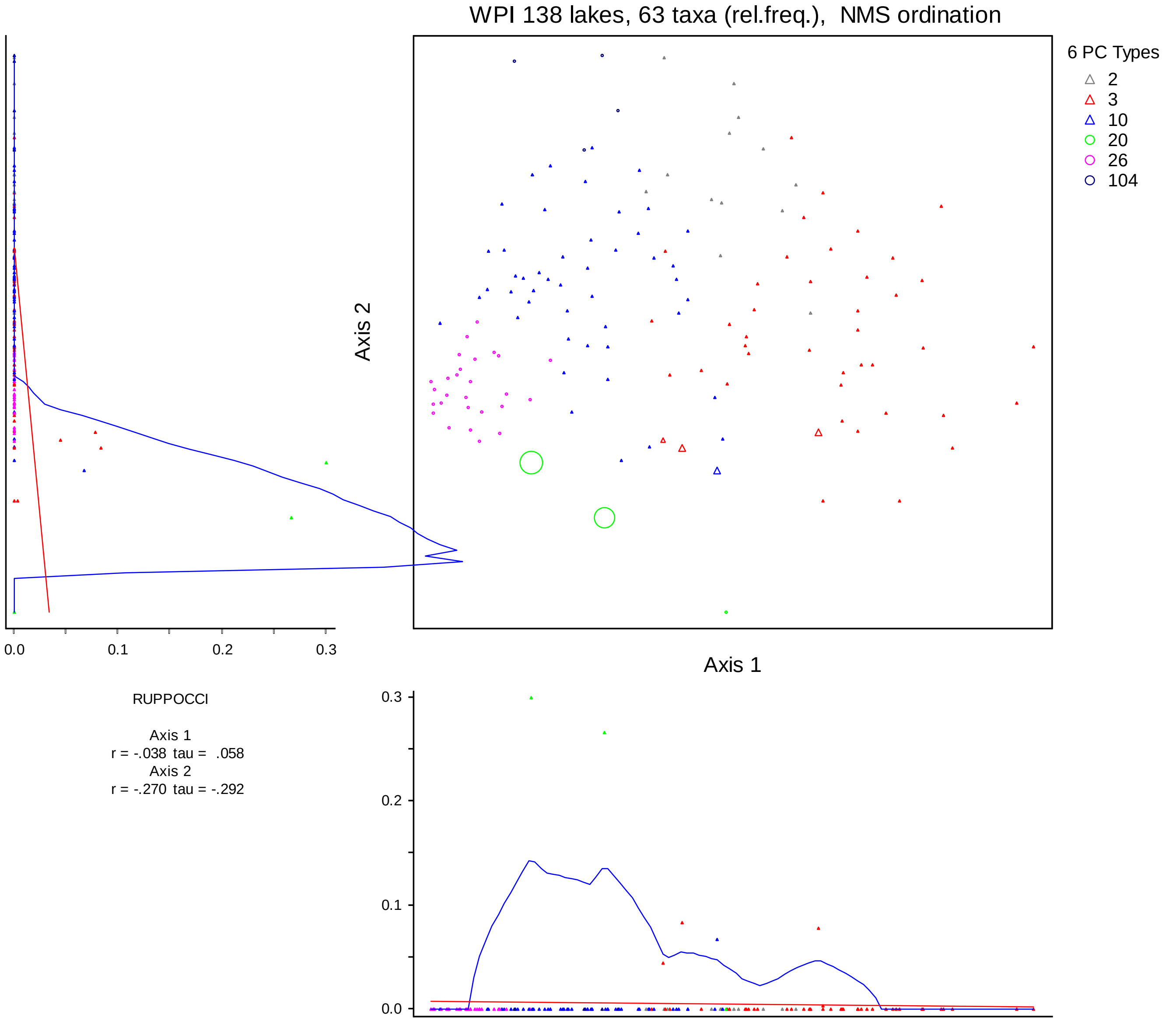


Figure WPI-14: NMS ordination of WPI dataset, overlay of favorability of lakes for **widgeon grass** (*Ruppia occidentalis*). Large symbols represent lakes with higher relative frequency values for the taxon. Symbol colors and shapes represent the 6 PC types recognized in analysis of the WPI dataset.

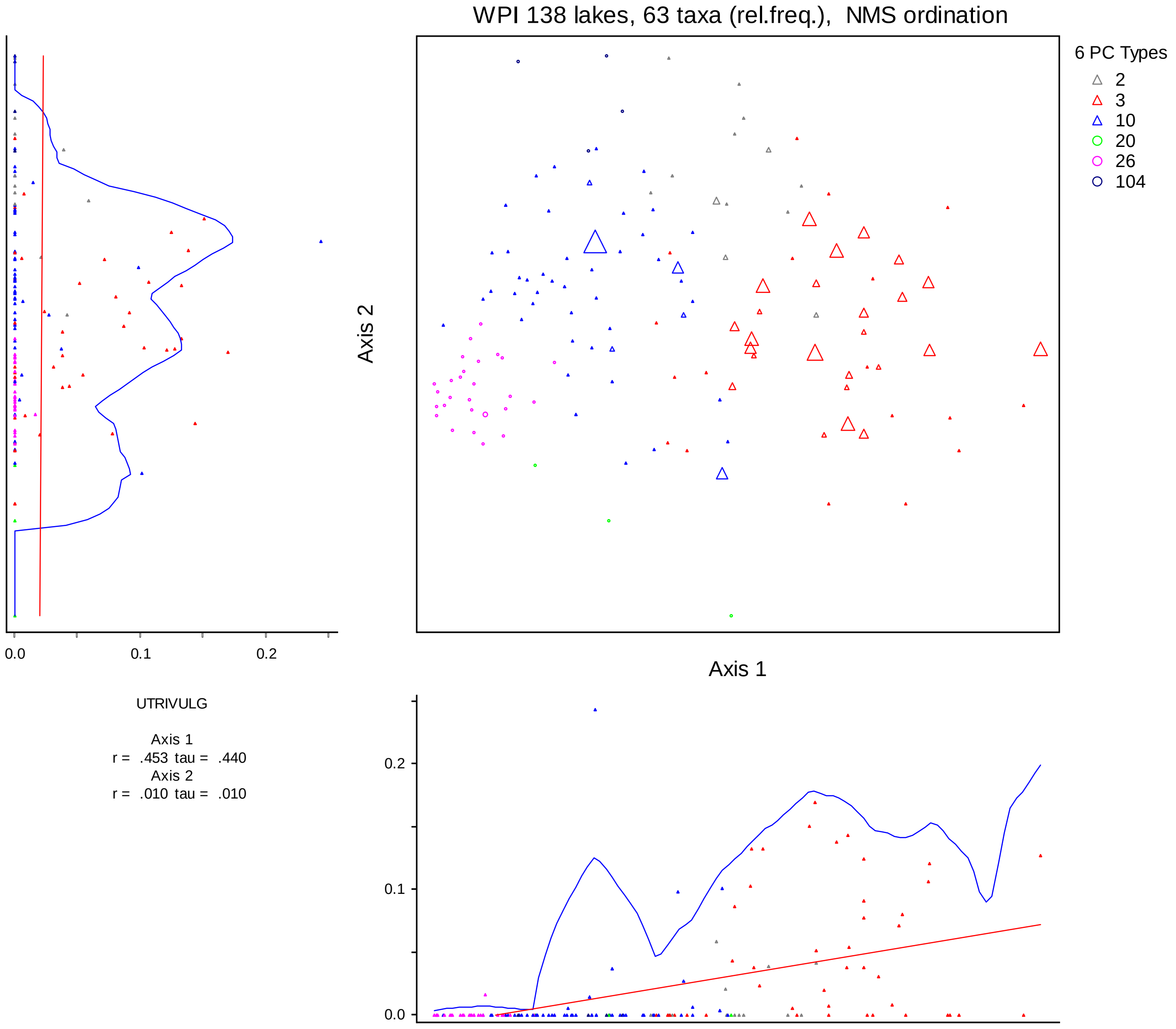


Figure WPI-15: NMS ordination of WPI dataset, overlay of favorability of lakes for **common bladderwort** (*Utricularia vulgaris* = *U. macrorhiza*). Large symbols represent lakes with higher relative frequency values for the taxon. Symbol colors and shapes represent the 6 PC types recognized in analysis of the WPI dataset.

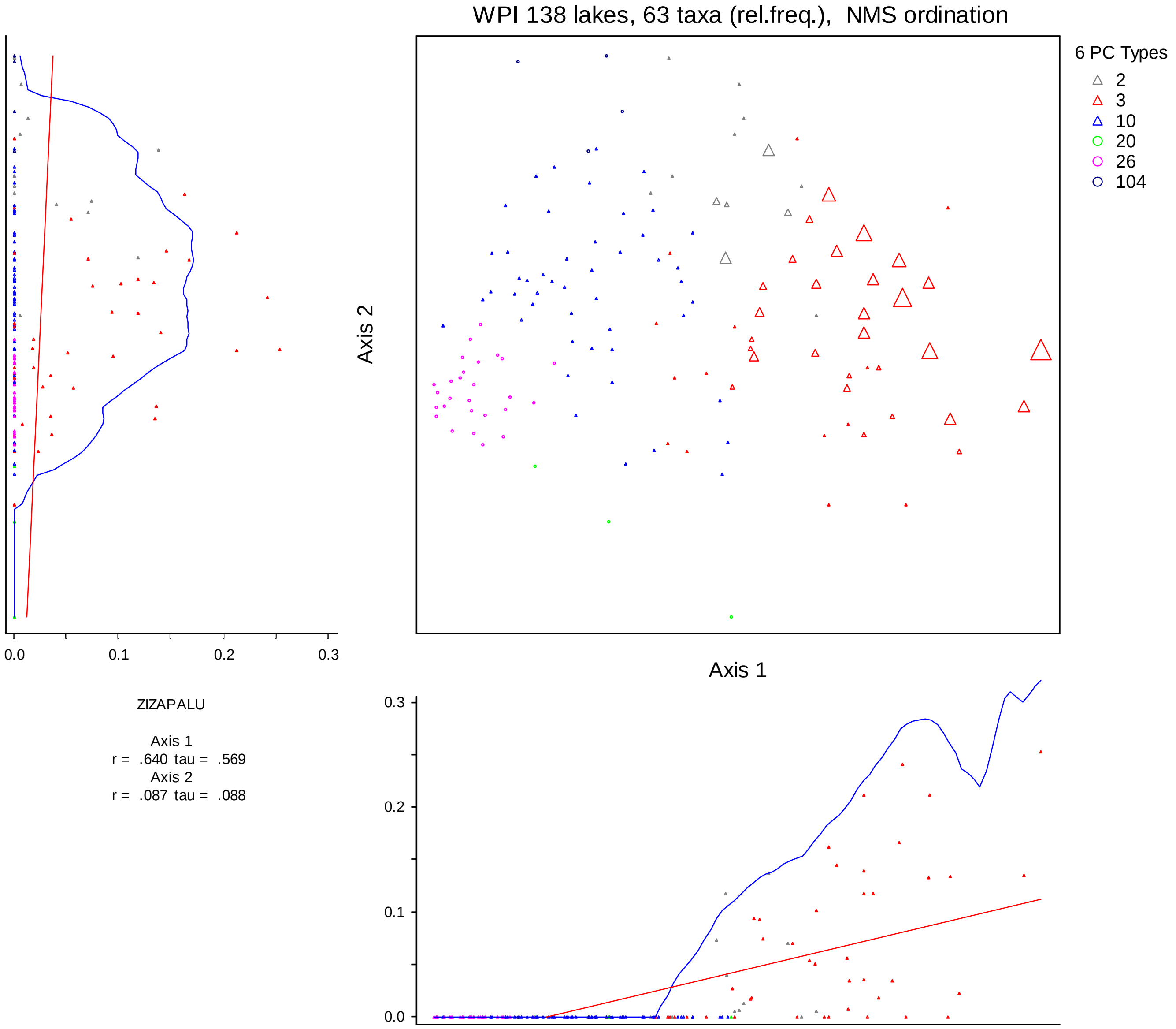


Figure WPI-16: NMS ordination of WPI dataset, overlay of favorability of lakes for **wild rice** (*Zizania palustris* and *Z. aquatica*). Large symbols represent lakes with higher relative frequency values for the taxon. Symbol colors and shapes represent the 6 PC types recognized in analysis of the WPI dataset.

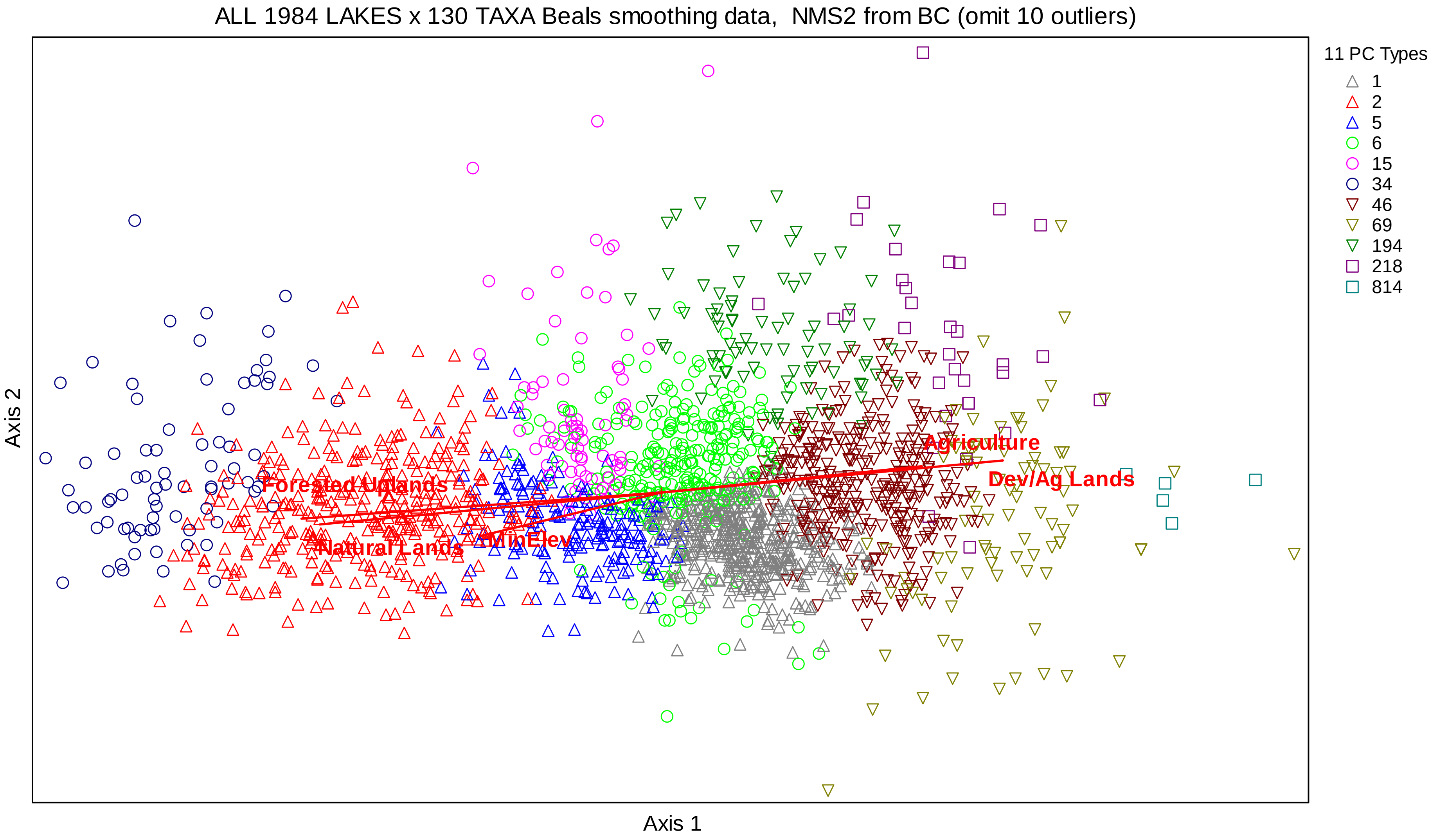


Figure ALLKS-1: NMS ordination of ALLKS dataset; joint plot showing lines of strongest correlation with quantitative environmental variables. The strongest correlations with the left end of Axis 1 are with proportion of **Forested Lands** in the 1 km buffer around each lake, proportion of **Natural Lands** in the buffer, and **Minimum Elevation** of buffer (approximates elevation of the lake). The strongest correlations with the right end of Axis 1 are with proportion of **Agriculture Lands** and **Developed/Ag Lands** in the buffer. Symbol colors and shapes represent the 11 PC types recognized in analysis of the ALLKS dataset.

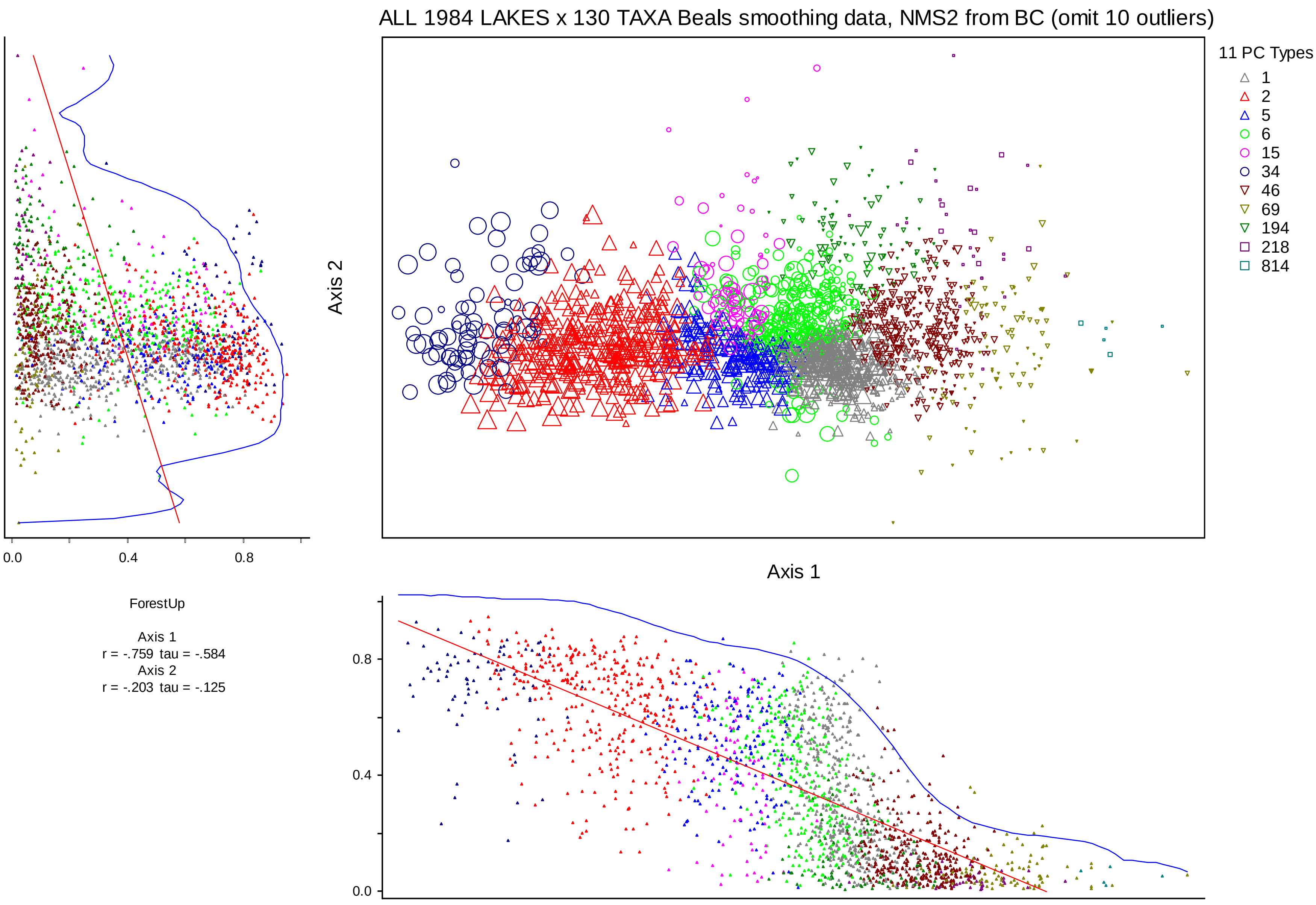
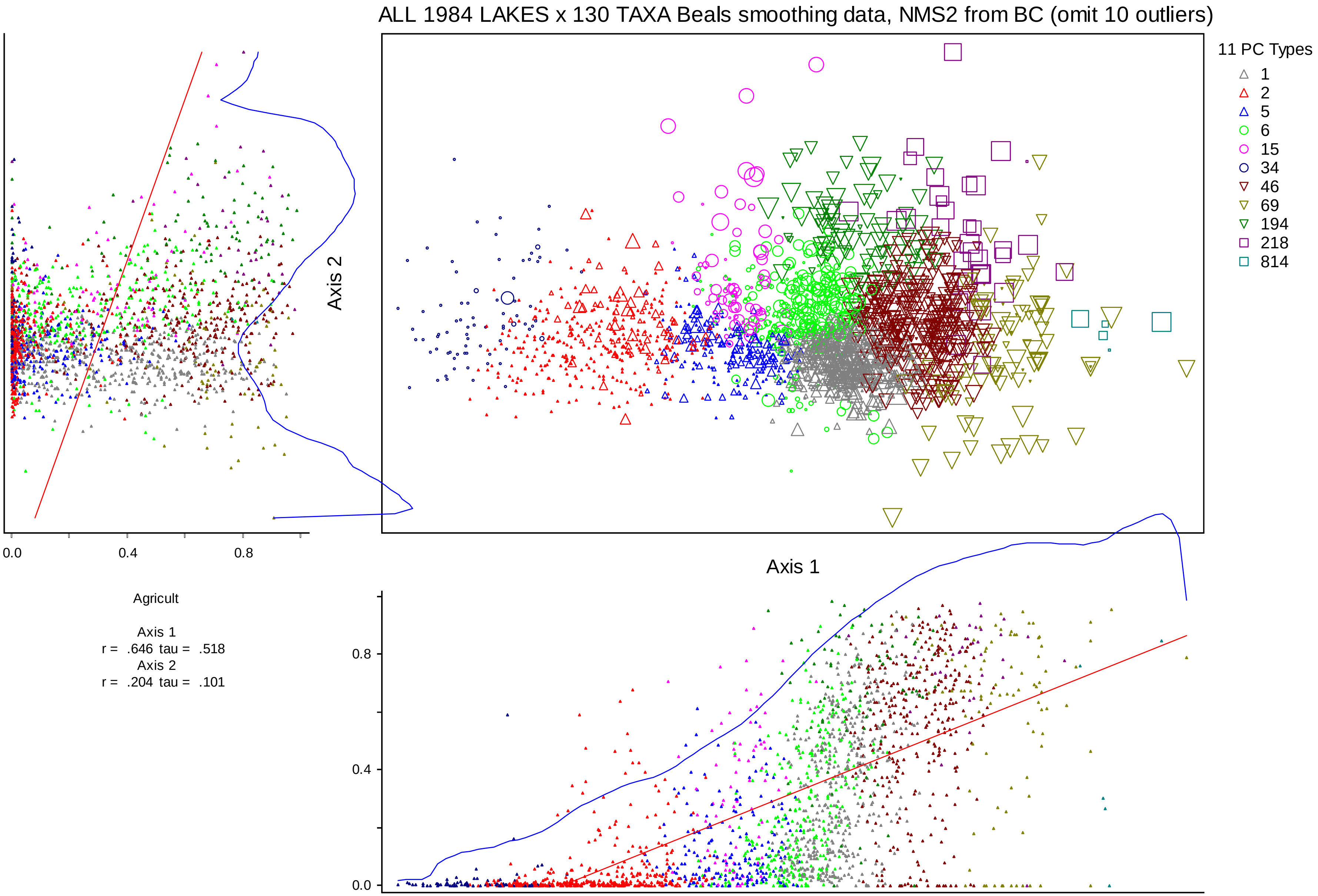


Figure ALLKS-2: NMS ordination of ALLKS dataset, overlay of proportion of 1 km buffer around each lake that is **Forested Uplands**. Large symbols represent lakes in landscapes with higher proportions of forest in the buffer. Symbol colors/shapes represent ALLKS PC Types.



ALL 1984 LAKES x 130 TAXA Beals smoothing data, NMS2 from BC (omit 10 outliers)

Figure ALLKS-3: NMS ordination of ALLKS dataset, overlay of proportion of 1 km buffer of each lake that is **Agriculture Lands**. Large symbols represent lakes in landscapes with higher proportions of agriculture in buffer. Symbol colors/shapes represent ALLKS PC Types.

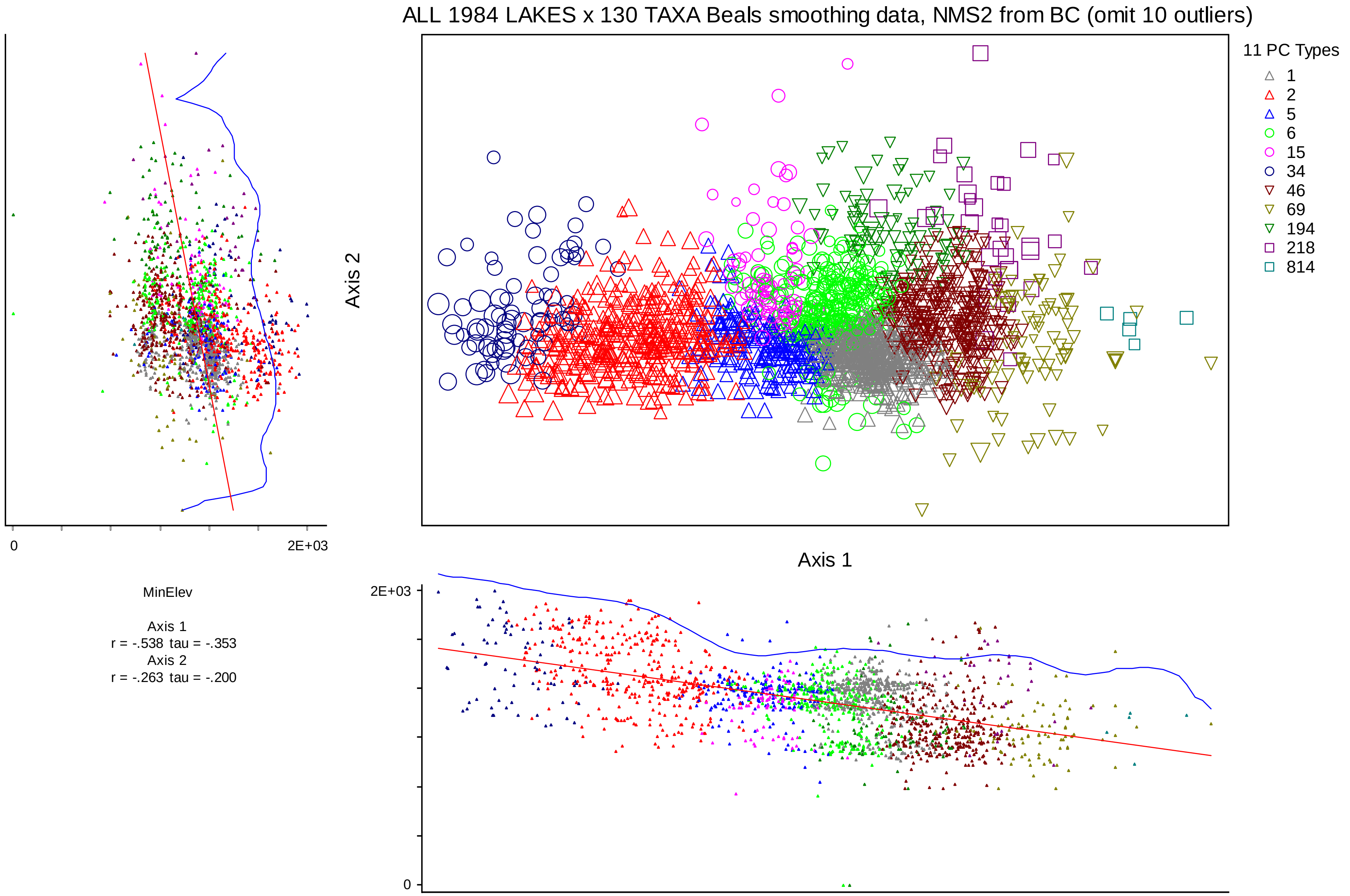


Figure ALLKS-4: NMS ordination of ALLKS dataset, overlay of **Minimum Elevation** of surrounding 1 km buffer area. Large symbols represent lakes in landscapes with higher Minimum Elevation in the buffer. Symbol colors/shapes represent ALLKS PC Types.

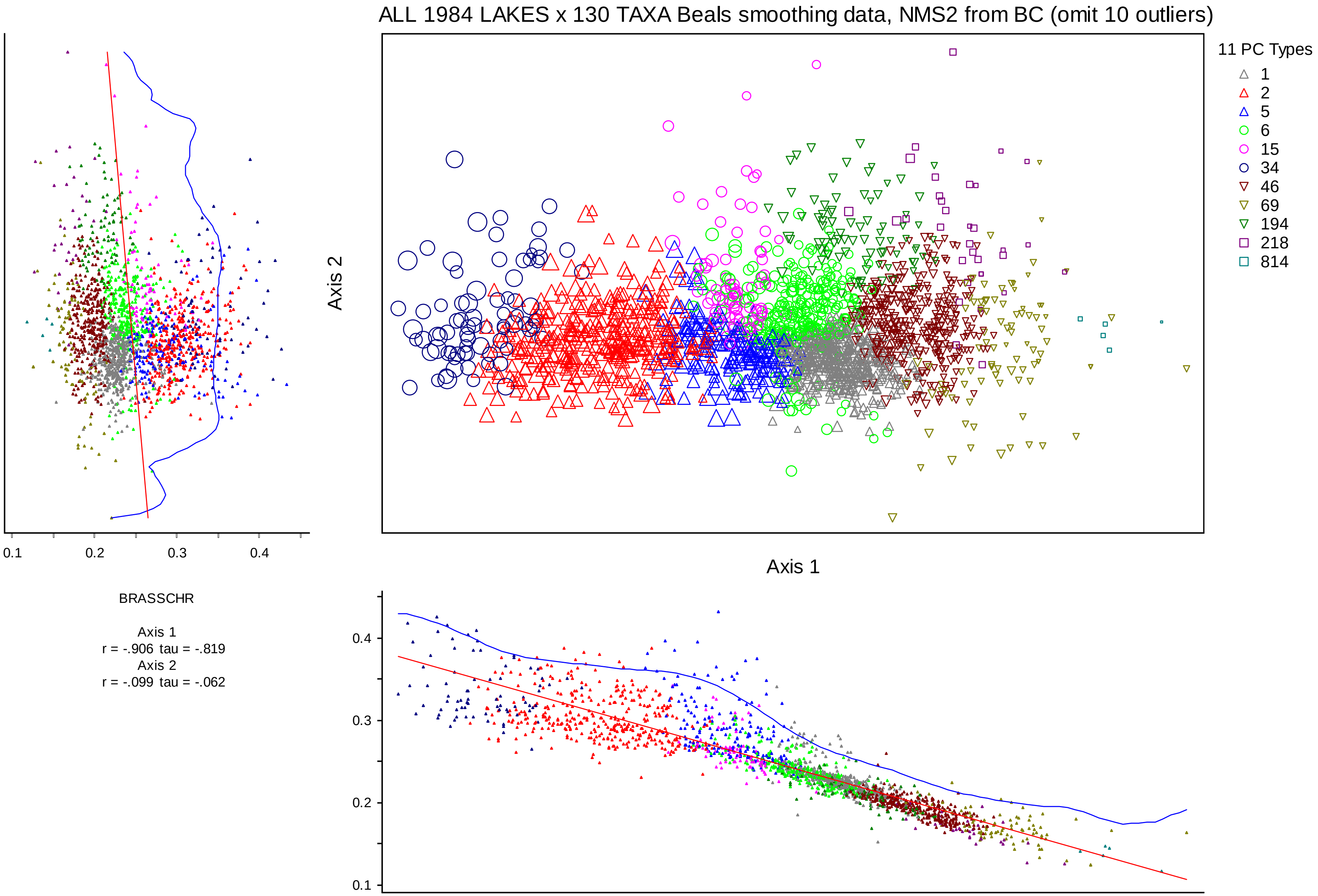


Figure ALLKS-5: NMS ordination of ALLKS dataset, overlay of favorability of lakes for **water shield** (*Brasenia schreberi*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent 11 ALLKS PC Types.

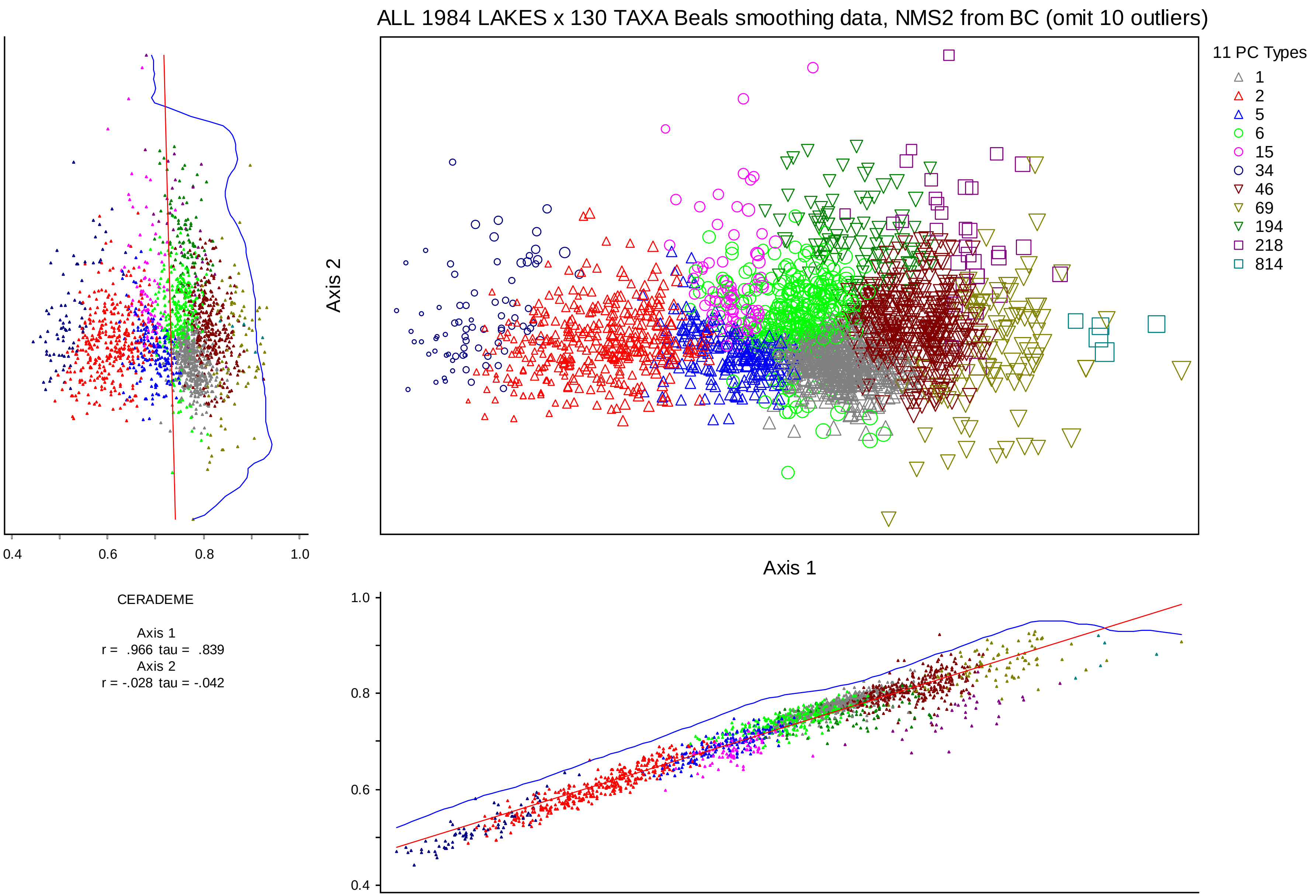


Figure ALLKS-6: NMS ordination of ALLKS dataset, overlay of favorability of lakes for **coontail** (*Ceratophyllum demersum*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent 11 ALLKS PC Types.

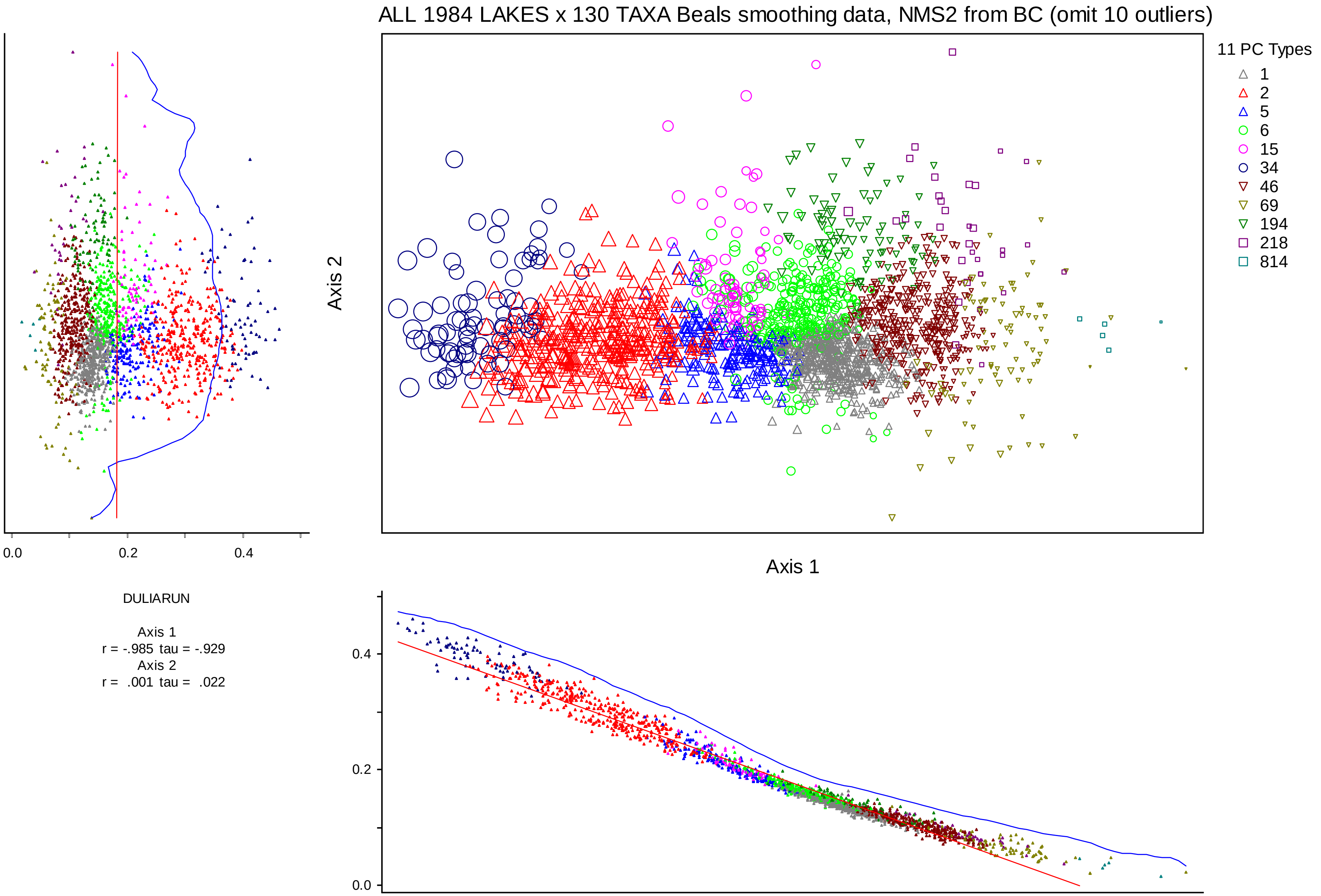


Figure ALLKS-7: NMS ordination of ALLKS dataset, overlay of favorability of lakes for **three-way sedge** (*Dulichium arundinaceum*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent 11 ALLKS PC Types.

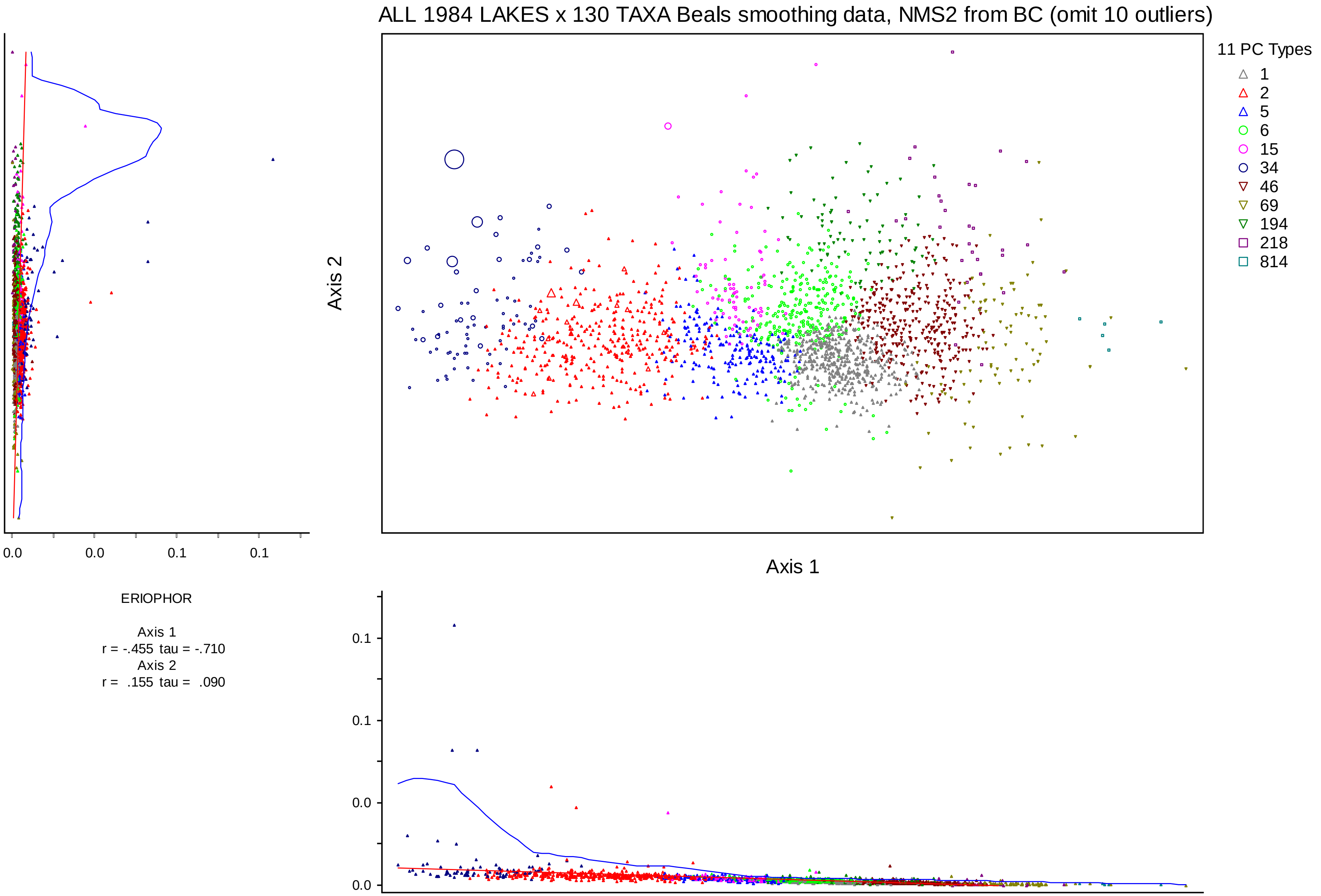


Figure ALLKS-8: NMS ordination of ALLKS dataset, overlay of favorability of lakes for **cotton grass** (*Eriophorum* spp.). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent 11 ALLKS PC Types.

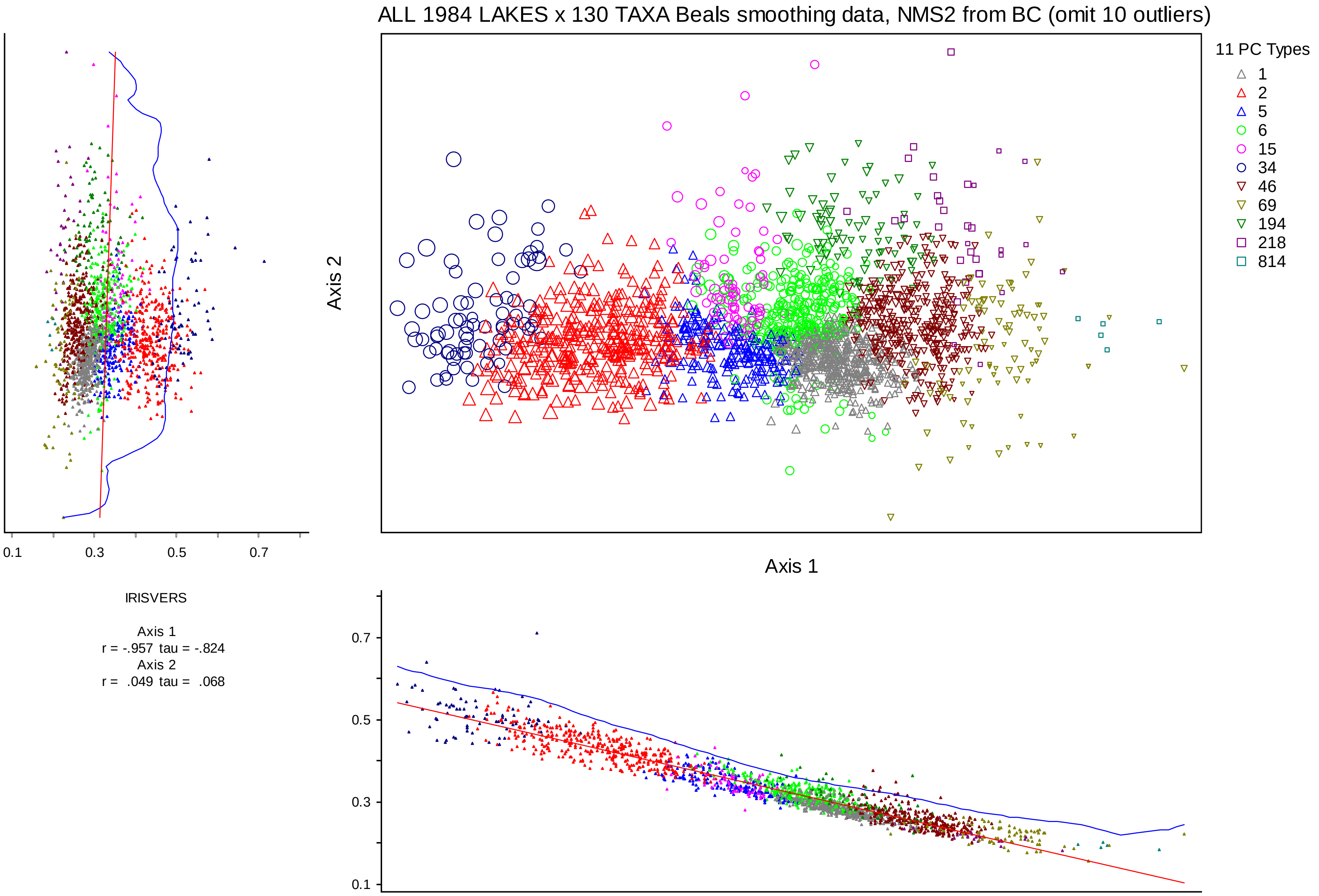


Figure ALLKS-9: NMS ordination of ALLKS dataset, overlay of favorability of lakes for **blue flag iris** (*Iris versicolor*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent 11 ALLKS PC Types.

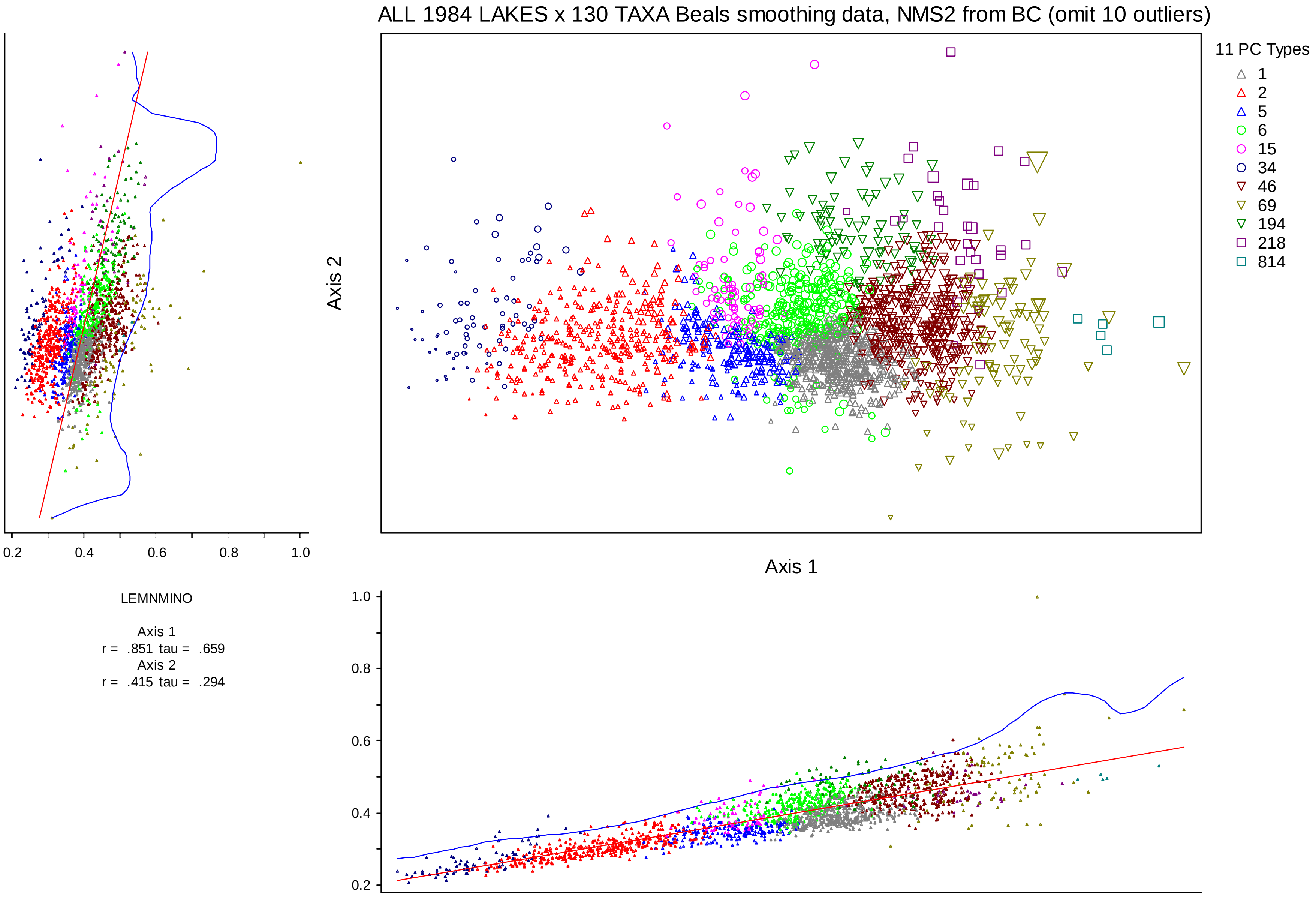


Figure ALLKS-10: NMS ordination of ALLKS dataset, overlay of favorability of lakes for **lesser duckweed** (*Lemna minor*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent 11 ALLKS PC Types.

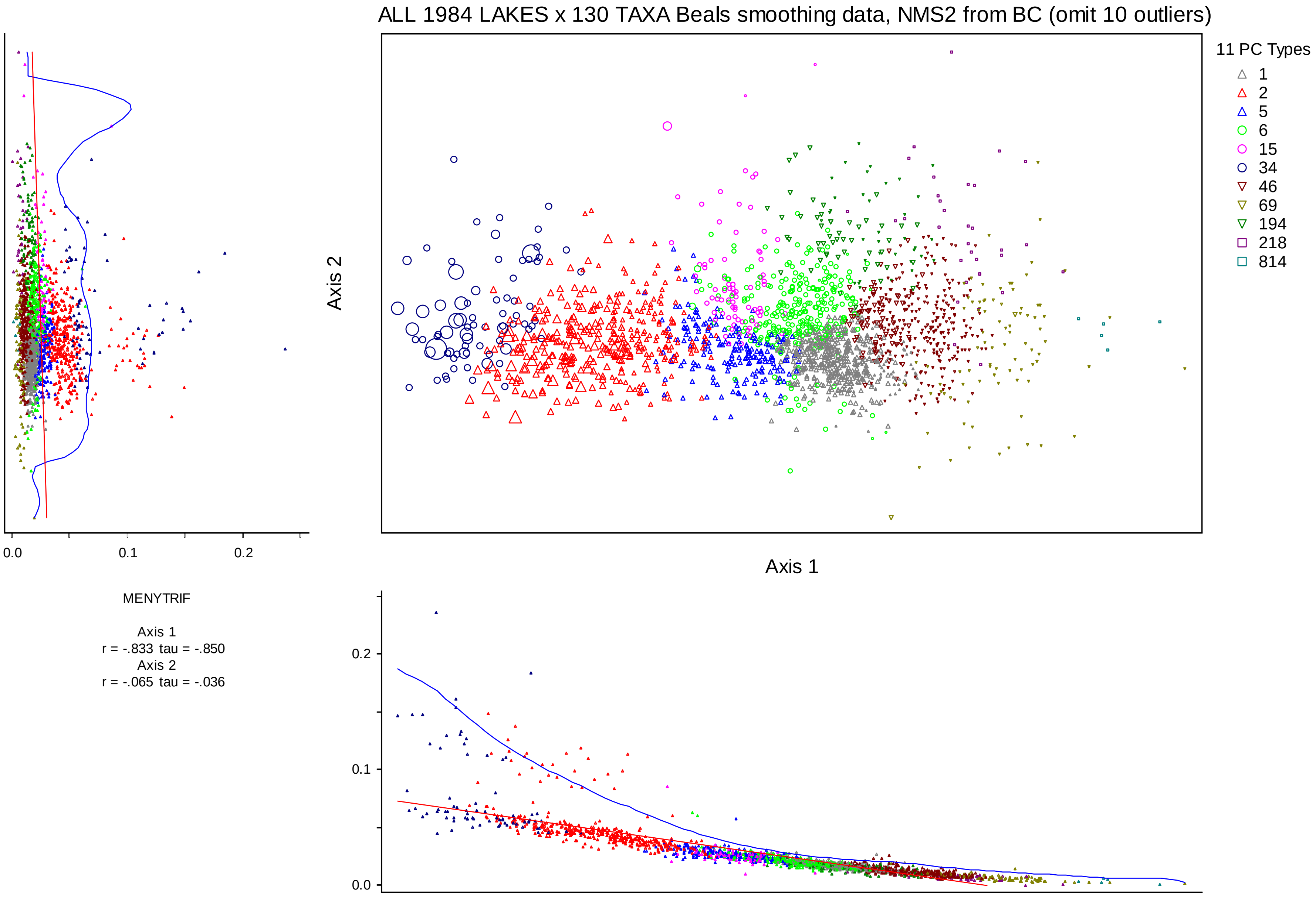


Figure ALLKS-11: NMS ordination of ALLKS dataset, overlay of favorability of lakes for **buckbean** (*Menyanthes trifoliata*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent 11 ALLKS PC Types.

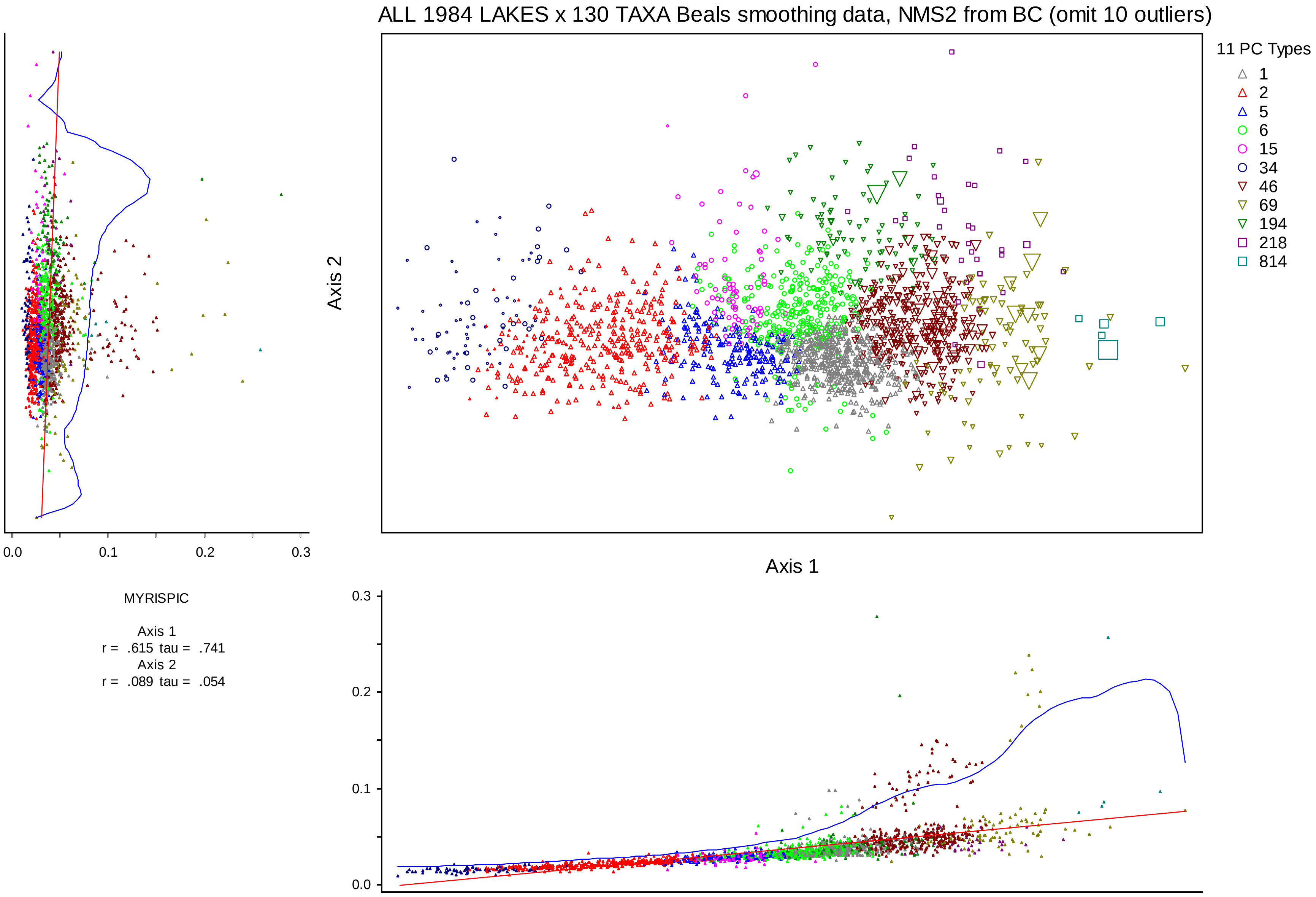


Figure ALLKS-12: NMS ordination of ALLKS dataset, overlay of favorability of lakes for exotic **Eurasian milfoil** (*Myriophyllum spicatum*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent 11 ALLKS PC Types.

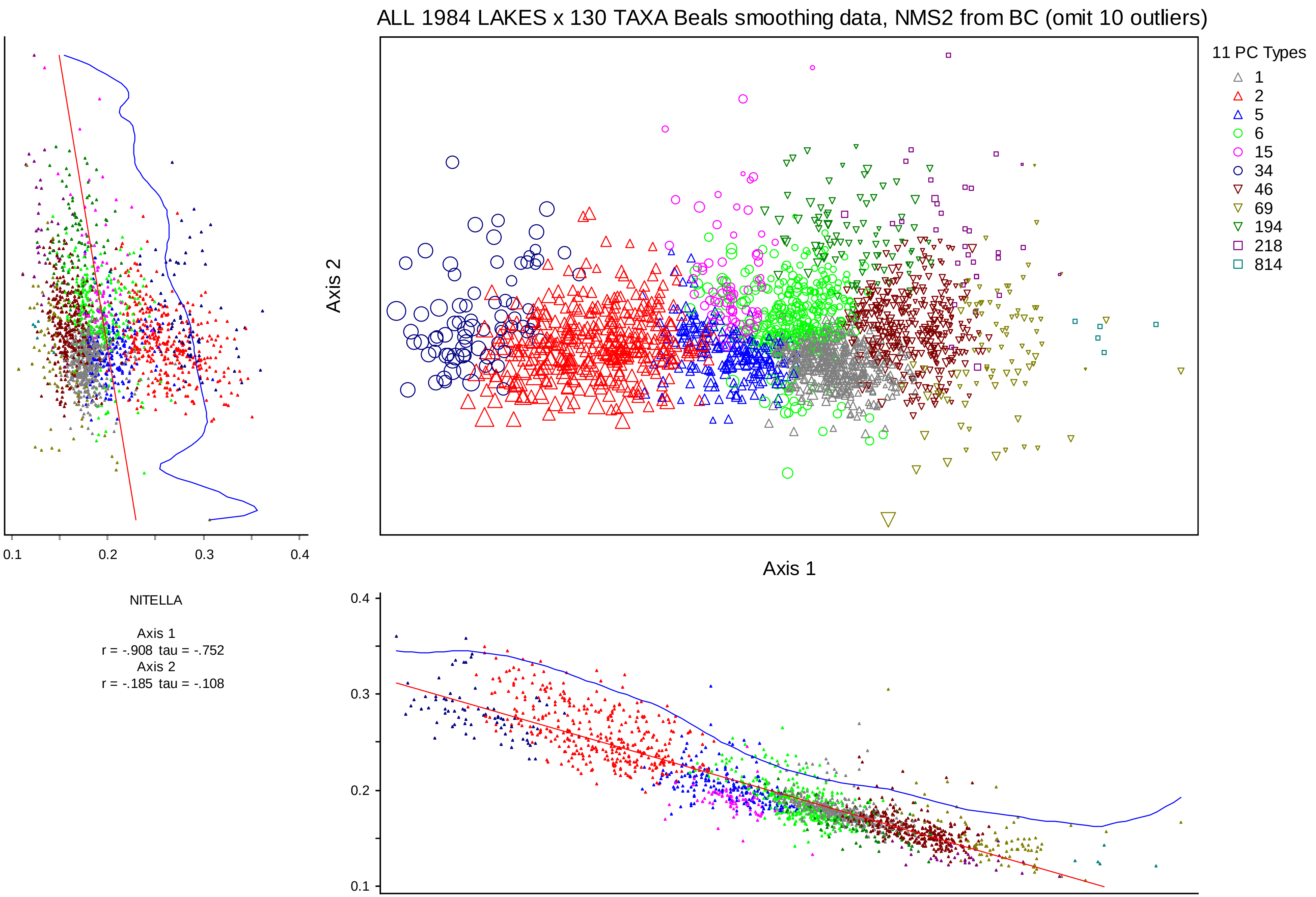


Figure ALLKS-13: NMS ordination of ALLKS dataset, overlay of favorability of lakes for **stoneworts** (*Nitella spp.*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent 11 ALLKS PC Types.

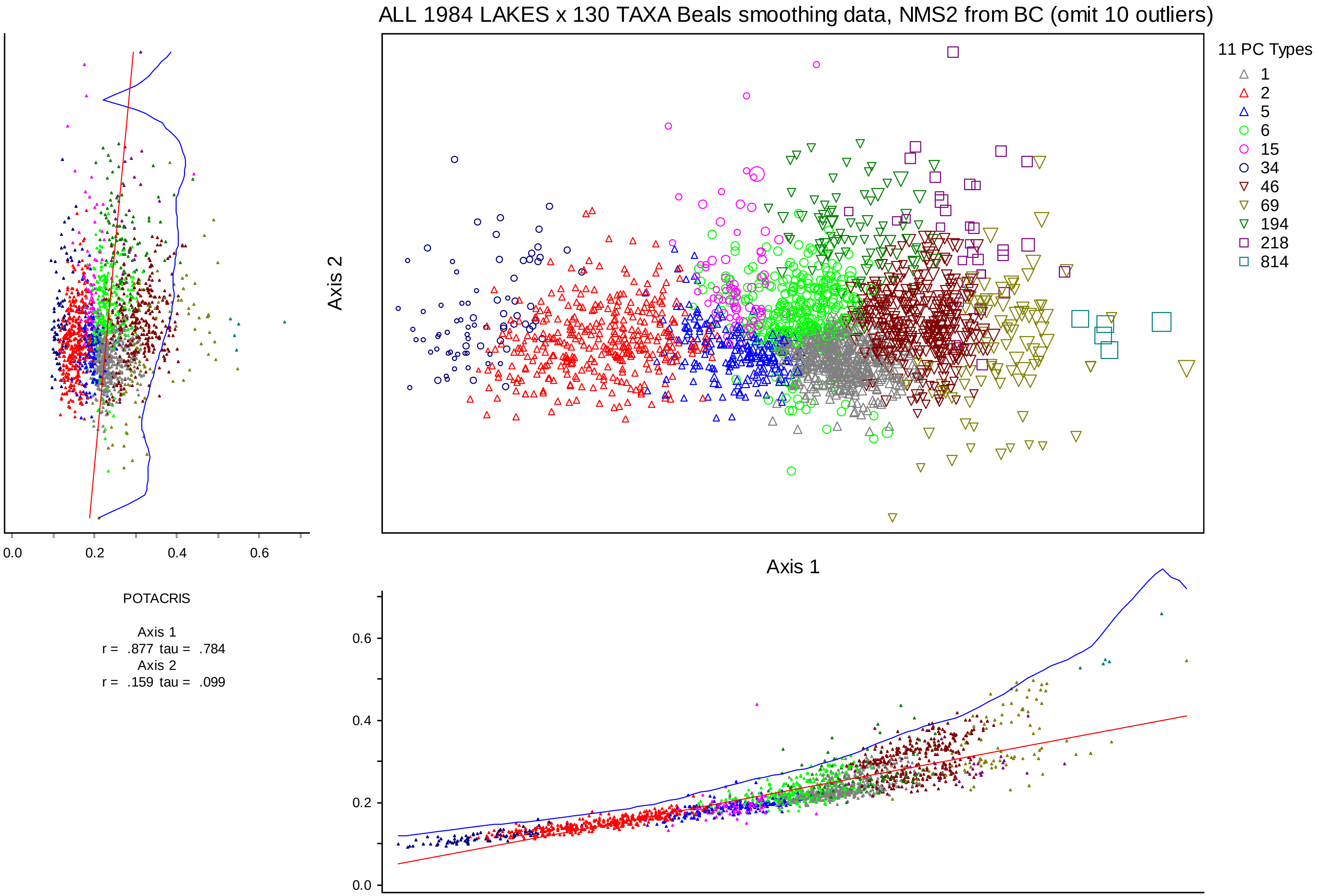


Figure ALLKS-14: NMS ordination of ALLKS dataset, overlay of favorability of lakes for exotic **curly-leaf pondweed** (*Potamogeton crispus*). Large symbols represent lakes with higher favorability for the taxon. Symbol colors and shapes represent 11 ALLKS PC Types.

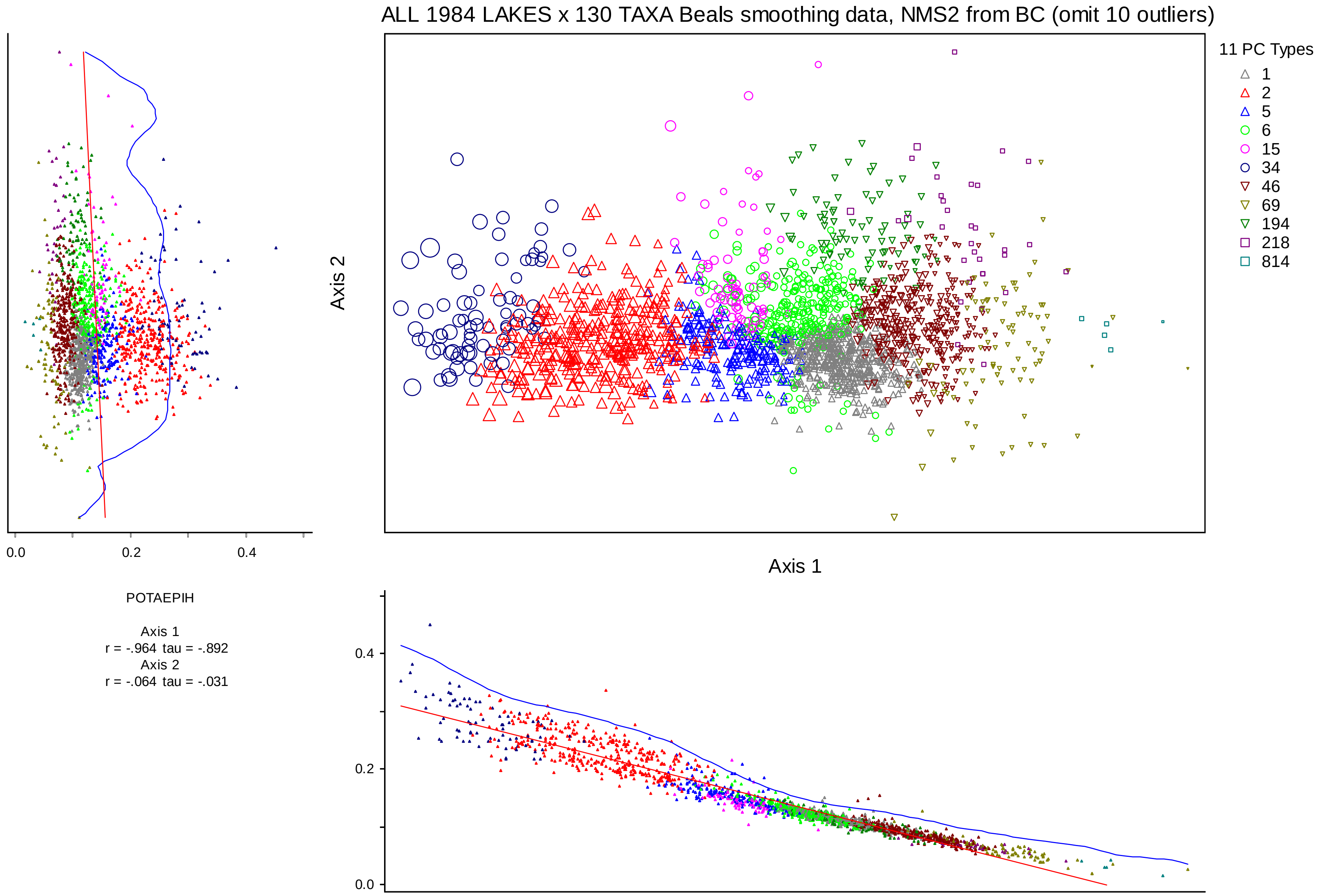


Figure ALLKS-15: NMS ordination of ALLKS dataset, overlay of favorability of lakes for **Nuttall's (ribbon-leaf) pondweed** (*Potamogeton epihydrus*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent 11 ALLKS PC Types.

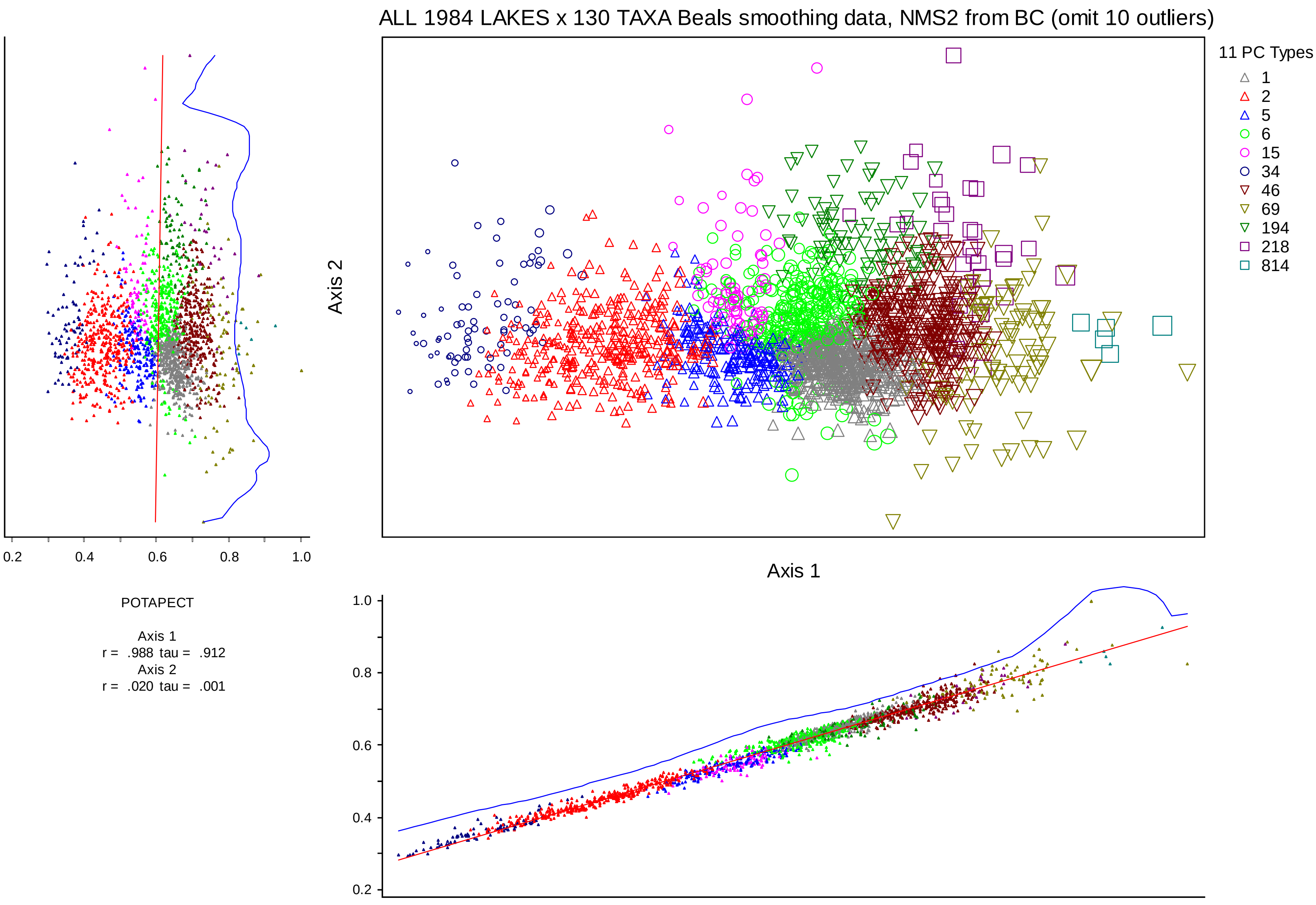


Figure ALLKS-16: NMS ordination of ALLKS dataset, overlay of favorability of lakes for **Sago pondweed** (*Potamogeton pectinatus*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent 11 ALLKS PC Types.

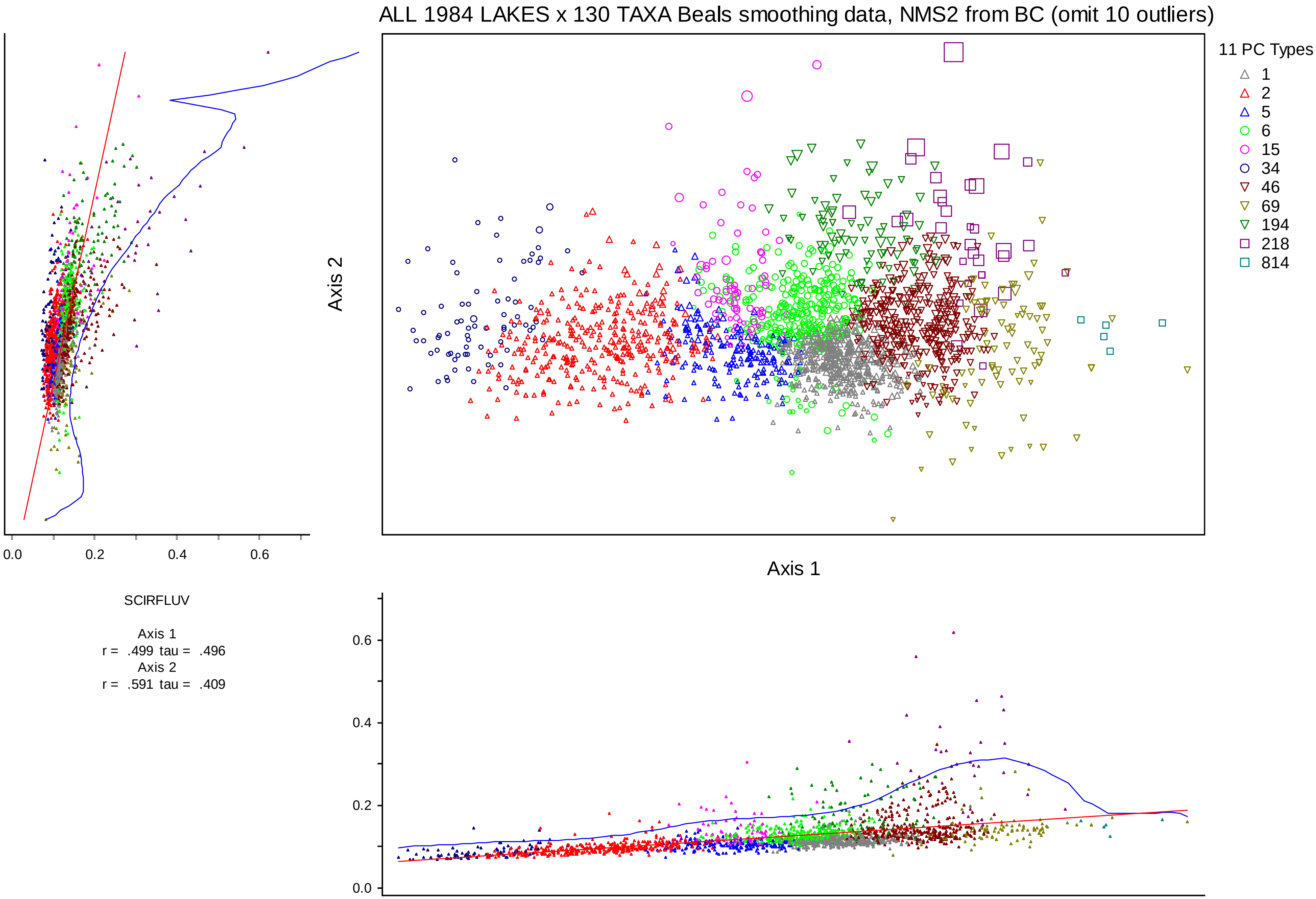


Figure ALLKS-17: NMS ordination of ALLKS dataset, overlay of favorability of lakes for **river bulrush** (*Scirpus fluviatilis*). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent 11 ALLKS PC Types.

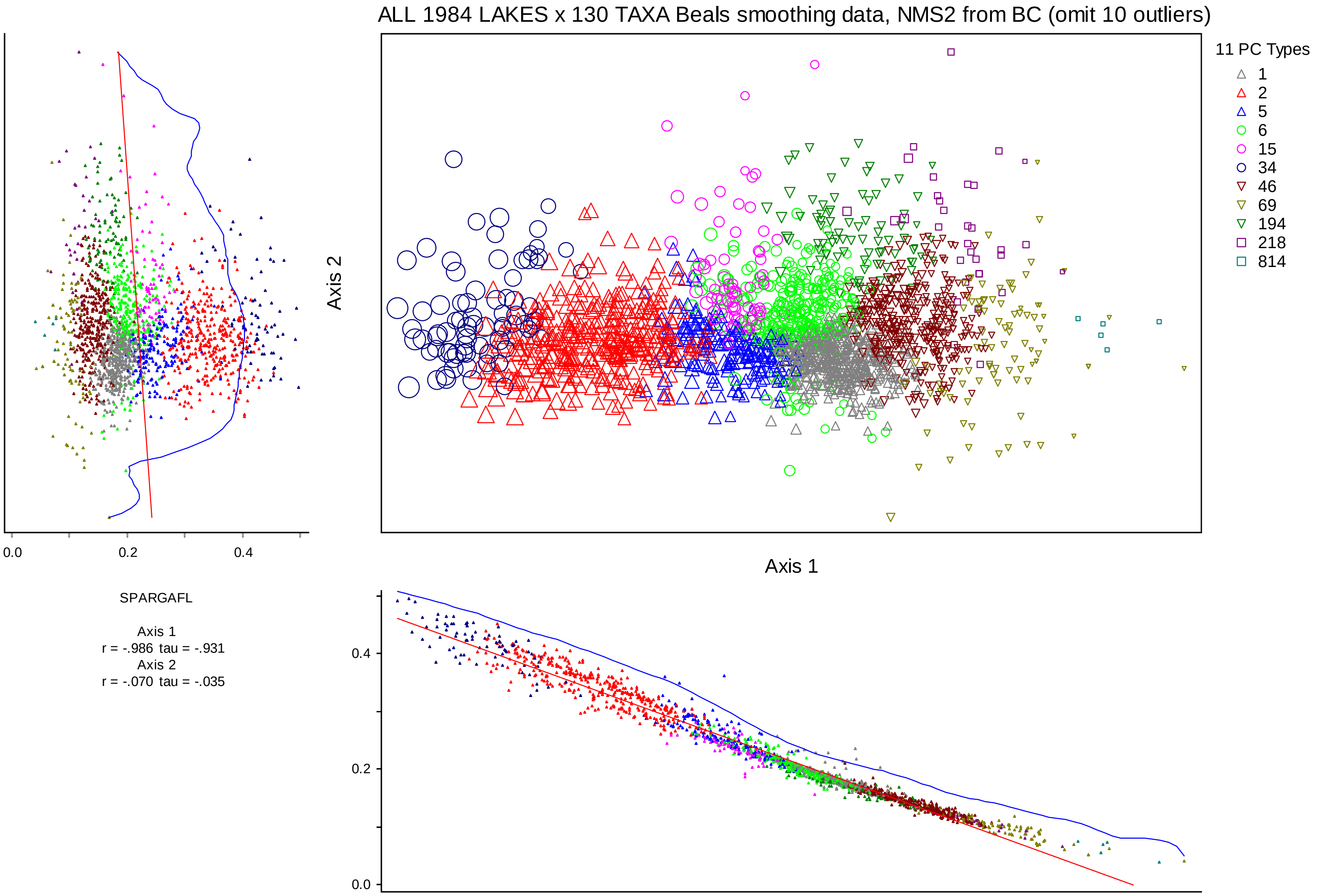


Figure ALLKS-18: NMS ordination of ALLKS dataset, overlay of favorability of lakes for **floating-leaved burreeds** (*Sparganium* spp.). Large symbols represent lakes with higher favorability values for the taxon. Symbol colors and shapes represent 11 ALLKS PC Types.

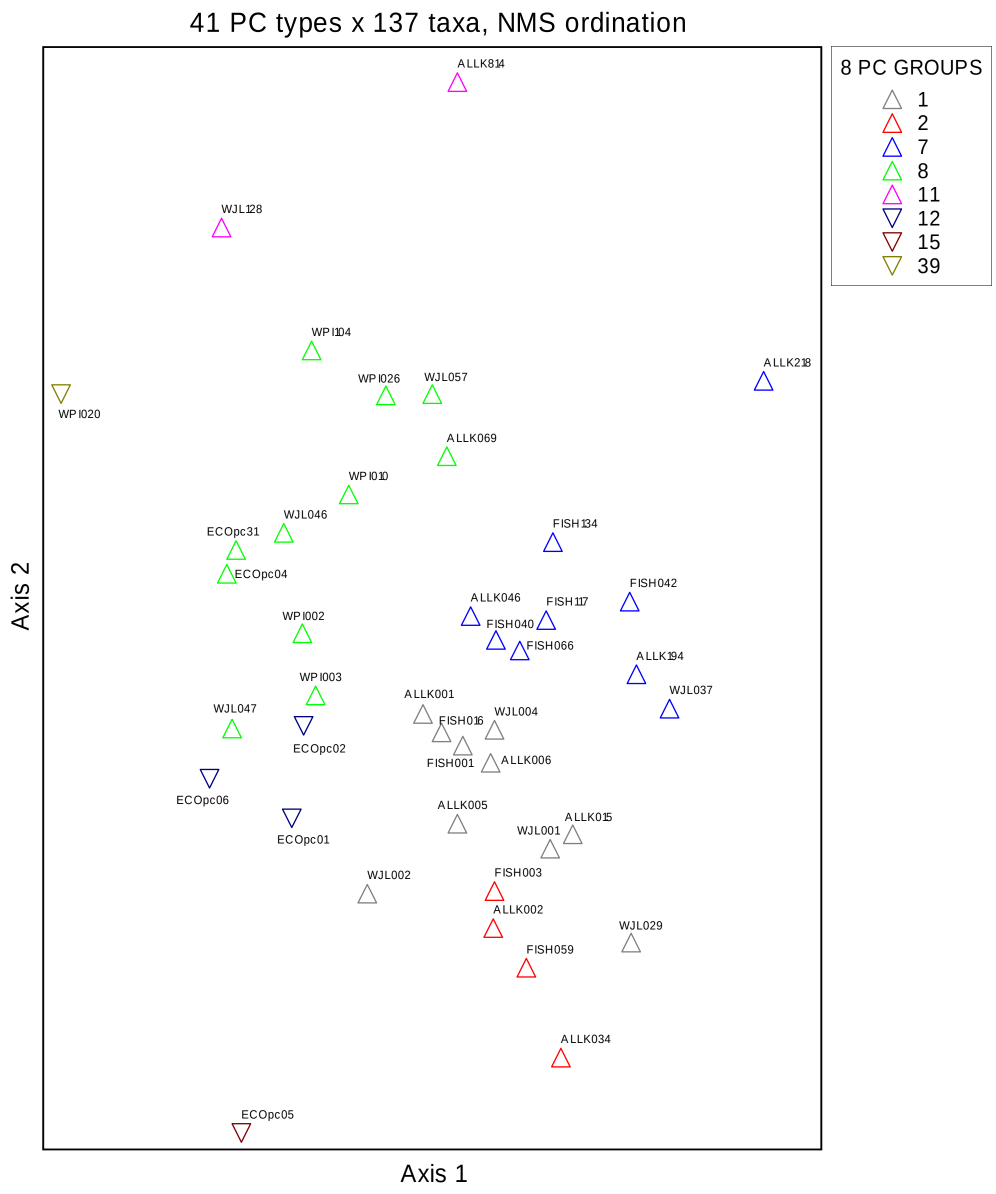


Figure ALLKS-19: NMS ordination of PCTYPES dataset. Symbol colors and shapes represent 8 PC Groups. Each symbol represents one PC type recognized in ordinations of the ECO, FISH, WJL, WPI, or ALLKS datasets; symbols are labeled to show which PC type they represent.

Map Plates

Map Plate 1. Distribution of lakes from four datasets by Omernik Level III Ecoregions.

Map Plate 2. Distribution of lakes from four datasets by MN DNR Ecological Sections.

Map Plate 3. Percentage of Agricultural land use within 1 km buffer of sampled lakes.

Map Plate 4. Percentage of Forested land use within 1 km wide buffer of sampled lakes.

Map Plate 5. Percentage of Developed land use within 1 km buffer of sampled lakes.

Map Plate 6. Percentage of Wetland land use within 1 km wide buffer of sampled lakes.

Map Plate 7. Population density (persons / km^2) within 1 km buffer of sampled lakes.

Map Plate 8. Road density (km road/ km^2 area) within 1 km buffer of sampled lakes.

Map Plate 9. Species Richness in sampled lakes.

Map Plate 10. Shannon diversity (H’) in sampled lakes.

Map Plate 11. Taxa Richness in sampled lakes.

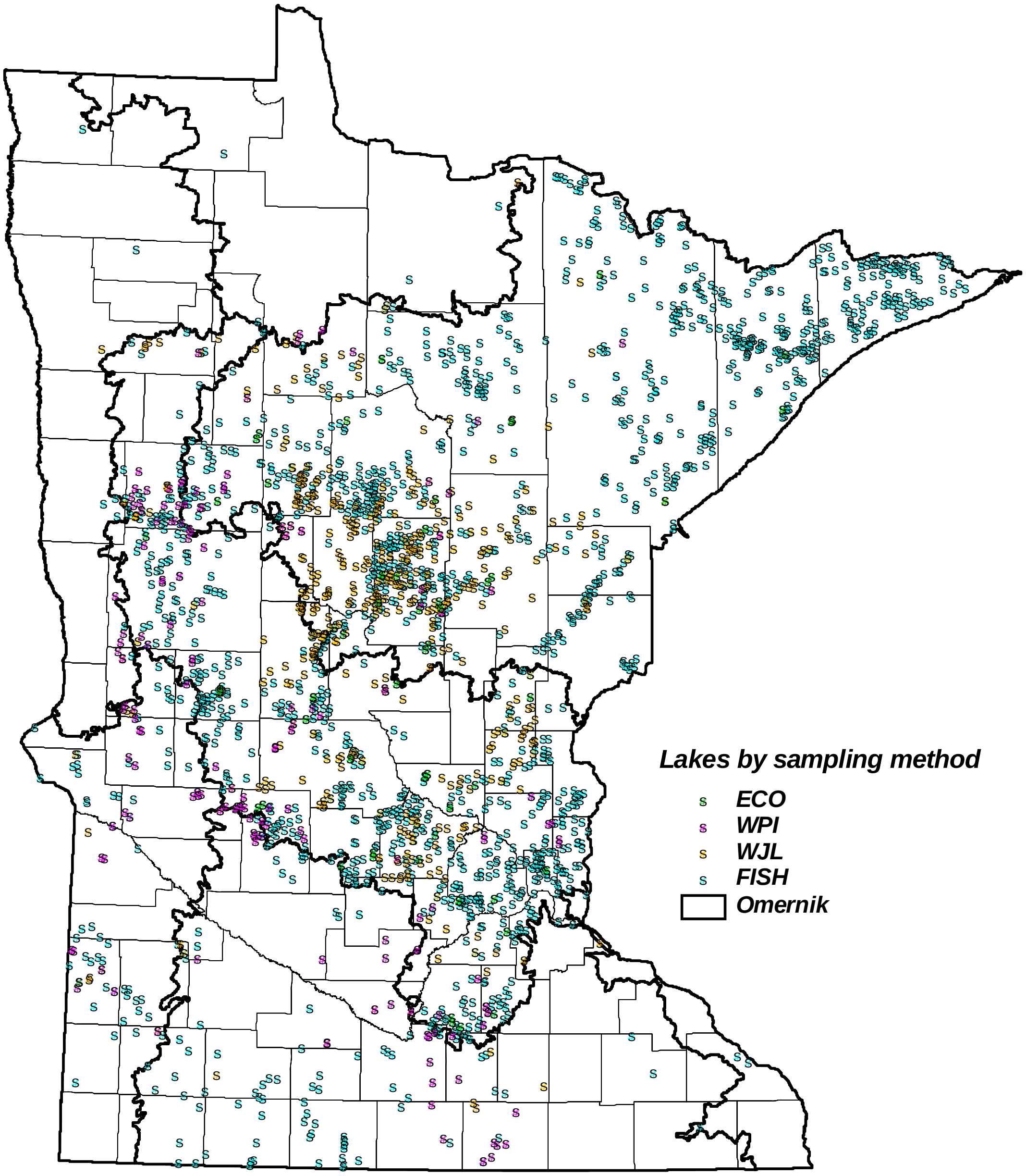
Map Plate 12. Aquatic plant community types based on analysis of ECO dataset.

Map Plate 13. Aquatic plant community types based on analysis of FISH dataset.

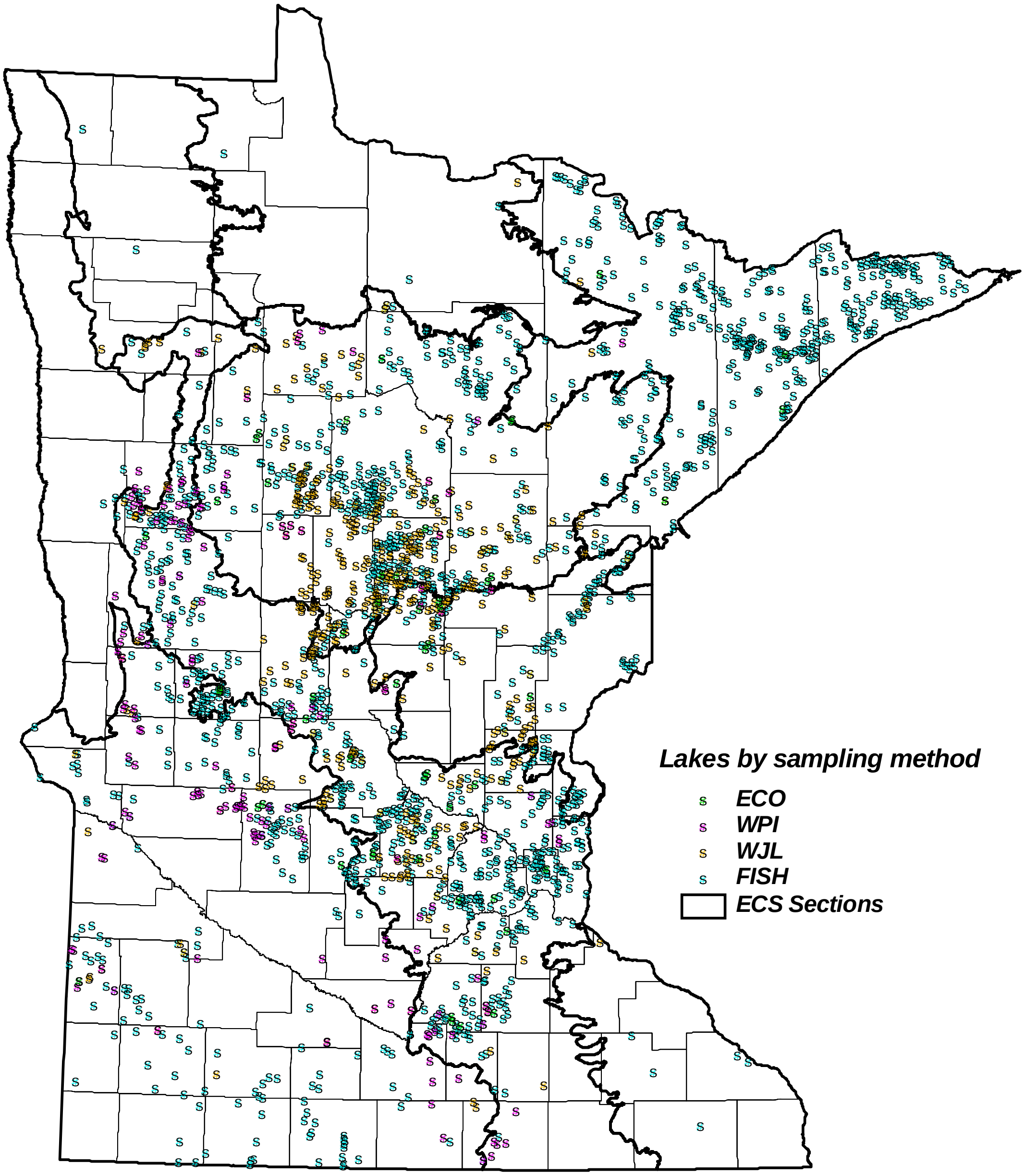
Map Plate 14. Aquatic plant community types based on analysis of WJL dataset.

Map Plate 15. Aquatic plant community types based on analysis of WPI dataset.

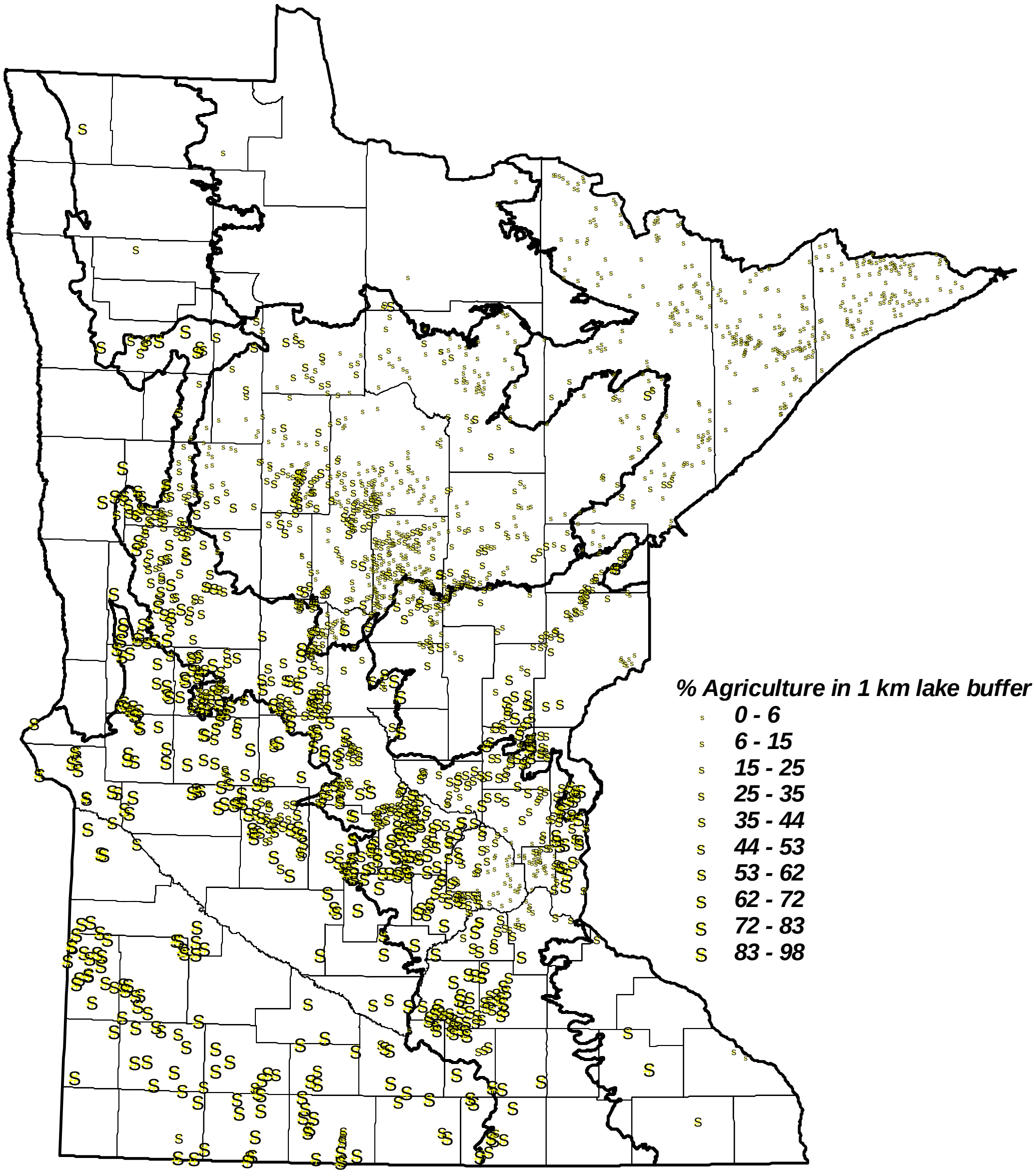
Map Plate 16. Aquatic plant community types based on analysis of ALLKS dataset.



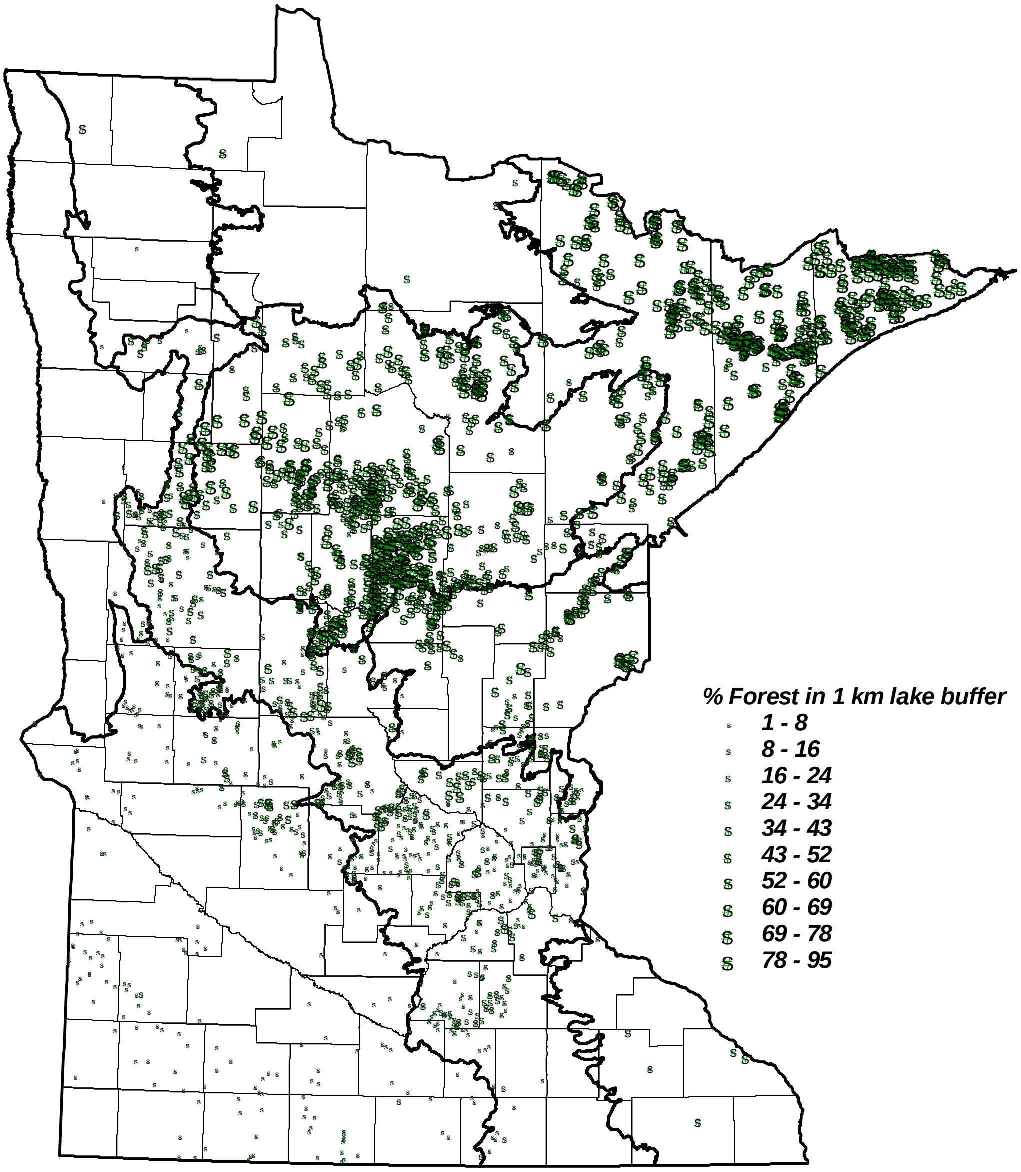
Map Plate 1. Distribution of lakes from four datasets by Omernik Level III Ecoregions.



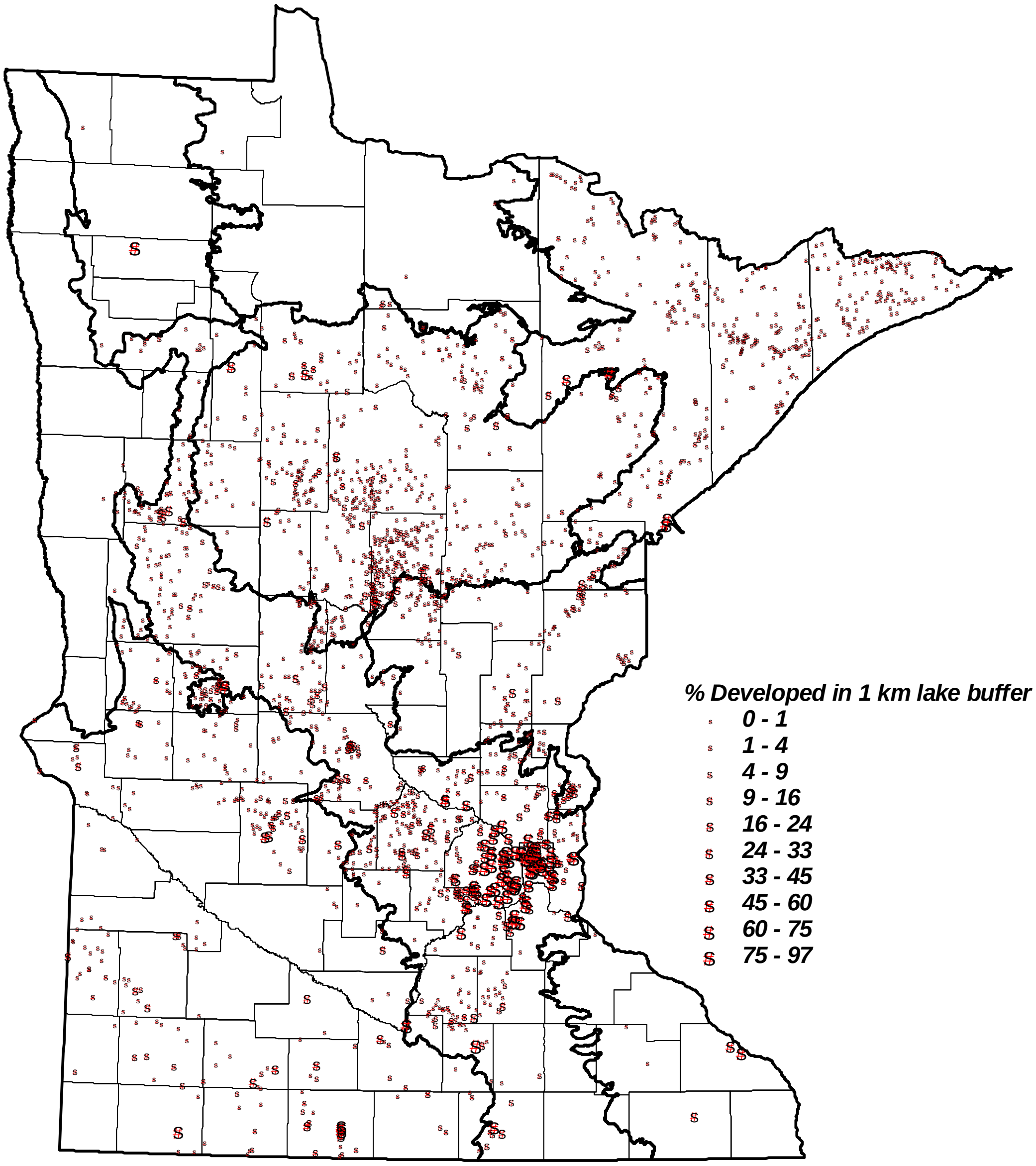
Map Plate 2. Distribution of lakes from four datasets by MN DNR Ecological Sections.



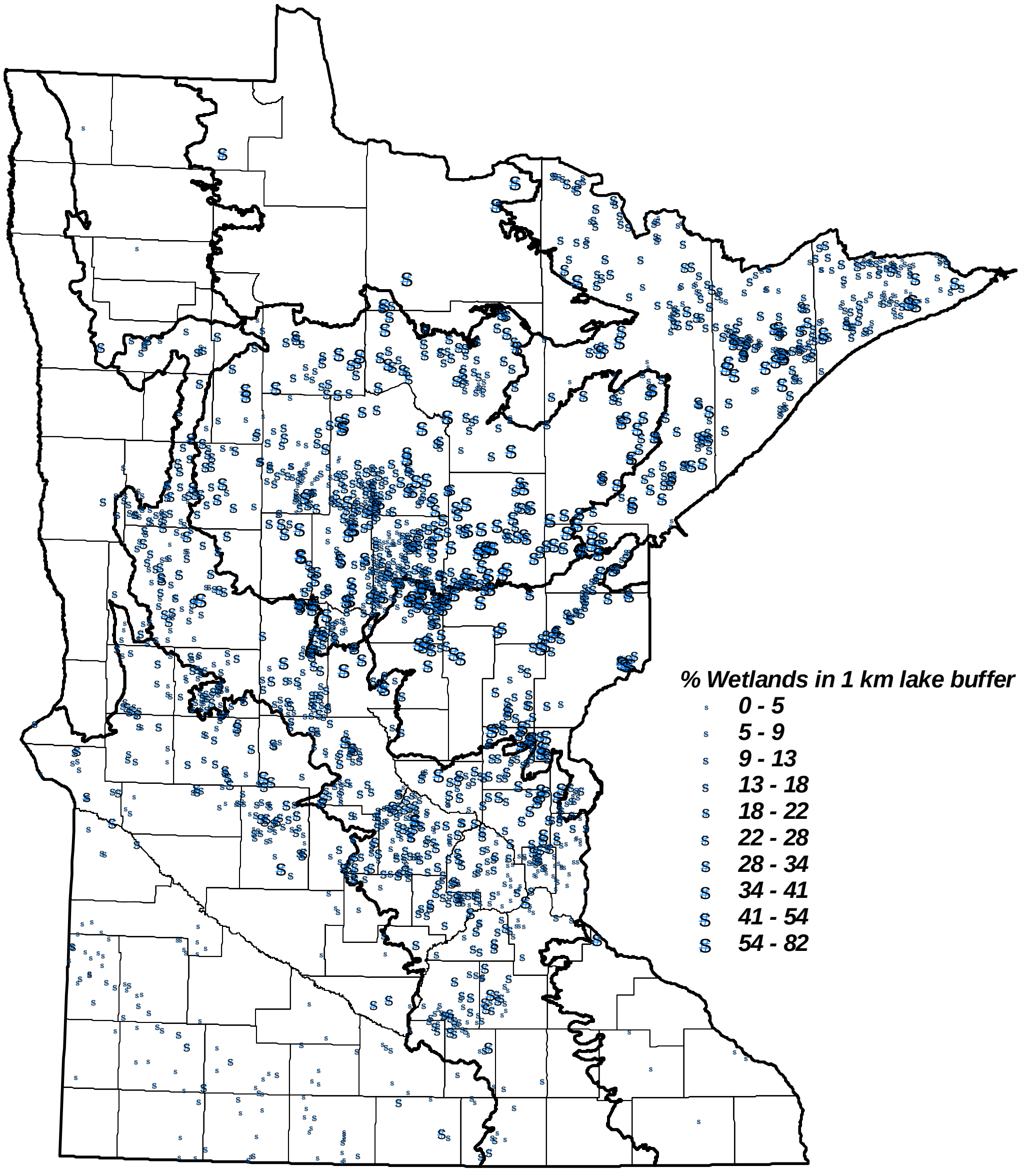
Map Plate 3. Percentage of Agricultural land use within 1 km buffer of sampled lakes.



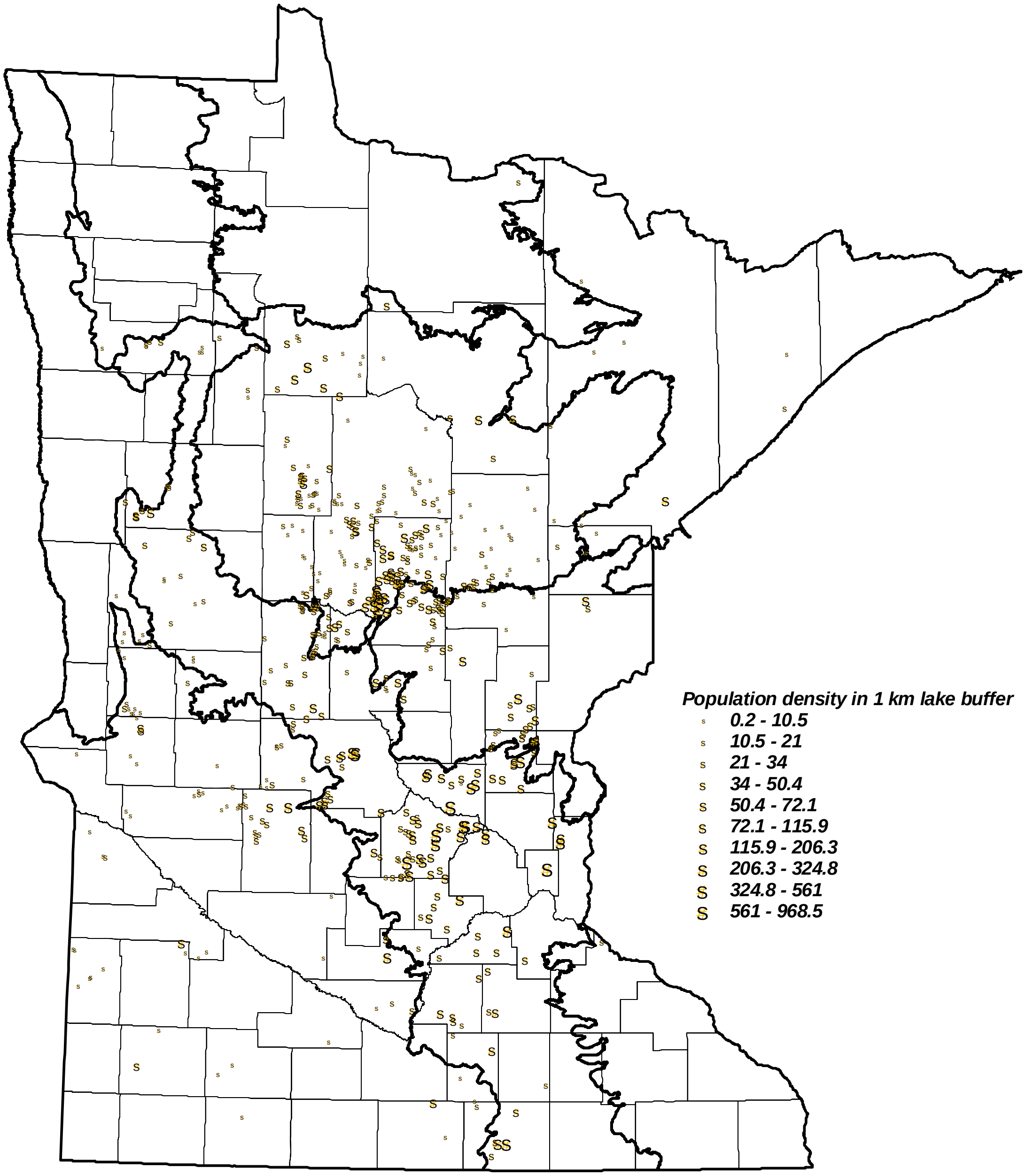
Map Plate 4. Percentage of Forested land use within 1 km wide buffer of sampled lakes.



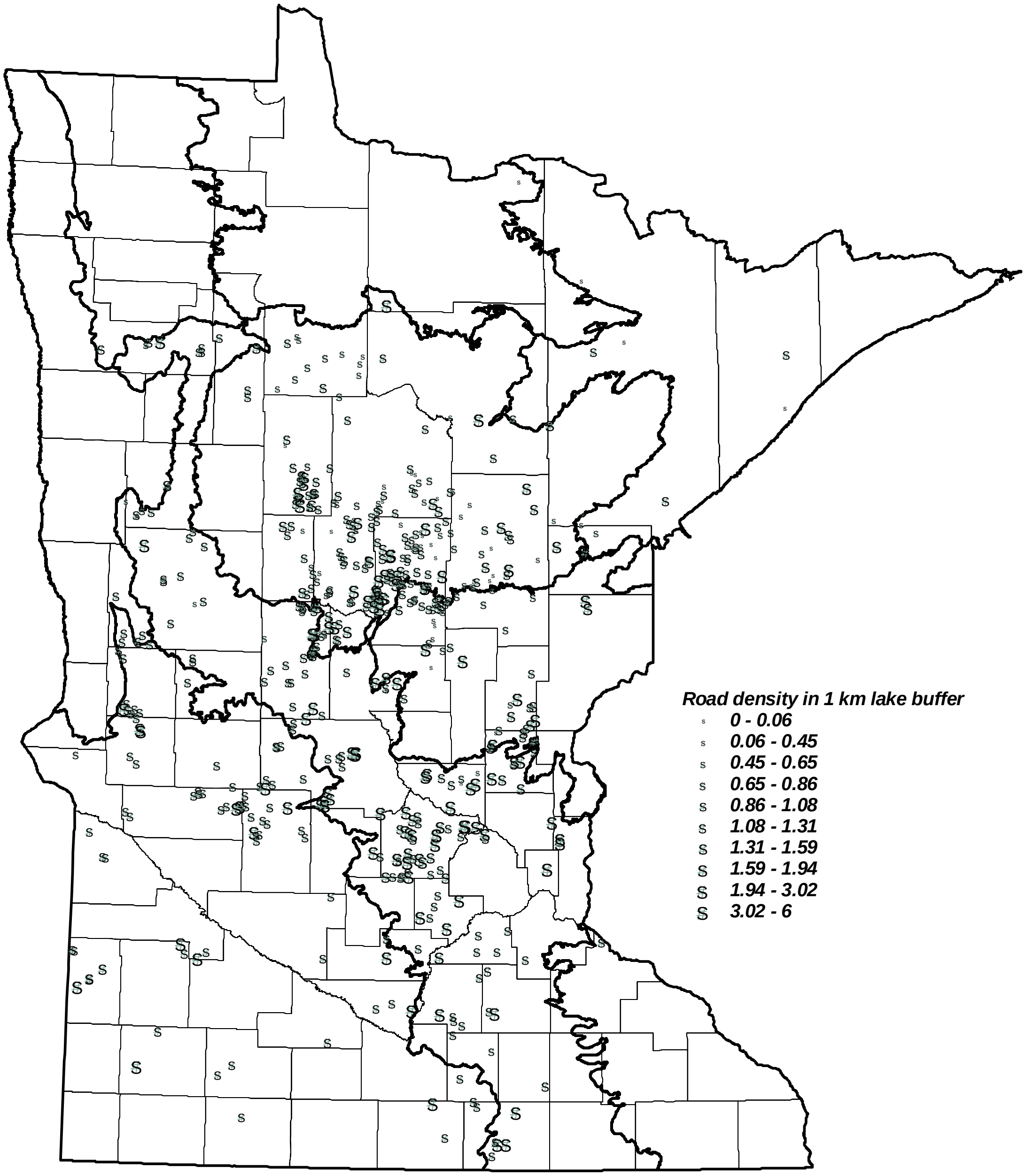
Map Plate 5. Percentage of Developed land use within 1 km buffer of sampled lakes.



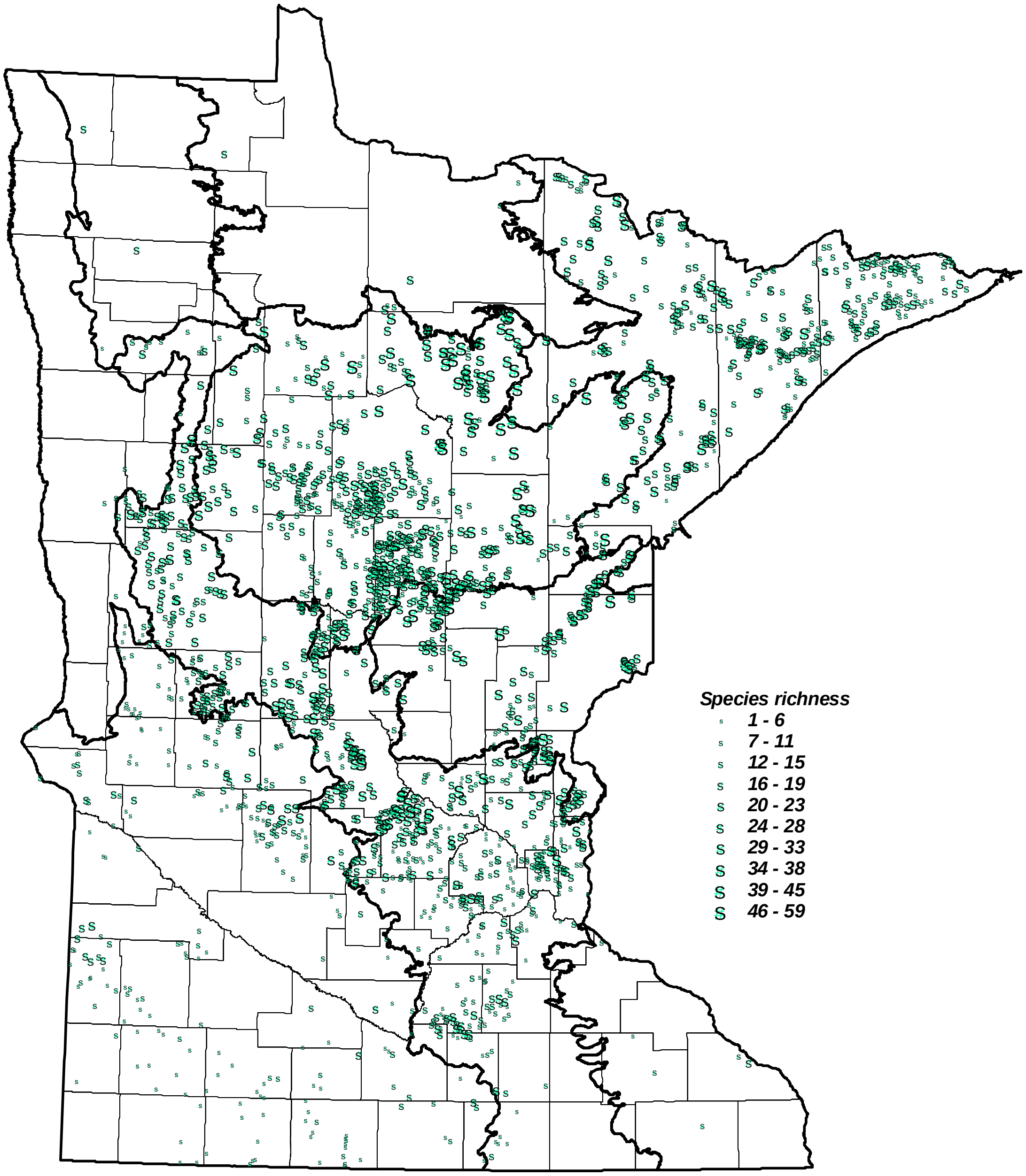
Map Plate 6. Percentage of Wetland land use within 1 km wide buffer of sampled lakes.



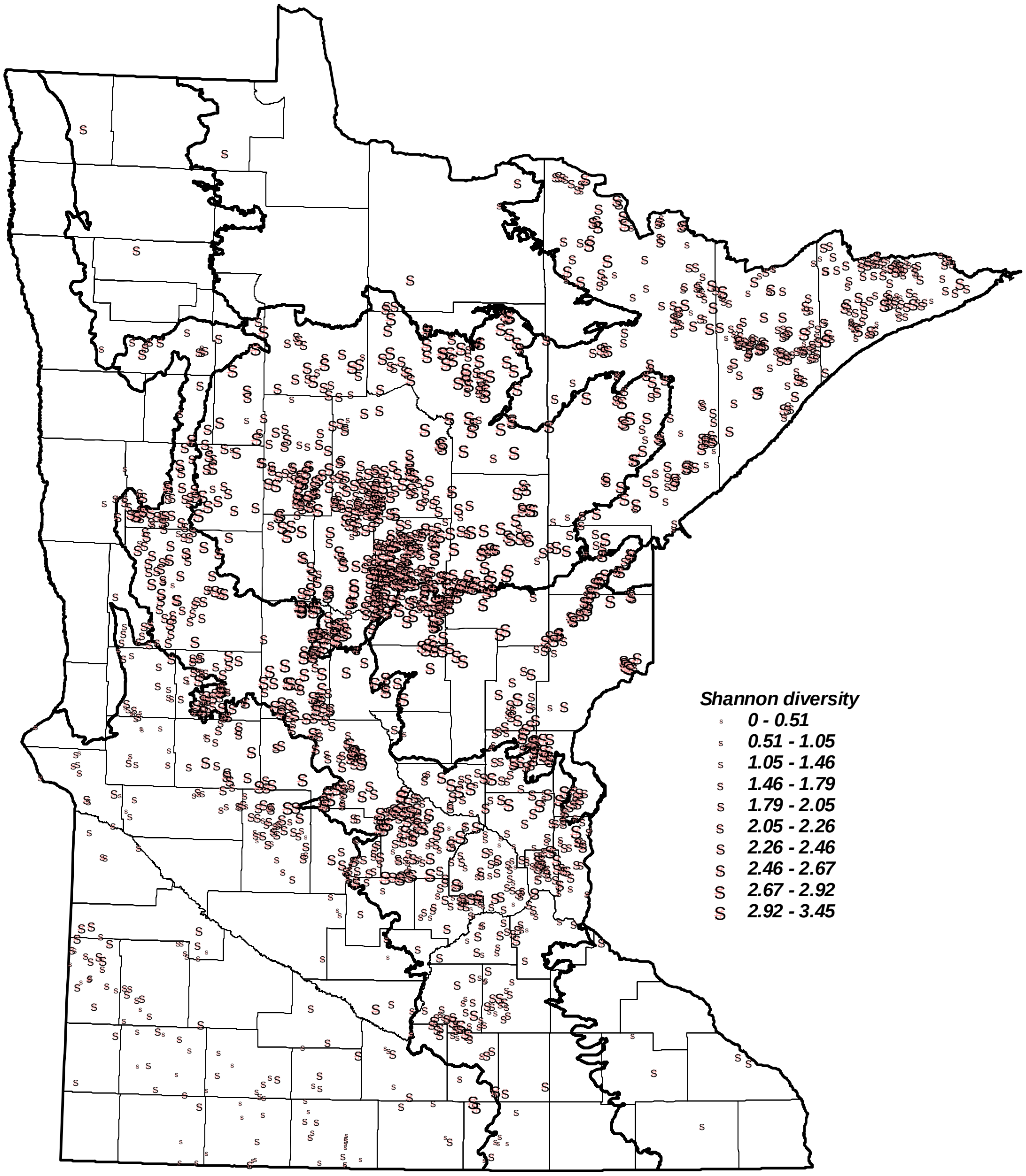
Map Plate 7. Population density (persons / km^2) within 1 km buffer of sampled lakes.



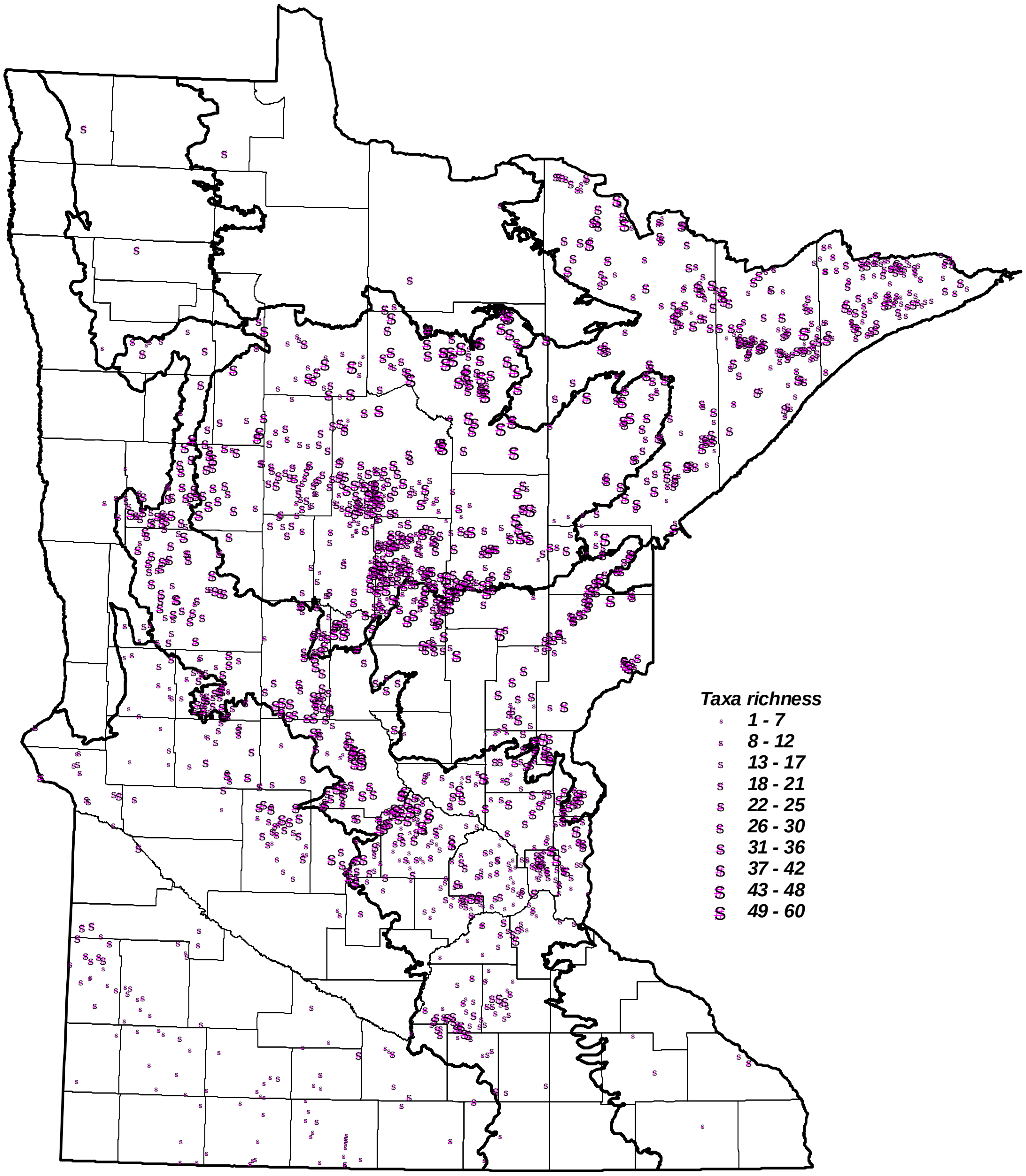
Map Plate 8. Road density (km road/ km^2 area) within 1 km buffer of sampled lakes.



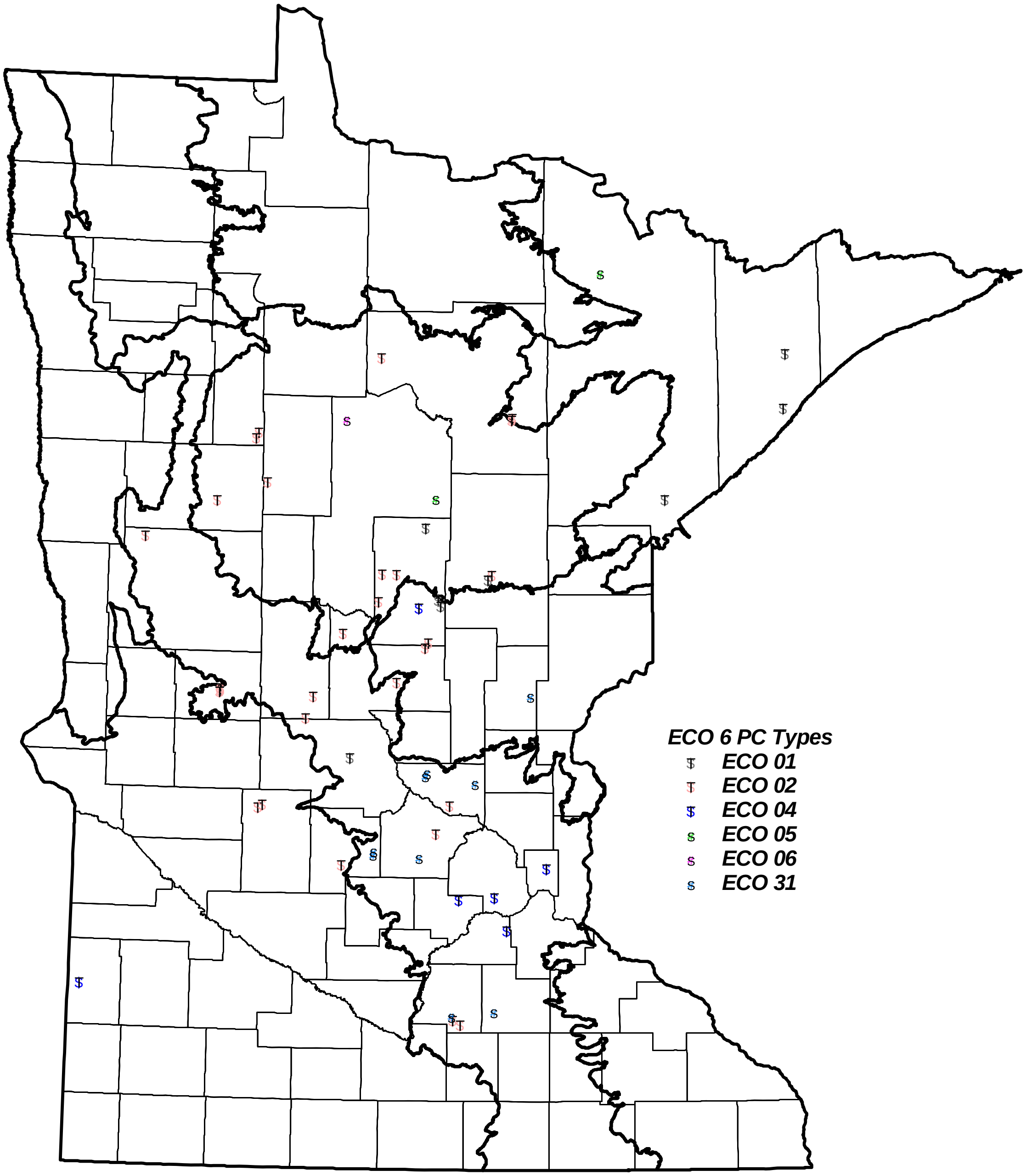
Map Plate 9. Species Richness in sampled lakes.



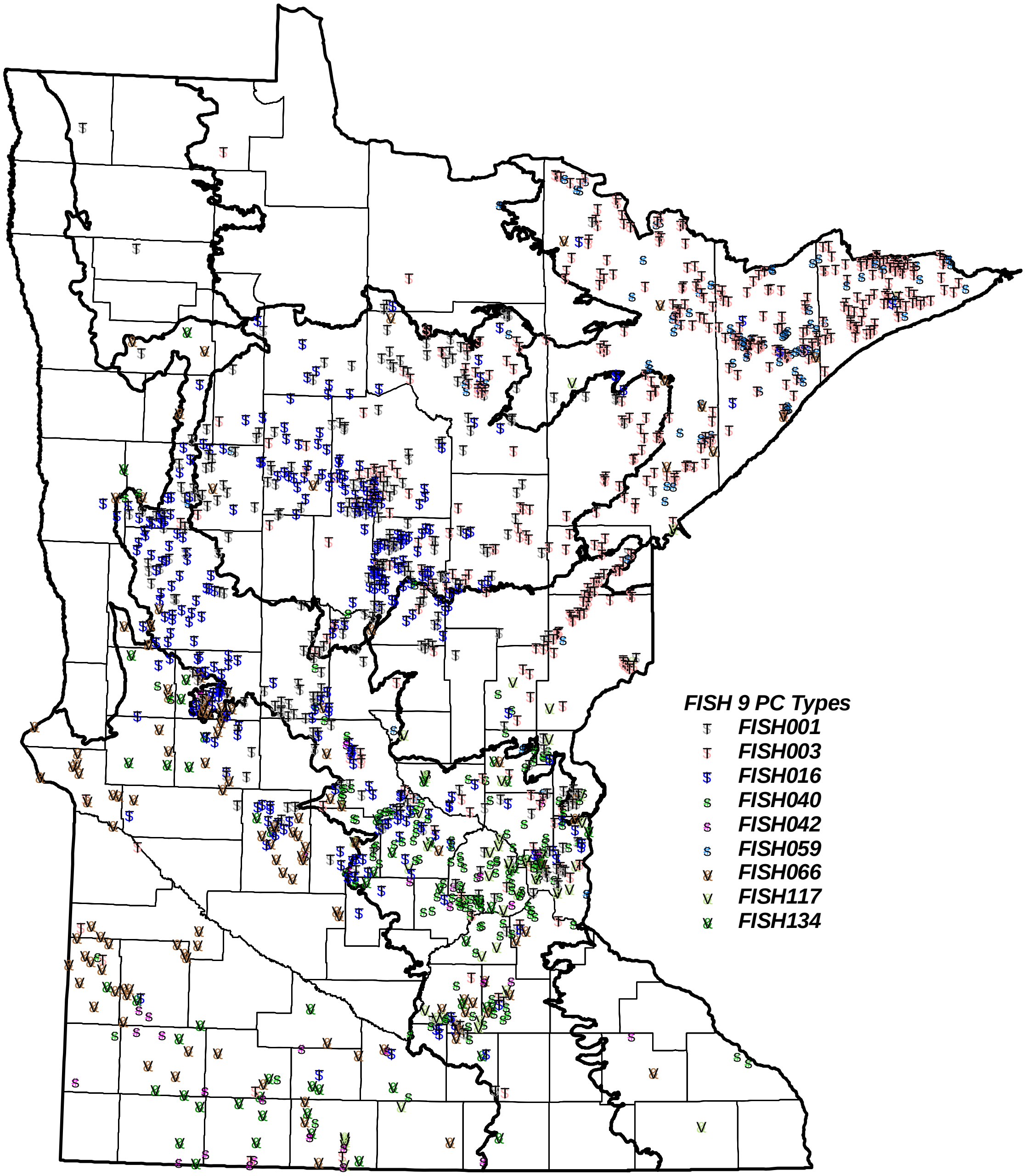
Map Plate 10. Shannon diversity (H’) in sampled lakes.



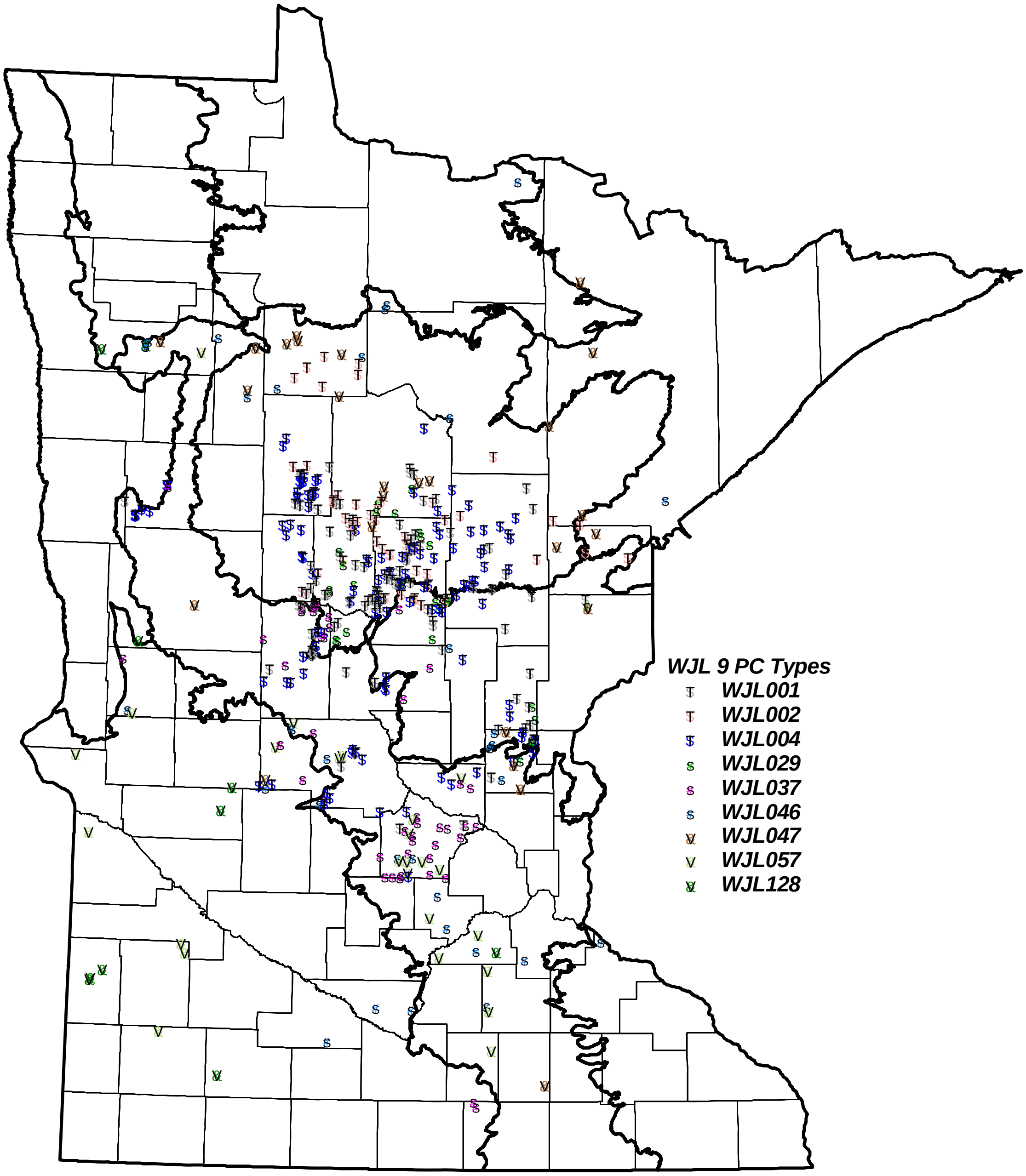
Map Plate 11. Taxa Richness in sampled lakes.



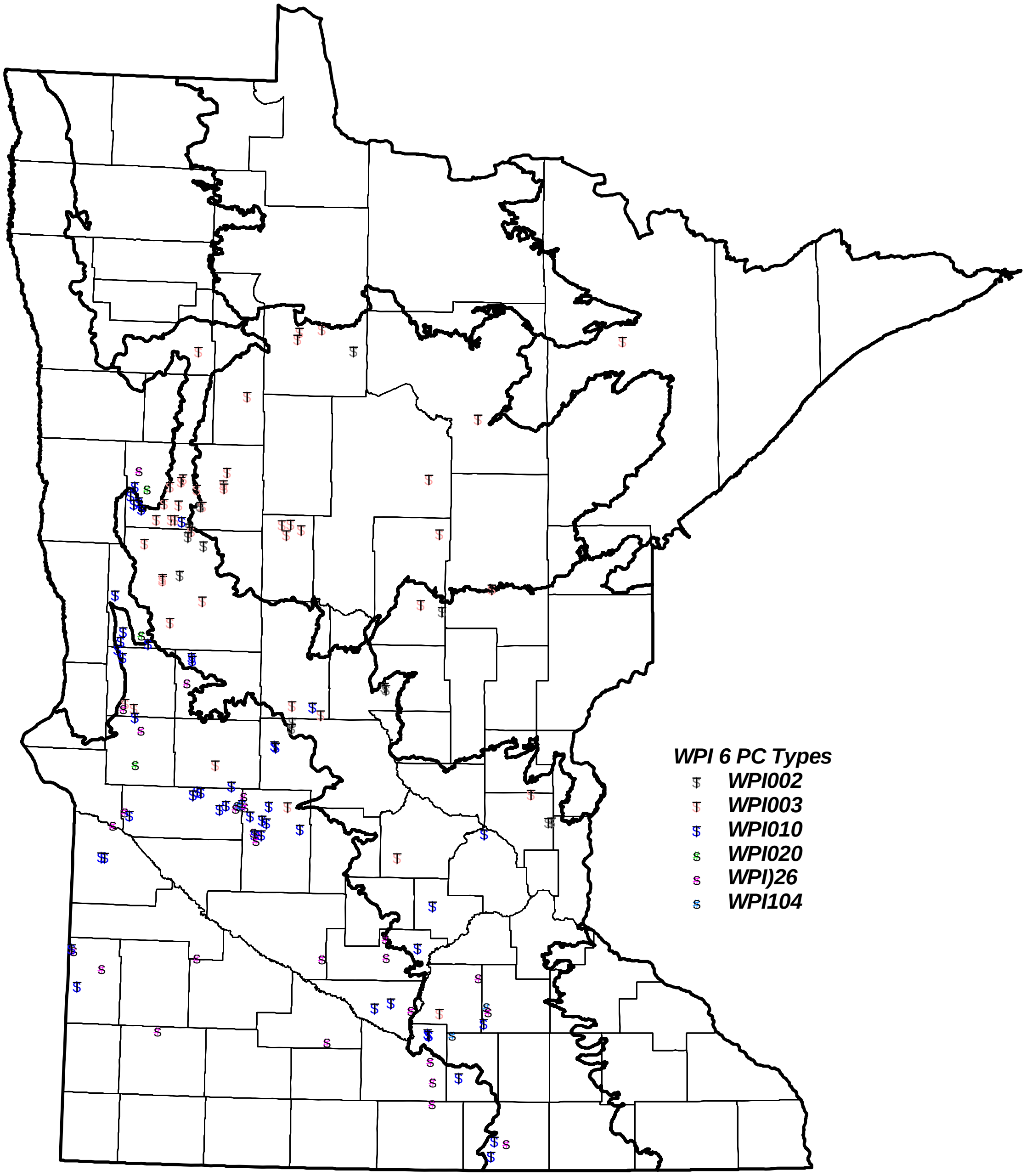
Map Plate 12. Aquatic plant community types based on analysis of ECO dataset.



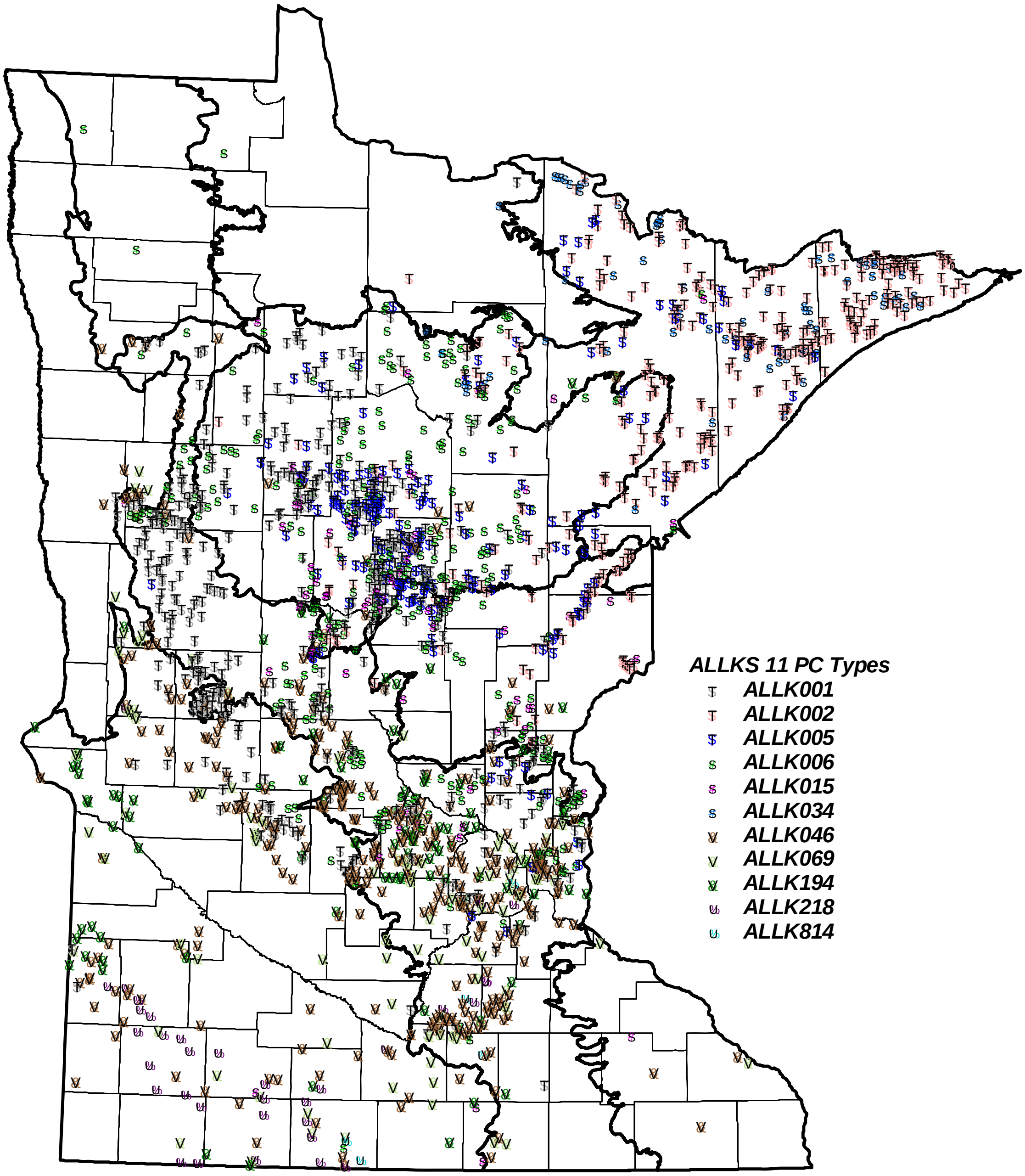
Map Plate 13. Aquatic plant community types based on analysis of FISH dataset.



Map Plate 14. Aquatic plant community types based on analysis of WJL dataset.



Map Plate 15. Aquatic plant community types based on analysis of WPI dataset.



Map Plate 16. Aquatic plant community types based on analysis of ALLKS dataset.