

Supplement B: Example Analysis Pathways

Example 1: Wildlife Habitat Summarization

Objective: Summarize a suite of environmental variables to characterize the conditions present at sites supporting a species of concern and for use as predictor variables in habitat modeling.

Main Matrix structure: 26 sites × 24 variables

Main Matrix content: Mixture of discrete, continuous, and dummy variables (Q variables)

Second Matrix structure: None

What was measured? One-time measure of metrics characterizing habitat extent and quality, some moderately correlated, all different units. Highly redundant calculated metrics were pared down to a single metric. Dummy variables made for nominal variables.

What do the zeros mean? Very few zeros, zero means zero.

Non-zeros: Made comparable (centered and standardized) automatically in PCA.

Heterogeneity: Data are not very heterogeneous: 14% zeros, average skewness of variables = 0.5, coefficient of variation of site totals = 9%. Centering reduces zeros to 0%.

Sources of noise: Measurement error.

Outliers: There are no obvious outliers in the dataset.

Distance measure: Euclidean because it should perform well on these data and tolerates negative values resulting from centering.

Transformations? After reviewing boxplots, two variables were log-transformed to improve normality.

Relationships? The environmental variables are hypothesized to have a linear relationship with the ordination axes summarizing them.

Model form is thus linear.

Tool to use: PCA (free ordination).

First look: Ran PCA using a correlation cross-products matrix, identified three interpretable axes with strong associations with (i.e., loadings of) six particular variables.

Confirmation: Ran NMS on Autopilot (on medium) using Euclidean distances after relativizing by standard deviates; identified three interpretable axes having strong associations with four of the six particular variables.

Tool to report: PCA. Re-ran the PCA, saving the Column Graph and Results Files this time and then saving those with the Main Matrix as a Project File.

Interpretation: Took the % variance explained from the PCA Result File. Graphed the ordination scores, using Main Matrix Overlay. Applied a square-root transformation to three variables to improve linearity of relationship with ordination axes. Calculated correlation coefficients of all variables with the ordination axes. PCA axes used as predictor variables in habitat modeling (outside of *PC-ORD*).

Story: 24 variables reduced down to 3 composite variables that strongly reflect the patterns of six variables. Future sampling should focus on these six variables.

Summaries to present: Descriptive statistics (mean, variation, range) for at least the six influential variables. % variance explained on each axis, correlation coefficients of variables with axes.

Graphics to present: Unrotated PCA ordination diagrams for Axis 1×2 and Axis 1×3 showing overlays of at least the six most influential variables.

The PCA output indicated that four axes 'passed the Monte Carlo test' and were statistically unusual. However, the fourth axis contributed only an additional 8.75% of variation explained. Since the objective was to obtain only a few summary variables, three was deemed a sufficient number of axes.

Result Example1

***** PRINCIPAL COMPONENTS ANALYSIS*****

VARIANCE EXTRACTED, FIRST 10 AXES

AXIS	Eigenvalue	% of Variance	Cum.% of Var.	Broken-stick Eigenvalue
1	5.589	23.287	23.287	3.776
2	4.809	20.037	43.324	2.776
3	3.043	12.679	56.003	2.276
4	2.100	8.750	67.254	1.943
5	1.619	6.745	73.998	1.693
6	1.482	6.174	80.173	1.493
7	1.215	5.061	85.234	1.326
8	0.963	4.014	89.248	1.183
9	0.655	2.728	91.976	1.058
10	0.581	2.420	94.396	0.947

RANDOMIZATION RESULTS
999 = number of randomizations

Axis	Eigenvalue from real data	Eigenvalues from randomizations			p *
		Minimum	Average	Maximum	
1	5.5889	2.7158	3.3339	4.1408	0.001000
2	4.8089	2.3782	2.8407	3.4030	0.001000
3	3.0429	2.0926	2.4839	2.9793	0.001000
4	2.1002	1.8611	1.9012	2.1384	0.001000
5	1.6188	1.6056	1.9460	2.2964	0.999000
6	1.4818	1.4278	1.7214	2.0758	0.992000
7	1.2146	1.2347	1.5221	1.9074	1.000000
8	0.96345	1.0415	1.3359	1.6967	1.000000
9	0.65481	0.94322	1.1697	1.4202	1.000000
10	0.58078	0.80476	1.0190	1.2545	1.000000

Main Matrix overlays indicated six particular variables (one shown for each axis here) with strong linear associations with the three ordination axes; i.e., showed which variables contributed most strongly to the redundant pattern summarized by these three PCA axes.

The final ordination diagrams for presentation included overlays of the six particularly influential variables, which were put in a second matrix to facilitate use of the joint-plot view for simultaneous overlay.

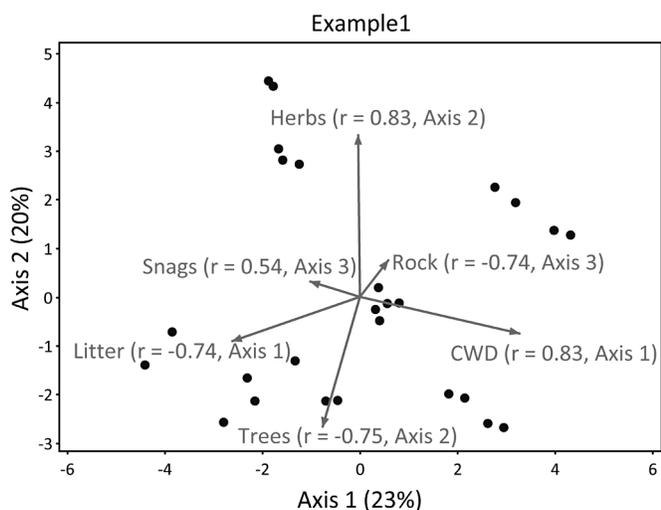
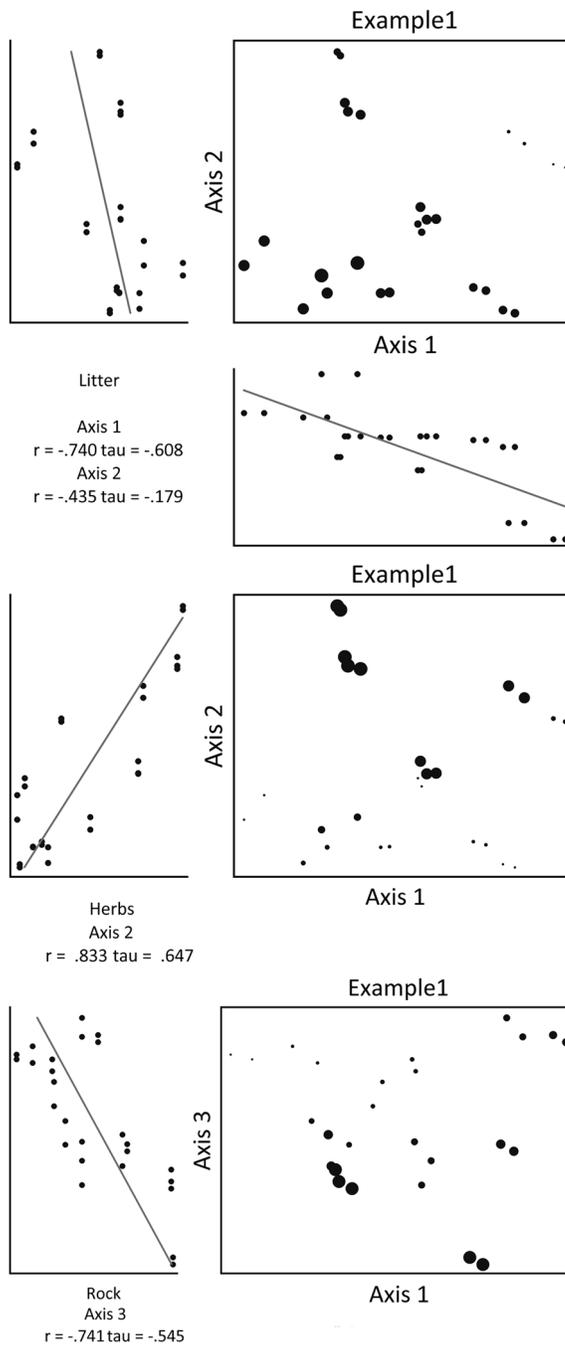


Figure 5. PCA ordination diagram showing associations of the six most influential variables defining the three synthetic summary axes. Vectors indicate the direction of increasing abundance and their length reflects the magnitude of the association with the ordination axes.

The ordination axes (i.e., three summary variables) were later used as independent predictor variables in habitat modeling (using *HyperNiche*).



Example 2: Terrestrial Habitat Delineation

Objective: Identify distinct habitat zones based on a combination of structural vegetation variables and species assemblages.

Main Matrix structure: 36 sites × 68 variables

Main Matrix content: Mixture of continuous variables (Q variables)

Second Matrix structure: None

What was measured? One-time measure of forest stand metrics (many different units) plus species abundance data for three strata of vascular vegetation (herb, shrub, and tree layers), recorded as percent cover for herbs, density for shrubs, and basal area for trees. In this case the vegetation data are viewed as habitat variables rather than as responses to the environment.

What do the zeros mean? Slightly zero-rich; most zeros mean zero.

Non-zeros: Made more comparable by centering and standardizing.

Heterogeneity: Data are not very heterogeneous following relativization: 0% zeros, average skewness of variables = 0.8.

Sources of noise: Stochastic variability in species abundance, measurement error.

Outliers: Three slight site outliers were identified and retained for observation.

Distance measure: Euclidean because it tolerates negative values resulting from centering.

Transformations? No transformations were deemed helpful.

Relationships? No theoretical expectations.

Model form: No one model form will fit all variables, so nonparametric preferred.

Tool to use: Cluster analysis combined with NMS (free ordination).

First look: Ran cluster analysis using Euclidean distances and group-average linkage, wrote several different grouping variables to the Second Matrix, ran NMS twice on Autopilot set to slow & thorough using Euclidean distances, selected a 3-D solution, observed clean separation in ordination space of six of the groups identified by clustering.

Confirmation: Ran cluster analysis using Euclidean distances and flexible beta = -0.25 linkage and obtained comparable groupings. Ran polar ordination using objective (variance-regression) endpoint selection and again observed separation of at least six groups.

Tool to report: Cluster analysis and NMS. Re-ran the cluster analysis and the NMS, saving the Column Graph and Results Files this time and then saving those with the Main Matrix as a Project File.

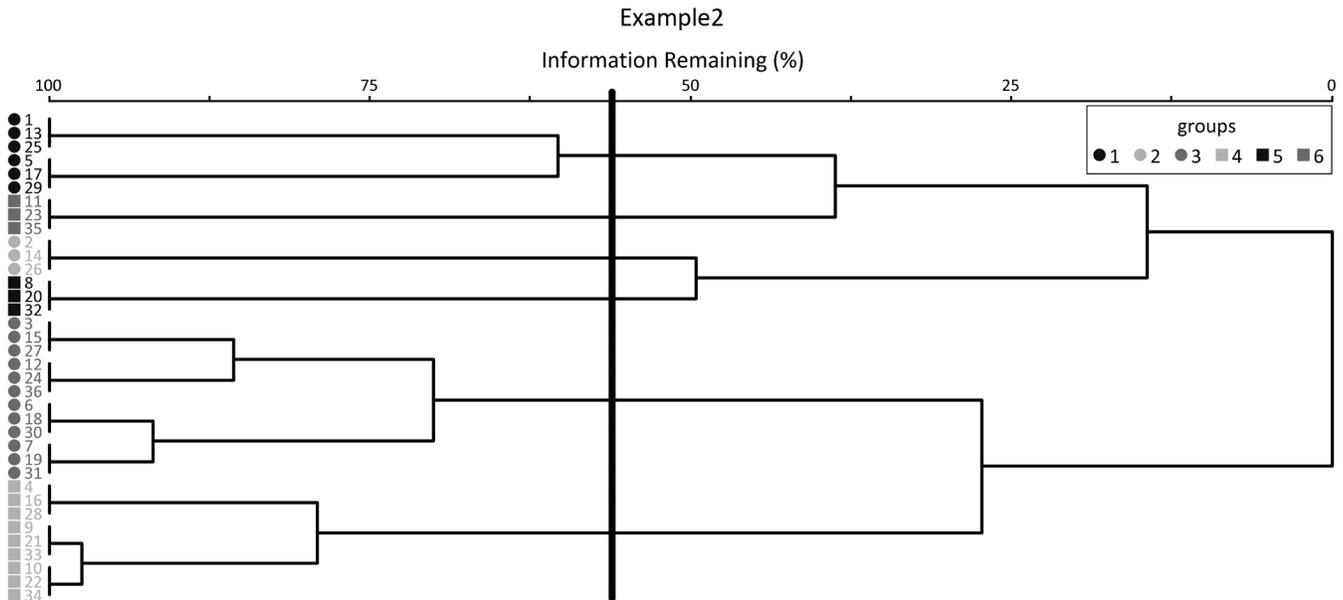
Interpretation: Graphed both the dendrogram and ordination with an overlay of the grouping variable showing six groups. Graphed the ordination with an overlay of the main matrix to explore which variables/species were most influential in defining these habitats. Applied Indicator Species Analysis to determine which species were most frequent and abundant in the different groups. Calculated individual univariate *F*-ratios using Sum*F* to determine which of the non-species variables differed most strongly among groups.

Story: 68 biotic and abiotic variables were used to delineate six habitats.

Summaries to present: Descriptive statistics (mean, variation, range) for each variable in each group. Indicator Species Analysis IV and p-values.

Graphics to present: Dendrogram showing separation of groups, NMS ordination diagram with overlay of grouping variable and most influential habitat variables.

The dendrogram indicated considerable separation (determined from the long horizontal lines) among several groups. After considering the strength of separation, verifying separation in the ordination space, and examining additional site data (not shown), six groups were chosen.



Indicator species analysis helped identify the sixteen species that were statistically significantly (all $p < 0.01$) more frequent and abundant in some groups than others. For instance, ferns were most frequent and abundant in groups 1, 3, and 6, whereas balsam fir (ABIBAL) was most frequent and abundant in group 2. An IV difference among groups of at least 20 was chosen as a cutoff for determining if a species 'indicated' a particular group. Most species indicated two to three groups, all groups had some indicator species, and group 3 had the most indicator species (12). The non-species variables with larger than average univariate F -ratios included those characterizing the abundance of conifers and yellow birch and the weight of the soil O-horizon.

Result

***** Indicator Values *****

INDICATOR VALUES

Column	Sequence:	Group						
		1	2	3	4	5	6	
1 ABIBAL	16 77	2	8	77	6	4	1	2
4 allies	16 50	4	4	1	41	50	0	0
5 ARANUD	16 43	3	2	17	43	16	18	2
6 ASACAN	16 93	11	3	0	1	0	0	93
7 ASTMAC	17 57	3	1	3	57	7	30	1
8 ATHFIL	17 36	11	30	0	24	9	0	36
10 CLACAR	16 38	1	38	17	8	1	0	32
11 CLIBOR	17 62	3	8	12	62	16	1	1
12 DRYSPI	16 59	3	32	1	59	2	0	2
13 ferns	17 36	3	27	0	36	10	0	27
14 graminoi	17 39	8	14	11	23	7	39	7
15 LYCANN	15 59	4	5	1	23	59	0	0
16 MAICAN	17 32	3	11	21	32	14	20	1
17 MOSS	17 40	3	4	28	40	17	6	5
20 STRROS	17 24	11	20	18	19	11	7	24
21 VIOSPP	17 31	8	23	4	27	10	31	5

The accompanying NMS ordination diagram was used to visually demonstrate the separation of groups, which are shown here using an overlay of convex hulls to 'connect the dots' of the ordination space defined by each group. In addition, some of the most influential habitat variables are shown using a joint-plot overlay.

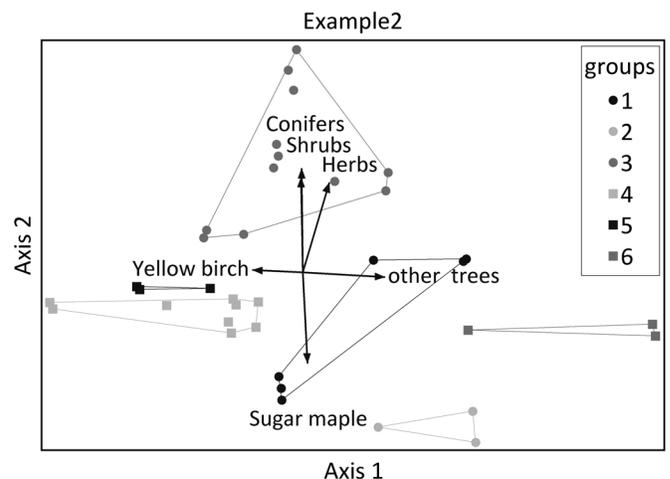


Figure 2. NMS ordination diagram showing the separation of the six chosen groups in the ordination space and the associations of six of the most influential habitat variables. Vectors indicate the direction of increasing abundance and the magnitude of the association with the ordination axes.

Example 3: Water Quality Gradient Analysis

Objective: Characterize the change in macroinvertebrate composition along a known gradient in water quality.

Main Matrix structure: 6 sites × 16 macroinvertebrate species

Main Matrix content: Total (sum) macroinvertebrate density counts (Q variables)

Second Matrix structure: None

What was measured? One-time measure of species captured in subplots, summed to the site-level; comparable units but varying life-history traits.

What do the zeros mean? Zeros do not necessarily mean zero.

Non-zeros: Made comparable among species by applying a relativization by column maxima.

Heterogeneity: Data are slightly heterogeneous following relativization: 18% zeros, average skewness of species = 1.5, coefficient of variation of site totals = 53%, CV of species totals = 28%.

Sources of noise: Stochastic variability in species abundance.

Outliers: No outliers were identified.

Distance measure: Sørensen because of skewness and comparable meaning of zeros.

Transformations? No transformations were deemed necessary.

Relationships? Relationships between species and the ordination axes may be unimodal or highly variable.

Model form: No one model form will fit all species.

Relativization? Relativized by species maxima to equalize influence of varying life-history traits and equalize the influence of common and rare species due to potential sampling bias against rare species.

Tool to use: polar ordination (guided ordination).

First look: Ran polar ordination using Sørensen distances and subjectively selecting the first axis (identifying as endpoints the sites with the greatest and least water quality); one additional axis was calculated using objective (variance-regression) endpoint selection. Observed strong associations of several species with both ordination axes.

Confirmation: Ran NMS on Autopilot set to slow & thorough using Sørensen distances and again observed associations with these species.

Tool to report: Polar ordination. Re-ran the polar ordination and saved the Column Graph and Results Files and then saved those with the Main Matrix as a Project File.

Interpretation: Took the percentage of variance from the Result File, graphed the ordination using Main Matrix Overlay, identified the most influential species, and verified linear relationships of unrelativized species' abundances to axes to ensure validity of calculated correlation coefficients (no transformations were necessary).

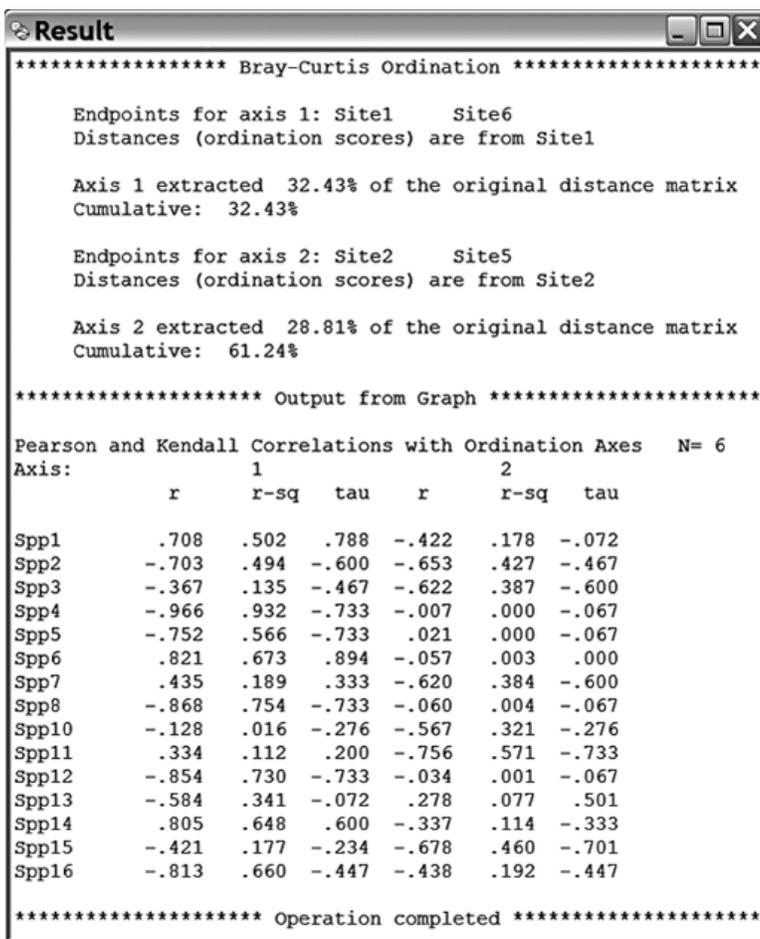
Story: Several macroinvertebrate species declined, and other increased, in abundance along the gradient from pristine to polluted water quality. The species associated with Axis 2 were all warm-water species, stimulating further questions regarding riparian vegetation management.

Summaries to present: Total abundance of species by site. Correlation coefficients for abundance of each species along the water quality gradient (i.e., with Axis 1). % variance explained for each axis.

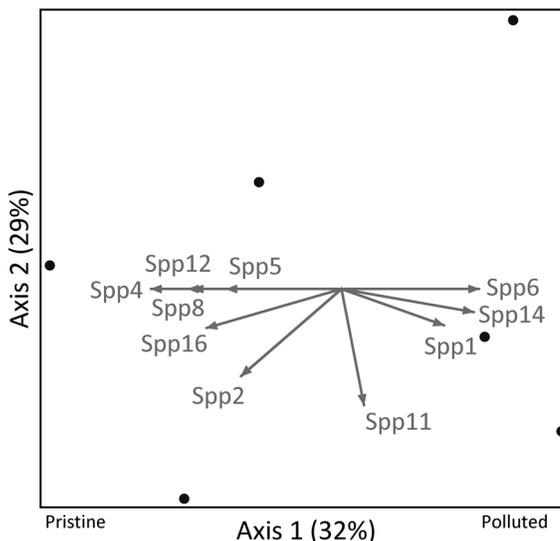
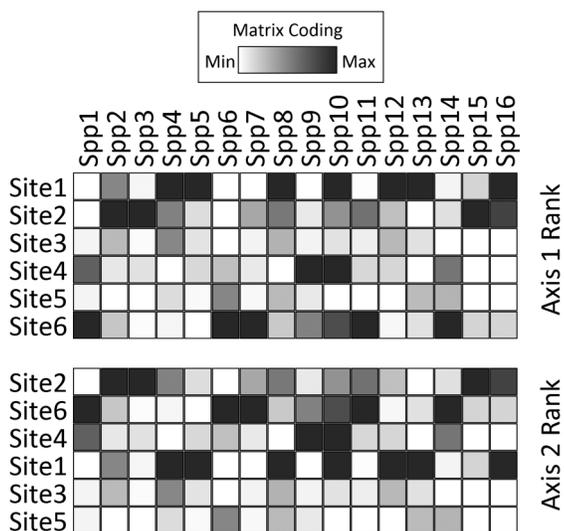
Graphics to present: Polar ordination diagram with an overlay of at least some of the species (with a table listing correlation coefficients) and/or an ordered main matrix showing abundance patterns of species along the both ordination axes.

The % variance explained along the water quality gradient is that for the subjectively defined Axis 1. An additional 29%, captured on Axis 2, must represent another major trend in the macroinvertebrate data.

Examining correlations of the individual species with each axis helps clarify these trends. Three species had higher abundances in the more polluted sites, while at least seven species were more abundant in the more pristine sites. In addition, the gradient on Axis 2 reflects a notable decrease in abundance among at least six species with no species showing an increase. Considering the habitat requirement of these species, it was hypothesized that this gradient may reflect stream temperature; a hypothesis that can be tested in future sampling.



These patterns can be graphically displayed in at least two ways. Here two ordered Main Matrices are coded by 'relative values by column' to show the distribution of abundance of each individual species among the six sites (this figure was created by cutting and pasting (in Photoshop) two figures that had been created by ordering the Main Matrix by both axes). The ordination diagram shown here used a joint-plot overlay of unrelativized species abundances read from a Second Matrix.



Example 4: Descriptive Biogeography

Objective: Characterize the variation in slime mold species composition across the landscape.

Main Matrix structure: 58 sites × 27 slime mold species

Main Matrix content: Slime mold presence/absence (Q variables)

Second Matrix structure: 58 sites × 44 environmental variables

Second Matrix content: Mixture of discrete, continuous, and nominal variables (C and Q variables)

What was measured? One-time measure of slime mold species presence; comparable units and life-history traits.

What do the zeros mean? Very zero-rich; zeros do not necessarily mean zero.

Non-zeros: Presence values are comparable by design.

Heterogeneity: Data are moderately heterogeneous: 68% zeros, average skewness of species = 2.9, coefficient of variation of site totals = 25%, CV of species totals = 90%.

Sources of noise: Stochastic variability in species presence, measurement error.

Outliers: Three strong outliers were identified but retained under observation.

Distance measure: Sørensen because data are zero-rich and outliers influential.

Transformations? No transformations were deemed helpful.

Relationships? Relationships between species and the ordination axes may be unimodal or highly variable.

Model form: No one model form will fit all species, so nonparametric preferred.

Relativization? No relativizations were deemed helpful.

Tool to use: NMS (free ordination).

First look: Ran NMS on Autopilot set to slow & thorough using Sørensen distances several times but unable to extract a non-random structure (i.e., 'failed' Monte Carlo test, all $p > 0.1$). Noticed that 16 of the 27 species only occurred in 1 or 2 of the 58 sites. Deleted all species occurring in only 1 site and ran NMS again; still no structure. Deleted all species occurring in only 2 sites and ran NMS again; a 2-D solution was recommended. Final matrix then had 58 sites x 11 species.

Confirmation: Ran NMS on Autopilot set to medium using Sørensen distances three more times; all recommended 2-D solution. Compared all three graphs and found adequate consistency.

Tool to report: NMS. Re-ran the NMS manually for 2-D and saved the Column Graph and Results Files and then saved those with the Main Matrix as a Project File.

Interpretation: Graphed the ordination using Main Matrix Overlay, identified the most influential species, verified linear relationships to axes to ensure validity of correlation coefficients, calculated the After-the-fact % variance explained. Graphed the ordination using Second Matrix Overlay and checked linearity, calculated correlation coefficients. Identified moderate associations with seven environmental variables.

Story: The eleven slime molds that are sufficiently abundant on the landscape to analyze varied in composition across the landscape. These patterns of variation were associated with variation in seven environmental variables that capture trends in moisture.

Summaries to present: Table of species' constancy. After-the-fact % variance explained for the relevant axes. Correlation coefficients of environmental variables with ordination axes.

Graphics to present: NMS ordination diagram with an overlay of the seven environmental variables.

Example 5: Management Treatment Effects

Objective: Test for compositional differences among treatments and determine which species, if any, are more frequent and abundant in one treatment than another.

Main Matrix structure: 16 sitextrt (4 treatments in 4 sites) × 276 herbaceous plant species

Main Matrix content: Percent cover (Q variables)

Second Matrix structure: 16 sitextrt × 2 coding variables

Second Matrix content: Coding variable for SITE and another for TRT (C variables)

What was measured? One-time measure of plant species captured in subplots, averaged to the sitextrt level; adequately comparable units and life-history traits.

What do the zeros mean? Fairly zero-rich; zeros do not necessarily mean zero.

Non-zeros: Cover is adequately comparable by design.

Heterogeneity: Data are very heterogeneous: 95% zeros, average skewness of species = 3.7, coefficient of variation of site totals = 65%, CV of species totals = 480%.

Sources of noise: Stochastic variability in species abundance, variability in application of treatments, measurement error.

Outliers: No outliers were identified.

Distance measure: Sørensen because data are zero-rich and heterogeneous.

Transformations? No transformations were deemed helpful.

Relativization? Relativized by species maxima within sites to reduce overall heterogeneity and make the influence of individual species more equal despite variability in abundances (i.e., downweight dominants).

Tool to use: PerMANOVA for randomized complete blocks (RBC), Indicator Species Analysis.

First look: Ran PerMANOVA for RBC with SITE as the first (block) factor and TRT as the second (fixed) factor, using Sørensen distances. Both SITE and TRT were marginally significant but their interaction was not. Indicator Species Analysis showed that most species had non-significant IVs. Noted that 87 of the 276 herb species only occurred in a single sitextrt; deleted those that only occurred in one of the two sites. Final matrix down to 16 sitextrt x 239 species.

Tool to report: PerMANOVA, Indicator Species Analysis. Re-ran the PerMANOVA including pair-wise comparisons. Ran Indicator Species Analysis on TRT.

Interpretation: Compositional differences among TRT now more significant, while those for SITE are now less. 98 herb species had significant IVs, 42 of which were notably greater in one treatment than another.

Story: Herbaceous species composition differed among treatments, with 12 species most frequent and abundant in treatment A, 28 species most frequent and abundant in treatment B, and 2 species most frequent and abundant in treatment C; no species indicated treatment D.

Summaries to present: Average cover of species by treatment. PerMANOVA *F*- and *p*-values, indicator species analysis IV and *p*-values for 42 species.

Graphics to present: An NMS ordination could be performed to supplement this analysis with a diagram showing separation of the treatment groups in species space if desired.

Example 6: New Species Delineation

Objective: Determine whether or not two populations of freshwater guppies are sufficiently morphologically different to warrant investigation to determine species status.

Main Matrix structure: 72 fish × 8 morphological variables

Main Matrix content: various measures of body dimensions (Q variables)

Second Matrix structure: 72 fish × 1 coding variable

Second Matrix content: coding variable for fish population, POP (C variable)

What was measured? Collection of morphological measures taken from fish captured from two ends of a large lake over a period of a year.

What do the zeros mean? No zero values.

Non-zeros: Made comparable among fish by applying a relativization by fish standard deviates.

Heterogeneity: Data are not very heterogeneous following standardization: 0% zeros, average skewness of fish = 0.8, coefficient of variation of site totals = 21%, CV of fish totals = 67%.

Sources of noise: Measurement error.

Outliers: No outliers were identified.

Distance measure: Euclidean because it should perform well on these data and tolerates negative values resulting from the relativization.

Transformations? No data transformations were deemed helpful.

Relativization? No additional relativization was performed.

Tool to use: MRPP due to unequal sample sizes.

First look: Ran an MRPP on POP using Euclidean distances and observed a small p -value but also a relatively small A value.

Tool to report: MRPP. Recorded the p -value and A values.

Interpretation: Although there is evidence to suggest that these two populations of guppies are morphologically distinct from one another, there is a considerable amount of variation within each population.

Story: There is sufficient evidence to indicate that these two populations represent morphologically distinct groups of individuals. The high heterogeneity within each group, however, sheds doubt on the proximity of their lineage. Further analyses are warranted.

Summaries to present: Descriptive statistics (mean, variation, range) for each variable in each population. MRPP p - and A values.

Graphics to present: Boxplots of each population might be useful summary graphs for display. A cluster analysis within each of the two populations might be helpful to further explore within-group heterogeneity. PCA could be used to graphically display both between- and within-group variation.

Example 7: Spatial Systematics

Objective: Determine if spatial proximity implies genetic proximity in invasive species (i.e., representing only one versus multiple introduction events).

Main Matrix structure: 46 earthworms × 12 genetic markers

Main Matrix content: presence/absence (Q variables)

Second Matrix structure: 46 earthworms × 12 spatial locations

Second Matrix content: easting and northing UTM coordinates (Q variables)

What was measured? Presence of specific genetic markers were determined for invasive earthworms whose precise location of capture was recorded using GPS.

What do the zeros mean? Main Matrix: some zeros, zeros are meaningful. Second Matrix: no zeros.

Non-zeros: Comparable within each matrix by design.

Heterogeneity: Main Matrix: data are slightly heterogeneous: 40% zeros, average skewness of markers = 1.8, coefficient of variation of worm totals = 24%, CV of markers = 56%. Second Matrix not heterogeneous: 0% zeros, skew 0.4, CVs 6% and 14%, respectively.

Sources of noise: Stochastic variability in species presence

Outliers: 1 outlier worm was identified and determined to be mis-identified and omitted.

Distance measure: Sørensen for the zero-rich Main Matrix, Euclidean for the Second matrix.

Transformations? No transformations were deemed helpful.

Relativization? No relativizations were deemed helpful.

Tool to use: Mantel Test (group testing).

First look: Ran the Mantel Test using Sørensen distances for the Main Matrix and Euclidean distances for the Second Matrix and Mantel's asymptotic approximation.

Tool to report: Mantel Test. Recorded the r - and p -values.

Interpretation: The highly non-significant ($p = 0.4$) Mantel Test does not indicate a correspondence between the spatial distribution of the earthworms and their genetic similarity.

Story: The lack of spatial and genetic correspondence is consistent with either a single introduction or a long period of genetic mixing following multiple introductions. The short time frame since invasion would indicate the former.

Summaries to present: The geographic range under consideration, the frequency of each genetic marker, the Mantel Test r - and p -values.

Graphics to present: A cluster analysis of the main matrix could supplement this analysis and provide additional detail on population variability (i.e., the presence or absence of subpopulations).

Example 8: Predicted Management Impacts

Objective: Predict the change in mussel species composition with an increase in water flow rate below a dam.

Main Matrix structure: 20 trial×flow (5 reps of 4 flow rates) × 22 mussel species

Main Matrix content: mussel density counts (Q variables)

Second Matrix structure: 20 trial×flow × 1 variable

Second Matrix content: flow rate (Q variable)

What was measured? Mussel densities in one river, downstream of four dams with different flow rates. Site and rate are unavoidably confounded. Species vary considerably in their life-history traits.

What do the zeros mean? Very zero-rich; zeros do not necessarily mean zero.

Non-zeros: Non-comparable abundances of species (due to varying life history traits) standardized by applying a general relativization by column (mussel) totals within sites.

Heterogeneity: Data are still somewhat heterogeneous following relativization: 53% zeros, average skewness of mussels = 0.9, coefficient of variation of trial×flow totals = 36%, CV of mussel totals = 58%.

Sources of noise: Stochastic variability in species abundance.

Outliers: One notable trial×flow outlier was observed. RA revealed that this observation was highly influential in shaping the ordination space. The observation, which had very few species, low total abundance, and very high skewness, was omitted from the analysis.

Distance measure: An examination of species' distributions indicated that almost all were unimodally distributed and therefore the Chi-squared distance measure of CCA is appropriate.

Transformations? No transformations were deemed necessary.

Relationships? Relationships between the most abundant mussels and flow rate are largely unimodal.

Model form: Unimodal is adequate.

Relativization? The general relativization tended to equalize common and uncommon species.

Tool to use: CCA (guided ordination)

First look: Ran a CCA using centering and normalizing, optimizing columns, and using LC scores for graphing.

Tool to report: CCA. Saved Result File and Graph File and then saved with Main Matrix as a Project.

Interpretation: The significant randomization tests indicate that the Main and Second Matrices do covary. The flow rate variable had a large canonical coefficient and biplot score on Axis 1.

Story: Although true replication is impossible, the strong association of the flow rate variable indicates that this is potentially a very influential factor affecting mussel composition. The directionality of the canonical coefficient for flow rate on Axis 1 indicates that increasing the flow rate will shift the mussel composition toward the right of the species space and the magnitude of the value indicates that this shift will be considerable. The final scores, and correlation coefficients, indicate that several species would be affected by changes in flow rate.

Summaries to present: Descriptive statistics (mean, variation, range) for each mussel across the entire dataset. CCA % variance explained, canonical coefficient for flow (and std), randomization test p-values.

Graphics to present: CCA ordination diagram with a biplot showing species and flow rate as vectors indicating the directionality of their association.

Result

***** Canonical Correspondence Analysis *****

AXIS SUMMARY STATISTICS

	Axis 1	Axis 2	Axis 3
Eigenvalue	0.430	0.128	0.089
Variance in species data			
% of variance explained	34.0	8.3	5.1

RANDOMIZATION TEST RESULTS -- EIGENVALUES FOR INDIVIDUAL AXES

Axis	Real data	Randomized data			p
	Eigenvalue	Mean	Minimum	Maximum	
1	0.430	0.277	0.083	0.458	0.0010

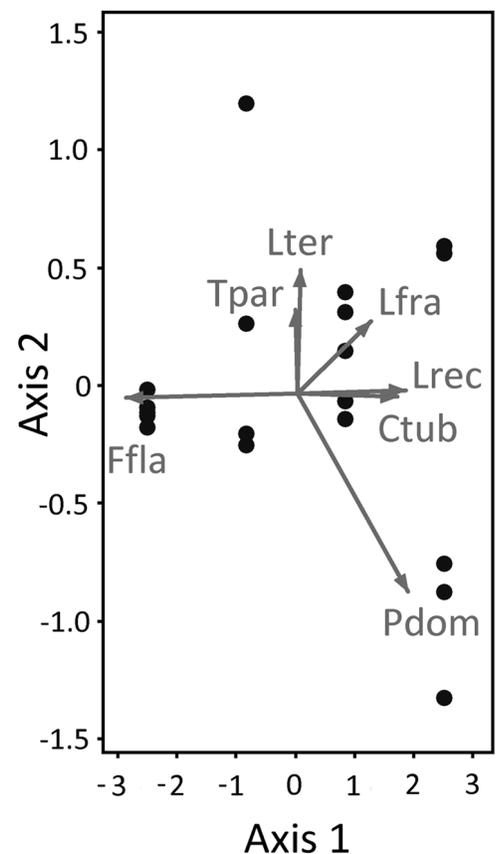
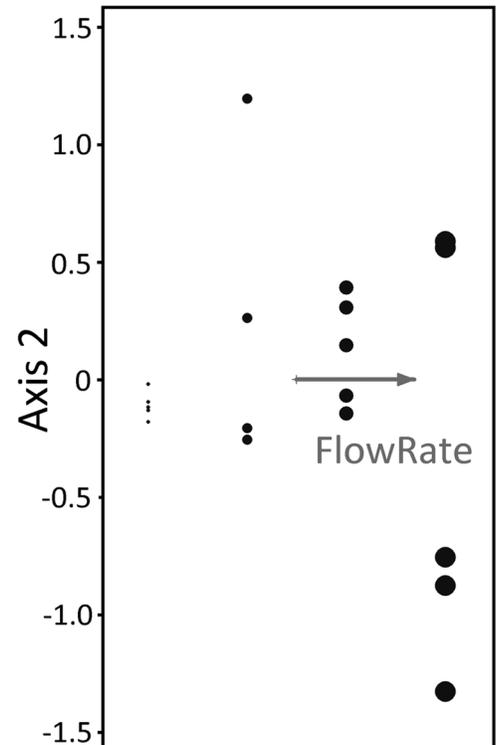
RANDOMIZATION TEST RESULTS -- SPECIES-ENVIRONMENT CORRELATIONS

Axis	Real data	Randomized data			p
	Spp-Envr Corr.	Monte Carlo test, 998 runs	Mean	Minimum	
1	0.899	0.716	0.470	0.923	0.0060

The CCA output indicated that a considerable amount of variation was associated with the FlowRate axis (34% on Axis 1) and that the extracted pattern was statistically significant ($p = 0.001$). Further, the association between the FlowRate variable and the variation in species abundances ($r = 0.899$) was also statistically unusually strong ($p = 0.006$). These results confirm that the FlowRate variable can be used as a predictor of species composition.

An overlay of the FlowRate variable onto the CCA ordination diagram (top) showed the correspondence between the four flow rates of each dam and the Axis 1 ordination scores for samples taken below each of those dams. Because flow rate alone does not explain all of the variation in the mussel community, samples have variable scores along Axis 2, as indicated by the vertical spread in the diagram.

The coefficients for predicting changes in mussel species abundances (i.e., 'Final scores' for species) were listed in the Result File. An overlay of mussel species' abundances onto the ordination diagram (bottom) provides a graphical representation of those relationships. Here we can see that *Ffla* had its highest abundance below the low-flow rate dams, and thus is most likely to decrease in abundance as flow rate increases. In contrast, *Pdom*, *Lrec*, *Ctub*, and *Lfra* showed the opposite pattern.



Example 9: Forest Structure over Time

Objective: Track changes in forest composition over time under three different management regimes.

Main Matrix structure: 12 timextrt (4 periods after 3 treatments) × 25 structural variables

Main Matrix content: mixture of discrete, continuous, and nominal variables (Q & C variables)

Second Matrix structure: 12 timextrt × 2 coding variables

Second Matrix content: 1 code for TIME, 1 for TRT (C variables)

What was measured? Prior to treatment and then once every 5 years, a variety of structural characteristics (with differing units) were measured for different forest strata in several subplots at one site. Subplot values averaged to timextrt. Highly redundant variables ($r > 0.9$) were pared down to a single variable.

What do the zeros mean? Few zeros, most of which mean zero.

Non-zeros: Made comparable by centering and standardizing.

Heterogeneity: Data are not very heterogeneous following relativization: 2% zeros, average skewness of variables = 0.8, coefficient of variation of variable totals = 48%.

Sources of noise: Measurement error, temporal and stochastic variability in response abundance.

Outliers: No outliers were identified.

Distance measure: Euclidean will perform well on these data and tolerate negative values resulting from centering.

Transformations? Two variables were log+1 transformed prior to relativization to improve normality.

Relationships? Relationships between response variables and time and treatment may be linear, unimodal, or polymodal.

Model form: Either linear or nonparametric.

Relativization? The standardization tended to equalize variables of varying abundance.

Tool to use: PCA or NMS (free ordination), PerMANOVA (group testing)

First look: Ran a PCA using correlation and obtained one significant and interpretable axis. Ran NMS on Autopilot using Euclidean and later Sørensen and neither produced very interpretable results. Ran PerMANOVA using the Euclidean distance measure with a randomized complete block design (TIME & TRT) and observed significant differences across TIME.

Tool to report: PCA, PerMANOVA. Saved Graph File, saved Result File from PCA, appended Result File from PerMANOVA, and then saved all with Main Matrix as a Project.

Interpretation: A graph of the ordination diagram using successional vectors (row pattern) showed grouping of TIME1 (pre-treatment) toward the left of Axis 1. Vectors then extended in slightly different but inconsistent directions, with different lengths, for each TRT for the next two TIME, and reconverged for the final TIME. PerMANOVA verified differences across TIME but not among TRT.

Story: Resilience in forest structural composition was observed in that treatment effects were short-lived.

Summaries to present: Descriptive statistics (mean, variation, range) for each variable across each treatment over time. PCA % variance explained for each relevant axis. Also (outside PC-ORD), repeated-measures ANOVA results for a contrast of vector lengths among treatments.

Graphics to present: PCA ordination diagram with successional vectors connecting time periods.

The PCA results indicated that one statistically significant ($p = 0.001$) axis was available for interpretation, which explained the majority of variation in the data (84%).

This axis corresponded to a time gradient, with structural composition varying linearly from Time 1 (pre-treatment) to Time 2 and Time 3. The structural composition at Time 4 was more similar to Time 1 & 2 than Time 3, which is consistent with recovery over time. No pattern with TRT was seen. The variation in Time, and the lack of pattern with respect to TRT, was verified in the PerMANOVA, which indicated statistically unusual differences in structural composition across time periods ($p = 0.017$) but not among treatments ($p = 0.6$).

The challenge with this project was determining how to best graphically represent this information.

***** PRINCIPAL COMPONENTS ANALYSIS *****			
VARIANCE EXTRACTED, FIRST 10 AXES			
AXIS	Eigenvalue	% of Variance	Cum.% of Var.
1	21.114	84.457	84.457
2	2.942	11.767	96.224

RANDOMIZATION RESULTS					
999 = number of randomizations					
Axis	Eigenvalue real data	Eigenvalues from randomizations			p
		Minimum	Average	Maximum	
1	21.114	4.0229	5.0581	7.0813	0.001
2	2.9418	3.3101	4.1284	5.1633	1.000

***** Permutation based MANOVA (PerMANOVA) *****					
Source	d.f.	SS	MS	F	p
TRT	2	9.5445	4.7722	0.64494	0.648600
TIME	3	221.06	73.686	9.9584	0.016600
Residual	6	44.397	7.3994		
Total	11	275.00			

Because the PCA solution obtained only one statistically significant axis, the ordination 'space' consisted of a one-dimensional line. Although it was possible to present a horizontal axis only, a two-dimensional view was chosen in order to permit an overlay of a successional vector connecting the four TIME periods that would show the back-tracking of the vector on Axis 1 between Time 3 and Time 4. Because there was no treatment effect, the three treatments sampled in each time period were grouped together (with connecting lines forming a 'convex hull'). The second axis, in this case, represents random noise introduced on the Y-axis (a 'jiggered' horizontal axis). The ordination was also reflected such that time increased from left to right.

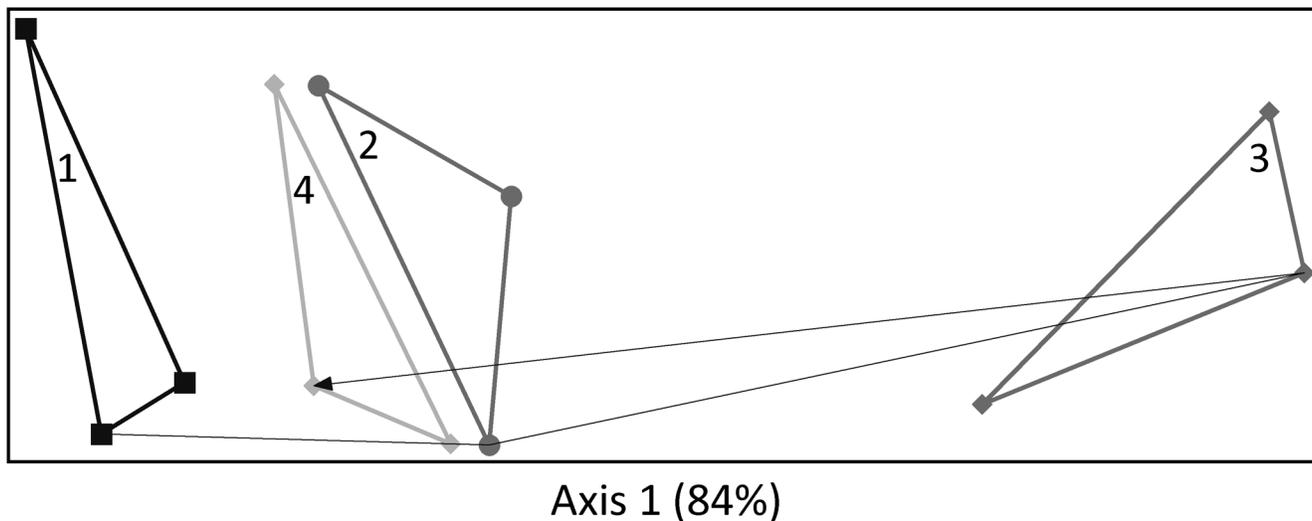


Figure 1. PCA summarized the variation in the structural characteristics into one dominant gradient. This one-dimensional ordination axis corresponded to a time gradient, shifting from left to right over the first three time periods. Each triangle represents the three treatments, which did not differ, at each of the four time periods. A single successional vector is shown indicating how structural composition shifted to the right along Axis 1 with time (time periods 1 to 2, then 2 to 3) but then reversed trajectory by the final time period (3 to 4). The Y-axis was jiggered to provide a two-dimensional view.

Example 10: Lichen Habitat Characterization

Objective: Quantify the *relative* influence of habitat type on lichen community composition.

Main Matrix structure: 27 sitexhab (9 reps of 3 habitats) × 175 lichen species

Main Matrix content: average cover class midpoints (Q variables)

Second Matrix structure: 27 sitexhab × 3 variables (1 C coding variable converted to 3 Q variables)

Second Matrix content: 1 code for GRND, 1 for BASE, 1 for BOLE (Q binary dummy variables)

What was measured? One time measure of species abundance in quadrats placed in three different habitat types at nine replicate sites. Subplot values averaged to sitexhab. Comparable units and life-history traits.

What do the zeros mean? Very zero rich; zeros do not necessarily mean zero.

Non-zeros: Made more comparable among species by applying a relativization by column maxima.

Heterogeneity: Data are still moderately heterogeneous following relativization: 66% zeros, average skewness of variables = 3.5, coefficient of variation of variable totals = 71%.

Sources of noise: Stochastic variability in species abundance, measurement error.

Outliers: No outliers were identified.

Distance measure: Sørensen because of skewness and noncomparability of zero values.

Transformations? No transformations were deemed helpful.

Relationships? Relationships between species and habitat type will be linear due to the binary nature of the predictor variables.

Model form: Either linear or nonparametric.

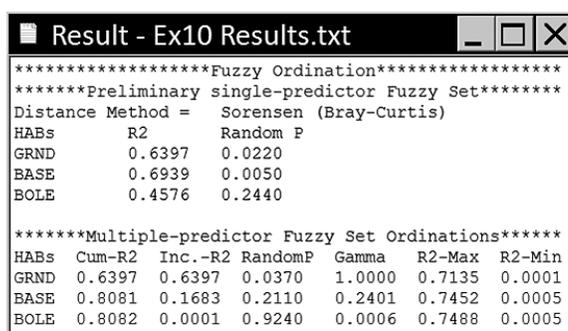
Relativization? The relativization by species maxima tended to equalize species of varying abundance.

Tool to use: FSO (guided ordination)

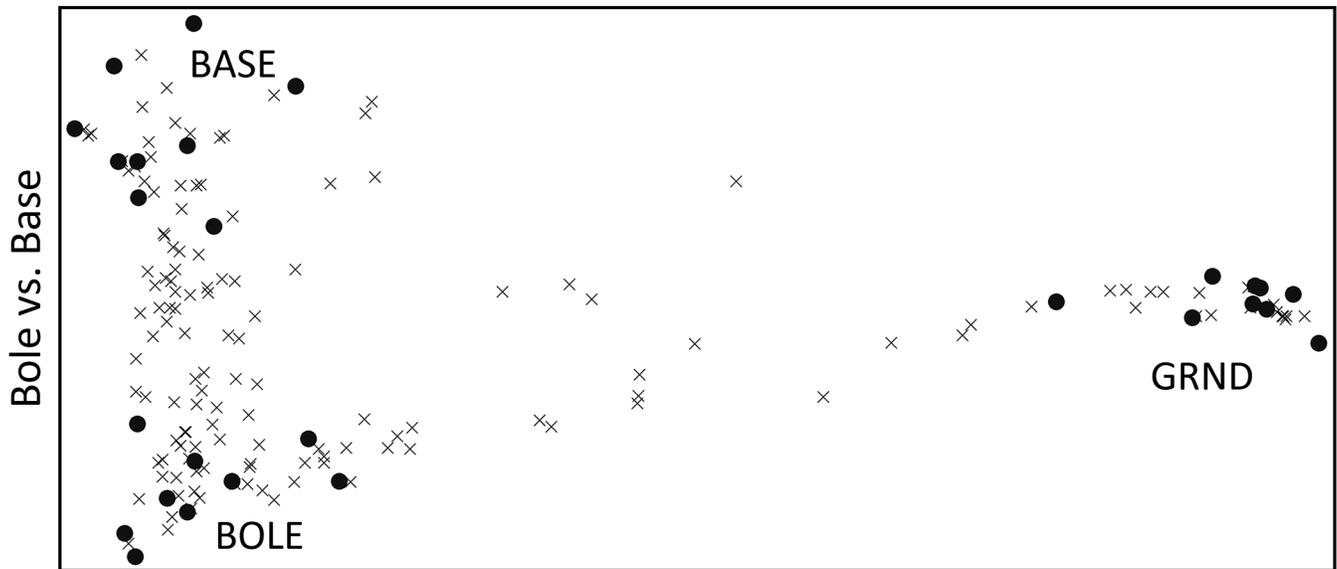
First look: Ran FSO using the Sørensen distance measure, the default order of predictor variables, and the randomization test for individual predictors and found that BASE explained a bit more variation than GRND. Re-ran FSO changing order to BASE, GRND, BOLE, but interpretation did not change so retained original order because it corresponded better to the ordination diagram. Ran NMS on Autopilot using Sørensen and confirmed strong separation by habitat type.

Tool to report: FSO. Saved the Graph File and Result File from FSO and then saved them with the Main Matrix and Second Matrix as a Project.

Interpretation: In this case, the ordination diagram (Figure 1 below) provides little more than visual confirmation of the output given in the Result file. [Note that when using continuous predictor variables, the ordination diagram has greater interpretational potential.] Here the diagram shows the same pattern was observed with free ordination, i.e. the separation of the three habitat types in lichen composition ordination space. The model results, however, permit a quantification of the relative importance of those differences. First, the single-predictor results show that the ground and base variables were nearly equally important, because they represent opposite ends of a conditional gradient. In this case, the third habitat type (bole) is essentially redundant to the second (base) and is therefore nonsignificant (this is an example of how multicollinearity renders all but one of the highly correlated variables nonsignificant). This is further reflected in the marginal incremental R2 for



```
Result - Ex10 Results.txt
*****Fuzzy Ordination*****
*****Preliminary single-predictor Fuzzy Set*****
Distance Method = Sorensen (Bray-Curtis)
HABs      R2      Random P
GRND      0.6397   0.0220
BASE      0.6939   0.0050
BOLE      0.4576   0.2440
*****Multiple-predictor Fuzzy Set Ordinations*****
HABs Cum-R2 Inc.-R2 RandomP Gamma R2-Max R2-Min
GRND 0.6397 0.6397 0.0370 1.0000 0.7135 0.0001
BASE 0.8081 0.1683 0.2110 0.2401 0.7452 0.0005
BOLE 0.8082 0.0001 0.9240 0.0006 0.7488 0.0005
```



Base/Bole vs. Ground

Figure 1. FSO summarization of lichen compositional differences related to habitat type. Axis 1 reflected the influence of the GRND variable and shows a strong contrast in composition between samples taken from the ground (right) and those taken from tree bases and boles (left). Axis 2 captured the lesser contrast between tree bases and boles.

BOLE in the multiple model. Second, the multiple-predictor results show that neither base nor bole is significant when accounting for the influence of ground, indicating that the distinction between lichen community composition on bases and boles is very weak but the contrast between bases/boles and the ground is quite strong and can be quantified with an R^2 of 0.64.

Story: Ground lichen community composition not only differs from epiphytic lichen community composition, but the relative distinction between the communities on tree bases vs. boles is marginal *compared to* the contrast between that on the ground vs. on trees. [Note that this is a relational and thus a relative contrast. You could use PerMANOVA to assess the absolute difference in species composition among the three habitat types.]

Summaries to present: Average cover midpoints of species by habitat type. FSO single and multiple predictor statistics (at least incremental Inc.-R2 and Random P for each predictor variable).

Graphics to present: Optionally, FSO ordination diagram showing contrast of habitat types.

Example 11: Harvesting Impacts on Vegetation by Tolerance

Objective: Evaluate the impact of overstory harvesting on dominant understory vegetation with respect to species' environmental tolerances (to moisture, nutrients, heat, and light).

Main Matrix structure: 32 sitextrt (4 reps of 2 treatments in 4 sites) × 15 herbaceous species

Main Matrix content: average percent cover (Q variables)

Second Matrix structure: 32 sitextrt × 10 environmental variables

Second Matrix content: mixture of discrete, continuous, and nominal variables (Q & C variables)

What was measured? One time measure of species abundance in quadrats placed in different treatment areas at different sites; comparable units and life-history traits.

What do the zeros mean? Many zeros; zeros do not necessarily mean zero.

Non-zeros: Non-zero values are made more comparable by relativizing by species maxima within each level of treatment.

Heterogeneity: Data are still somewhat heterogeneous following relativization: 41% zeros, average skewness of variables = 1.9, coefficient of variation of variable totals = 0.3%.

Sources of noise: Stochastic variability in species abundance, variability in treatment application, measurement error.

Outliers: Two sitextrt and one species variables were identified as moderate outliers.

Distance measure: N/A.

Transformations? No transformations were deemed helpful.

Relationships? Relationships among species, environmental variables, and traits may be linear, unimodal, or polymodal.

Model form: Nonparametric.

Relativization? The relativization tended to equalize variables of varying abundance.

Tool to use: FCA (association assessment)

First look: Ran FCA using several different randomization options, then chose Combined (Row first, then Column) as the most suited to the analysis objective. Re-ran FCA.

Tool to report: FCA. Saved Result File with all three Matrices as a Project.

Interpretation: The positive association (R) of NUTR with TRT1, and the negative association with TRT2 (both $p < 0.05$), indicate that the abundance of species with high nutrient scores decreased under TRT2 as compared to TRT1. No other significant relationships were observed, although the nutrient trait was suggestively positively associated with light availability (as measured by LAI).

Story: The more nutrient demanding herbaceous species were less abundant under the overstory harvest treatment as compared to the control.

Summaries to present: Descriptive statistics for each species by treatment. Relevant FCA R and p -values.

Result - RESULT.TXT							
***** Fourth Corner Problem *****							
Random model =		Combined row-column					
P Adjustment =		Benjamini & Hochberg					
***** Numerical Environment and Traits *****							
Env	Traits	R	P-1 tail	P-2 tail	AdjP-1 tail	AdjP-2 tail	
LAI	MOIST	0.053	0.2550	0.5450	0.9990	0.9990	
LAI	NUTR	0.550	0.0280	0.0430	0.0698	0.0743	
LAI	HEAT	-0.022	0.4180	0.8190	0.7837	0.9450	
LAI	LIGHT	-0.088	0.1770	0.3490	0.9990	0.9990	
BA	MOIST	-0.006	0.4810	0.9380	0.7800	0.9990	
BA	NUTR	0.040	0.3150	0.6630	0.8591	0.9945	
BA	HEAT	-0.046	0.2860	0.5320	0.8171	0.9990	
BA	LIGHT	0.030	0.3680	0.7340	0.7123	0.9990	
***** Categorical Environment Numerical Traits *****							
Sp	Trts	F	P	AdjP			
	MOIST	0.183	0.5360	0.9990			
	NUTR	2.847	0.0430	0.0761			
	HEAT	0.018	0.9410	0.9990			
	LIGHT	0.489	0.3520	0.9990			
Env	Sp	Trts	D	P	AdjP		
1		MOIST	0.443	0.2280	0.9990		
1		NUTR	0.468	0.3420	0.9990		
1		HEAT	0.468	0.3260	0.9990		
1		LIGHT	0.461	0.3480	0.9990		
2		MOIST	0.250	0.8180	0.9624		
2		NUTR	0.197	0.3340	0.9990		
2		HEAT	0.264	0.9430	0.9926		
2		LIGHT	0.165	0.1400	0.9990		
Env	Sp	Trts	R	P-1 tail	P-2 tail	AdjP-1 tail	AdjP-2 tail
1		MOIST	0.026	0.4030	0.7580	0.6909	0.9990
1		NUTR	0.460	0.0030	0.0110	0.0475	0.0892
1		HEAT	0.005	0.4620	0.9520	0.6930	0.9990
1		LIGHT	-0.129	0.0690	0.1540	0.6900	0.9990
2		MOIST	-0.078	0.1710	0.3240	0.9327	0.9990
2		NUTR	-0.331	0.0057	0.0142	0.0585	0.0947
2		HEAT	-0.024	0.3760	0.7670	0.6635	0.9990
2		LIGHT	0.080	0.1790	0.3590	0.8262	0.9990