


Using Multivariate Methods to Investigate Changes in Species Assemblages at Restoration Sites. An Introduction.

Lucinda Tear

Challenges in Monitoring Effectiveness

- Restoration action embedded in a landscape
 - Landscape affected by multiple factors besides restoration action
 - Costly and difficult to monitor all factors that affect restoration outcomes so that effect of restoration alone can be determined
- 

Analogy

- Patient with congestive heart failure
- Work with patient to affect: smoking, diet, drug regime, exercise, sleep, heart-related emotional problems
- How to determine effect of individual factors?

Options

- Change each factor, one by one, over time (takes a long time, correlation over time of treatments)
- Treat a variety of patients with one treatment, summarize results (may take less time, but very costly and results likely to be varied and hard to interpret and synthesize)

Can you really test all factors?

- If some people remain congested was it because some were exposed to
 - a flu virus
 - Second hand smoke
 - Increased stress from new sources
 - Withdrawal reactions



In short

- To determine best course of action and monitor results of treatment takes time and money
- Is it really possible to sort out effects of restoration actions and of complex ecological interactions on restoration actions unless we devote extensive resources?
- How can we draw useful inference from restoration monitoring?
- Multivariate methods can help evaluate multiple response variables, but good design and consideration of landscape context still important.

Types of Monitoring

- **Status:** Pre-treatment
- **Compliance:** Actions completed successfully?
- **Effectiveness:** Physical outcomes achieved?
Proximate effects, direct measures
- **Validation:** Ecological outcomes achieved? (Ultimate effects, indirect measures)

At all levels must consider:

- Spatial scale (grain and extent)
- Replication (within and between treatments)
- Temporal scale (frequency, duration, longevity)
- What level of each is needed to draw conclusions

As we have heard, must clearly answer:

- What are the goals and scope of your restoration action?
- What are the goals and scope of your monitoring program?
- Two separate questions.
- Priorities depend on monitoring resources.

Importance of Conceptual Model

- **Always** develop a conceptual model to get people on the same page and determine restoration and monitoring priorities. (Thom and Wellman 1996)



- From the conceptual model determine:
 - What are the information priorities?
 - Can these be made into direct, testable hypotheses?
 - Do you have the resources to measure something from each section of model to help interpret relationships?
 - If not, what can you measure and how will you have to interpret the data?

Types of Monitoring

- **Status:** Pre-treatment
- **Compliance:** Actions completed successfully?
- **Effectiveness:** Physical outcomes achieved?
Proximate effects, direct measures
- **Validation:** Ecological outcomes achieved?
Ultimate effects, indirect measures

Effectiveness Monitoring

- Testable hypothesis related to means
- Power test to determine sample size

$$n = \frac{2s^2 p}{d^2} \left(t_{\alpha, \nu} + t_{\beta(1)\nu} \right)^2$$

- Sample size (n) related to:
 - Variance of both populations (s^2_p)
 - Difference you want to detect (d)
 - Willingness to be wrong (α, β)

Effectiveness Monitoring

- Can convert formula so
- variance (s^2_p) is expressed as the CV and d is expressed as a fraction of the mean (f) or
- d is expressed as a number of standard deviations (s)

$$n \geq \frac{2s^2_p}{d^2} \left(t_{\alpha,v} + t_{\beta(1)v} \right)^2 = \frac{2CV^2}{f^2} \left(t_{\alpha,v} + t_{\beta(1)v} \right)^2$$

- Make graphs that show how n changes with assumptions about d , s , α , β

What if you are interested in relative abundances and pattern?

- **Fewer opportunities for testable hypotheses**
 - rely on simulation, best-professional judgement, results from other studies to determine if results from your project are different than what would be expected by chance alone
- **Need to use multiple analytical techniques**
 - Univariate, bivariate, multivariate techniques
 - “multimetric” techniques

Multiple techniques

- Approach data from multiple angles
 - **Exploratory univariate and bivariate graphs**
 - Histograms, cumulative frequency distributions, boxplots
 - Scatterplots, comparisons of univariate graphs
 - **Univariate and Bivariate “models” (parametric and nonparametric)**
 - ANOVA, regression, correlation, contingency tables, etc.
 - **Multivariate analyses (parametric and nonparametric)**
 - Multiple input variables, single response variable
 - Single input variable, multiple response variables
 - Multiple input variables, multiple response variables
 - Physical and biological variables

Multivariate Techniques

- Multiple input variables, single response

$$y = a + bx_1 + cx_2 + dx_3$$

$$y = a + b^x + c^{x^2}$$

Parametric and nonParametric forms of

- Multifactor ANOVA
- Multiple regressions
 - Generalized Linear, Nonlinear, Logistic
- Multiple Discriminant Function Analysis
- Regression and Classification Trees
 - Error variance (normal, nonnormal)
 - Shape of model (linear, curvilinear, additive, multiplicative)
 - Type of independent and dependent variable (binary, categorical, continuous)

Multivariate Techniques

- Multiple input variables, single response
 - Diversity indices (richness, alpha, beta, gamma diversity, evenness)
 - Cluster analyses (discover groups based on multivariate data)



Multivariate Techniques

- **Single input variable, multiple response**
 - Single-factor MANOVA
 - E.g., Do sites differ based on mean abundance of multiple species?
 - Discriminant Function Analysis
 - E.g., put data into predefined groups based on only one independent variable
 - Determine group and probability of membership in group

Multivariate Techniques

➤ Multiple input variables, multiple response variables

Ordination

- Indirect gradient analysis
 - PCA, NMDS, Bray Curtis, Weighted Averaging/multimetric, Correspondence Analysis/Reciprocal Averaging, Detrended Correspondence Analysis
- Direct Gradient Analysis
 - Gradient Analysis, Canonical Correspondence Analysis

A few simple guidelines in describing communities

- Ordinations and Cluster Analyses are helpful exploratory tools – they summarize “distances” between species or plots in multivariate space
- But, different:
 - clustering and ordination techniques /algorithms,
 - distance measures,
 - data relativisations and standardizationsbring out very different views of a data set.
- Always:
 - Try several approaches.
 - Observe which results are consistent, and which are brought out by a particular technique
 - Try to determine whether results that are unique to a method are due to inappropriate or appropriate application of the method

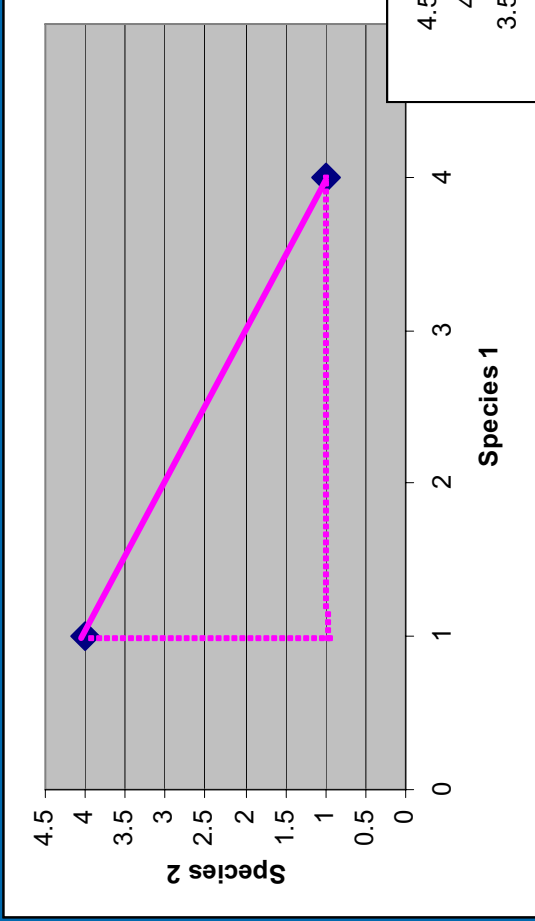
Modifications of Data

- **Standardizations: each data point**
 - Power (can create presence/absence, square root)
 - Logarithmic
 - Arcsine and Arcsine square root
- **Relativizations: by column and/or row**
 - Max, Total, Mean, Median, Mean and Stdev
 - Binary wrt Mean or Median (or any percentile)

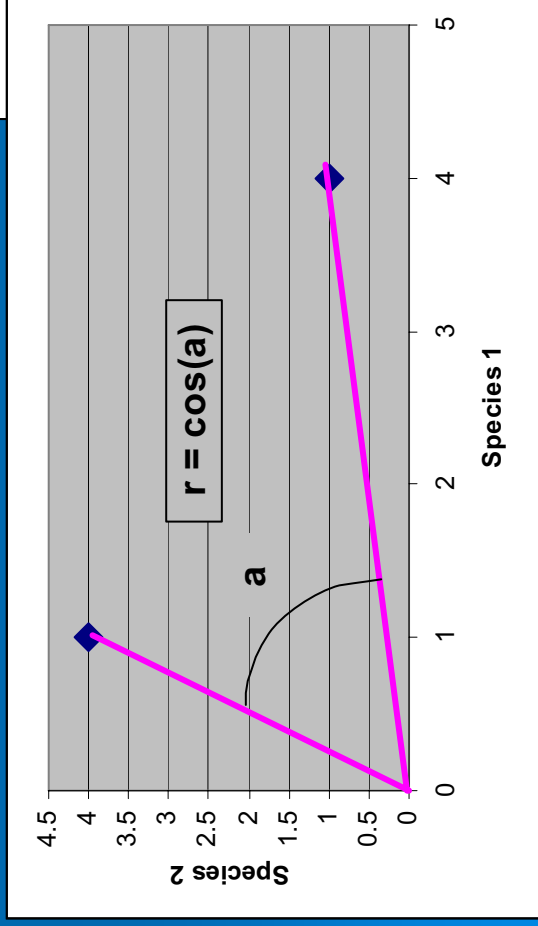
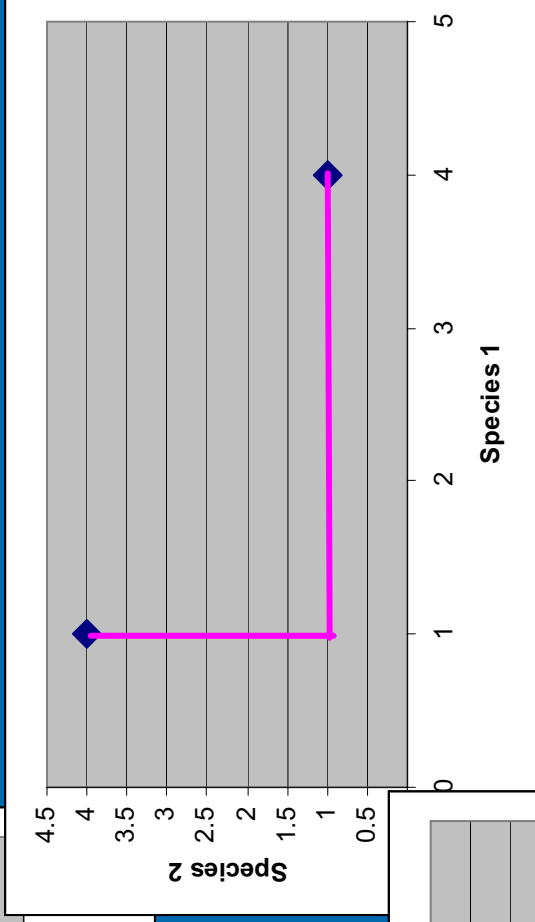
Distance Measures

- Euclidean (Pythagorean theorem applied to multiple dimensions)
 - Chi squared
 - Relative Euclidean (chord distance)
- City Block and Proportional City block
 - Sorenson (Bray and Curtis)
 - Relative Sorenson
 - Jaccard
- Correlation (Centroid)
- Mahalanobis

Euclidean



City Block



Correlation