

Design & Analysis of Ecological Data

Landscape of Statistical Methods: Part 3

Topics:

1. Multivariate statistics
2. Finding groups - cluster analysis
3. Testing/describing group differences
4. Unconstrained ordination
5. Constrained ordination

Landscape of Statistical Methods...

The Landscape

The basic statistical model:

$$Y = \underbrace{\text{deterministic part}}_{\substack{\uparrow \\ \text{Univariate} \\ \text{Multivariate}}} + \underbrace{\text{stochastic part}}_{\substack{\uparrow \\ \text{Distribution} \\ \text{Heterogeneity} \\ \text{Autocorrelation} \\ \text{Multiple levels} \\ \text{Random noise}}}$$



- Univariate
- Multivariate

- Linear
- Nonlinear
- Smoothed

- Distribution
- Heterogeneity
- Autocorrelation
- Multiple levels
- Random noise

Landscape of Statistical Methods...

Multivariate statistics

Why do we need multivariate statistics?



- Reflect more accurately the true multidimensional nature of natural systems
- Provide a way to handle large data sets with large numbers of variables
- Provide a way of summarizing redundancy in large data sets
- Provide rules for combining variables in an "optimal" way

Landscape of Statistical Methods...

Multivariate statistics

Why do we need multivariate statistics?



- Provide a means of detecting and quantifying truly multivariate patterns that arise out of the correlational structure of the variable set
- Provide a means of exploring complex data sets for patterns and relationships from which hypotheses can be generated and subsequently tested experimentally

Landscape of Statistical Methods...

What is multivariate statistics?

$y = x_1 + x_2 + \dots + x_j$ \longrightarrow Regression
Analysis of Variance
Contingency Tables, etc.

$y_1 + y_2 + \dots + y_i = x$ \longrightarrow Multivariate ANOVA
Discriminant Analysis
CART, MRPP, MANTEL

$y_1 + y_2 + \dots + y_i = x_1 + x_2 + \dots + x_j$ \longrightarrow Canonical Corr. Analysis
Constrained ordination

$y_1 + y_2 + \dots + y_i$ \longrightarrow Unconstrained ordination
Cluster Analysis

Multivariate Statistics

Landscape of Statistical Methods...

Multivariate methods

- Finding groups (Cluster analysis) \longrightarrow • Large family of techniques with similar goals; operating on data sets for which *pre-specified, well-defined groups do "not" exist*; characteristics of the data are used to assign entities into artificial groups
- Testing for groups (e.g., MRPP, MANTEL)
- Discriminating among groups (e.g., DA, ISA, mCART)
- Unconstrained ordination (e.g., PCA, CA, NMDS)
- Constrained ordination (e.g., RDA, CCA, CAPS)

Landscape of Statistical Methods...

Finding groups – cluster analysis



- Can we organize sampling entities (e.g., sites) into discrete classes, such that within-group similarity is maximized and among-group similarity is minimized?

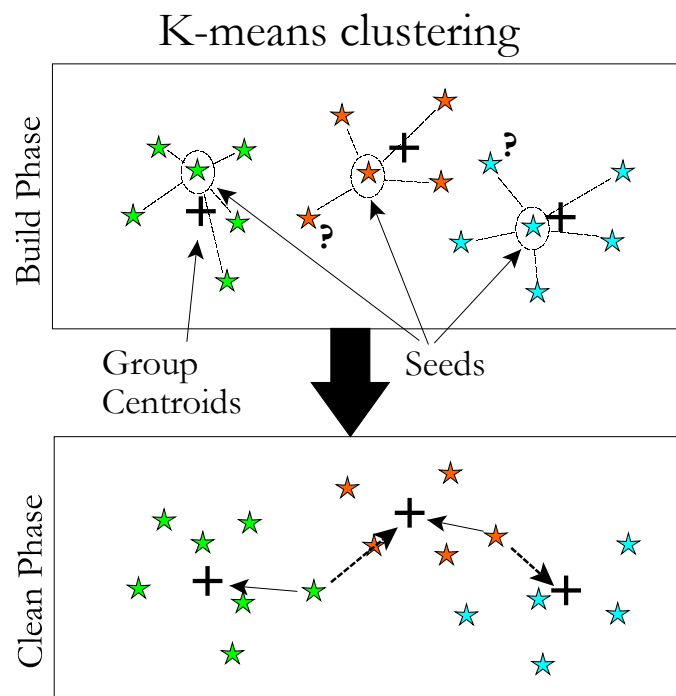
Sites	Species			
	A	B	C	D
1	1	9	12	1
2	1	8	11	1
3	1	6	10	10
4	10	0	9	10
5	10	2	8	10
6	10	0	7	2

Landscape of Statistical Methods...

Finding groups – cluster analysis

Nonhierarchical clustering:

- NHC methods merely assign each entity to a cluster, placing similar entities together in order to *maximize within-cluster homogeneity*

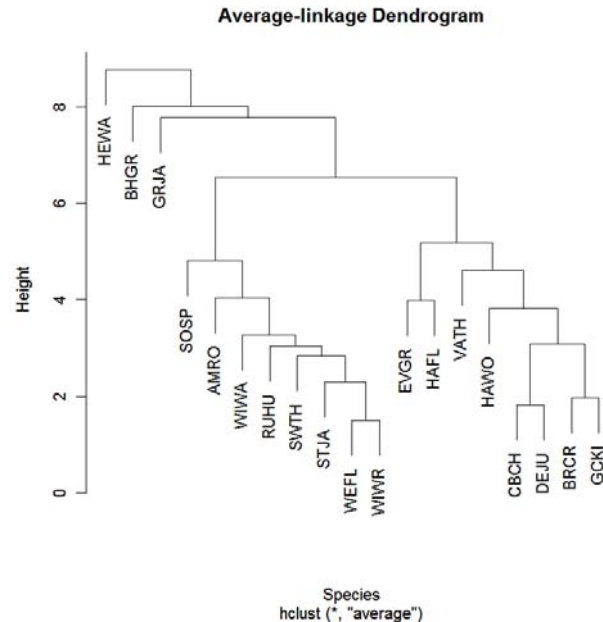


Landscape of Statistical Methods...

Finding groups – cluster analysis

Hierarchical clustering:

- HC methods combine similar entities into classes or groups and arrange these groups into a *hierarchy* that reveals relationships among the entities classified



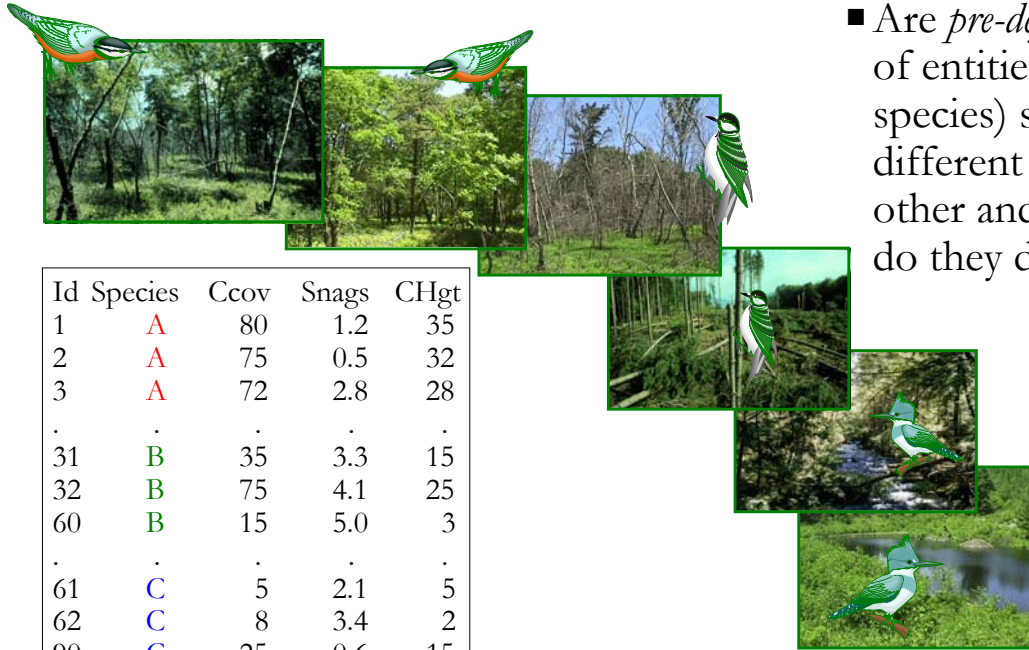
Landscape of Statistical Methods...

Multivariate methods

- Finding groups (Cluster analysis)
- Testing for groups (e.g., MRPP, MANTEL) → • Family of different methods for testing and/or describing differences among *pre-specified, well-defined groups* based on a set of discriminating variables
- Discriminating among groups (e.g., DA, ISA, mCART) →
- Unconstrained ordination (e.g., PCA, CA, NMDS)
- Constrained ordination (e.g., RDA, CCA, CAPS)

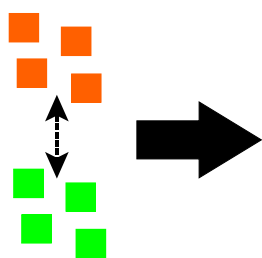
Landscape of Statistical Methods...

Discriminating among groups

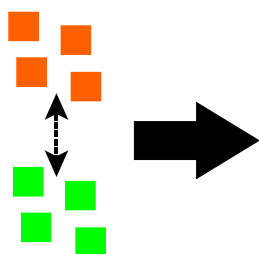


Landscape of Statistical Methods...

Testing for group differences



- Are groups significantly different? (How valid are the groups?)
 - ▶ Multivariate Analysis of Variance (MANOVA)
 - ▶ Multi-Response Permutation Procedures (MRPP)
 - ▶ Analysis of Group Similarities (ANOSIM)
 - ▶ Mantel's Test (MANTEL)



- How do groups differ? (Which variables best distinguish among the groups?)
 - ▶ Discriminant Analysis (DA)
 - ▶ Classification and Regression Trees (CART)
 - ▶ Logistic Regression (LR)
 - ▶ Indicator Species Analysis (ISA)

Landscape of Statistical Methods...

Multivariate methods

- Finding groups (Cluster analysis)
 - Testing for groups (e.g., MRPP, MANTEL)
 - Discriminating among groups (e.g., DA, ISA, mCART)
 - Unconstrained ordination (e.g., PCA, CA, NMDS) →
 - Constrained ordination (e.g., RDA, CCA, CAPS)
- A family of different methods for organizing sampling entities (e.g., species, sites, observations, etc.) along continuous gradients based on a set of interdependent variables

Landscape of Statistical Methods...

Unconstrained ordination



- Can we organize entities (e.g., sites) along one or more gradients based on their relationships among the interdependent variables?

Sites	Species A		Species B		Species C		Species D		Species E	
1	0	(1)	5	(1)	1	(1)	10	(4)	10	(4)
2	2	(3)	8	(3)	4	(3)	12	(6)	20	(6)
3	8	(6)	20	(6)	10	(6)	1	(2)	3	(2)
4	4	(5)	11	(5)	8	(5)	11	(5)	14	(5)
5	1	(2)	6	(2)	2	(2)	2	(3)	6	(3)
6	3	(4)	10	(4)	6	(4)	0	(1)	0	(1)

Landscape of Statistical Methods...

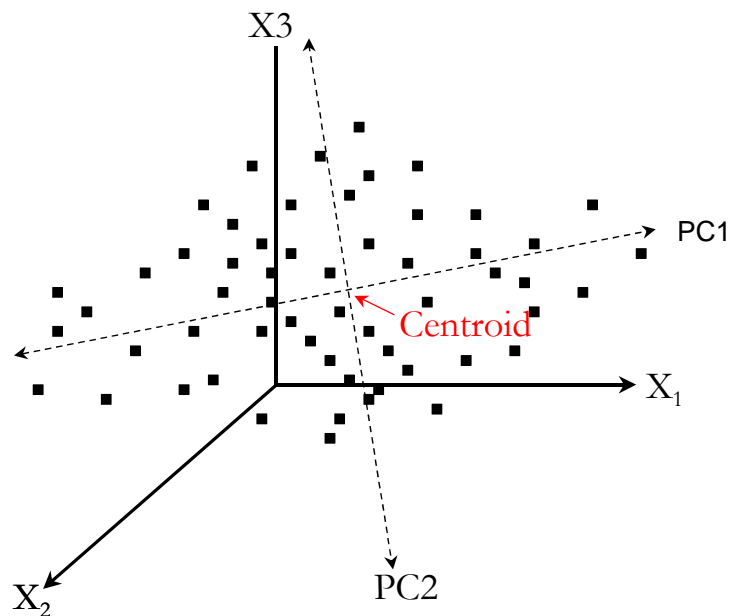
Unconstrained ordination

Obs	Canopy Cover	Snag Density	Canopy Height
1	80	1.2	35
2	75	0.5	32
3	72	0.8	28
.	.	.	.
N	25	0.6	15



$$PC1 = .8x_1 - .4x_2 + .1x_3$$

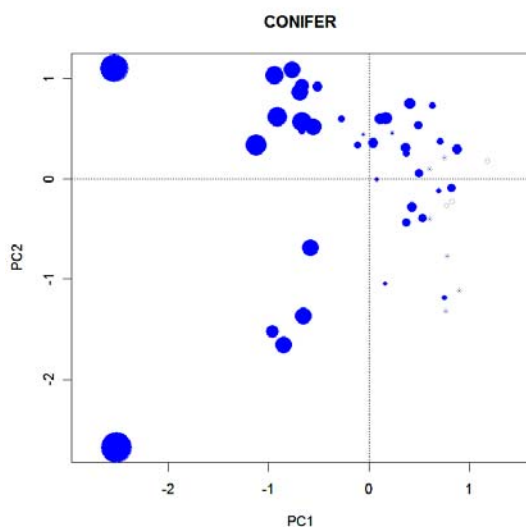
$$PC2 = -.1x_1 - .1x_2 + .9x_3$$



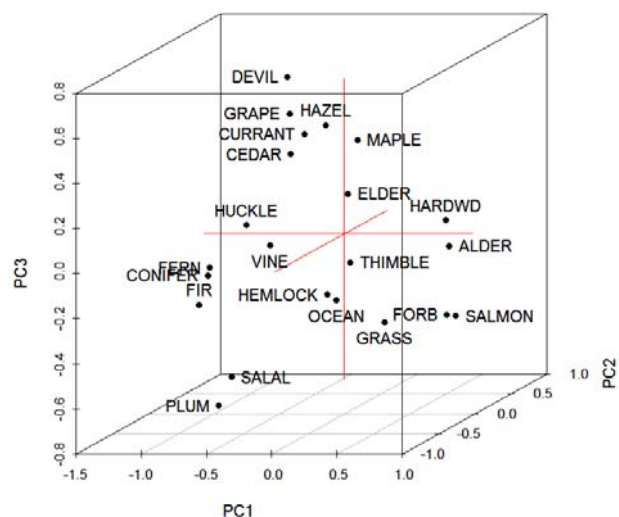
Landscape of Statistical Methods...

Unconstrained ordination

2d ordi bubble plot



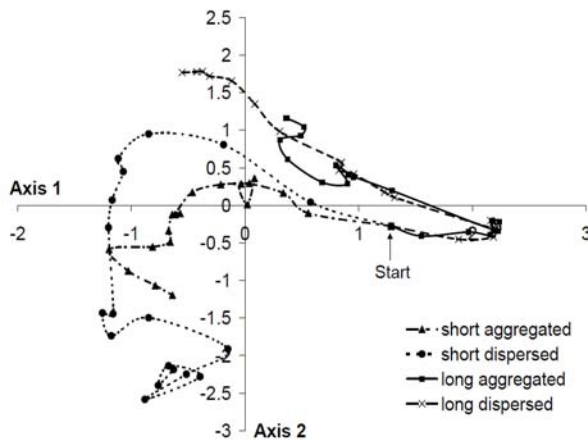
3d ordi scatter plot



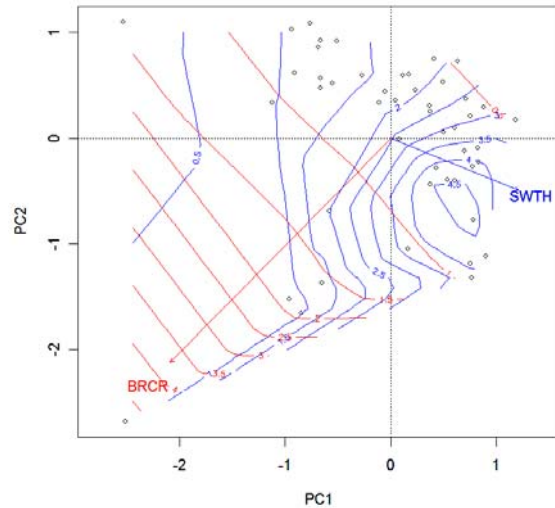
Landscape of Statistical Methods...

Unconstrained ordination

Ordi trajectories



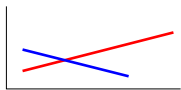
Ordi overlays



Landscape of Statistical Methods...

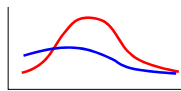
Unconstrained ordination

Linear



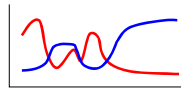
- Principal components analysis (PCA)
- Factor analysis (FA)
- Multidimensional scaling (MDS/PCO)
- **ML**-Unconstrained linear ordination (ULO)

Quadratic



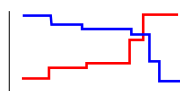
- Correspondence analysis (CA & DCA)
- **ML**-Unconstrained quadratic ordination (UQO)

Smooth



- **ML**-Unconstrained additive ordination (UAO)

Nonlinear



- Nonmetric multidimensional scaling (NMDS)

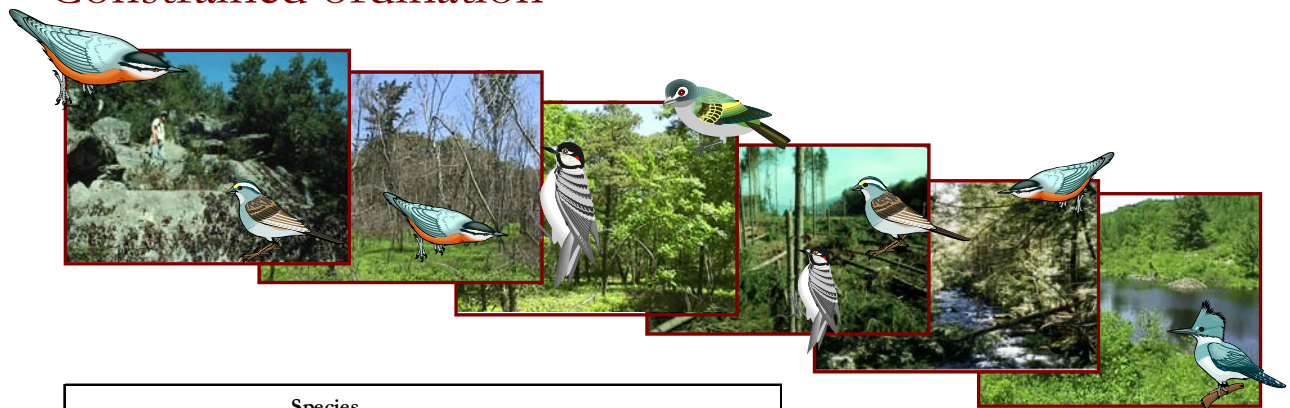
Landscape of Statistical Methods...

Multivariate methods

- Finding groups (Cluster analysis)
 - Testing for groups (e.g., MRPP, MANTEL)
 - Discriminating among groups (e.g., DA, ISA, mCART)
 - Unconstrained ordination (e.g., PCA, CA, NMDS)
 - Constrained ordination (e.g., RDA, CCA, CAPS) →
- A family of different methods for extending unconstrained ordination in which the solution is constrained to be expressed by ancillary variables

Landscape of Statistical Methods...

Constrained ordination



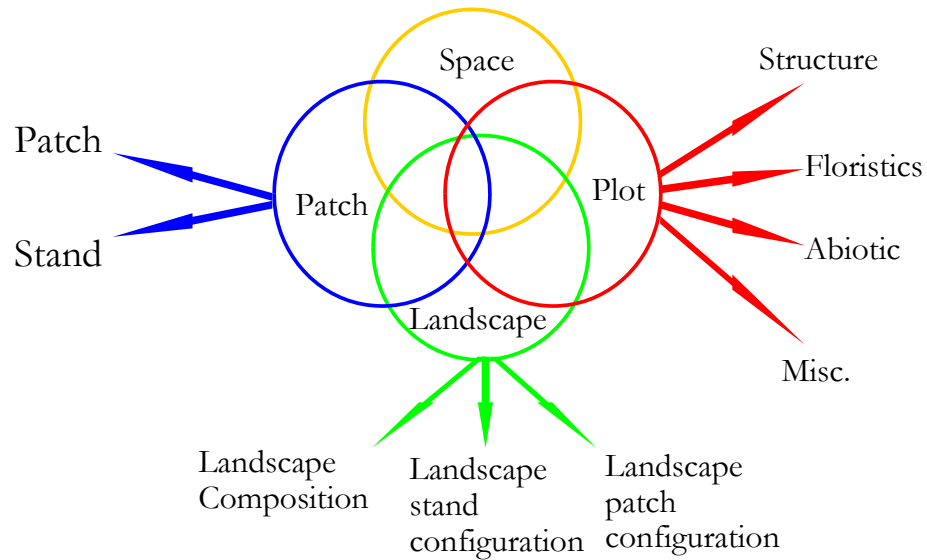
Sites	Species					Tree cover	Snag density	Shrub cover
	A	B	C	D	E			
1	0	5	1	10	10	15	0.2	30
2	2	8	4	12	20	55	0.5	45
3	8	20	10	1	3	55	2.3	22
4	4	11	8	11	14	75	1.8	31
5	1	6	2	2	6	85	0.3	15
6	3	10	6	0	0	60	0.8	10

- Can bird community patterns be explained by measured environmental variables?

Landscape of Statistical Methods...

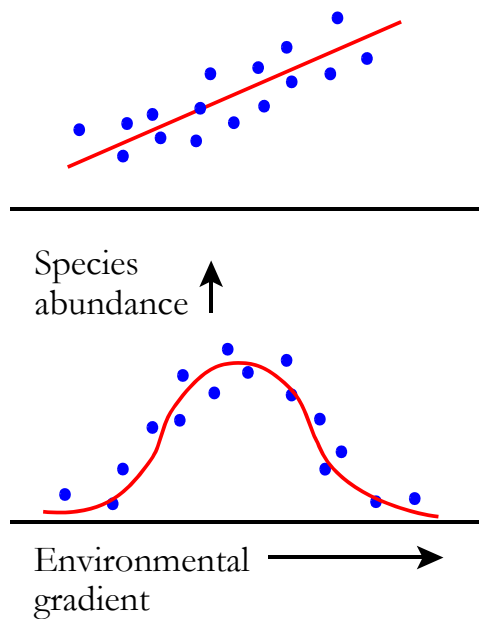
Constrained ordination

Variance partitioning



Landscape of Statistical Methods...

Constrained ordination



- Constrained analysis of principal coordinates (CAP)
- Redundancy analysis (RDA)

- Canonical correspondence analysis (CCA)