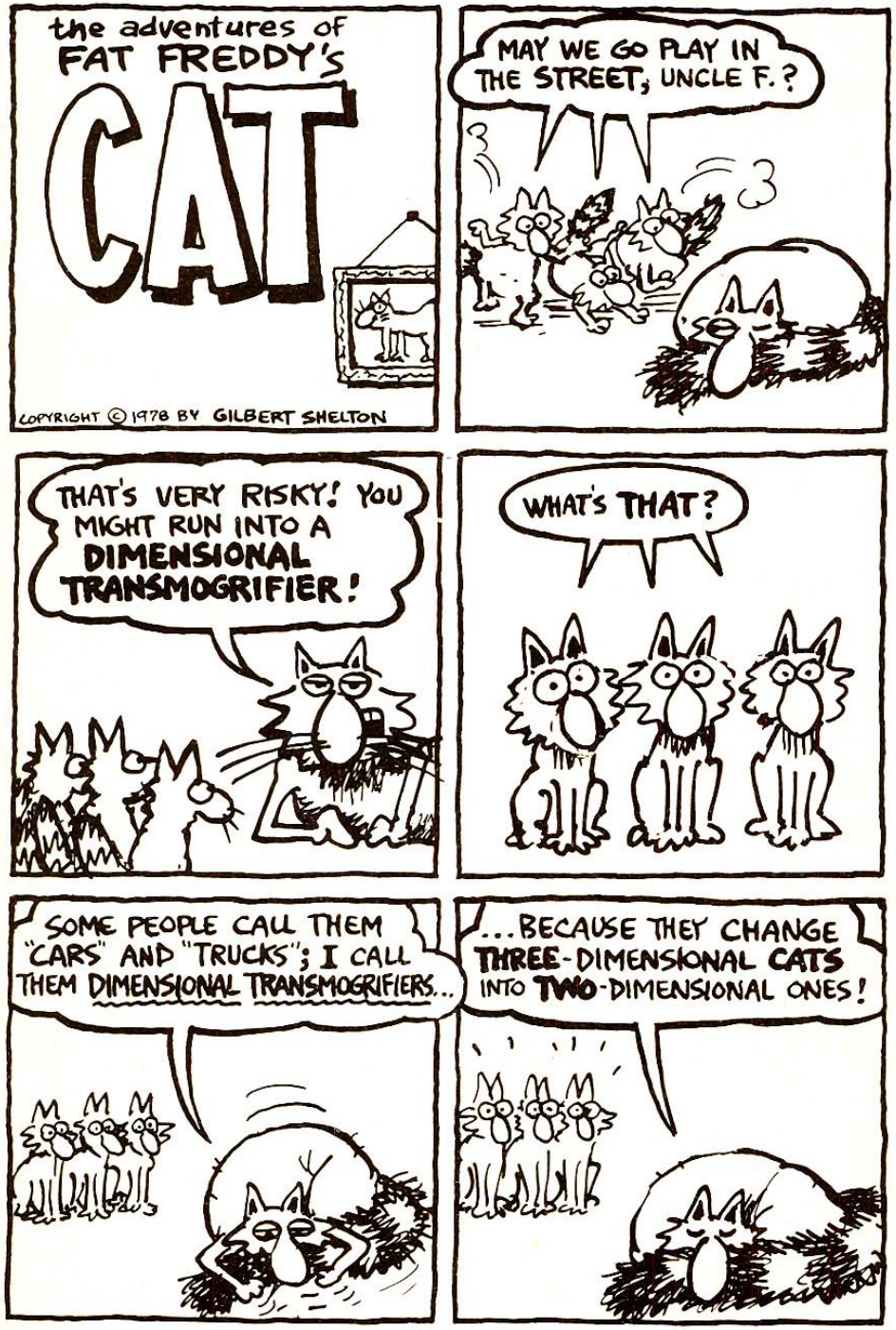
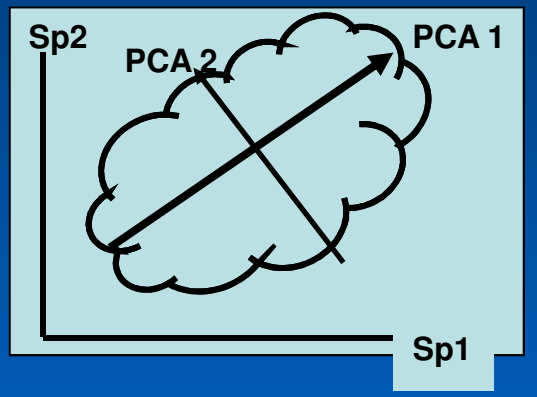
Introduction to ordination

Gary Bradfield
Botany Dept.



Ordination

"…there appears to be no word in English which one can use as an antonym to "classification"; I would like to propose the term "ordination." (Goodall, D. W. 1954. Amer. J. Bot. 2: p.323)

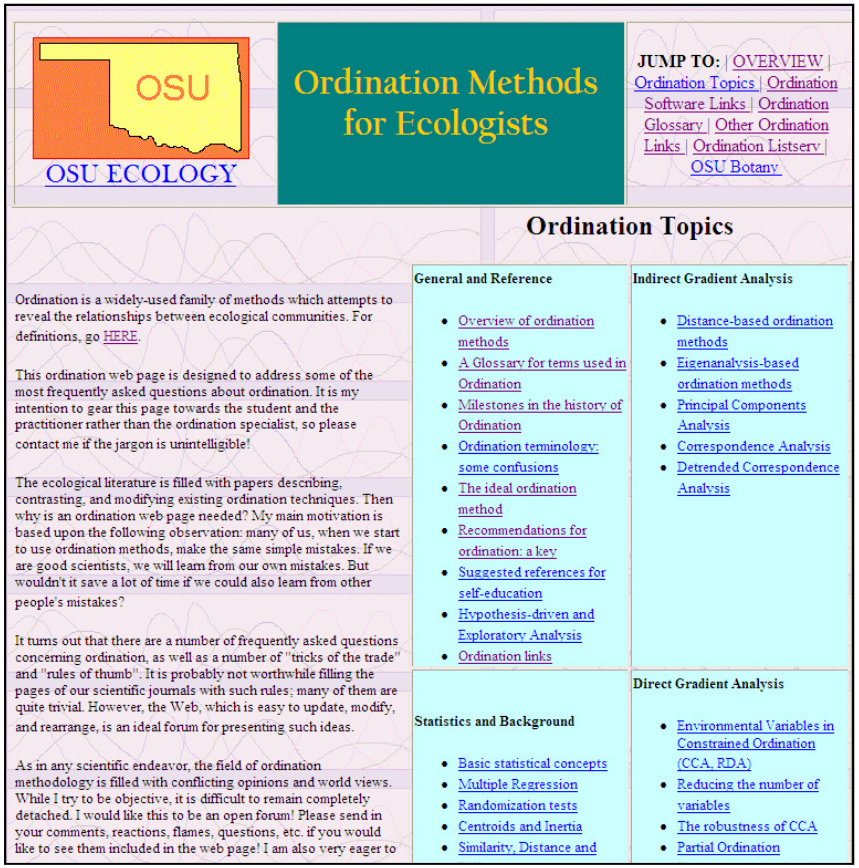


MAIN USES:

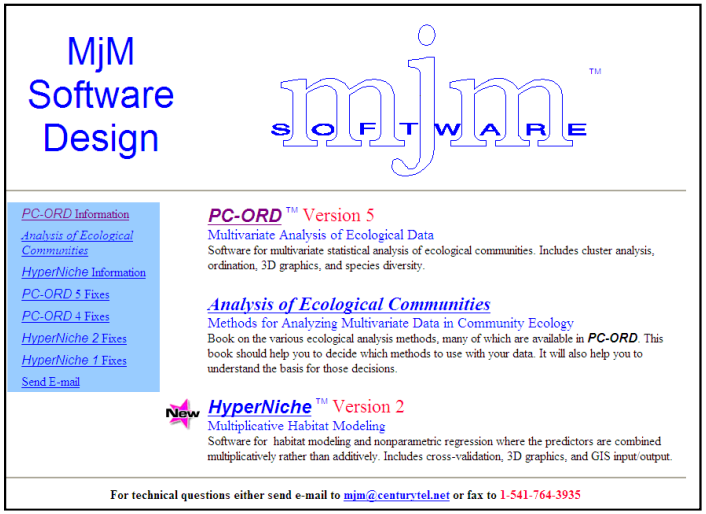
- Data reduction and graphical display
- Detection of main structure and relationships
- Hypothesis generation
- Data transformation for further analysis

Ordination info & software

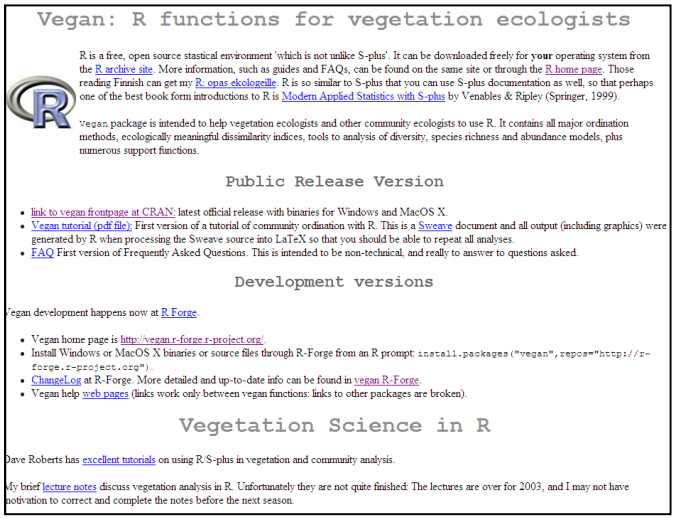
http://ordination.okstate.edu/



http://home.centurytel.net/~mjm/index.htm



http://cc.oulu.fi/~jarioksa/softhelp/vegan.html



Ordination background:

Community-unit hypothesis: "classification" of discrete variation

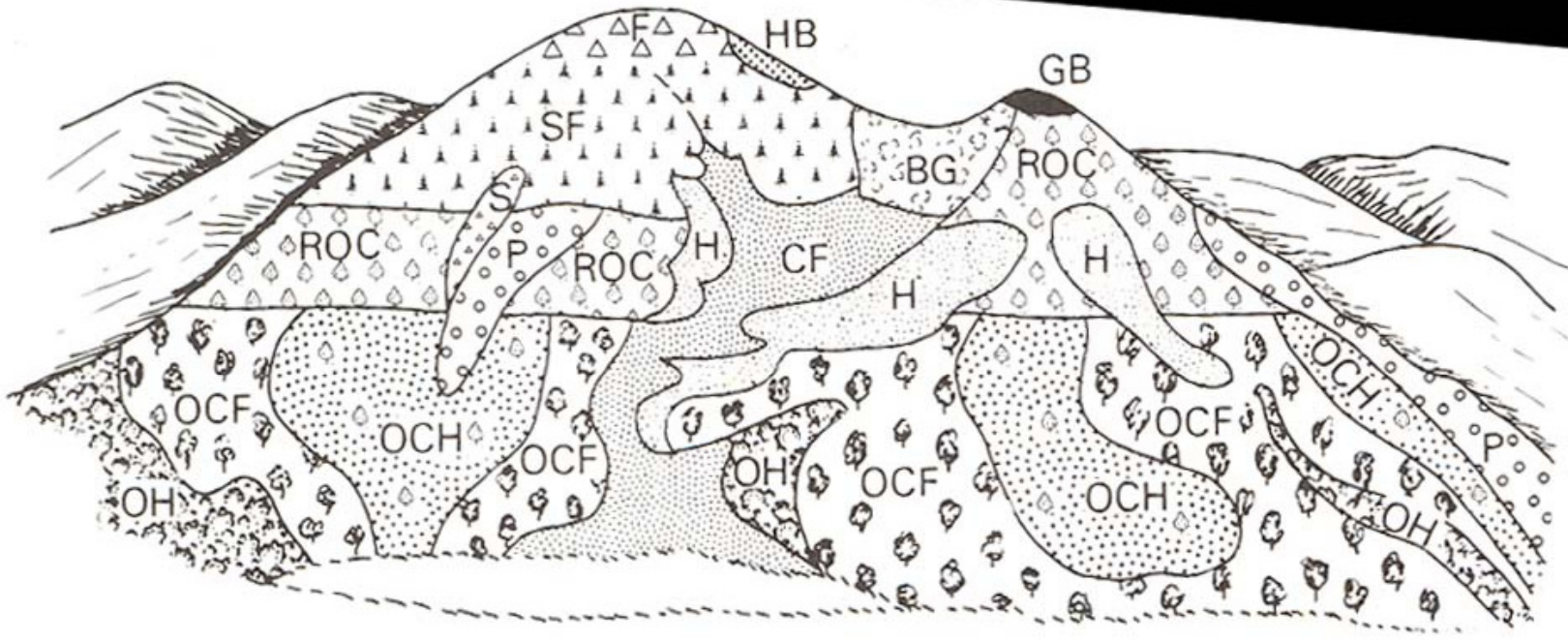
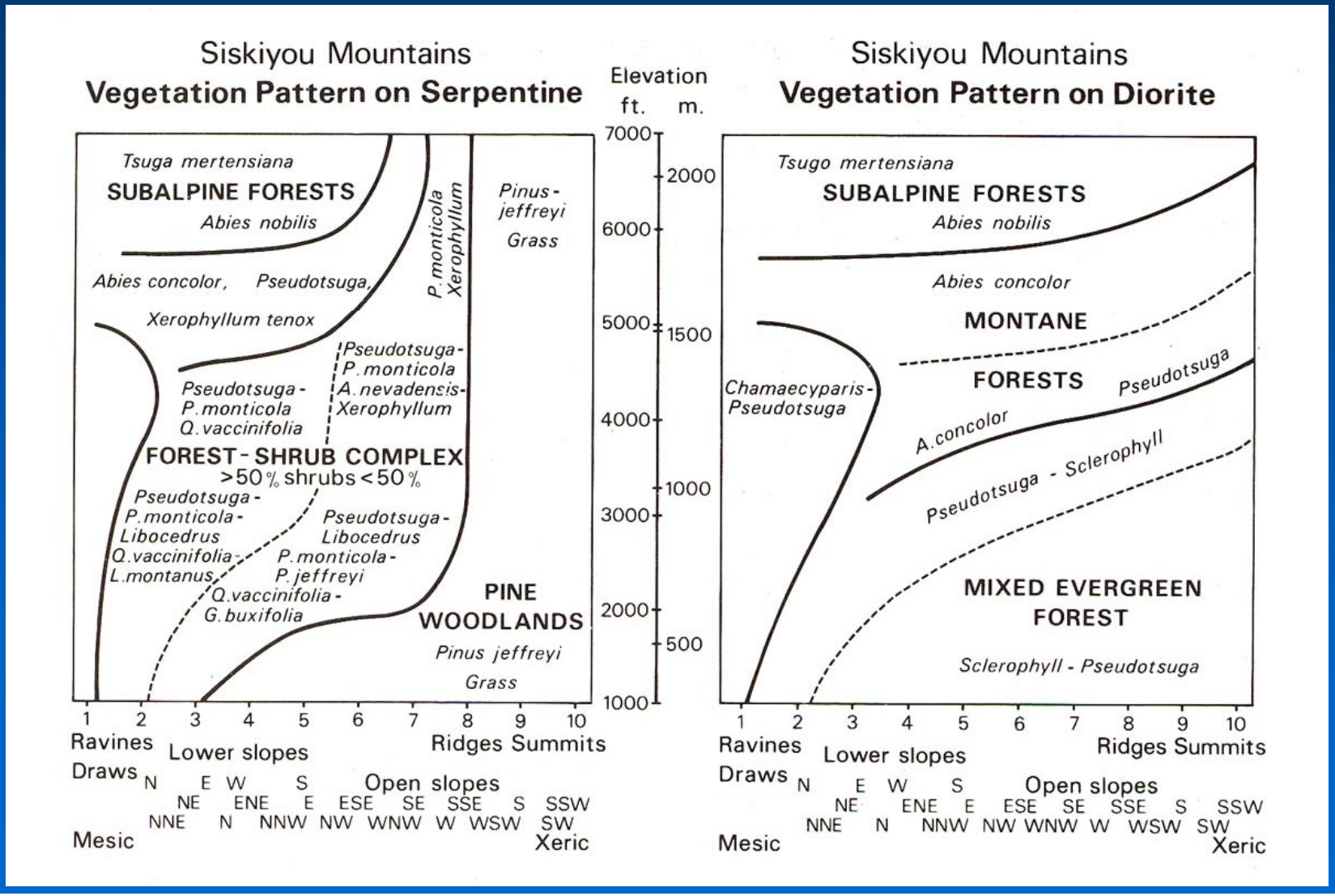


Figure 1.9 Topographic distributions of vegetation types on an idealised west-facing mountain and valley in the Great Smokey Mountains, USA. Vegetation types: BG, beech gap; CF, cove forest; F, Fraser fir forest; GB, grassy bald; H, hemlock forest; HB, heath bald; OCF, chestnut oak–chestnut forest; OCH, chestnut oak–chestnut heath; OH, oak–hickory forest; P, pine forest and pine heath; ROC, red oak–chestnut forest; S, spruce forest; SF, spruce–fir forest. (Redrawn from Whittaker, 1956; with kind permission of *Ecological Monographs*)

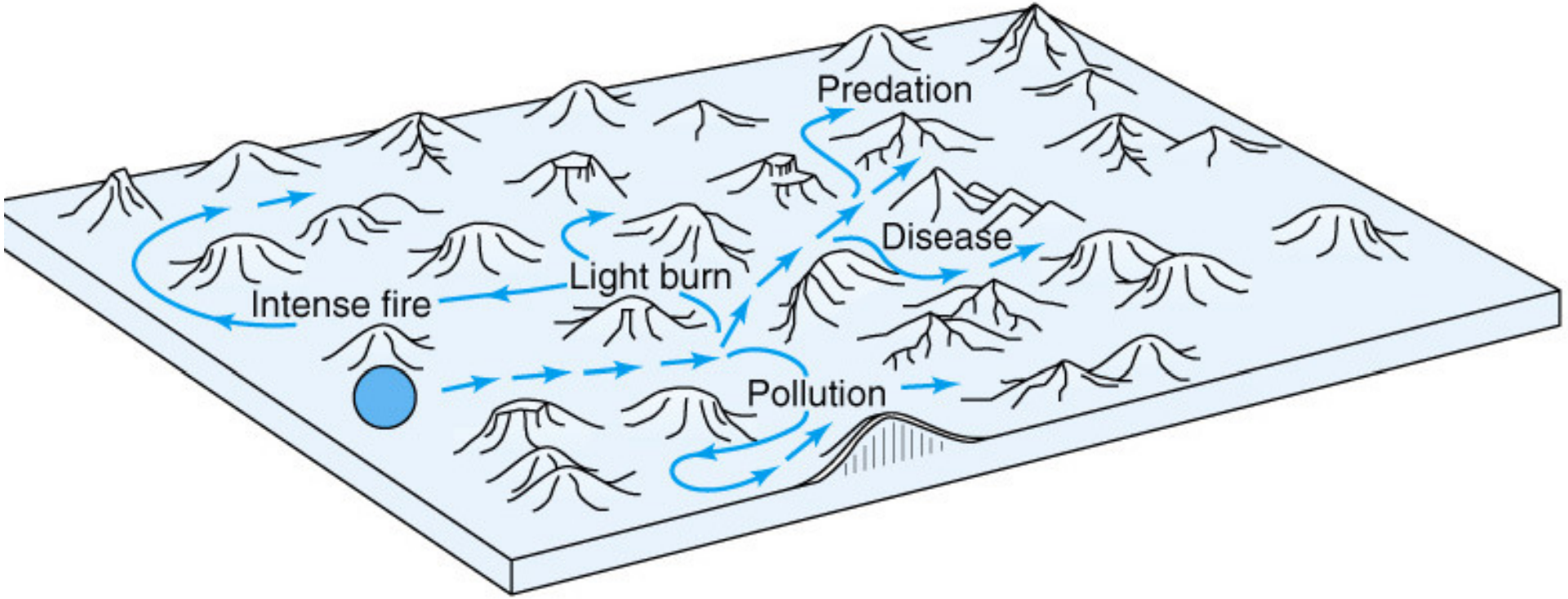
Ordination background:

Individualistic hypothesis: "ordination" of continuous variation



Ordination background:

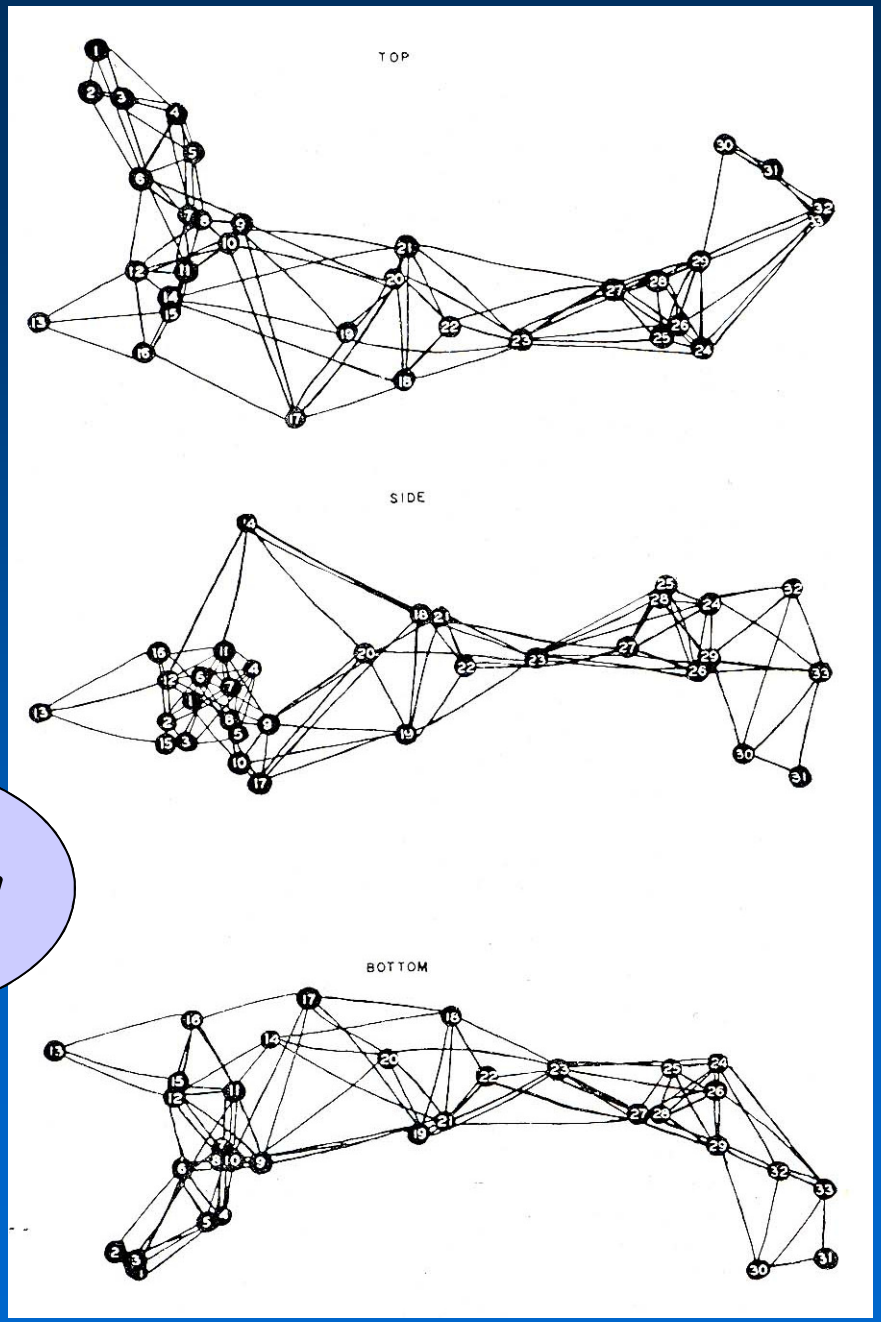
Nonequilibrium landscape model



- continuous interplay of spatial & temporal processes
- consistent with ordination approach to analysis

Early ordinations:

Plexus diagram of plant species in Saskatchewan (Looman 1963)

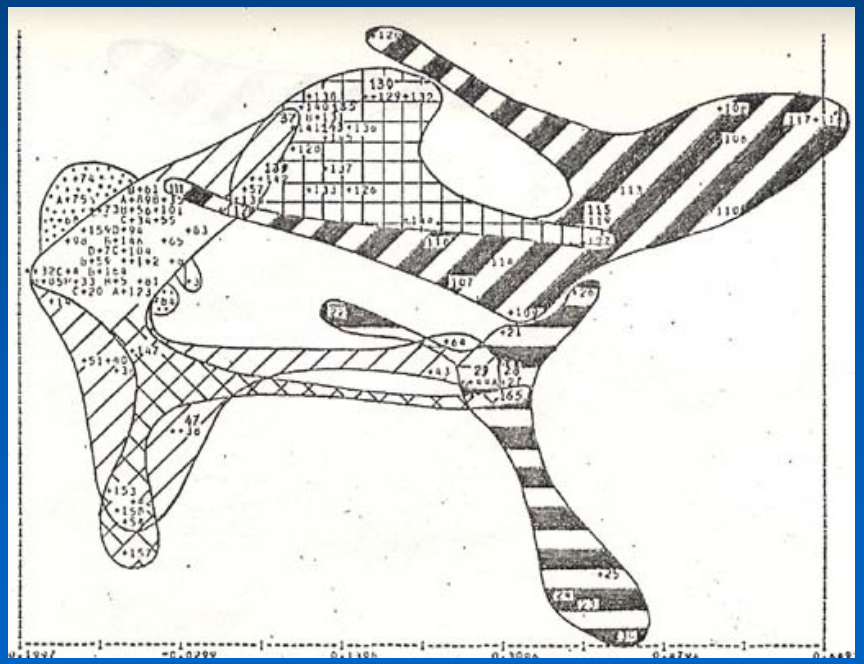


Bow-wow

Early ordinations:

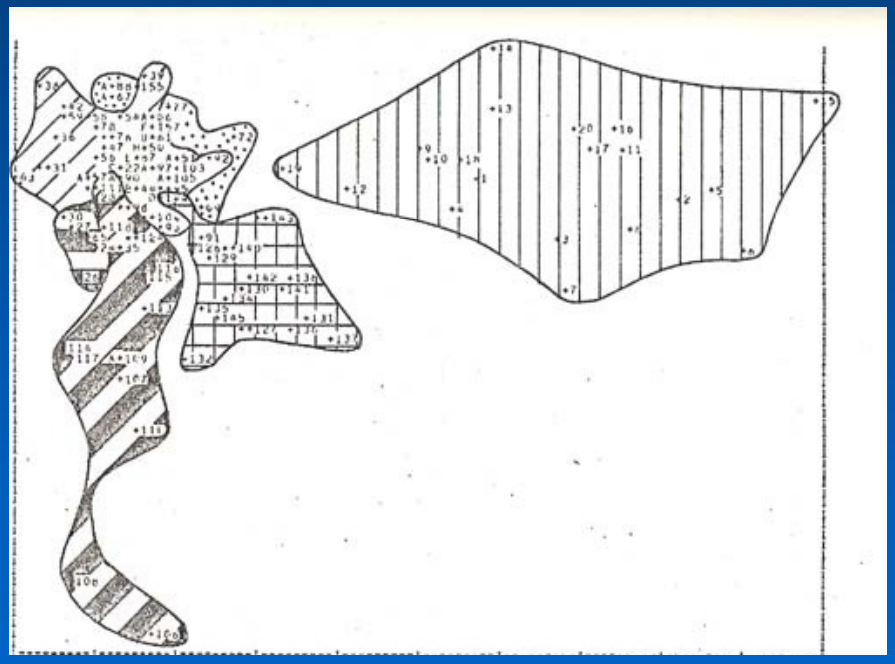
PCA of *Eucalyptus* forest localities after fire in S.E. Australia (Bradfield 1977)

Species covariance



shrub cover

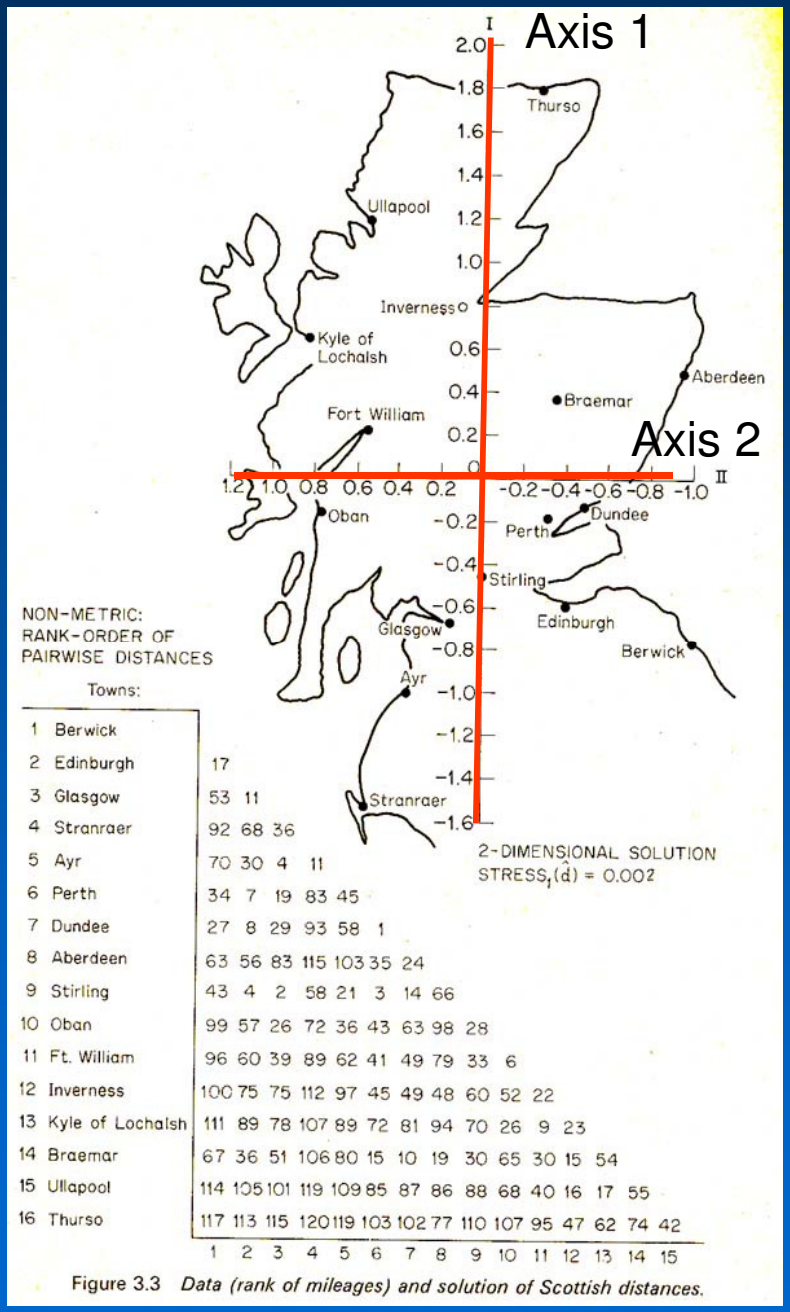
Species correlation



 # rare species

Early ordinations:

NMS ordination of Scottish cities (Coxon 1982)



NON-METRIC:
RANK-ORDER OF
PAIRWISE DISTANCES

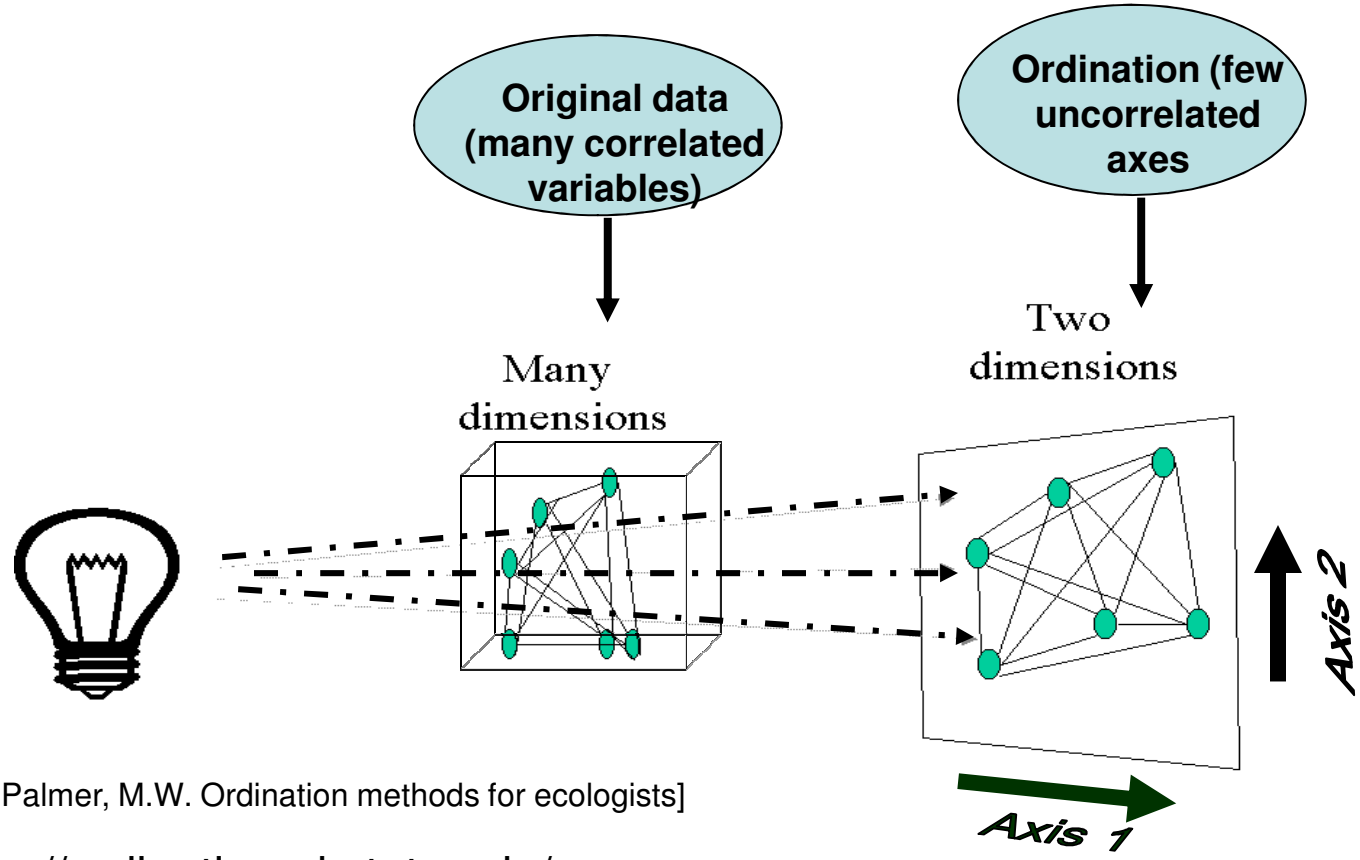
Towns:

1	Berwick													
2	Edinburgh	17												
3	Glasgow	53	11											
4	Stranraer	92	68	36										
5	Ayr	70	30	4	11									
6	Perth	34	7	19	83	45								
7	Dundee	27	8	29	93	58	1							
8	Aberdeen	63	56	83	115	103	35	24						
9	Stirling	43	4	2	58	21	3	14	66					
10	Oban	99	57	26	72	36	43	63	98	28				
11	Ft. William	96	60	39	89	62	41	49	79	33	6			
12	Inverness	100	75	75	112	97	45	49	48	60	52	22		
13	Kyle of Lochalsh	111	89	78	107	89	72	81	94	70	26	9	23	
14	Braemar	67	36	51	106	80	15	10	19	30	65	30	15	54
15	Ullapool	114	105	101	119	109	85	87	86	88	68	40	16	17	55
16	Thurso	117	113	115	120	119	103	102	77	110	107	95	47	62	74	42
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Figure 3.3 *Data (rank of mileages) and solution of Scottish distances.*

Matrix of ranked distances between cities

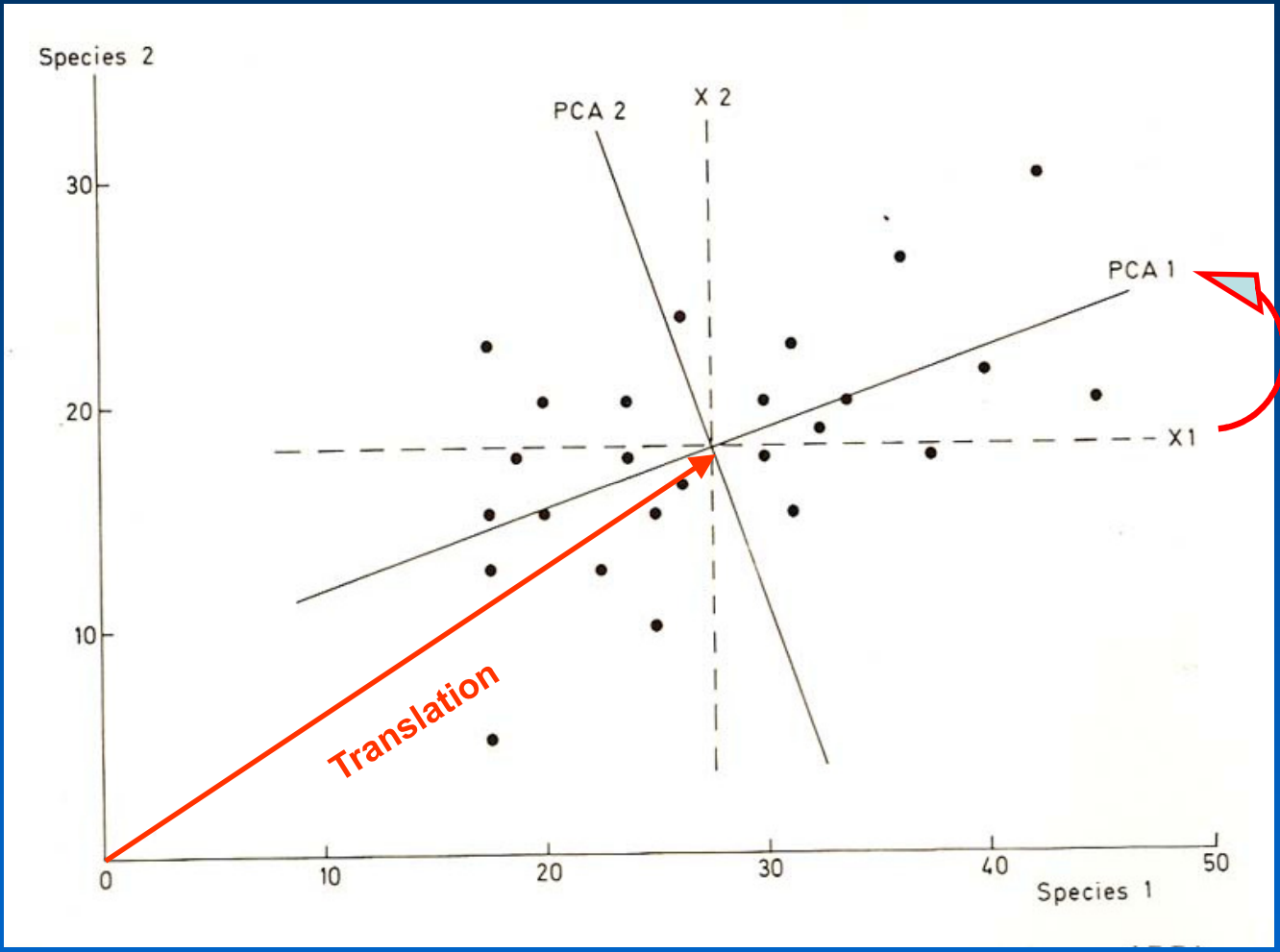
Basic idea of ordination:



[Source: Palmer, M.W. Ordination methods for ecologists]

http://ordination.okstate.edu/

Geometric model of PCA



PCA assumes linear relations among species

Linear

Non-linear

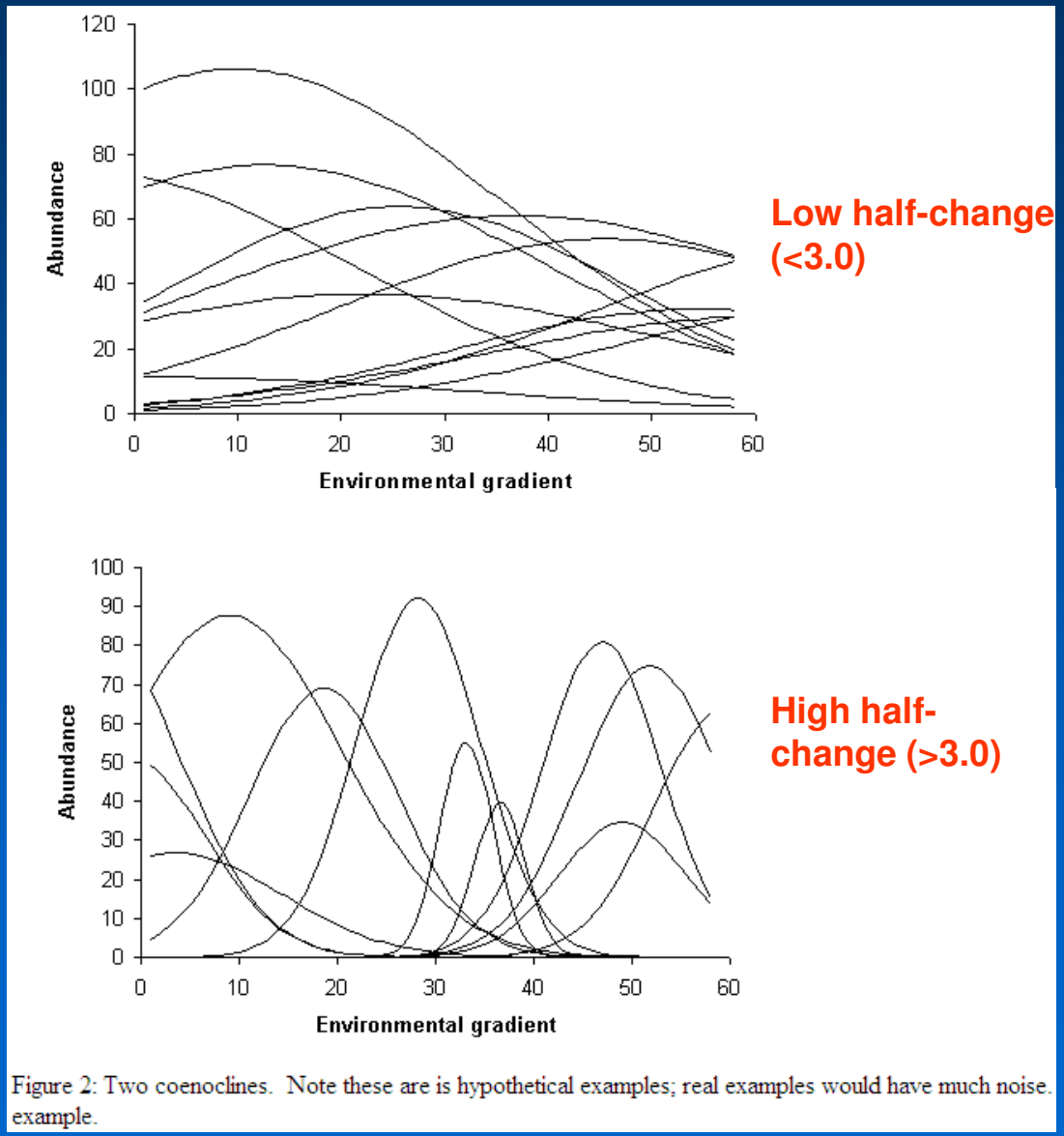


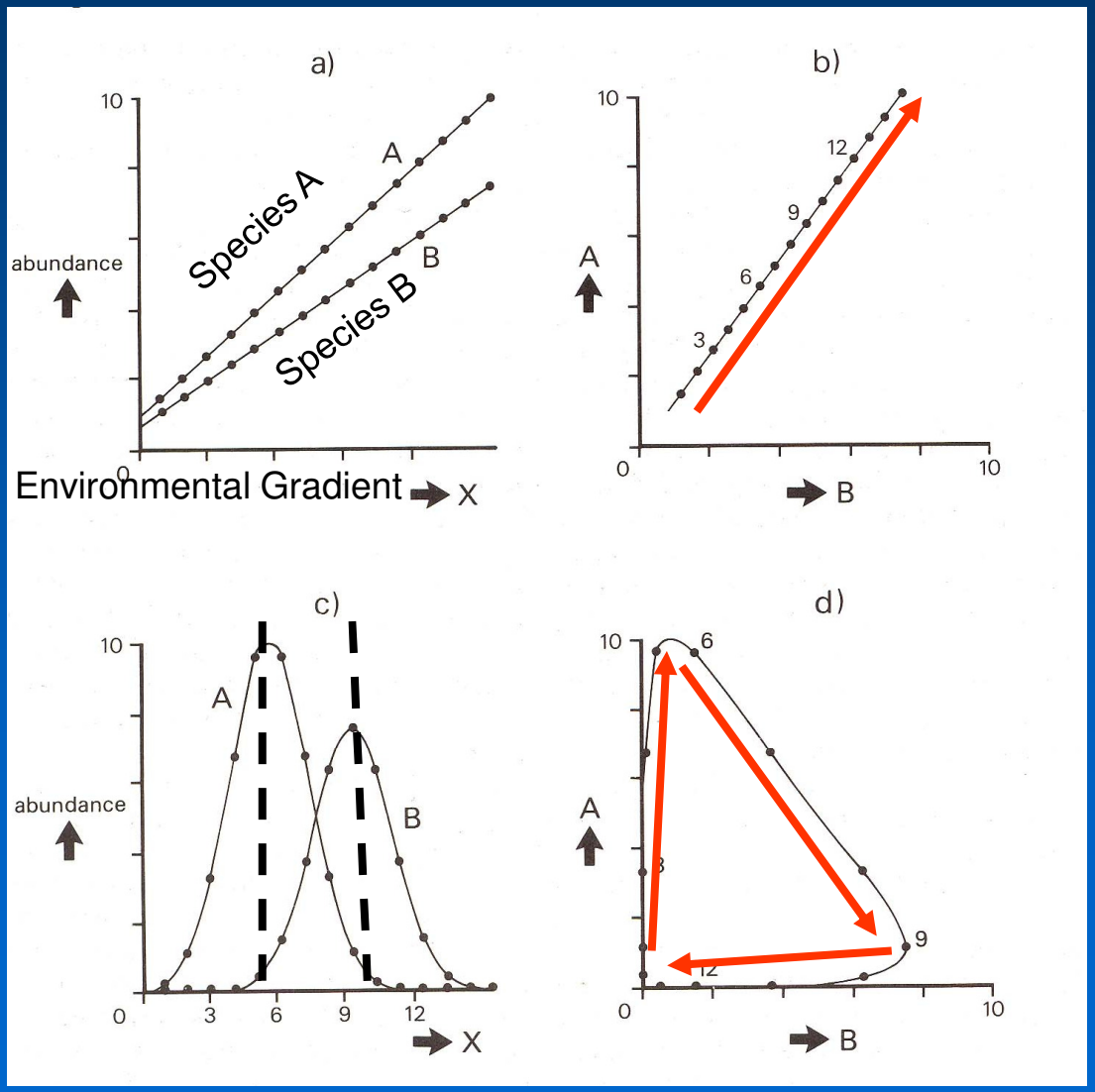
Figure 2: Two coenoclines. Note these are is hypothetical examples; real examples would have much noise. example.

PCA assumes linear relations among species

Environment space Species space

Linear

Non-linear



CHOOSING AN ORDINATION METHOD

Unconstrained methods	Constrained methods
Methods to describe the structure in a single data set: • PCA (principal component analysis on a covariance matrix or a correlation matrix) • CA (correspondence analysis, also known as reciprocal averaging) • DCA (detrended correspondence analysis) • NMS (nonmetric multidimensional scaling, also known as NMDS)	Methods to explain one data set by another data set (ordinations constrained by explanatory variables): • RDA (redundancy analysis, the canonical form of PCA) • CCA (canonical correspondence analysis, the canonical form of CA) • CANCOR (canonical correlation analysis) • "Partial" analysis (methods to describe the structure in a data set after accounting for variation explained by a second data set i.e.covariable data)

NMS (Nonmetric multidimensional scaling)

- Goal of NMS is to position objects in a space of reduced dimensionality while preserving rank-order relationships as well as possible (i.e. make a nice picture)
- Wide flexibility in choice of distance coefficients
- Makes no assumptions about data distributions
- Often gives "better" 2 or 3 dimensional solution than PCA (but NMS axes are arbitrary)
- Success measured as that configuration with lowest "stress"

$$\sqrt{\sum_{i<j}(d_{ij}-\hat{d}_{ij})^2 \Big/ \sum_{i<j} d_{ij}^{\,2}}$$

Stress	Fit
0.40	Poor
0.30	Fair
0.20	Good
0.10	Excellent
0.00	Perfect

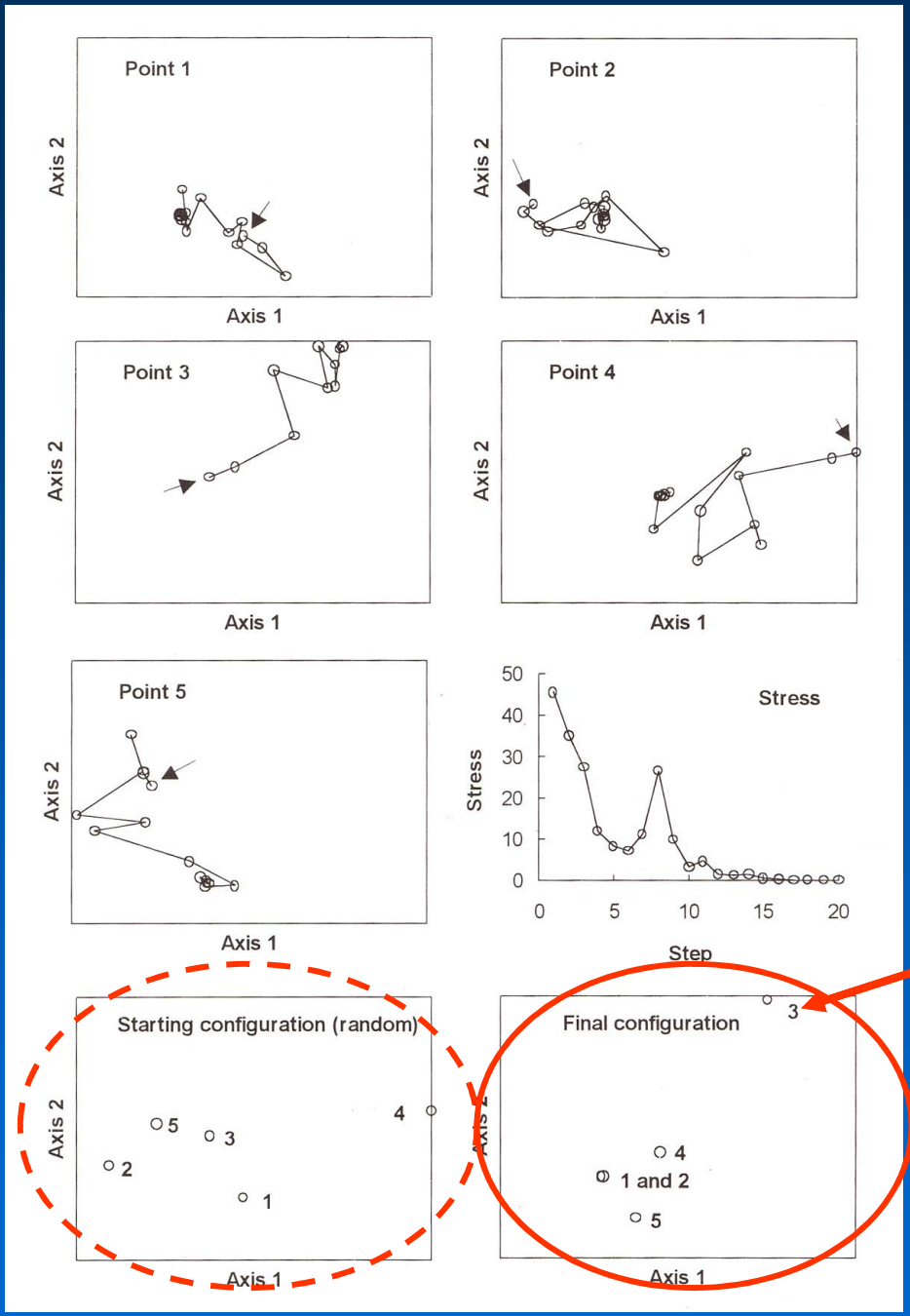
NMS illustration (McCune & Grace 2002)

Table 16.5. Abundance of six species in each of five sample units.

	Species					
SU	sp1	sp2	sp3	sp4	sp5	sp6
1	1	2	3	4	5	5
2	1	3	2	4	6	6
3	0	3	0	1	0	1
4	1	2	2	2	3	4
5	5	2	1	3	5	6

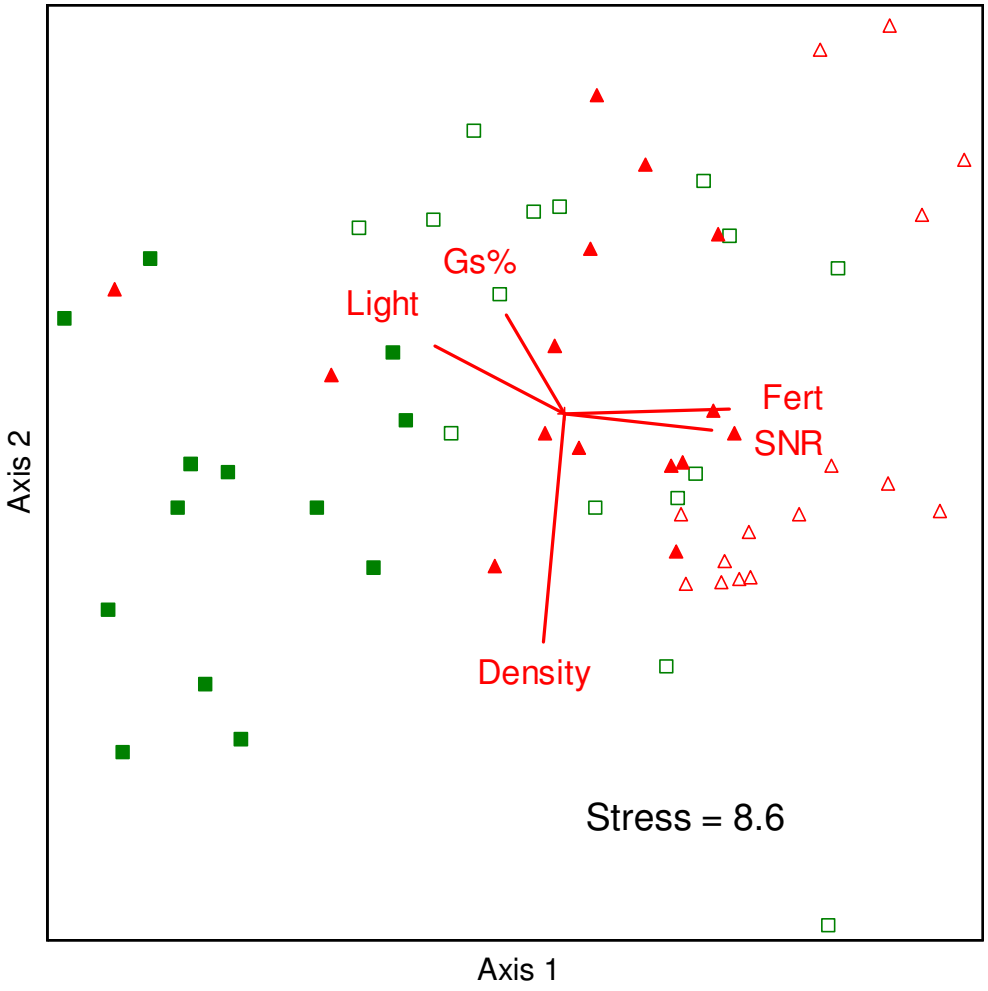
Table 16.6. Sørensen distances among the five sample units from Table 16.5.

	SU				
SU	1	2	3	4	5
1	0				
2	0.0952	0			
3	0.6800	0.6296	0		
4	0.1765	0.2222	0.5789	0	
5	0.1905	0.1818	0.7037	0.2778	0



Example: Planted hemlock trees – northern Vancouver Island (Shannon Wright MSc thesis)

NMS (Nonmetric Multidimensional Scaling)

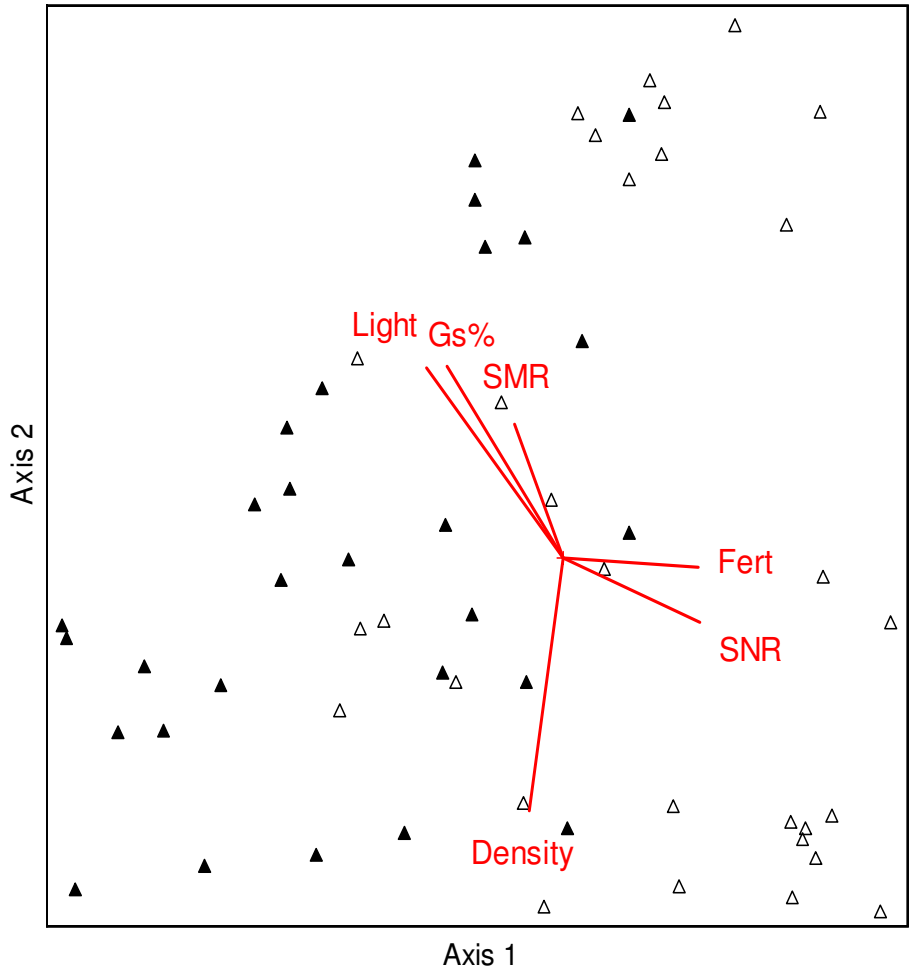


Original Lewis Classification
- ▲ HA
- △ HA fert
- ■ CH
- □ CH fert

Variable	Axis 1	Axis 2
Environment correlations		
Fertilization	0.605	0.109
Density	-0.213	-0.712
SNR	0.571	-0.192
Gs%	-0.361	0.468
Light	-0.535	0.391
Tree response correlations		
Form	-0.410	0.215
Vigour	0.710	-0.121
Canopy Closure	0.502	-0.499
Top Height	0.883	0.157
Vol / tree	0.945	0.273
DBH	0.906	0.063

Example: Planted hemlock trees – northern Vancouver Island (Shannon Wright MSc thesis)

CCA (Canonical Correspondence Analysis)



Variable	Axis 1	Axis 2
Environment intraset correlations		
Fertilization	0.620	-0.125
Scarifiication	-0.163	-0.433
Density	-0.314	-0.892
SMR	-0.094	0.531
SNR	0.588	-0.327
FFcm	-0.236	0.267
Gs%	-0.496	0.757
Rs%	0.123	0.009
For Flr	0.279	-0.320
Light	-0.532	0.776
Tree response correlations		
Form	0.040	0.080
Vigour	0.744	-0.158
Canopy Closure	0.469	-0.721
Top Height	0.834	-0.132
Vol / tree	0.812	-0.050
DBH	0.870	0.056

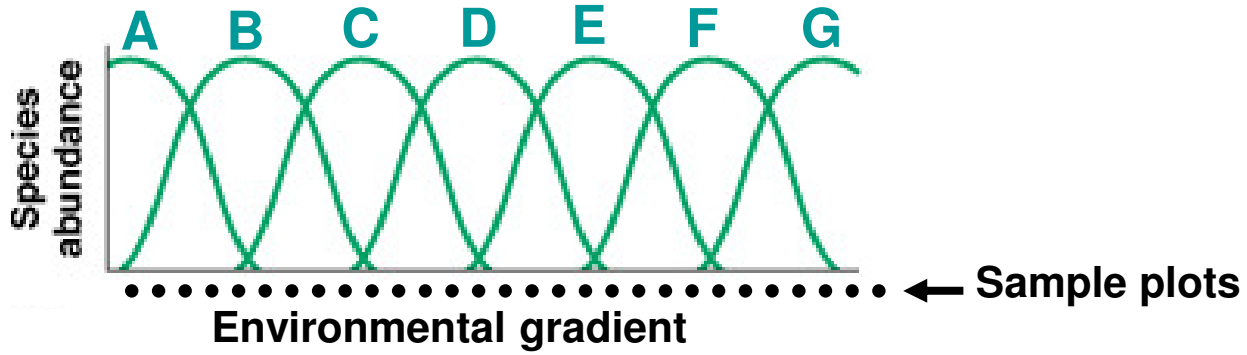
Evaluating an ordination method:

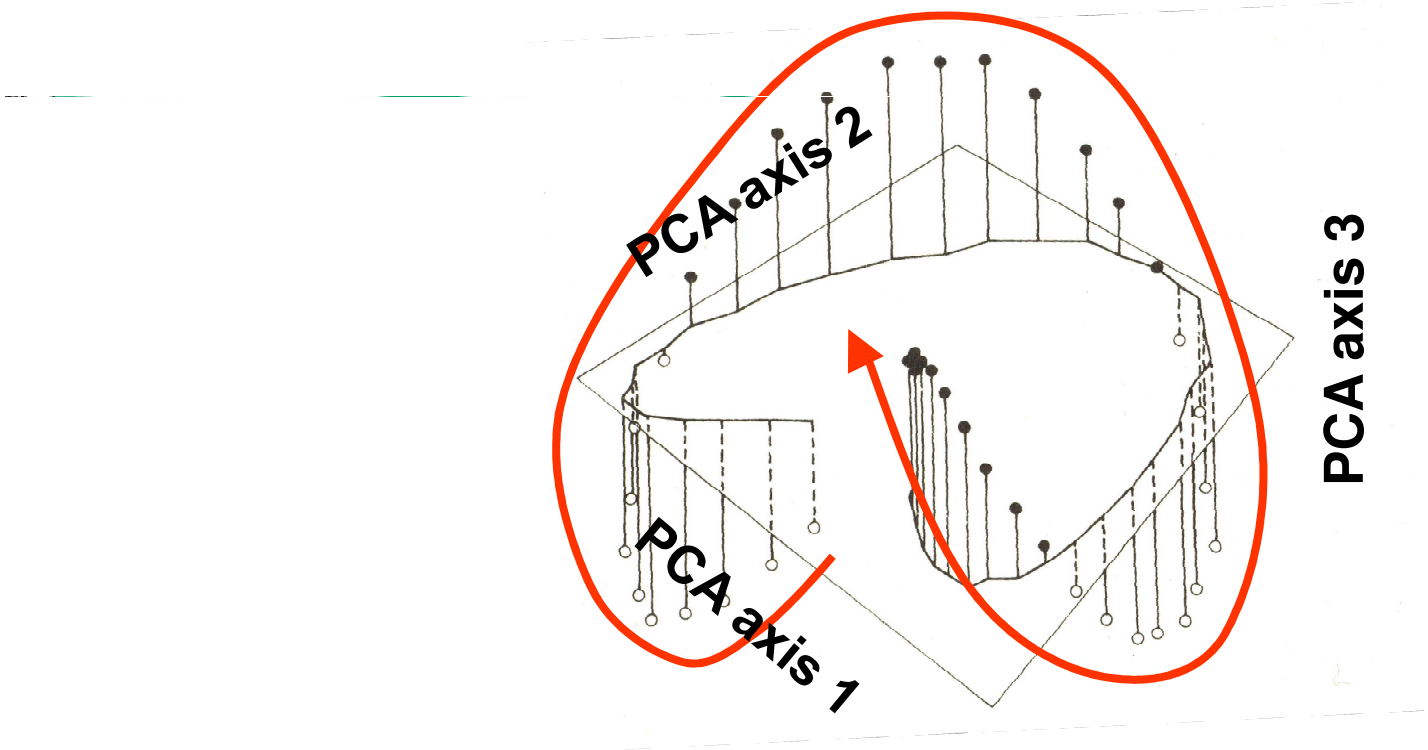
- "Eyeballing" – Does it make sense?
- Summary stats:
  - variance explained (PCA) $(\lambda_i / \Sigma \lambda_i)$ * **100%**
  - correlations with axes (all methods)
  - stress (NMS)

$$\sqrt{\sum_{i<j} (d_{ij} - \hat{d}_{ij})^2 / \sum_{i<j} d_{ij}^2}$$

- Performance with simulated data:
  - coenocline: single dominant gradient
  - coenoplane: two (orthogonal) gradients

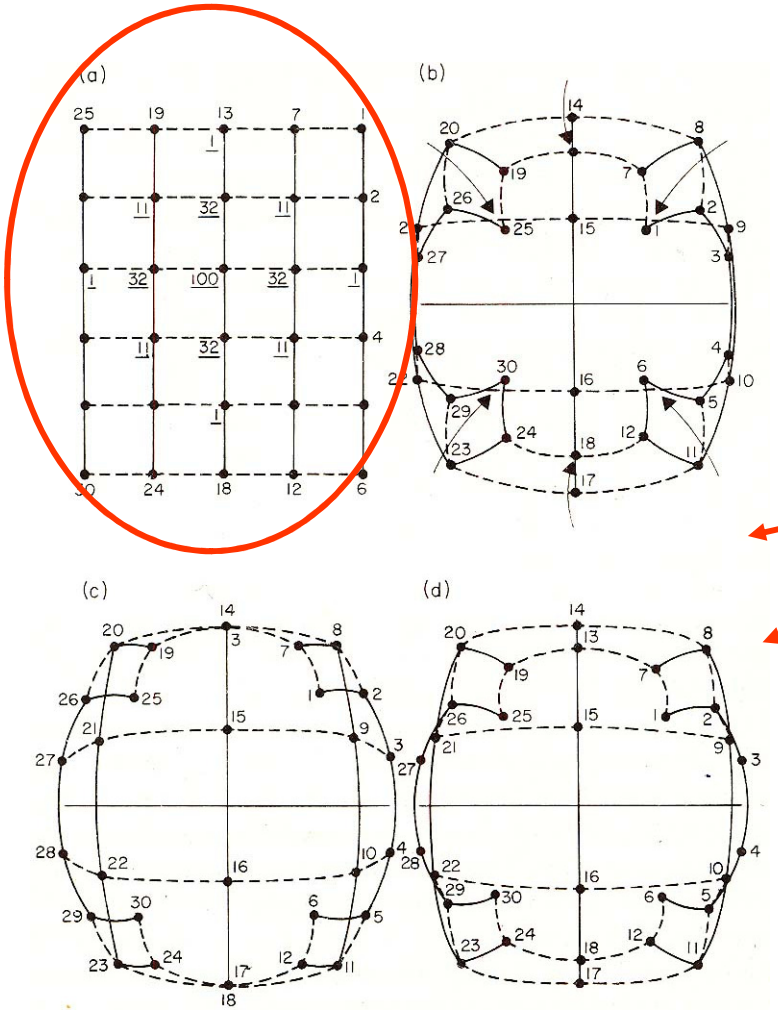
Simulated data: 1-D coenocline (>2 species, 1 gradient)



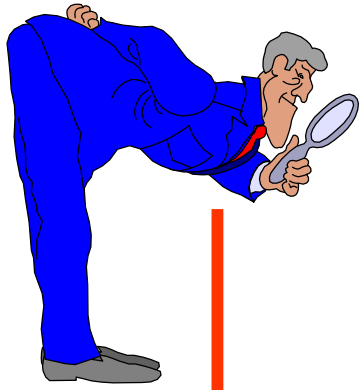


Simulated data: 2-D coenoplane (>2 species, 2 gradients)

Sampling grid (30 plots x 30 species)



PCA ordinations (various data standardizations)

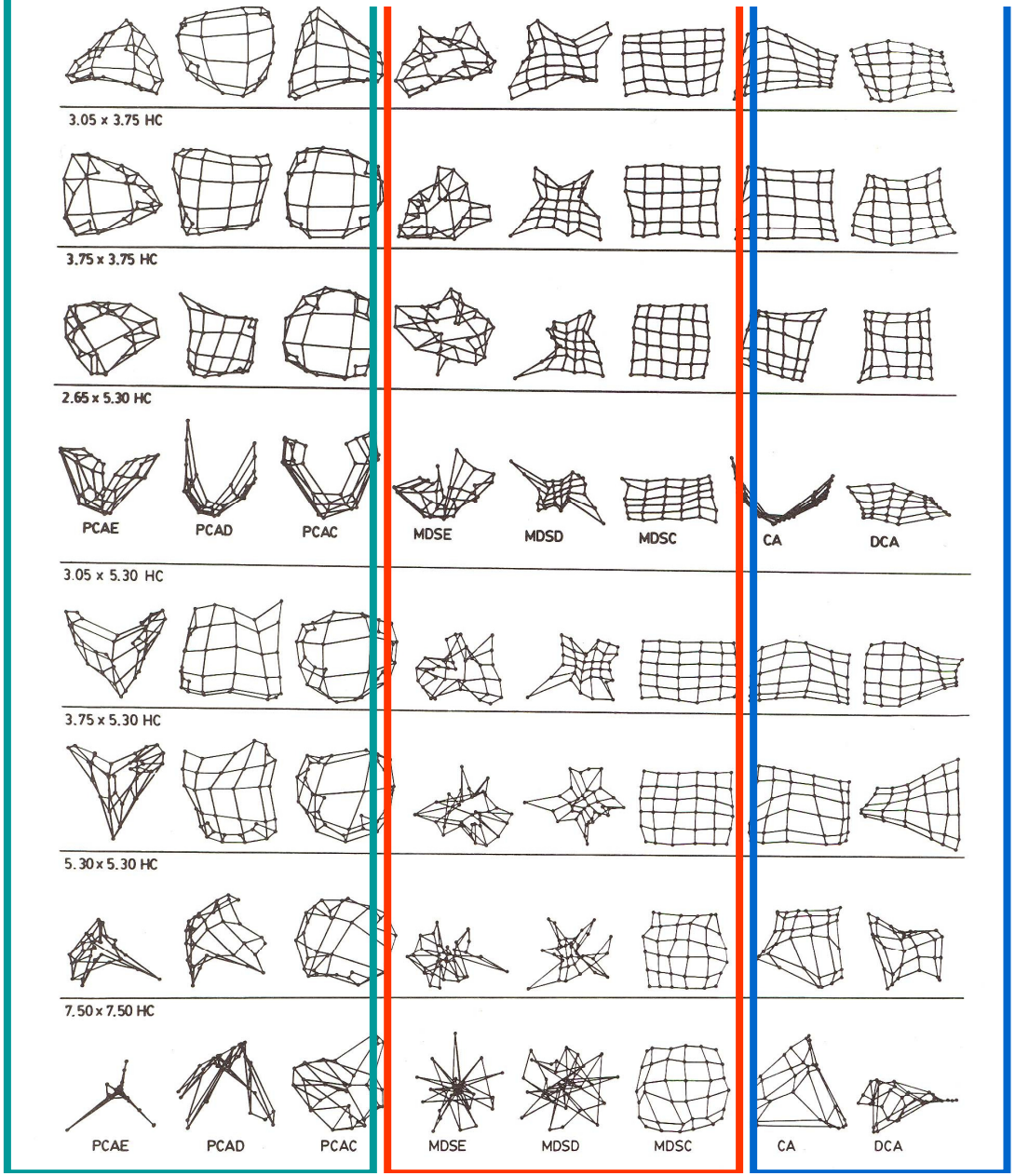


Increasing half-changes

Applying Metric and Nonmetric Multidimensional Scaling to Ecological Studies: Some New Results

N. C. Kenkel and L. Orlóci

Ecology, Vol. 67, No. 4 (Aug., 1986), pp. 919-928



PCA

MDS

CA & DCA

SUMMARY : ORDINATION STRATEGY

1. Data transformation.
2. Standardization of variables and/or sampling units.
3. Selection of ordination method.

CHOICES AT STEPS 1 and 2 ARE AS IMPORTANT AS CHOICE AT STEP 3.

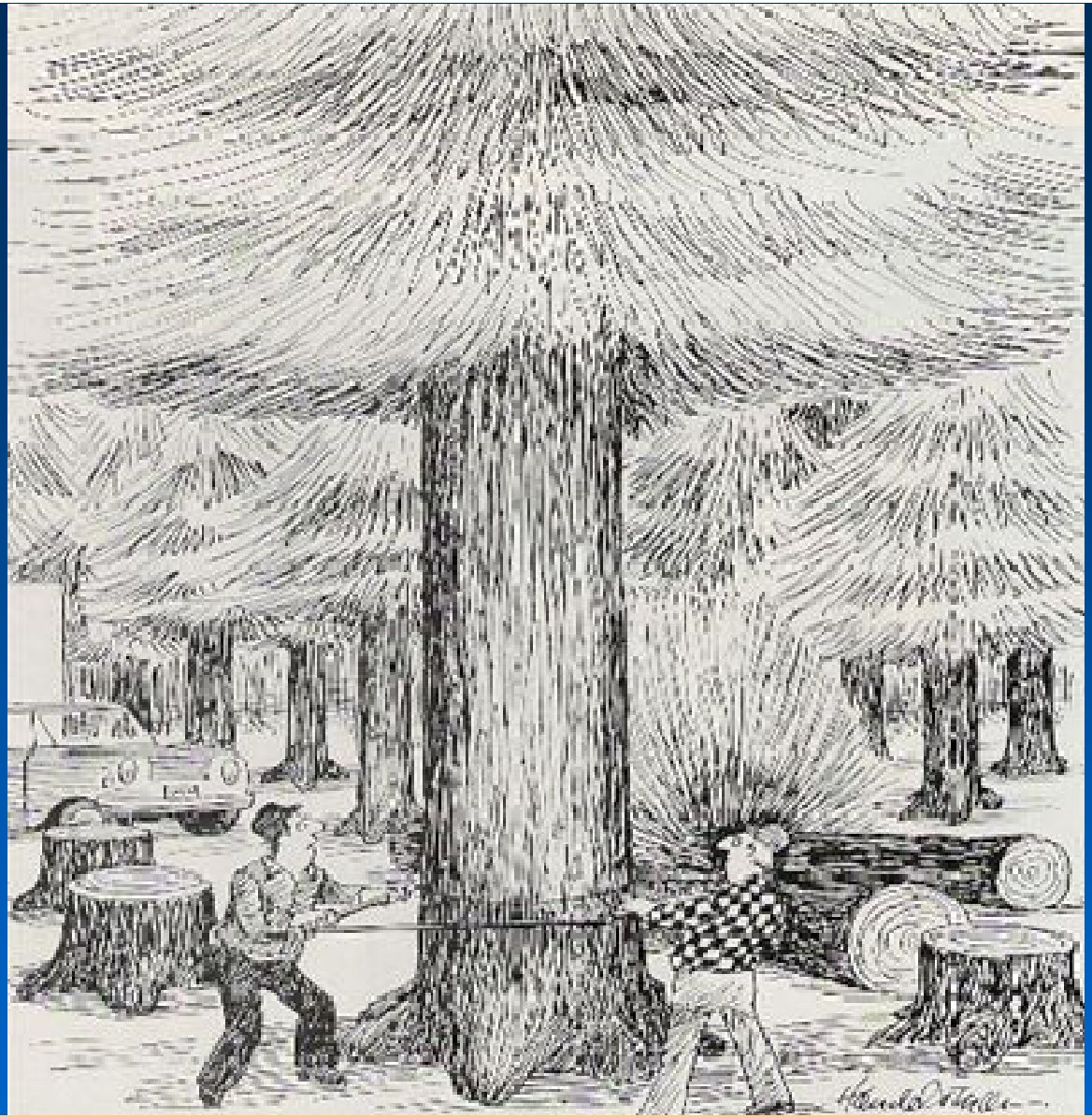
SUMMARY: ORDINATION RECOMMENDATIONS

- Abiotic (environment) survey data:
  - Principal Component Analysis.
  - Standardize variables to “z-scores” (correlation).
  - Log-transform data (continuous variables).

- Biotic (species) survey data:
  - Principal Component Analysis.
  - Do not standardize variables.
  - Log-transform data (continuous variables).
  - Examine results carefully for evidence of unimodal species responses. If so, try correspondence analysis (CA) but be aware that infrequent species may dominate.

NON-METRIC MULTIDIMENSIONAL SCALING
also good but…

- Limitations:
  - Iterative method: solution is not unique and may be sub-optimal or degenerate.
  - Ordination axes merely define a coordinate system: order and direction are meaningless concepts.
  - Variable weights (biplot scores) are not produced.
  - Ordination configuration is based only on ranks, not absolutes.
  - User must choose distance measure, and solution is highly dependent on measure chosen.



"That's life. You stand straight and tall and proud for a thousand years and the next thing you know, you're junk mail."