

Comunidades de Ambientes lênticos

Aula 8 – Introdução à análise
multivariada de dados

- **Objetivos Principais**
 - Principal direção de variação dos dados, correlação de matrizes, diferenças entre grupos
 - Índices de dissimilaridade
- **Análises Exploratórias x Testes**
 - Hipóteses (expectativas)

MULTIVARIATE ANALYSIS IN ECOLOGY AND SYSTEMATICS: PANACEA OR PANDORA'S BOX?

Frances C. James

Department of Biological Science, Florida State University, Tallahassee, Florida
32306

Charles E. McCulloch

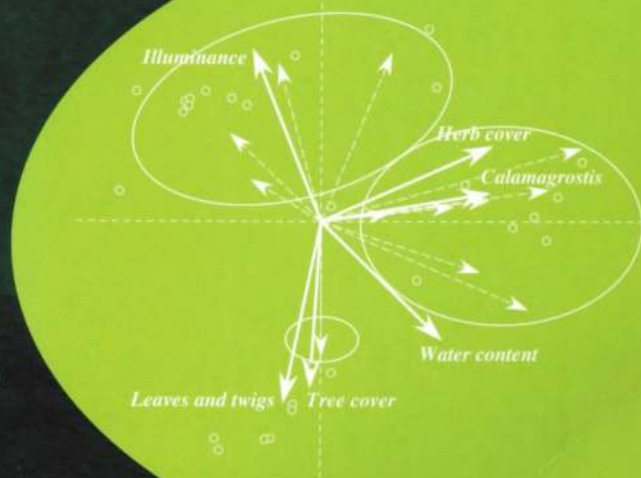
Biometrics Unit, Cornell University, Ithaca, New York 14853



Third English
Edition

Numerical Ecology

Pierre Legendre
Louis Legendre

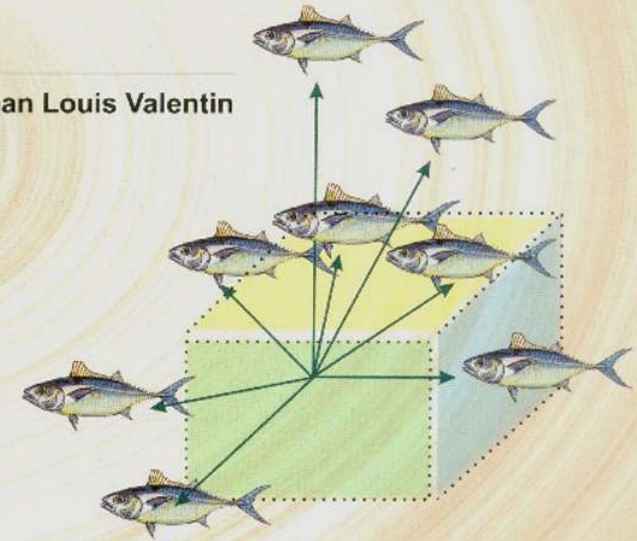


Ecologia numérica

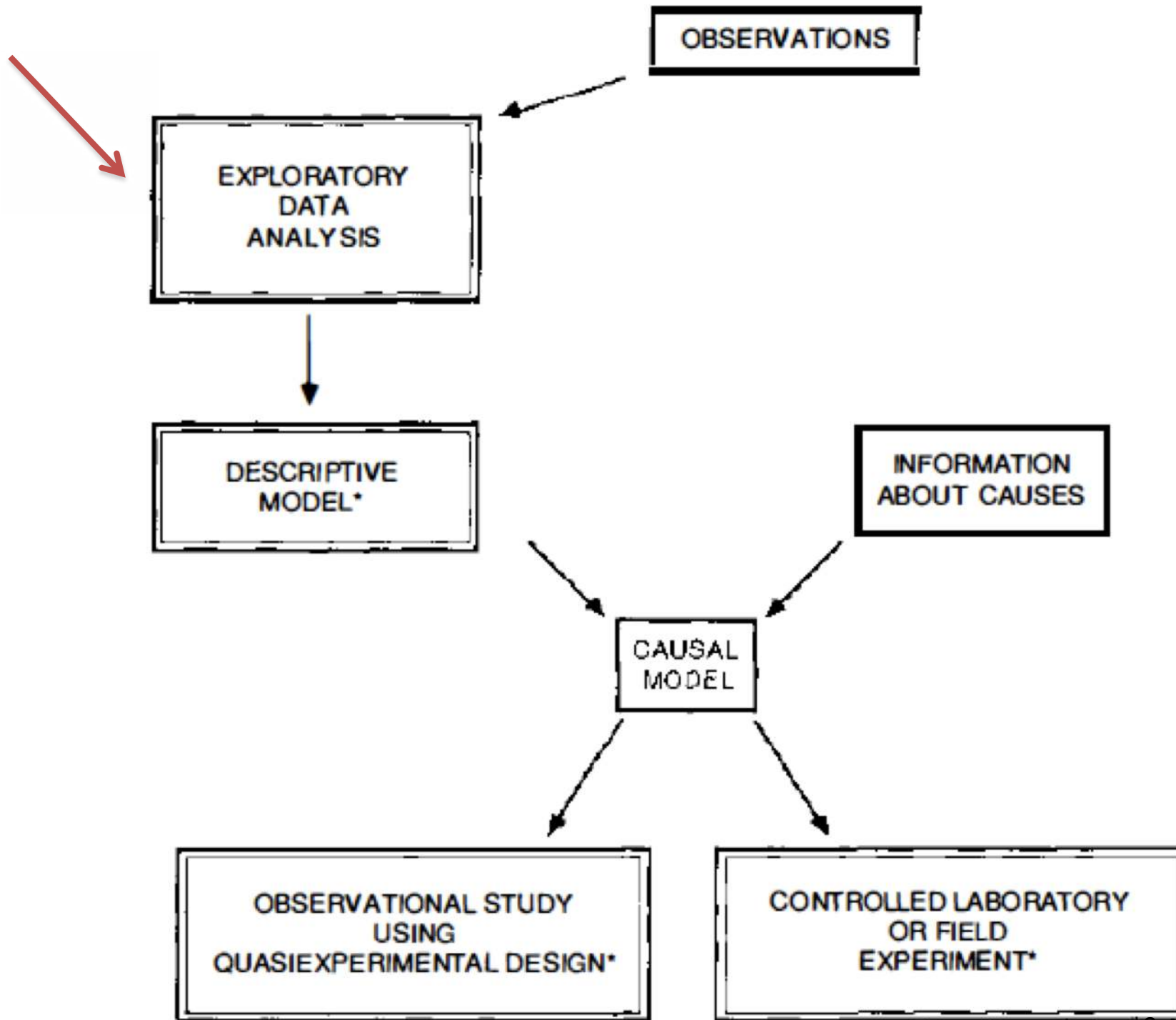
UMA INTRODUÇÃO À ANÁLISE MULTIVARIADA
DE DADOS ECOLÓGICOS

2ª Edição

Jean Louis Valentin



EDITORA INTERCIÊNCIA

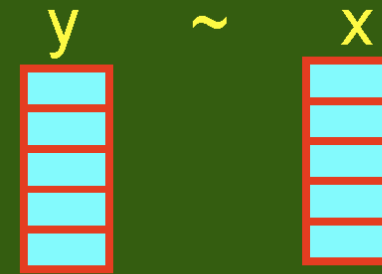


- Organização de dados
 - Variáveis na coluna e unidades amostrais nas linhas
- Tipos de testes

Análises Multivariadas

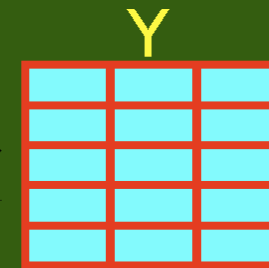
Univariadas $\rightarrow y \sim x$

Regressão, Anova, Reg. Logística



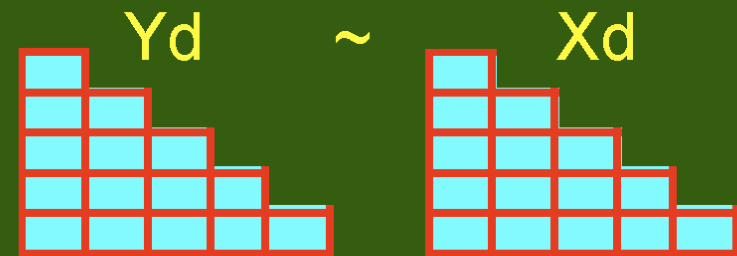
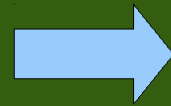
Multivariadas exploratórias $\rightarrow Y$

Ordenação e Classificação



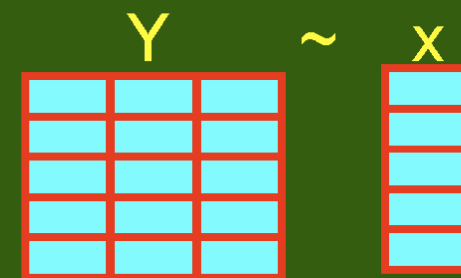
Multivariadas (testes) $\rightarrow Y_d \sim X_d$

Teste de Mantel



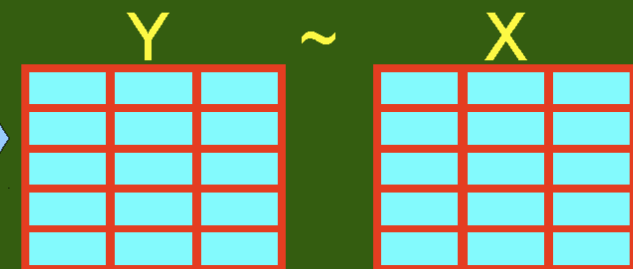
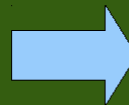
Multivariadas (testes) $\rightarrow Y \sim x$

Manova, db-Manova



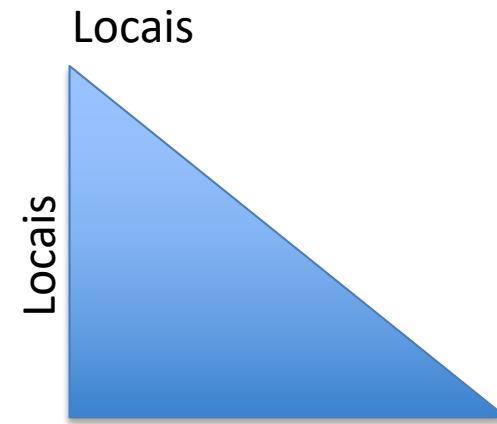
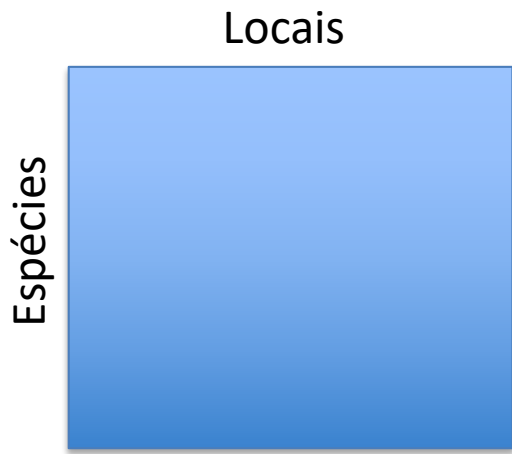
Multivariada (ordenações restritas) $\rightarrow Y \sim X$

CCA, RDA, CapScale



Índices de dissimilaridade ou distância

- Vários índices que medem a diferença entre as espécies ou unidades amostrais num espaço n-dimensional
- A escolha depende do tipo de dado



Redução de dimensões
Visualização



Table 1 Dissimilarity coefficients compared in this article

Dissimilarity	Abundance-based	Incidence-based	References	Coefficient no. in L&L*
Euclidean distance	$\sqrt{\sum_{j=1}^p [y_{1j} - y_{2j}]^2}$	$\sqrt{p \left(\frac{b+c}{a+b+c+d} \right)} = \sqrt{b+c}$		D ₁
Manhattan distance	$\sum_{j=1}^p y_{1j} - y_{2j} $	$p \left(\frac{b+c}{a+b+c+d} \right) = b+c$		D ₇
Modified mean character difference	$\frac{1}{pp} \sum_{j=1}^p y_{1j} - y_{2j} $	$\frac{b+c}{a+b+c}$	Legendre & Legendre (2012)	D ₁₉
Species profile distance	$\sqrt{\sum_{j=1}^p \left[\frac{y_{1j}}{y_{1+}} - \frac{y_{2j}}{y_{2+}} \right]^2}$	$\sqrt{\frac{b+c}{(a+b)(a+c)}}$	Legendre & Gallagher (2001)	D ₁₈
Hellinger distance	$\sqrt{\sum_{j=1}^p \left[\sqrt{\frac{y_{1j}}{y_{1+}}} - \sqrt{\frac{y_{2j}}{y_{2+}}} \right]^2}$	$\sqrt{2 \left(1 - \frac{a}{\sqrt{(a+b)(a+c)}} \right)}$	Rao (1995)	D ₁₇
Chord distance	$\sqrt{\sum_{j=1}^p \left[\frac{y_{1j}}{\sqrt{\sum_{k=1}^p y_{1k}^2}} - \frac{y_{2j}}{\sqrt{\sum_{k=1}^p y_{2k}^2}} \right]^2}$	$\sqrt{2 \left(1 - \frac{a}{\sqrt{(a+b)(a+c)}} \right)}$	Orlóci (1967)	D ₃
Chi-square distance	$\sqrt{y_{++} \sum_{j=1}^p \frac{1}{y_{+j}} \left[\frac{y_{1j}}{y_{1+}} - \frac{y_{2j}}{y_{2+}} \right]^2}$	NA [†]	Lebart & Fénelon (1971)	D ₁₆
Coefficient of divergence	$\sqrt{\frac{1}{pp} \sum_{j=1}^p \left(\frac{y_{1j} - y_{2j}}{y_{1j} + y_{2j}} \right)^2}$	$\sqrt{\frac{b+c}{a+b+c}}$	Clark (1952)	D ₁₁
Canberra metric [‡]	$\frac{1}{pp} \sum_{j=1}^p \frac{ y_{1j} - y_{2j} }{(y_{1j} + y_{2j})}$	$\frac{b+c}{a+b+c}$	Lance & Williams (1967), Stephenson <i>et al.</i> (1972) for $1/pp$	D ₁₀
Whittaker's index of association	$\frac{1}{2} \sum_{j=1}^p \left \frac{y_{1j}}{y_{1+}} - \frac{y_{2j}}{y_{2+}} \right $	$\frac{1}{2} \left(\frac{b}{a+b} + \frac{c}{a+c} + \left \frac{a}{a+b} - \frac{a}{a+c} \right \right)$	Whittaker (1952)	D ₉
Percentage difference (<i>alias</i> Bray-Curtis dissimilarity [§])	$\frac{\sum_{j=1}^p y_{1j} - y_{2j} }{y_{1+} + y_{2+}}$	$\frac{b+c}{2a+b+c}$	Odum (1950)	D ₁₄
Wishart coefficient = (1-similarity ratio)	$1 - \left[\frac{\sum_{j=1}^p y_{1j} y_{2j}}{\sum_{j=1}^p y_{1j}^2 + \sum_{j=1}^p y_{2j}^2 - \sum_{j=1}^p y_{1j} y_{2j}} \right]$	$\frac{b+c}{a+b+c}$	Wishart (1969), Janssen (1975)	
D = (1-Kulczynski coefficient)	$1 - \frac{1}{2} \left[\frac{\sum_{j=1}^p \min(y_{1j}, y_{2j})}{y_{1+}} + \frac{\sum_{j=1}^p \min(y_{1j}, y_{2j})}{y_{2+}} \right]$	$1 - \frac{1}{2} \left(\frac{a}{a+b} + \frac{a}{a+c} \right)$	Kulczynski (1928)	1 - S ₁₈
Abundance-based Jaccard [¶]	$\left(1 - \frac{UV}{U+V-UV} \right)$	$\frac{b+c}{a+b+c}$	Chao <i>et al.</i> (2006)	
Abundance-based Sorensen [¶]	$\left(1 - \frac{2UV}{U+V} \right)$	$\frac{b+c}{2a+b+c}$	Chao <i>et al.</i> (2006)	
Abundance-based Ochiai [¶]	$\left(1 - \sqrt{UV} \right)$	$\left(1 - \frac{a}{\sqrt{(a+b)(a+c)}} \right)$	Chao <i>et al.</i> (2006)	

*L&L: Legendre & Legendre (2012).

†NA: No binary form with parameters *a*, *b* and *c* for this coefficient, although it can be computed for presence-absence data.

‡Division by *pp* (number of species excluding double zeros) introduced by Stephenson *et al.* (1972) and adopted by Oksanen *et al.* (2012).

§Coefficient first described by Steinhaus in the 1940s, then by Odum (1950) as the *percentage difference*. The Bray & Curtis (1957) paper described a new ordination method; the index described and used by these authors was Whittaker's dissimilarity, not the percentage difference which is more general. It is incorrect to attribute this coefficient to these authors.

¶*U* and *V* notation: see Chao *et al.* (2006).

Idea and Perspective

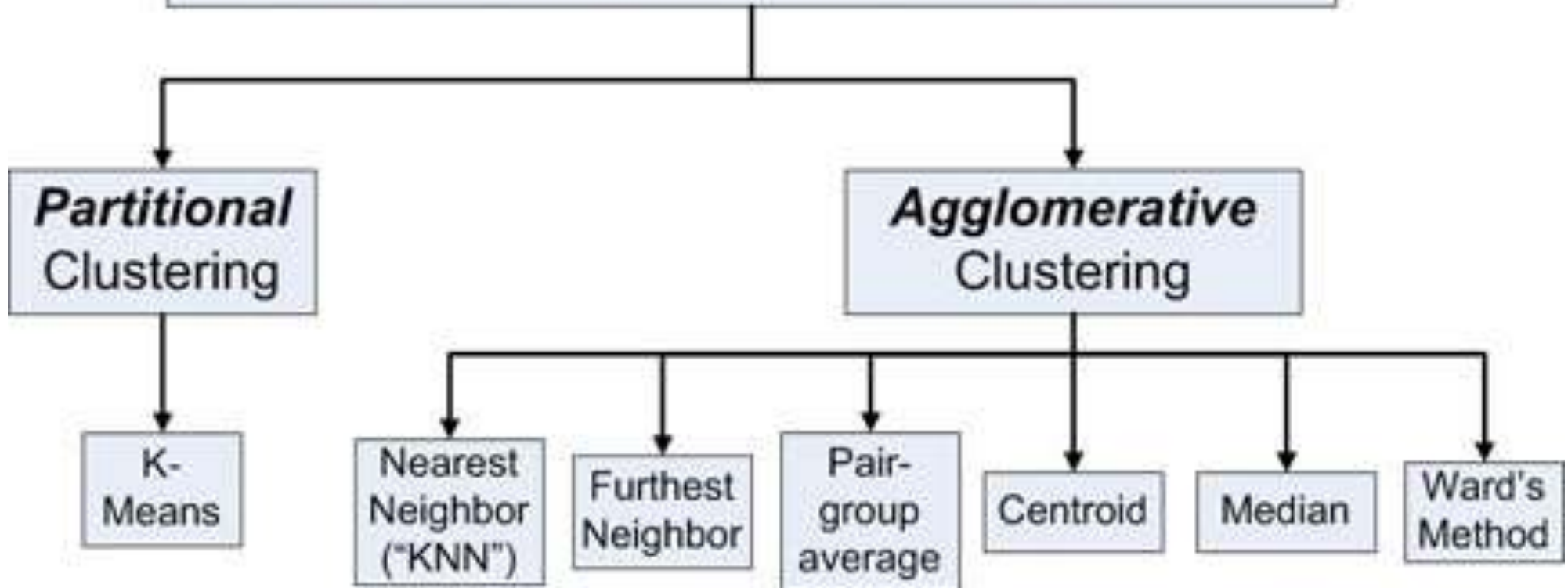
Beta diversity partitioning

Classificação

Agrupamento

- Análise de Cluster
- Análise de espécie indicadora (IndVal)

HIERARCHICAL CLUSTER ANALYSIS



Análise de cluster

- Resumir um grande volume de informação
- Agrupa objetos por grau de similaridade
- Busca por descontinuidades no conjunto de dados para formar grupos
- Descontinuidade X Gradiente

- Hierárquico
- Métodos de agrupamento
 - UPGMA e Ward
- Índices de Similaridade/Distância
 - Índices binários assimétricos: Jaccard, Sørensen
 - Índices quantitativos assimétricos: Bray-Curtis, Morisita-Horn, Gower
 - Índices métricos: Distância Euclidiana, Manhattan, Canberra

- Como avaliar a representatividade do dendrograma=> coef. Correlação cofenética
 - Correlação de Pearson entre a matriz de similaridade e a matriz cofenética. Regra do 0,8 .
- Nível de corte => Vários critérios (Bootstrap dos nós, 50% similaridade, inspeção visual etc.
 - Pacote pvclut

The Annals of Statistics

2004, Vol. 32, No. 6, 2616–2641

DOI 10.1214/009053604000000823

© Institute of Mathematical Statistics, 2004

**APPROXIMATELY UNBIASED TESTS OF REGIONS USING
MULTISTEP-MULTISCALE BOOTSTRAP RESAMPLING¹**

BY HIDETOSHI SHIMODAIRA

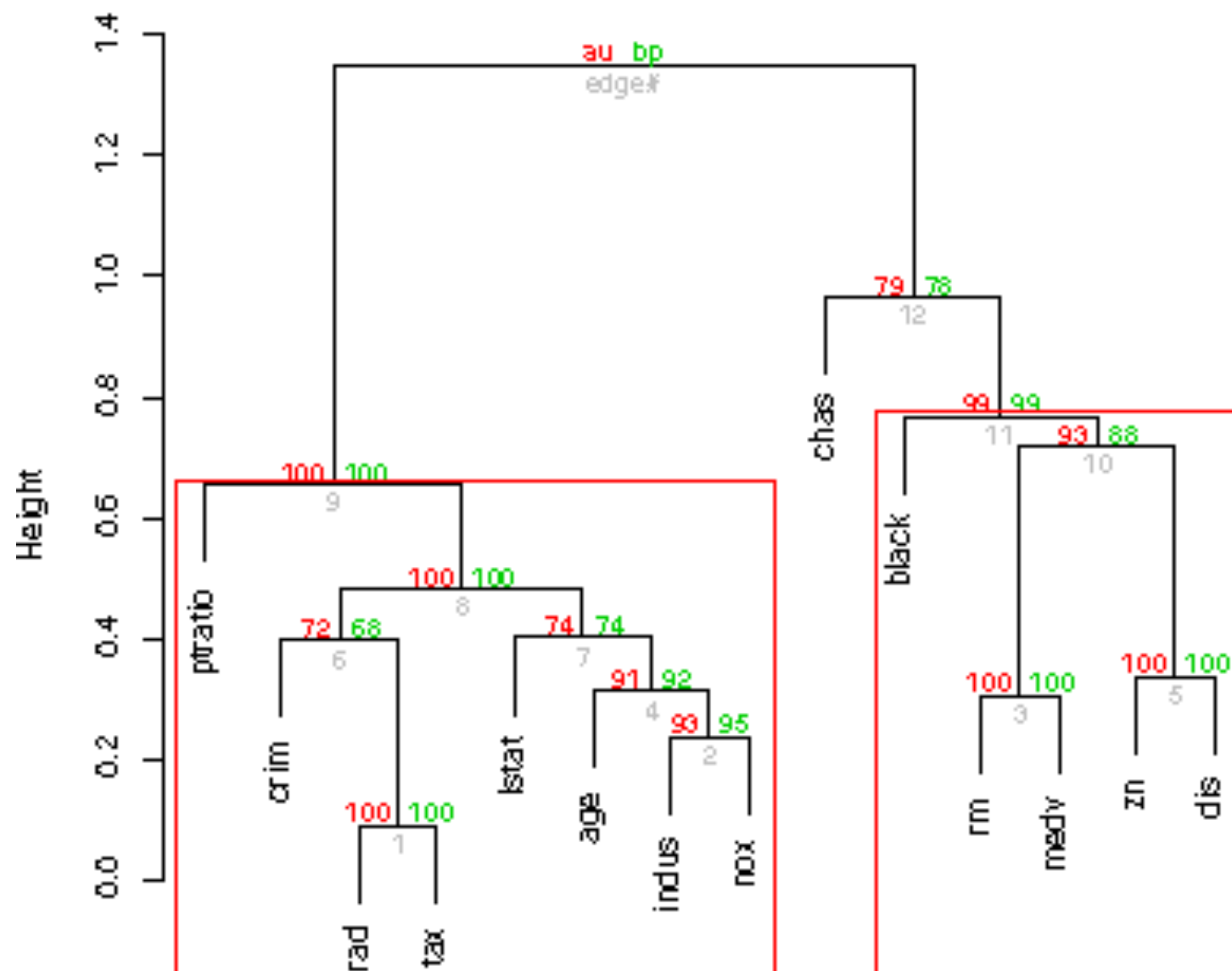
PVCLUST

- Reamostragem utilizando bootstrap em multiescala.
 - Utiliza 2 estimadores para calcular o valor de p:
AU e **BP**
 - Verifica se os agrupamentos (clusters) se manteriam se o tamanho amostral fosse aumentado.
 - função pvrect circula grupos com **AU** maiores que 95%

```
> result <- pvclust(lung, method.dist="cor", method.hclust="average", nboot=1000)
Bootstrap (r = 0.5)... Done.
Bootstrap (r = 0.6)... Done.
Bootstrap (r = 0.7)... Done.
Bootstrap (r = 0.8)... Done.
Bootstrap (r = 0.9)... Done.
Bootstrap (r = 1.0)... Done.
Bootstrap (r = 1.1)... Done.
Bootstrap (r = 1.2)... Done.
Bootstrap (r = 1.3)... Done.
Bootstrap (r = 1.4)... Done.
```

?pvclust

Cluster dendrogram with AU/BP values (%)



Distance: correlation
Cluster method: average

Análise de espécie indicadora (IndVal)

Ecological Monographs, 67(3), 1997, pp. 345–366
© 1997 by the Ecological Society of America

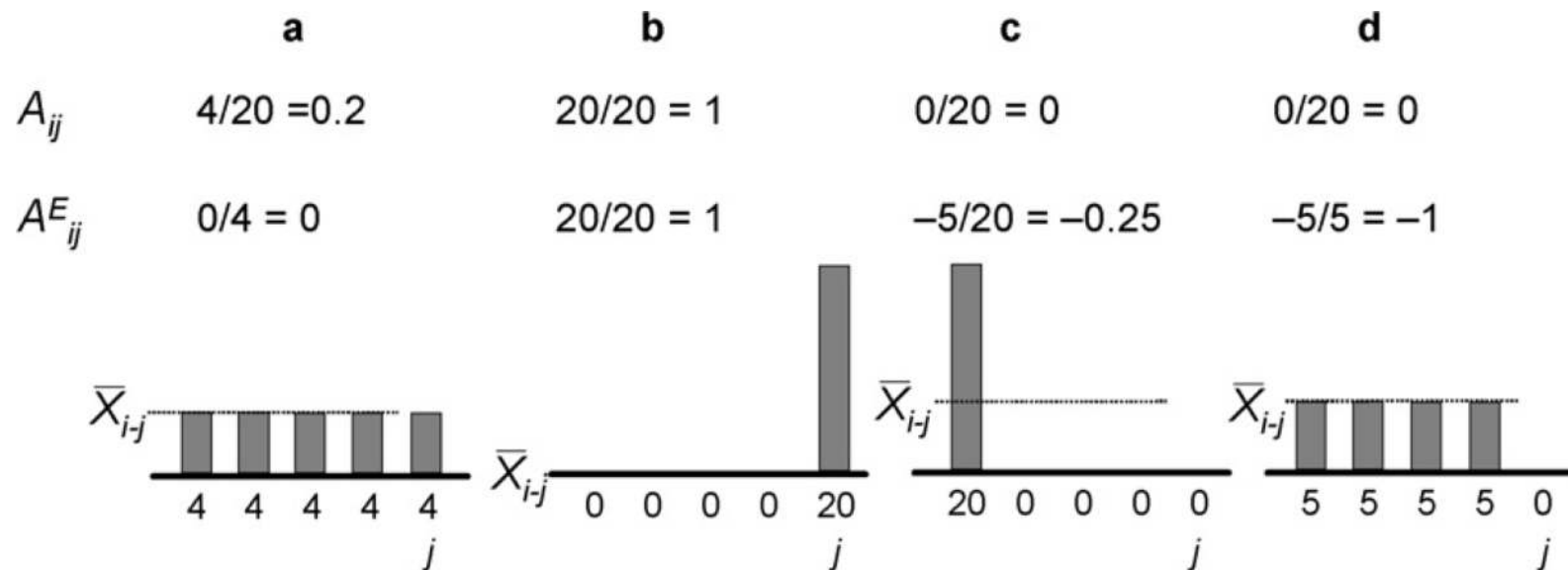
**SPECIES ASSEMBLAGES AND INDICATOR SPECIES:
THE NEED FOR A FLEXIBLE ASYMMETRICAL APPROACH**

MARC DUFRÊNE¹ AND PIERRE LEGENDRE²

Veja também De Cáceres & Legendre 2009, De Cáceres et al. 2012

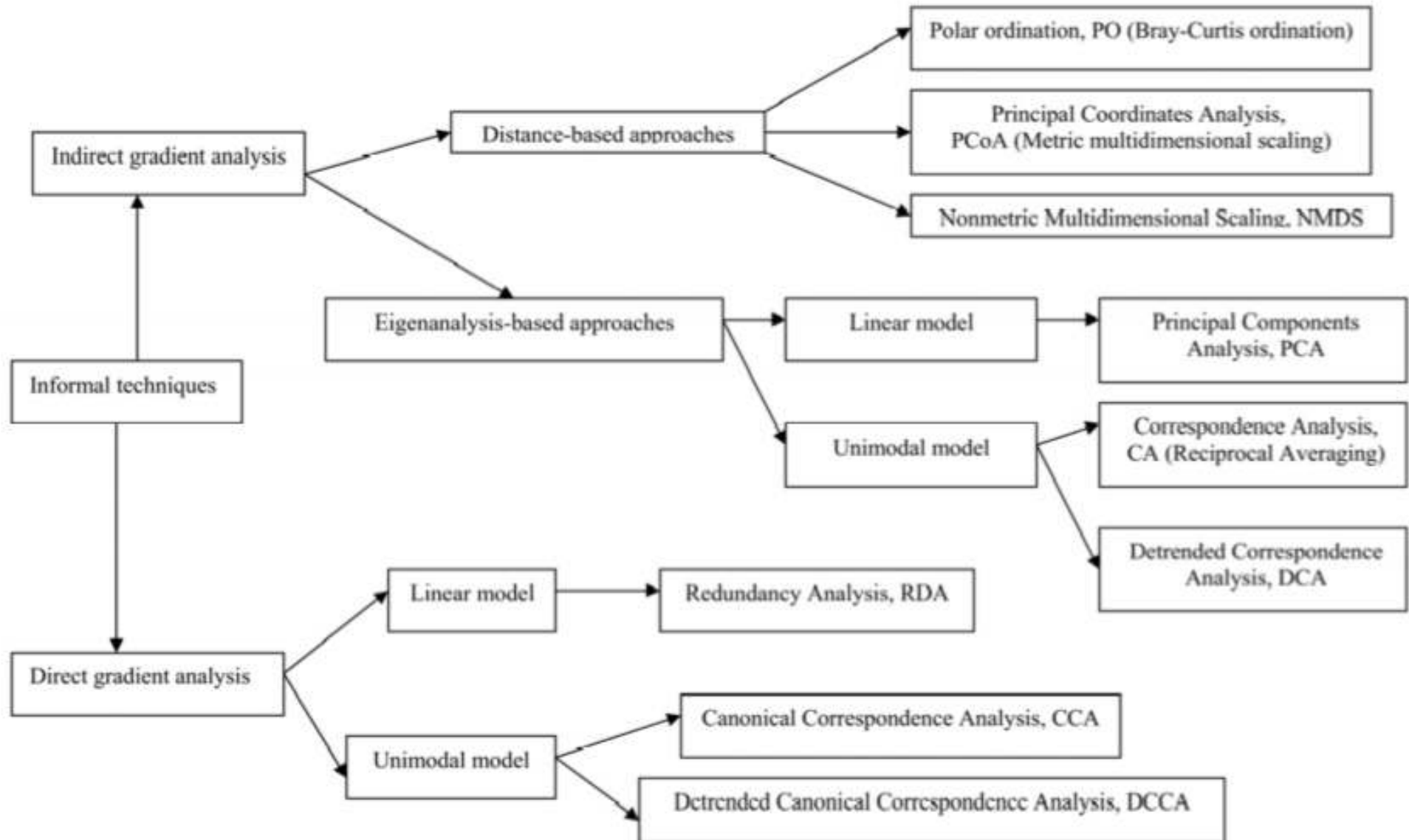
IndVal

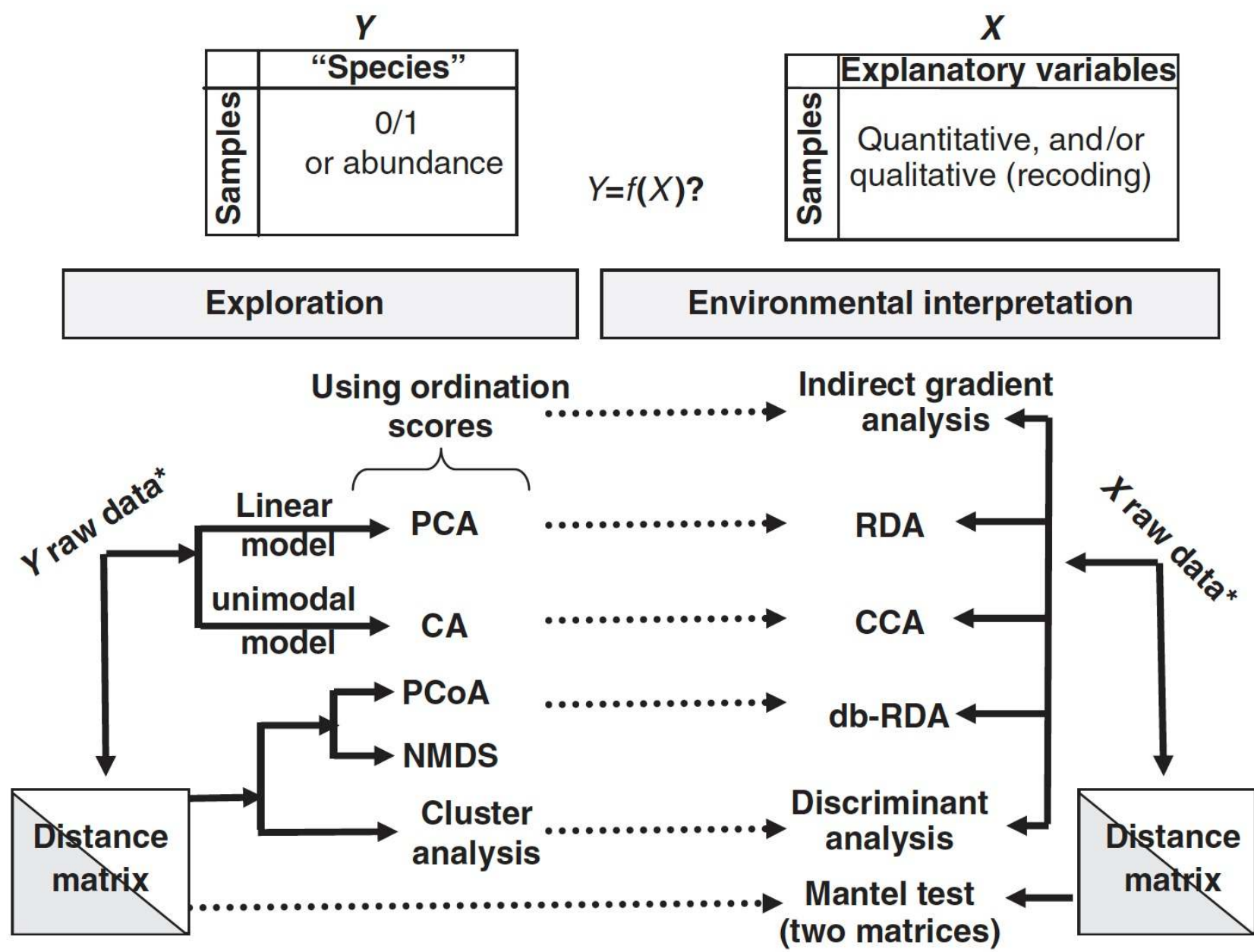
- Conceitos
 - Especificidade => todas as u.a. do grupo
 - Fidelidade => só em 1 grupo



Ordenação

Árvore de decisão de análises de ordenação





Introdução

- Concebidas para análises psicométricas e sociológicas.
- Introduzidas na ecologia para classificação de comunidades vegetais.



Biol. Rev. (1967), 49, pp. 207-264

207

GRADIENT ANALYSIS OF VEGETATION*

By

R. H. WHITTAKER



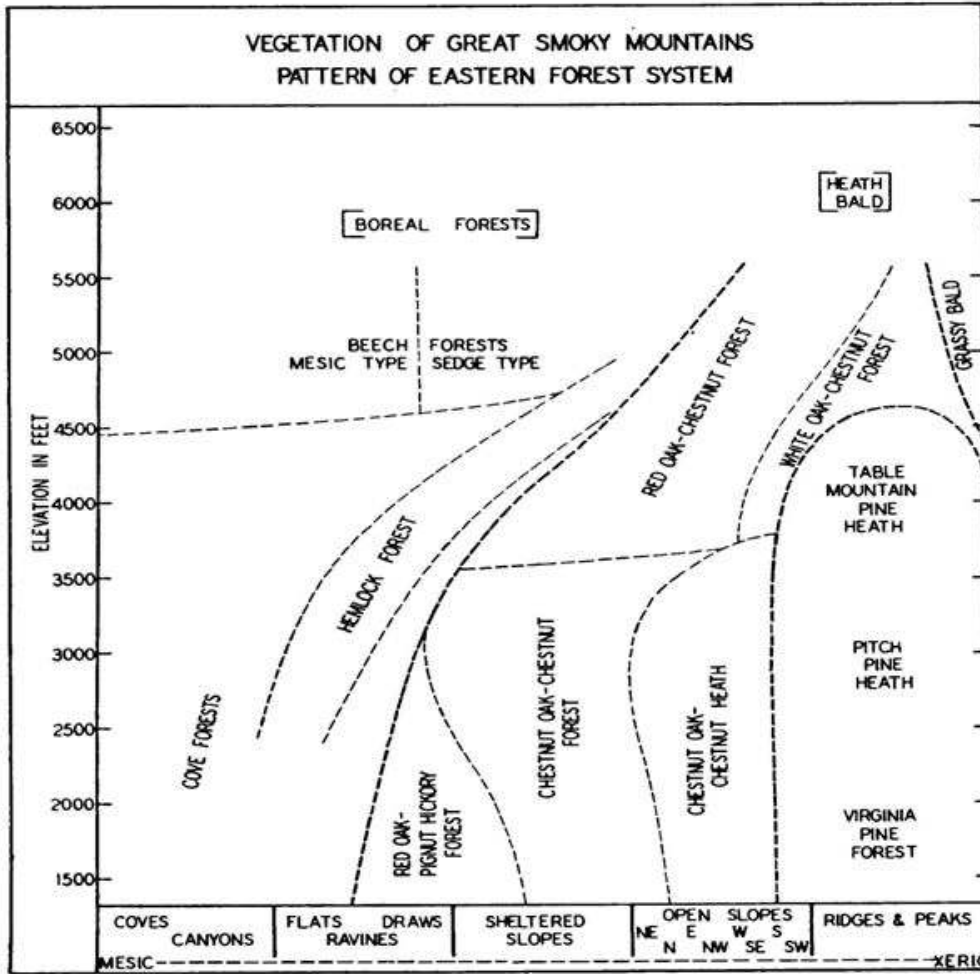
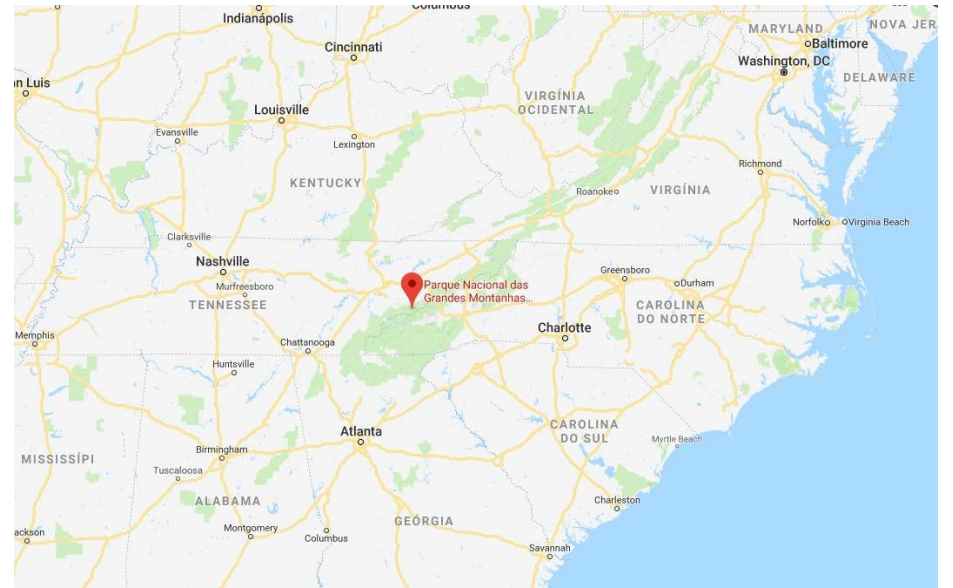


FIG. 19. (Vegetation of Great Smoky Mountains, pattern of Eastern Forest System.)

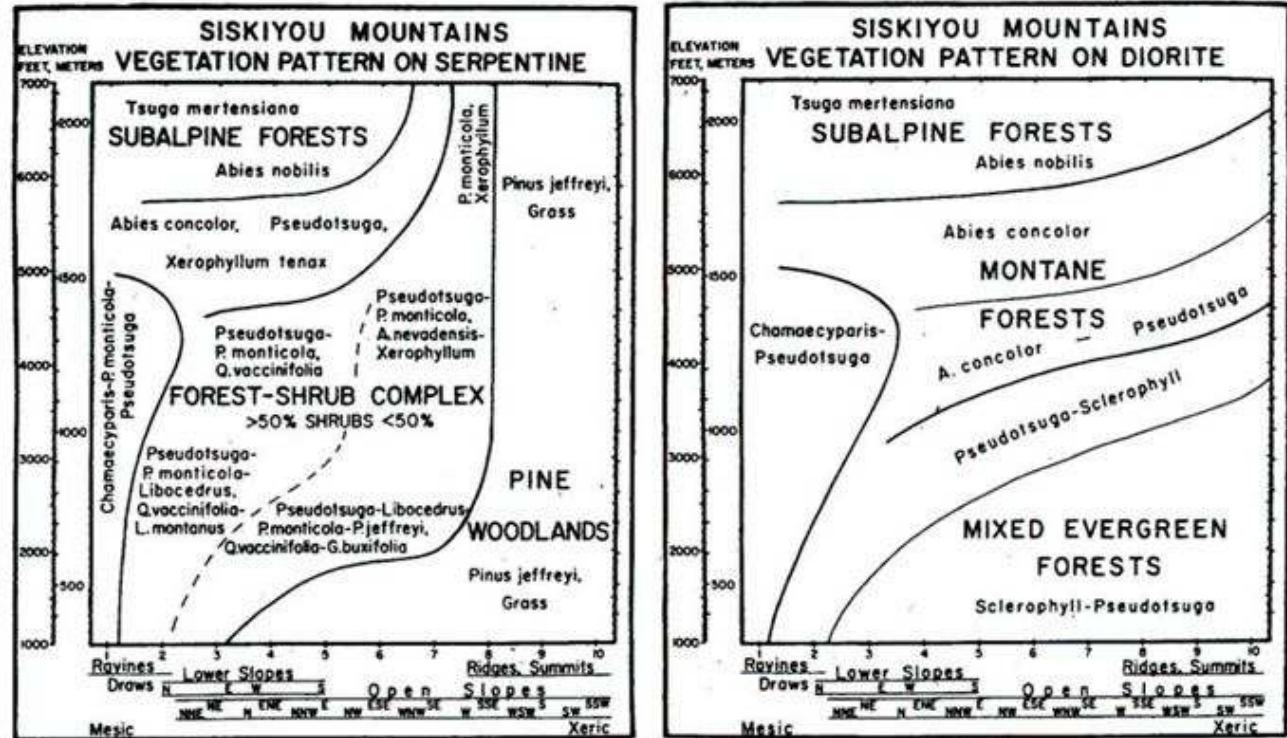


Fig. 5. Mosaic diagrams for vegetation on quartz diorite (right) and serpentines (left) in the Siskiyou Mountains, southwestern Oregon (WHITTAKER 1960, cf. WHITTAKER 1956, WHITTAKER & NIERING 1965, 1968b). Vegetation samples were classified (into dominance-types) and plotted by elevation and topographic position on the chart. Boundaries for community-types were drawn at average positions of transitions between types. Vegetation on each parent material is considered a complex pattern of continuously intergrading communities (WHITTAKER 1960).

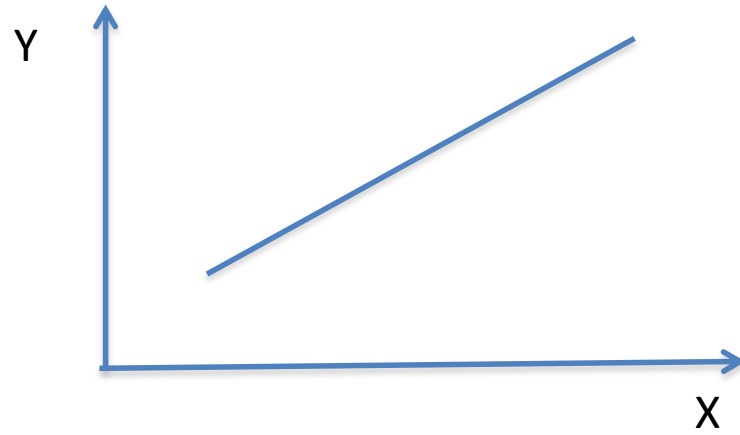
-
-

Não se perca na sopa de
letrinhas

Ordenação irrestrita

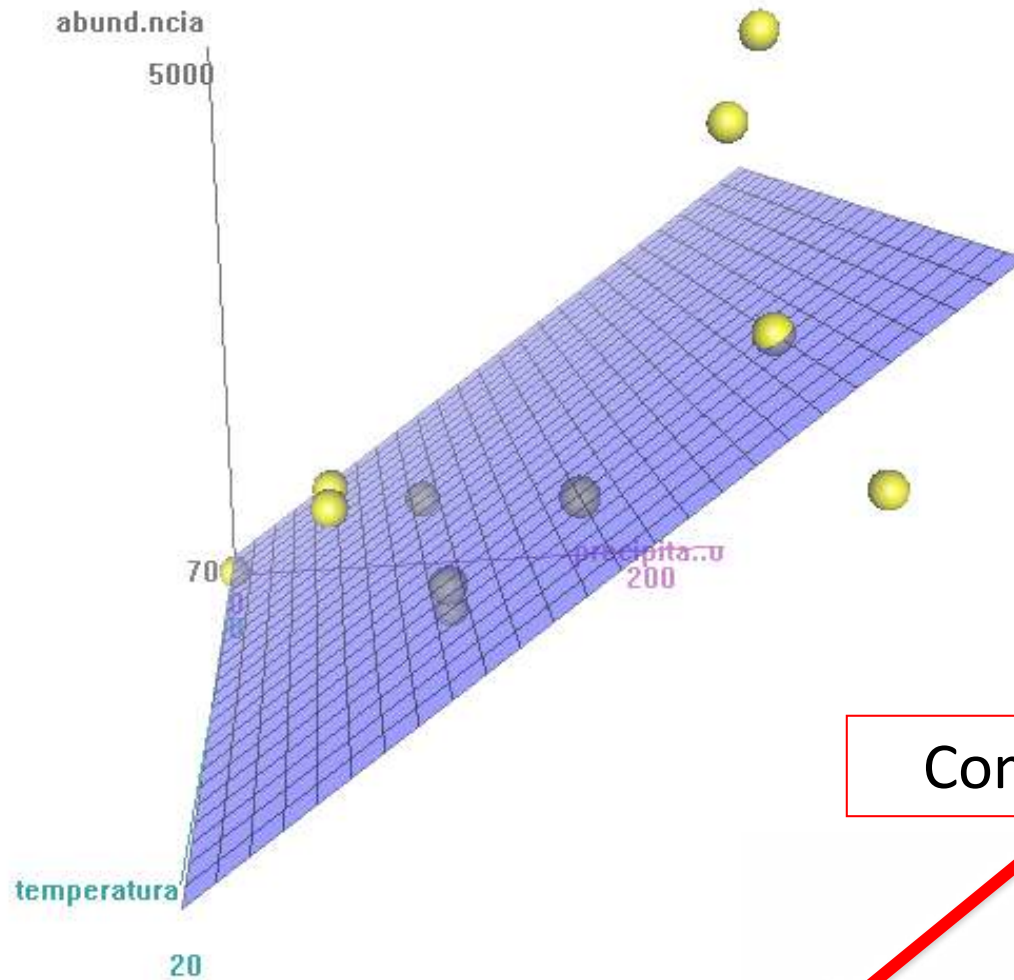
- PCA = Análise de Componentes Principais
- PCoA= Análise de Coordenadas Principais
 - Também chamada de Escalonamento multidimensional métrico

PCA



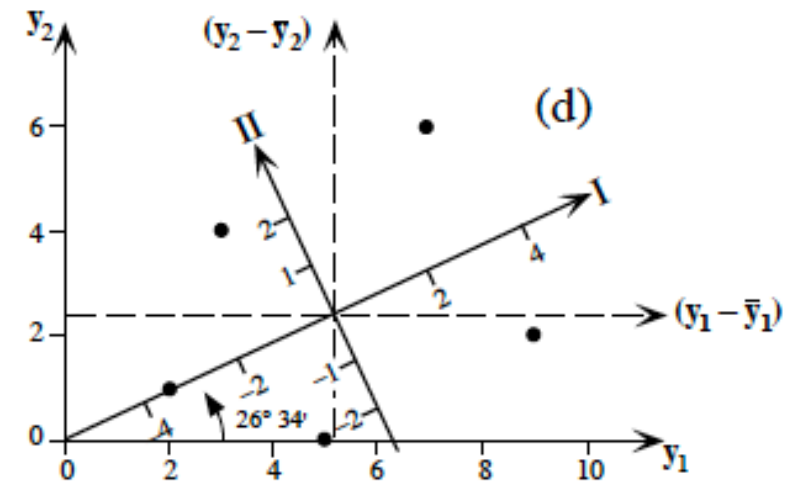
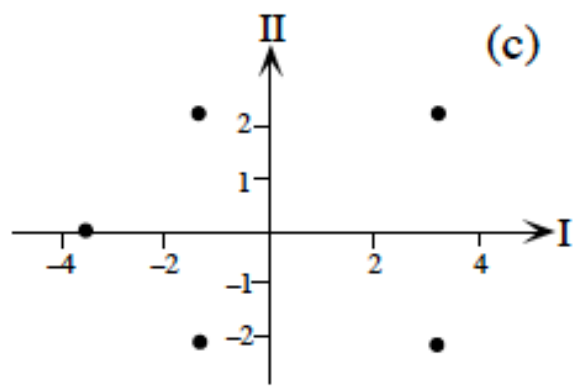
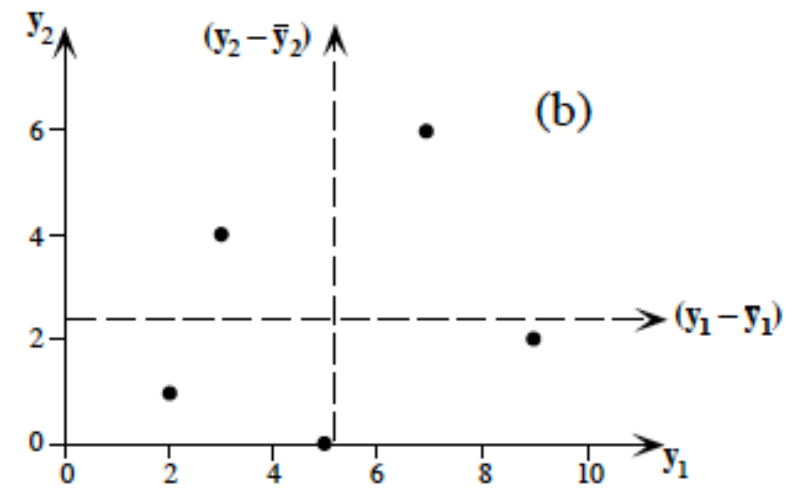
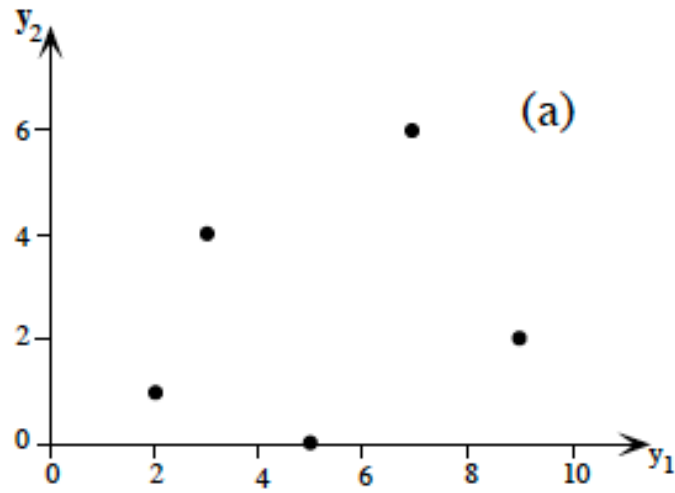
$$Y = a + bX$$

PCA



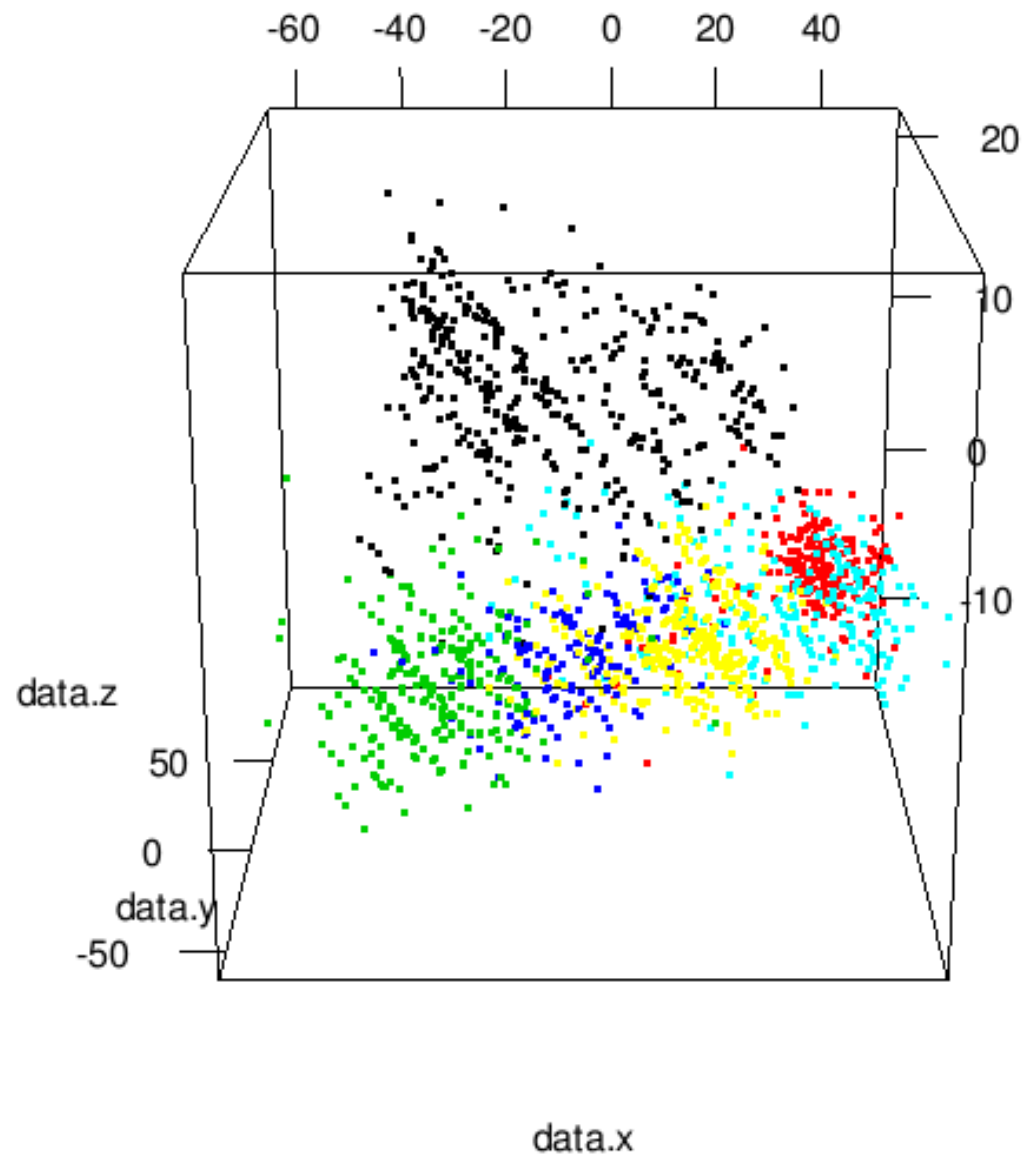
Combinação Linear

$$Y = a + bX_1 + bX_2 + \dots + bX_n$$

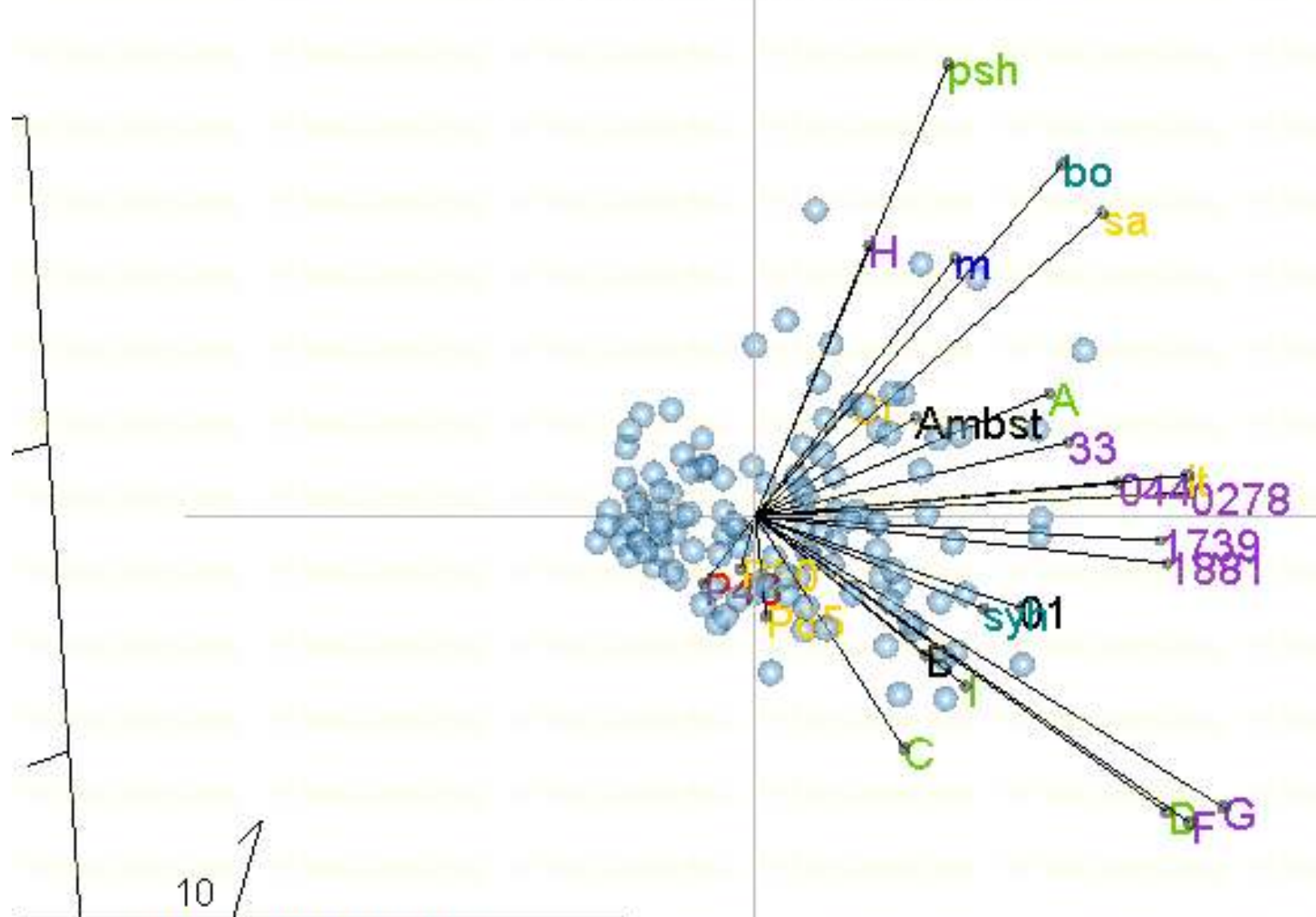


PCA

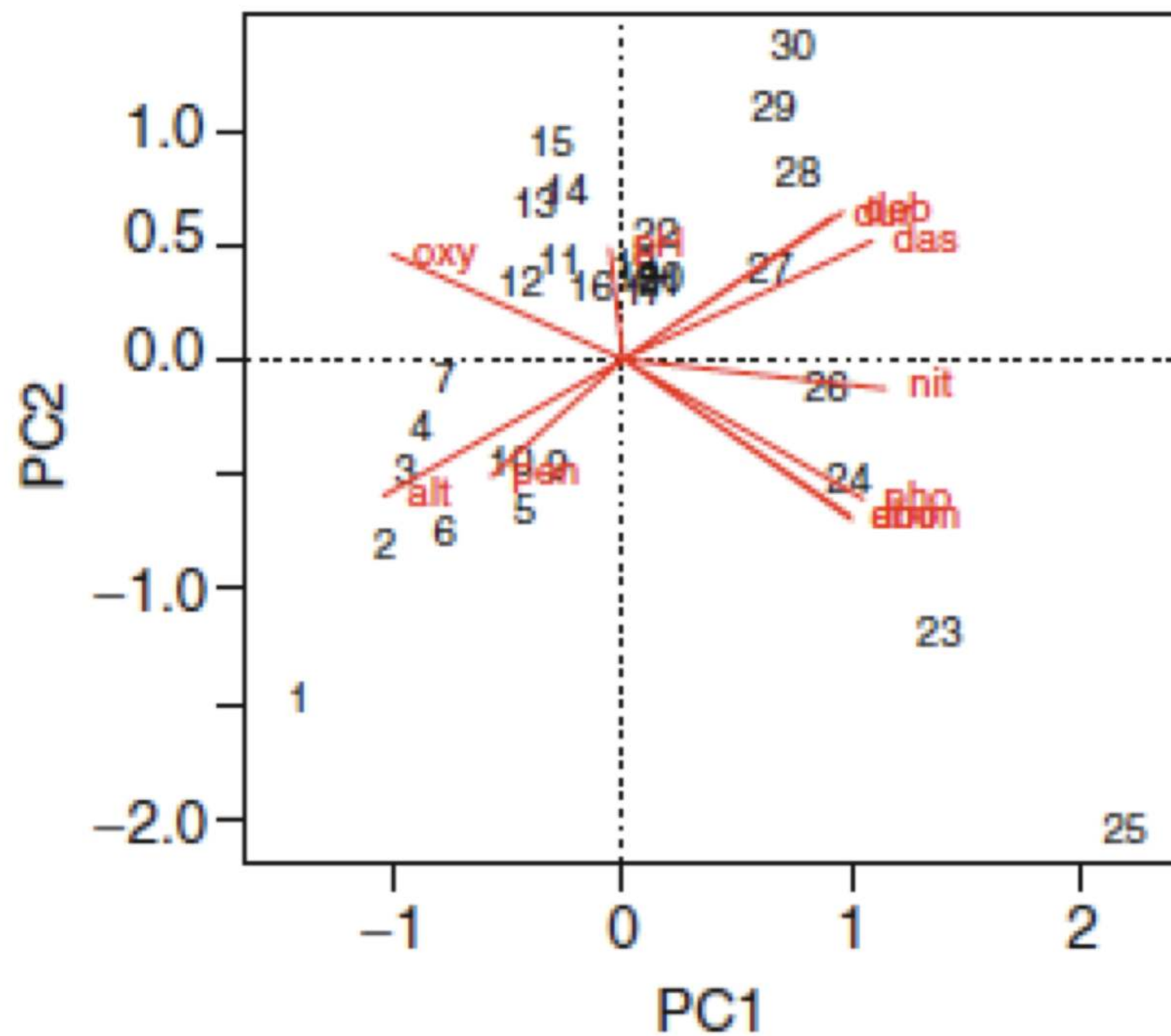
- Só trabalha com distâncias Euclidianas
 - Dados contínuos (Medidas de distância, biomassa, etc)



PCA Biplot 1 Thessalonians



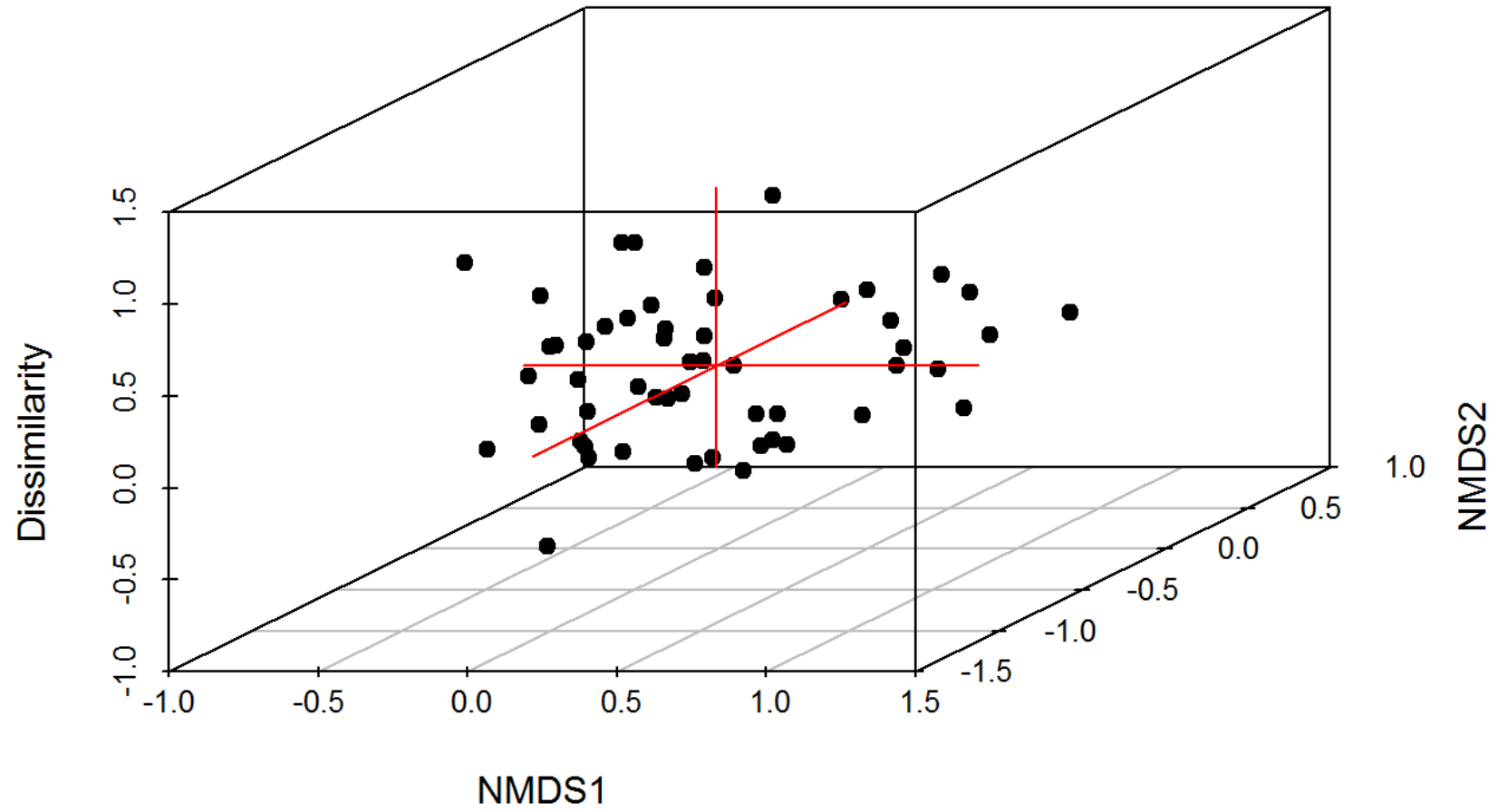
Witness Colors by Century							
	II		IV		VI		VIII
	III		V		VII		IX >



PCoA

- Similar à PCA mas pode trabalhar com qualquer medida de distância
 - Abundância, frequência relativa, % cobertura, etc

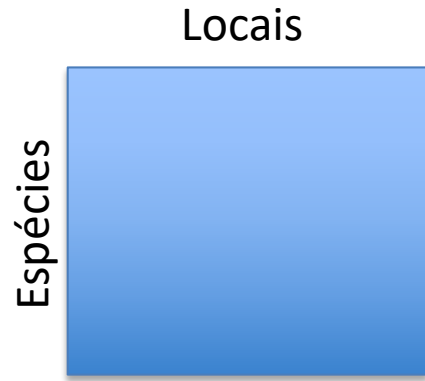
3-Dimensions



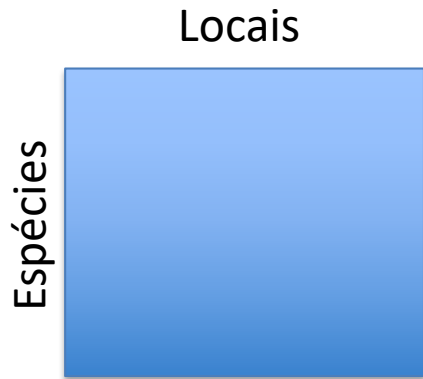
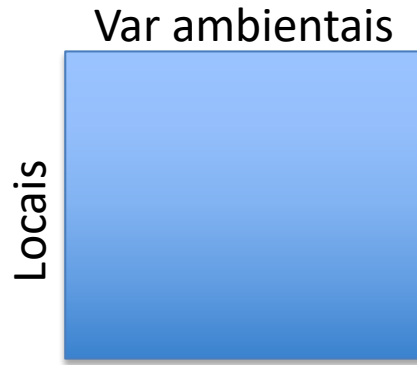
Ordenação restrita

- CCA (Distância de χ^2)
- RDA (Distância Euclidiana)

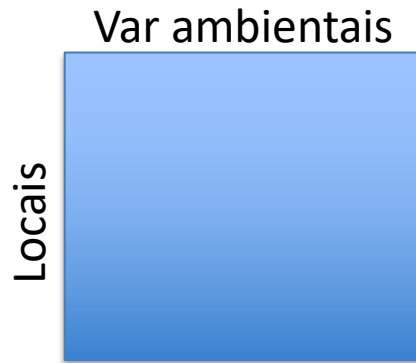
CCA



~



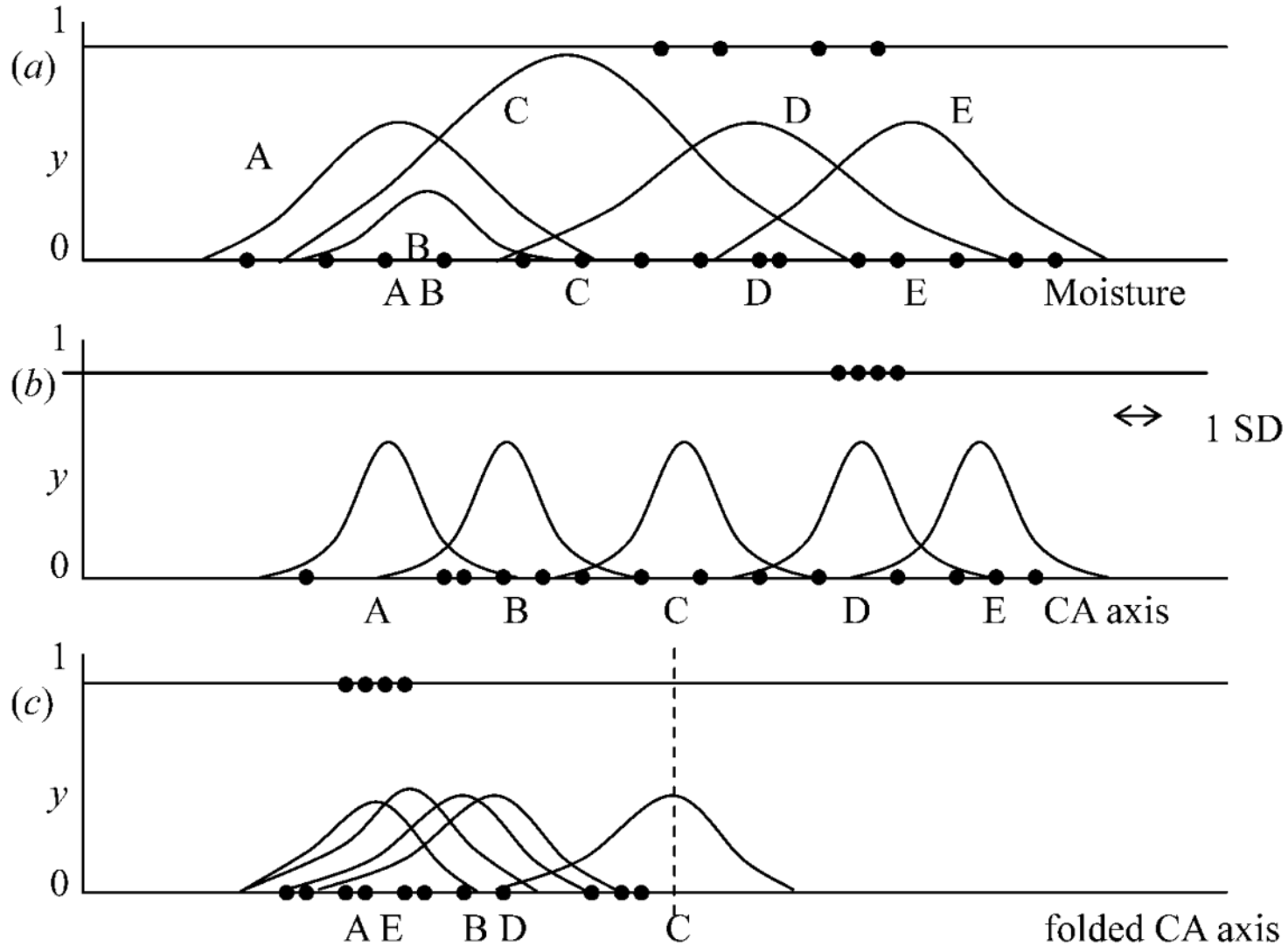
~

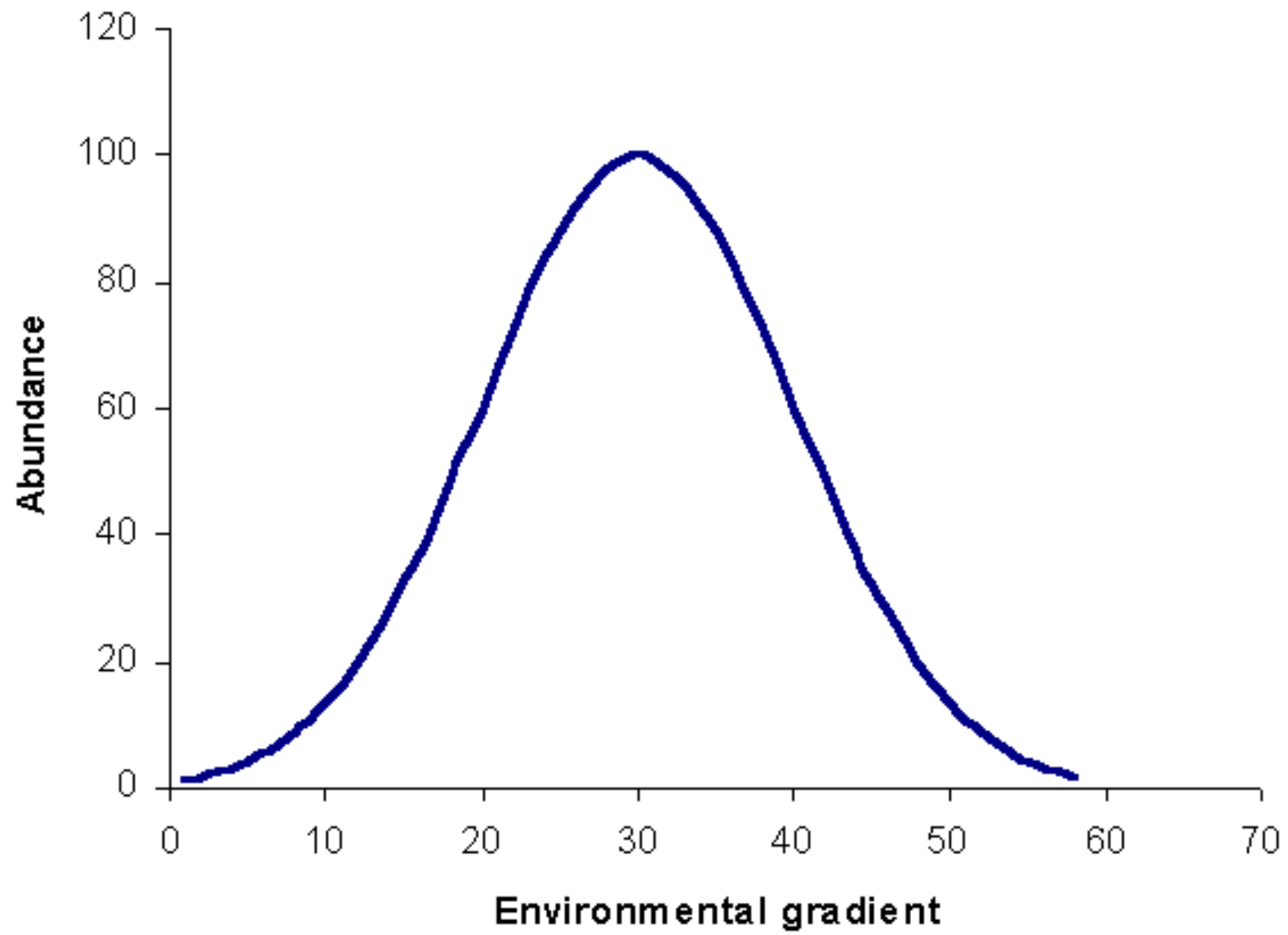


+



CCA





RDA

- Mesmo raciocínio da CCA, mas assume uma relação linear entre o gradiente ambiental e a distribuição de espécies
 - Deve ser usado quando o gradiente é curto (maioria dos casos)

Meu deus
Não entendi
quero
Quem pode

MIGA SUA LOCA



SIACALME

