



Respira fundo
e continua.



Aula 5: Ordenações restritas (canônicas)

Cap. 11 Legendre & Legendre

Ao final desta aula você deverá ter compreendido

- CCA
- RDA
- LDA
- Análises parciais (pRDA, pCCA)
- RDA de consenso (Breve comentário)
- db-RDA

Objetivo

- Usada principalmente em estudos ecológicos para avaliar a relação das espécies com o ambiente.



Chamadas de análises diretas de gradiente, que utiliza gráficos simples para reduzir e revelar a estrutura de dados multivariados.

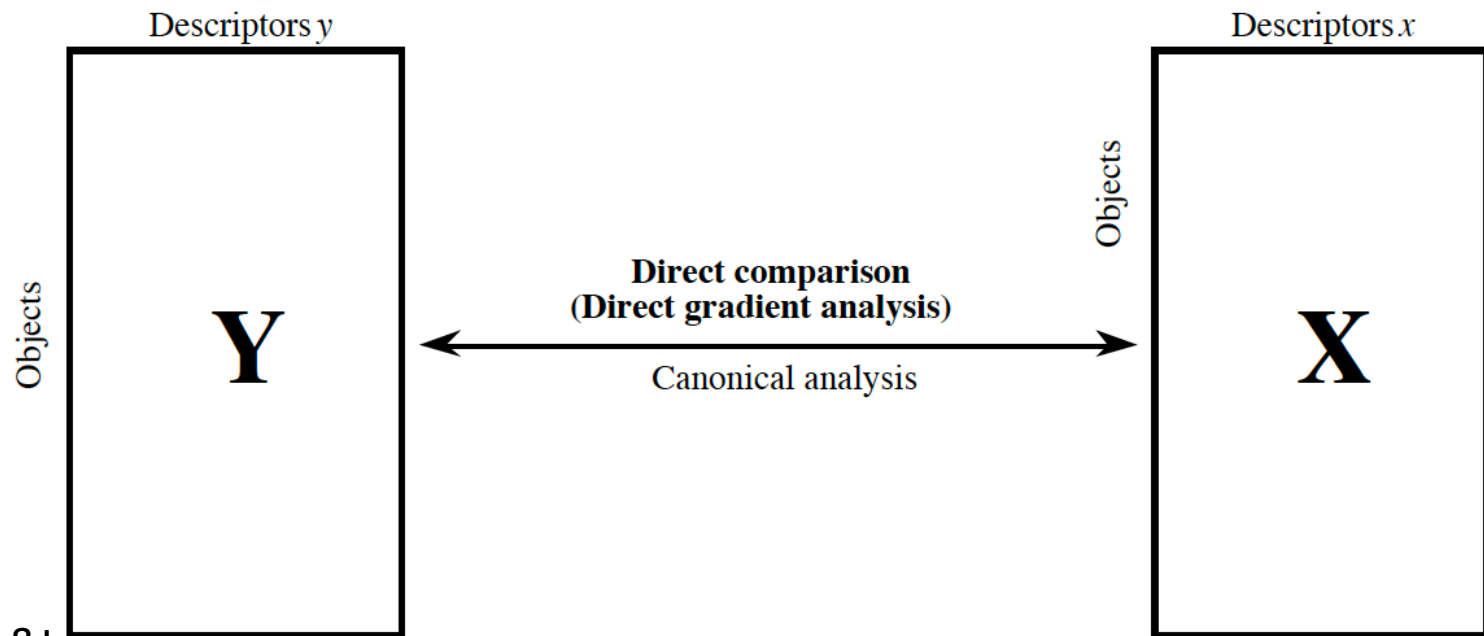
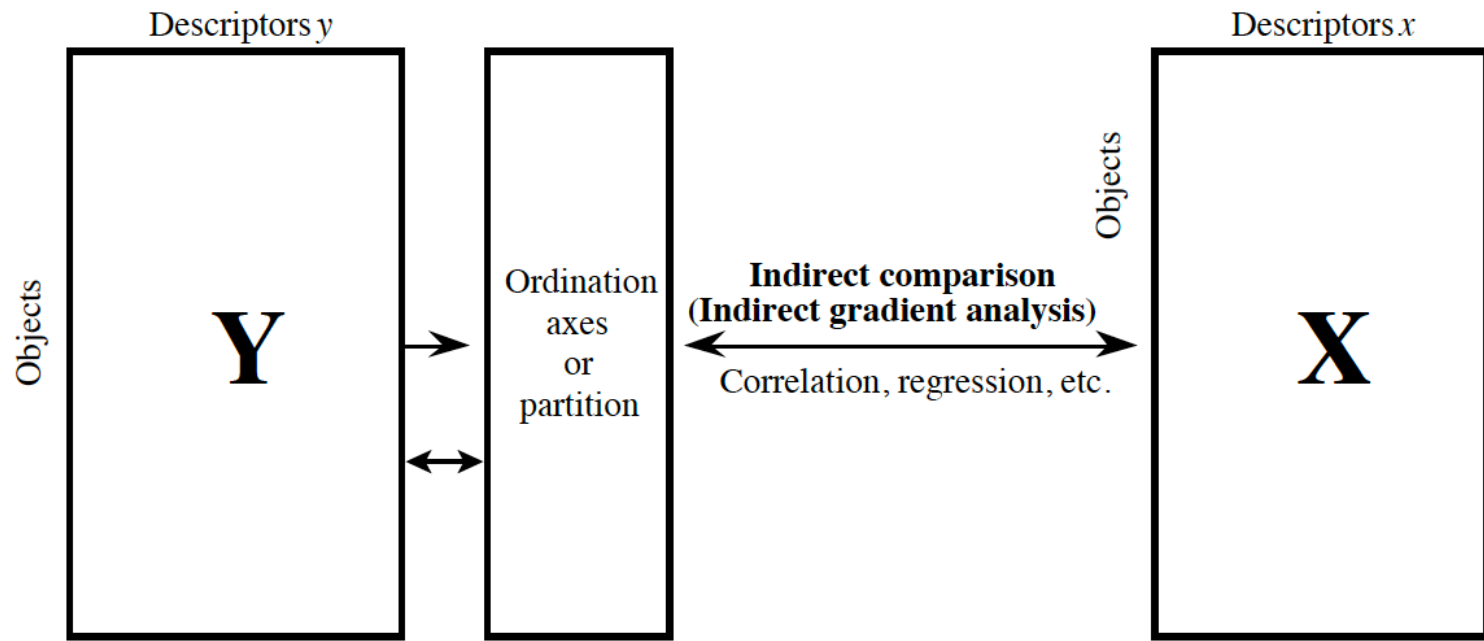
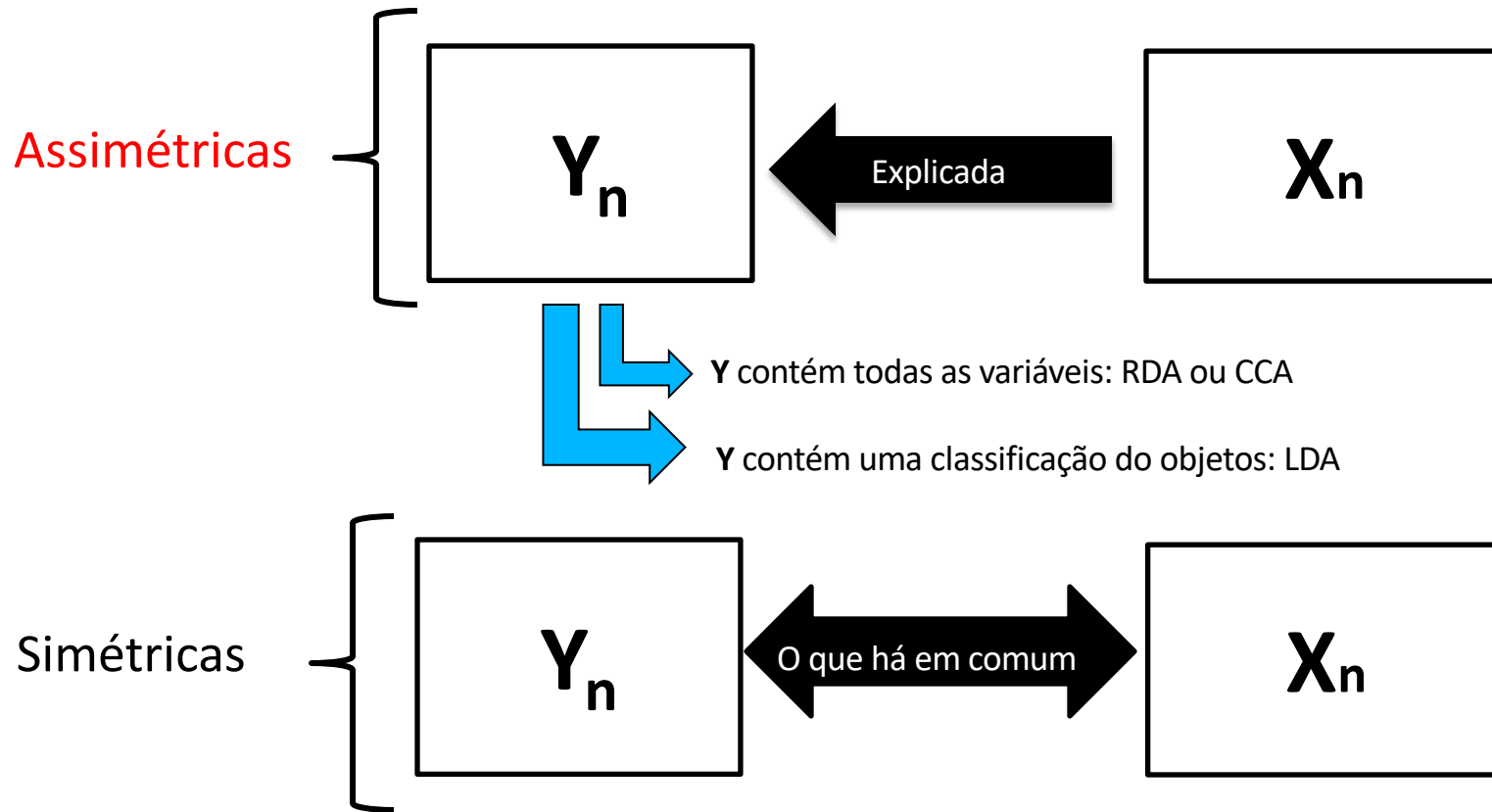


Fig. 10.4 L&L

Quando usar?



Análises canônicas

Assimétricas:

- ✓ Análise de Correspondência Canônica;
- ✓ Análise de Redundância;
- ✓ Análise de Discriminante Linear;

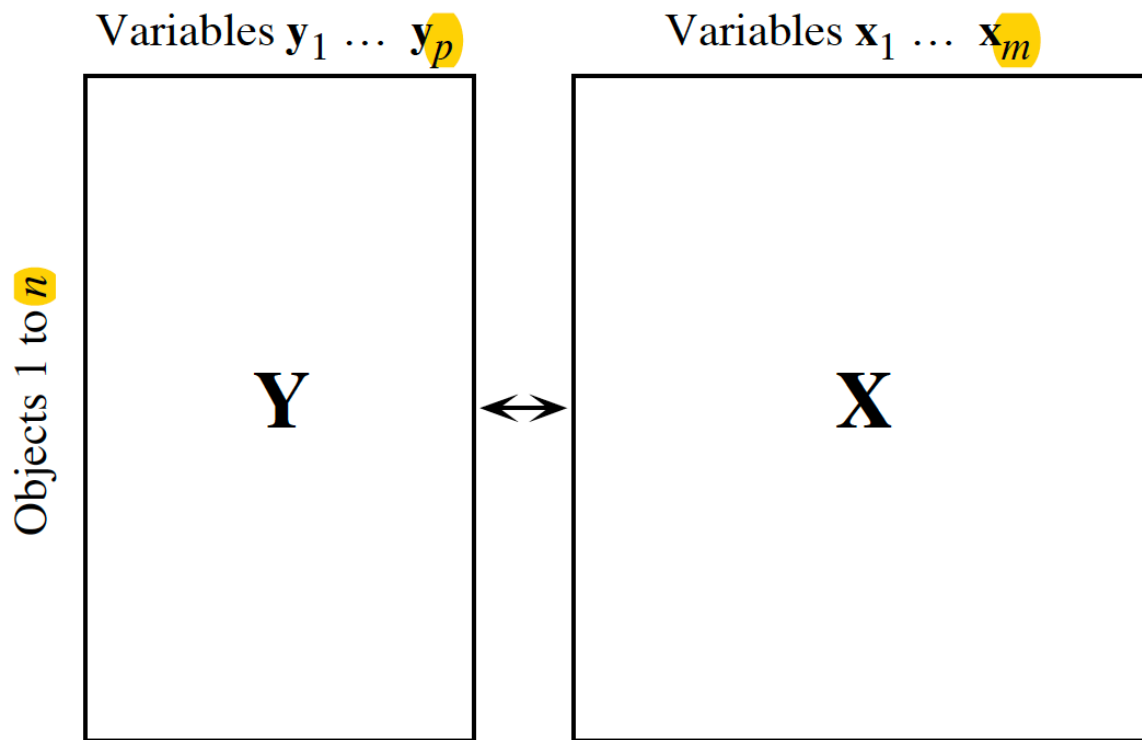
Simétricas:

- ✓ Análise de Correlação Canônica;
- ✓ Análise de Co-Inércia.
- ✓ Procrustes

O que as análises assimétricas fazem?

- Diferentemente da análise de gradiente indireto, utilizam uma matriz de variáveis preditoras (gradientes) para explicar a distribuição dos dados de uma matriz resposta;
- Combinam os conceitos de ordenação e regressão;

(c) Ordination of \mathbf{Y} under constraint of \mathbf{X} :
redundancy analysis (RDA)
canonical correspondence analysis (CCA)



letras (tamanho das matrizes)
entram no cálculo dos g.l.

Qual o propósito dessas análises?

- Comumente usadas para entender a relação entre composição da comunidade e variáveis ambientais;
- Incorporam dados ambientais na análise para produzir um diagrama de ordenação mostrando a relação entre as espécies e variáveis ambientais, reduzindo a dimensionalidade dos dados.
- Na ordenação, os eixos são combinações lineares das variáveis ambientais (gradiente) e as espécies são a resposta a esse gradiente.

Vantagens

- Assume uma **relação causal** entre a variável **dependente e independente**;
- Pode analisar os dados de maneira **exploratória**;
- Permite **testar** hipóteses definidas *a priori*.

O que é necessário?

- 1. Matriz dependente (**Y**) – usualmente composta por dados de composição de espécies (binários ou quantitativos)
- 2. Matriz independente (**X**) – usualmente um conjunto de variáveis ambientais.

Obrigatoriamente as matrizes deverão conter o mesmo numero de linhas, ou seja, de réplicas e a ordem deve ser a mesma entre as duas matrizes **Y** e **X**.

Pressupostos

- Dados dimensionalmente homogêneos, ou seja, as variáveis ambientais **devem** estar padronizadas;
- Preditoras podem ser quantitativas ou qualitativas (multiestado ou binárias/dummy)
- Ausência de multicolinearidade;
- **A matriz com as variáveis independentes (preditoras) não deve conter mais do que 1/3 da quantidade de variáveis da matriz de dados dependentes.**

Redundancy Analysis - RDA

Assume uma **relação linear** entre a variável **dependente** e **independente**

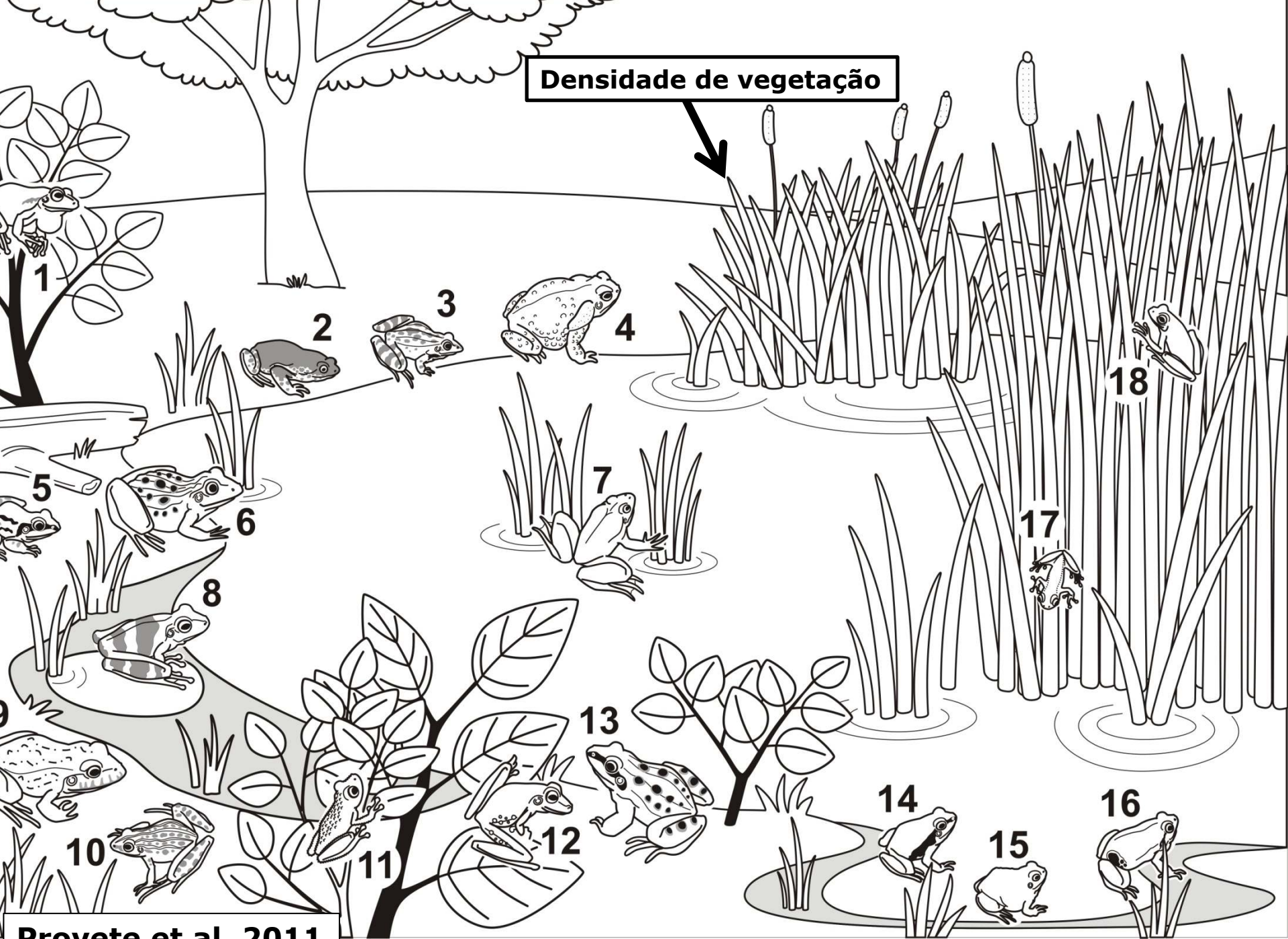
Redundância = variância explicada

Canonical Correlation Analysis - CCA

Assume uma **relação unimodal** entre a variável resposta (e.g., abundância de espécies) e variável dependente (variáveis ambientais)

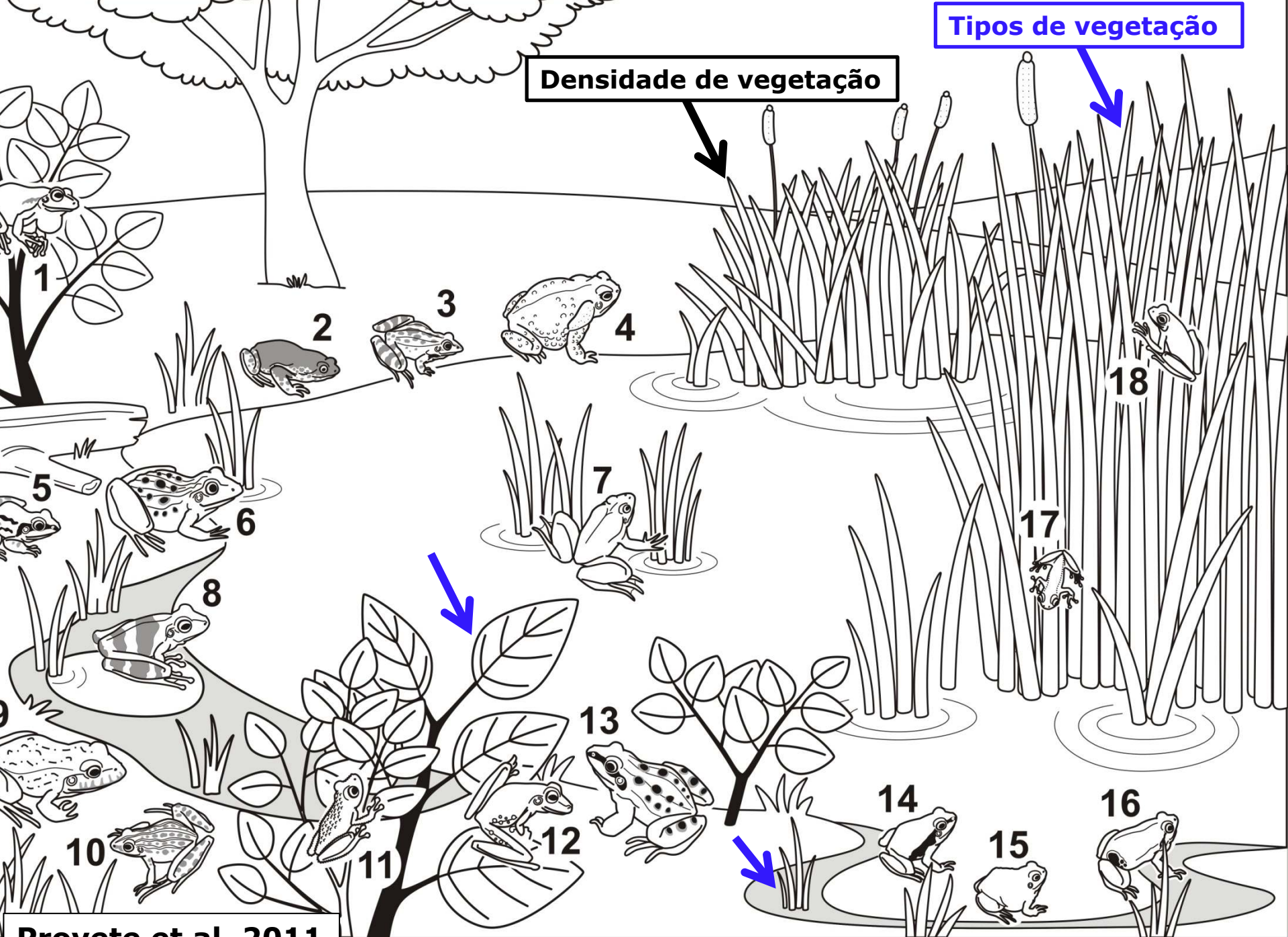


Densidade de vegetação



Tipos de vegetação

Densidade de vegetação

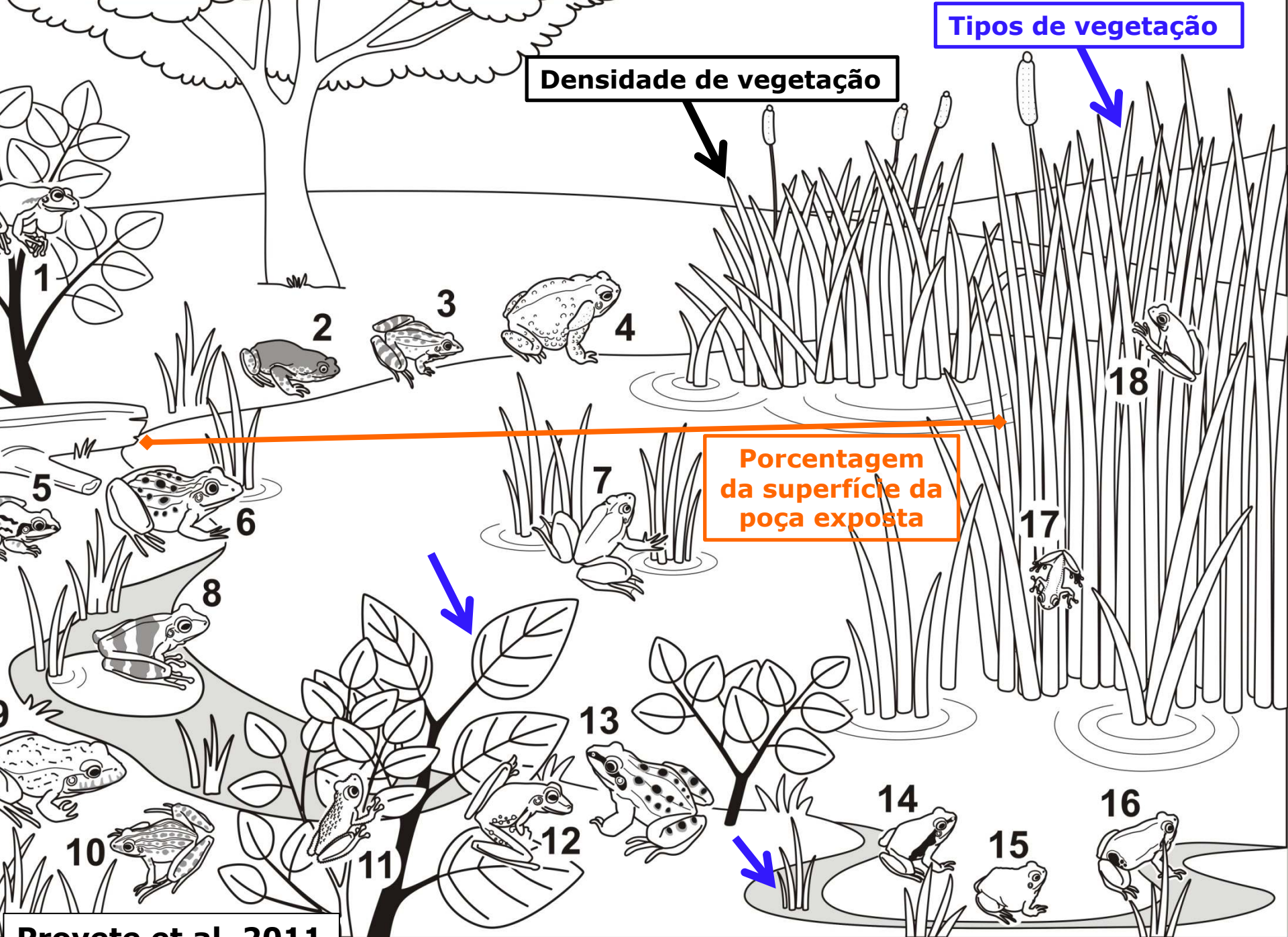


Provete et al. 2011

Tipos de vegetação

Densidade de vegetação

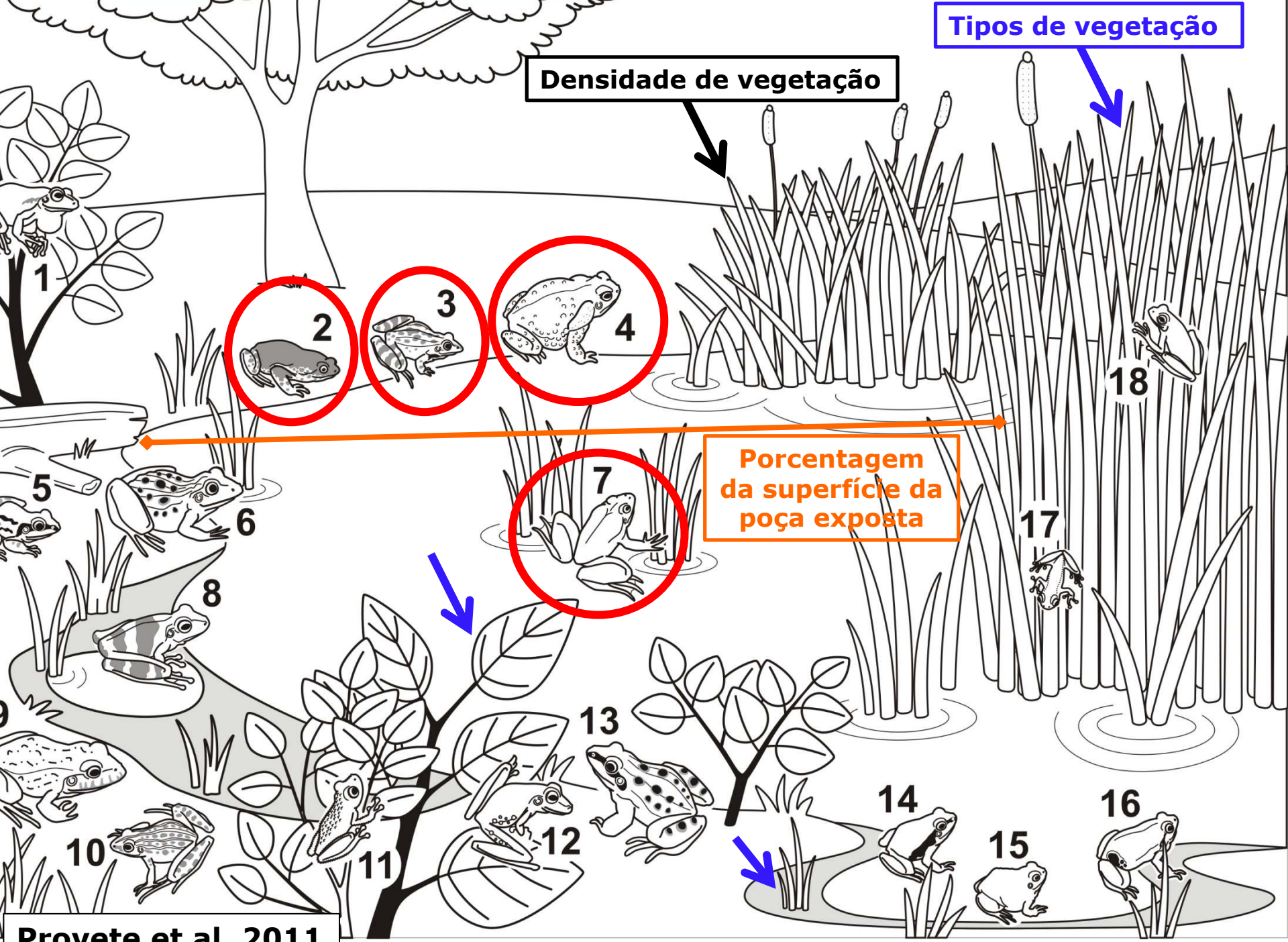
Porcentagem da superfície da poça exposta



Tipos de vegetação

Densidade de vegetação

Porcentagem da superfície da poça exposta



Tipos de vegetação

Densidade de vegetação

Porcentagem da superfície da poça exposta

X

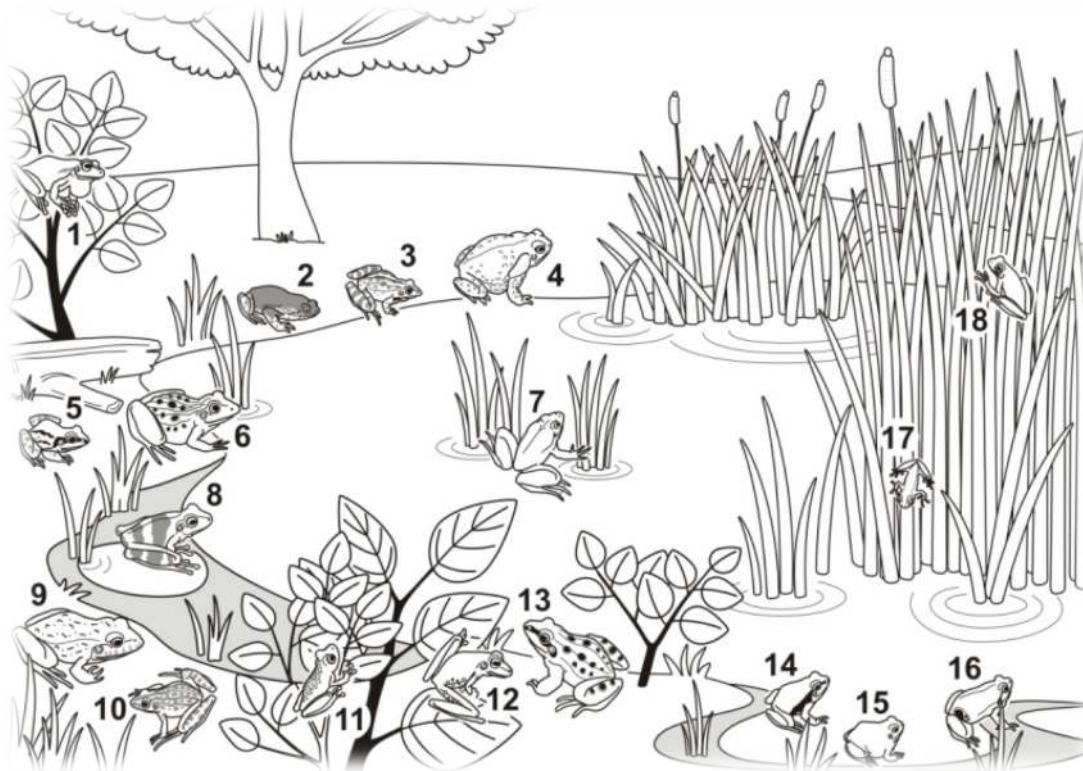


Y

Resposta de comunidades biológicas

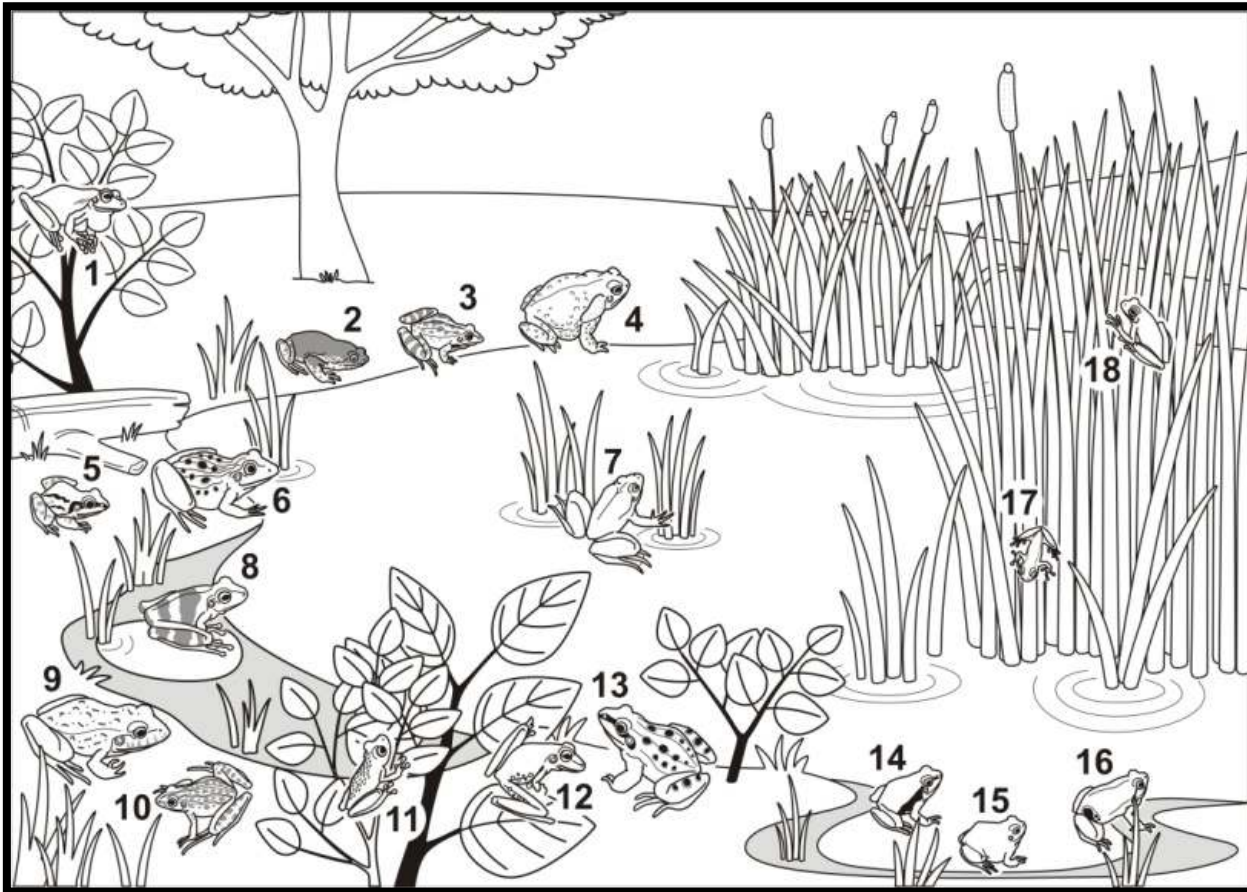
Principal questão

A variação ambiental influencia a composição de espécies?

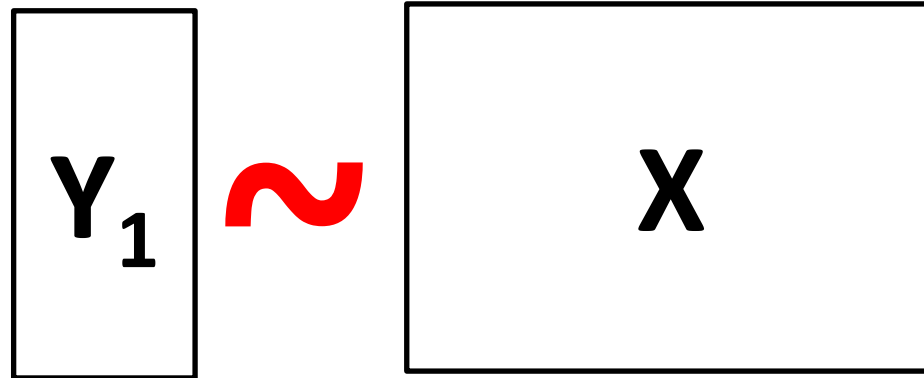


Redundancy Analysis - RDA

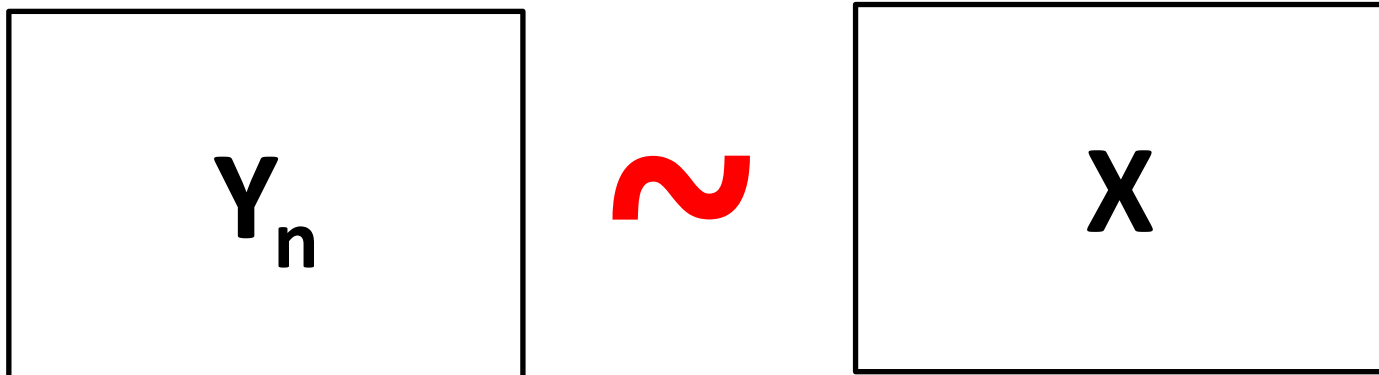
Utilizada para testar a relação entre a **composição de espécies** e as **características ambientais**

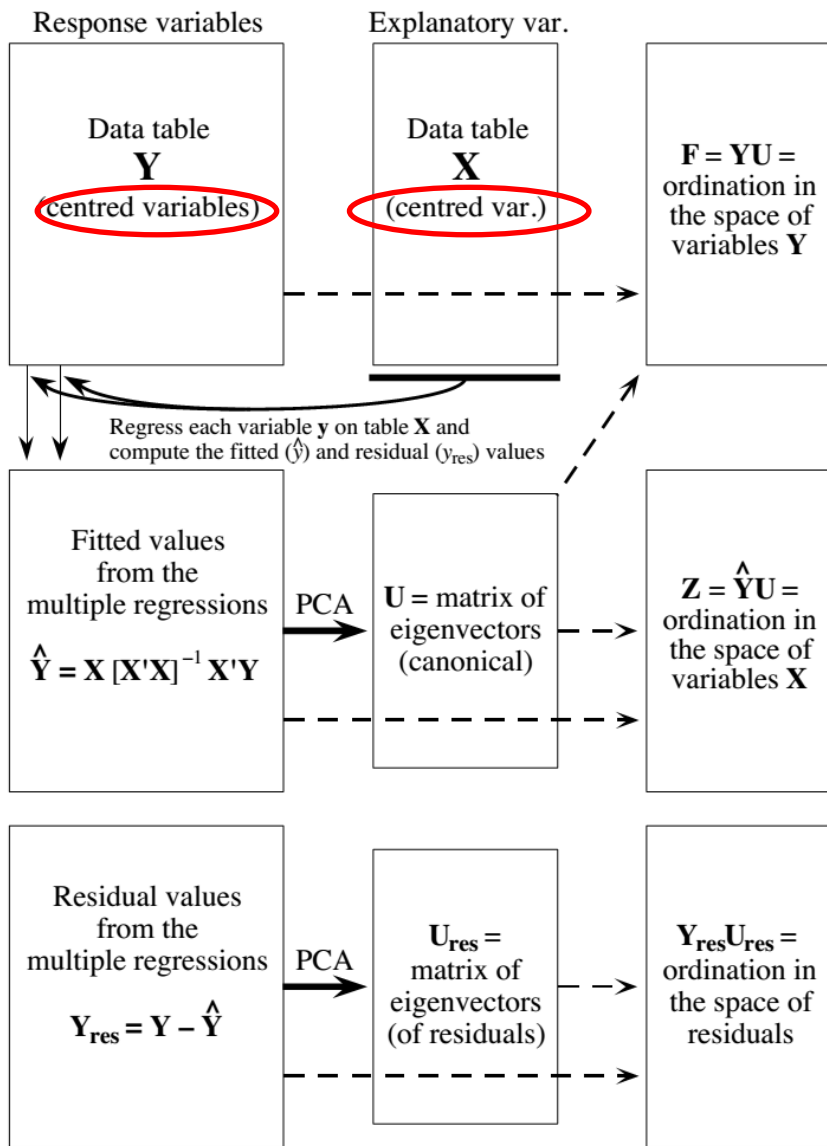


Regressão múltipla – variável dependente UNIVARIADA (**vetor**)

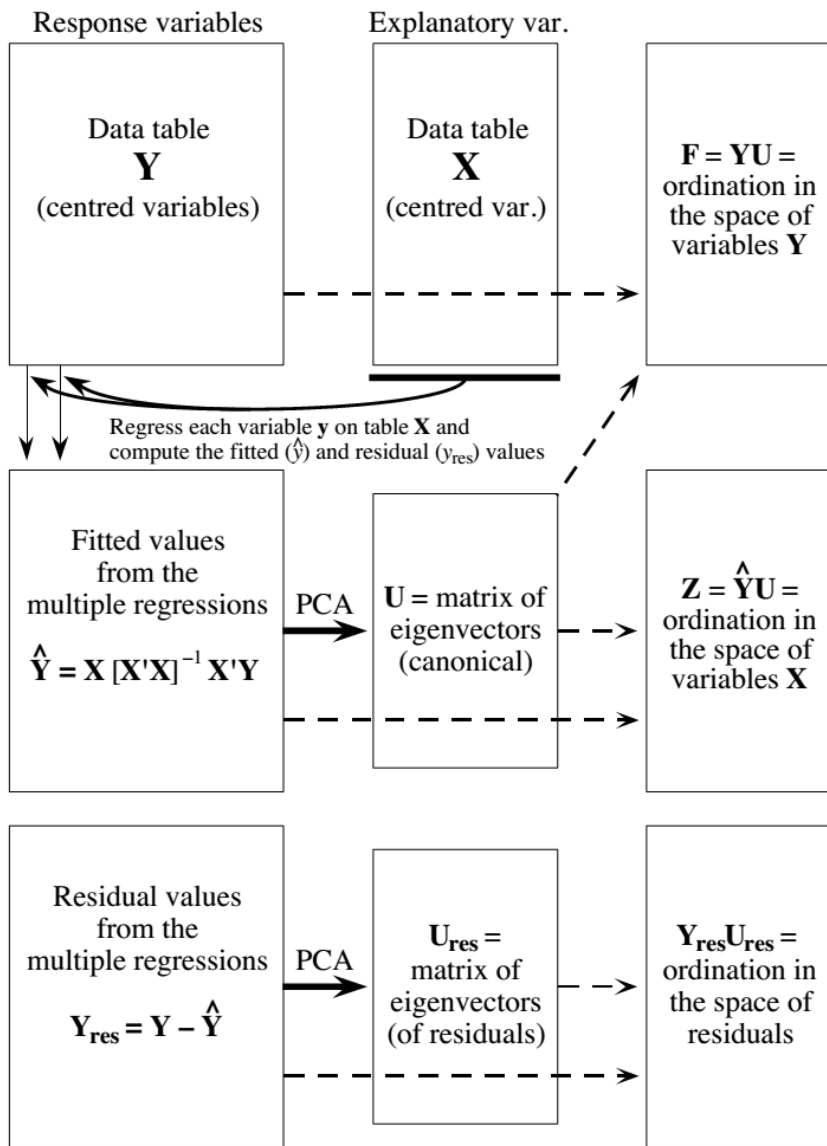


RDA – variável dependente MULTIVARIADA (**matriz**)





!!Lembre-se de padronizar as variáveis predictoras e eliminar multicolinearidade!!



Simplex assim?



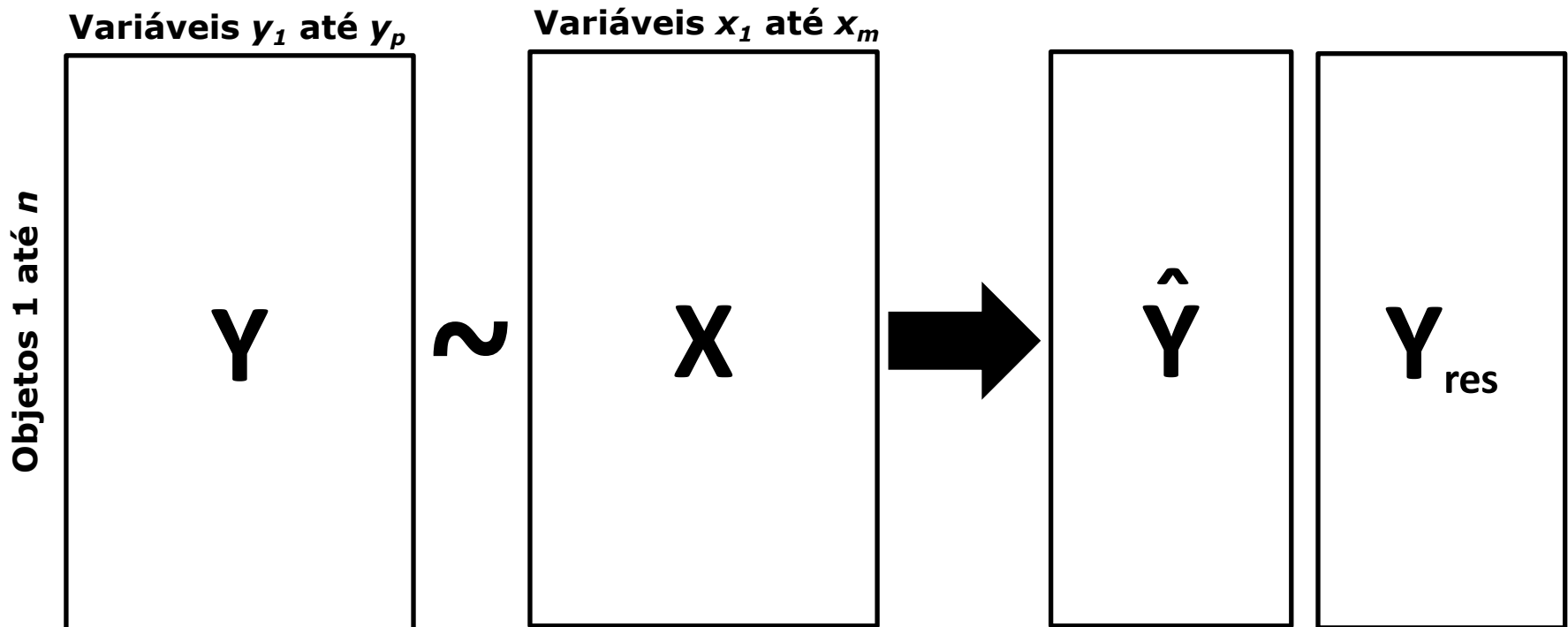


num intendi nada que
você disse ai fera

Tudo bem então, vejamos a RDA passo
a passo...

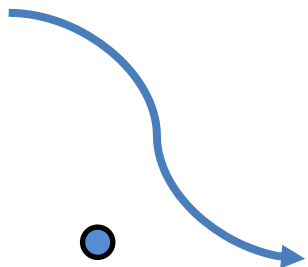
RDA

Passo 1: Regressão múltipla de cada variável de **Y** contra a matriz **X** (descritores).



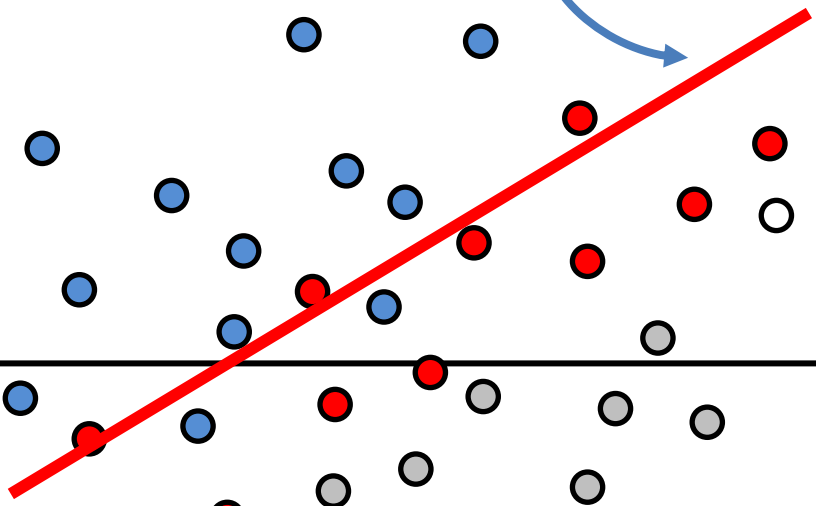
Y_1

Imagine que essa linha é um plano hiperdimensional (não me pergunte como se faz isso)



X_2

X_1



\hat{Y}

Escrevendo a equação de um modelo linear (OLS) na forma de matriz

The General Linear Model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}$$

Component	Dimension	Description
\mathbf{Y}	$n \times p$	Data matrix with n observations for p variables
\mathbf{X}	$n \times k$	Linear model design matrix with n observations for k parameters
$\boldsymbol{\beta}$	$k \times p$	Matrix of coefficients expressing change in values for the k model parameters for each of p variables
\mathbf{E}	$n \times p$	Matrix of residuals (error) for n observations for p variables

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}$$

$$\mathbf{X}^T \mathbf{Y} = \mathbf{X}^T \mathbf{X} \boldsymbol{\beta}$$

$$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} \boldsymbol{\beta}$$

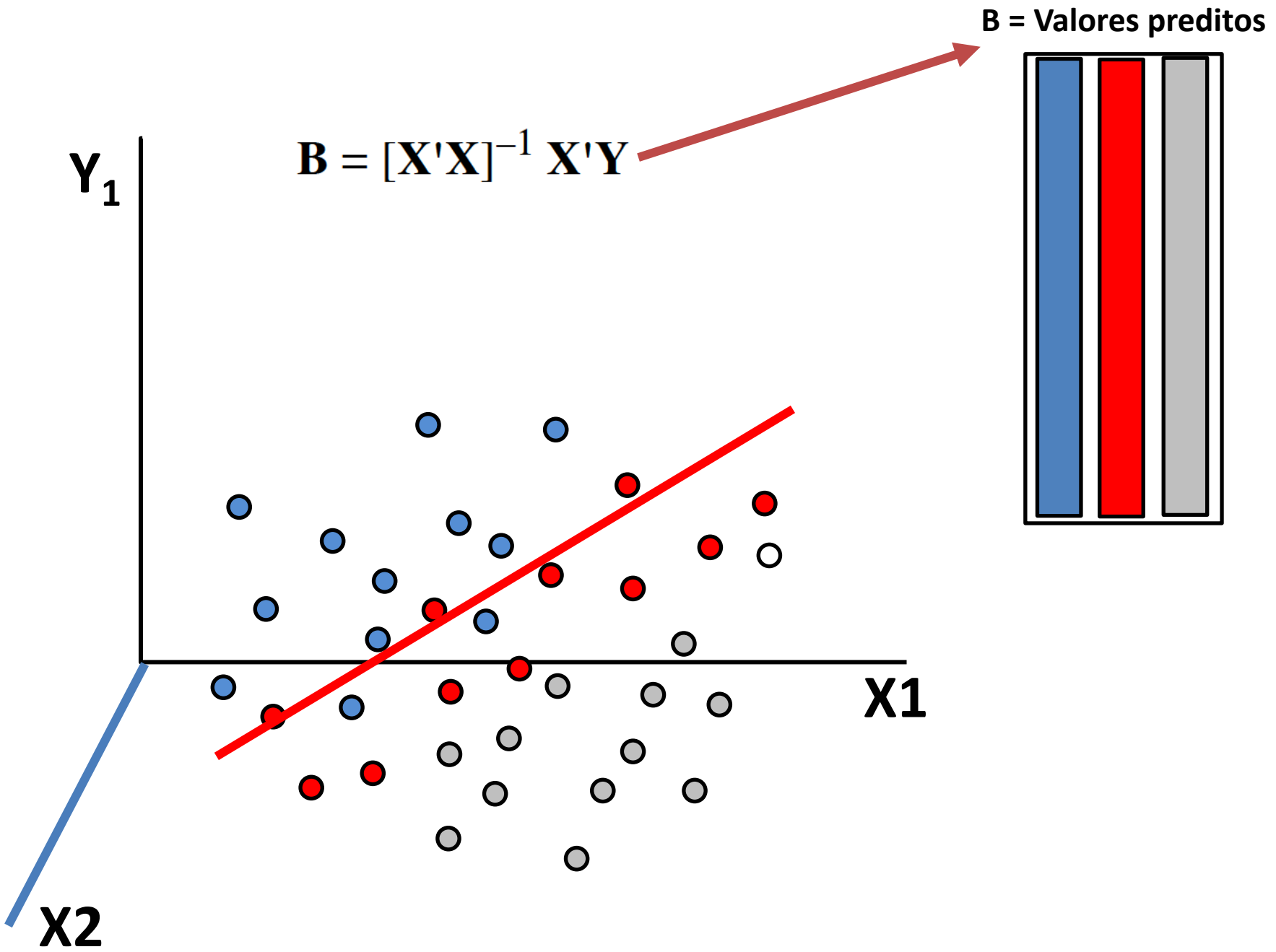
$$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} = \mathbf{I} \boldsymbol{\beta}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{Y})$$

Regressão múltipla (OLS) em cada variável individual de \mathbf{Y}

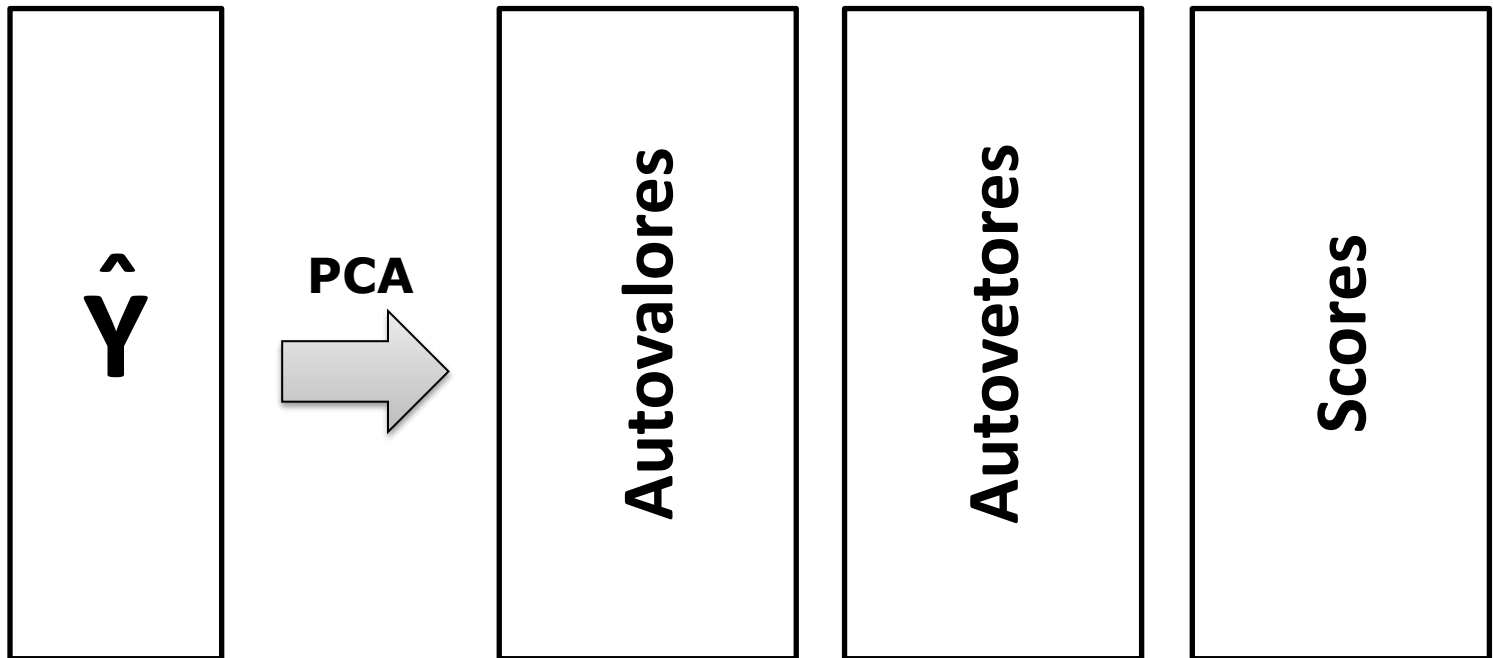
$$\hat{\mathbf{Y}} = \mathbf{X} [\mathbf{X}'\mathbf{X}]^{-1} \mathbf{X}'\mathbf{Y}$$

Como as variáveis nas matrizes \mathbf{X} e \mathbf{Y} são centralizadas pela média, não existe intercepto na equação



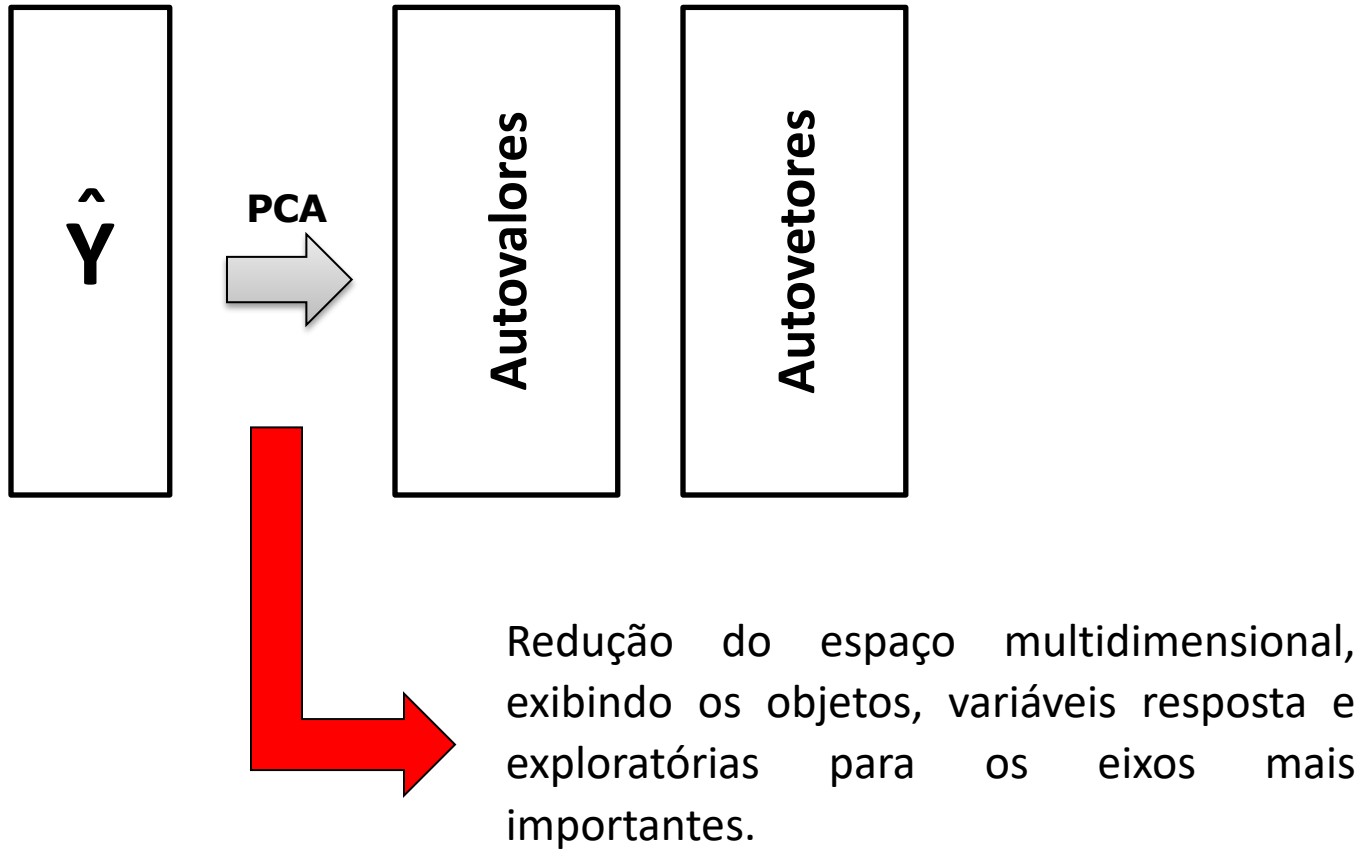
RDA

Passo 2: PCA dos valores preditos de \mathbf{Y} : $\hat{\mathbf{Y}}$.

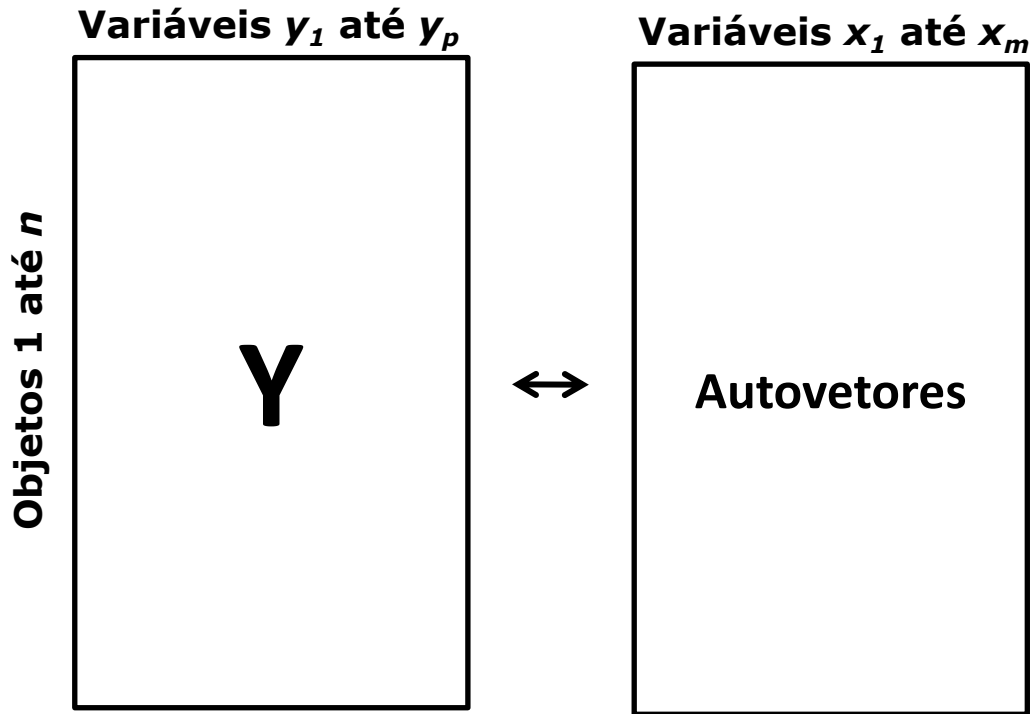


RDA e CCA

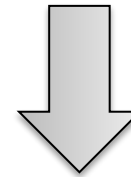
Passo 2: Ordenação dos valores preditos de \mathbf{Y} .



RDA

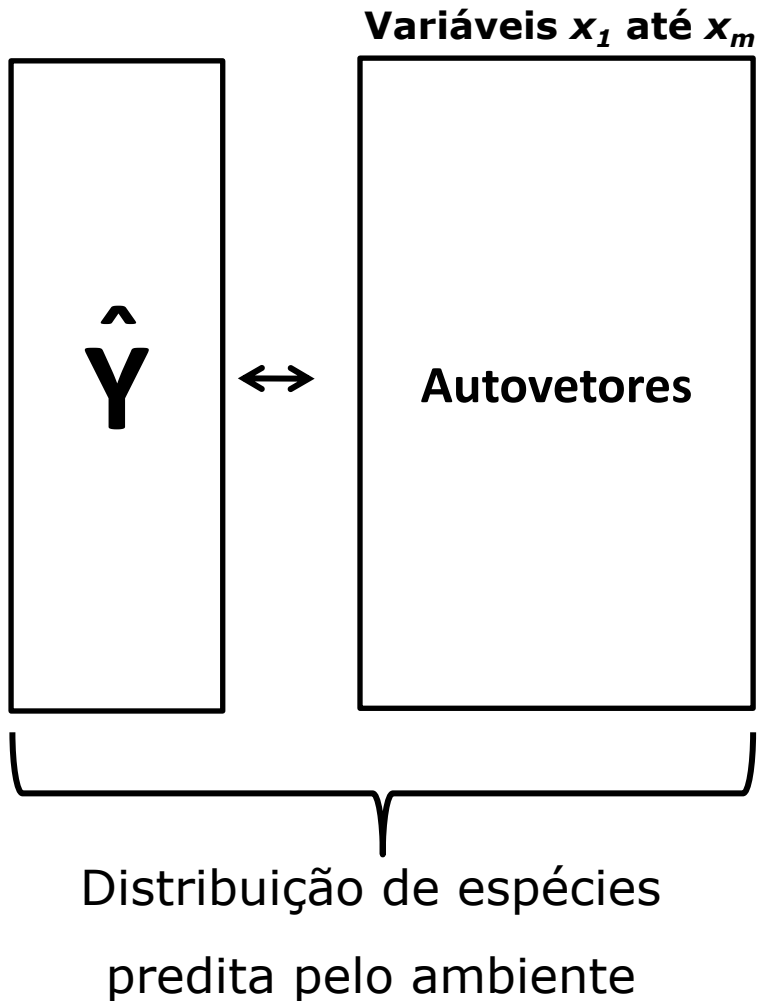


Passo 2: Combinação linear de **Y** restrita pelos autovetores da matriz resposta.

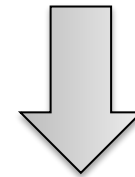


No vegan chamado de "**site scores**"

RDA



Passo 2: Combinação linear dos valores preditos de \mathbf{Y} com os autovetores

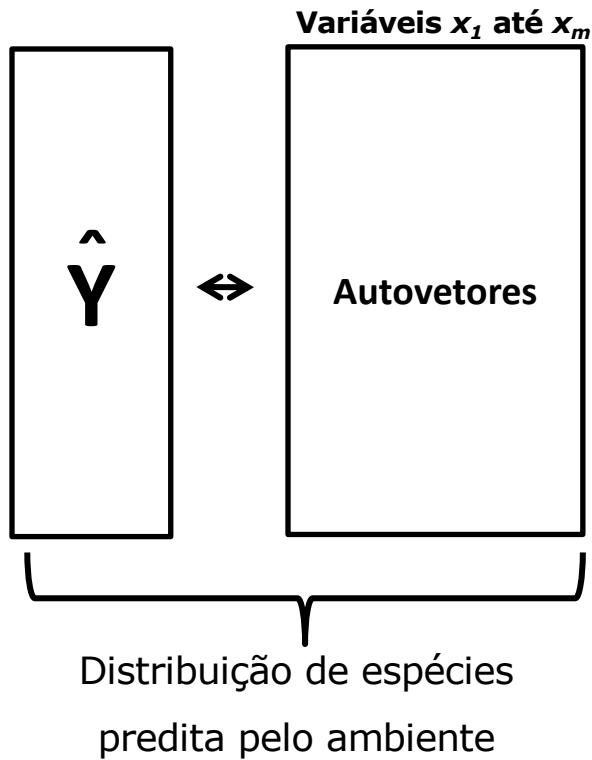


Fitted site constraints



No vegan chamado de "**site constraints**"

RDA e CCA



RDA: o resultado são combinações lineares entre **Y** e **X**.

CCA: o resultado são relações ponderadas entre **Y** e **X**.

Diferença entre RDA e CCA

- CCA é a forma ponderada da RDA
 - Regressão múltipla ponderada
 - Combinações lineares são ponderadas pelas linhas da matriz preditora \mathbf{X}

$$\mathbf{B} = [\mathbf{X}_{stand}' \mathbf{D}(p_{i+}) \mathbf{X}_{stand}]^{-1} \mathbf{X}_{stand}' \mathbf{D}(p_{i+})^{1/2} \bar{\mathbf{Q}}$$

$$\hat{\mathbf{Y}} = \mathbf{D}(p_{i+})^{1/2} \mathbf{X}_{stand} [\mathbf{X}_{stand}' \mathbf{D}(p_{i+}) \mathbf{X}_{stand}]^{-1} \mathbf{X}_{stand}' \mathbf{D}(p_{i+})^{1/2} \bar{\mathbf{Q}}$$

$$\bar{\mathbf{Q}}_{res} = \bar{\mathbf{Q}} - \hat{\mathbf{Y}} .$$

- Preserva a distância de χ^2 entre linhas da matriz dependente

Como avaliar se o modelo é adequado?

A PCA do Passo 2 só é feita se for encontrada uma relação significativa entre a matriz \mathbf{X} e \mathbf{Y} . Mas como medir a proporção de variação explicada?

$$R^2 \text{ canônico} \quad R_{\hat{\mathbf{Y}}|\mathbf{X}}^2 = \frac{SS(\hat{\mathbf{Y}})}{SS(\mathbf{Y})}$$

$$R^2 \text{ canônico ajustado} \quad R_a^2 = 1 - (1 - R_{\hat{\mathbf{Y}}|\mathbf{X}}^2) \frac{(n-1)}{(n-m-1)}$$

R^2 ajustado canônico

- Valor mínimo não é zero, depende do número de variáveis preditoras e amostras

$$E(R^2) = m/(n - 1)$$

- O R^2 ajustado, junto com o teste de significância do eixo, pode também ser usado para escolher entre CCA vs. RDA
 - Gradiente curto vs. longo
 - Prefira o modelo que tenha maior explicação

Teste F

- É possível também calcular uma estatística F, e seu valor de P associado

$$F = \frac{R_{\mathbf{Y}_{stand}|\mathbf{X}}^2 / mp}{(1 - R_{\mathbf{Y}_{stand}|\mathbf{X}}^2) / (n - m - 1) p}$$

- Testa significância da relação entre matrizes no modelo RDA/CCA
- Usado para determinar a significância do modelo

Testar significância dos eixos

- Testa se a quantidade de variação capturada no eixo é maior do que o esperado ao acaso
- Quantos eixos devem ser interpretados
 - `vegan::anova.cca(..., by="axis")`
 - `vegan::anova.cca(..., by="margin")`
 - Fornece o valor de P para cada variável explanatória

Como interpretar um Triplot?

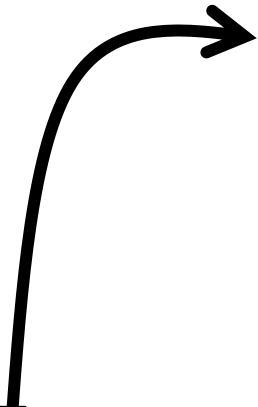
Como interpretar um Triplot?

- 3 informações: locais, variáveis preditoras (e.g., variáveis ambientais) e respostas (e.g., espécies)
- Distâncias no diagrama aproximam distância Euclidiana na RDA e χ^2 na CCA
- Dois tipos de scaling
 1. Foca nos objetos
 1. Útil quando se tem muita variáveis categóricas
 2. Foca nas variáveis
 1. Útil para variáveis quantitativas

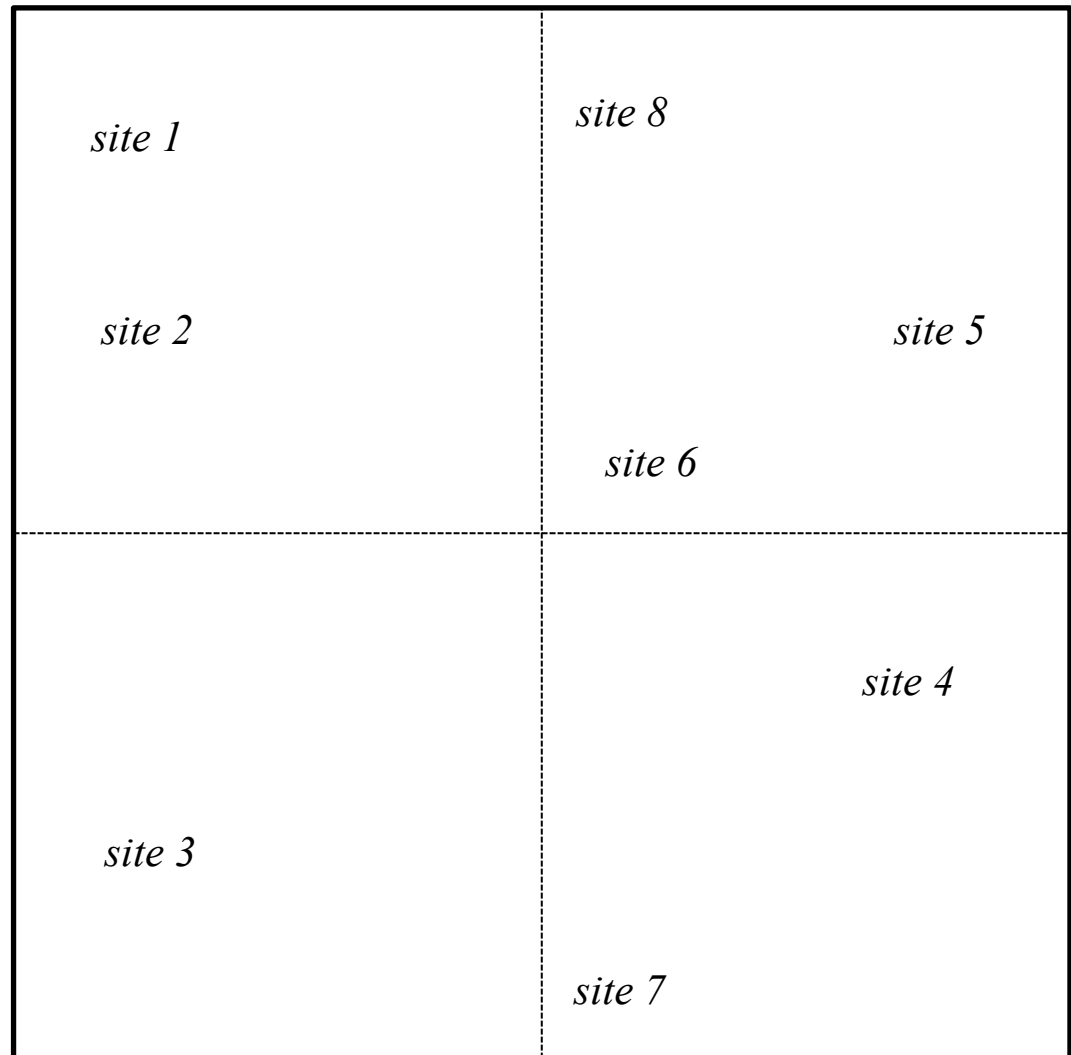
O segundo passo gera:

- *score dos locais*

Scores



RDA 2

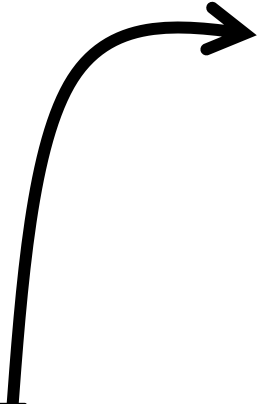


RDA 1

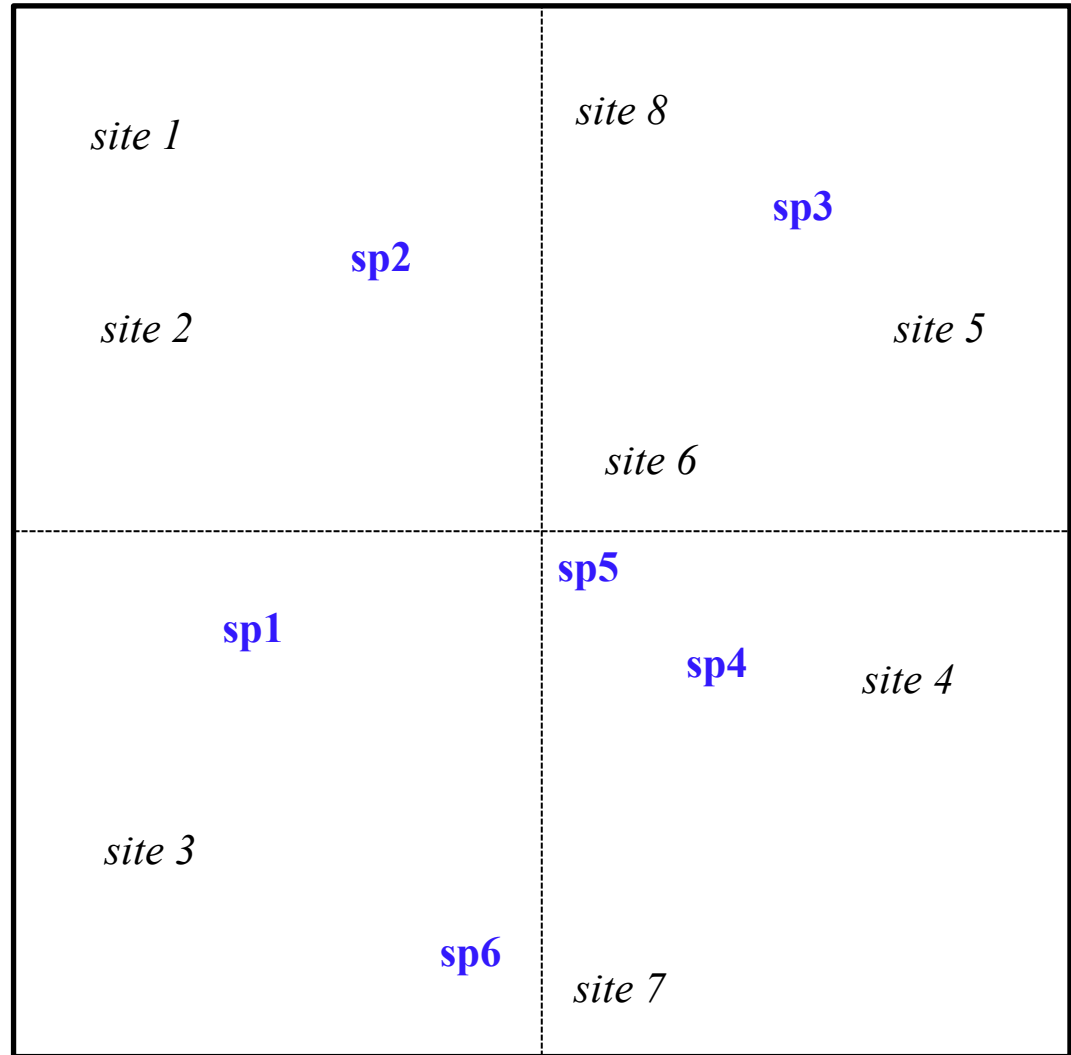
O segundo passo gera:

- *score dos locais*
- *score das espécies*

Scores



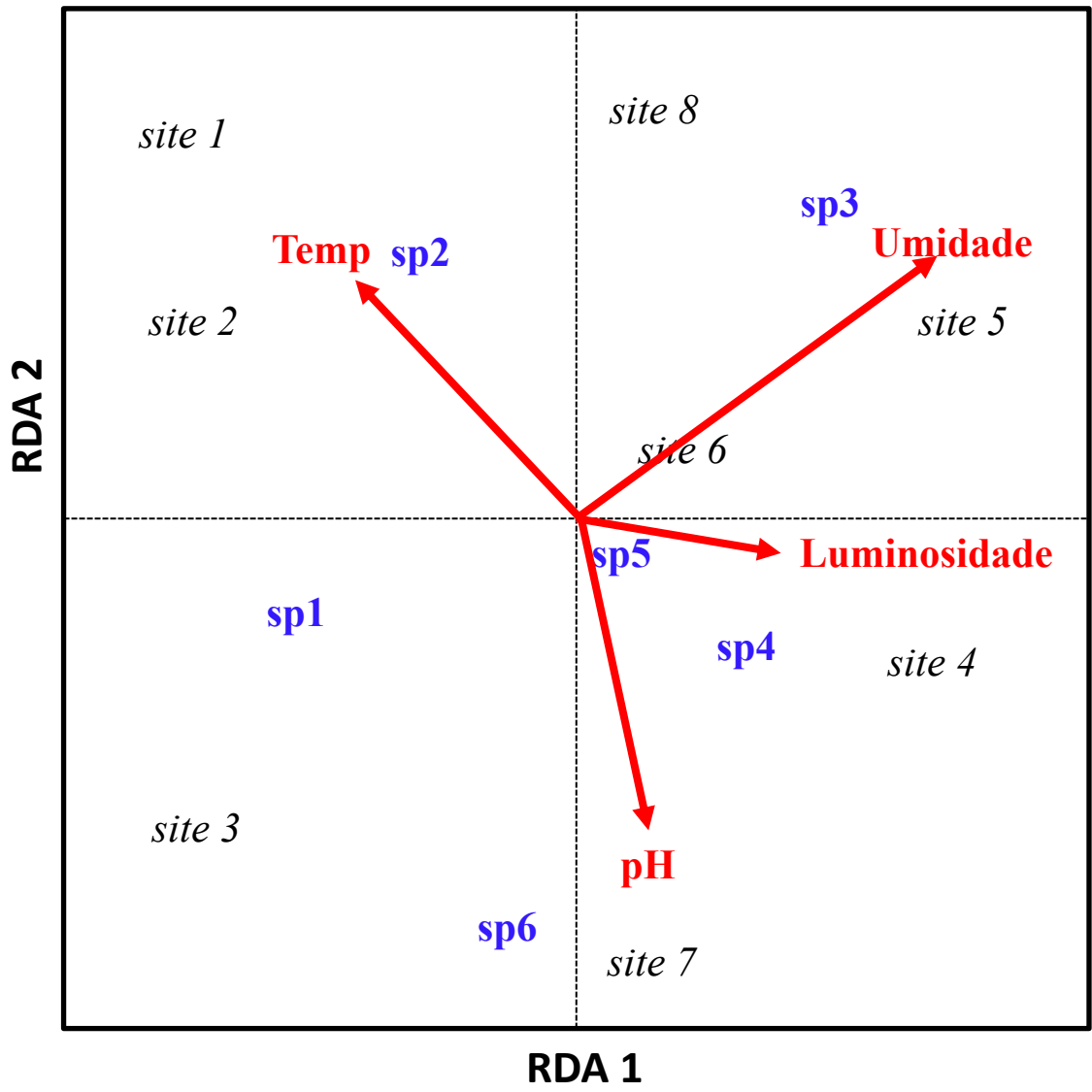
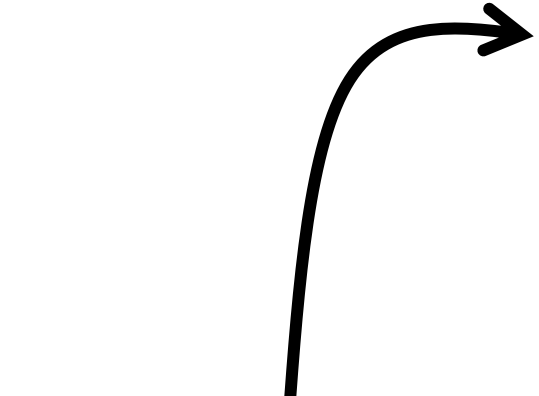
RDA 2



RDA 1

O segundo passo gera:

- *score dos locais*
- *score das espécies*
- *score das abióticas*

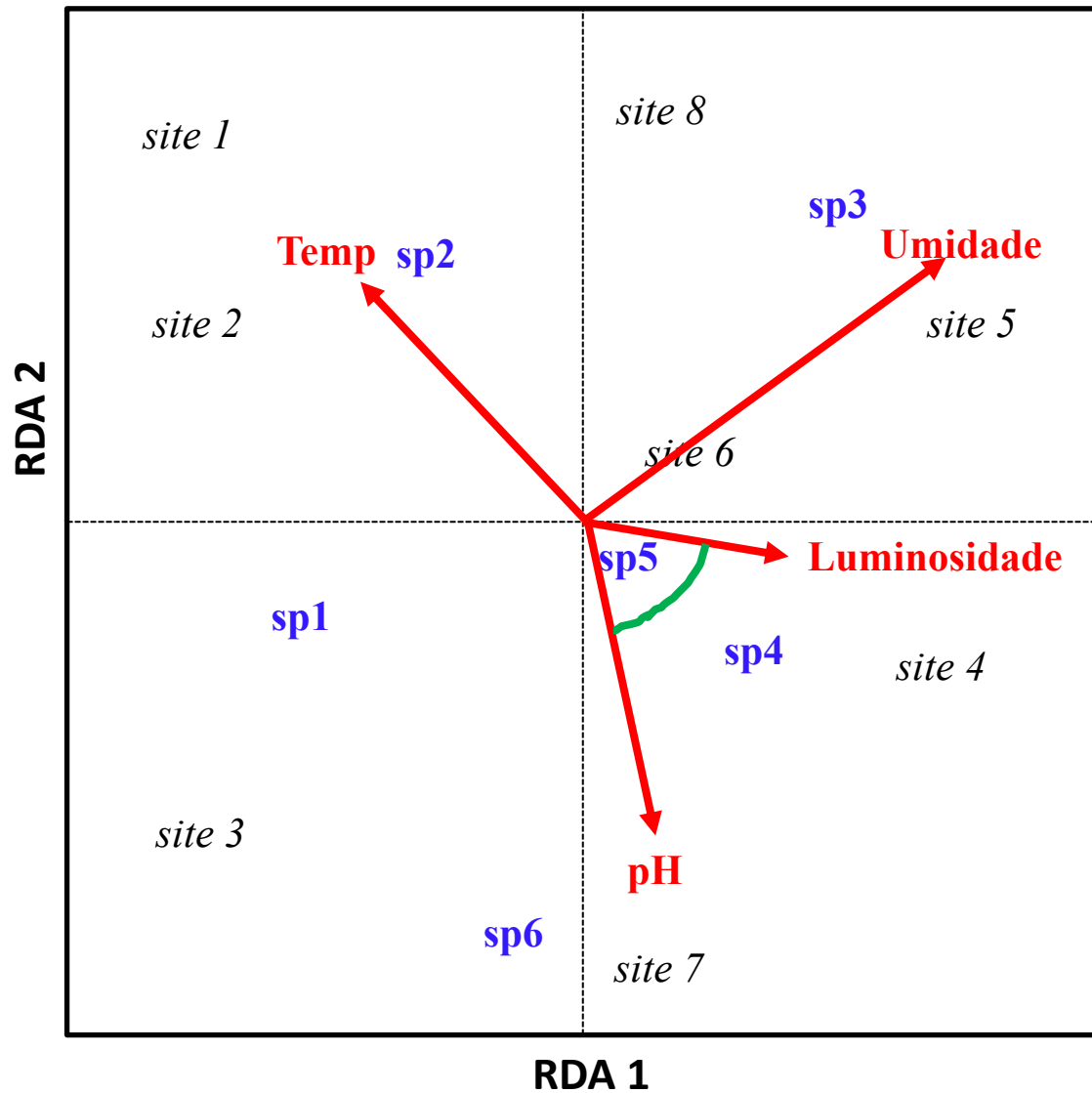


O segundo passo gera:

- *score dos locais*
- *score das espécies*
- *score das abióticas*

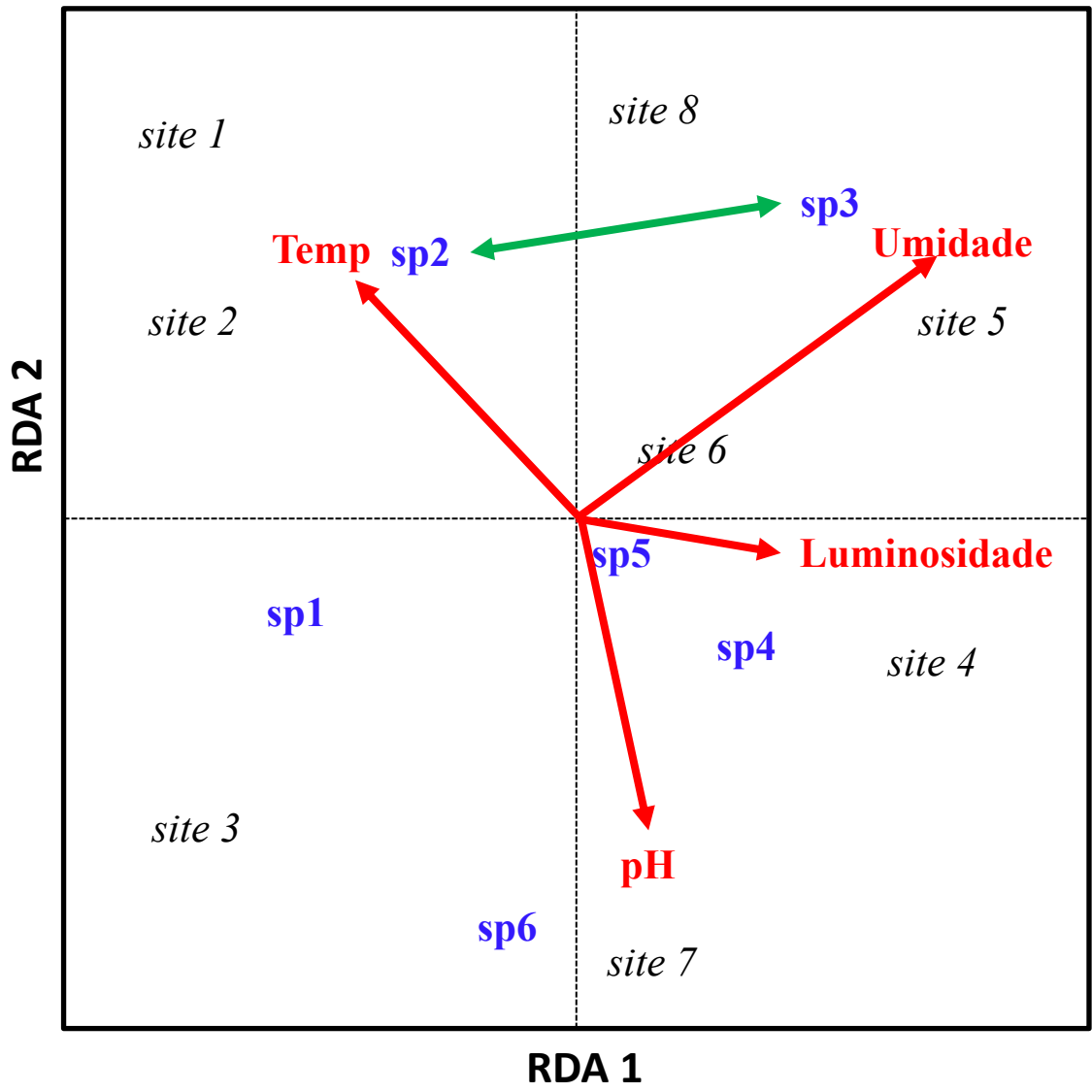
Scaling 2:

Ângulo entre
variáveis reflete
sua correlação



O segundo passo gera:

- *score dos locais*
- *score das espécies*
- *score das abióticas*



Scaling 1:

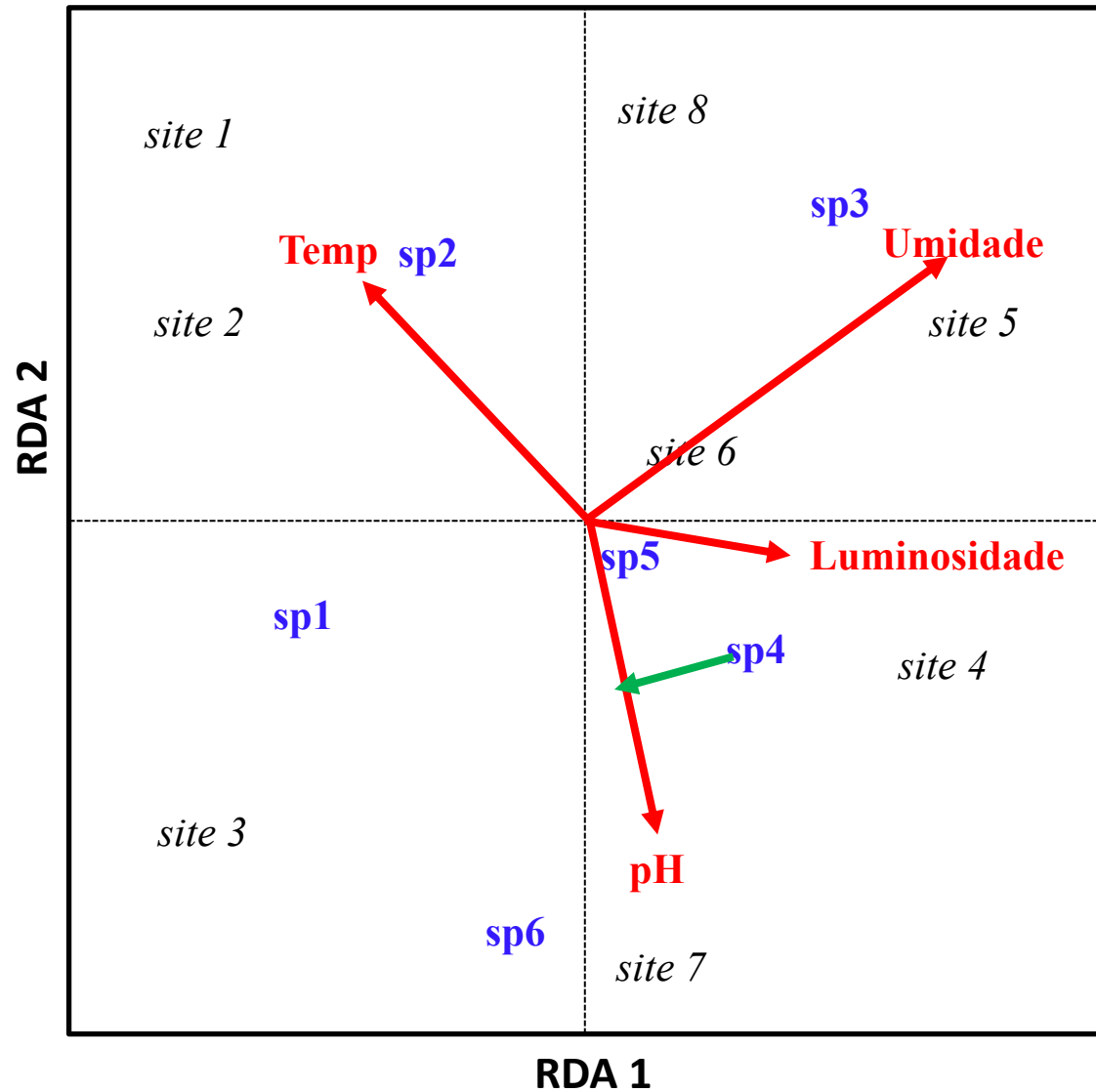
Distância entre objetos são sua distância Euclidiana

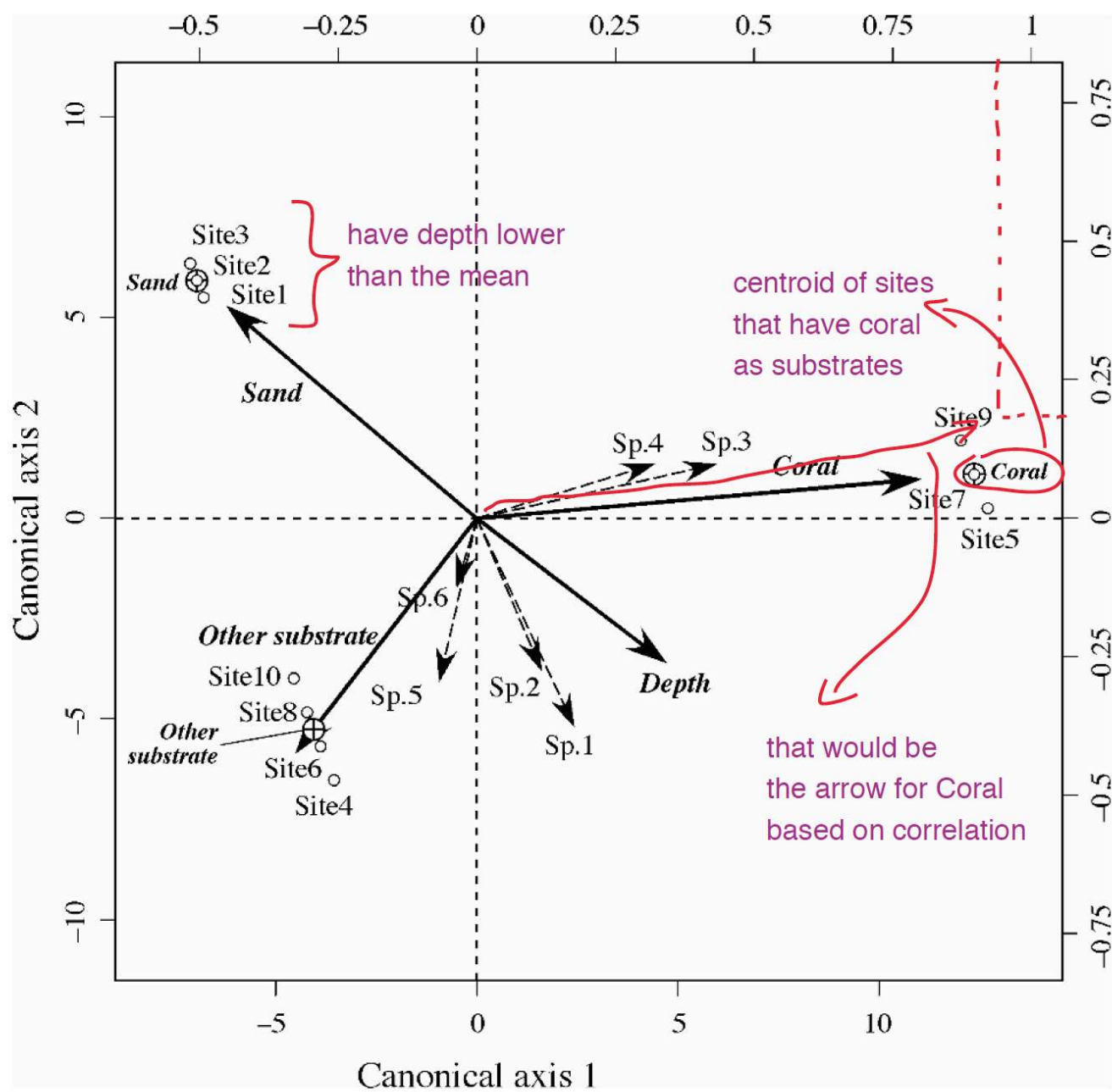
O segundo passo gera:

- *score dos locais*
- *score das espécies*
- *score das abióticas*

Scaling 1:

Abundância da espécie ao longo da variável





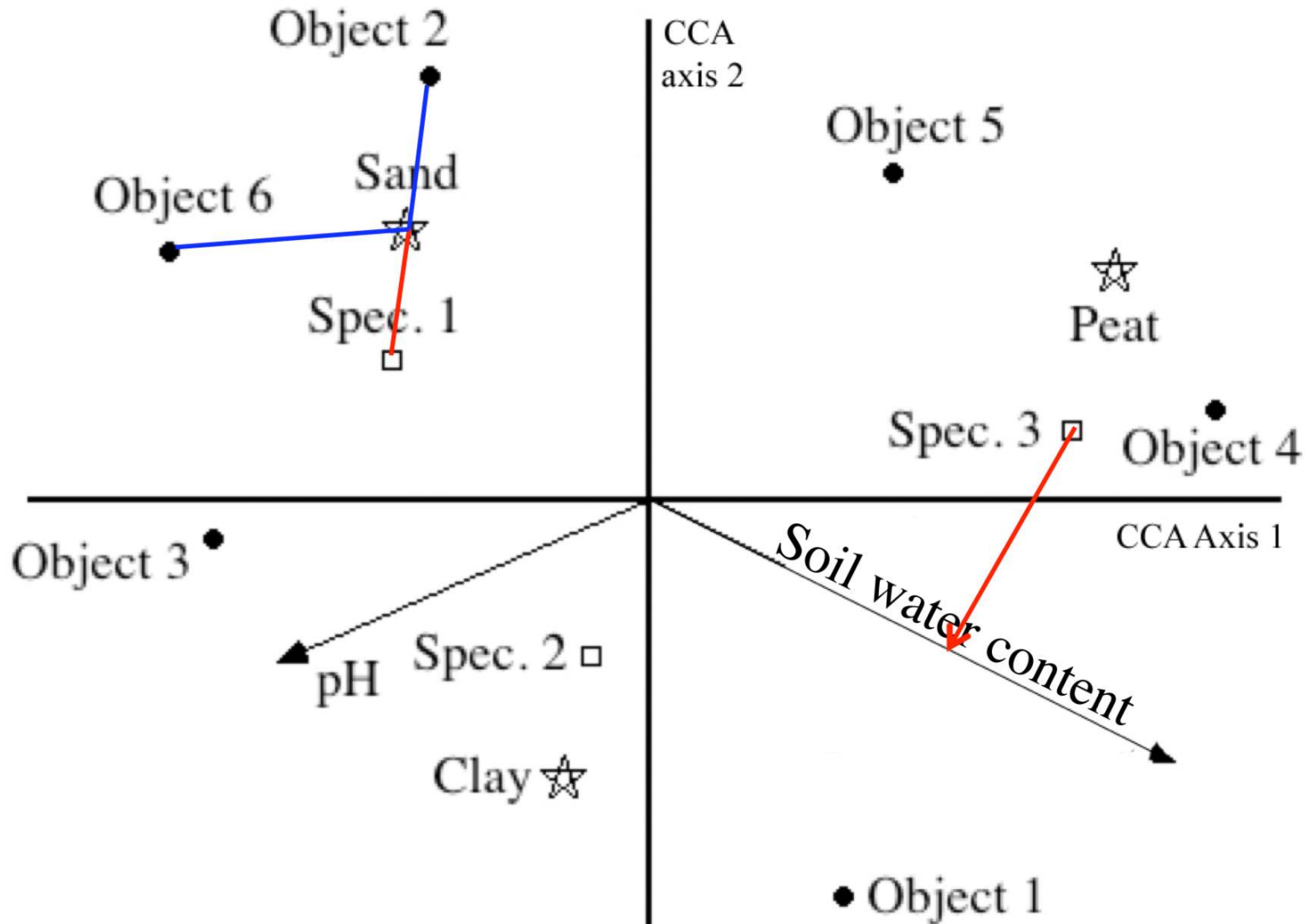
Projetando uma variável indica a correlação dela com o eixo de ordenação

Figure 11.3 RDA triplot of the data in Table 11.3, scaling 1; the numerical results are in Table 11.4. Open

Como interpretar um triplot de CCA?

- No Scaling 2
- Projetar o objeto perpendicularmente à seta da variável quantitativa fornece o valor ótimo da espécie (nicho) ao longo daquela variável

3. CCA: example with scaling 2

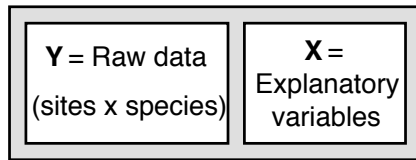


Abundância de espécies na RDA

- Como vimos na aula anterior, matrizes com abundância de espécies geralmente contêm muitos zeros, isso pode gerar problemas em análises canônicas
- Torna as distâncias não-métricas
- RDA preserva distâncias Euclidianas
- Solução: transformar matriz de composição de espécies por Hellinger ou Chord (veja aula 2)

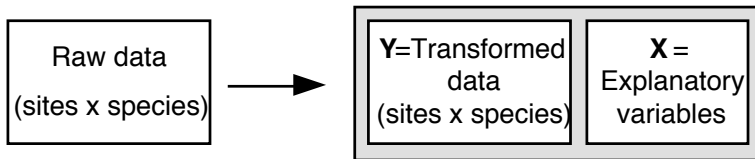
Constrained ordination of species data

(d) Classical approach

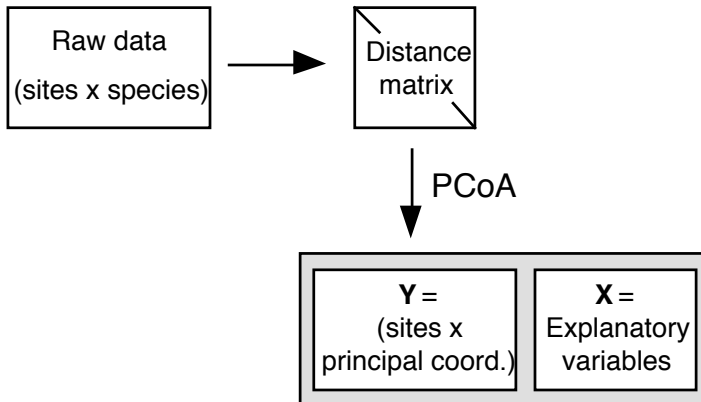


Short gradients: CCA or RDA
 Long gradients: CCA

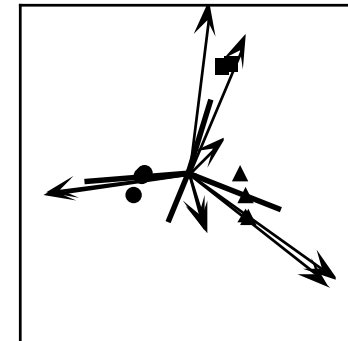
(e) Transformation-based approach (tb-RDA)



(f) Distance-based approach (db-RDA)



Canonical ordination triplot



Representation of elements:
 Species = arrows
 Sites = symbols
 Explanatory variables = lines

Quando usar uma CCA ou uma RDA?

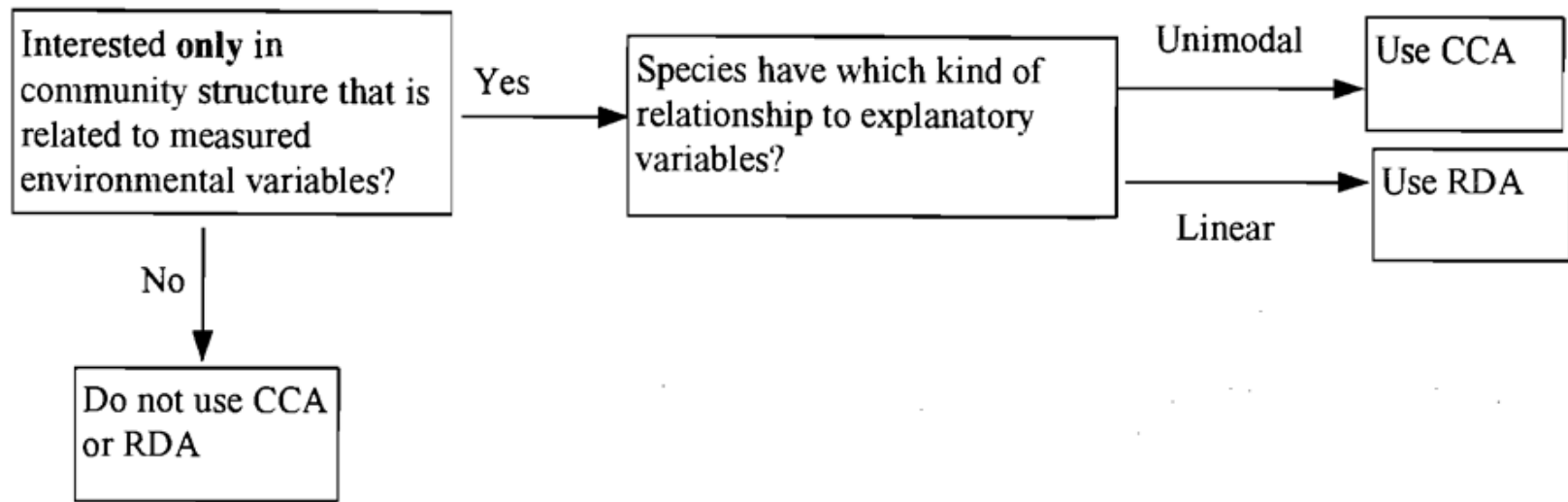
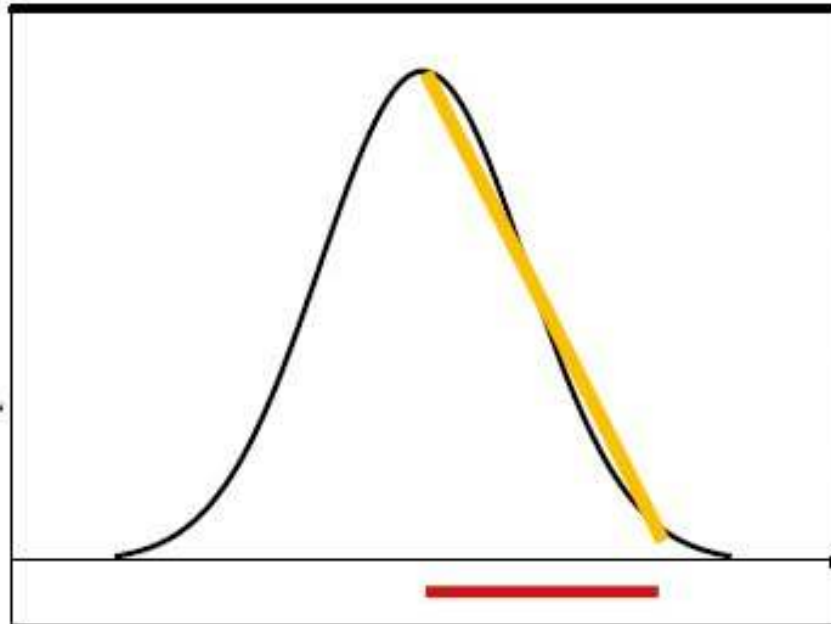


Figure 21.1. Decision tree for using CCA for community data. Assume that we have a site \times species matrix and a site \times environment matrix and that chi-square distances are acceptable. RDA is a constrained ordination method based on a linear model (see "Variations" below).

Isto é um trabalho para:



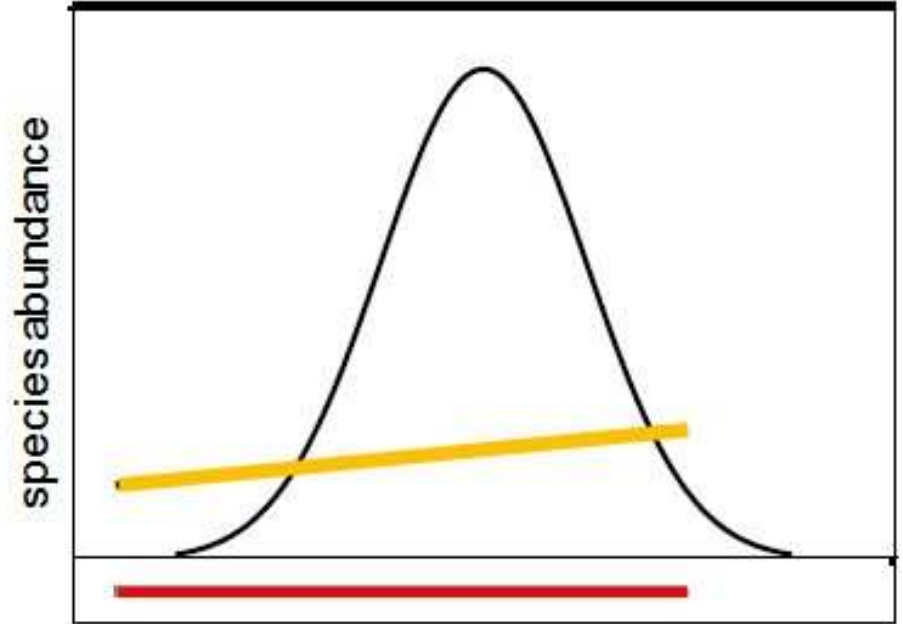
short ecological gradient



environmental gradient (e.g. elevation)



long ecological gradient



environmental gradient (e.g. elevation)

Como vou saber se o gradiente é longo ou curto, ora bolas?

- Faça uma DCA
- O comprimento do DCA1 é muito útil para determinar o comprimento do gradiente ambiental
 - Gradiente unimodal => 4 SD
 - Comprimento aproximado em que espécies substituem totalmente uma a outra
 - Gradiente linear => 1-3 SD
 - Mudança de metade da composição ocorre 1 - 1.4 SD

Quais e quantas variáveis incluir no modelo?

5. Forward selection of environmental variables

The principle of forward selection is as follows:

1. Compute, in turn, the independent contribution of each of the m explanatory variables to the explanation of the variation of the response data table. This is done by running m separate canonical analyses.
2. Test the significance of the contribution of the best variable.
3. If it is significant, include it into the model as a first explanatory variable.

5. Forward selection of environmental variables

4. Compute (one at a time) the **partial** contributions (conditional effects) of the $m-1$ remaining explanatory variables, controlling for the effect of the one already in the model.
5. Test the significance of the best partial contribution among the $m-1$ variables.
6. If it is significant, include this variable into the model.
7. Compute the partial contributions of the $m-2$ remaining explanatory variables, controlling for the effect of the two already in the model.
8. The procedure goes on until no more significant partial contribution is found.

5. Forward selection of environmental variables

- a) First of all, **forward selection is too liberal** (i.e., it allows too many explanatory variables to enter a model).

Before running a forward selection, **always perform a global test** (including **all** explanatory variables). If, *and only if* the global test is significant, run the forward selection.

5. Forward selection of environmental variables

- b) Even if the global test is significant, forward selection is too liberal. Simulations have shown that, in addition to the usual alpha level, one must add a **second stopping criterion** to forward selection: the model under construction must not have an R^2_{adj} higher than that of the global model (i.e., the model containing all explanatory variables).

Blanchet F. G., P. Legendre, and D. Borcard. 2008. Forward selection of explanatory variables. *Ecology* 89: 2623-2632.

5. Forward selection of environmental variables

- f) Forward selection *can help* when several candidate explanatory variables are strongly correlated, but the choice has no *a priori* ecological validity. In this case it is often advisable to eliminate one of the intercorrelated variables on ecological basis rather than on statistical basis.
- g) In cases where several correlated explanatory variables are present, without clear *a priori* reasons to eliminate one or the other, one can examine the **variance inflation factors (VIF)**.

5. Forward selection of environmental variables

In **R**, variable selection for ecological data can be run with the following functions:

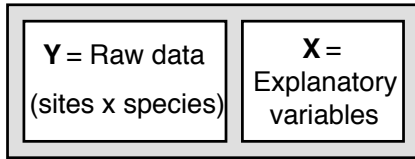
	Forward	Backward	R^2_a	Factors
<code>forward.sel()</code> {adespatial}	✓		✓	
<code>ordistep()</code> {vegan}	✓	✓		✓
<code>ordiR2step()</code> {vegan}	✓		✓	✓

distance-based RDA

- Permite usar qualquer matriz de distância para relacionar com variáveis ambientais ou espaciais (e.g., MEMs)

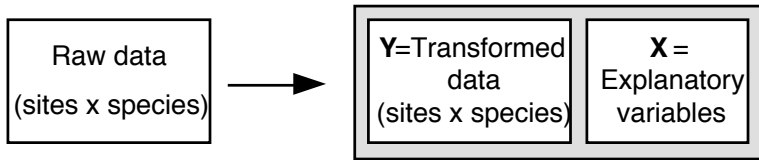
Constrained ordination of species data

(d) Classical approach

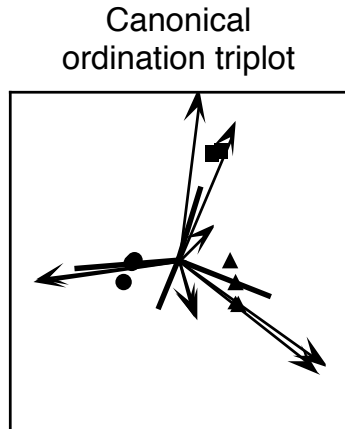
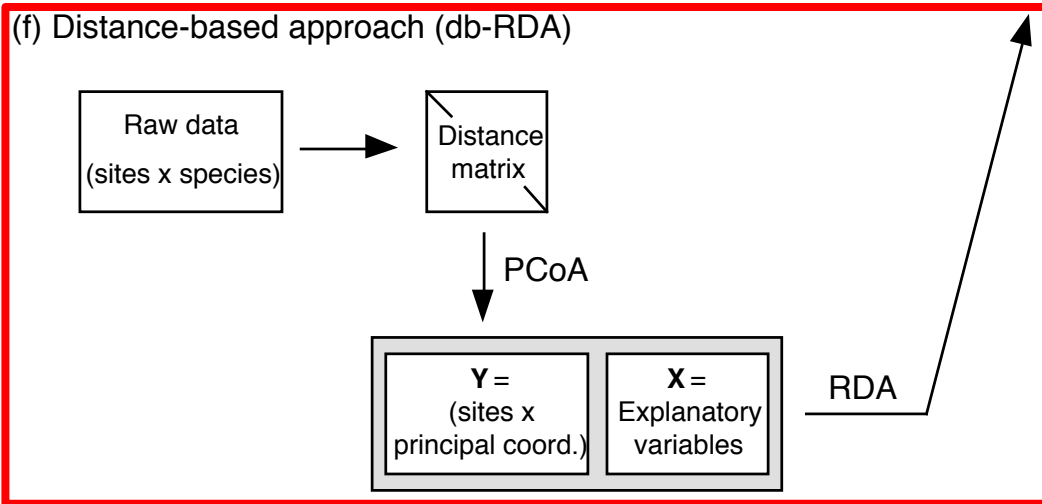


Short gradients: CCA or RDA
Long gradients: CCA

(e) Transformation-based approach (tb-RDA)



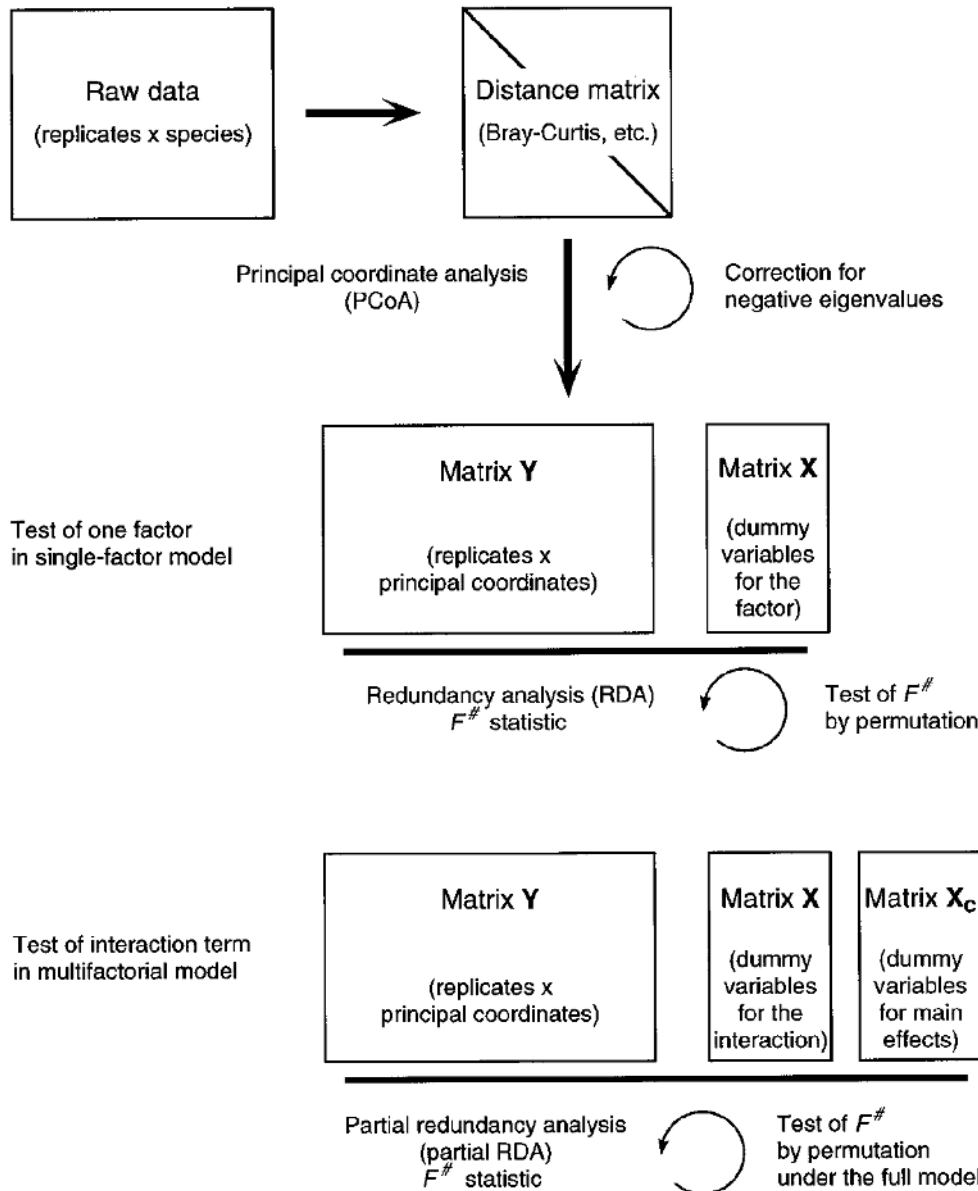
(f) Distance-based approach (db-RDA)



Representation of elements:
Species = arrows
Sites = symbols
Explanatory variables = lines

DISTANCE-BASED REDUNDANCY ANALYSIS: TESTING MULTISPECIES RESPONSES IN MULTIFACTORIAL ECOLOGICAL EXPERIMENTS

PIERRE LEGENDRE¹ AND MARTI J. ANDERSON²



- Permite particionar a SQ de uma matriz de distância (analisada com PCoA) entre fatores.
- Também é útil quando a matriz preditora contém variáveis contínuas
- Artigo original fornece método para escolher quantos eixos incluir no modelo final

CANONICAL ANALYSIS OF PRINCIPAL COORDINATES: A USEFUL METHOD OF CONSTRAINED ORDINATION FOR ECOLOGY

MARTI J. ANDERSON^{1,3} AND TREVOR J. WILLIS^{1,2}

TABLE 2. Methods of constrained ordination relating response variables, **Y** (species abundance variables) with predictor variables, **X** (such as quantitative environmental variables or qualitative variables that identify factors or groups as in ANOVA).

Name of method (acronyms, synonyms)	Distance measure preserved	Relationship of ordination axes with original variables	Takes into account correlation structure†
Redundancy Analysis (RDA)	Euclidean distance	linear with X , linear with fitted values, $\hat{\mathbf{Y}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$... among variables in X , but <i>not</i> among variables in Y
Canonical Correspondence Analysis (CCA)§	chi-square distance	linear with X , approximately unimodal with Y , linear with fitted values, $\hat{\mathbf{Y}}^*$... among variables in X , but <i>not</i> among variables in Y
Canonical Correlation Analysis (CCorA, COR)	Mahalanobis distance	linear with X , linear with Y	... among variables in X , and among variables in Y
Canonical Discriminant Analysis (CDA; Canonical Variate Analysis, CVA; Discriminant Function Analysis, DFA)	Mahalanobis distance	linear with X , linear with Y	... among variables in X , and among variables in Y
Canonical Analysis of Principal Coordinates (CAP; Generalized Discriminant Analysis)	any chosen distance or dissimilarity	linear with X , linear with \mathbf{Q}_m ; unknown with Y (depends on distance measure)	... among variables in X , and among principal coordinates \mathbf{Q}_m

PCoA+CCorA (CDA)=CAP

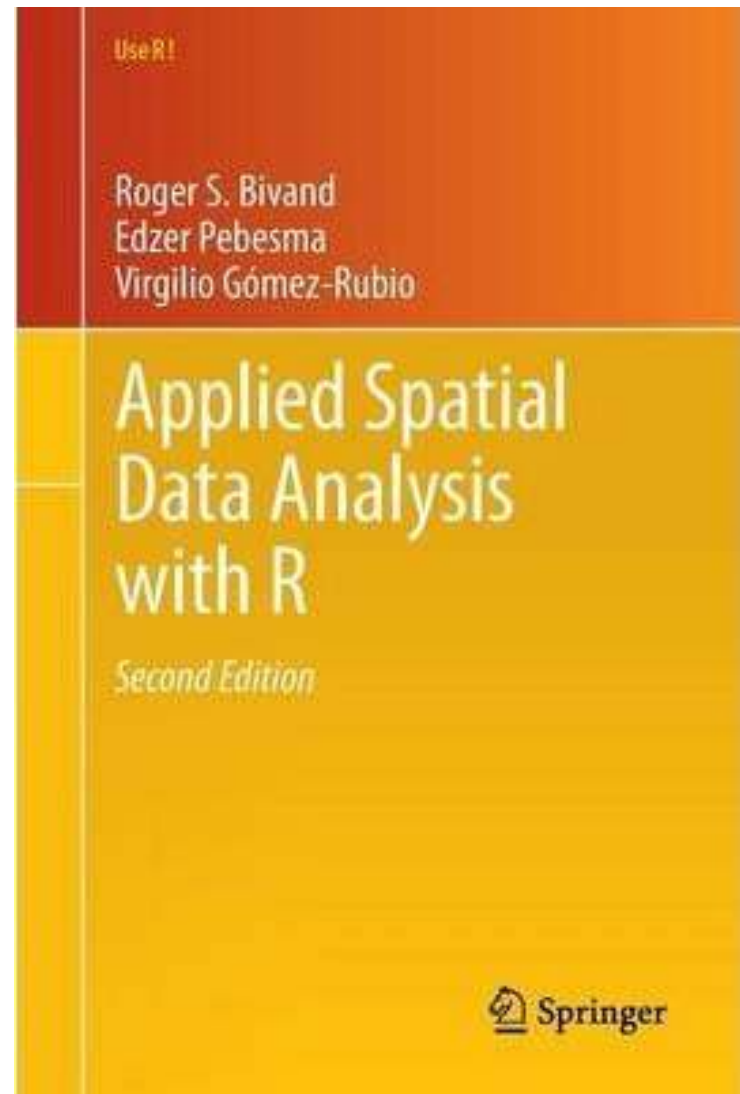
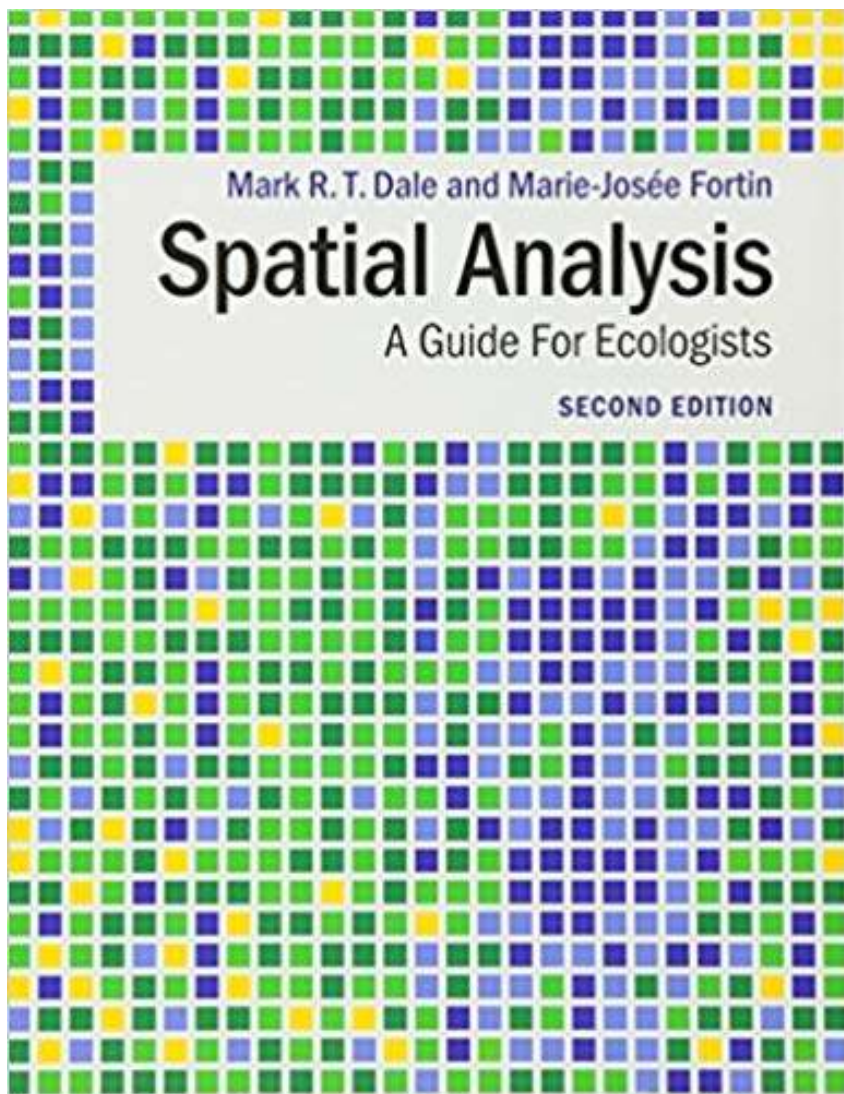
TABLE 2. Extended.

Criterion for drawing ordination axes	Matrix for eigenvalue decomposition‡
finds axis of maximum correlation between Y and some linear combination of variables in X (i.e., multivariate multiple regression of Y on X , followed by PCA on fitted values, $\hat{\mathbf{Y}}$)	$\mathbf{Y}'_c \mathbf{X}_c (\mathbf{X}'_c \mathbf{X}_c)^{-1} \mathbf{X}'_c \mathbf{Y}_c$
same as RDA, but Y are transformed to \mathbf{Y}^* and weights (square roots of row sums) are used in multiple regression	$\mathbf{C}^{-1/2} \mathbf{Y}' \mathbf{X}_c (\mathbf{X}'_c \mathbf{R} \mathbf{X}_c)^{-1} \mathbf{X}'_c \mathbf{Y} \mathbf{C}^{-1/2}$
finds linear combination of variables in Y and X that are maximally correlated with one another	$\mathbf{Y}'_s \mathbf{X}_c (\mathbf{X}'_c \mathbf{X}_c)^{-1} \mathbf{X}'_c \mathbf{Y}_s$
finds axis that maximizes differences among group locations. Same as CCorA when X contains group identifiers. Equivalent analysis is regression of X on Y , provided X contains orthogonal contrast vectors	as for CCorA, but CDA eigenvalues are $\delta_i^2 / (1 - \delta_i^2)$, and corresponding eigenvectors are $(1 - \delta_i^2)^{-1/2}$ times the CCorA eigenvectors¶
finds linear combination of axes in \mathbf{Q}_m and in X that are maximally correlated, or (if X contains group identifiers) finds axis in PCO space that maximizes differences among group locations	$\mathbf{Q}'_m \mathbf{X}_c (\mathbf{X}'_c \mathbf{X}_c)^{-1} \mathbf{X}'_c \mathbf{Q}_m$

Como escolher entre tb-RDA e db-RDA?

- As duas fornecem resultados equivalentes
 - Quando usado distância Euclidiana=RDA simples
- db-RDA deve ser usada somente em análises envolvendo funções de distância que não podem ser obtidas por transformações (veja aula 2), p.ex., dados binários
 - Jaccard
 - Sørensen
 - Distância geodésica
 - Percentage difference (Bray-Curtis)
 - Whittaker
 - Canberra
 - Mahalanobis
 - Gower

RDA & CCA parciais



PRIMEIRA LEI DA GEOGRAFIA: “Tudo está relacionado com tudo, mas coisas próximas são mais parecidas do que coisas distantes”

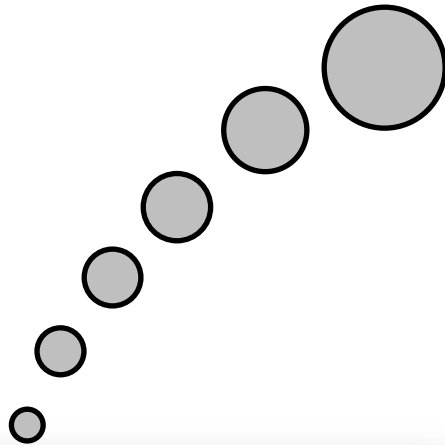


SPATIAL AUTOCORRELATION: TROUBLE OR NEW PARADIGM?¹

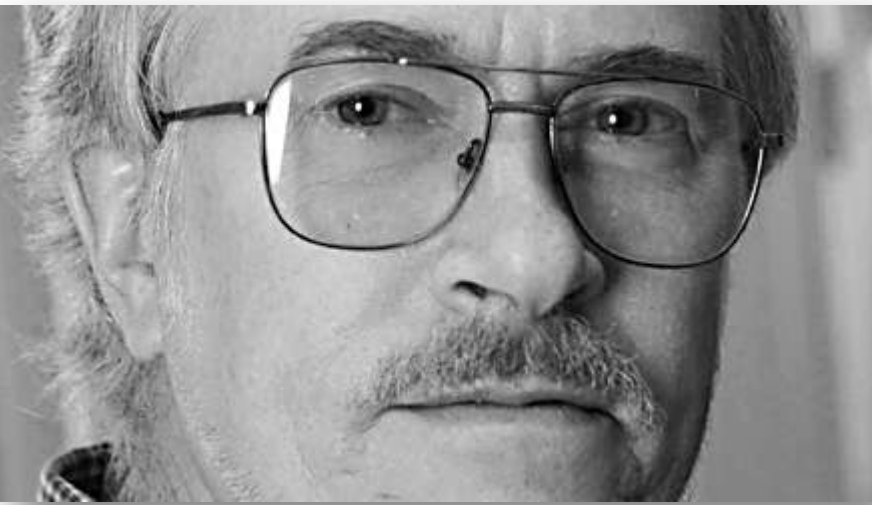
PIERRE LEGENDRE

*Département de sciences biologiques, Université de Montréal, C.P. 6128, succursale A,
Montréal, Québec, Canada H3C 3J7*

Ecology, 74(6), 1993, pp. 1659–1673



A **autocorrelação espacial** é um problema para testes estatísticos pois dados autocorrelacionados **violam** o **pressuposto de independência** da maioria dos procedimentos estatísticos.



All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices

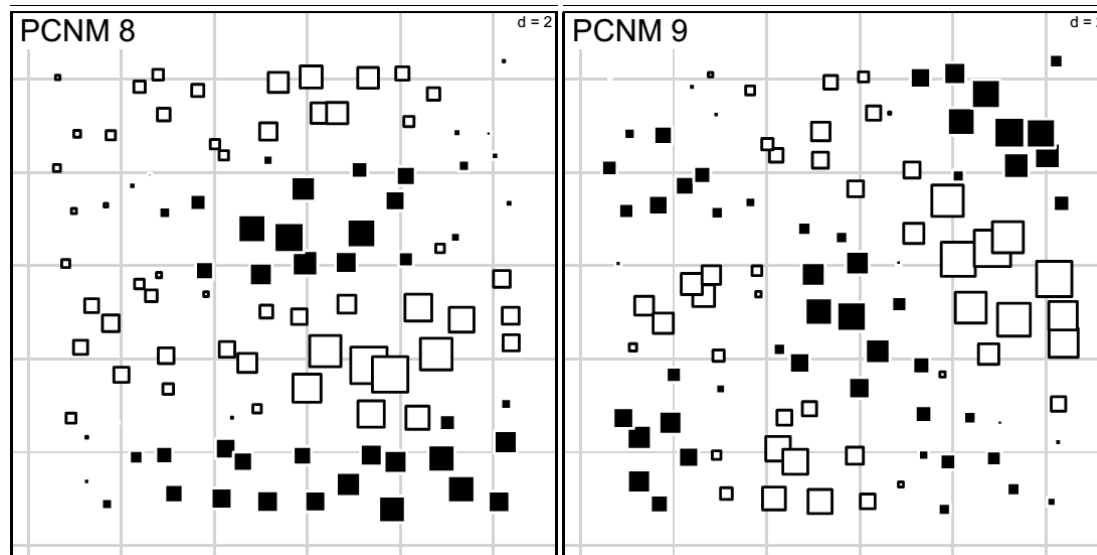
Daniel Borcard *, Pierre Legendre

2002

DISSECTING THE SPATIAL STRUCTURE OF ECOLOGICAL DATA AT MULTIPLE SCALES

2004

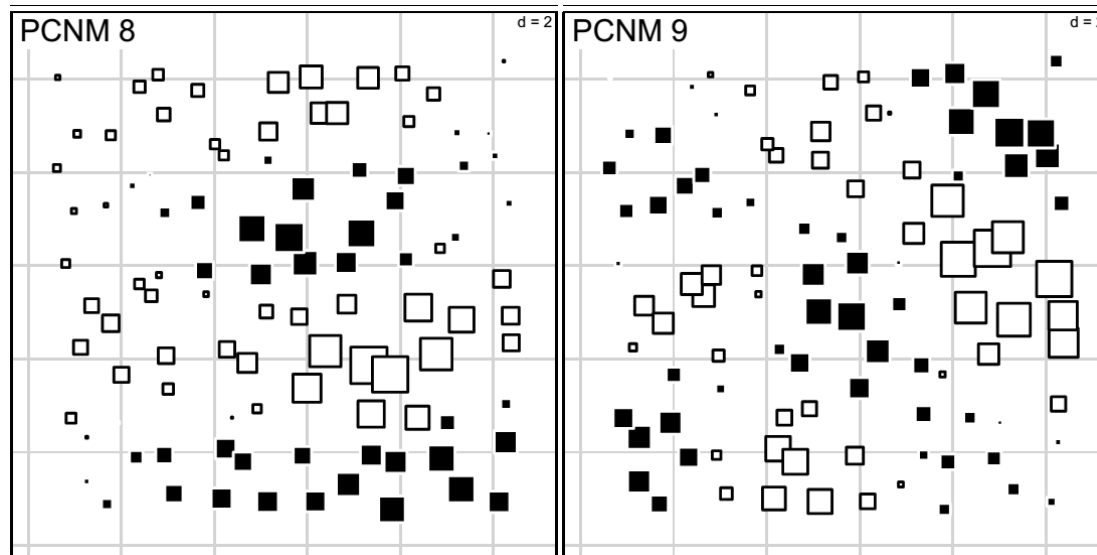
DANIEL BORCARD,^{1,4} PIERRE LEGENDRE,¹ CAROL AVOIS-JACQUET,^{1,2} AND HANNA TUOMISTO³



All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices

Daniel Borcard *, Pierre Legendre

PCNM ===== dbMEM



dbMEM através de um exemplo numérico

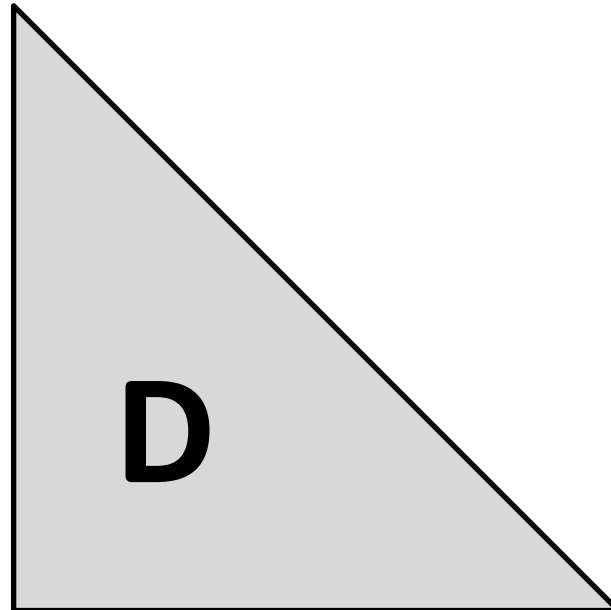


dbMEM através de um exemplo numérico

1º passo: calcular a distância entre os pontos



dbMEM através de um exemplo numérico



dbMEM através de um exemplo numérico

2º passo: escolher a distância de truncamento ou de separação *baseado numa Minimum Spanning Tree = truncate distance*

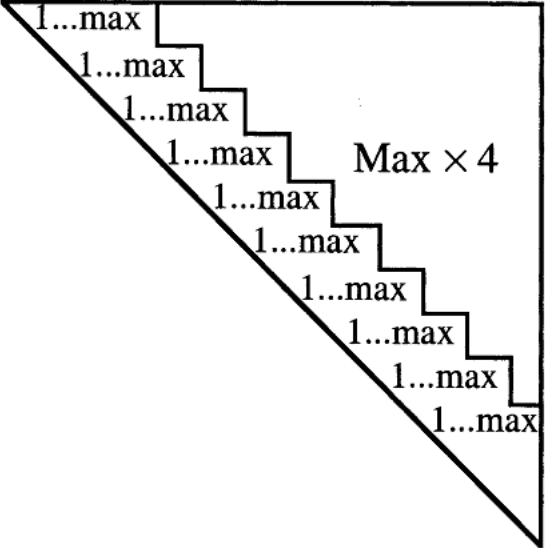
- o valor que separa é chamado de *thresh*



O valor *thresh* separa as distâncias em dois grupos:

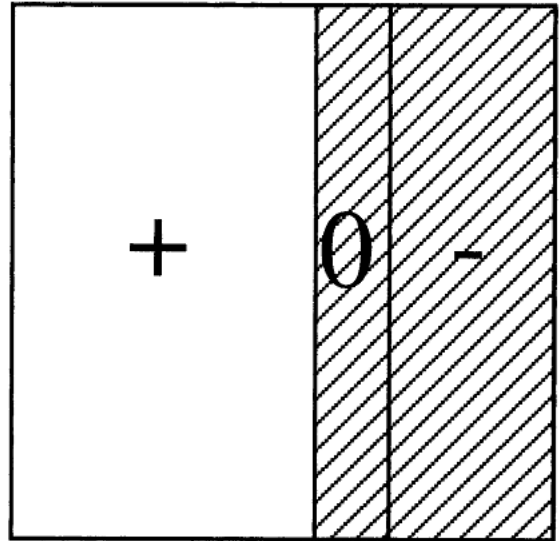
pequena e grande

Truncated Euclidean distance matrix
= neighbour matrix



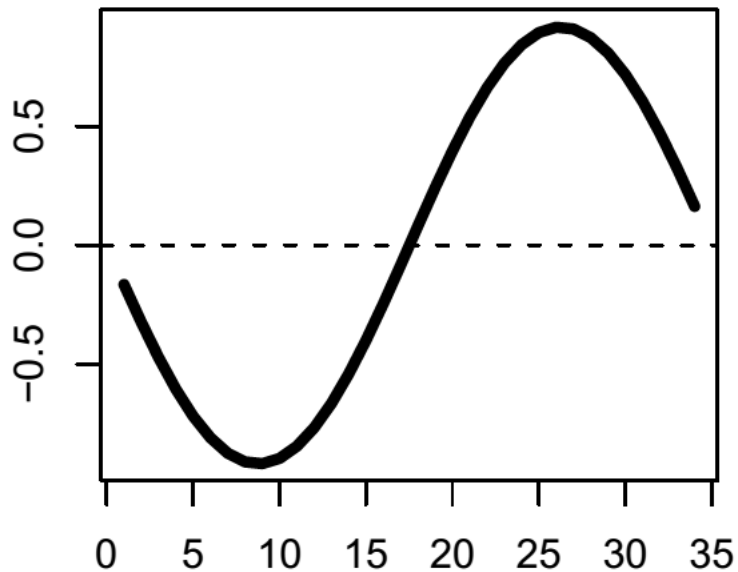
PCoA →

Principal coordinate analysis

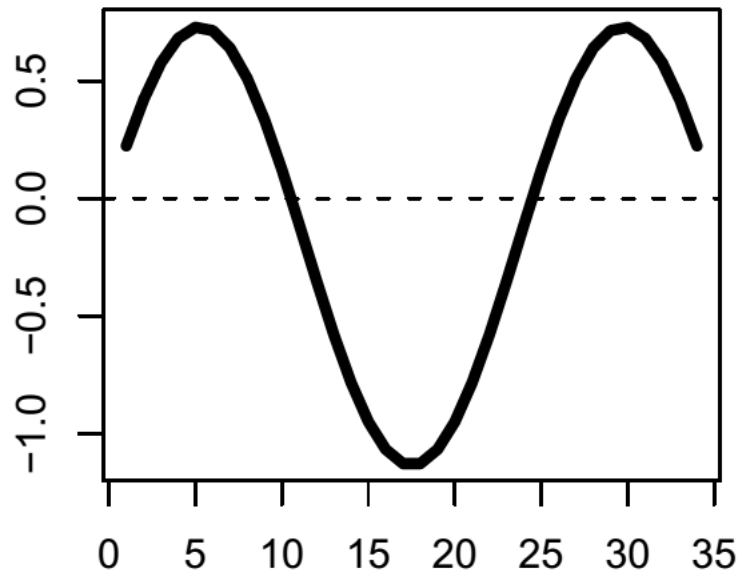


Eigenvectors

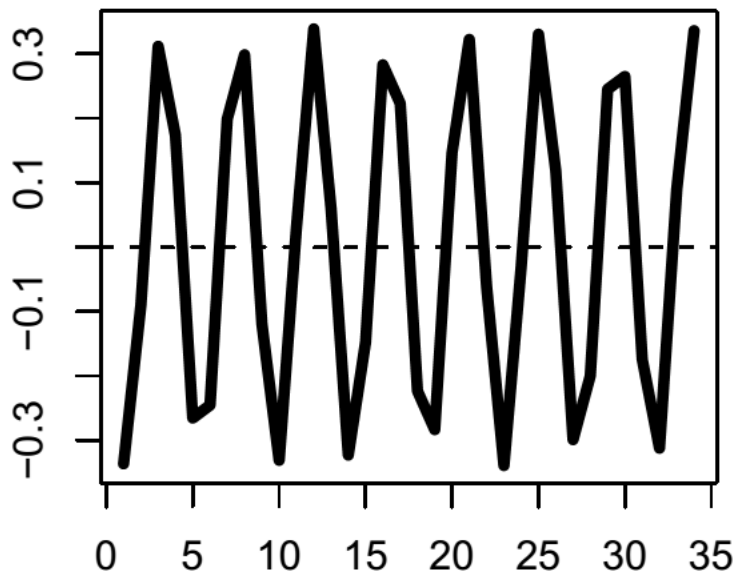
dbMEM 1



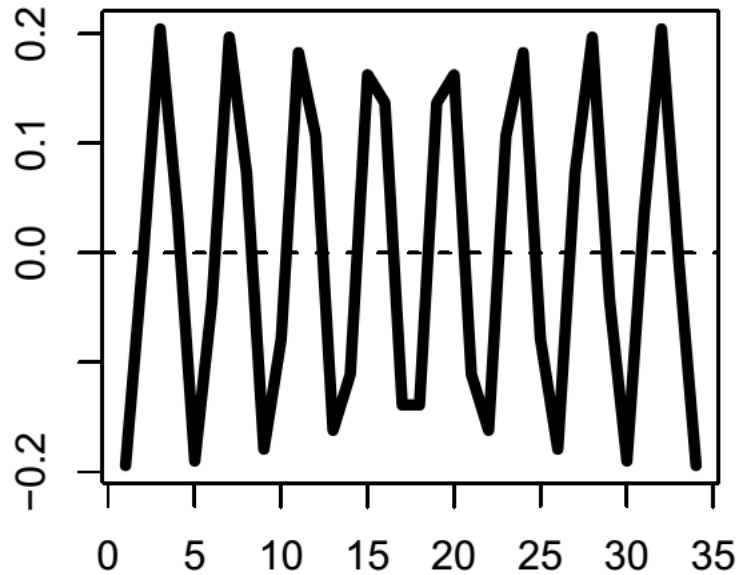
dbMEM 2



dbMEM 15



dbMEM 16



Index

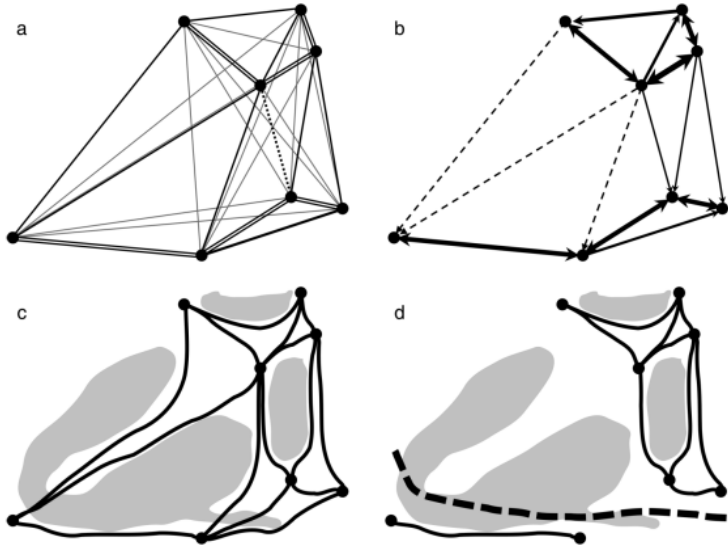
Index

Community ecology in the age of multivariate multiscale spatial analysis

S. DRAY,^{1,16} R. PÉLISSIER,^{2,3} P. COUTERON,² M.-J. FORTIN,⁴ P. LEGENDRE,⁵ P. R. PERES-NETO,⁶ E. BELLIER,^{7,8}
R. BIVAND,⁹ F. G. BLANCHET,¹⁰ M. DE CÁCERES,¹¹ A.-B. DUFOUR,¹ E. HEEGAARD,¹² T. JOMBART,^{1,13} F. MUNOZ,²
J. OKSANEN,¹⁴ J. THIOULOUSE,¹ AND H. H. WAGNER¹⁵

- Generalização dos dbMEM para qualquer rede de distância
- Dray et al. (2006) => Unifica a abordagem dos filtros espaciais (Griffith 2000) e Borcard & Legendre (2002)
- Passam a se chamar MEM => Moran Eigenvector Maps

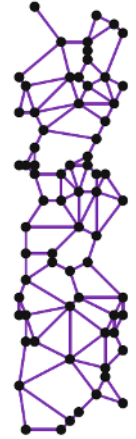
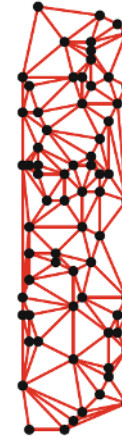
Contexto da paisagem Dray et al. 2012



Redes de conectividade irregulares

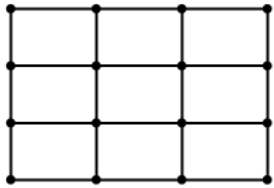
Delaunay triangulation

Gabriel graph

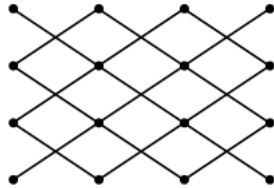


Redes de conectividade regulares

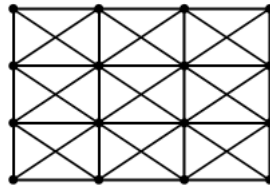
(a) Rectangular grid, rook connection



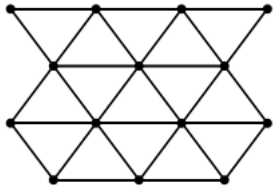
(b) Rectangular grid, bishop connection



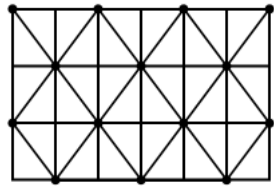
(c) Rectangular grid, king (or queen) connection



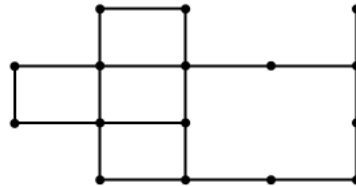
(d) Staggered rows, six neighbours



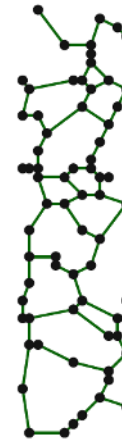
(e) Staggered rows, eight neighbours



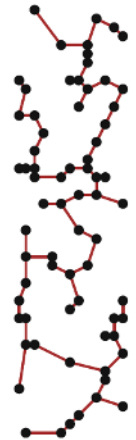
(f) Hollow grid, rook connection



Relative neighbourhood



Minimum spanning tree



Qual método posso relacionar a contribuição relativa da **variação ambiental e **espacial** na composição de espécies?**





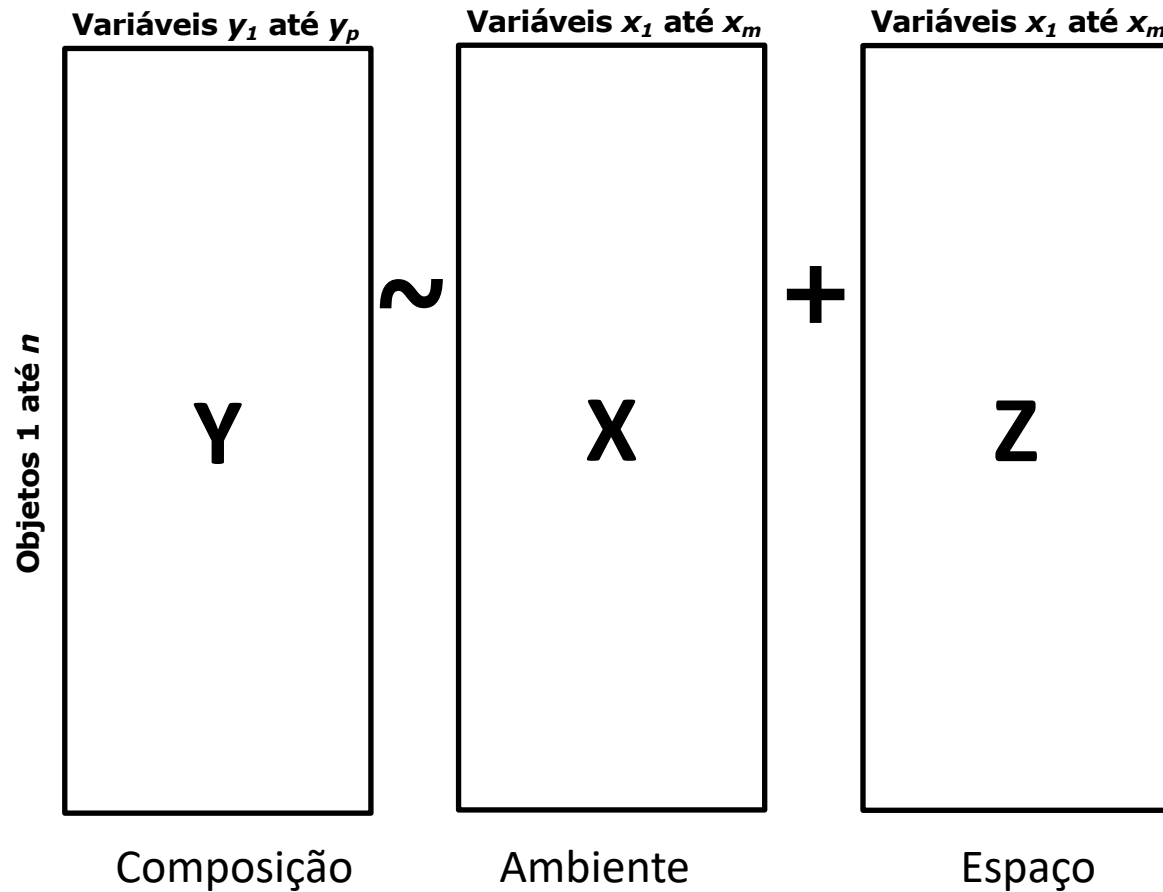
NÃO SEI

**MELHOR PERGUNTAR NO POSTO
IPIRANGA**

Partição de variância com RDA parcial

- Permite *controlar covariáveis* que quando analisando o efeito de **X** em **Y**:
 - Ex.: TEMPO, ESPAÇO, filogenia
- Saber o efeito isolado de um conjunto de variáveis
- Controlar para desenhos aninhados, medidas repetidas, e blocos

Análises canônicas parciais



PARTIALLING OUT THE SPATIAL COMPONENT OF
ECOLOGICAL VARIATION¹

DANIEL BORCARD

Institut de Zoologie, Université de Neuchâtel, Chantemerle 22, CH-2007 Neuchâtel, Switzerland

PIERRE LEGENDRE AND PIERRE DRAPEAU

*Département de Sciences biologiques, Université de Montréal, C.P. 6128, succursale A,
Montréal, Québec, Canada H3C 3J7*

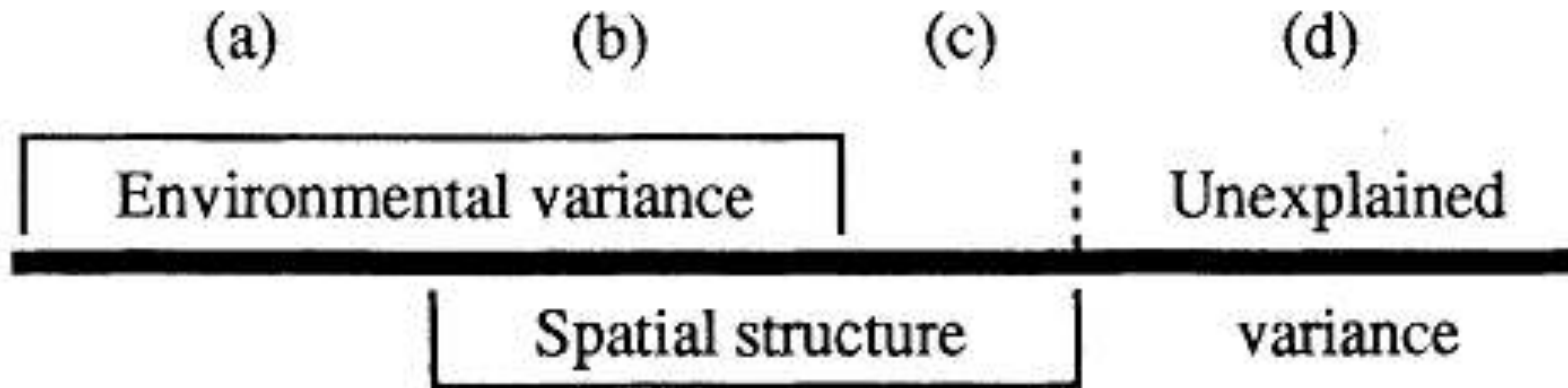
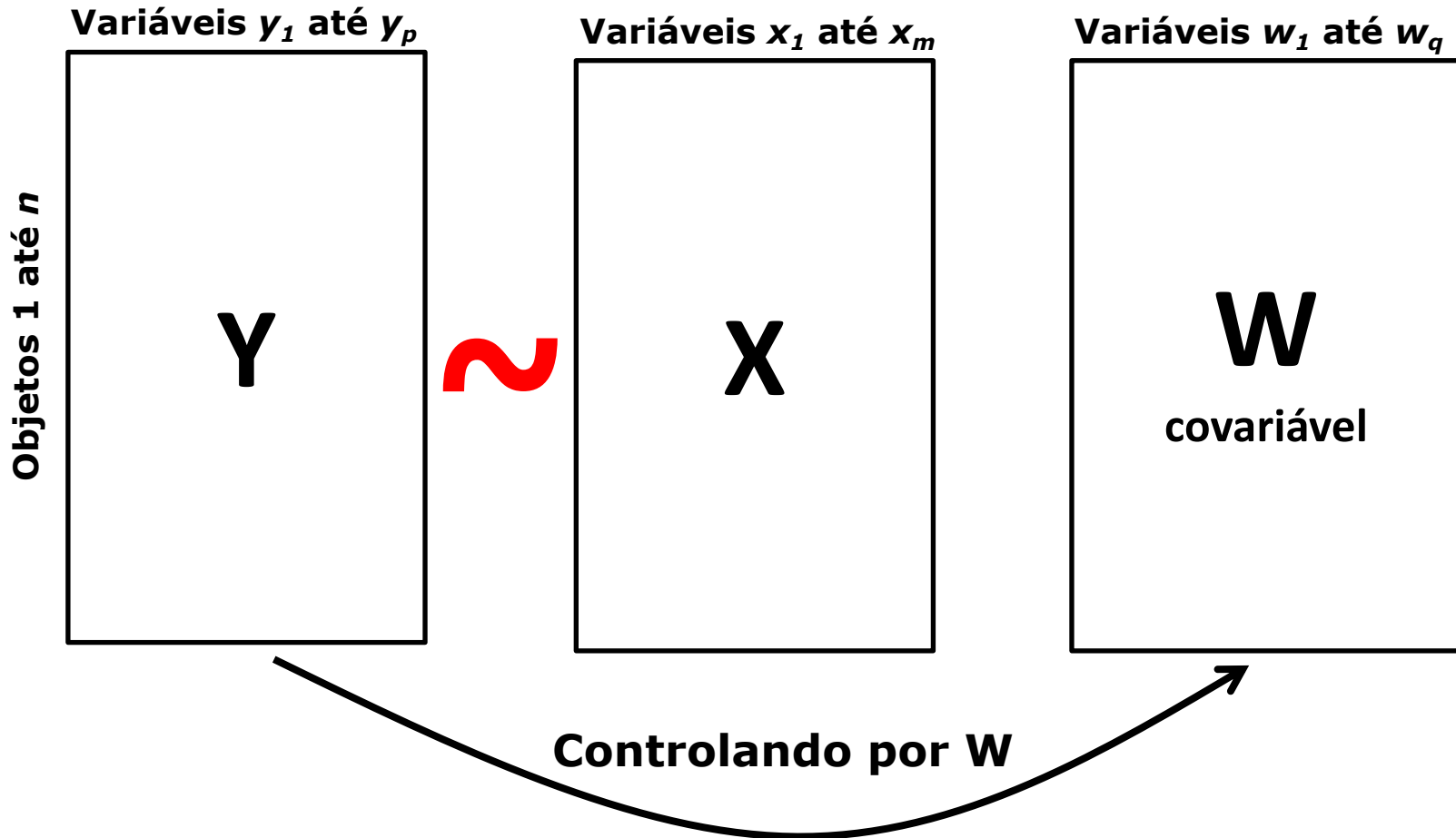


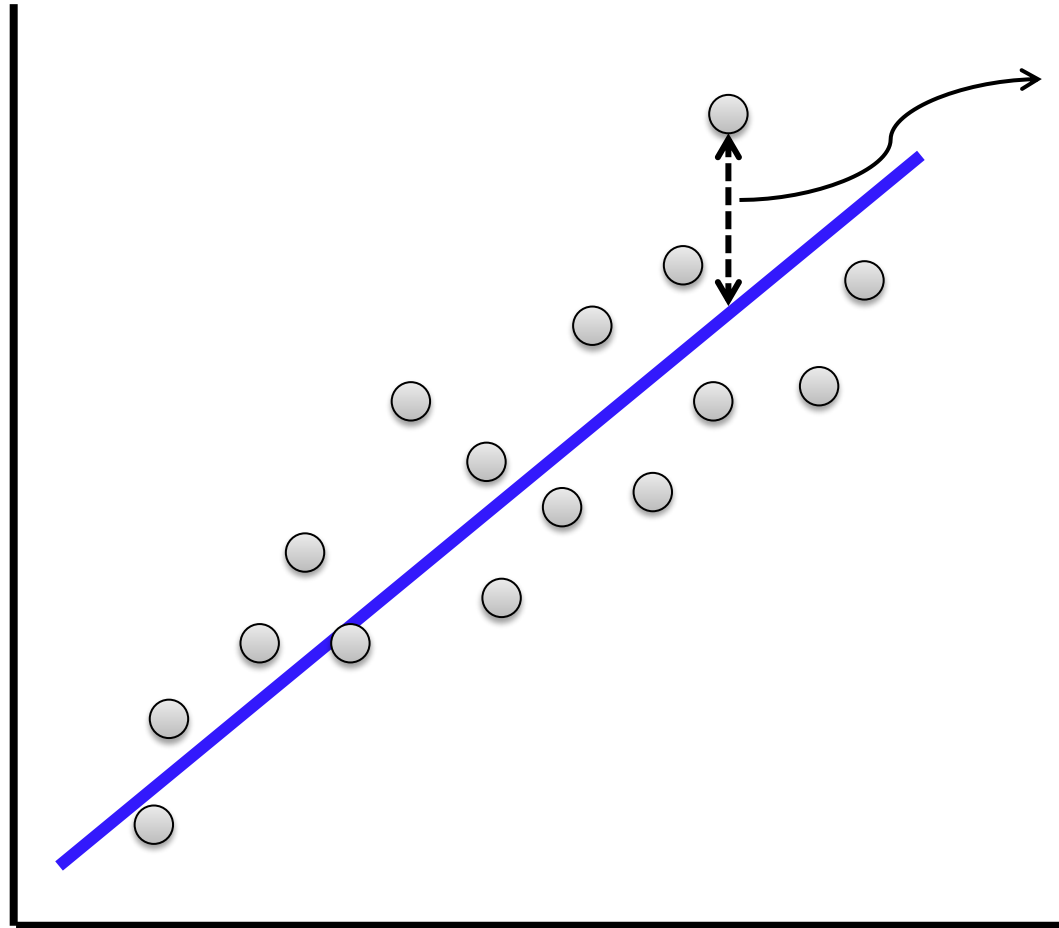
FIG. 2. Variation partitioning of a species data table, showing that fraction (b) is the intersection of the environmental and spatial components of the species variation.

RDA parcial



1º passo

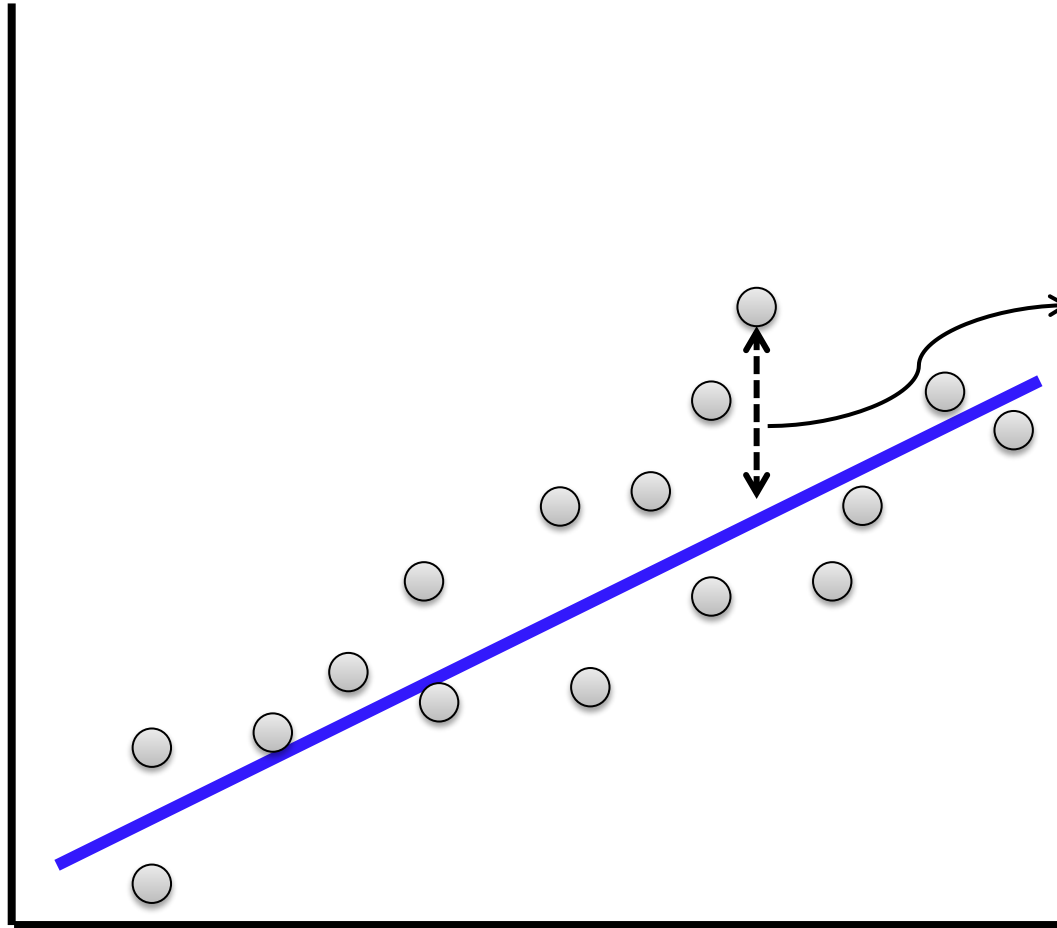
Y



W
covariável

2º passo

X



Resíduo = $X_{res|w}$

W
covariável

3º passo

$Y_{res|w}$

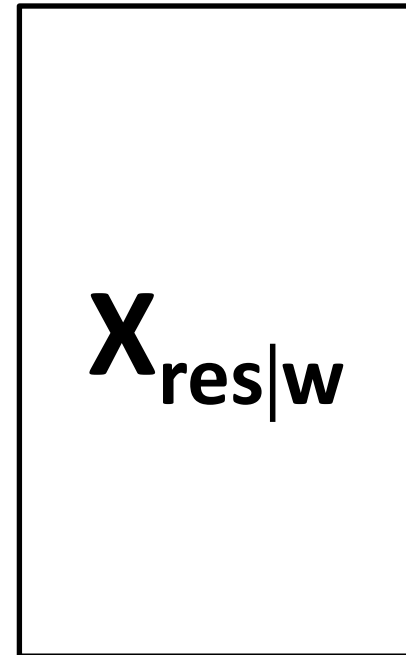
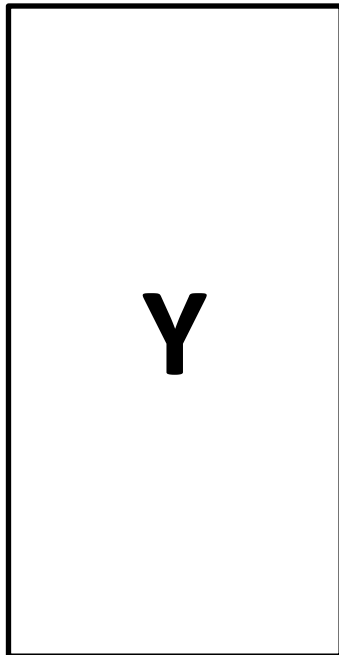
RDA →

$X_{res|w}$

$X_{res|w}$

Varição ambiental **não**
explicada pelo espaço

OU 3º passo

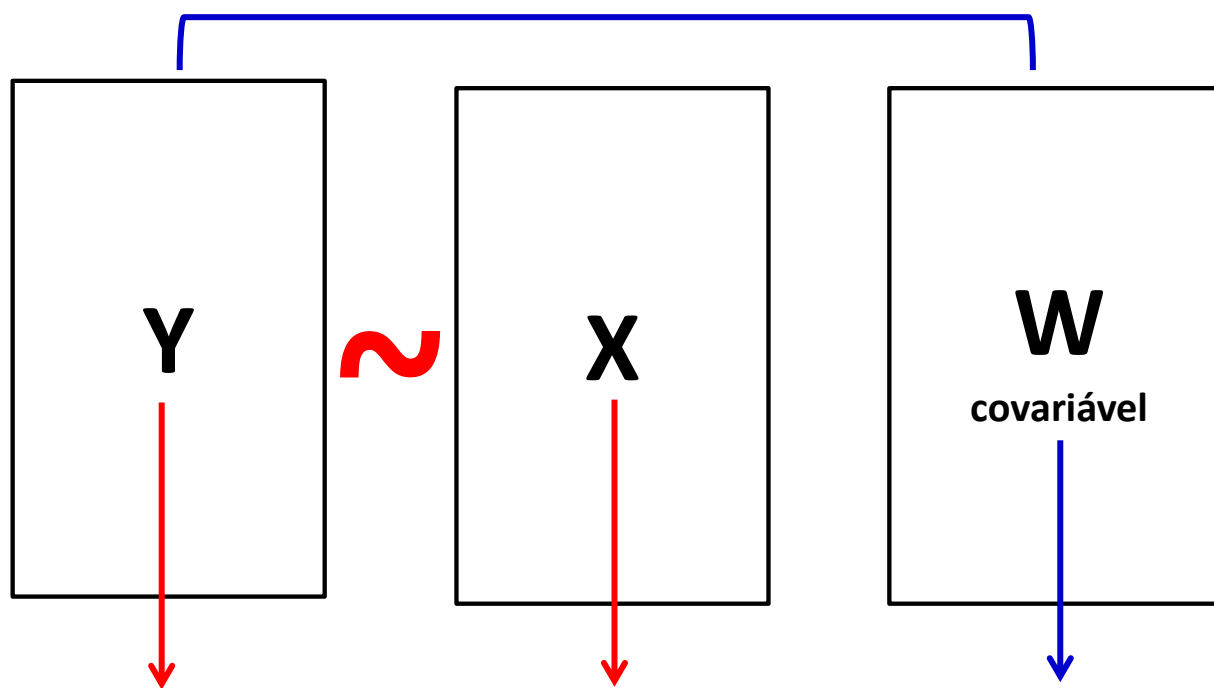


X_{res|w}

Varição ambiental **não**
explicada pelo espaço

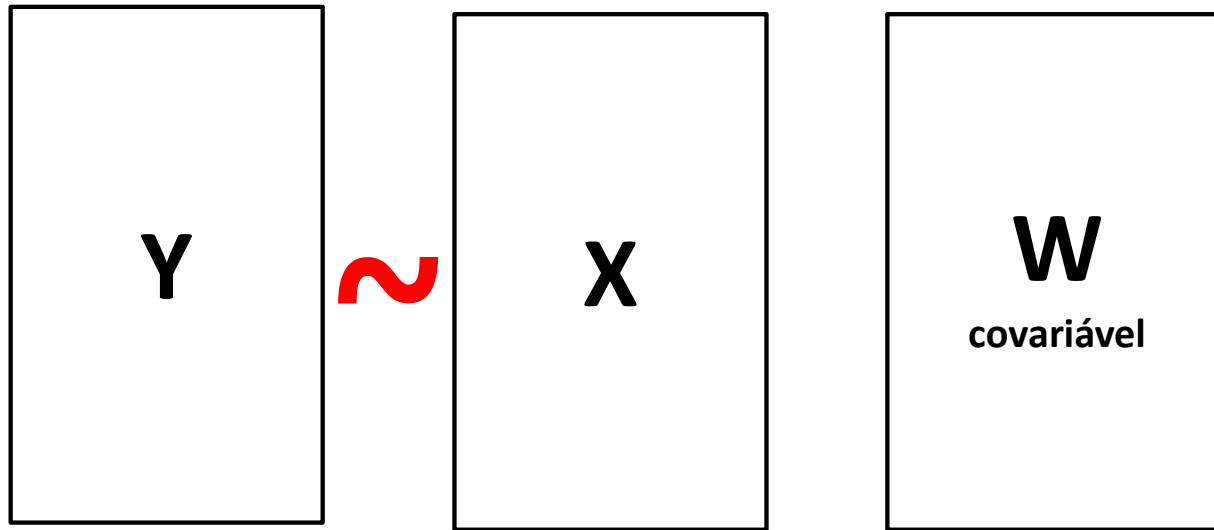
Em quais casos devo usar a RDA ou CCA parcial?

- Ambas são utilizadas quando desejamos controlar ou particionar uma variável que também afeta a variável resposta;
- RDA – assume relação linear entre espécies-ambiente;
- CCA – assume relação unimodal entre espécies-ambiente



```
rda (mite.pad, mite.env, mite.pcnm)
```

```
rda (mite.pad~SubsDens+WatrCont+Condition  
(V1+V2+V3) )
```



	Inertia	Proportion	Rank
Total	<u>0.39428</u>	1.00000	
Conditional	<u>0.24563</u>	0.62300	22
Constrained	<u>0.02000</u>	0.05072	2
Unconstrained	0.12865	0.32628	35

Inertia is variance

Variância total explicada

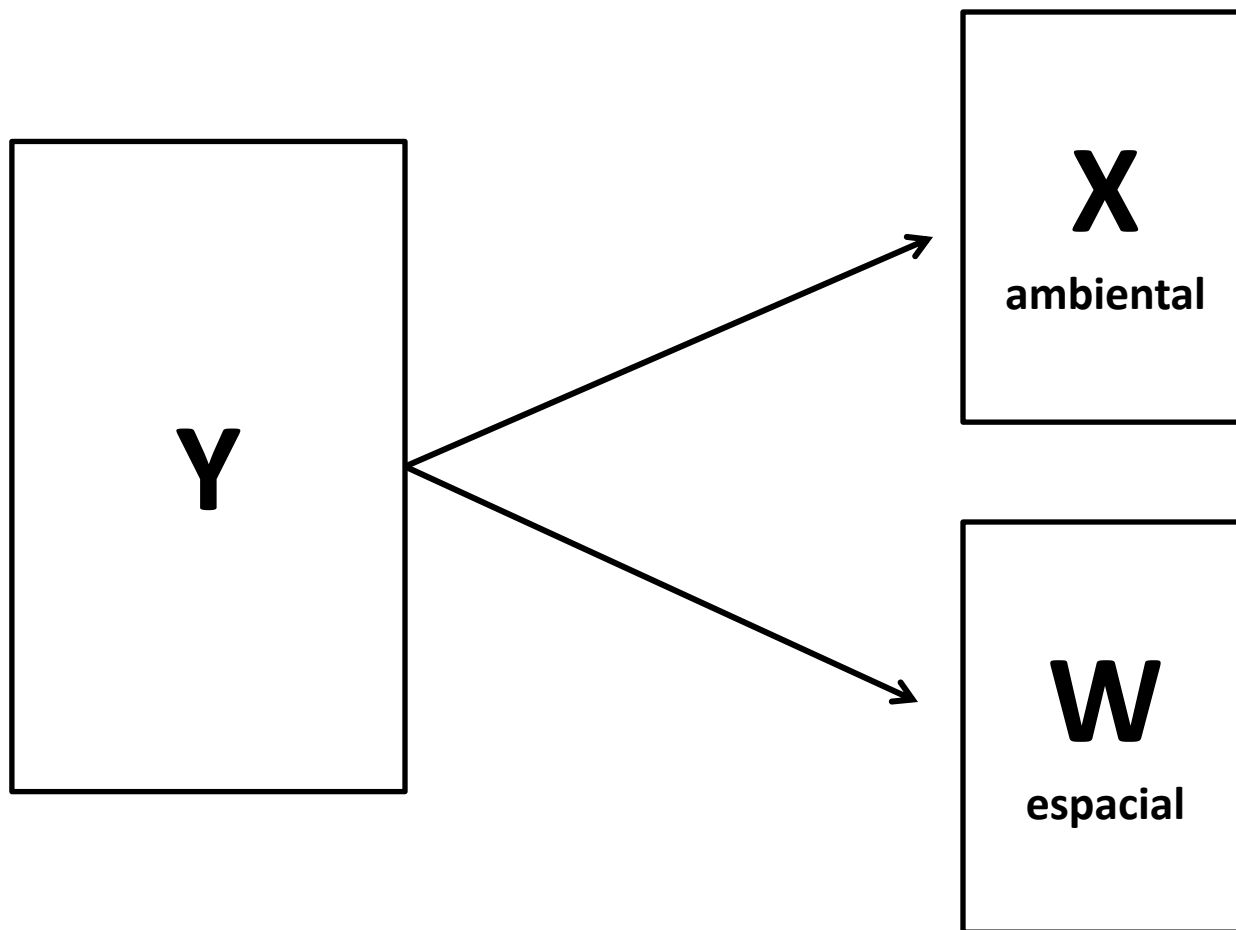
Variância explicada por X e W

Variância explicada **SOMENTE** por X

Partição de variância

- Testa se **dois diferentes grupos** de variáveis (X e W) são **redundantes** ou se cada um deles explica **aspectos únicos** da composição de espécies.

Partição de variância





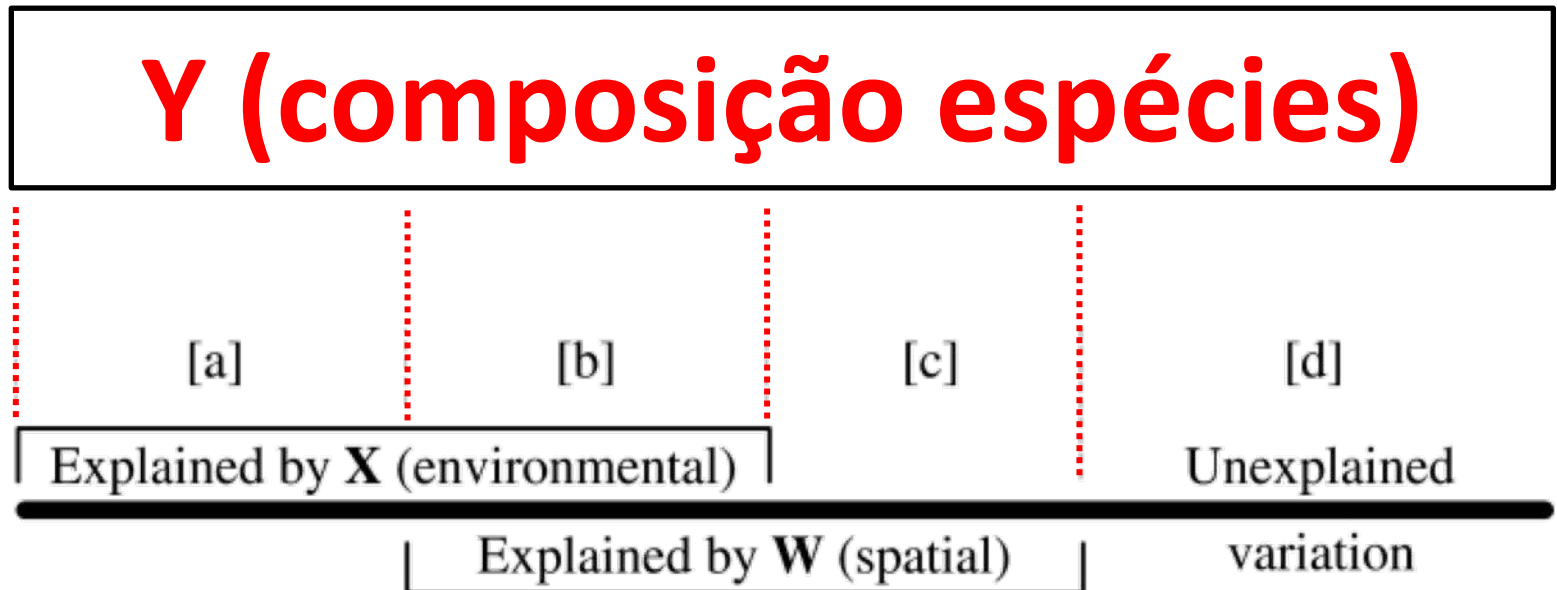
Estimating and controlling for spatial structure in the study of ecological communities

Pedro R. Peres-Neto^{1*} and Pierre Legendre²

VARIATION PARTITIONING OF SPECIES DATA MATRICES: ESTIMATION AND COMPARISON OF FRACTIONS

PEDRO R. PERES-NETO,¹ PIERRE LEGENDRE, STÉPHANE DRAY, AND DANIEL BORCARD

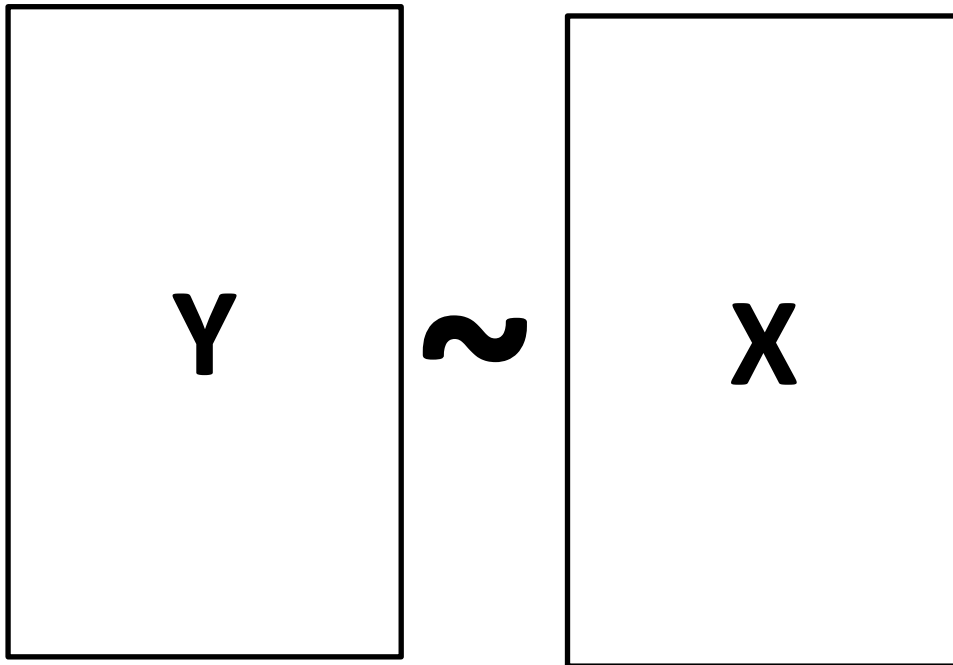
Partição de variância e cálculo do R² canônico parcial



R^2 canônico parcial

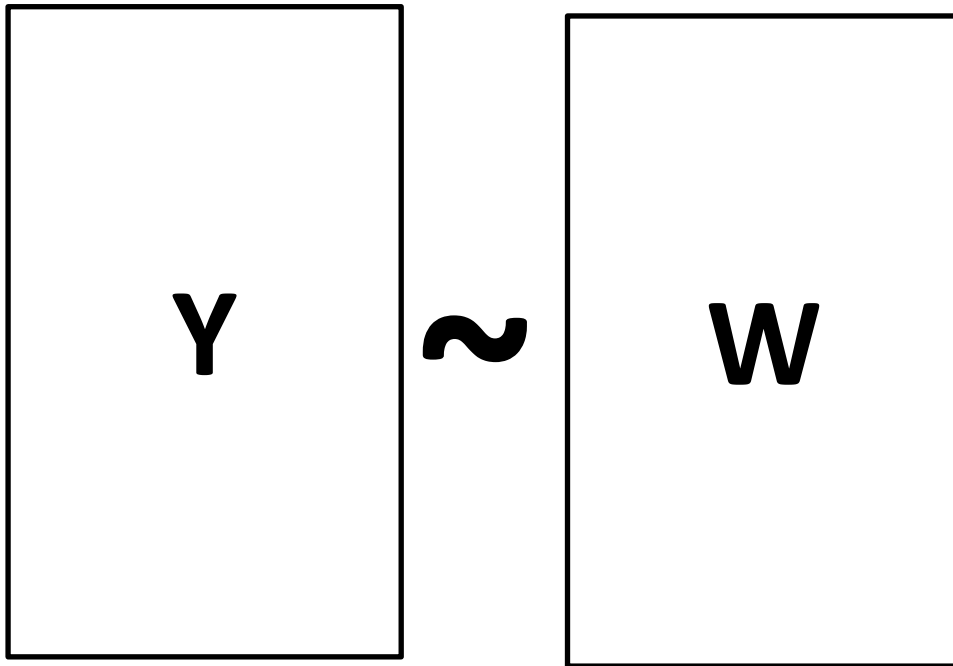
$$R^2_{\mathbf{Y}_{\text{res}|\mathbf{W}}|\mathbf{X}_{\text{res}|\mathbf{W}}} = \frac{SS(\mathbf{Y}_{\text{fit}})}{SS(\mathbf{Y}_{\text{res}|\mathbf{W}})}$$

Passo 1: regressão de $Y \sim X$



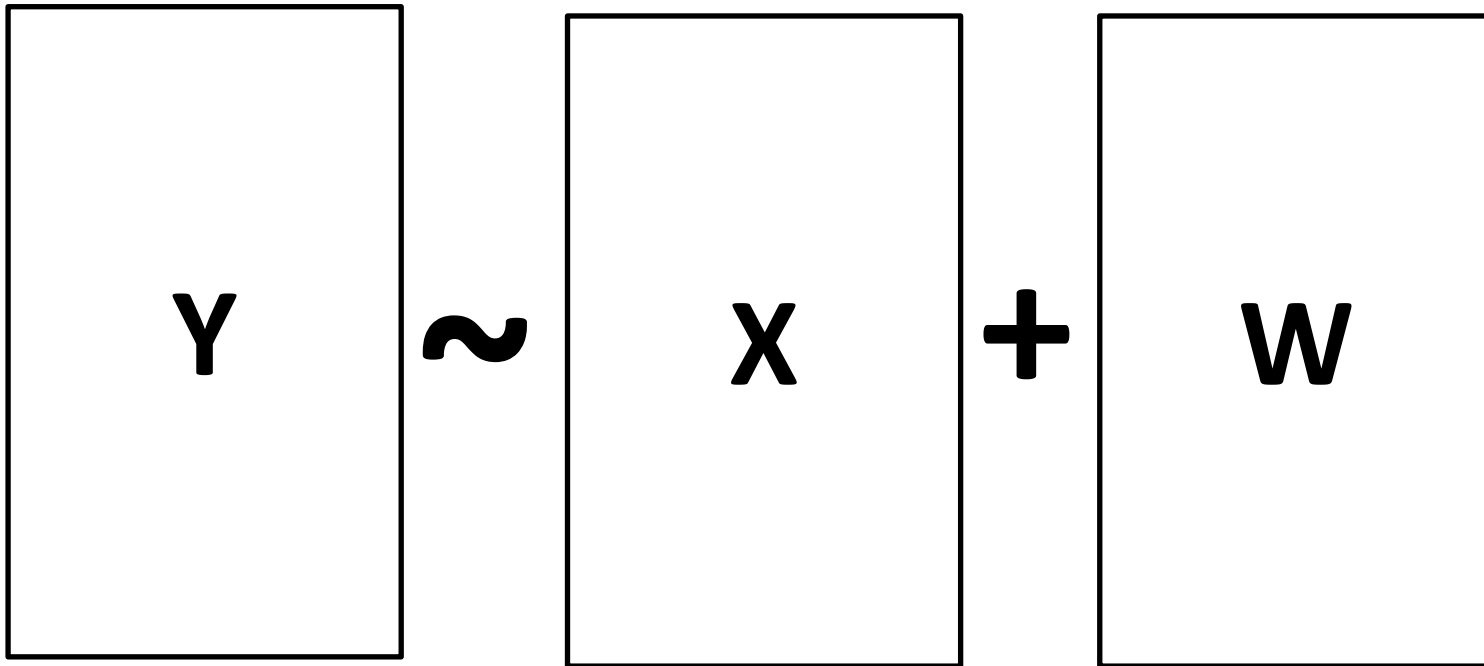
$$R^2_{YX} = a + b$$

Passo 2: regressão de $Y \sim W$




$$R^2_{YW} = b + c$$

Passo 3: regressão de $Y \sim X$



$$R^2_{YXW} = a + b + c$$

Passo 4: valores individuais - a, b, c

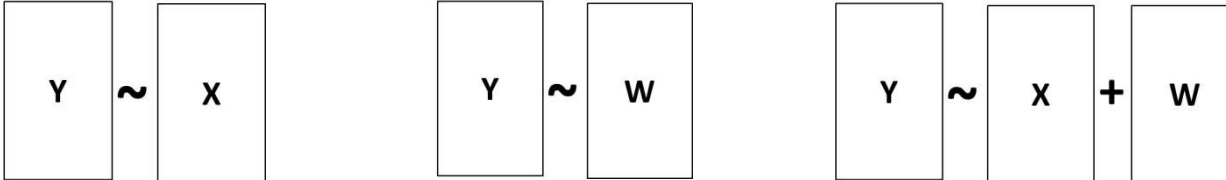
$$[a] = R^2_{YXW} - R^2_{YW}$$


The diagram consists of two parts. The left part shows three boxes: the first contains 'Y', followed by a tilde symbol (~), then a box containing 'X', a plus sign (+), and a box containing 'W'. The right part shows two boxes: the first contains 'Y', followed by a tilde symbol (~), and a box containing 'W'. The red text R^2_{YXW} is positioned above the first part, and R^2_{YW} is positioned above the second part.

Passo 4: valores individuais - a, b, c

$$[a] = R_{YXW}^2 - R_{YW}^2$$


The diagram for [a] consists of two parts. The first part shows a vertical box labeled 'Y' followed by a tilde symbol, then three vertical boxes labeled 'X', '+', and 'W'. The second part shows a vertical box labeled 'Y' followed by a tilde symbol, then a vertical box labeled 'W'. Both parts are rendered in a light gray color.

$$[b] = R_{YX}^2 + R_{YW}^2 - R_{YXW}^2$$


The diagram for [b] consists of three parts. The first part shows a vertical box labeled 'Y' followed by a tilde symbol, then a vertical box labeled 'X'. The second part shows a vertical box labeled 'Y' followed by a tilde symbol, then a vertical box labeled 'W'. The third part shows a vertical box labeled 'Y' followed by a tilde symbol, then three vertical boxes labeled 'X', '+', and 'W'. All parts are rendered in black.

Passo 4: valores individuais - a, b, c

$$[a] = R_{YXW}^2 - R_{YW}^2$$


Diagram illustrating the components of [a]:

- Left term: A box containing $Y \sim X + W$.
- Right term: A box containing $Y \sim W$.

$$[b] = R_{YX}^2 + R_{YW}^2 - R_{YXW}^2$$


Diagram illustrating the components of [b]:

- Left term: A box containing $Y \sim X$.
- Middle term: A box containing $Y \sim W$.
- Right term: A box containing $Y \sim X + W$.


$$[c] = R_{YXW}^2 - R_{YX}^2$$


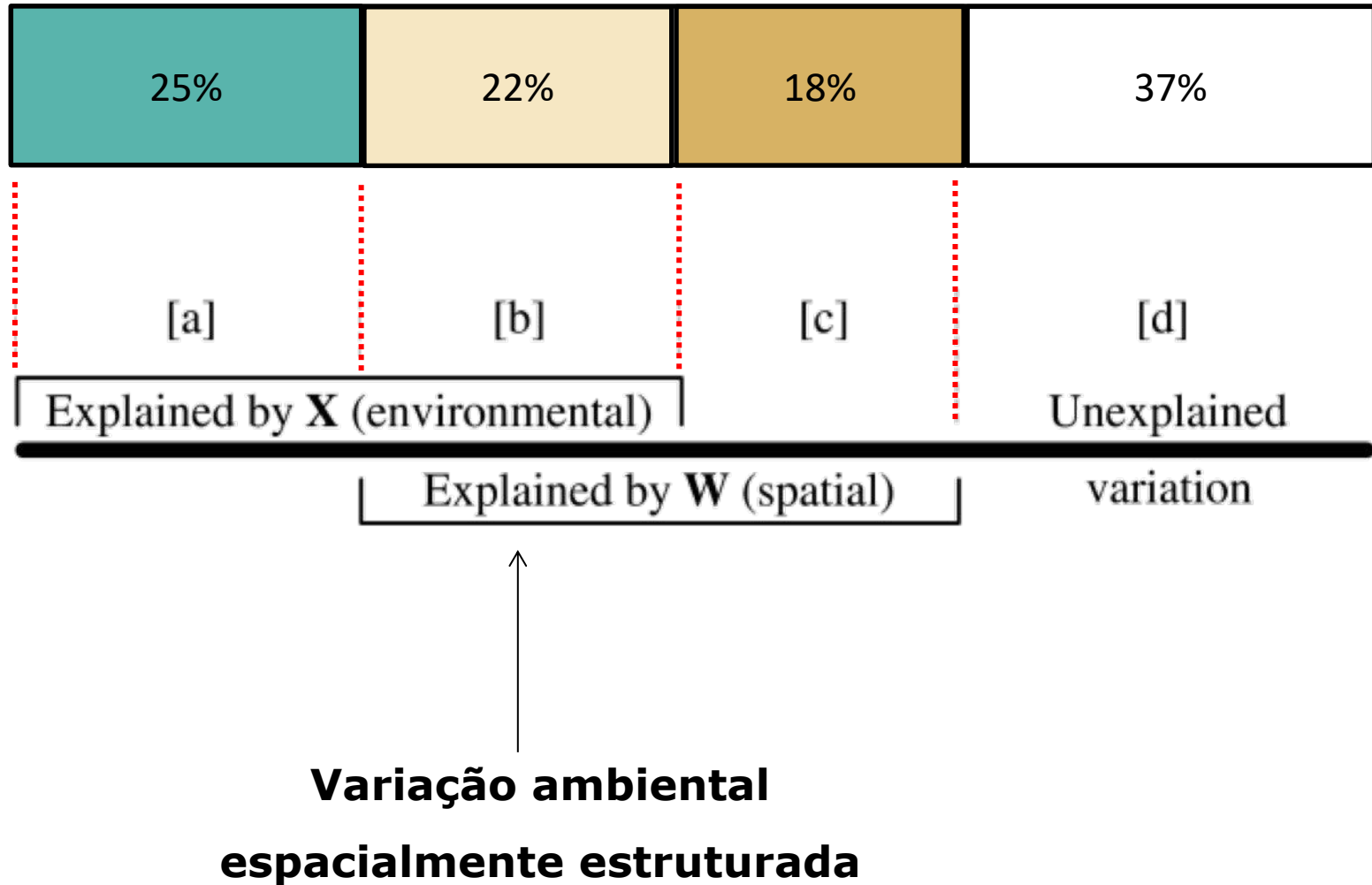
Diagram illustrating the components of [c]:

- Left term: A box containing $Y \sim X + W$.
- Right term: A box containing $Y \sim X$.

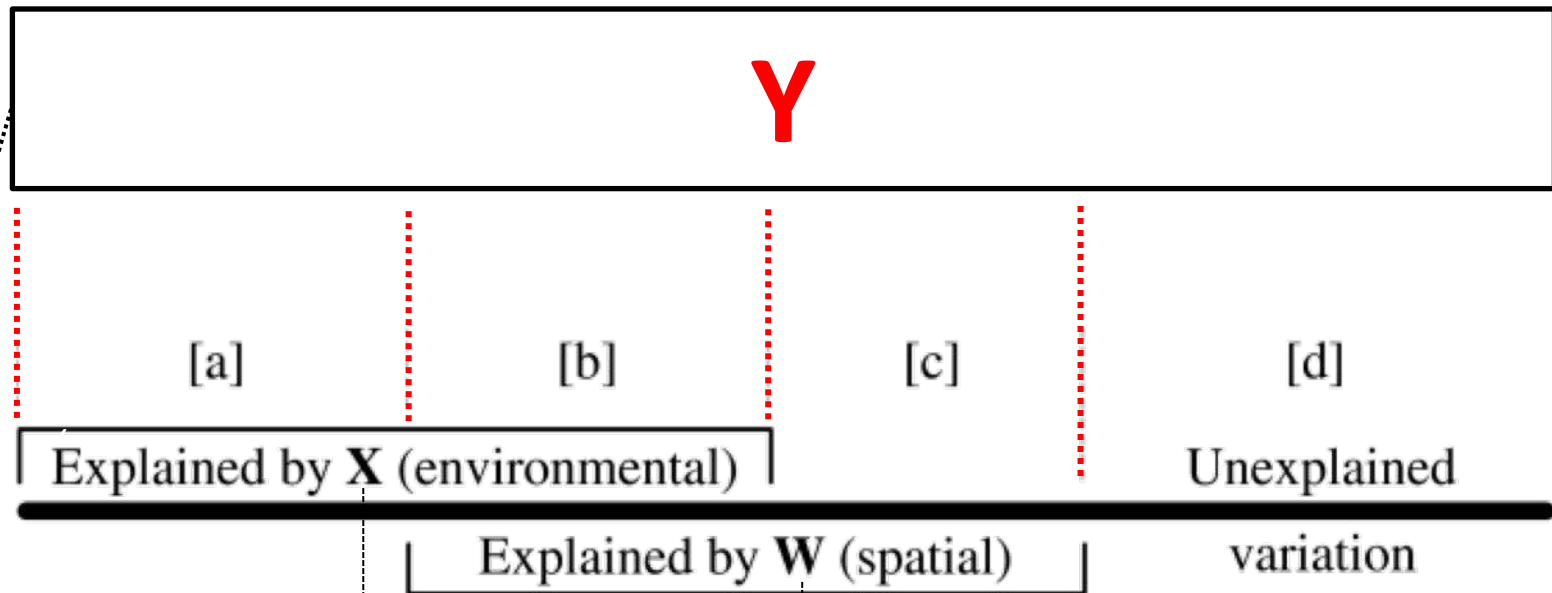
Passo 5: variação residual = d

$$[d] = 1 - ([a]+[b]+[c])$$

Partição de variância



Partição de variância



```
varpart(mite, ~., mite.pcnm, data=mite.env, transfo="hel")
```

Linear Discriminant Analysis (LDA)

Quando usar?

- Testar o pertencimento de objetos a grupos
- Discriminar descritores (e.g., variáveis ambientais) entre grupos de locais para encontrar padrão comum a cada grupo
- Testar se indivíduos pertencem a OTUs ou grupos de espécies
- Tem sido muito usada recentemente em aprendizado de máquina para prever comportamentos (alô Netflix)

Como funciona?

- Sinônimo de Canonical Variate Analysis (CVA)
- Objetos da matriz resposta \mathbf{Y} são divididos em k grupos, descritos por uma variável qualitativa
- O método então procura combinações lineares de variáveis explanatórias da matriz \mathbf{X} que expliquem a classificação em \mathbf{Y} maximizando a dispersão dos centroides dos grupos k
 - Ou seja, tenta separar o máximo possível os objetos
 - *Preserva a distância de Mahalanobis*

Distância de Mahalanobis

Vetor dos valores absolutos das diferenças entre as médias das variáveis nos grupos

$$D_5^2(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{d}_{12} \mathbf{V}^{-1} \mathbf{d}'_{12}$$

Variância conjunta dentro dos dois grupos de objetos

X_1 e x_2 são os grupos de objetos

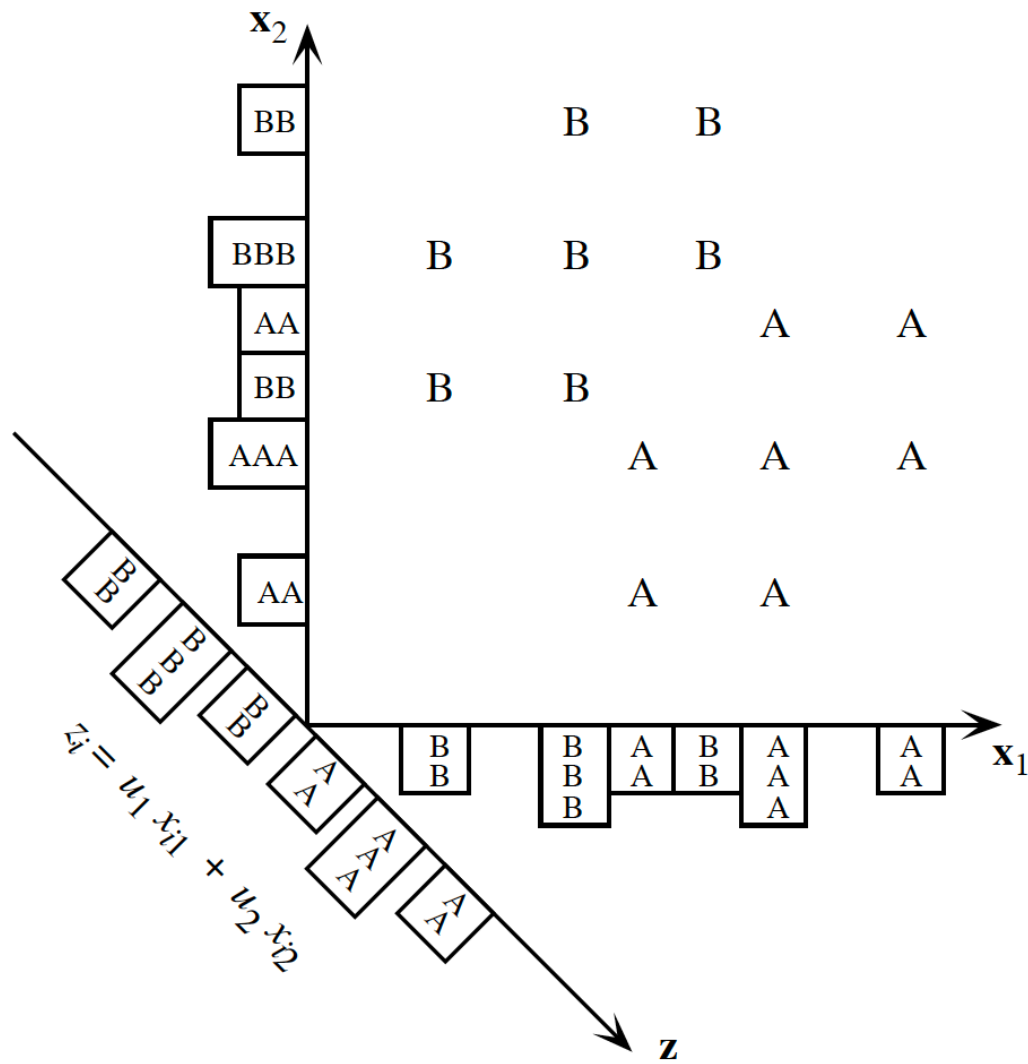


Figure 11.11 Two groups, A and B, with 6 objects each, are overlapping on both axes, x_1 and x_2 , as shown by the histograms on the axes. They are perfectly separated, however, along a discriminant axis z . The position of each object i is calculated along z using the equation $z_i = (\cos 45^\circ) x_{i1} - (\cos 45^\circ) x_{i2}$. Adapted from Jolicoeur (1959).

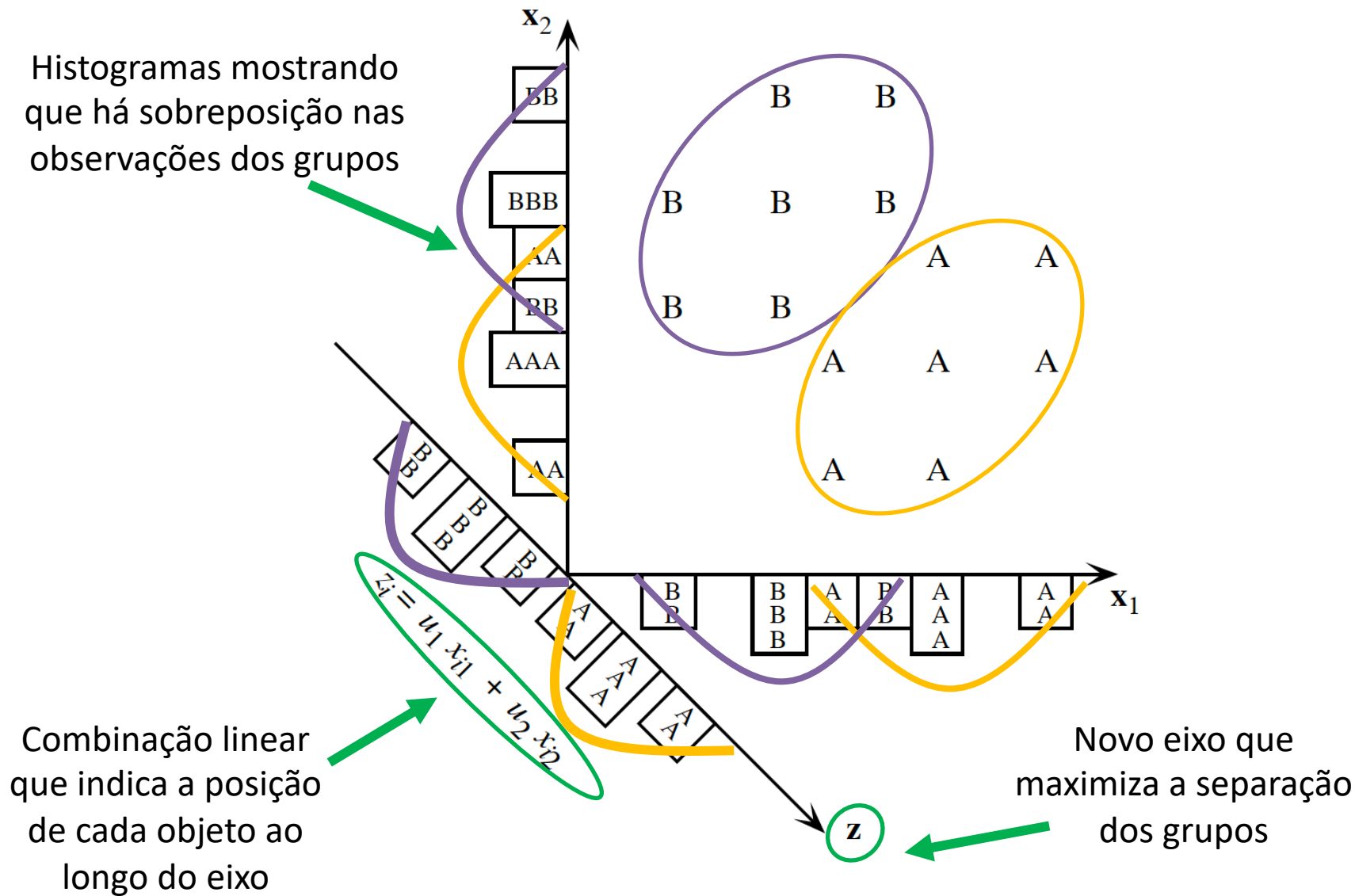


Figure 11.11 Two groups, A and B, with 6 objects each, are overlapping on both axes, x_1 and x_2 , as shown by the histograms on the axes. They are perfectly separated, however, along a discriminant axis z . The position of each object i is calculated along z using the equation $z_i = (\cos 45^\circ) x_{i1} - (\cos 45^\circ) x_{i2}$. Adapted from Jolicoeur (1959).

1. Calcula a separação entre diferentes grupos
(=estimativa da variância entre grupos)



2. Calcula variância dentro dos grupos
(=distância entre cada observação e o centroide)



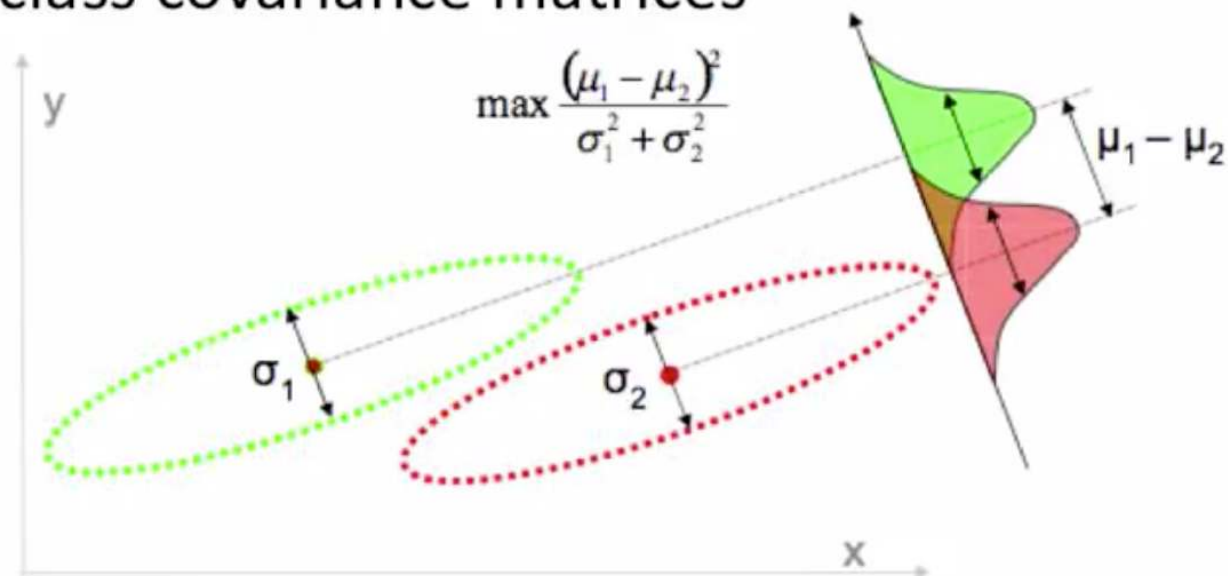
3. Constrói um espaço reduzido (canônico) que maximiza o Passo 1 (variância entre grupos) e minimiza Passo 2 (variância dentro de grupos).



4. Encontrar a posição dos objetos (scores) no espaço reduzido definido pelos eixos.
O LDA1 indica a direção de maior variação entre centroides dos grupos

Linear Discriminant Analysis

- LDA: pick a new dimension that gives:
 - maximum separation between means of projected classes
 - minimum variance within each projected class
- Solution: eigenvectors based on between-class and within-class covariance matrices



Como funciona?

- Estatísticas da LDA
- Lambda de Wilks' => testa se a posição dos centroides dos grupos diferem significativamente
 - Varia de 0 (máxima dispersão) a 1 (ausência de dispersão entre grupos)

Podemos usar LDA para classificar novos objetos

- Discriminant (identification) functions
 - Usadas para posicionar novos objetos no espaço reduzido
 - Calcula score, objeto é atribuído ao grupo com maior score
 - Mesmo raciocínio da regressão linear (múltipla)
- Classification table
 - Compara a classificação original dos grupos com a atribuição feita pela função de classificação
 - Podemos testar a efetividade da classificação calculando a distância de Mahalanobis entre grupos
 - Distâncias > 3 indicam boa separação

Observed group	Assigned to group			Total and % correct
	1	2	3	
1	3	0	0	3 (100%)
2	0	2	0	2 (100%)
3	0	0	2	2 (100%)
Total	3	2	2	7 (100%)

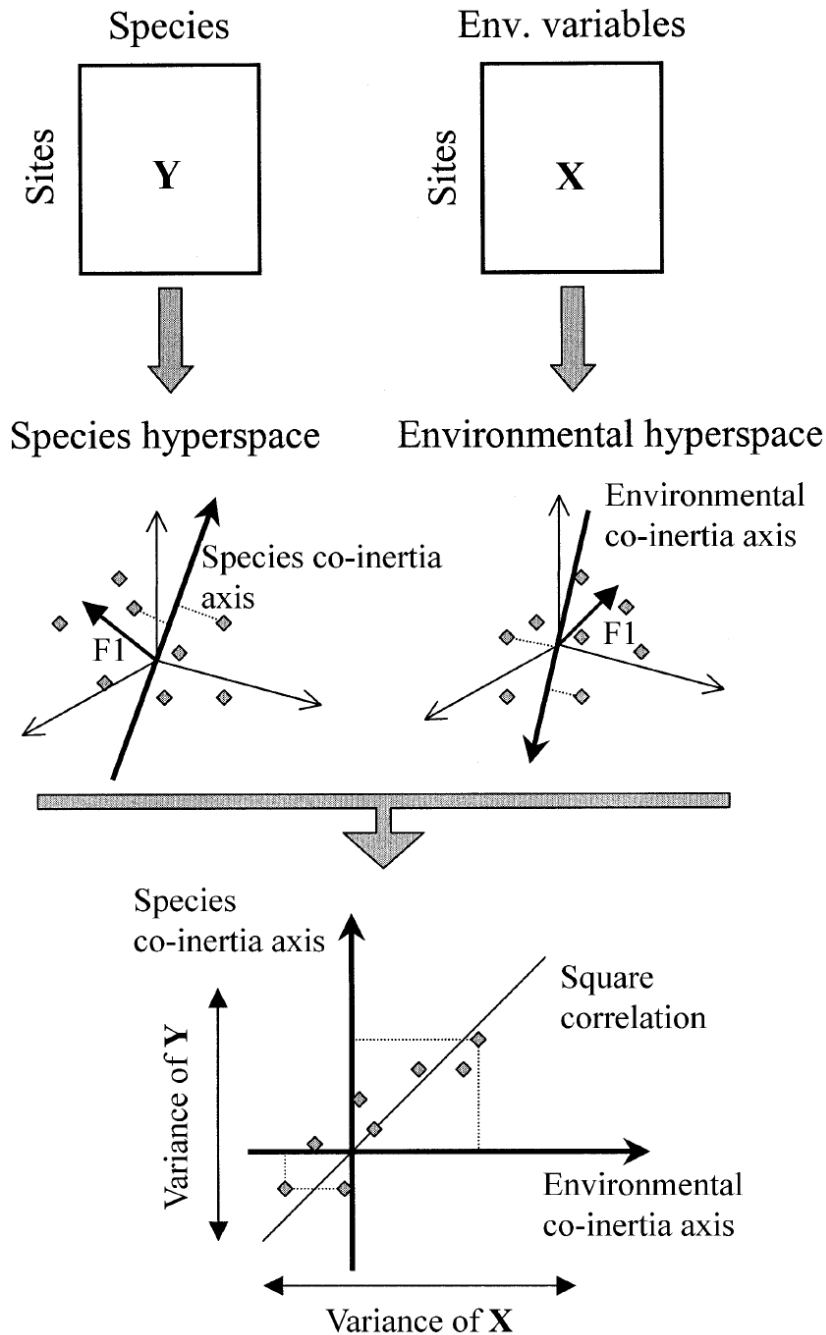
Pressupostos

- Distribuição de frequência dentro dos grupos segue distribuição normal (Gaussiana)
- Homogeneidade de variância dentro de grupos
 - `vegan::betadisper`

Rápido comentário sobre métodos alternativos

Análise de *Correlação* Canônica (CCorA)

- Requer número pequeno de espécies e locais
- *O número de variáveis ambientais precisa ser menor do que o número de locais*
- Como esses critérios raramente são encontrados em estudos ecológicos, esta análise não é muito usada

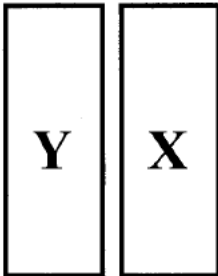
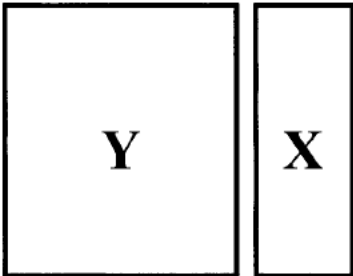
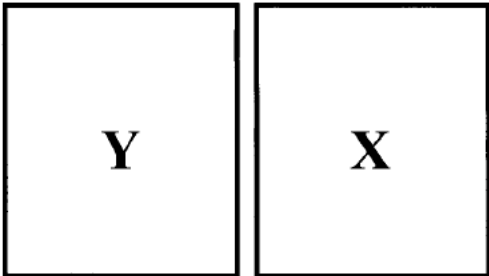


CO-INERTIA ANALYSIS AND THE LINKING OF ECOLOGICAL DATA TABLES

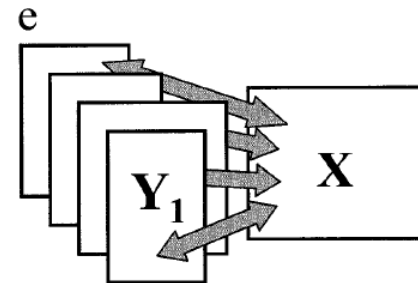
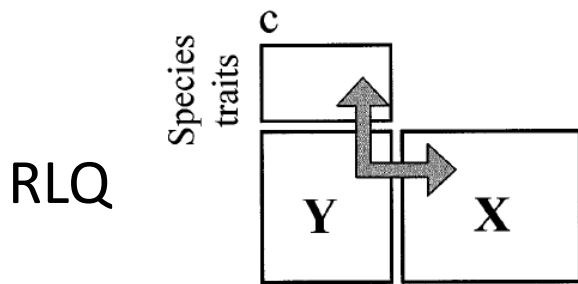
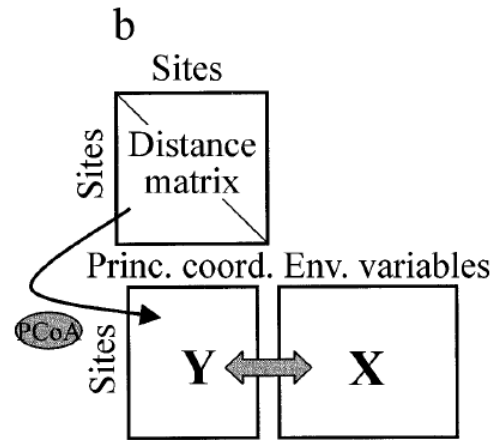
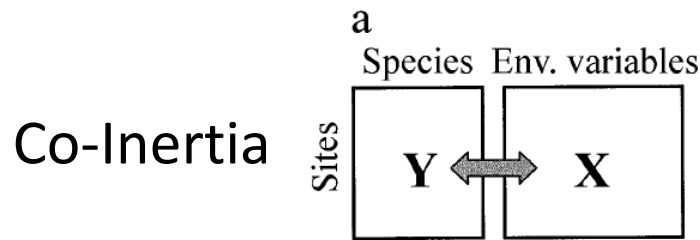
STÉPHANE DRAY,¹ DANIEL CHIESSEL, AND JEAN THIOULOUSE

- Mede a **concordância** entre duas matrizes
- Maximiza as combinações lineares por Mínimos Quadrados Parciais (Partial Least Squares), diferentemente da RDA que usa Mínimos Quadrados
- CCA analisa a matriz de espécies por CA, com locais sendo pesados pela sua riqueza, já ColAI permite que qualquer método seja usado
- Funciona bem quando as variáveis ambientais são altamente correlacionadas
- Permite que se use variáveis fuzzy e qualitativas, além das quantitativas
- Permite esquemas de peso (weighting) para locais e espécies

Diferença entre CCA/RDA e Co-Inertia

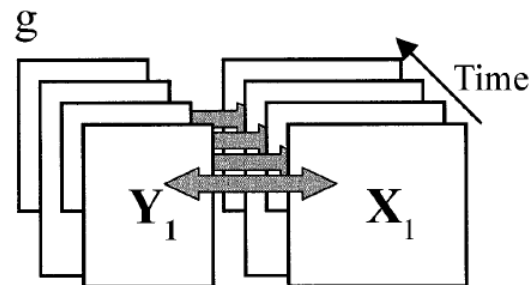
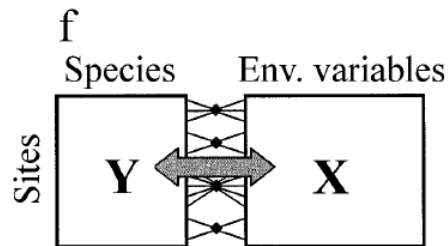
Maximization of $\text{cov}^2(\mathbf{XQ}\mathbf{u}_i, \mathbf{YR}\mathbf{v}_i) = \text{corr}^2(\mathbf{XQ}\mathbf{u}_i, \mathbf{YR}\mathbf{v}_i) \times \text{var}(\mathbf{XQ}\mathbf{u}_i) \times \text{var}(\mathbf{YR}\mathbf{v}_i)$		
with the constraints		
$\text{var}(\mathbf{XQ}\mathbf{u}_i) = 1$ $\text{var}(\mathbf{YR}\mathbf{v}_i) = 1$	$\text{var}(\mathbf{XQ}\mathbf{u}_i) = 1$	No constraint
Canonical correlation analysis	Analysis with respect to instrumental variables (CCA, RDA)	Co-inertia analysis
		

Análises derivadas da Co-Inertia



Co-Inertia
Múltipla

Análise de
concordância

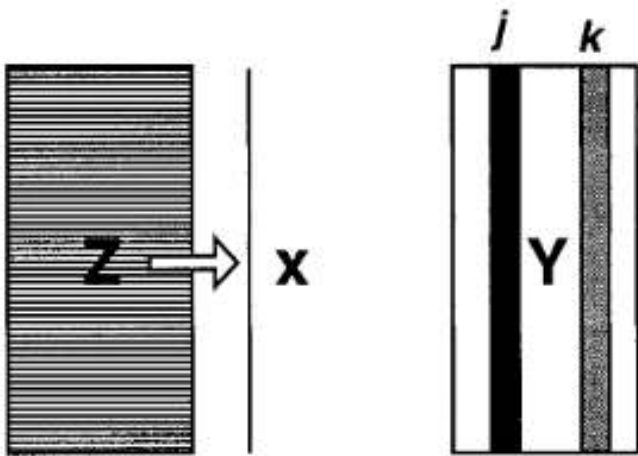


Co-Inertia
Within &
between class

STATICO

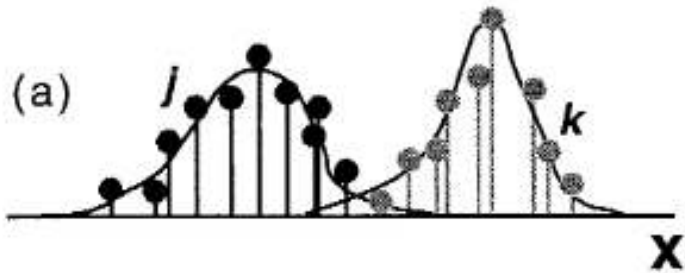
Outlying Mean Index

- Criado por Dolédec et al. (2000): ordenação restrita que permite calcular largura e sobreposição de nicho num espaço multidimensional
 - Permite calcular se as variáveis ambientais incluídas no modelo representam satisfatoriamente o nicho das espécies
- “mistura de CCA com RDA”
 - Permite testar relações espécie-ambiente tanto lineares quanto unimodais
- Muito útil para investigar nicho climático
 - E especialmente evolução de nicho climático



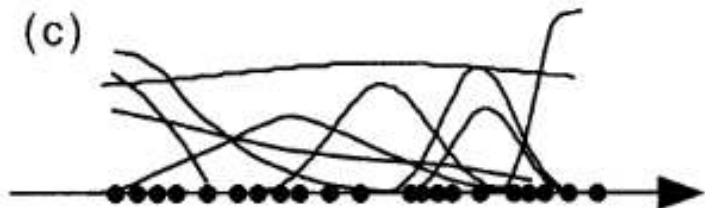
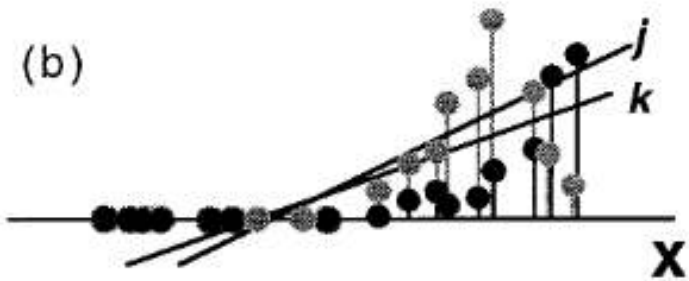
Inercia Total = Sobreposição de nicho

$$\text{Tr}(\mathbf{X}^T \mathbf{D}_t \mathbf{X} \mathbf{I}_p) = \sum_{j=1}^I f_{\cdot j} m_a(j).$$



Tolerância = Largura de nicho

$$T_m(j) = \sum_{i=1}^n f_{ij} \|G_j - m_i\|_{I_p}^2.$$



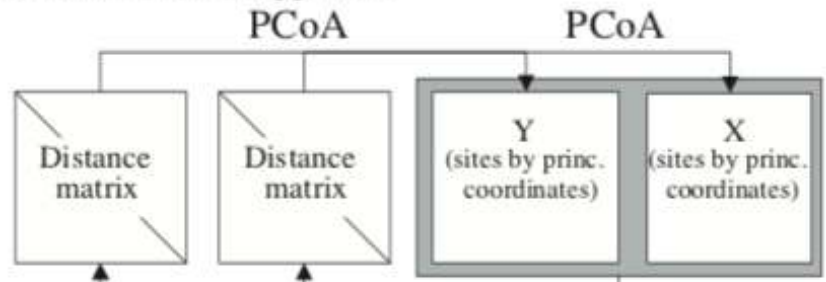
Relações lineares e unimodais

Veja também
Karasiewicz et
al. 2017

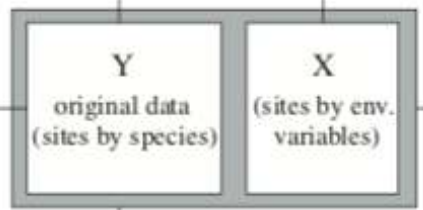
Procrustean co-inertia analysis for the linking of multivariate datasets¹

Stéphane DRAY², Daniel CHESSEL & Jean THIOULOUSE, UMR CI

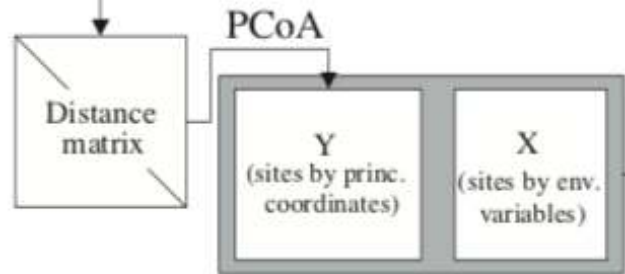
a) Double distance-based approach



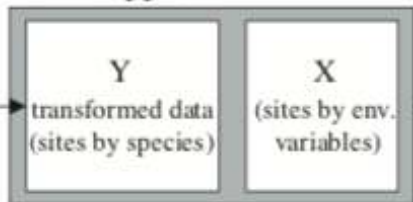
b) Classical approach



c) Distance-based approach



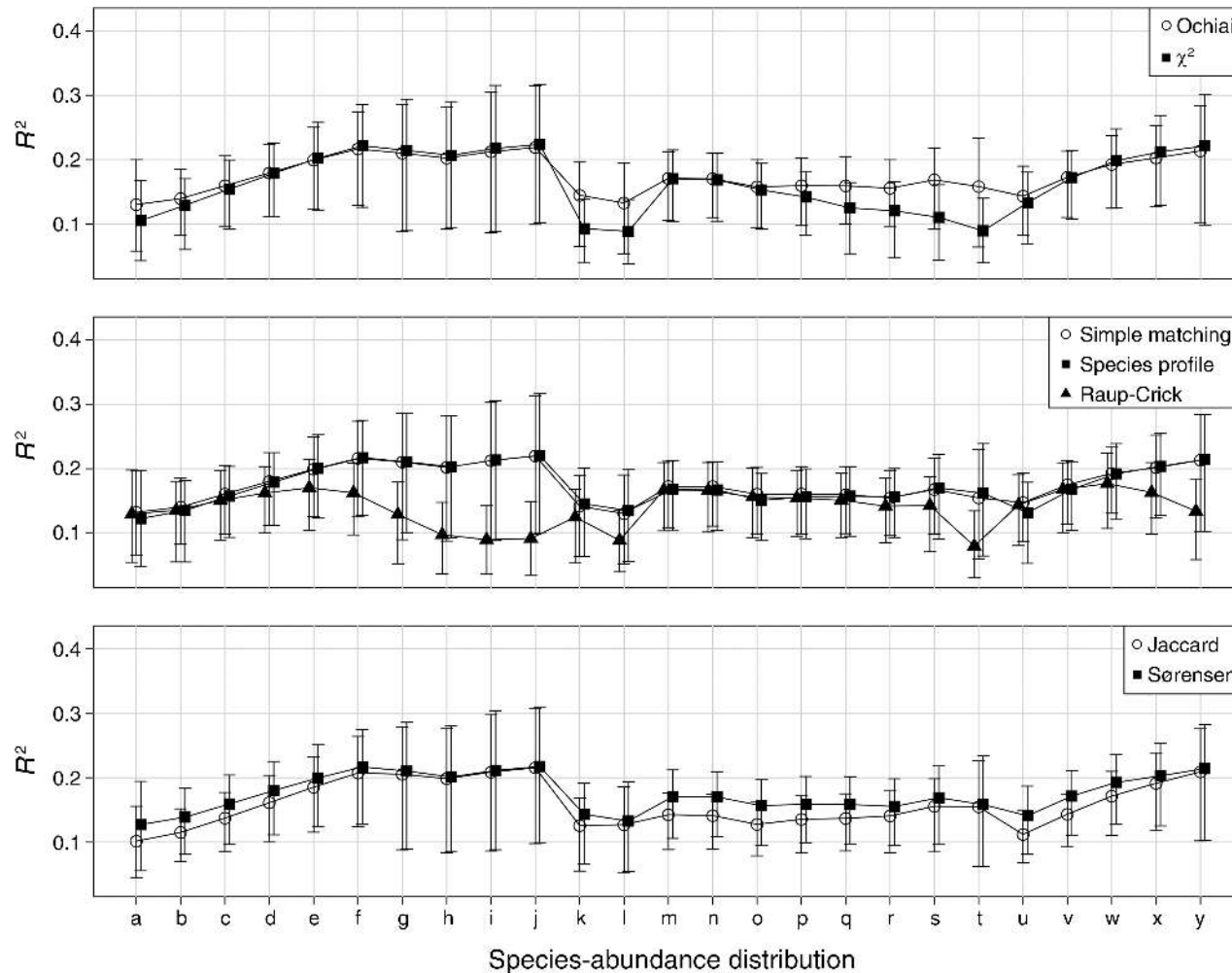
d) Transformed data approach



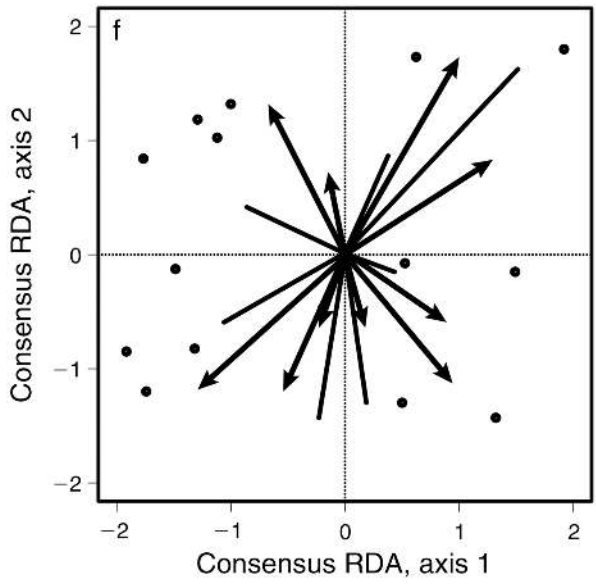
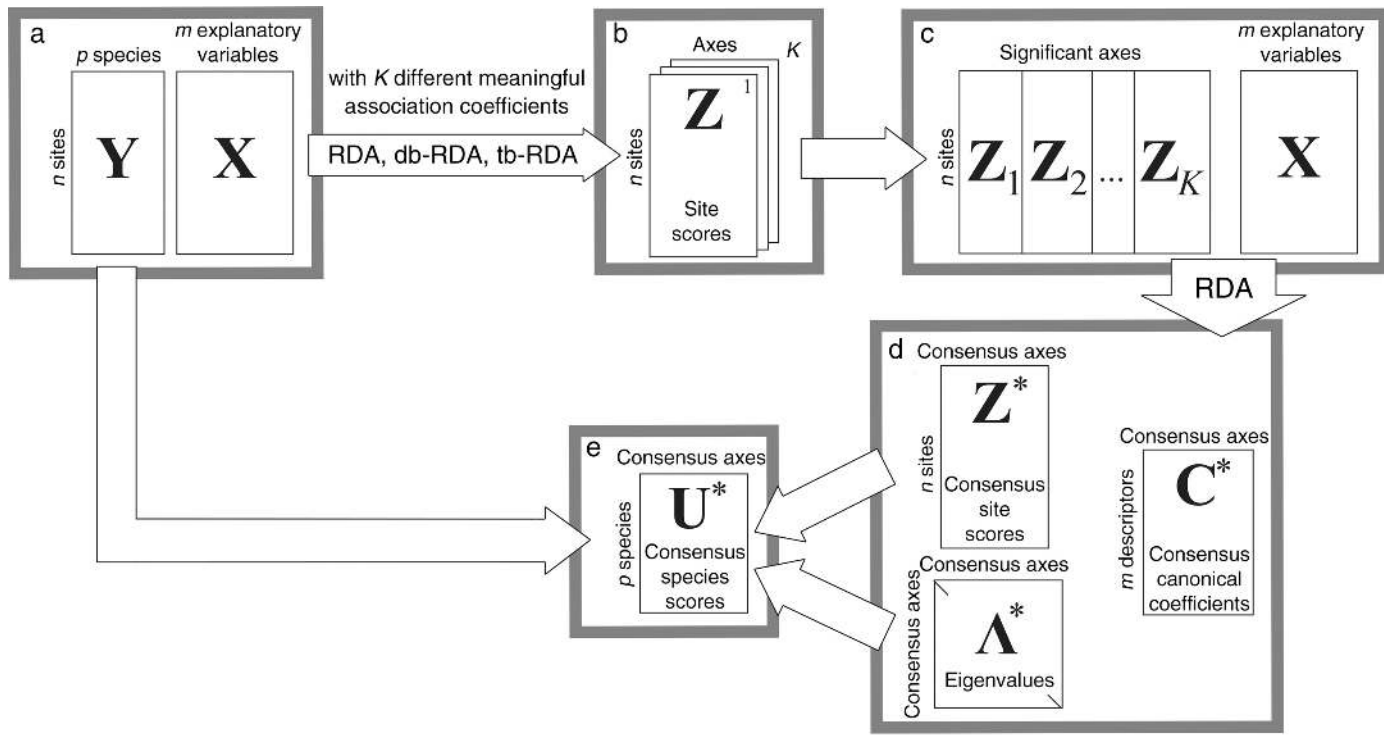
Procrustean co-inertia analysis

Consensus RDA across dissimilarity coefficients for canonical ordination of community composition data

F. GUILLAUME BLANCHET,^{1,2,3,5} PIERRE LEGENDRE,⁴ J. A. COLIN BERGERON,³ AND FANGLIANG HE³



Comparação do R^2 da ordenação feita com vários índices de similaridade para abundância

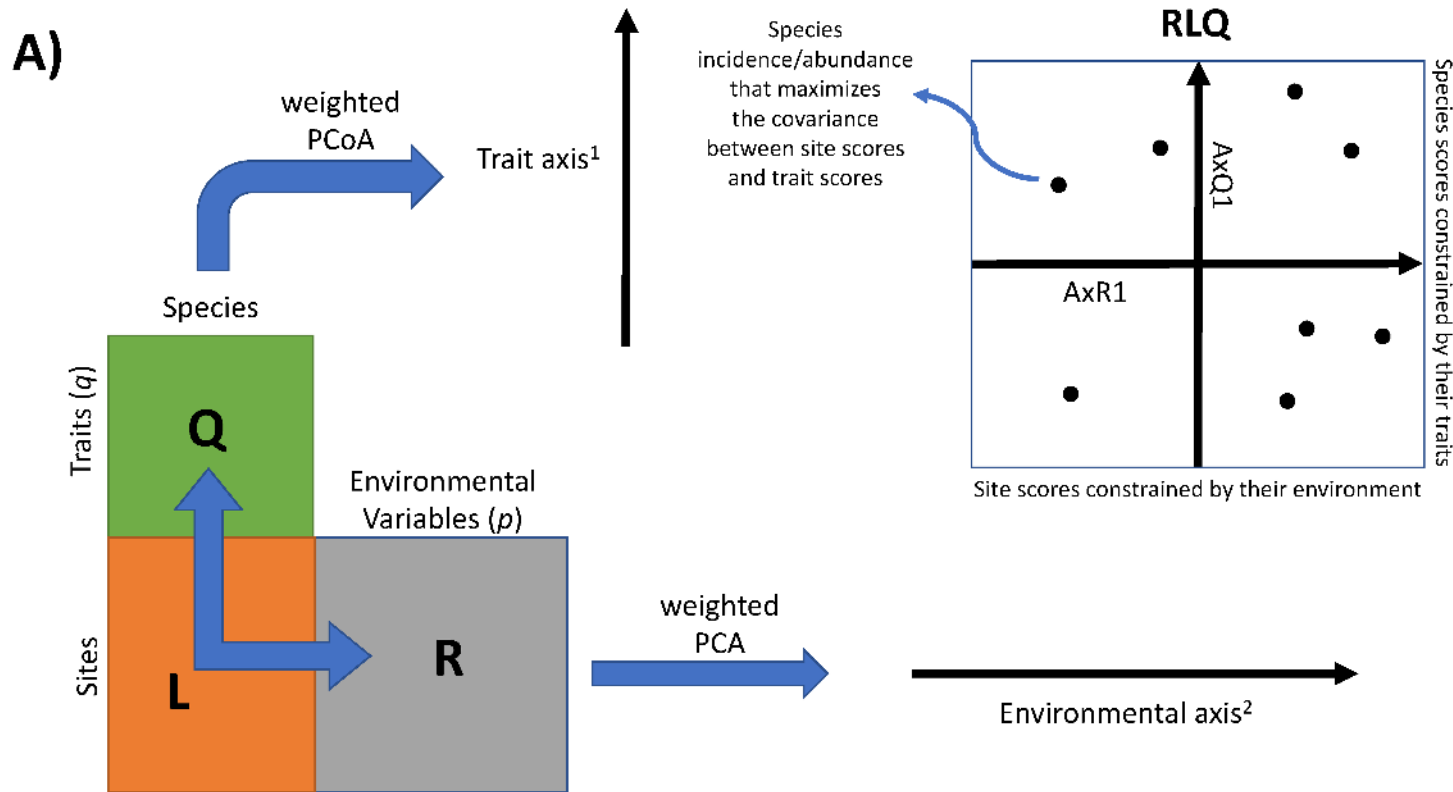


Análise para ligar atributos à variáveis ambientais

(veja também Pavoine et al. 2009, Dray et al. 2014, Peres-Neto et al. 2017, ter Braak et al. 2018a,b)

Matching species traits to environmental variables: a new three-table ordination method

S. DOLÉDEC¹, D. CHESSEL¹, C.J.F. TER BRAAK² and S. CHAMPELY³



1 - Weighted by species abundance/incidence derived from Correspondence Analysis
2 - Weighted by sites derived from Correspondence Analysis

Meu deus d
Não entend
uma so
Eu quer
Quem poder

MIGA SUA LOCA



SIACALME