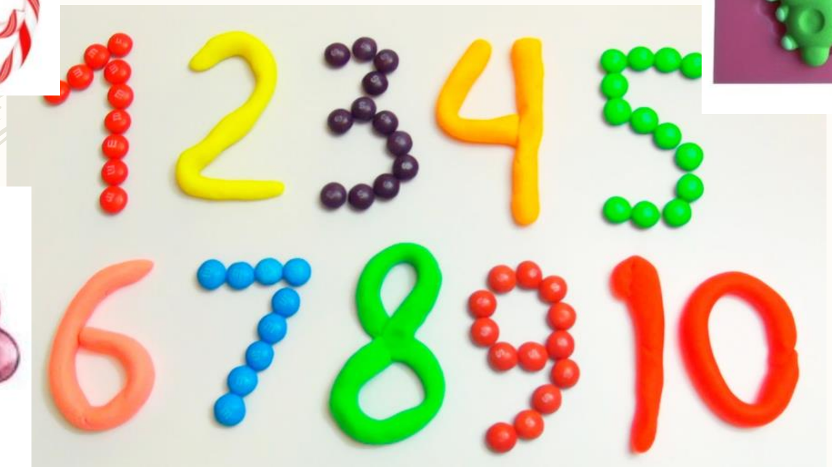
Aula 25 Goodies*



* Goodies related to animals, plants and numbers…

A nice resource:

http://library.open.oregonstate.edu/monitoring/chapter/data-analysis-in-monitoring/

Monitoring Animal Populations and Their Habitats: A Practitioner's Guide

Brenda McComb, Benjamin Zuckerberg, David Vesely, Christopher Jordan

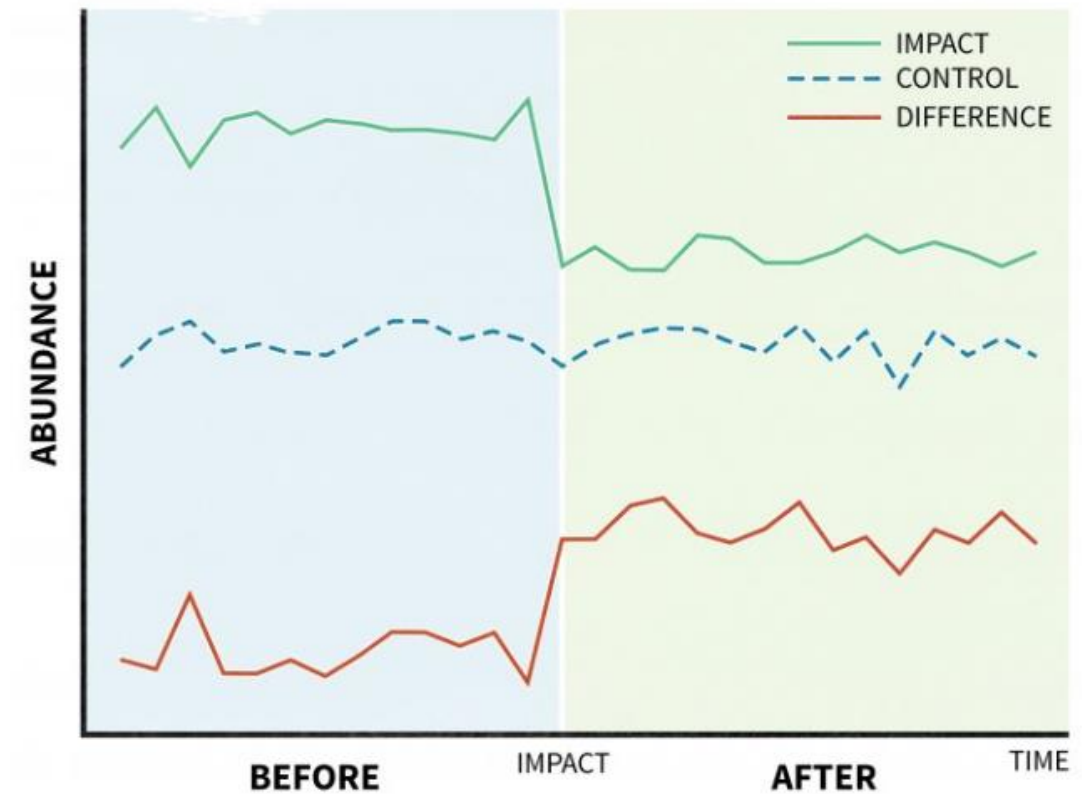
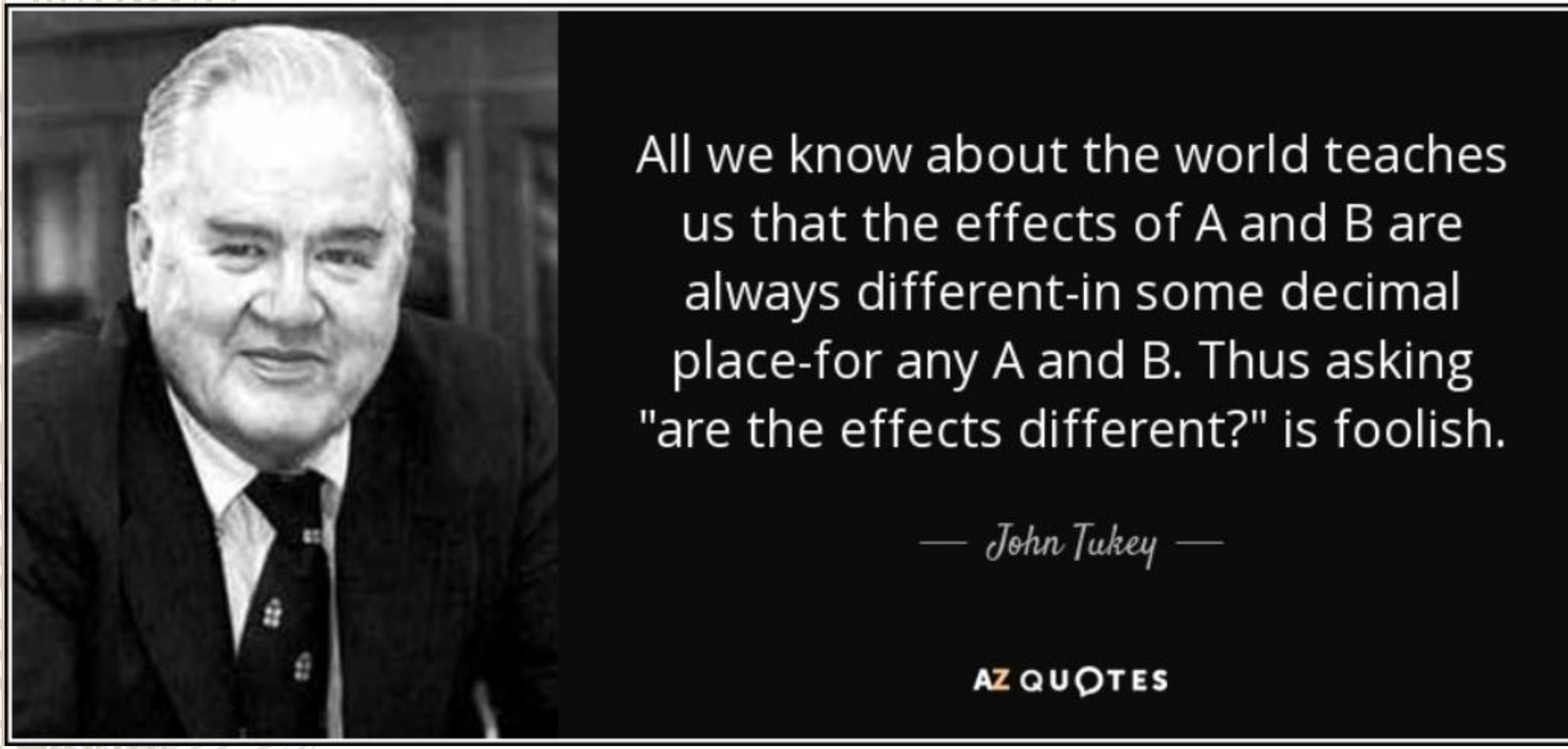


Figure 11.5. A hypothetical example of a BACI analysis where abundance samples are taken at control impact sites before and after the impact and compared to a control site (redrafted from Stewart-Oaten et al. 1986).

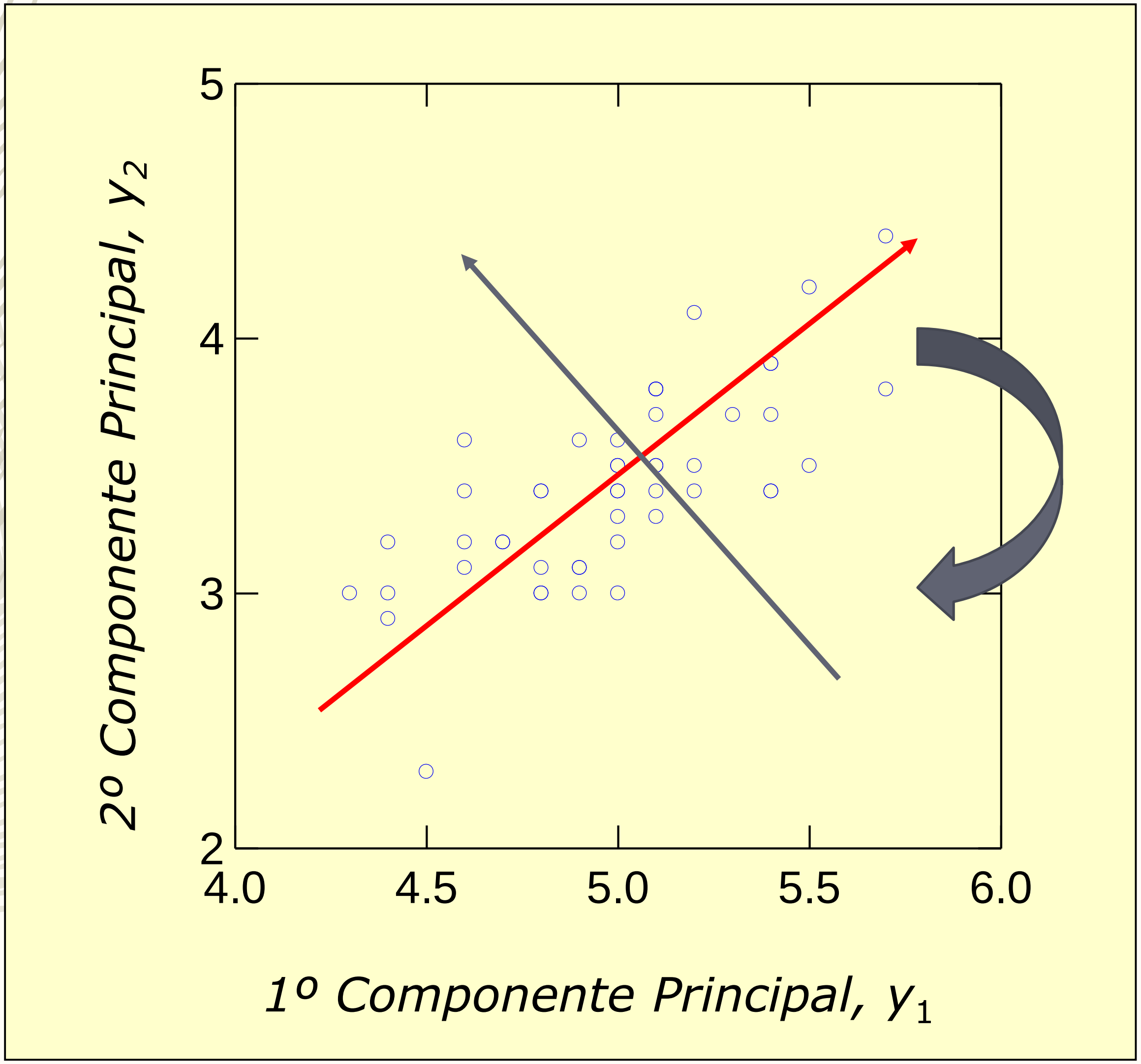
Aqui no fénix!

Ecologia Numérica - Aula Teórica 25 – 10-12-2018



https://www.azquotes.com/quote/1399812

Análise de Componentes Principais



ordenação

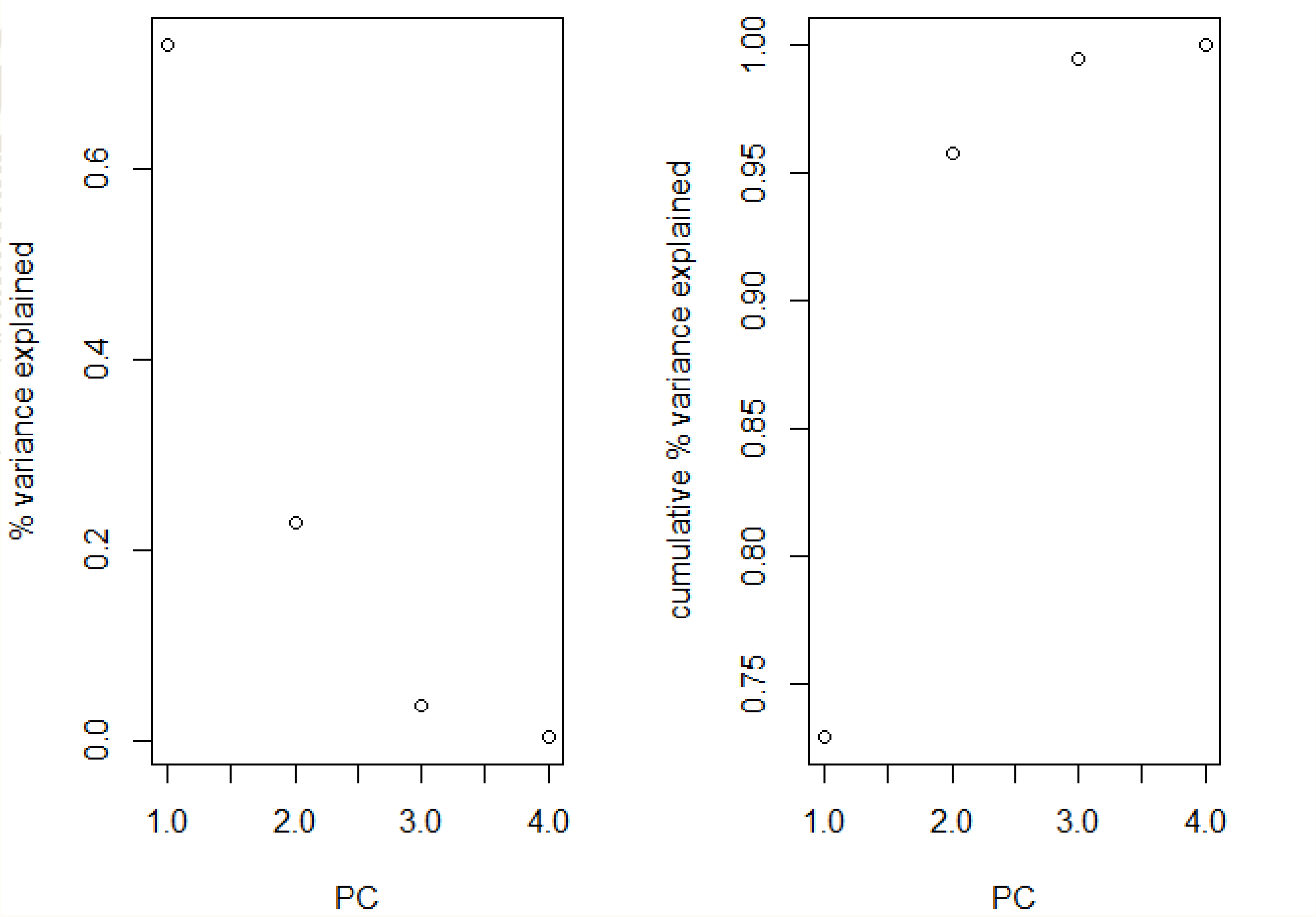
Análise de Componentes Principais

Então, sendo

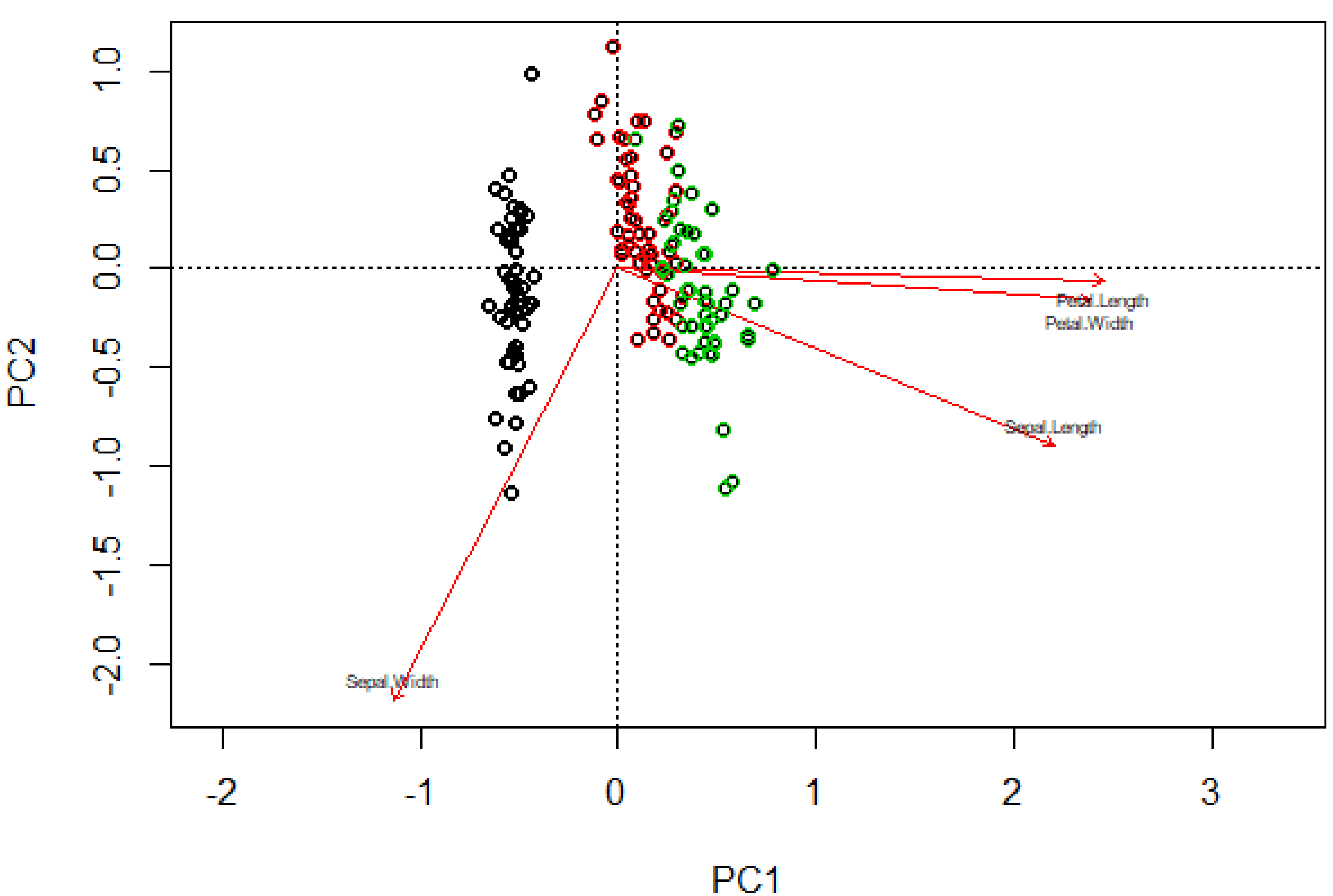
$y_1 = a_{11}x_1 + a_{12}x_2 + ... + a_{1k}x_k$
$y_2 = a_{21}x_1 + a_{22}x_2 + ... + a_{2k}x_k$
...
$y_k = a_{k1}x_1 + a_{k2}x_2 + ... + a_{kk}x_k$

xj's são padronizados (redução e centragem) se for utilizada a matriz de correlação (no R, `scale(dados)`)

```
#another example, now with data iris
data(iris)
myPCAiris=rda(scale(iris[,1:4]))
par(mfrow=c(1,2),mar=c(4,4,0.5,2.5))
eigvali=myPCAiris$CA$eig
plot(1:4,eigvali/sum(eigvali),ylab="% variance explained",xlab="PC")
plot(1:4,cumsum(eigvali/sum(eigvali)),
ylab="cumulative % variance explained",xlab="PC")
```

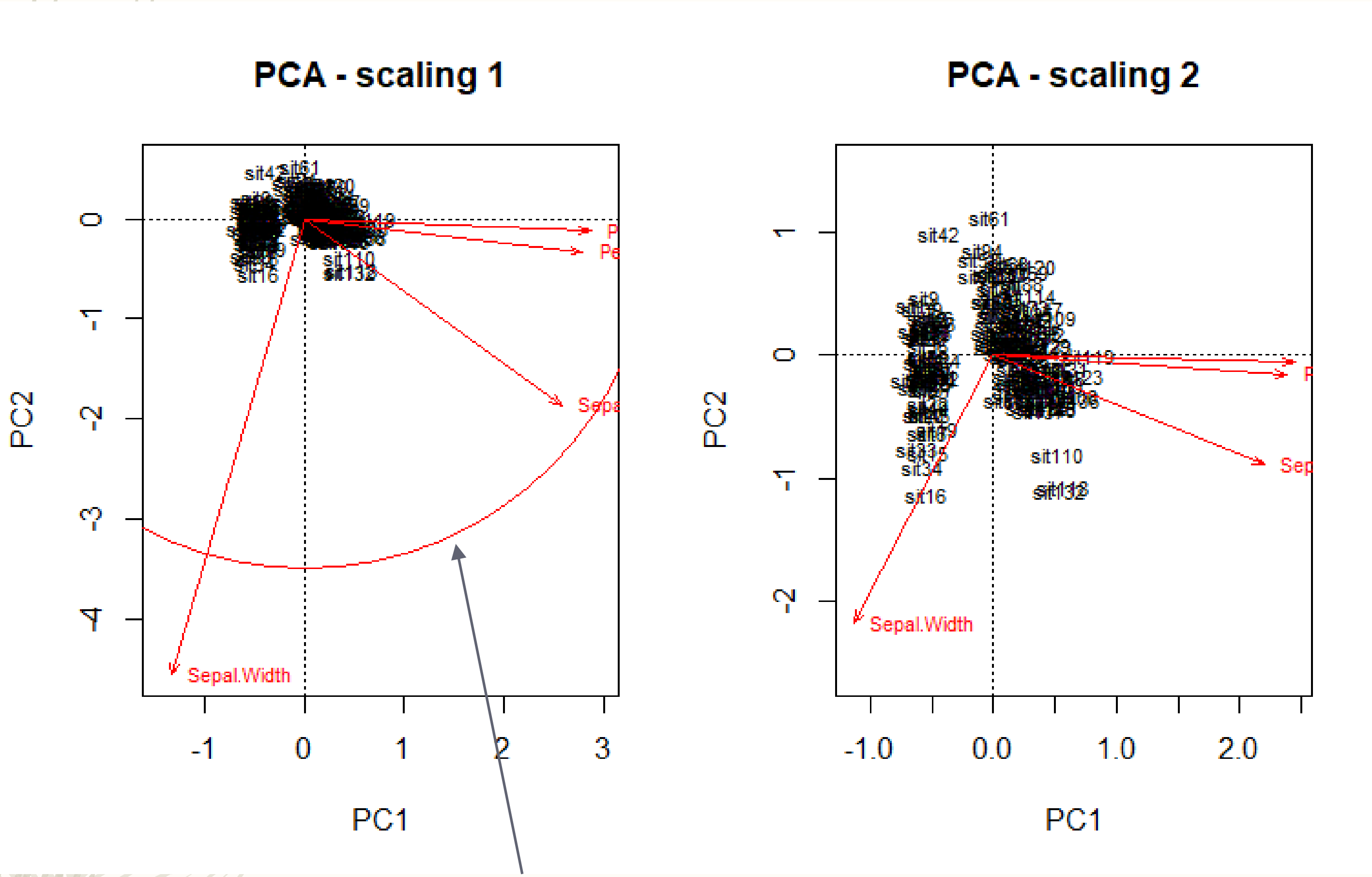


```
par(mfrow=c(1,1),mar=c(4,4,2.5,0.5))
biplot(myPCAiris)
#add colors to points
points(summary(myPCAiris)$sites[,1:2],col=iris[,5])
#for some reason vars names not appearing
text(summary(myPCAiris)$species[,1],
summary(myPCAiris)$species[,2]+c(0.1,0.1,-0.1,-0.1),
names(iris[,1:4]),cex=0.5)
```



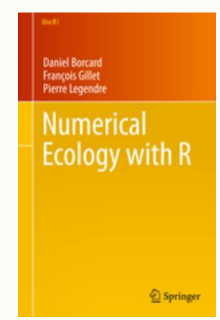
Function in file `brocardfunctions.R`

```
cleanplot.pca(myPCAiris)
```



First, the *scaling 1 biplot* displays a feature that must be explained. The circle is called a *circle of equilibrium contribution*. Its radius is equal to $\sqrt{d/p}$, where d is the number of axes represented in the biplot (usually $d=2$) and p is the number of dimensions of the PCA space (i.e. usually the number of variables of the data matrix).[2] The radius of this circle represents the length of the vector representing a variable that would contribute equally to all the dimensions of the PCA space. Therefore, for any given pair of axes, the variables that have vectors longer than this radius make a higher contribution than average and can be interpreted with confidence.

`Sepal.Width` seems the most important variable in this case



Aspectos importantes a considerar numa PCA

O j-ésimo componente principal é o j-ésimo vector próprio da matriz de covariância/correlação;

Os coeficientes, a_{jk}, são elementos dos vectores próprios que se relacionam com as variáveis originais (padronizados se se utilizar a matriz de correlação) com os componentes principais;

As coordenadas (scores) são a referenciação das observações sobre o sistema de eixos dado pelos componentes principais;

Aspectos importantes a considerar numa PCA

A quatidade de variância explicada por cada componente principal é dada pelo respectivo valor próprio, λ_j

A proporção da variância total explicada por cada componente principal é dada por $\lambda_j / \Sigma \lambda_j$

O coeficiente da k-ésima variável original sobre o j-ésimo componente principal é dado por $a_{jk}\lambda_j$ – corresponde à correlação entre a variável original e o componente principal.

Como analisar os resultados de uma PCA

Valores próprios (eigenvalues);

```
> summary(myPCA)

Call:
rda(X = specbyloc)

Partitioning of variance:
              Inertia Proportion
Total          0.6105          1
Unconstrained  0.6105          1

Eigenvalues, and their contribution to the variance

Importance of components:
                         PC1    PC2     PC3     PC4     PC5
Eigenvalue            0.3321 0.1322 0.07077 0.04450 0.03090
Proportion Explained  0.5440 0.2166 0.11593 0.07289 0.05062
Cumulative Proportion 0.5440 0.7605 0.87648 0.94938 1.00000
```

Como analisar os resultados de uma PCA

Coordenadas das observações e das variáveis nos componentes principais (factor scores).

Site scores (weighted sums of species scores)

	PC1	PC2	PC3	PC4	PC5
sit1	-0.19907	-0.03120	0.6650329	0.32418	-0.47610
sit2	-0.33919	-0.20040	0.7132389	-0.07777	0.32573
sit3	-0.23530	-0.18085	0.4698606	0.19166	-0.21479
sit4	-0.33530	0.04034	0.5953796	0.34422	-0.31426
sit5	0.01362	-0.43530	-0.2011587	0.17201	-0.61423
sit6	-0.75018	0.77007	-0.1595154	-0.24311	0.47097
sit7	0.37983	-0.03070	-0.2370972	1.06487	0.73899
sit8	0.64916	0.36089	0.0623928	0.14717	0.38908
sit9	-0.56018	0.43185	0.2478845	0.05844	0.09041
sit10	-0.01358	-0.93701	0.0836581	-0.75562	0.81353
sit11	-0.51997	0.44798	-0.4557238	-0.22207	0.17623
sit12	-0.01359	-0.71406	-0.0001287	-0.35936	0.22691
sit13	0.17008	-0.26355	-0.1532435	0.26574	-0.20534
sit14	-0.15991	-0.08141	-0.6924916	-0.09650	-0.32679
sit15	0.11658	-0.29127	-0.1918964	0.30923	-0.43401
sit16	-0.19225	-0.02316	-0.7794093	-0.15294	-0.25904
sit17	0.04268	-0.05440	-0.4719042	0.13008	-0.13305
sit18	0.65536	0.44280	0.2760325	-0.87884	-0.54287
sit19	0.63908	0.38554	0.1639703	-0.36514	-0.10109
sit20	0.65213	0.36385	0.0651186	0.14374	0.38970

Species scores

	PC1	PC2	PC3	PC4	PC5
spA	0.6864	0.3982	0.165674	-0.32131	-0.08841
spB	-0.0551	-0.4192	0.125099	-0.26016	0.27789
spC	-0.6222	0.4580	-0.359557	-0.15194	0.12191
spD	-0.6879	0.3259	0.471583	0.07006	0.05505
spE	0.7197	0.2956	-0.008506	0.22214	0.26359

Como analisar os resultados de uma PCA

Correlações entre os CP's e as variáveis originais (factor loadings);

```
> loadings(myPCA2)

Loadings:
     Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
spA   0.504  0.464  0.264  0.645  0.213
spB         -0.488  0.199  0.522 -0.669
spC  -0.457  0.533 -0.572  0.305 -0.294
spD  -0.505  0.379  0.750 -0.141 -0.133
spE   0.529  0.344        -0.446 -0.635
```

Análise de Componentes Principais

Matriz de covariância:

- Variáveis têm que ser medidas nas mesmas unidades;
- Dá ênfase às variáveis com maior variância.

Análise de Componentes Principais

Matriz de correlação:

- Variáveis padronizadas (média=0 e d.p.=1);
- Variáveis podem ter unidades diferentes;
- Todas as variáveis têm contribuição idêntica na análise;
- Valor próprio médio=1.

Potenciais problemas

- Falta de independência

  Não constitui um problema

- Normalidade

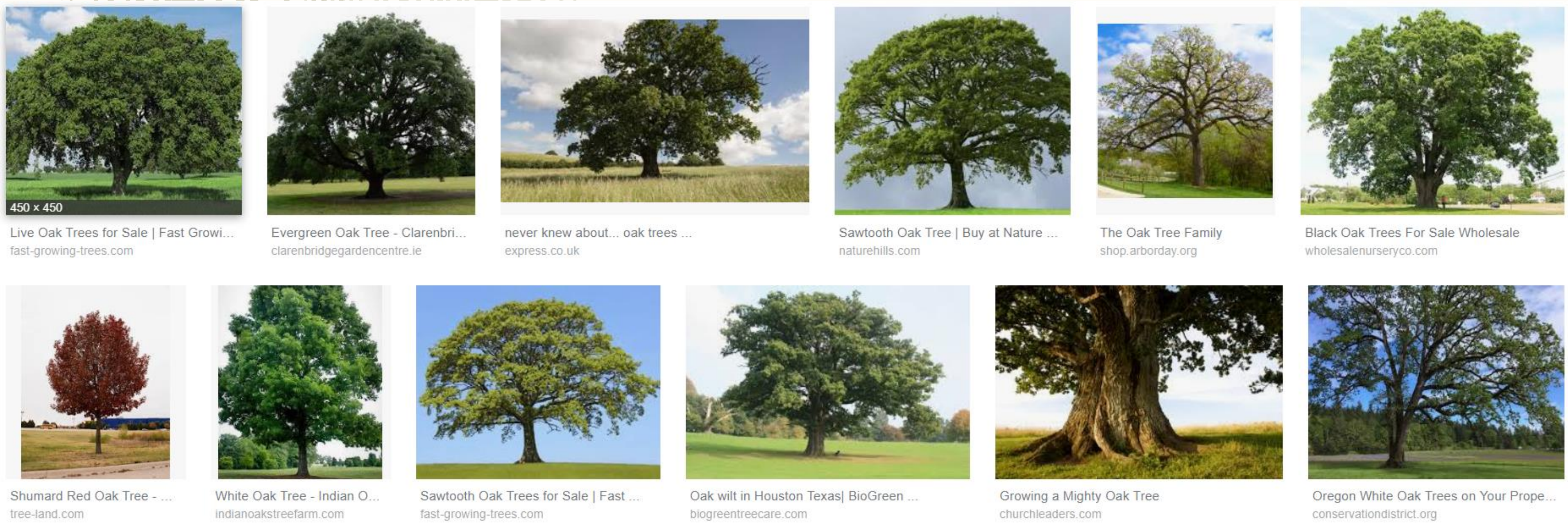
  A normalidadeé desejável mas não essencial

- Excessivo nº de zeros na matriz inicial

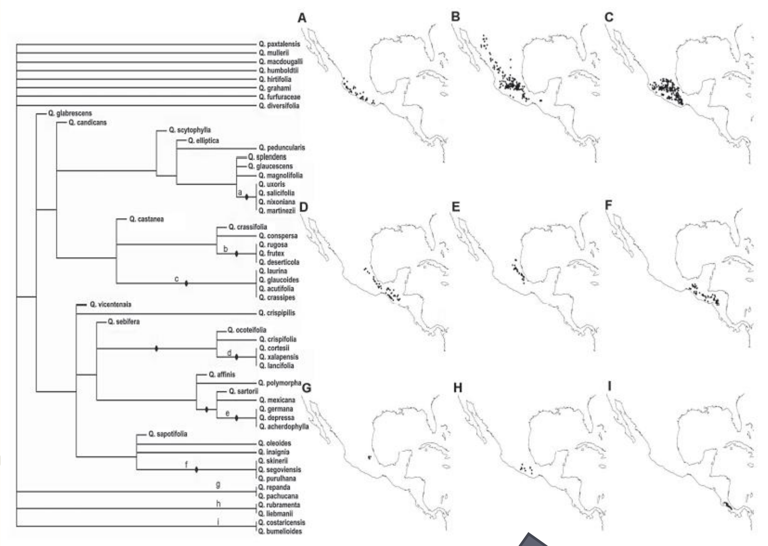
  Acarreta problemas – considerar utilizar outras técnicas, em particular análise de correspondências

Budaev, S. V. 2010 Using Principal Components and Factor Analysis in Animal Behaviour Research: Caveats and Guidelines *Ethology* **116**: 472-480

EXEMPLO 1


450 × 450
Live Oak Trees for Sale | Fast Growi...
fast-growing-trees.com
Evergreen Oak Tree - Clarenbri...
clarenbridgegardencentre.ie
never knew about... oak trees ...
express.co.uk
Sawtooth Oak Tree | Buy at Nature ...
naturehills.com
The Oak Tree Family
shop.arborday.org
Black Oak Trees For Sale Wholesale
wholesalenurseryco.com
Shumard Red Oak Tree - ...
tree-land.com
White Oak Tree - Indian O...
indianoakstreefarm.com
Sawtooth Oak Trees for Sale | Fast ...
fast-growing-trees.com
Oak wilt in Houston Texas| BioGreen ...
biogreentreecare.com
Growing a Mighty Oak Tree
churchleaders.com
Oregon White Oak Trees on Your Prope...
conservationdistrict.org

ordenação



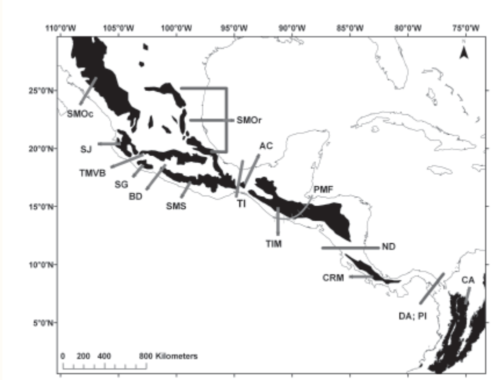


Fig. 1 Study area and its principal geological elements. TMVB = Trans-Mexican Volcanic Belt; SMOc = Sierra Madre Occidental; SMOr = Sierra Madre Oriental; SJ = Serranías de Jalisco; SG = Serranías de Guerrero; SMS = Sierra Madre del Sur; AC = Altos de Chiapas; TI = Tehuantepec Isthmus; TIM = Trans-Isthmian Mountains; PMF = Polochic-Motagua Fault; ND = Nicaraguan Depression; CRM = Costa Rican mountains; PI = Panamanian Isthmus; DAR = Darién region; CA = Colombian Andes. Black areas represent mountainous systems (>1000 m asl).

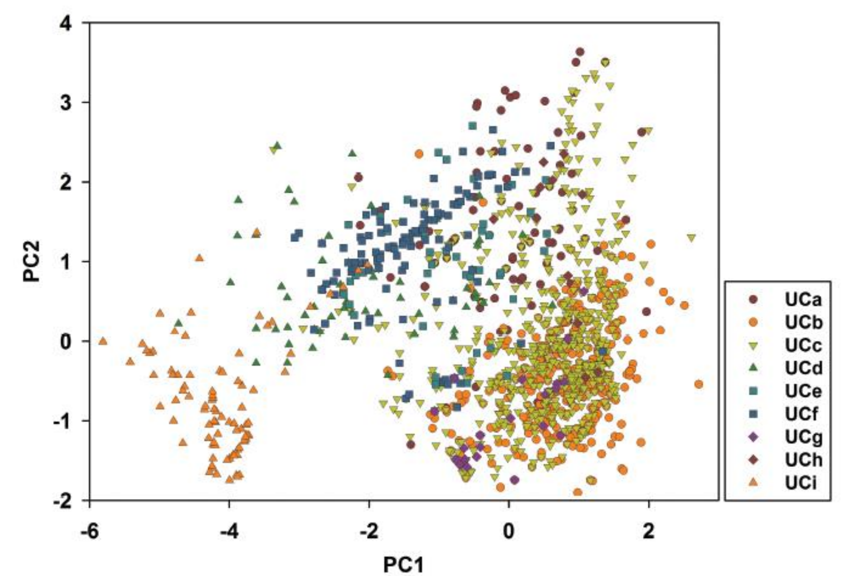
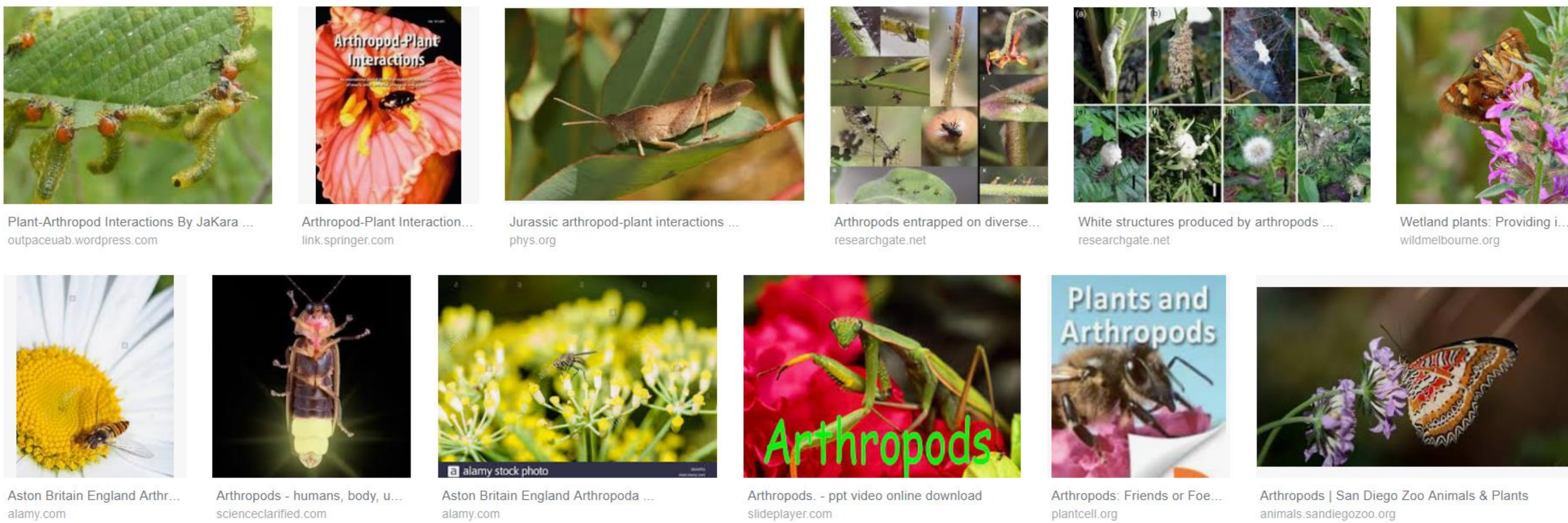


Fig. 5 Principal component (PC) analysis showing ecological niche envelopes for groups of species constituting the nine units of co-occurrence.

Rodriguez-Correa, H.; Oyama, K.; MacGregor-Fors, I. & González-Rodriguez, A. 2015 How Are Oaks Distributed in the Neotropics? A Perspective from Species Turnover, Areas of Endemism, and Climatic Niches *International Journal of Plant Sciences* **176**: 222-231

EXEMPLO 2

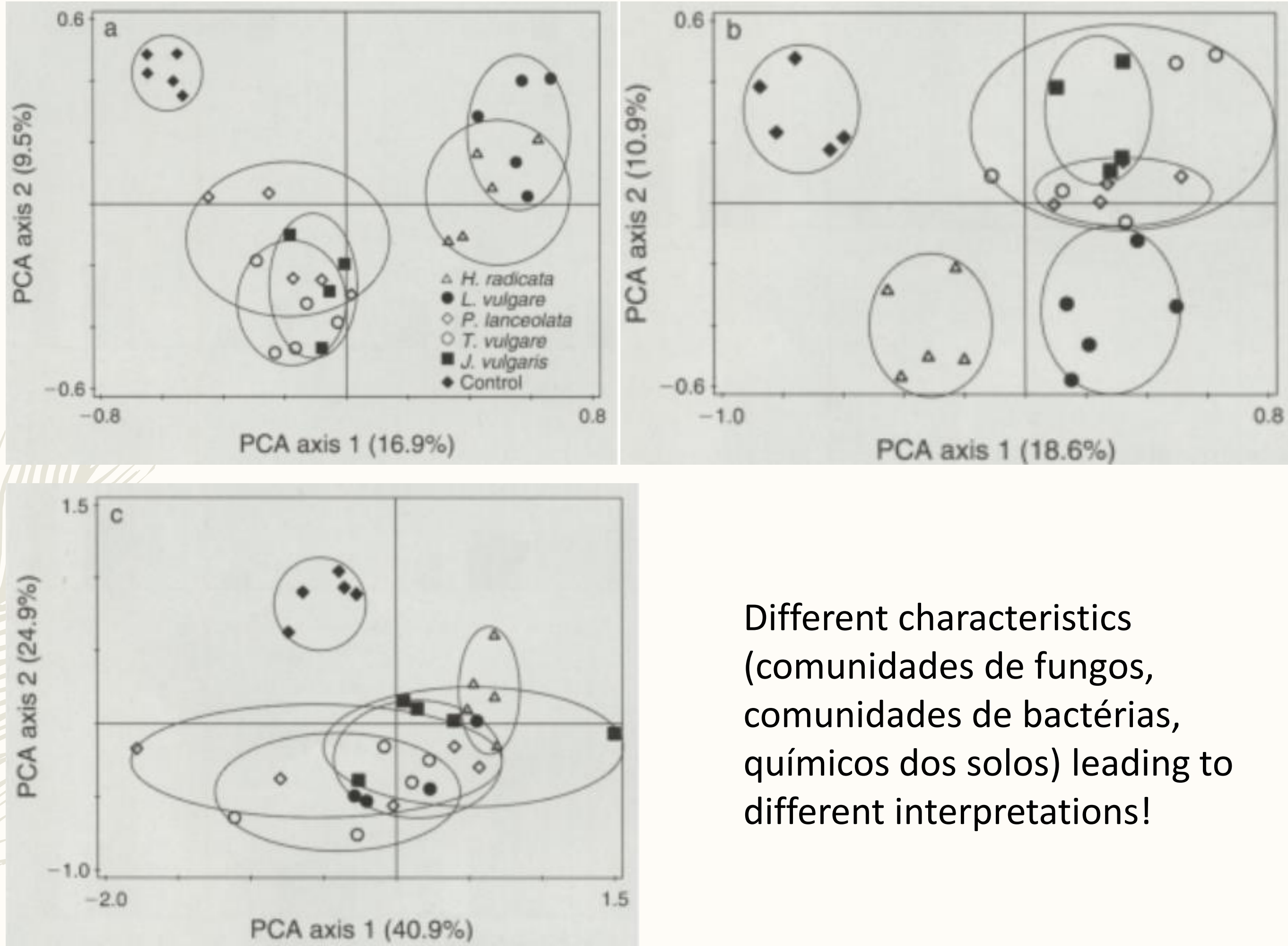


Arthropod-Plant Interactions
Plant-Arthropod Interactions By JaKara ...
outpaceuab.wordpress.com
Arthropod-Plant Interaction...
link.springer.com
Jurassic arthropod-plant interactions ...
phys.org
Arthropods entrapped on diverse...
researchgate.net
White structures produced by arthropods ...
researchgate.net
Wetland plants: Providing i...
wildmelbourne.org
alamy stock photo
Arthropods
Plants and Arthropods
Aston Britain England Arthr...
alamy.com
Arthropods - humans, body, u...
scienceclarified.com
Aston Britain England Arthropoda ...
alamy.com
Arthropods. - ppt video online download
slideplayer.com
Arthropods: Friends or Foe...
plantcell.org
Arthropods | San Diego Zoo Animals & Plants
animals.sandiegozoo.org

Disentangling above- and belowground neighbor effects on the growth, chemistry, and arthropod community on a focal plant

Martine Kos, Tibor Bukovinszky, Patrick P. J. Mulder and T. Martijn Bezemer

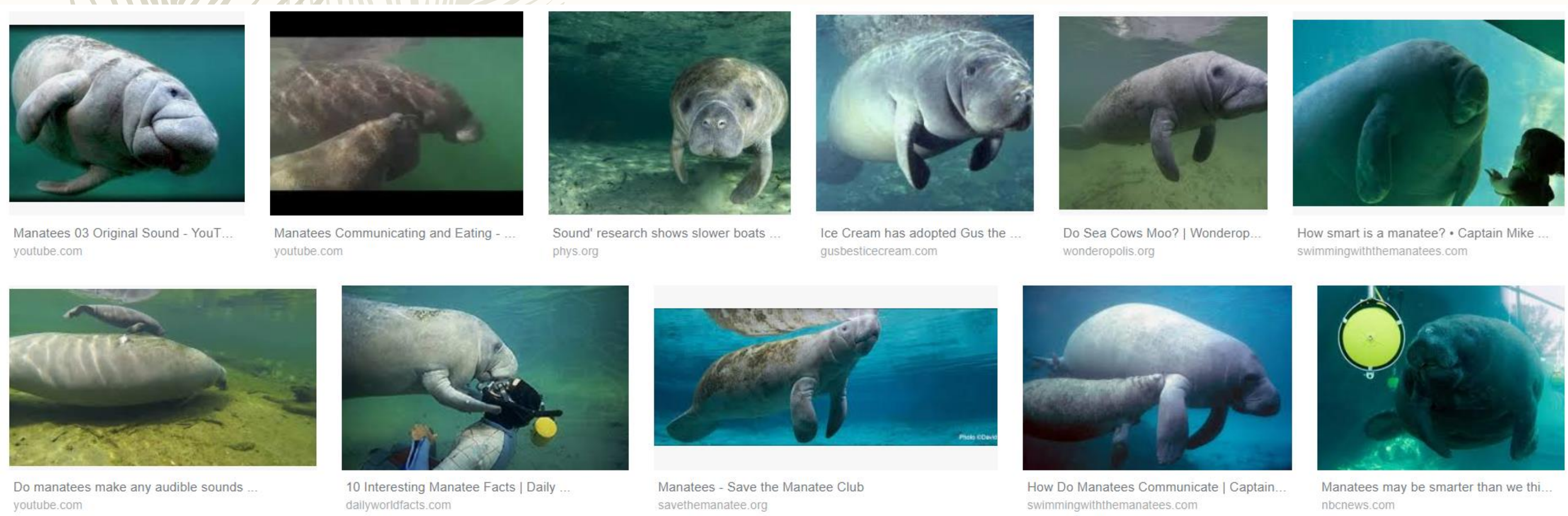
Ecology
Vol. 96, No. 1 (January 2015), pp. 164-175 (12 pages)
Published by: Wiley on behalf of the Ecological Society of America



Different characteristics (comunidades de fungos, comunidades de bactérias, químicos dos solos) leading to different interpretations!

FIG. 2. Ordination diagram of principal component analysis (PCA) of (a) fungal community composition, as assessed by terminal restriction fragment length polymorphism analysis (T-RFLP), (b) bacterial T-RFLP community composition, and (c) chemistry of soil conditioned by the five neighboring plant species (*H. radicata*, *L. vulgare*, *T. vulgare*, *P. lanceolata*, and *J. vulgaris*) and of the control soil. Percentages of total explained variation by PCA axes are given in parentheses. The composition of fungal and bacterial communities and soil chemistry differed significantly between the different soil types.

EXEMPLO 3


Manatees 03 Original Sound - YouT...
youtube.com
Manatees Communicating and Eating - ...
youtube.com
Sound' research shows slower boats ...
phys.org
Ice Cream has adopted Gus the ...
gusbesticecream.com
Do Sea Cows Moo? | Wonderop...
wonderopolis.org
How smart is a manatee? • Captain Mike ...
swimmingwiththemanatees.com
Do manatees make any audible sounds ...
youtube.com
10 Interesting Manatee Facts | Daily ...
dailyworldfacts.com
Manatees - Save the Manatee Club
savethemanatee.org
How Do Manatees Communicate | Captain...
swimmingwiththemanatees.com
Manatees may be smarter than we thi...
nbcnews.com

The Journal of the Acoustical Society of America

Acoustic monitoring of manatee populations in Panamanian wetlands using PCA-based vocalization spectrogram representation and clustering

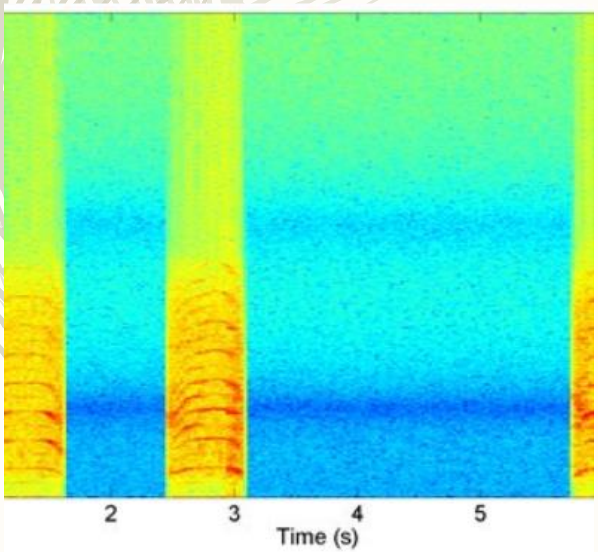
--Manuscript Draft--

For the clustering stage, the main contribution consists in analyzing the spectrograms of the vocalizations as an image with patterns. More specifically, it is proposed to represent the vocalizations spectrogram using a set of coefficients from its projection on a basis obtained by its principal component analysis (PCA). This approach is usually used in the context of face

Consequently, this scheme allows to recognize manatee vocalizations automatically to estimate manatee population size based on capture-recapture models using vocalizations instead of individual image or actual counts as other standard methods (Guzman and Condit,

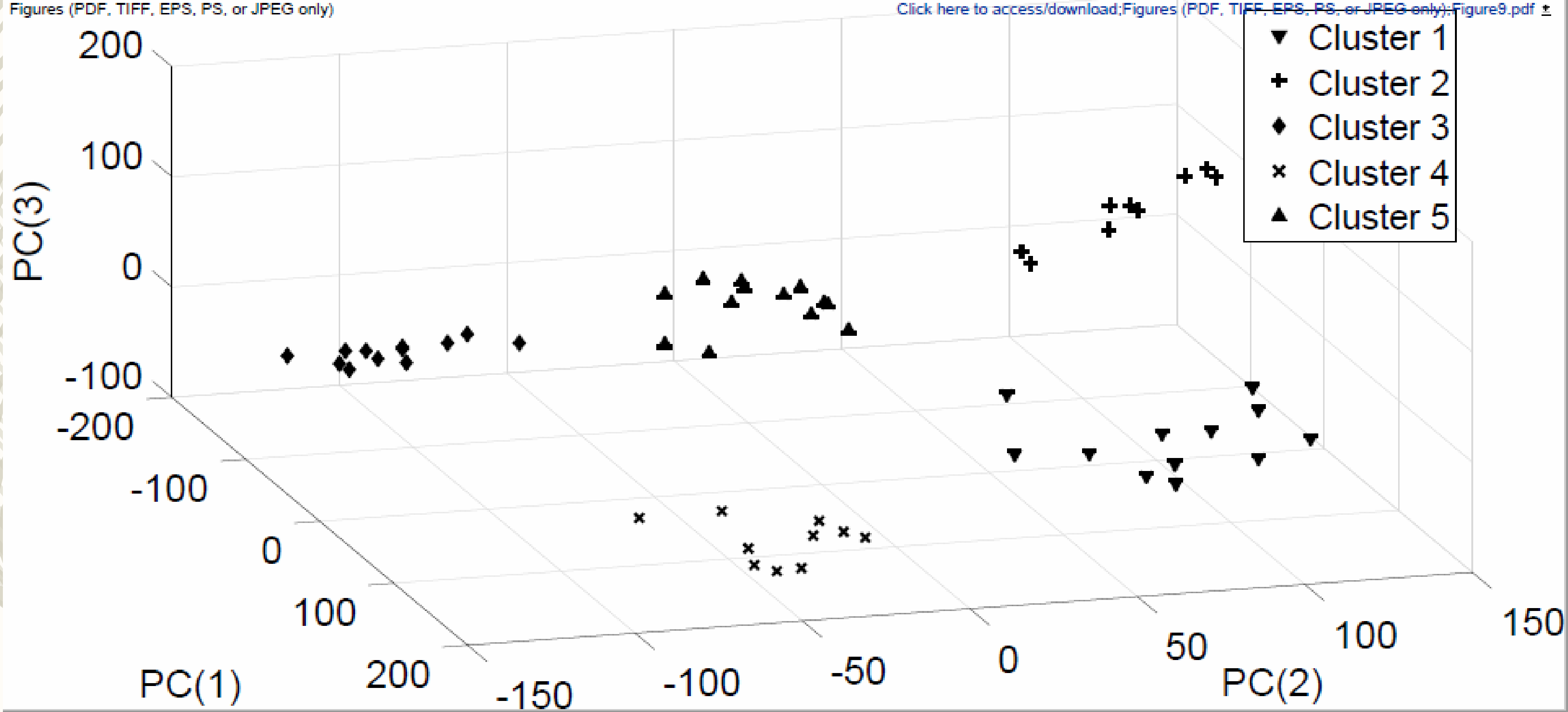
CONFIDENTIAL

Under review… but anonymized and shown just to illustrate that these are currently used to solve ecological questions of interest



Not from paper, but indeed a manatee call spectrogram

To validate the clustering stage, a data base consisting of $M = 56$ vocalizations of 5 individuals was prepared and analyzed. The number of vocalizations per individual can be further subdivided as follows: Manatee #1 with 9 vocalizations, Manatee #2 with 10 vocalizations, Manatee #3 and #4 with 12 vocalizations each and Manatee #5 with 13 vocalizations. It should be noticed that the algorithm does not consider which vocalizations corresponds to each individual or cluster (i.e. the algorithm is not provided with information of the true set of clusters).



EXEMPLO 4

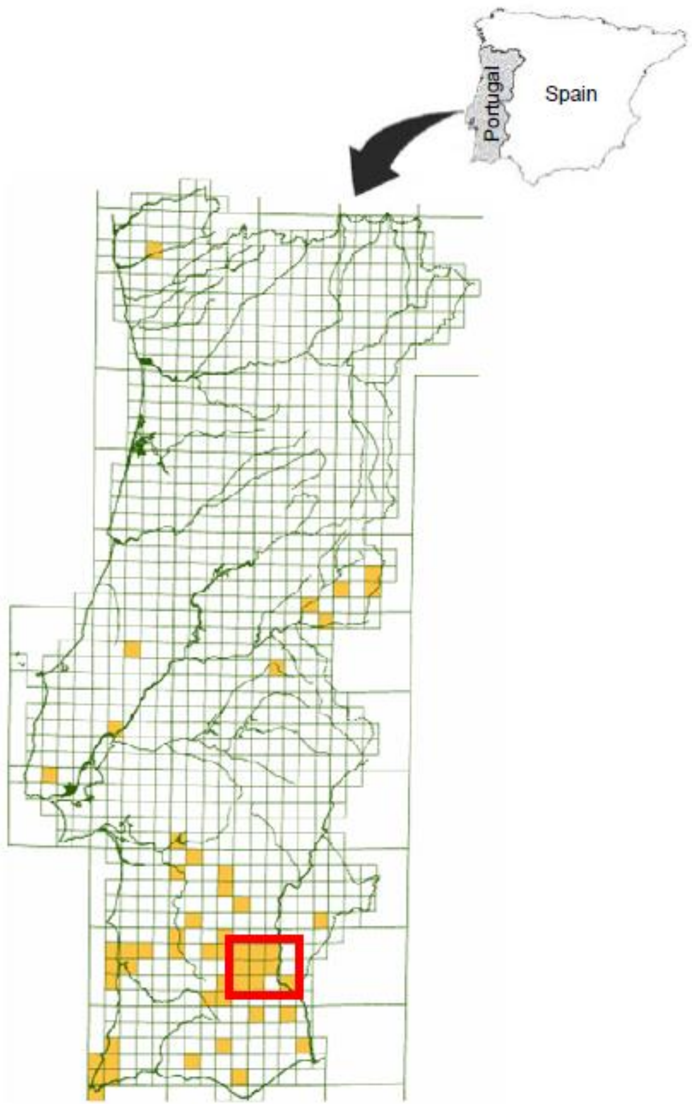


Figure 1. Distribution (UTM 10x10 km) of freshwater LBS sites in Portugal. The Guadiana region (GUAD) around Mértola and the Parque Natural do Vale do Guadiana are highlighted in red.

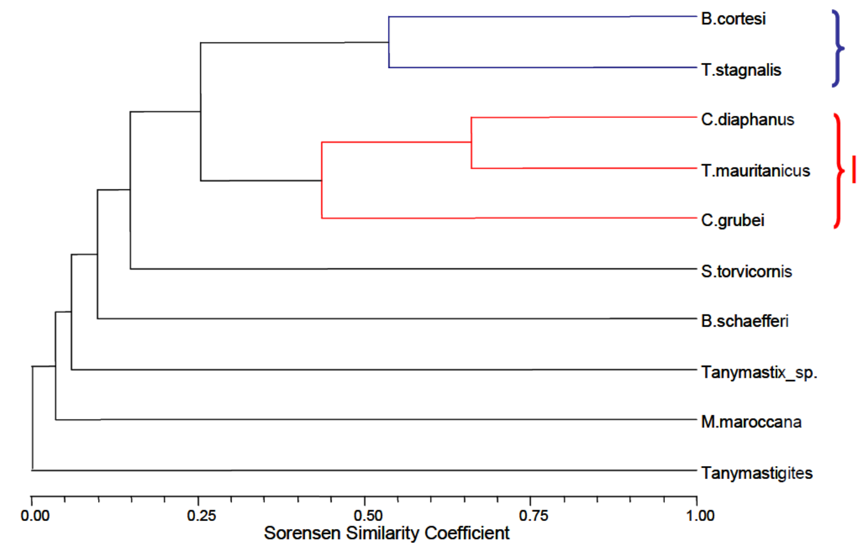


Figure 6. Cluster multivariate analysis using the Sørensen Similarity Coefficient (UPGMA method). Two main groups of large branchiopods can be identified (presence/absence LBS per pond data, R mode).

Cancela Da Fonseca, L.; Cristo, M.; Machado, M.; Sala, J.; Reis, J.; Alcazar, R. & Beja, P. 2008 Mediterranean temporary ponds in Southern Portugal: key faunal groups as management tools? *Pan-American Journal of Aquatic Sciences* **3**:304-320

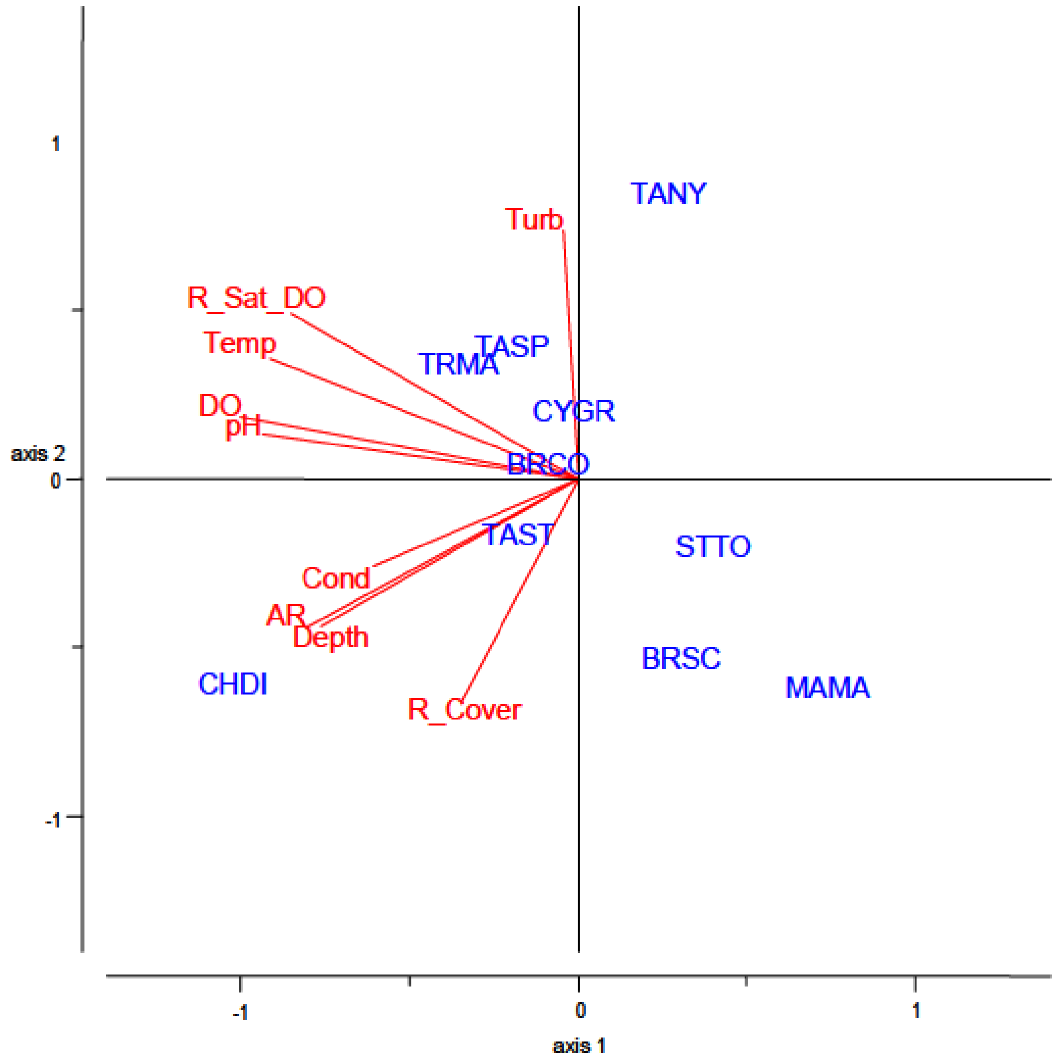


Figure 5. Ordination of large branchiopods (PCA) according to the range of environmental variables. Codes according to Tables I and II (axis 1 and 2 explain 75.8% of the variance).

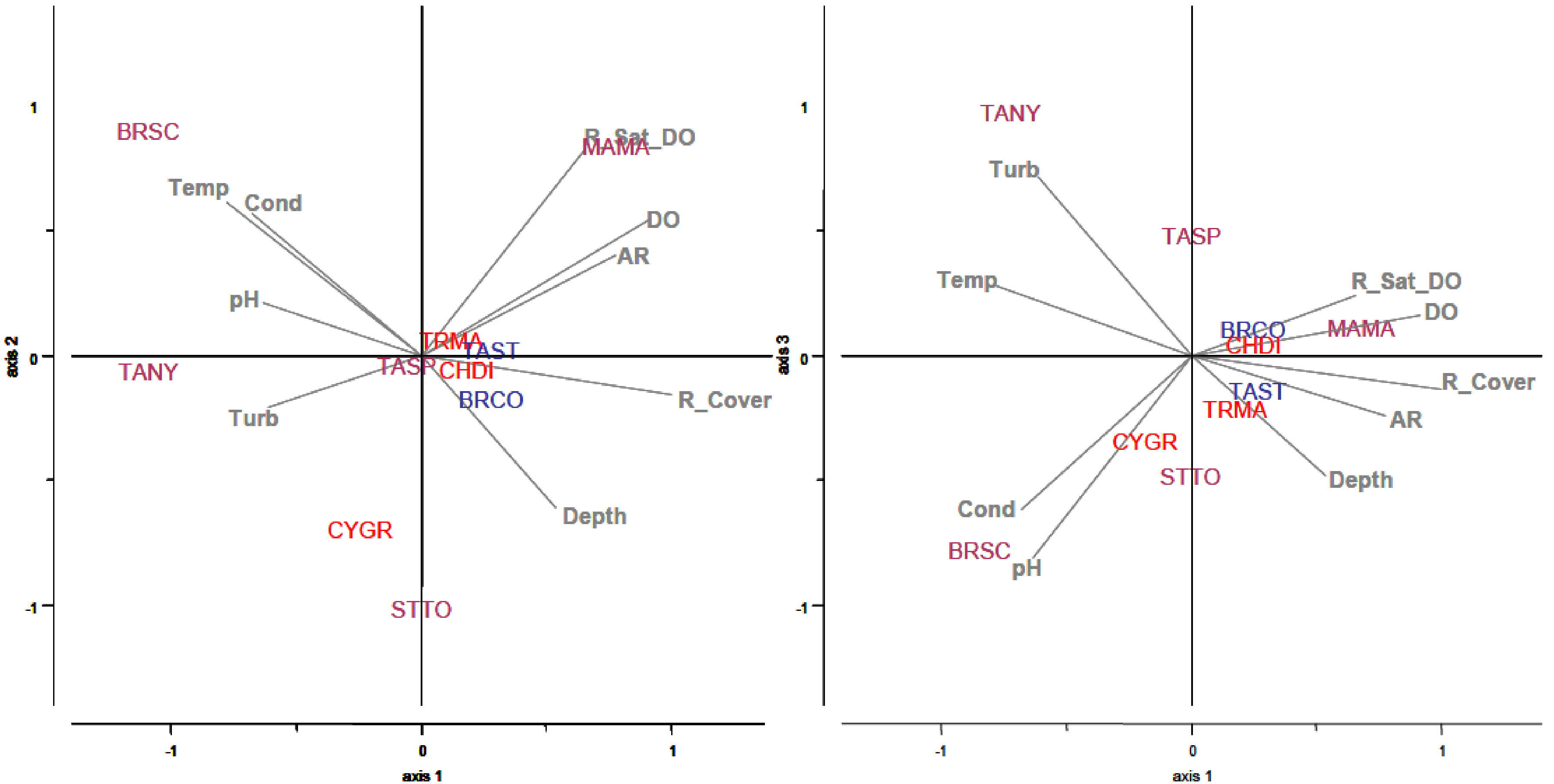


Figure 7. Ordination (PCA) of LBS in relation to average and median values of environmental variables. Codes according to Tables I and II. Blue and red are those referred as Groups I and II in presence/absence cluster analysis (axis 1, 2 and 3 explain 86.0% of the variance).

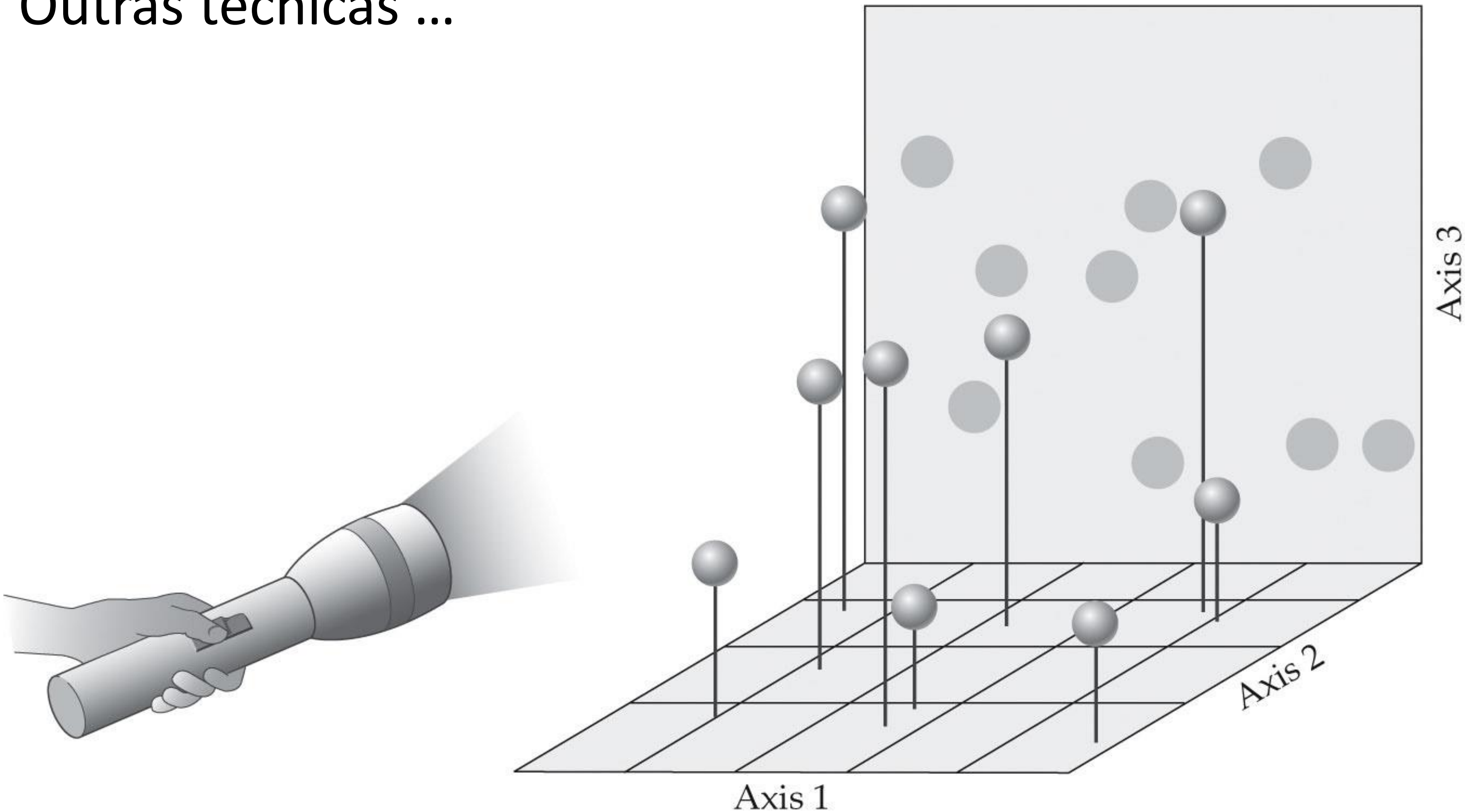
Uma aplicação comum da PCA é reduzir a dimensionalidade de um conjunto de dados com muitas variáveis explicativas (e que resultará tanto melhor quanto mais correlacionadas entre si forem as variáveis).

Depois poderemos por exemplo aplicar uma análise de regressão sobre as (*k* primeiras) componentes principais.

Note-se que com este procedimento o resultado da regressão que se segue será de interpretação difícil, o que quer dizer que só faz sentido fazer isto se o objetivo principal da análise de regressão for a **predição** e não a **explicação/interpretação**.

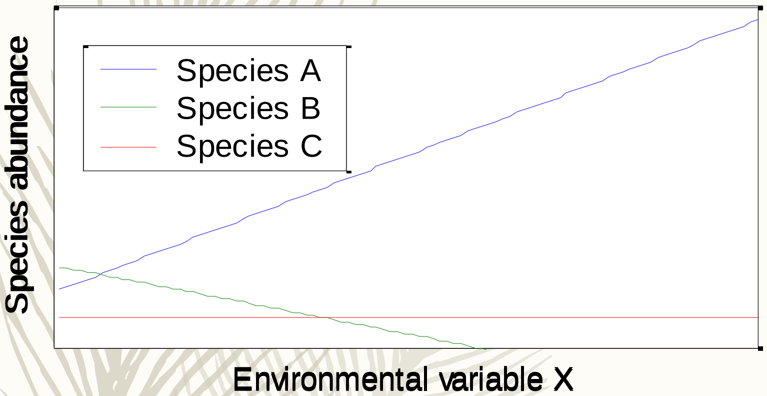
ordenação

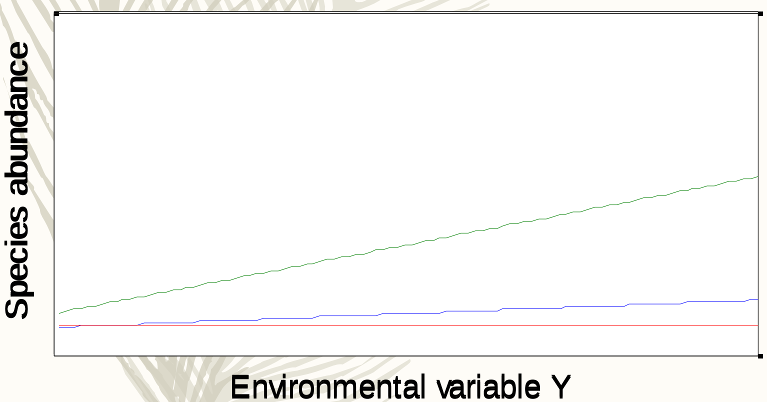
Outras técnicas ...



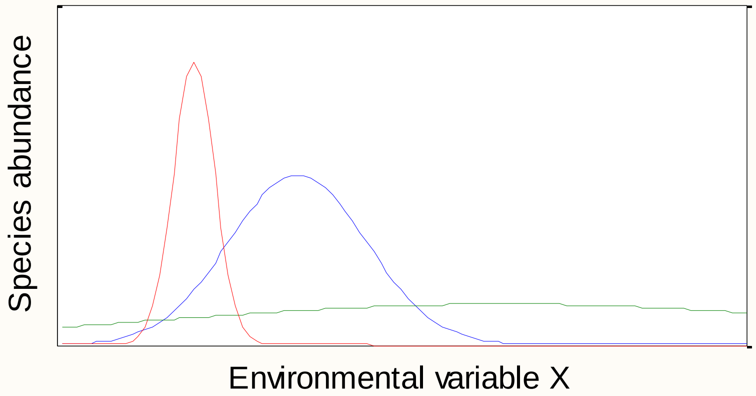
ordenação

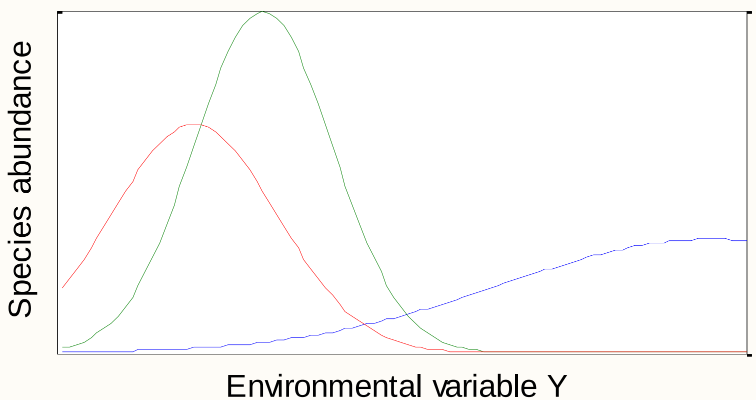
Relações lineares





Relações unimodais

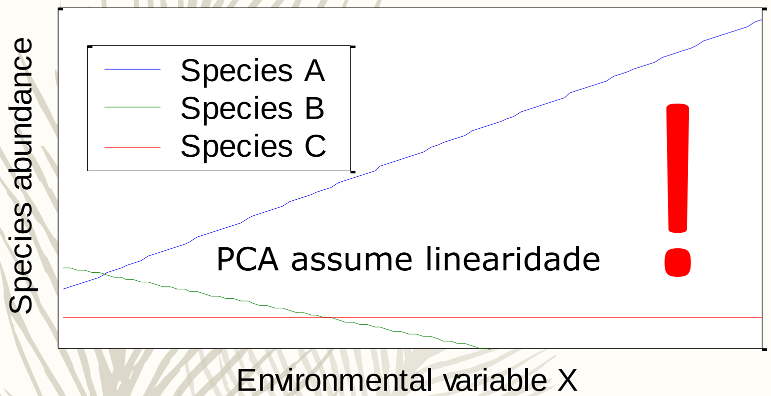


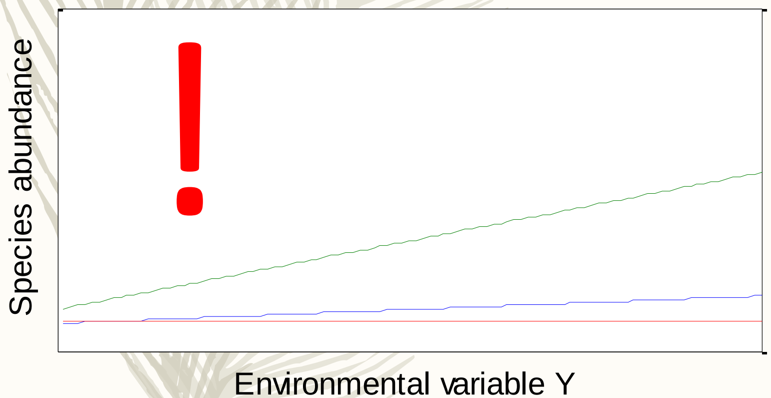


ou ... relações polimodais ou outras...

ordenação

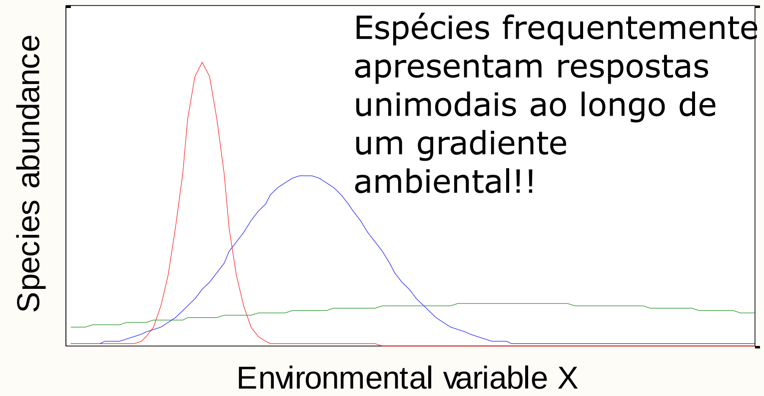
Relações lineares

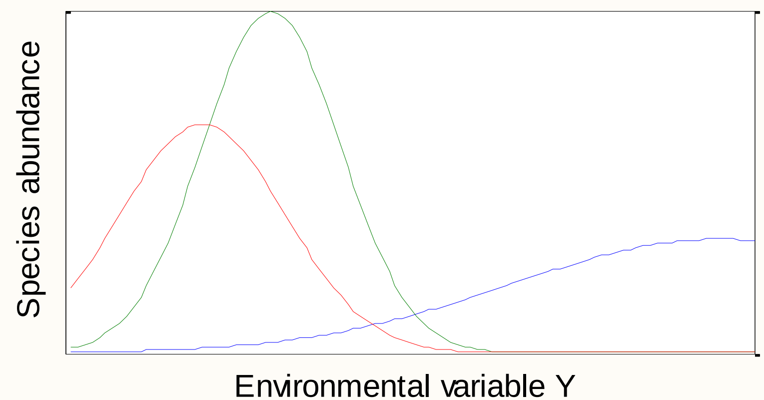




PCA

Relações unimodais





CA e CCA

https://www.researchgate.net/post/How_to_choose_ordination_method_such_as_PCA_CA_PCoA_and_NMDS

https://stats.stackexchange.com/questions/70533/using-principal-components-analysis-vs-correspondence-analysis

MOLECULAR ECOLOGY

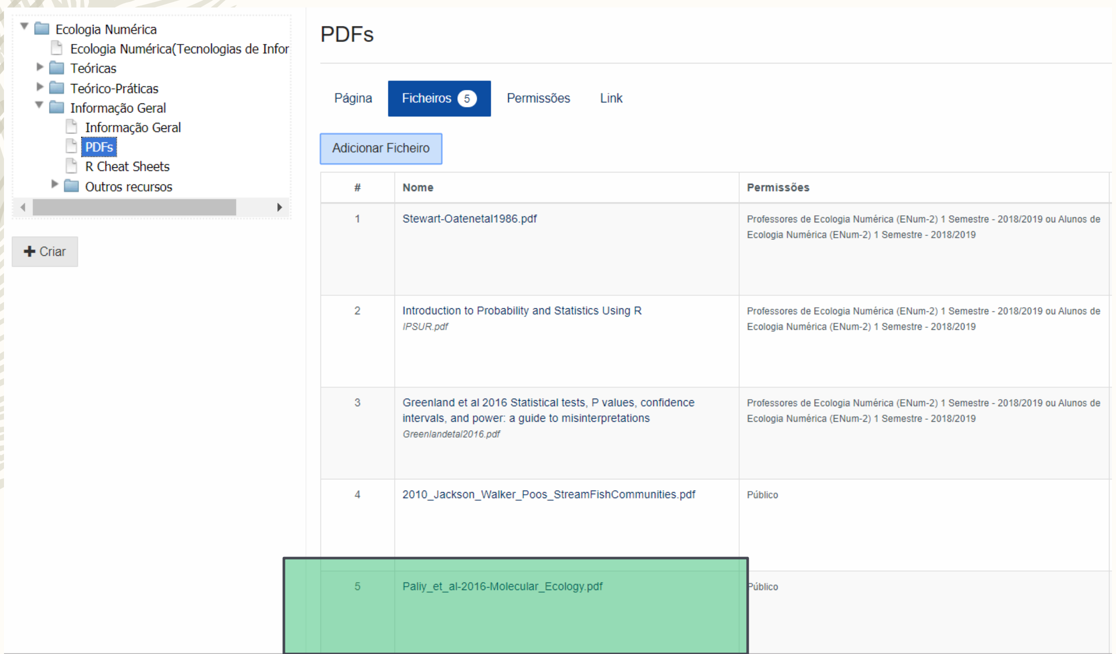
Molecular Ecology (2016) **25**, 1032–1057 doi: 10.1111/mec.13536

INVITED REVIEWS AND SYNTHESES

Application of multivariate statistical techniques in microbial ecology

O. PALIY and V. SHANKAR

Department of Biochemistry and Molecular Biology, Boonshoft School of Medicine, Wright State University, 260 Diggs Laboratory, 3640 Col. Glenn Hwy, Dayton, OH 45435, USA



Análise de correspondências (CA)

- Desenvolvida originalmente como método de decomposição de tabelas de contingência;
- O principal objectivo é sumariar a falta de independência duma tabela de contingência num sistema de eixos de dimensão reduzida;
- Os cálculos envolvem a determinação de valores esperados e ponderação por totais de linhas e colunas.

Análise de correspondências (CA)

	Q1	Q2	Q3	Q4	Q5
Espécie A	15	2	5	2	1
Espécie B	9	6	15	0	0
Espécie C	1	7	5	8	29
Total	25	15	25	10	30

Q4(10)< Q2 (15) < Q1 e Q3 (25) < Q5 (30) → Q1 e Q3: mais semelhantes
Q4 e Q5: mais diferentes

Olhando para as abundâncias: espécie que domina em Q1, está em menor abundância em Q3
Q4 e Q5: abundâncias diferentes, mas as espécies são as mesmas.

↓

Necessário atribuir um "peso" às abundâncias

Como decidir? Mais peso à espécie mais frequente? À mais abundante?

Análise de correspondências (CA)

Abordagem metodológica mais comum é denominada reciprocal averaging, segundo a qual as coordenadas das observações nos vários eixos são obtidas calculando as médias ponderadas das variáveis para essas observações e vice-versa.

Reciprocal averaging, num sequência de tentativa e erro, atribui pesos arbitrários às espécies e vai recalculando estes valores até convergirem num único valor. Ou seja, são os próprios dados que dão origem aos “pesos”, não sendo este imposto. Este “re-cálculo” é guiado por várias ponderações entre linhas (espécies) e colunas (quadrados): os scores dos quadrados passam a ser médias ponderadas das espécies, e as espécies passam a ser médias ponderadas dos quadrados. A partir daqui são calculados os scores para cada ponto.

Análise de correspondências (CA)

- O resultado é um biplot que representa simultaneamente observações (linhas da matriz) e variáveis (colunas da matriz);

- O diagrama de ordenação é interpretado de acordo com as associações entre os pontos que representam observações e variáveis.