

# NEPŘÍMÉ ORDINACE CA, DCA, PCoA A NMDS

# KORESPONDENČNÍ ANALÝZA

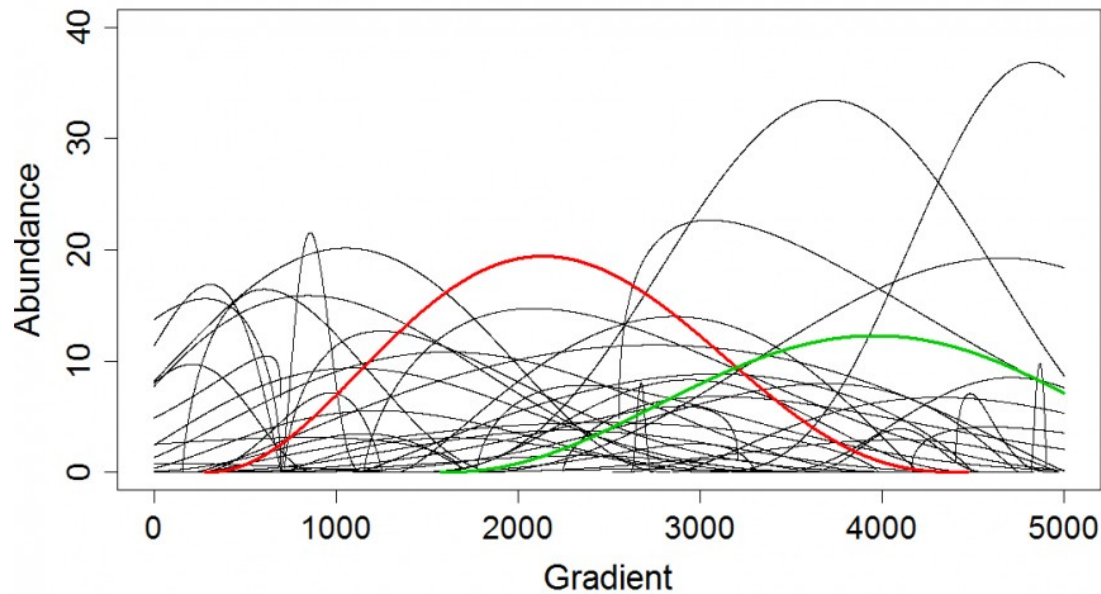
## *CORRESPONDENCE ANALYSIS (CA)*

- založena na konceptu niky (Hutchinson 1957)
  - druhy vykazují unimodální odpověď na ekologické gradienty, optima druhů stejně jako jejich niky se vzájemně liší
  - CA hledá takové gradienty, na nichž se niky druhů maximálně separují
- zachovává  $\chi^2$  (chi kvadrát) vzdálenosti
  - odráží rozdíly v relativním zasoupení druhů vážené celkovou početností
  - Netrpí problémem dvojitých absencí (double zeros)
  - možné použít přímo na data početností
  - nesmí být negativní hodnoty a **prázdné vzorky**
  - V ekologii CA zavedena Markem Hillem v roce 1973 pod názvem *reciprocal averaging* – což je algoritmus, který se používá pro konstrukci ordinačních os

# SIMULOVANÁ DATA

## JEDEN EKOLOGICKÝ GRADIENT

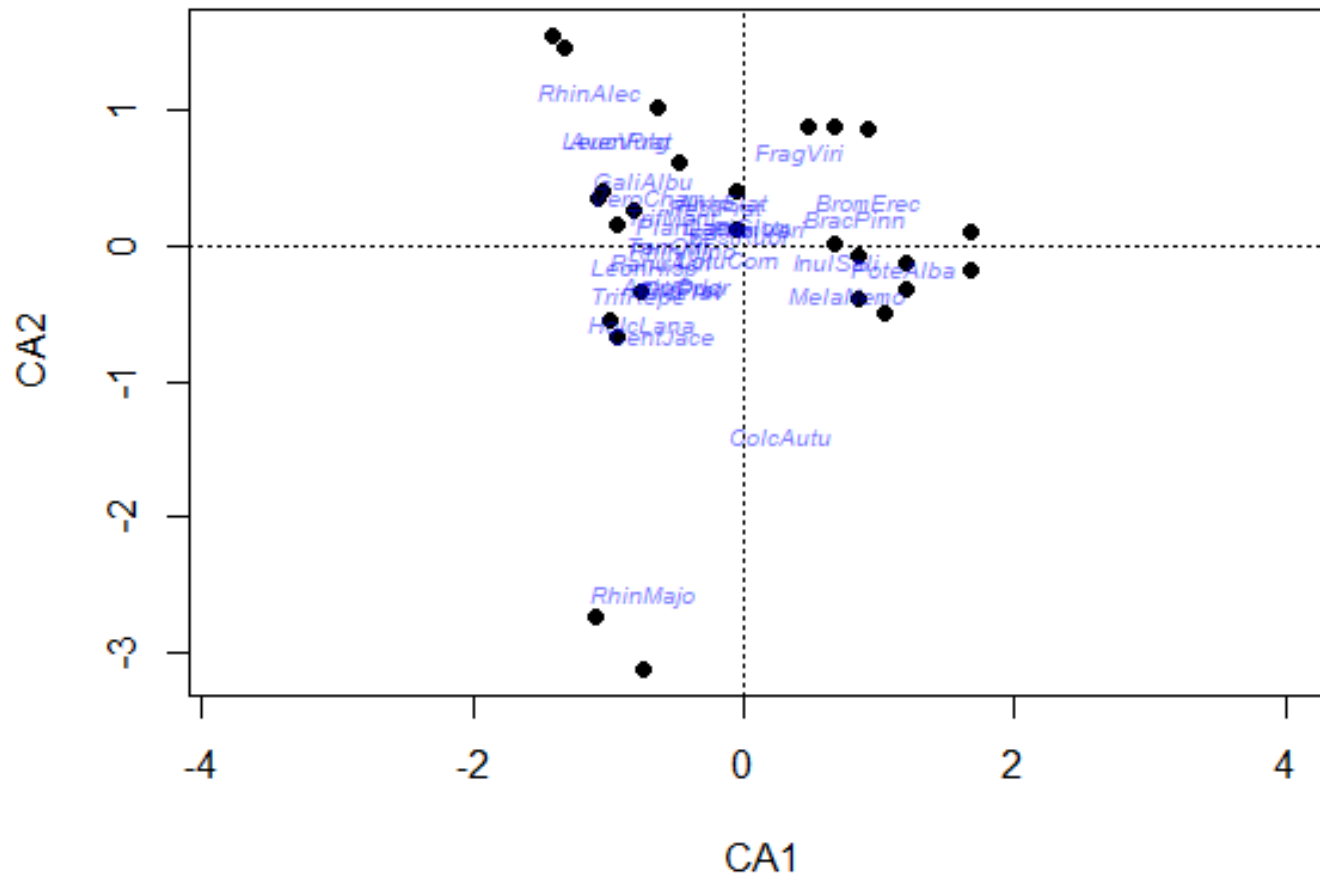
- simulovaný gradient dlouhý 5000 jednotek
- 300 druhů s unimodální odpovědí, různými šířkami nik
- 500 vzorků náhodně rozmístěných podél gradientu



## CA – ŠKÁLOVÁNÍ ORDINAČNÍCH SKÓRE

- stejně jako u PCA škálování ordinačních skóre
  - 1 – zaměřeno na vzorky, umístění vzorků odpovídá váženému průměru druhů, jsou zachovány  $\chi^2$  vzdálenosti mezi vzorky
  - 2 – zaměřeno na druhy, umístění druhů odpovídá váženému průměru vzorků, zachovává  $\chi^2$  vzdálenosti mezi druhy
  - 3 – symetrické zobrazující přibližně (ale nepřesně) oboje. Zachovává odmocninu  $\chi^2$  vzdáleností mezi druhy i vzorky

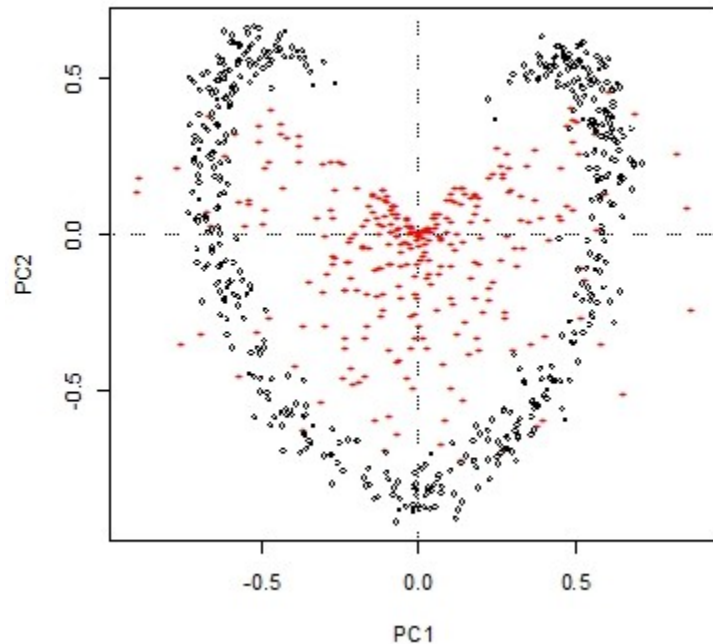
# CA – ORDINAČNÍ DIAGRAM (DRUHY A VZORKY)



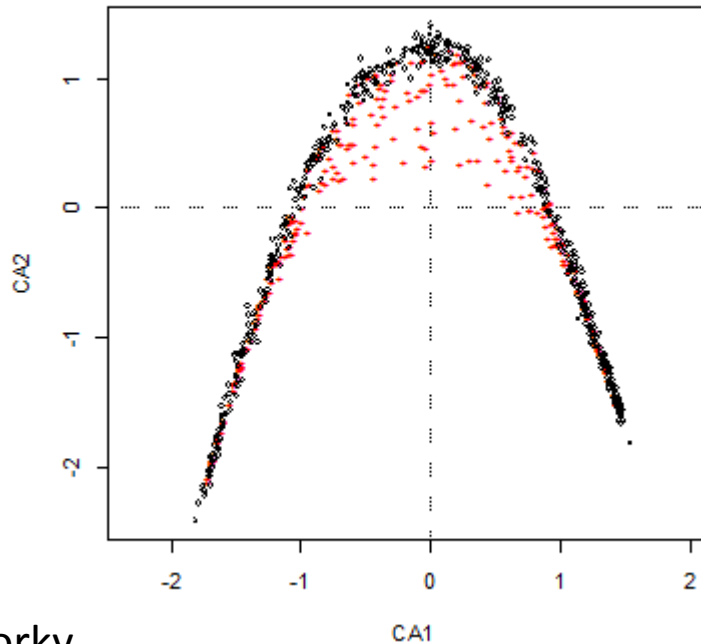
# ARTEFAKTY NEPŘÍMÝCH ORDINACÍ

## SIMULOVANÁ DATA

### PCA - podkova



### CA - oblouk

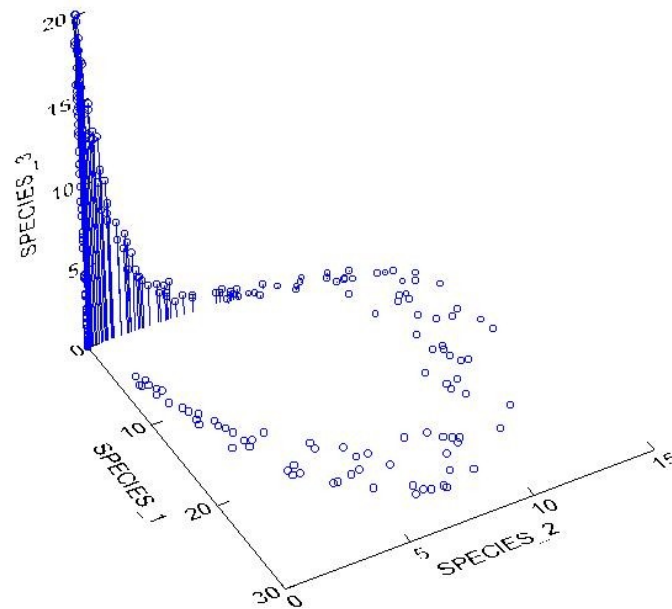
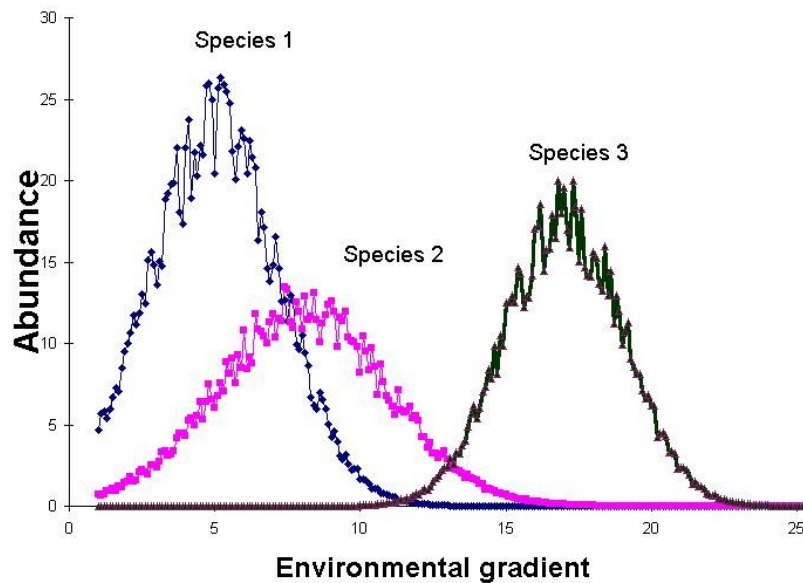


o vzorky  
+ druhy

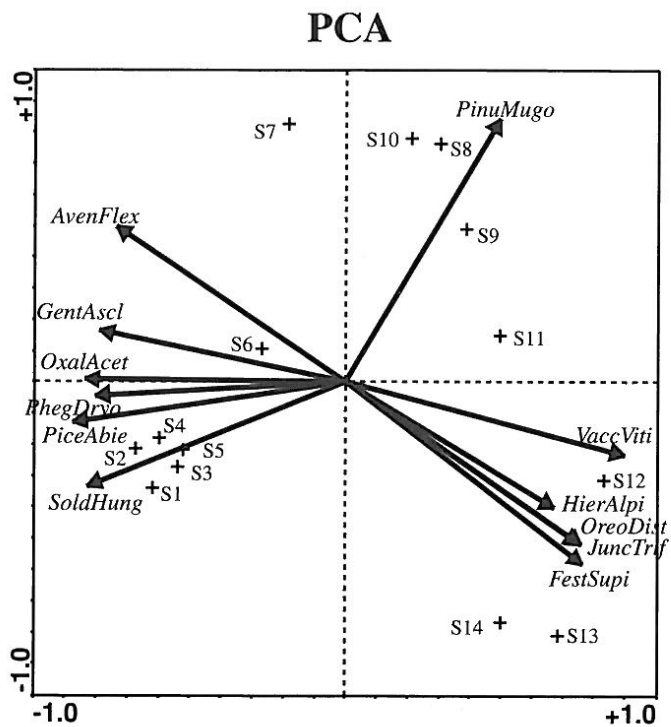
# ARTEFAKTY V ORDINACÍCH

## PŘÍČINY

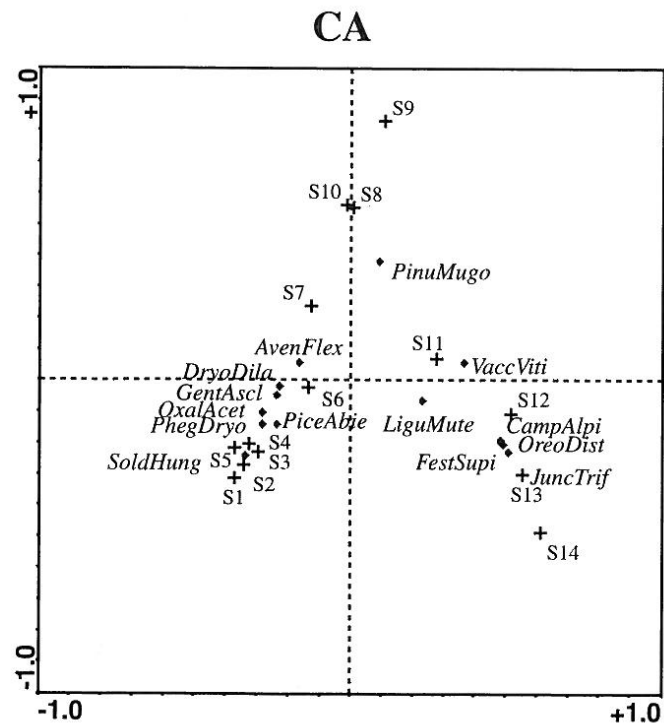
- důsledek algoritmu (lineární nezávislost všech os), ale kvadratická není ošetřena
- důsledek projekce (nelineární vztahy mezi druhy -> lineární prostor)



# ORDINAČNÍ DIAGRAMY



lineární metoda



unimodální metoda



# ARTEFAKTY V ORDINACÍCH

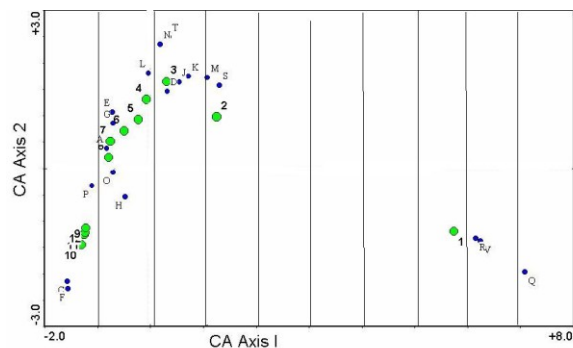
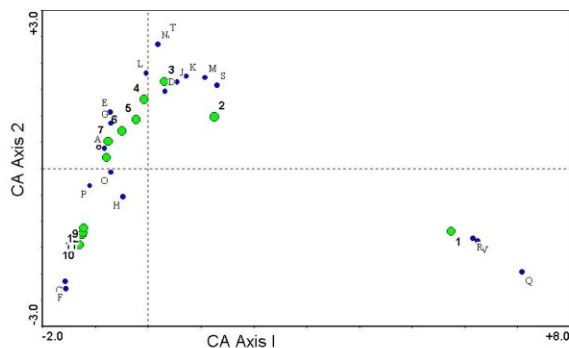
## MOŽNOSTI ŘEŠENÍ

- odstranění trendu z ordinačních os (***detrending***)
  - detrendovaná korespondenční analýza, ***Detrended Correspondence Analysis*** (DCA, Hill & Gauch 1980)
  - *detrending by segments* (nejčastější)
  - *detrending by polynomials*
- použití takových ordinačních technik, které umožňují ordinaci vzorků v prostoru pomocí jiných metrik než je Euklidovská distance (PCA) nebo  $\chi^2$ -kvadrát distance (CA)
  - analýza hlavních koordinát, ***Principal Coordinate Analysis*** (PCoA)
  - nemetrické mnohorozměrné škálování, ***Non-metric Multidimensional Scaling*** (NMDS)

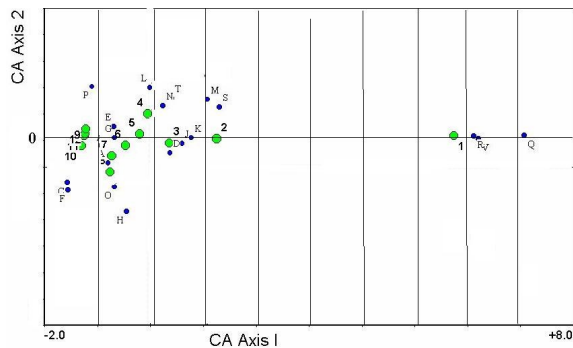
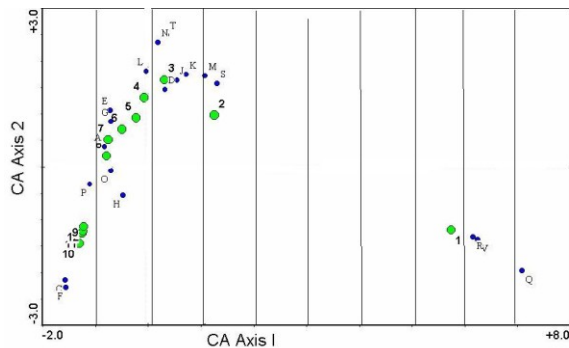
# DCA – ODSTRANĚNÍ TRENDU

(Detrended Correspondence Analysis, detrendovaná korespondenční analýza)

Krok 1 – rozdělení první osy na několik segmentů



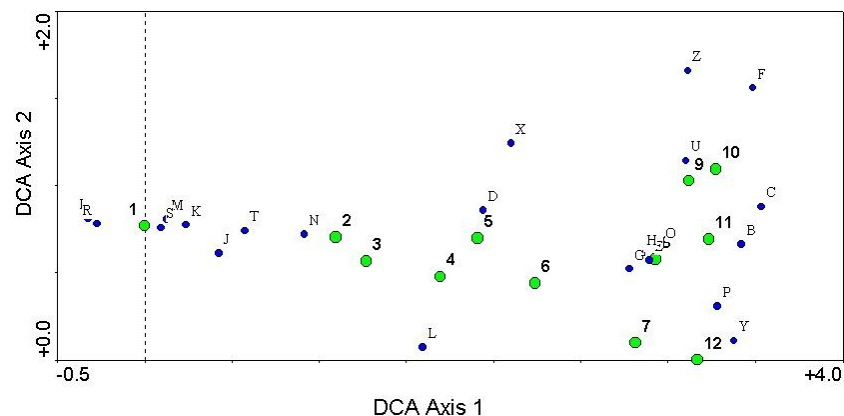
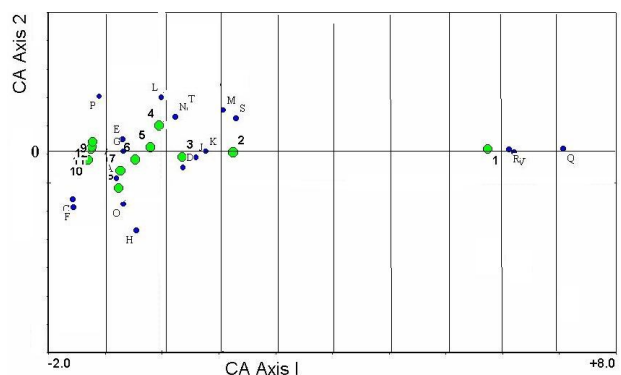
Krok 2 – vycentrování druhé osy každého segmentu kolem nuly



# DCA – ODSTRANĚNÍ TRENDU

(*Detrended Correspondence Analysis*, detrendovaná korespondenční analýza)

Krok 3 – nelineární přeškálování první osy



<http://ordination.okstate.edu>

Výsledek škálování:

- osy naškálované v jednotkách **směrodatné odchylky (SD)**
- celé druhové složení se obmění na **4 SD**

# VÝBĚR ORDINAČNÍ METODY NA ZÁKLADĚ DCA

## LINEÁRNÍ NEBO UNIMODÁLNÍ?

Pokud je délka 1. osy DCA

- **menší než 3 SD** – homogenní data - lineární metoda
- **větší než 4 SD** – heterogenní data - unimodální metoda
- **v rozmezí 3-4 SD** – obě techniky pracují rozumně

Platí jen pro detrendování po segmentech a délku **první osy!**

Obecně je ale lepší se zamyslet nad smysluplností dvojitého absencí pro podobnost. Dávají-li smysl, použiju lineární metody, jinak používám unimodální, příp. distance-based

# PCoA – PRINCIPAL COORDINATE ANALYSIS

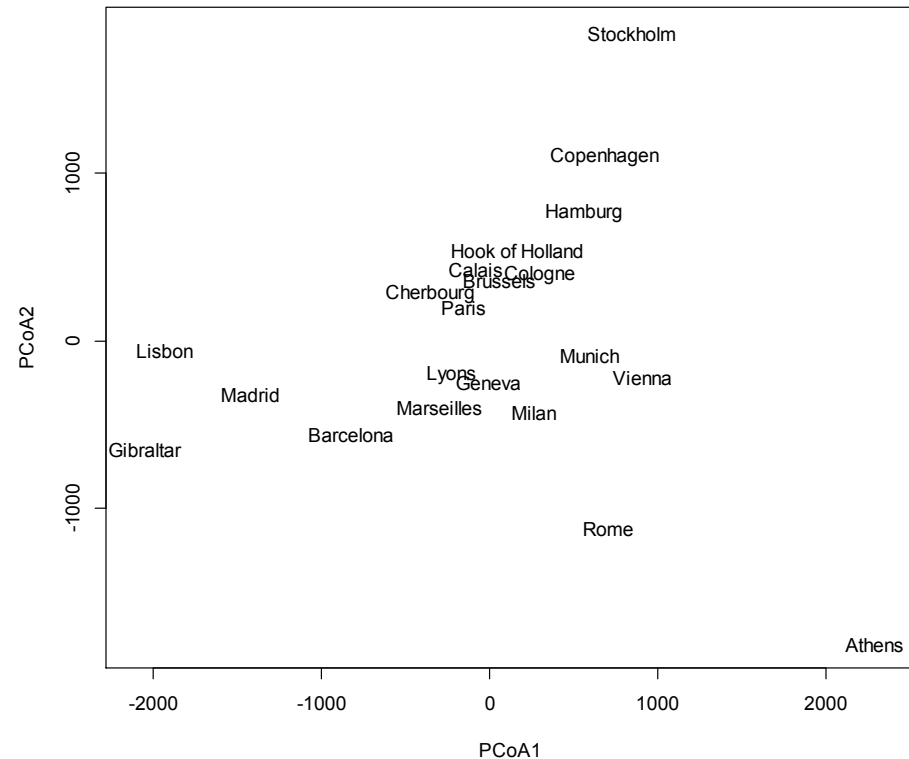
(*analýza hlavních koordinát*)

- metoda založená na nepodobnostech mezi vzorky
- vstupní data – matice nepodobností mezi vzorky
  - Bray-Curtis (percent dissimilarity) – vhodná metrika nepodobnosti pro abundanční data; netrpí problémem dvojitých nul
  - Sørensen, Jaccard index – pro presence/absence data; netrpí problémem dvojitých nul
  - pokud zvolím Euklidovskou vzdálenost -> obdoba PCA
  - pokud zvolím Chi-kvadrát vzdálenost -> obdoba CA
- umístí objekty na základě jejich nepodobností do Euklidovského prostoru (tvořeného souřadnicemi – skóre vzorků na osách)
- Při převodu nepodobností (B-C, Sör, Jacc) do Eukleidovského prostoru mohou vznikat osy s negativními eigenvalues a skóre tvořenými imaginárními čísly
  - Problém pro interpretaci výsledků (množství vysvětlené variability),
  - Často lze vyřešit odmocněním původní nepodobnosti (zejména u semimetrických nepodobností - B-C, Sör)
- synonymum MDS – (*metric*) MultiDimensional Scaling

# PCoA – PŘÍKLAD NA VZDÁLENOSTECH MEZI MĚSTY

## Vzdálenosti mezi městy (km)

	Athens	Barcelona	Brussels	...
Athens	0			
Barcelona	3313	0		...
Brussels	2963	1318	0	...
Calais	3175	1326	204	...
Cherbourg	3339	1294	583	...
Cologne	2762	1498	206	...
Copenhagen	3276	2218	966	...
Geneva	2610	803	677	...
Gibraltar	4485	1172	2256	...
Hamburg	2977	2018	597	...
...	...	...	...	...

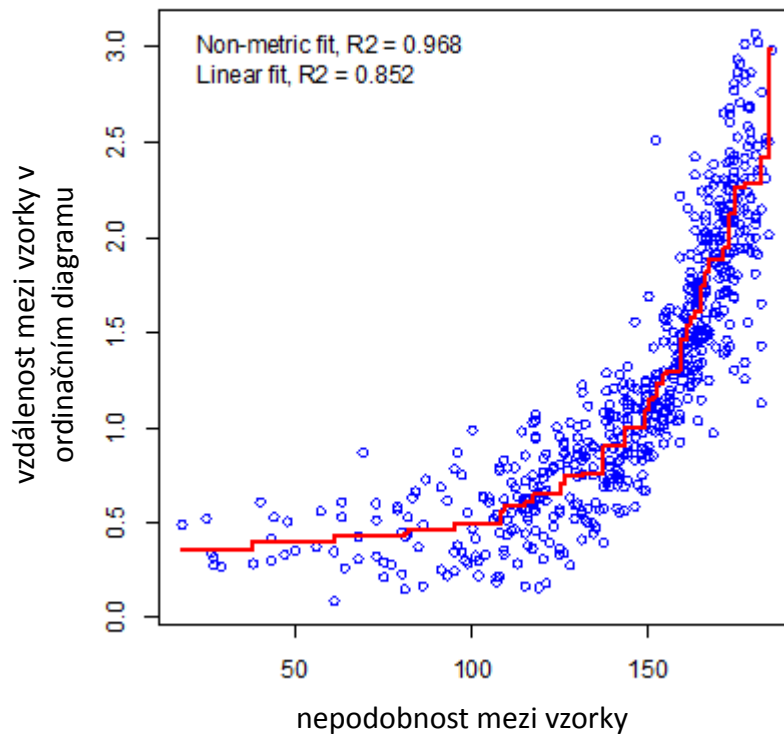


# NMDS - NON-METRIC MULTIDIMENSIONAL SCALING)

## NEMETRICKÉ VÍCEROZMĚRNÉ ŠKÁLOVÁNÍ

- nemetrická varianta PCoA (nepracuje přímo s distancemi mezi vzorky, ale s jejich pořadím)
- vstupní data – matice nepodobností mezi vzorky
- **iterativní algoritmus**, který nemusí pokaždé dojít ke stejnému výsledku (lokální optima)
- **Stress** – parametr kvality analýzy
- nutno určit počet dimenzí, se kterými bude metoda pracovat
  - Bud' počet zobrazovaných dimenzí nebo tento počet+1
  - Je dobré zkusit více hodnot  $k$  (typicky až 5) a sledovat, kde dochází k prudkému poklesu stresu. Toto  $k$  následně zvolíme pro konečnou analýzu
- na rozdíl od PCoA optimalizuje výsledné vzdálenosti mezi vzorky do několika málo (dvě – tři) dimenzí

# NMDS – SHEPARDŮV DIAGRAM



Pro stress-value přibližně platí:

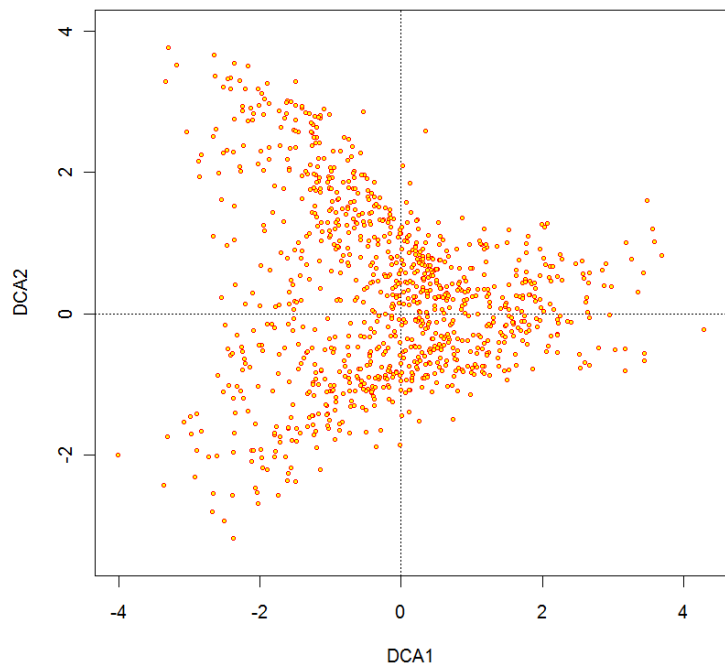
- < 0.05 – vynikající
- < 0.1 – výborný
- < 0.2 – dobrý
- > 0.3 – špatný

(Clarke & Warwick 2001)



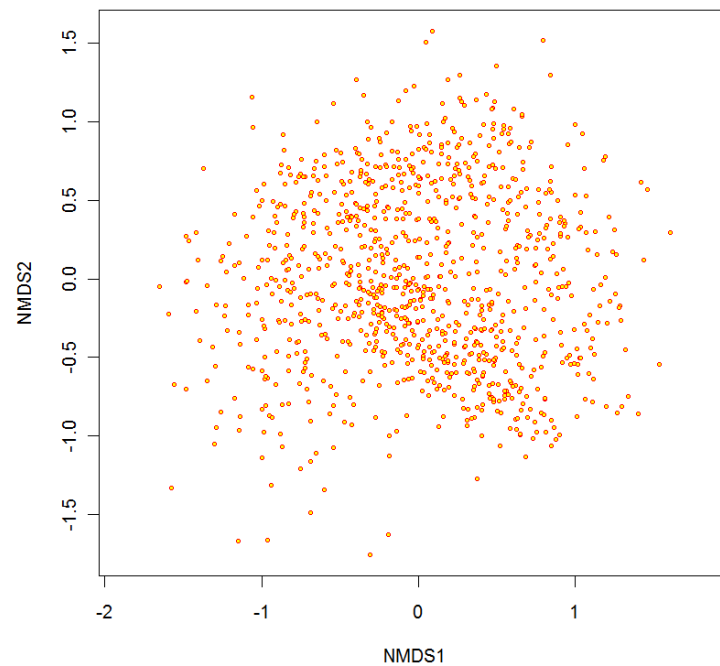
# POROVNÁNÍ METOD DCA A NMDS

## DCA



při větším počtu vzorků tvoří trojúhelník  
nebo pěticípou hvězdu (artefakt)

## NMDS



tradiční algoritmus má tendenci  
jakákoliv data zobrazit jako kouli

# KTEROU NEPŘÍMOU ORDINACI ZVOLIT?

	Měřené proměnné - Parametry prostředí - Morfometrická data	Druhové složení
PCA	ANO (se standardizací po proměnných!)	ANO (ale typicky vyžaduje standardizaci po vzorcích – např. Hellingerova standardizace)
DCA, CA	NE	ANO – dobře identifikuje gradienty v datech
PCoA	ANO, pokud nelze použít std-PCA; např. Gower dissimilarity pro kombinaci kategoriálních a kvantitativních proměnných	ANO (Bray-Curtis pro abundance; Jaccard nebo Sørensen pro pres/abs data) Dobře identifikuje gradienty.
NMDS	NE	ANO – dobře zobrazuje vzdálenosti mezi vzorky v 2D prostoru přes více (ale ne mnoho!) gradientů

# PROBLÉM VZÁCNÝCH DRUHŮ

- názor, že mají příliš velký relativní vliv na výslednou ordinaci v unimodálních metodách (Legendre and Gallagher 2001)
- Greenacre (2013) naopak demonstruje, že příspěvek vzácných druhů k  $\chi^2$  vzdálenostem je malý
- downweighting rare species
  - (nepříliš) sníží početnosti „vzácných“ druhů
  - vzácnost stanovena pomocí inverzního Simpsonova indexu
  - lweigh parameter v cca()/decorana()
  - Diskuze o downweighting na:

<http://r.789695.n4.nabble.com/ordination-in-vegan-what-does-downweight-do-td4010352.html>

- Obecně je dobré odstranit druhy s 1 výskytem – představují prakticky jen neinterpretovatelný šum v datech
- U velkých dat (tisíce vzorků) s omezenou kvalitou (fytocenologické databáze) i druhy s více výskyty –  $< 10$ ,  $< 20$  apod.
- Odstranění vzácných druhů může „stvořit“ prázdné vzorky!
  - Ty je pak potřeba vymazat