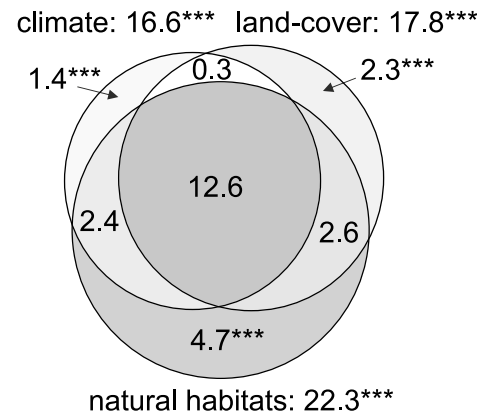
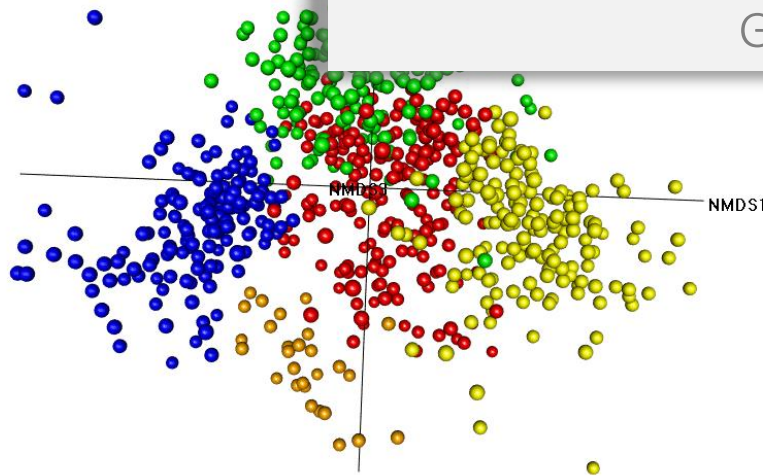


Metody fyzické geografie 3: Biogeografie & ekologie

Jan Divíšek
Geografický ústav & Ústav botaniky a zoologie



I TÝ SI ZAPIŠ NOVÝ
PŘEDMĚT Z 8055
METODY FYZICKÉ
GEOGRAFIE 3!



Ordinační metody

Ordinance – z německého *Ordnung* – uspořádat, vytvořit pořádek

Ordinace

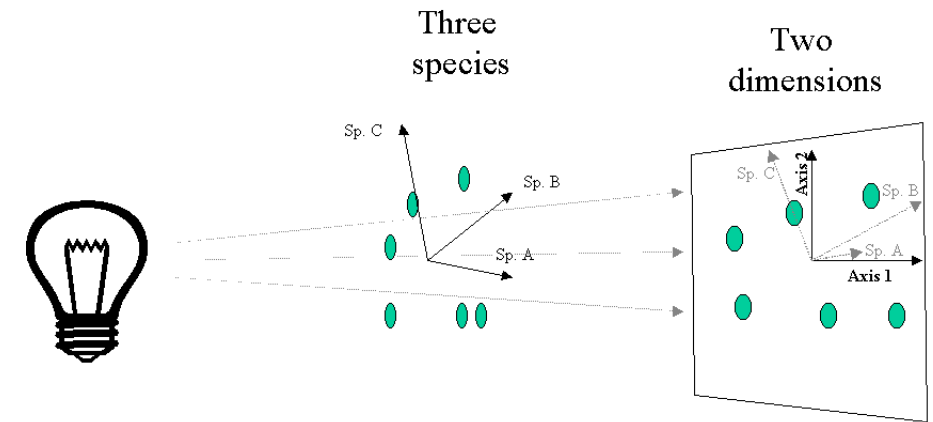
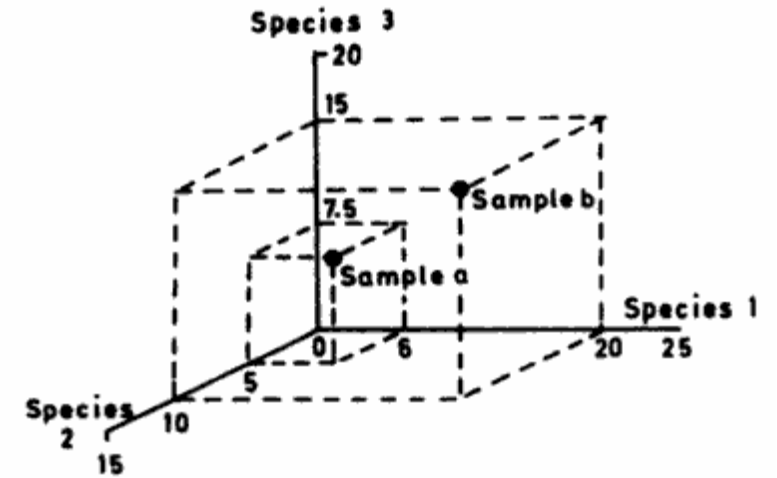
- Ekologické proměnné jsou často korelované
 - dvě korelované proměnné do jisté míry „říkají stejnou informaci“ – jsou **redundantní**
 - míra redundance závisí na míře korelace

→ ordinace agreguje soubor více či méně korelovaných proměnných do souboru vzájemně nekorelovaných ordinačních os
- Vzorčky jsou si vzájemně více či méně podobné hodnotami měřených proměnných
 - ordinace se snaží umístit objekty (vzorčky) v ordinačním prostoru tak, aby jejich rozmístění podél ordinačních os reflektovalo jejich vzájemnou podobnost
- Odhaluje nezávislé gradienty v datech a zjednodušuje tak jejich interpretaci
- *Pořádám exkurzi pro nadšené botaniky a chci, aby byla co nejzajímavější – kudy by měla vést trasa exkurze, abychom viděli co nejvíce různých společenstev a tedy co nejvíce druhů?*



Ordinance – princip

- Každý vzorek (případ, snímek) můžeme popsat jeho pozicí v mnohorozměrném prostoru (n dimenzí)
 - osami tohoto prostoru jsou jednotlivé proměnné (deskriptory)
- Více než 3D prostor (tj. prostor tvořený více než 3 deskriptory) je pro nás příliš složitý a neuchopitelný
- Ordinační analýza se tento problém snaží řešit redukcí dimenzionality dat „sloučením“ korelovaných proměnných do menšího počtu „faktorových“ proměnných → **osy ordinačního prostoru**



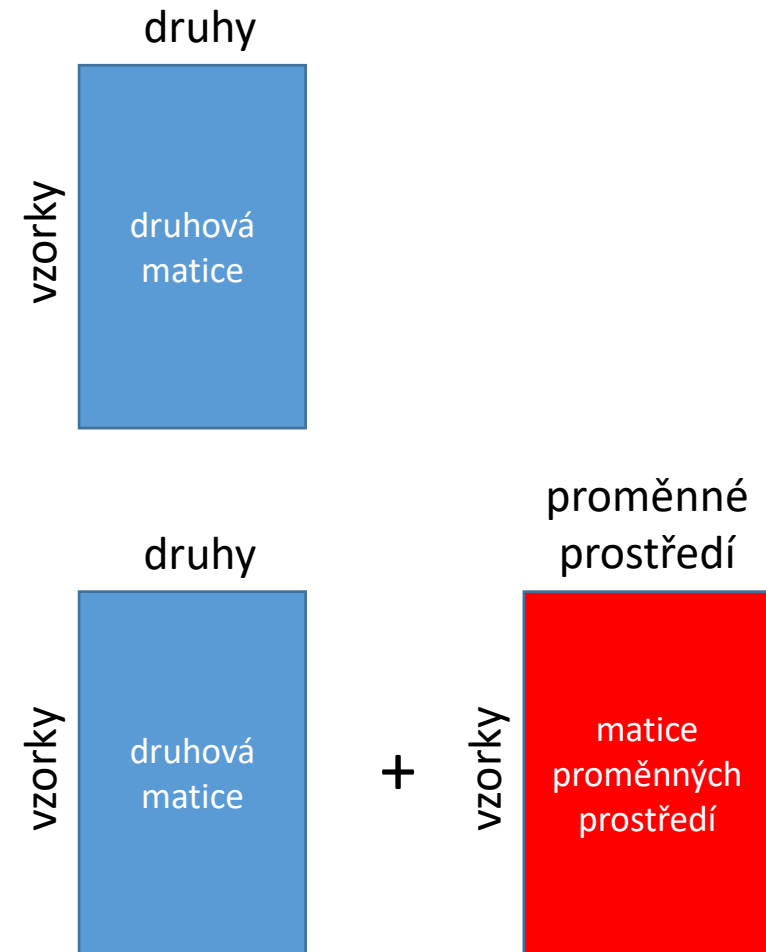
Nepřímá vs. přímá ordinace

Nepřímá ordinace (*unconstrained ordination*)

- Pouze druhová matice
- Ordinační osy (*unconstrained axes*) = směry největší variability dat
- Popis dat a generování hypotéz

Přímá ordinace (*constrained ordination*)

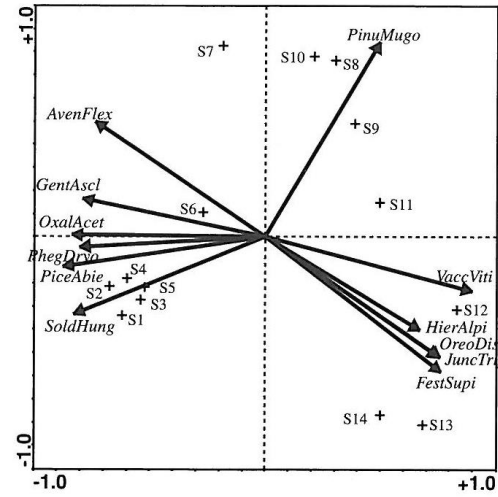
- Druhová matice a matice proměnných prostředí
- Ordinační osy (*constrained axes*) = variabilita vysvětlitelná danými proměnnými
- Testování hypotéz



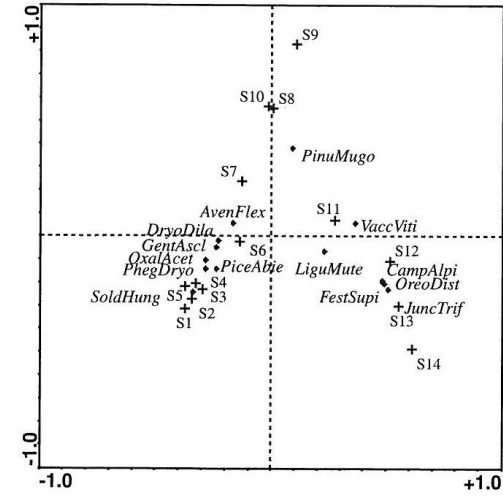
Ordinační diagram

Nepřímá ordinace

Lineární metoda
PCA

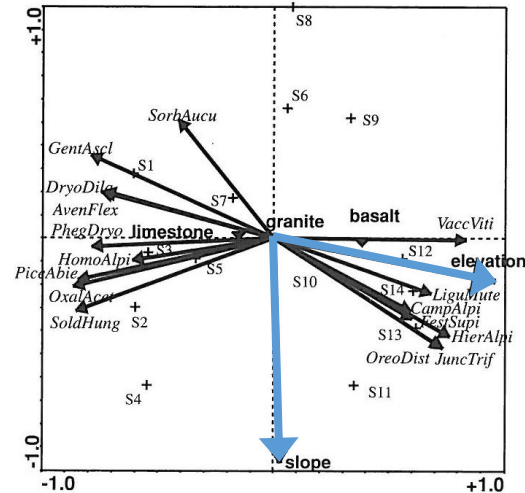


Unimodální metoda
CA

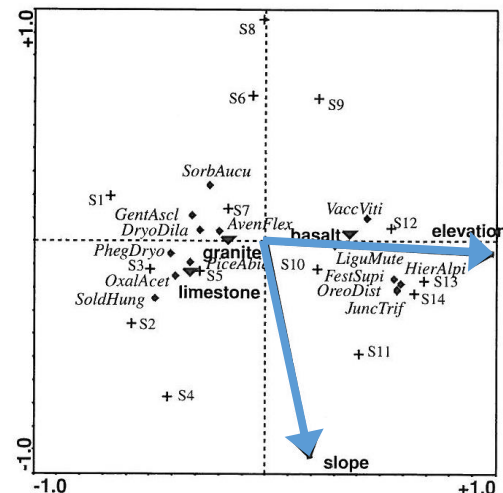


Přímá ordinace

RDA



CCA



Lepš & Šmilauer (2003)

Ordinační diagram – konvence v ekologii

- Zobrazení vzorků
 - Body
- Zobrazení druhů
 - Šipky v lineárních metodách
 - Body (centroidy) v unimodálních metodách
- Zobrazení ordinačních os
 - Vodorovná osa je vždy vyššího řádu (první)
- Typ ordinačního diagramu
 - Scatterplot – 1 typ dat (vzorky nebo druhy)
 - Biplot – 2 typy dat (např. vzorky a druhy)
 - Triplot – 3 typy dat (např. vzorky, druhy a proměnné prostředí v přímé ordinaci)

Vizualizace ordinací

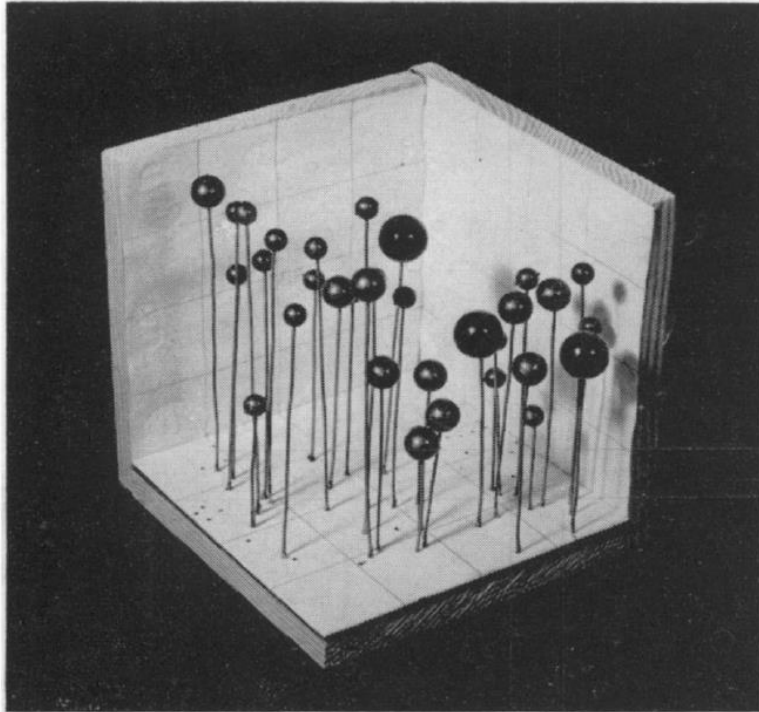
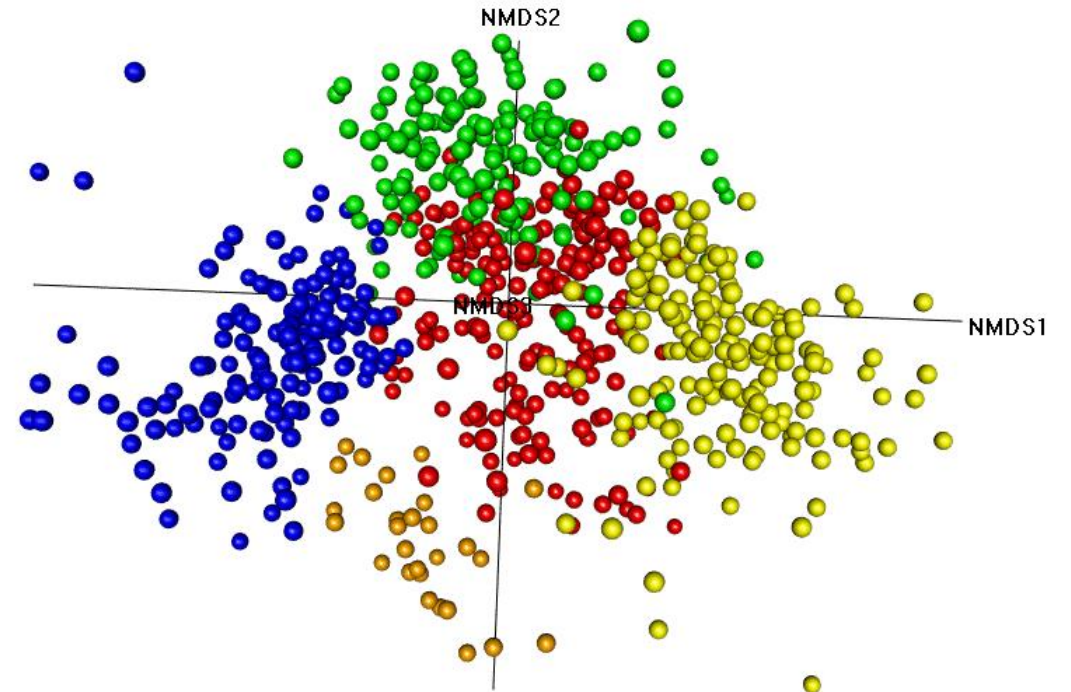


FIG. 7. Three-dimensional model of the dominance behavior of *Quercus borealis* within the ordination. The 3 sizes of spheres indicate the top 3 quartiles of dominance per acre. Stands of the lowest quartile and without the species are represented with holes which appear as dots in the figure. The x axis is on base of model at front from left to right; y axis on base from front to rear; z axis in vertical plane from below to above.

Bray & Curtis (1957)



Přehled metod ordinační analýzy

	<i>raw-data-based</i> (založené na primárních datech)		<i>transformation-based</i> (založené na transformovaných primárních datech)	<i>distance-based</i> (založené na distanční matici)
	<i>linear</i> (lineární)	<i>unimodal</i> (unimodální)		
<i>unconstrained</i> (nepřímé)	PCA (analýza hlavních komponent)	CA, DCA (korespondenční a detrendovaná korespondenční analýza)	tb-PCA (analýza hlavních komponent na transformovaných primárních datech)	PCoA (analýza hlavních koordinát) NMDS (nemetrické mnohorozměrné škálování)
<i>constrained</i> (přímé)	RDA (redundanční analýza)	CCA (kanonická korespondenční analýza)	tb-RDA (redundanční analýza na transformovaných primárních datech)	db-RDA (redundanční analýza založená na distanční matici)



Nepřímá ordinační analýza

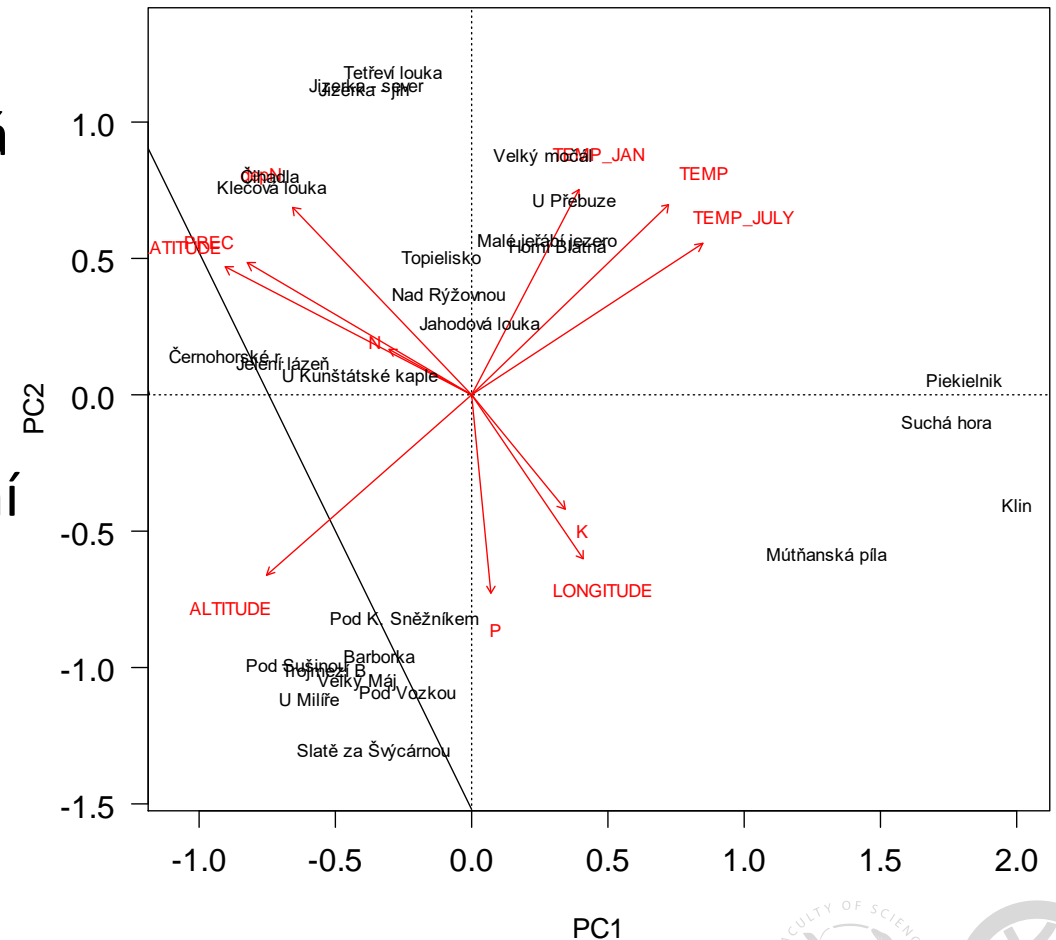
	<i>raw-data-based</i> (založené na primárních datech)		<i>transformation-based</i> (založené na transformovaných primárních datech)	<i>distance-based</i> (založené na distanční matici)
	<i>linear</i> (lineární)	<i>unimodal</i> (unimodální)		
<i>unconstrained</i> (nepřímé)	PCA (analýza hlavních komponent)	CA, DCA (korespondenční a detrendovaná korespondenční analýza)	tb-PCA (analýza hlavních komponent na transformovaných primárních datech)	PCoA (analýza hlavních koordinát) NMDS (nemetrické mnohorozměrné škálování)
<i>constrained</i> (přímé)	RDA (redundanční analýza)	CCA (kanonická korespondenční analýza)	tb-RDA (redundanční analýza na transformovaných primárních datech)	db-RDA (redundanční analýza založená na distanční matici)

PCA (*Principal Component Analysis; analýza hlavních komponent*)

- Vstupem je tabulka vzorků (řádky) a proměnných (sloupce)
- Pokud dělám PCA na druhových datech, která předem transformuji (např. Hellingerovou transformací) → **tb-PCA**
- Datový soubor by měl mít více vzorků než proměnných
- Snaží se najít hlavní gradienty v datech (hlavní komponenty, dimenze), které zahrnují větší podíl variability než připadá na původní proměnné
- Proměnné (deskriptory, sloupce) se zobrazují jako vektory a vzorky (řádky) jako body

`rda {vegan}` nebo `pca {labdsv}` nebo `prcomp ()`

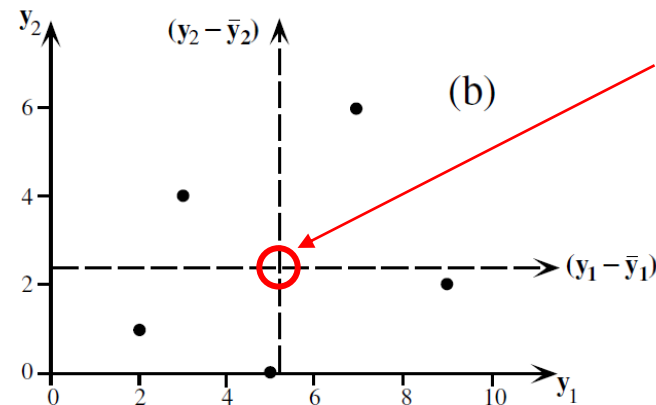
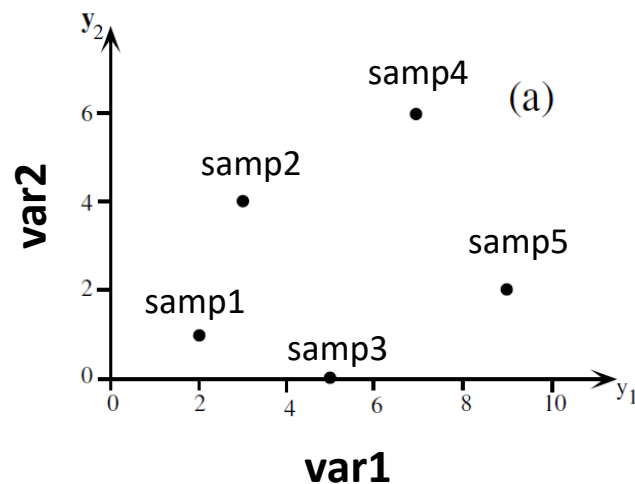
Environmentální data z rašelinišť ČR a Slovenska



Data z rašelinišť - Jiroušek (2008). Diplomová práce

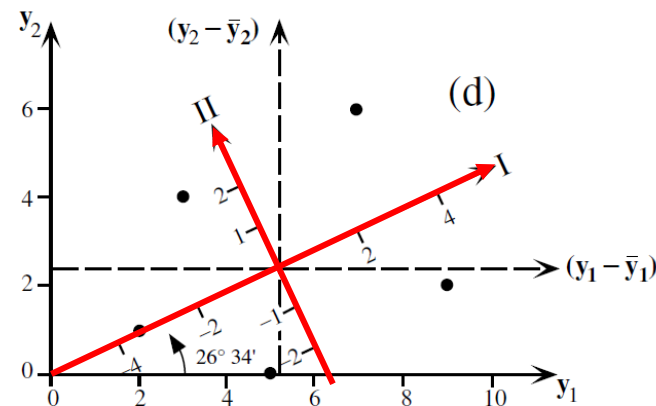
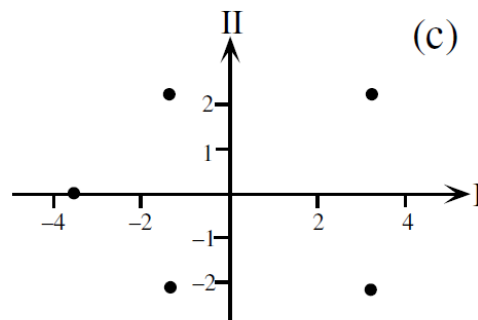
PCA – princip výpočtu

	var1	var2
samp1	2	1
samp2	3	4
samp3	5	0
samp4	7	6
samp5	9	2



centroid v n -
dimenzionálním
prostoru (počet
dimenzí = počtu
proměnných)

- a) rozmístění vzorků v prostoru definovaném proměnnými
- b) výpočet těžiště shluku
- c) centrování os
- d) rotace os



Výsledek PCA v R (vegan)

27 rašelinišť, 11 environmentálních proměnných

Celkový rozptyl dat

```
Call: rda(X = scale(env.ras))
```

	Inertia	Rank
Total	11	
Unconstrained	11	11

Inertia is variance

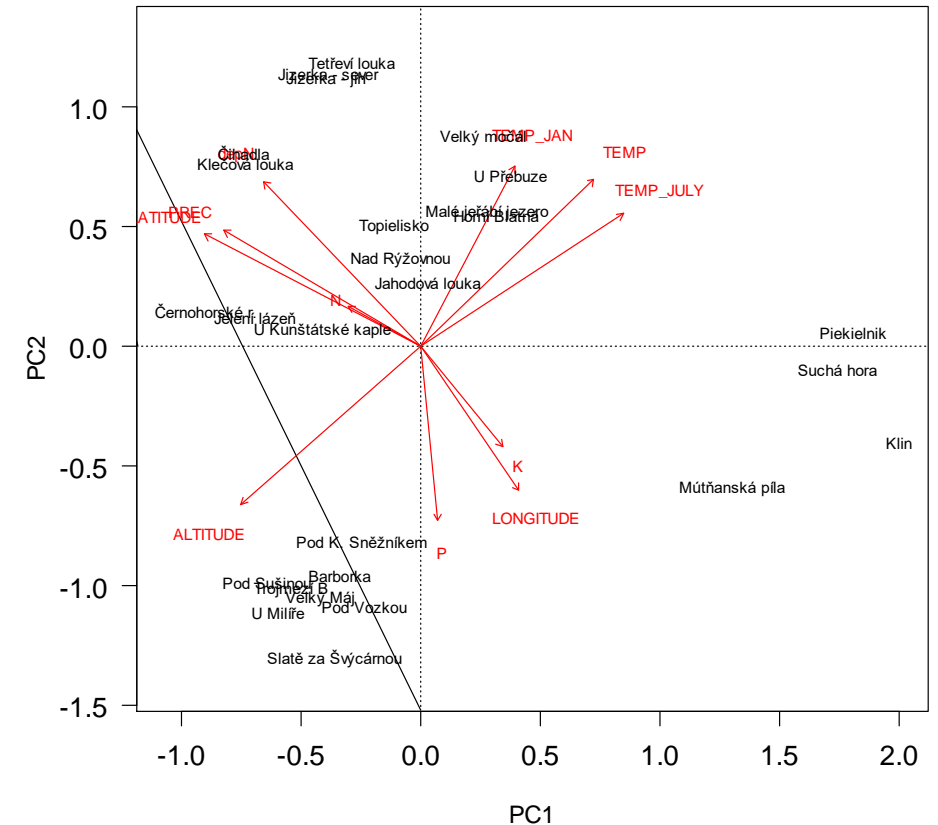
Rozptyl dat vysvětlený neomezenými osami

Eigenvalues for unconstrained axes:

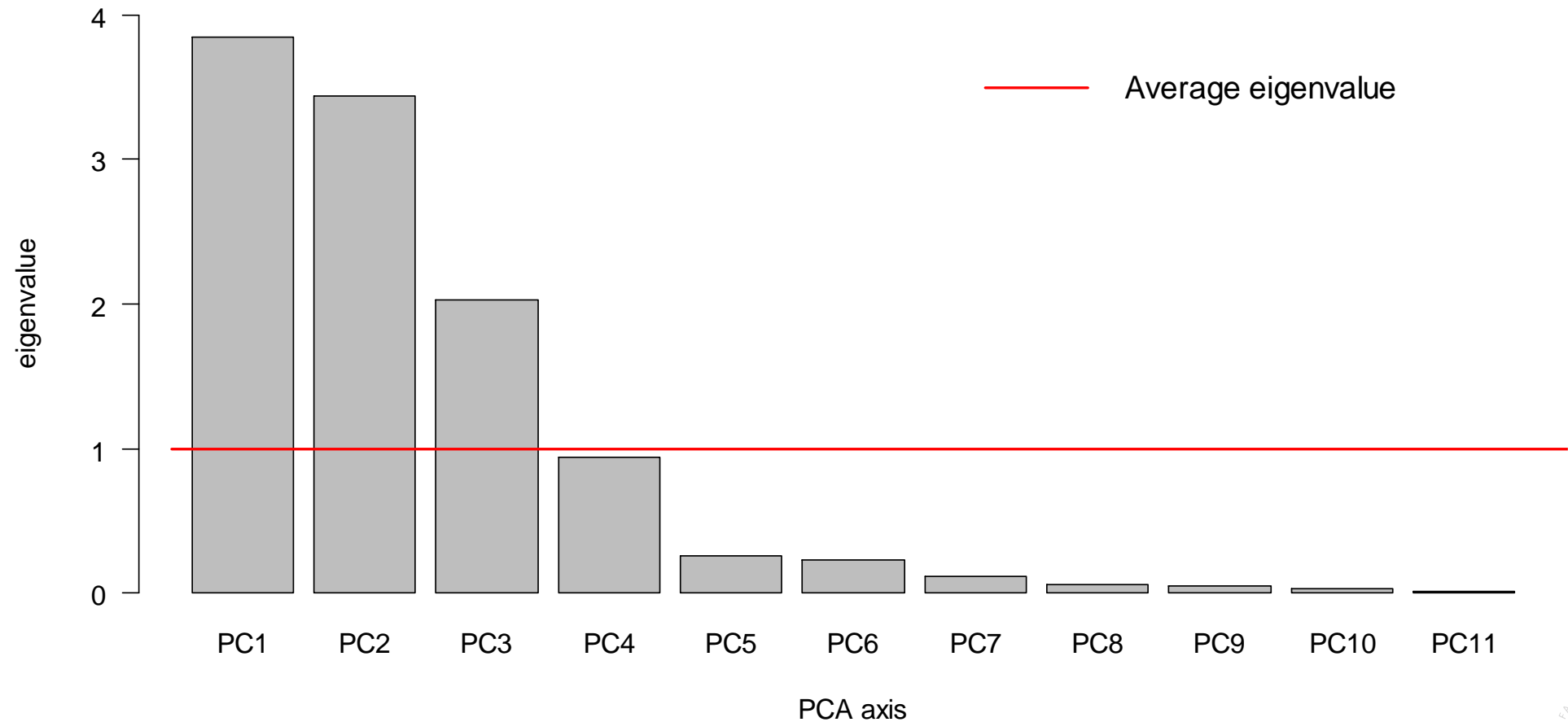
PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
3.847	3.439	2.027	0.937	0.259	0.230	0.117	0.058	0.045	0.032	0.010

Charakteristická (vlastní) čísla pro jednotlivé osy PCA.
eigenvalue/total inertia = variabilita vysvětlená danou osou

Environmentální data z rašelinišť ČR a Slovenska

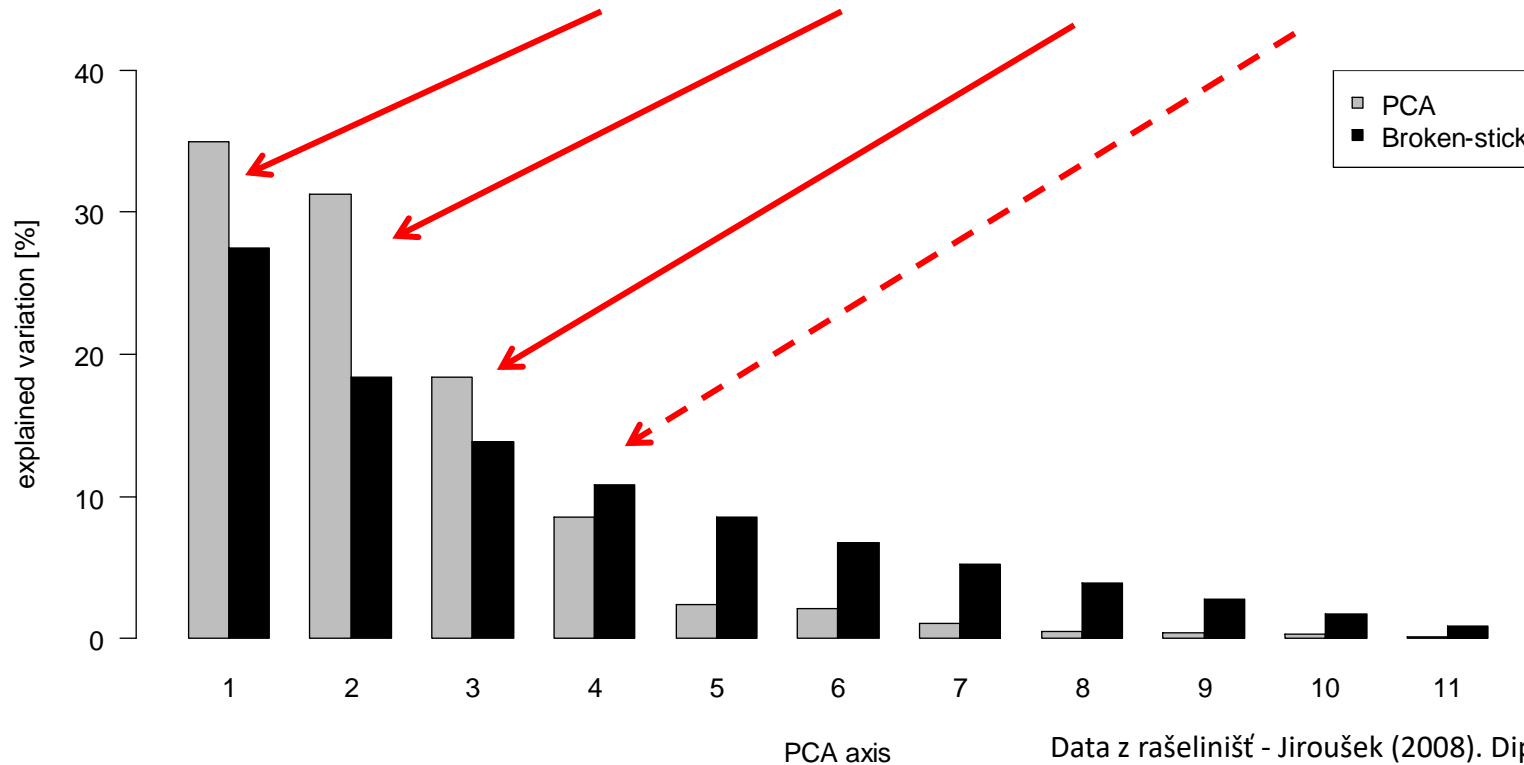


Které osy PCA jsou důležité?



Které osy PCA jsou důležité?

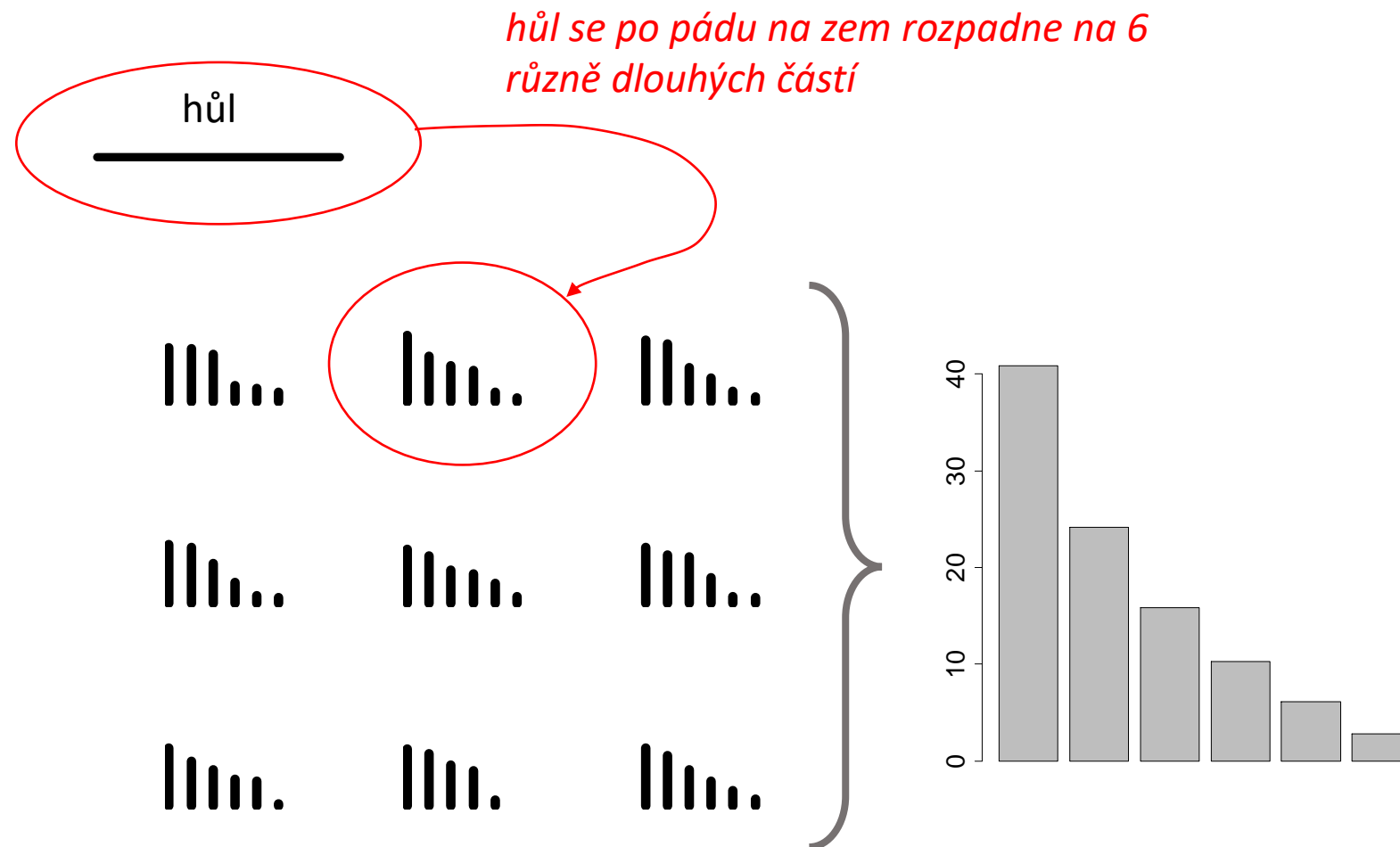
	1	2	3	4	5	6
eigenvalue	3.847153	3.439119	2.026698	0.9370375	0.2591531	0.2297798
percentage of variance	34.974120	31.264717	18.424530	8.5185231	2.3559372	2.0889069
cumulative percentage of variance	34.974120	66.238837	84.663367	93.1818902	95.5378274	97.6267344
broken-stick percentage	27.453430	18.362521	13.817067	10.7867637	8.5140365	6.6958547
broken-stick cumulative %	27.453430	45.815952	59.633018	70.4197822	78.9338187	85.6296734
% > bs%	1.000000	1.000000	1.000000	0.0000000	0.0000000	0.0000000
cum% > bs cum%	1.000000	1.000000	1.000000	1.0000000	1.0000000	1.0000000



Data z rašeliníšť - Jiroušek (2008). Diplomová práce

Broken-stick model

PCAsignificance {BiodiversityR}



Metody nepřímé ordinace založené na distanční matici

	<i>raw-data-based</i> (založené na primárních datech)		<i>transformation-based</i> (založené na transformovaných primárních datech)	<i>distance-based</i> (založené na distanční matici)
	<i>linear</i> (lineární)	<i>unimodal</i> (unimodální)		
<i>unconstrained</i> (nepřímé)	PCA (analýza hlavních komponent)	CA, DCA (korespondenční a detrendovaná korespondenční analýza)	tb-PCA (analýza hlavních komponent na transformovaných primárních datech)	PCoA (analýza hlavních koordinát) NMDS (nemetrické mnohorozměrné škálování)
<i>constrained</i> (přímé)	RDA (redundanční analýza)	CCA (kanonická korespondenční analýza)	tb-RDA (redundanční analýza na transformovaných primárních datech)	db-RDA (redundanční analýza založená na distanční matici)



PCoA (*Principal Coördiante Analysis, analýza hlavních koordinát*)

- Synonymem pro PCoA je MDS – Metric Dimensional Scaling
- Vstupem je matice vzdáleností (nepodobností) vzorků
 - Pokud zvolím Euklidovskou vzdálenost → identické s PCA
 - Pokud zvolím chi-square vzdálenost → obdoba korespondenční analýzy (CA)
- Umístí objekty na základě jejich vzdáleností (distancí) do Euklidovského prostoru (tvořeného souřadnicemi – skóre vzorků na osách)
- Použití nemetrických distancí může způsobit výskyt os ze zápornou hodnotou eigenvalue
 - korekce matice vzdáleností

```
cmdscale {vegan}
```

```
pcoa {ade4}
```



PCoA – evropská města

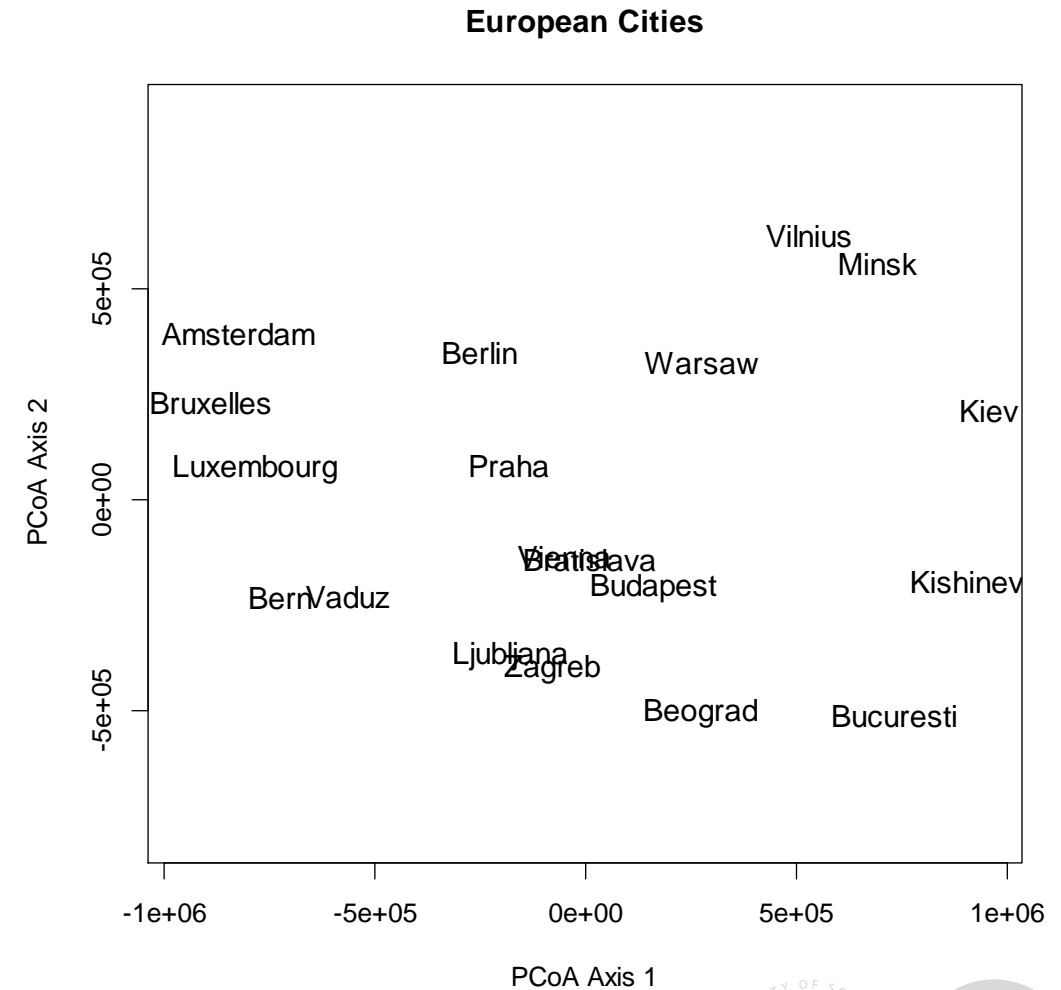
Souřadnice měst

	X	Y
Vilnius	4161503	6111292
Minsk	4324570	6048501
Berlin	3386469	5822208
Amsterdam	2812692	5853176
Warsaw	3910444	5807753
Kiev	4598405	5706653
Praha	3461154	5552876
Luxembourg	2858588	5535341
Bratislava	3658489	5337395

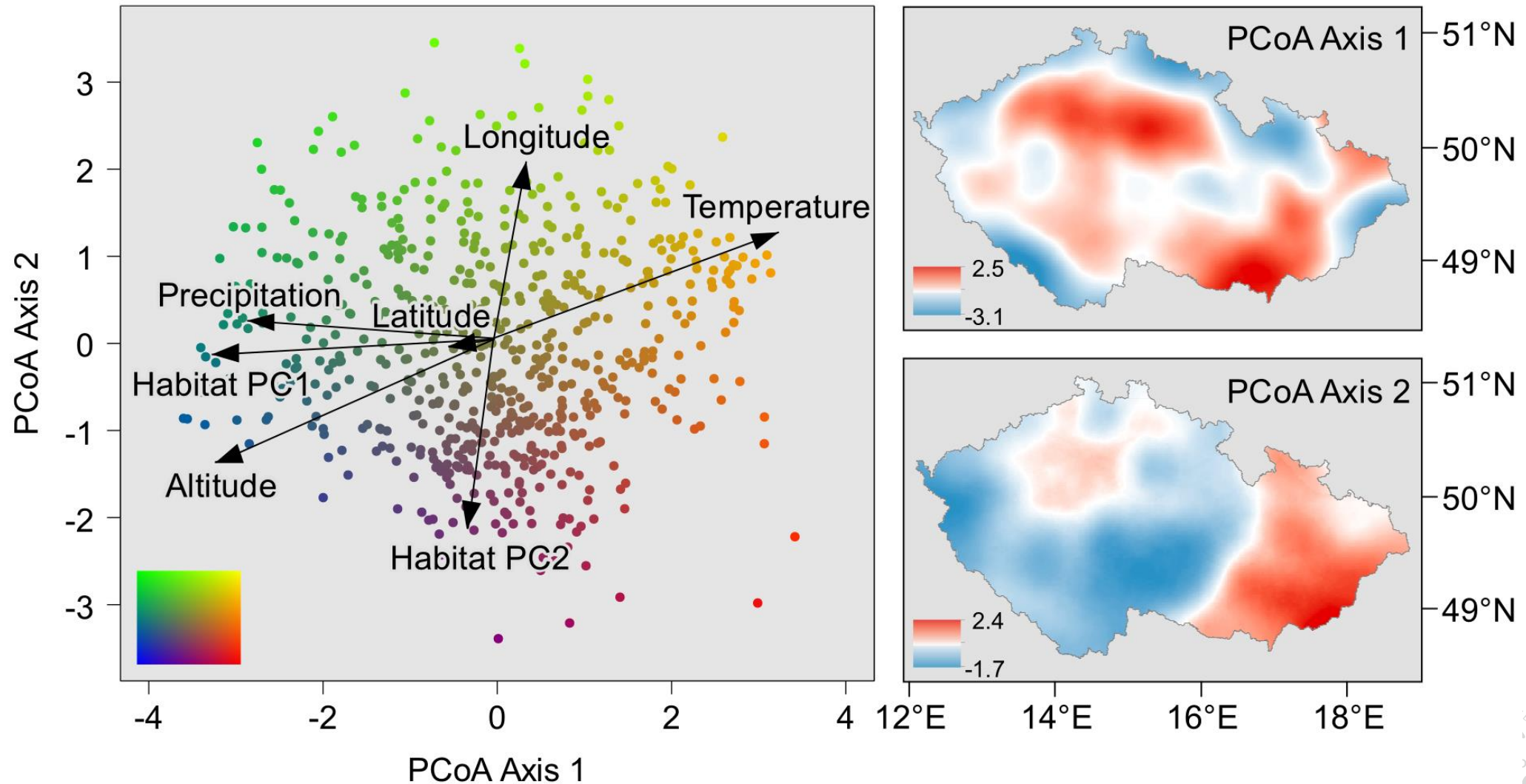


Matrice vzdáleností

	Vilnius	Minsk	Berlin	Amsterdam	Warsaw
Vilnius	0.0	174738.8	827192.1	1373286.3	393911.9
Minsk	174738.8	0.0	965008.3	1524443.2	479019.2
Berlin	827192.1	965008.3	0.0	574612.8	524174.4
Amsterdam	1373286.3	1524443.2	574612.8	0.0	1098692.1
Warsaw	393911.9	479019.2	524174.4	1098692.1	0.0



PCoA – společenstva terestrických obratlovců ČR

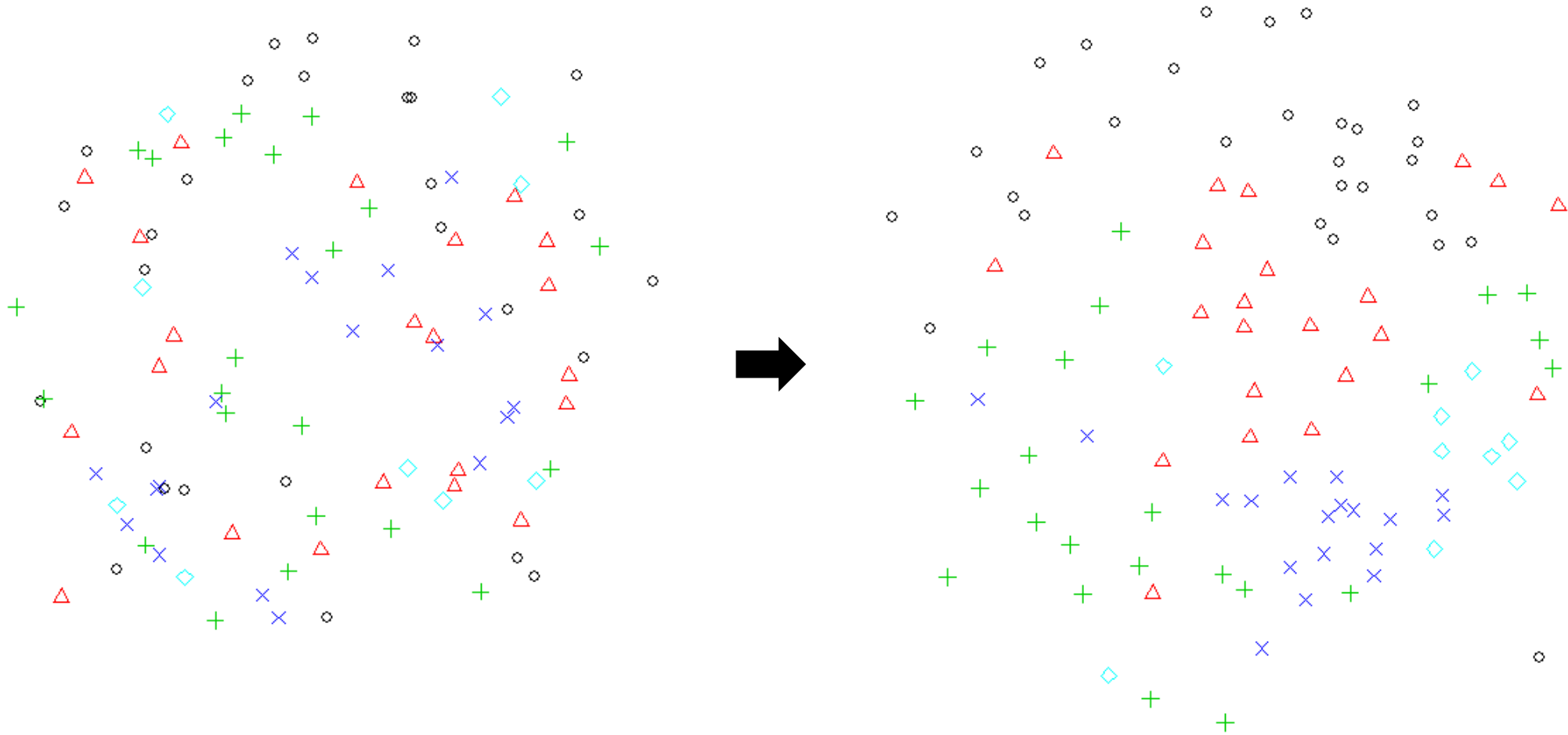


NMDS (*Non-metric Multidimensional Scaling, nemetrické mnohorozměrné škálování*)

- Nemetrická varianta PCoA (nepracuje přímo s distancemi mezi vzorky, ale s jejich pořadím)
- Vstupem je matice vzdáleností (nepodobností) vzorků
 - jakýkoliv index nepodobnosti
- **Iterativní algoritmus**, který nemusí pokaždé dojít ke stejnému výsledku (lokální optima)
- Nutno určit počet dimenzí, se kterými bude metoda pracovat
- Při větším množství dat VELMI časově náročná
- Na rozdíl od PCoA optimalizuje výsledné vzdálenosti mezi vzorky do několika málo (dvě – tři) dimenzí

metaMDS {vegan}

NMDS (*Non-metric Multidimensional Scaling, nemetrické mnohorozměrné škálování*)



náhodné rozmístění vzorků v prostoru

rozmístění vzorků v prostoru respektuje jejich nepodobnost

NMDS – evaluace

stressplot {vegan}

Call:

```
metaMDS(comm = beta.j, k = 2)
```

global Multidimensional Scaling using monoMDS

Data: beta.j

Distance: jaccard

Dimensions: 2

Stress: 0.1843335

Stress type 1, weak ties

Two convergent solutions found after 2 tries

Scaling: centring, PC rotation

Species: scores missing

Podle Clarke & Warwick (2001) pro stress value přibližně platí:

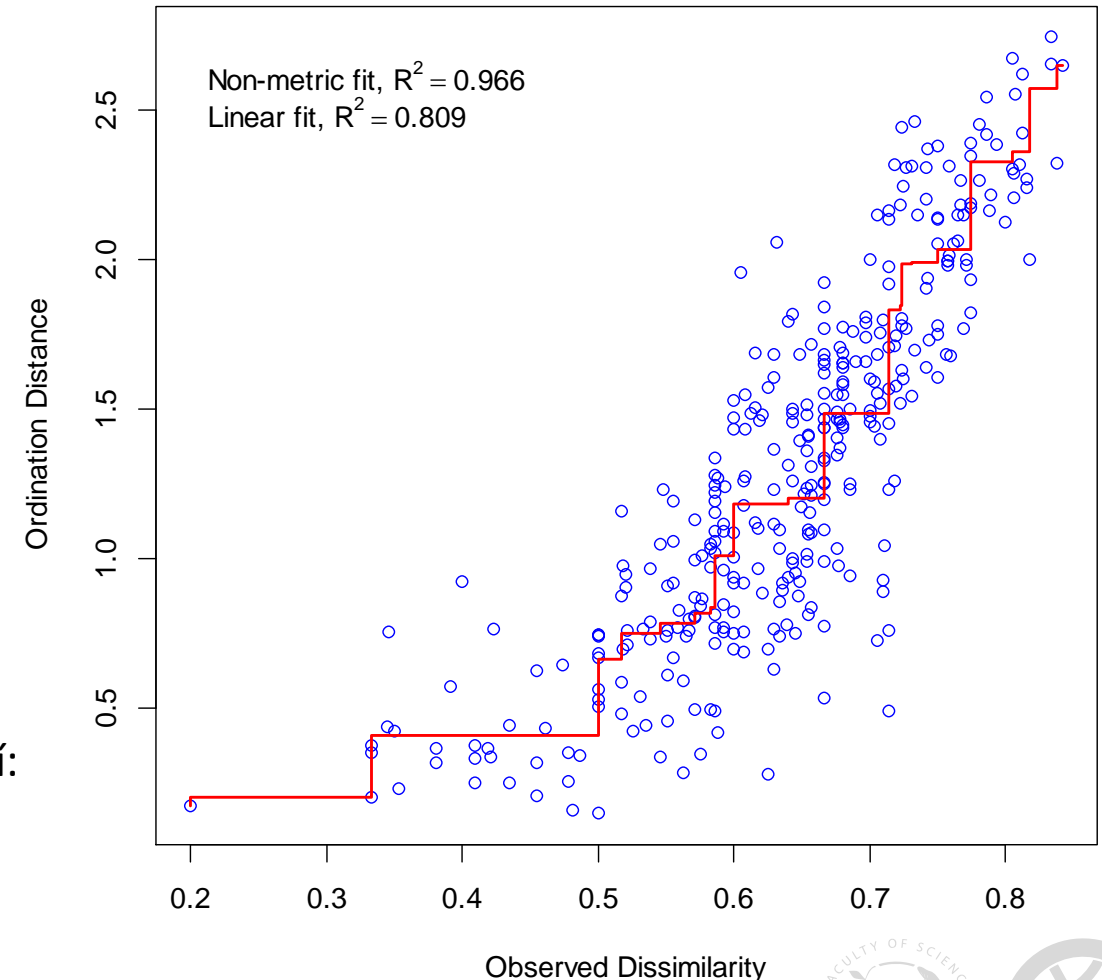
< 0.05 – vynikající

< 0.1 – výborný

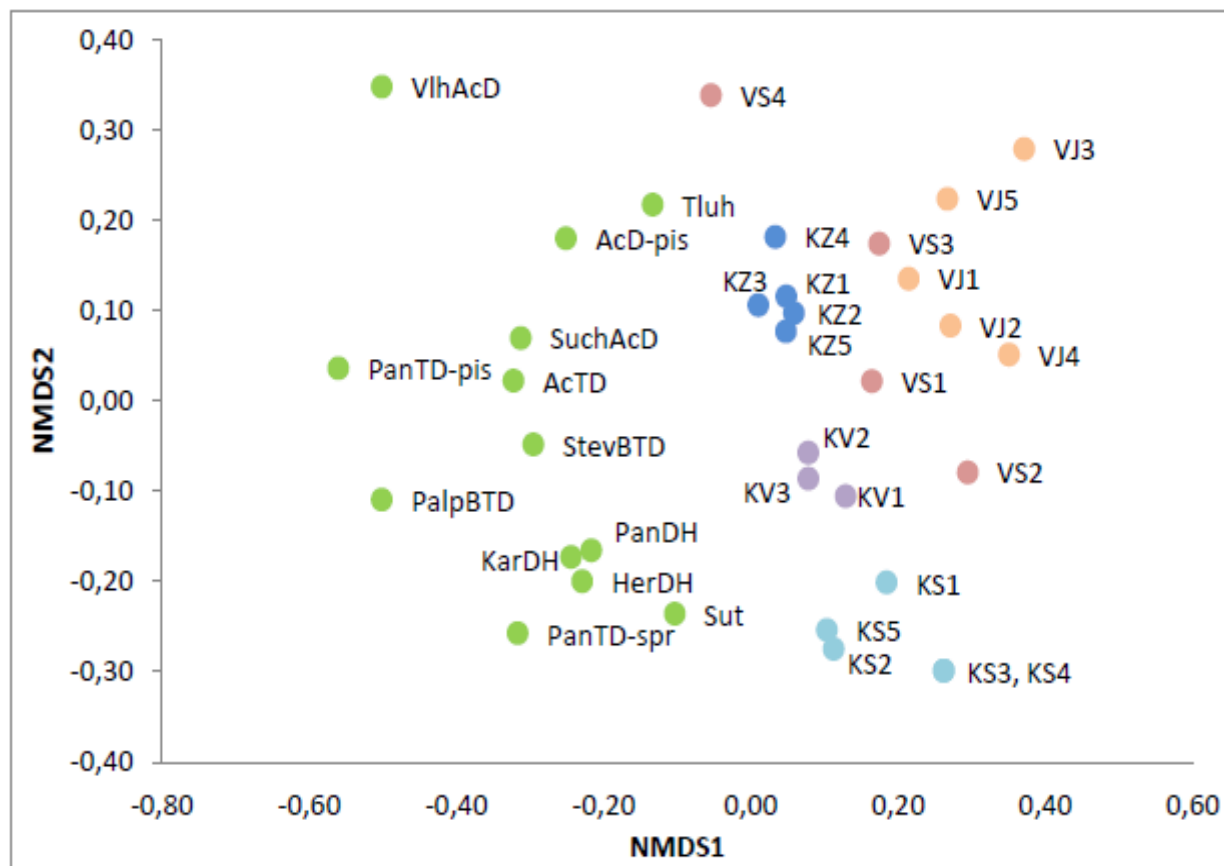
< 0.2 – dobrý

> 0.3 – špatný

Shepardův diagram

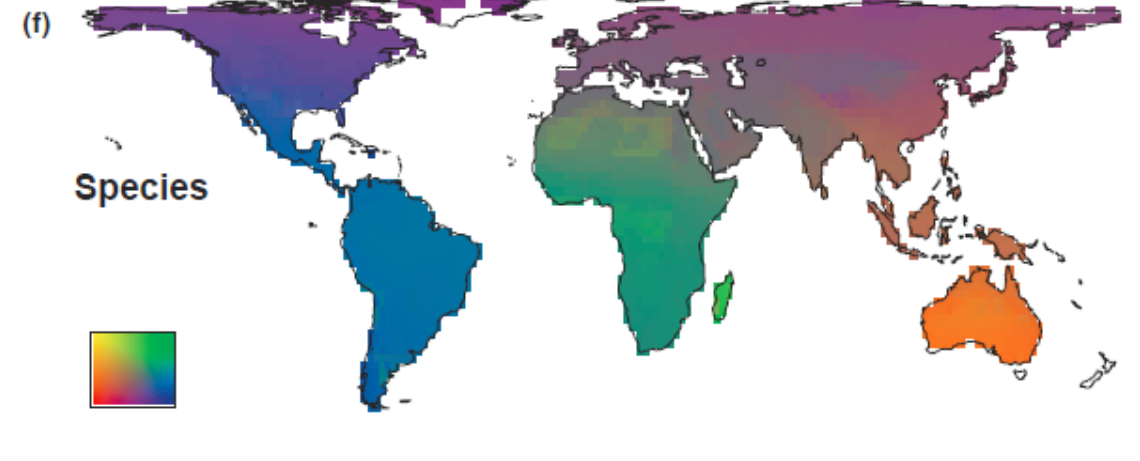
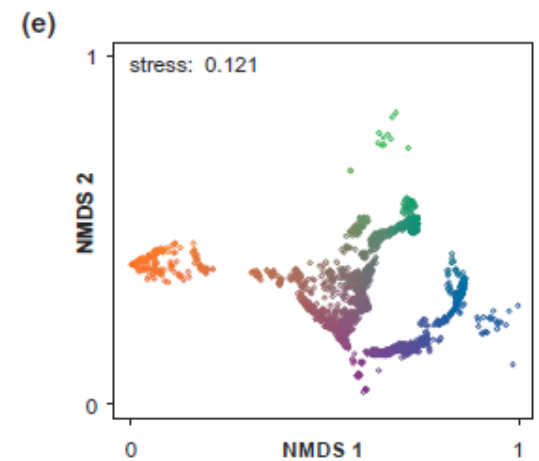
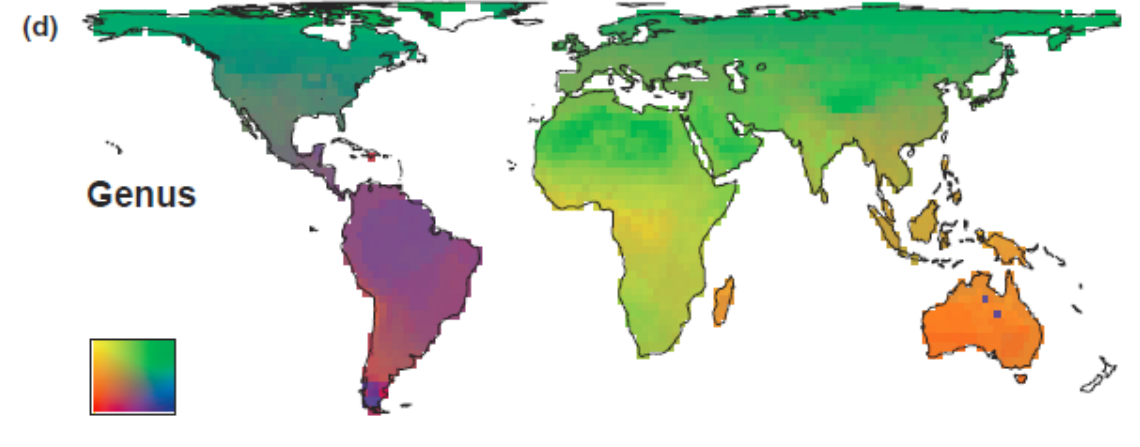
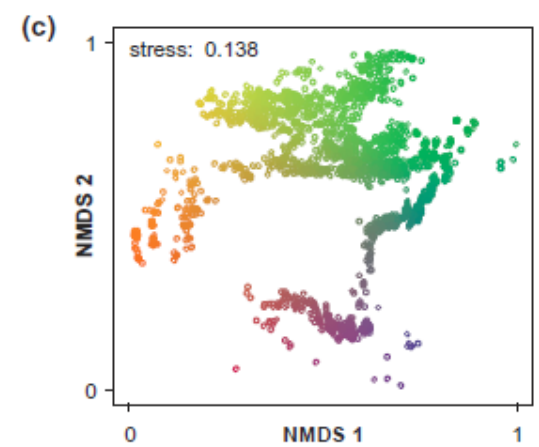
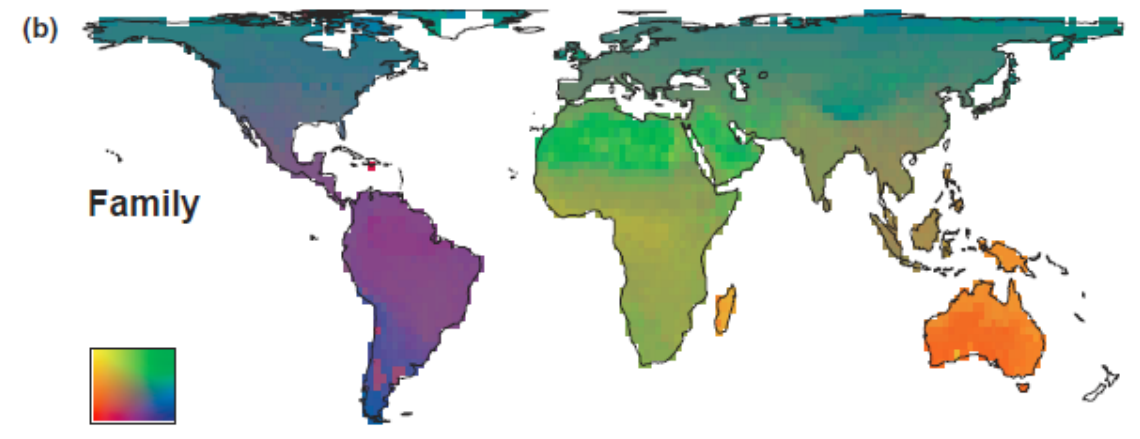
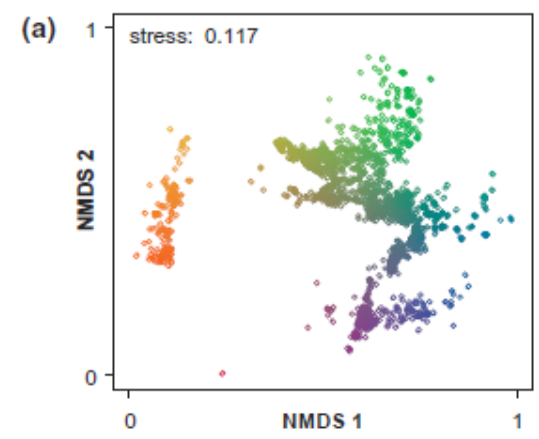


NMDS vegetačních snímků z biokoridorů



Obr. 14. Jaccardův index podobnosti bylinné skladby fytoecologických snímků z biokoridorů ve Vracově a Křižanovicích a referenčních biotopů znázorněný pomocí NMDS

Podobnost společenstev savců



Kreft & Jetz (2010)

Tři alternativní přístupy k nepřímé ordinaci

(a) Klasický přístup

Y = Raw data
(sites × species)

Short gradients: CA or PCA
Long gradients: CA

(b) Transformace dat (např. Hellingerova) (tb-PCA)

Raw data
(sites × species)

Y=Transformed data
(sites × species)

PCA

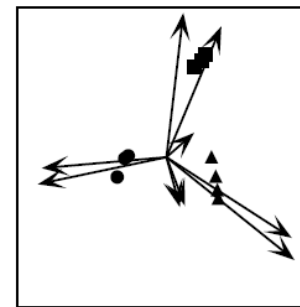
(c) Přes matici nepodobností (PCoA, NMDS)

Raw data
(sites × species)

Distance matrix

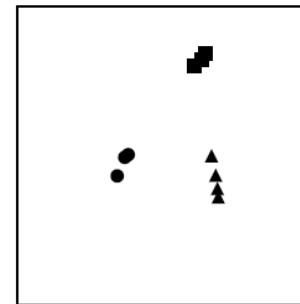
PCoA

Ordination biplot



Representation of elements:
Species = arrows
Sites = symbols

Ordination of sites

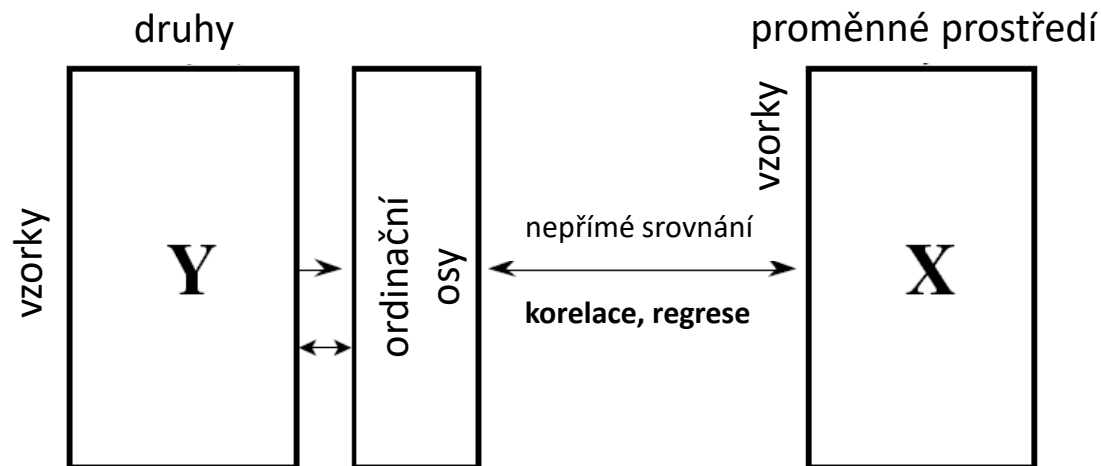


Representation of elements:
Sites = symbols

Proměnné prostředí v nepřímé ordinaci

Dva alternativní přístupy k použití vysvětlujících proměnných v ordinaci

**Pasivní „nařizování“
vysvětlujících
proměnných do
nepřímé ordinace**

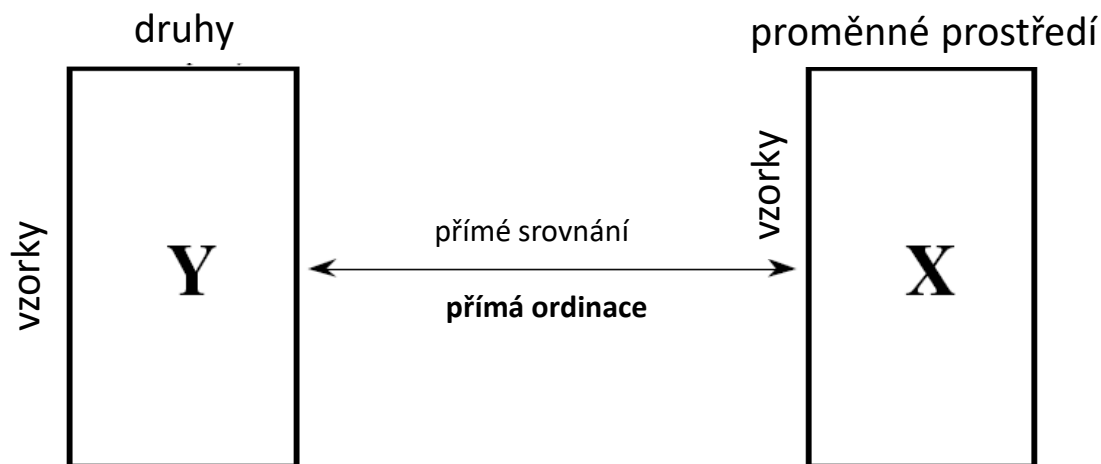


matic:

Y – druhové složení

X – proměnné prostředí

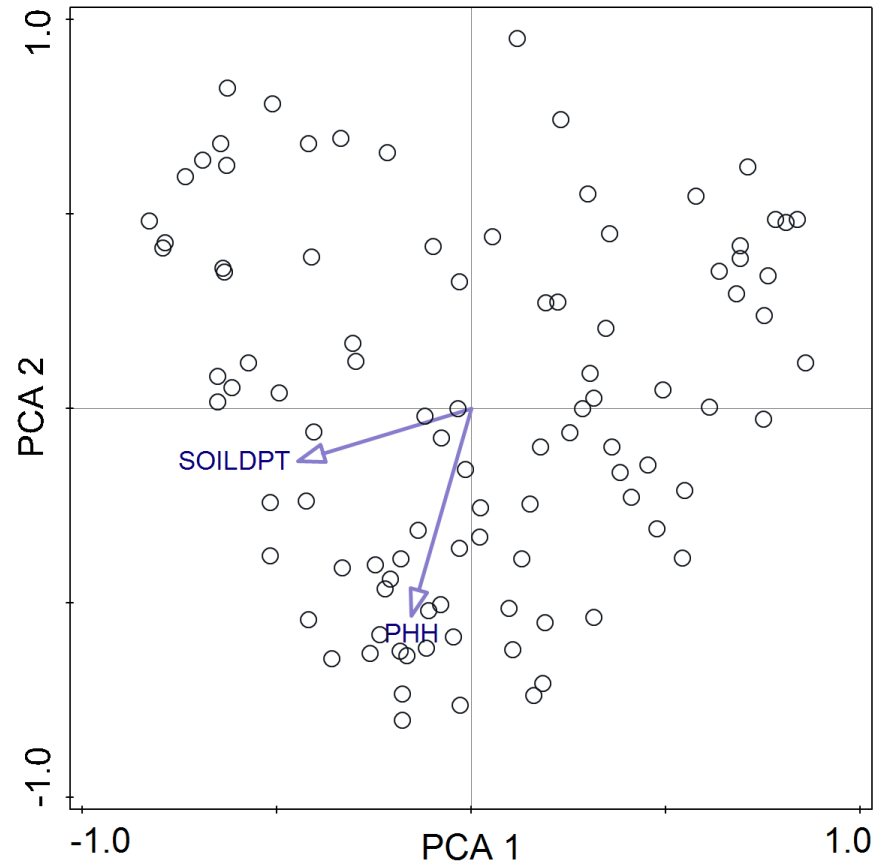
Přímá ordinace



**oba přístupy jsou
relevantní a navzájem
se doplňují!**

Pasivní proměnné prostředí v nepřímé ordinaci

- Pasivně promítnuté proměnné prostředí v nepřímé ordinaci – korelace (regrese) s ordinačními osami



Korelace (regrese) ordinačních os s proměnnými prostředí

matice druhových dat

	spe1	spe2	spe3	spe4	...
sam1					...
sam2					...
sam3					...
sam4					...
...

skóre vzorků na první a druhé ose PCA

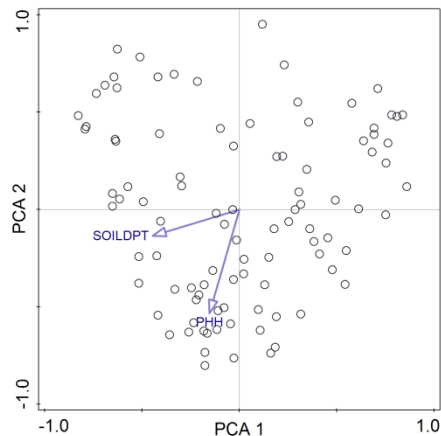
	PCA 1	PCA 2
sam1		
sam2		
sam3		
sam4		
...

proměnné prostředí

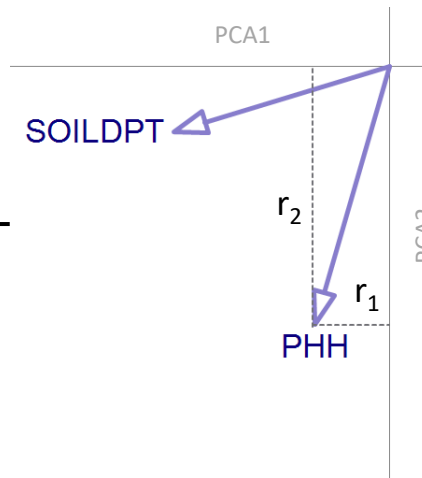
	PH	SOILDPT
sam1		
sam2		
sam3		
sam4		
...

PCA

korelace



ordinační diagram PCA



vztah proměnných prostředí (vektory) a ordinačních os

	PH	SOILDPT
PCA 1	r_1	r_3
PCA 2	r_2	r_4

korelace proměnných prostředí a ordinačních os

Korelace (regrese) ordinačních os s proměnnými prostředí

Korelace mezi proměnnou prostředí a skóre vzorků na ordinačních osách

- Pouze v ordinacích kde jsou skóre vzorků standardizované na jednotkovou varianci (PCA)
- V ostatních ordinacích, kde se variance os od sebe liší, je třeba použít (váženou) **mnohonásobnou regresi**:

$$\text{env} = \alpha + \beta_1 \times \text{score1} + \beta_2 \times \text{score2}$$

proměnná prostředí
(např. nadmořská výška)

Intercept = 0
(všechny proměnné
jsou centrovány)

Regresní koeficient

Skóre vzorků na první
ordinační ose



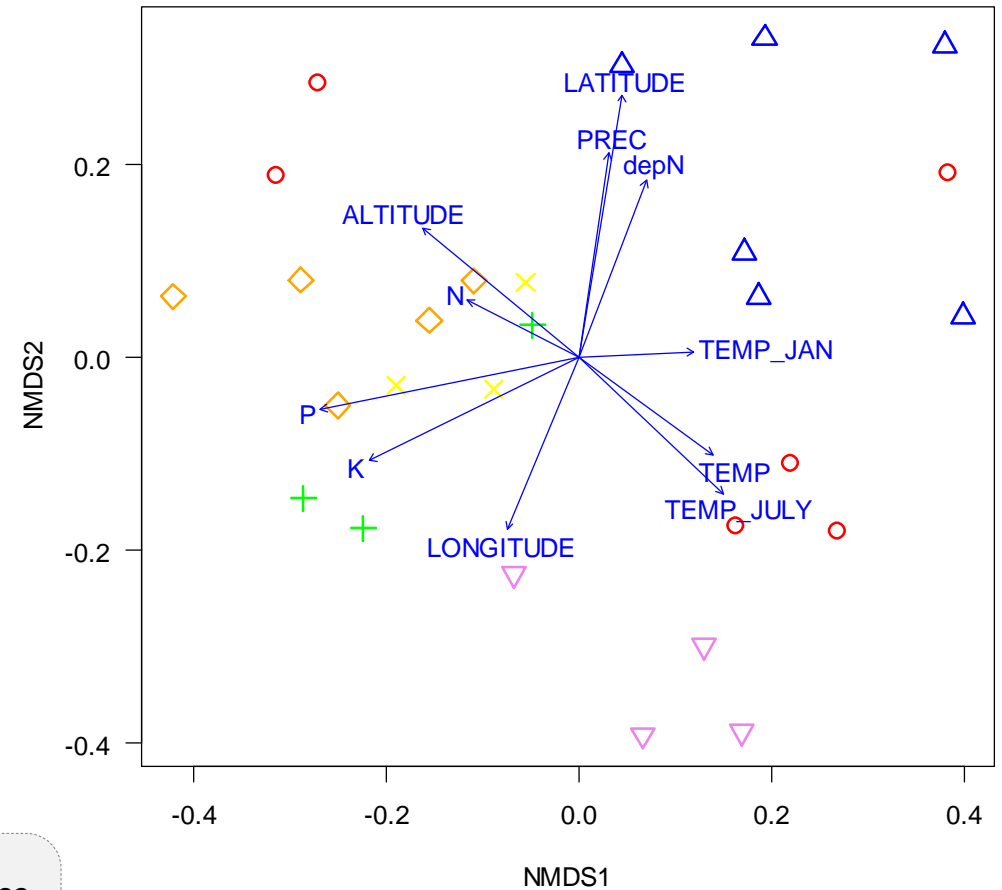
Pasivní promítnutí proměnných prostředí do NMDS

```
envfit {vegan}
```

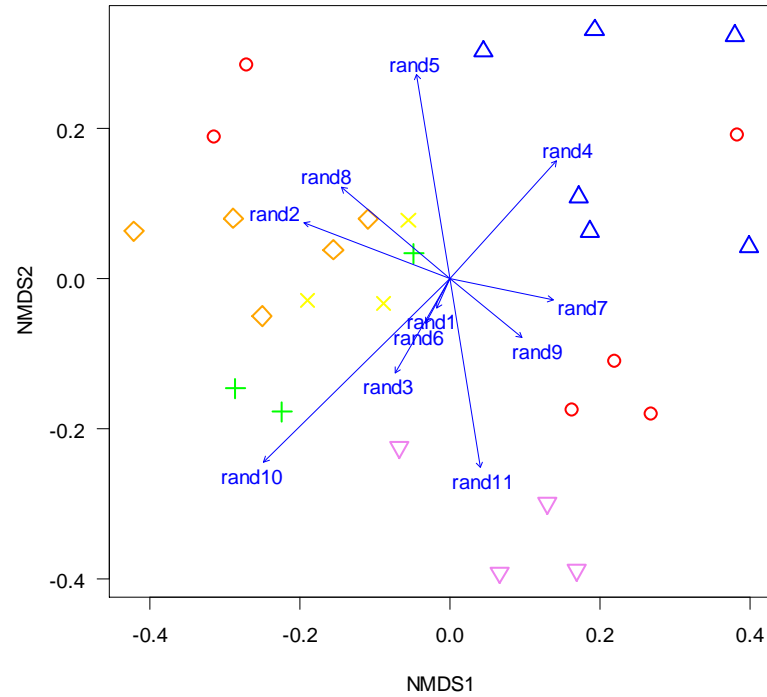
	NMDS1	NMDS2	r2	Pr (>r)	
LATITUDE	0.16394	0.98647	0.6000	0.001	***
LONGITUDE	-0.38731	-0.92195	0.2948	0.015	*
ALTITUDE	-0.77129	0.63648	0.3488	0.004	**
TEMP_JULY	0.72650	-0.68717	0.3341	0.009	**
TEMP_JAN	0.99907	0.04306	0.1121	0.246	
TEMP	0.80798	-0.58921	0.2329	0.045	*
PREC	0.14608	0.98927	0.3616	0.008	**
depN	0.35755	0.93389	0.3060	0.014	*
N	-0.89075	0.45449	0.1327	0.185	
P	-0.98021	-0.19794	0.5928	0.001	***
K	-0.89756	-0.44088	0.4656	0.001	***

Kolik variability vysvětlí první dvě osy NMDS z celkové variability dané environmentální proměnné (udává délku vektoru v ordinačním diagramu).

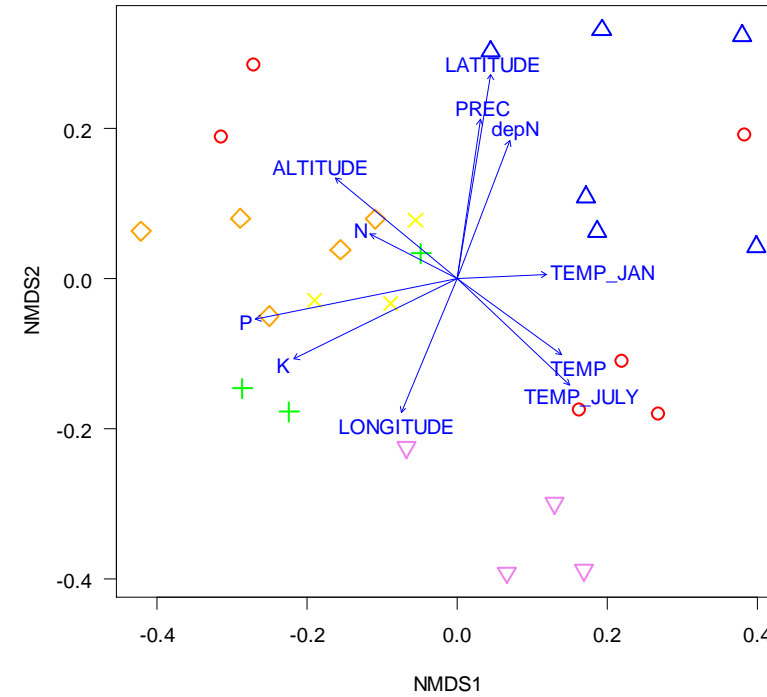
Statistická signifikance na základě permutačního testu.



Testování pasivních proměnných



	NMDS1	NMDS2	r2	Pr (>r)
rand1	-0.39339	-0.91937	0.0029	0.960
rand2	-0.93418	0.35681	0.0676	0.452
rand3	-0.50374	-0.86386	0.0331	0.677
rand4	0.67422	0.73853	0.0694	0.430
rand5	-0.16224	0.98675	0.1182	0.217
rand6	-0.48448	-0.87480	0.0071	0.925
rand7	0.97918	-0.20301	0.0307	0.699
rand8	-0.76753	0.64102	0.0553	0.478
rand9	0.77265	-0.63483	0.0241	0.727
rand10	-0.71217	-0.70200	0.1887	0.076
rand11	0.15929	-0.98723	0.1011	0.257

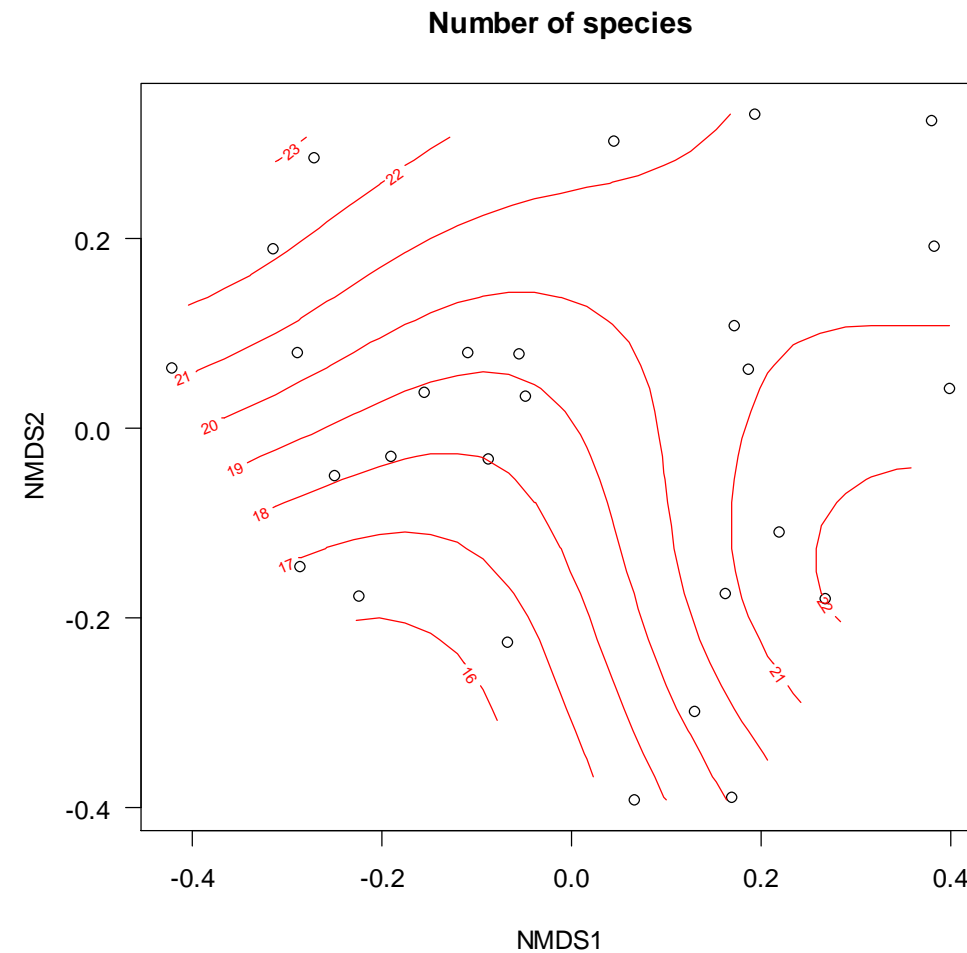
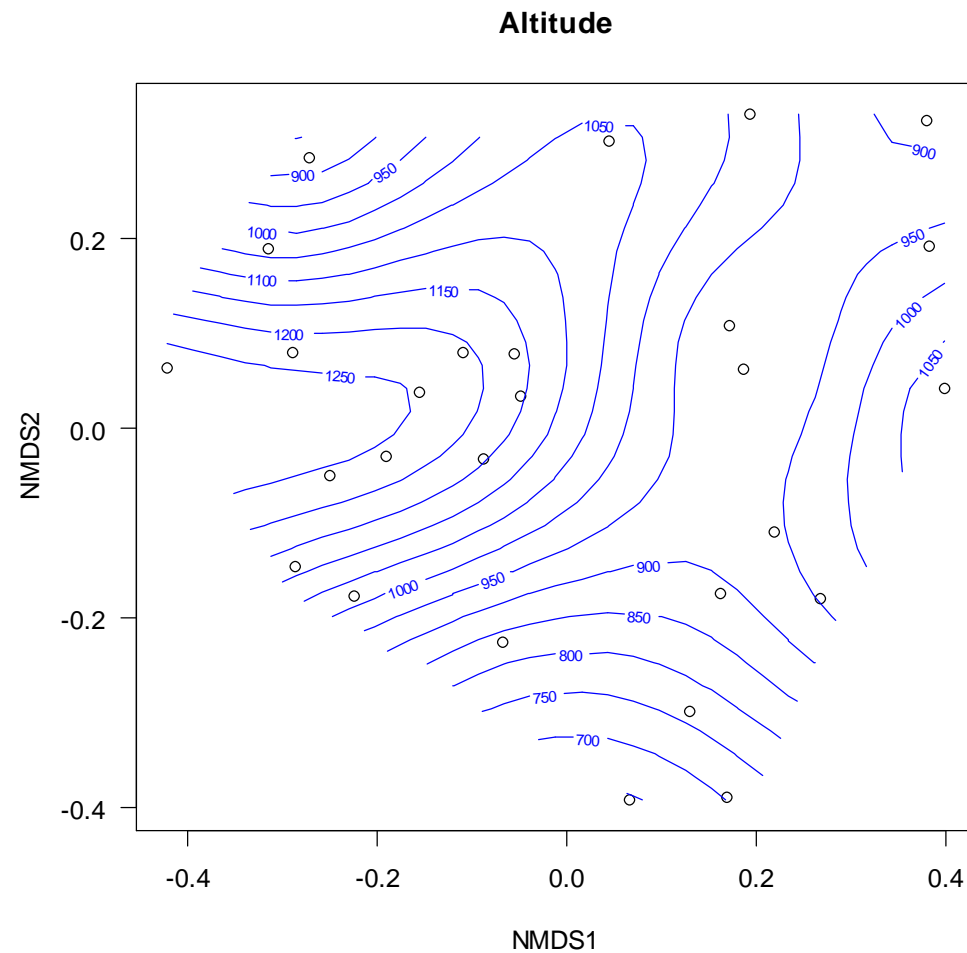


	NMDS1	NMDS2	r2	Pr (>r)
LATITUDE	0.16394	0.98647	0.6000	0.001 ***
LONGITUDE	-0.38731	-0.92195	0.2948	0.015 *
ALTITUDE	-0.77129	0.63648	0.3488	0.004 **
TEMP_JULY	0.72650	-0.68717	0.3341	0.009 **
TEMP_JAN	0.99907	0.04306	0.1121	0.246
TEMP	0.80798	-0.58921	0.2329	0.045 *
PREC	0.14608	0.98927	0.3616	0.008 **
depN	0.35755	0.93389	0.3060	0.014 *
N	-0.89075	0.45449	0.1327	0.185
P	-0.98021	-0.19794	0.5928	0.001 ***
K	-0.89756	-0.44088	0.4656	0.001 ***

Další možnost pasivního promítnutí proměnných

- Nelineární vztah zobrazený jako „vrstevnice“

`ordisurf {vegan}`



Data z rašeliníšť - Jiroušek (2008). Diplomová práce

vrstevnice jsou výsledkem GAM modelu

Literatura

- Legendre, P. & Legendre, L. (2012): Numerical ecology. Third Edition. Elsevier, Amsterdam.
- Borcard, D., Gillet, F. & Legendre, P. (2011): Numerical ecology with R. Springer, New York.
- Haruštiaková, D., Jarkovský, J., Littnerová, S. & Dušek, L. (2012): Vícerozměrné statistické metody v biologii. Akademické nakladatelství CERM, s.r.o., Brno.