

# **STATISTICKÉ METODY; ZÍSKÁVÁNÍ INFORMACÍ Z DRUHOVÝCH A ENVIRONMENTÁLNÍCH DAT**

---

# (NE)VÝHODY STATISTIKY

OTÁZKY si klást ještě před odběrem a podle nich naplánovat design, metodiku odběru (experimentální vs. pozorování), analýzy, grafy a testy.

# (NE)VÝHODY STATISTIKY

OTÁZKY si klást ještě před odběrem a podle nich naplánovat design, metodiku odběru (experimentální vs. pozorování), analýzy, grafy a testy.

Je důležité, aby se srovnávaly porovnatelné OBJEKTY (tj. takové, které se liší pouze, nebo hlavně studovaným jevem a v ostatních parametrech jsou si podobné).

# **(NE)VÝHODY STATISTIKY**

OTÁZKY si klást ještě před odběrem a podle nich naplánovat design, metodiku odběru (experimentální vs. pozorování), analýzy, grafy a testy.

Je důležité, aby se srovnávaly porovnatelné OBJEKTY (tj. takové, které se liší pouze, nebo hlavně studovaným jevem a v ostatních parametrech jsou si podobné).

Identifikace DRUHŮ není tolik důležitá z pohledu taxonomie, ale záleží na jednotnosti určování; pojmenovávat organismy jednotně napříč vzorky.

# (NE)VÝHODY STATISTIKY

OTÁZKY si klást ještě před odběrem a podle nich naplánovat design, metodiku odběru (experimentální vs. pozorování), analýzy, grafy a testy.

Je důležité, aby se srovnávaly porovnatelné OBJEKTY (tj. takové, které se liší pouze, nebo hlavně studovaným jevem a v ostatních parametrech jsou si podobné).

Identifikace DRUHŮ není tolik důležitá z pohledu taxonomie, ale záleží na jednotnosti určování; pojmenovávat organismy jednotně napříč vzorky.

**Možnost využití jiných taxonomických jednotek (OTU) než je druh (morfotypy, rody).**

# **(NE)VÝHODY STATISTIKY**

Nepoužívat PROGRAMY bez dobré znalosti metod.

# **(NE)VÝHODY STATISTIKY**

Nepoužívat PROGRAMY bez dobré znalosti metod.

Nedá se určit, která z METOD je obecně lepší záleží na povaze dat; často možnost použít více vhodných analýz.

# **(NE)VÝHODY STATISTIKY**

Nepoužívat PROGRAMY bez dobré znalosti metod.

Nedá se určit, která z METOD je obecně lepší záleží na povaze dat; často možnost použít více vhodných analýz.

**STATISTIKA není všemocná, někdy z dat “nic” nevyplývá.**



# **(NE)VÝHODY STATISTIKY**

Nepoužívat PROGRAMY bez dobré znalosti metod.

Nedá se určit, která z METOD je obecně lepší záleží na povaze dat; často možnost použít více vhodných analýz.

STATISTIKA není všemocná, někdy z dat “nic” nevyplývá.

Usuzování na parametry základního souboru pomocí parametrů VYBĚRU. Spolehlivost tohoto odhadu.

# (NE)VÝHODY STATISTIKY

Nepoužívat PROGRAMY bez dobré znalosti metod.

Nedá se určit, která z METOD je obecně lepší záleží na povaze dat; často možnost použít více vhodných analýz.

STATISTIKA není všemocná, někdy z dat “nic” nevyplývá.

Usuzování na parametry základního souboru pomocí parametrů VYBĚRU. Spolehlivost tohoto odhadu.

Nejjednodušší závislost je LINEÁRNÍ. To neznamena, že předpokládám, že svět se chová lineárně, ale to, že jej mohu lineárním modelem aproximovat (v určitém rozsahu hodnot).

# **(NE)VÝHODY STATISTIKY**

Nepoužívat PROGRAMY bez dobré znalosti metod.

Nedá se určit, která z METOD je obecně lepší záleží na povaze dat; často možnost použít více vhodných analýz.

STATISTIKA není všemocná, někdy z dat “nic” nevyplývá.

Usuzování na parametry základního souboru pomocí parametrů VYBĚRU. Spolehlivost tohoto odhadu.

Nejjednodušší závislost je LINEÁRNÍ. To neznamena, že předpokládám, že svět se chová lineárně, ale to, že jej mohu lineárním modelem aproximovat (v určitém rozsahu hodnot).

Podíl VARIABILITY jedné proměnné vysvětlené změnami druhé proměnné.

# OTÁZKY, KTERÉ SI LZE KLÁST

otázky spojené s organismy; nenáhodnost jevů

# OTÁZKY, KTERÉ SI LZE KLÁST

otázky spojené s organismy; nenáhodnost jevů  
DIVERZITA – porovnávání na různých úrovních

# OTÁZKY, KTERÉ SI LZE KLÁST

otázky spojené s organismy; nenáhodnost jevů

DIVERZITA – porovnávání na různých úrovních

SPOLEČENSTVA – výskyt druhů, podobnost, abundance

# OTÁZKY, KTERÉ SI LZE KLÁST

otázky spojené s organismy; nenáhodnost jevů

DIVERZITA – porovnávání na různých úrovních

SPOLEČENSTVA – výskyt druhů, podobnost, abundance

MORFOLOGIE – velikost, tvar, struktury

# OTÁZKY, KTERÉ SI LZE KLÁST

otázky spojené s organismy; nenáhodnost jevů

DIVERZITA – porovnávání na různých úrovních

SPOLEČENSTVA – výskyt druhů, podobnost, abundance

MORFOLOGIE – velikost, tvar, struktury

VÝZNAMNOST FAKTORŮ – fyzikálně-chemické  
parametry, vzdálenost lokalit/vzorků, sezónní změny



# TYPY DAT/PROMĚNNÝCH

druhová, morfologická a environmentální data

# TYPY DAT/PROMĚNNÝCH

druhová, morfologická a environmentální data

**KVANTITATIVNÍ** – diskrétní (např. jen určitá čísla), spojité (měření; počítání buněk druhu)

# TYPY DAT/PROMĚNNÝCH

druhová, morfologická a environmentální data

**KVANTITATIVNÍ** – diskrétní (např. jen určitá čísla), spojité (měření; počítání buněk druhu)

**SEMIKVANTITATIVNÍ** – odhad proměnné (procentuální; kategorie kvantit)

# TYPY DAT/PROMĚNNÝCH

druhová, morfologická a environmentální data

**KVANTITATIVNÍ** – diskrétní (např. jen určitá čísla), spojitě (měření; počítání buněk druhu)

**SEMIKVANTITATIVNÍ** – odhad proměnné (procentuální; kategorie kvantit)

**KVALITATIVNÍ** – binární (výskyt/nevýskyt), víceřadové (faktory, dummy variables), rozdělení kvantitativních do skupin/kategorií

# TERMINOLOGIE

VARIANCE/VARIABILITA/KOVARIANCE – variance hodnot určité proměnné/variabilita druhových dat (změny v druzích a/nebo v abundancích), variabilita morfologických dat (mnohorozměrná data o tvaru)/kovariance: společný vliv na variabilitu dat

# TERMINOLOGIE

VARIANCE/VARIABILITA/KOVARIANCE – variance hodnot určité proměnné/variabilita druhových dat (změny v druzích a/nebo v abundancích), variabilita morfologických dat (mnohorozměrná data o tvaru)/kovariance: společný vliv na variabilitu dat

**(NE)ZÁVISLÁ PROMĚNNÁ** - nezávislá proměnná: prediktor, faktor, kovariáta (odstínění vlivu proměnné na data); závislá proměnná: response variable, sledovaná proměnná

# TERMINOLOGIE

VARIANCE/VARIABILITA/KOVARIANCE – variance hodnot určité proměnné/variabilita druhových dat (změny v druzích a/nebo v abundancích), variabilita morfologických dat (mnohorozměrná data o tvaru)/kovariance: společný vliv na variabilitu dat

(NE)ZÁVISLÁ PROMĚNNÁ - nezávislá proměnná: prediktor, faktor, kovariáta (odstínění vlivu proměnné na data); závislá proměnná: response variable, sledovaná proměnná

DISTANCE/SIMILARITA – vzdálenost v prostoru dat/podobnost (distance = 1-hodnota podobnosti)

# TERMINOLOGIE

VARIANCE/VARIABILITA/KOVARIANCE – variance hodnot určité proměnné/variabilita druhových dat (změny v druzích a/nebo v abundancích), variabilita morfologických dat (mnohorozměrná data o tvaru)/kovariance: společný vliv na variabilitu dat

(NE)ZÁVISLÁ PROMĚNNÁ - nezávislá proměnná: prediktor, faktor, kovariáta (odstínění vlivu proměnné na data); závislá proměnná: response variable, sledovaná proměnná

DISTANCE/SIMILARITA – vzdálenost v prostoru dat/podobnost (distance = 1-hodnota podobnosti)

SIGNIFIKANCE - pravděpodobnost sebraných dat za podmínky neexistence závislosti (platnosti nulové hypotézy)



# METODY ANALÝZ

odhad typu a intensity závislosti proměnných

**JEDNOROZMĚRNÉ** – změny nebo vztahy několika málo proměnných; popisná statistika, analýza variance (ANOVA), korelace/regrese; druhy přes indexy diverzity nebo podobnosti

**MNOHOROZMĚRNÉ** – zjednodušení mnohorozměrných dat; shlukové analýzy, ordinační metody, mnohonásobné korelace/regrese, diskriminační analýzy

**TESTOVÁNÍ** – test jen předpokladem toho, aby bylo možné se o výsledcích vůbec bavit

# TESTOVÁNÍ

Vybrat vhodný TEST (hlavně na základě znalosti o povaze dat) a rozhodnout se pro něj před provedením vlastní analýzy (nikoli zkoušet mnoho testů naslepo).

# TESTOVÁNÍ

Vybrat vhodný TEST (hlavně na základě znalosti o povaze dat) a rozhodnout se pro něj před provedením vlastní analýzy (nikoli zkoušet mnoho testů naslepo).

Pokud nelineární vztah hledání různých modelů (funkce).

# TESTOVÁNÍ

Vybrat vhodný TEST (hlavně na základě znalosti o povaze dat) a rozhodnout se pro něj před provedením vlastní analýzy (nikoli zkoušet mnoho testů naslepo).

Pokud nelineární vztah hledání různých modelů (funkce).

*Parametrické* testy mají předpoklad určitého rozdělení (např. normální – není časté u menších výběrů a také obecně u biologických dat), *neparametrické testy* (permutace/randomizace, pořadí dle hodnot).

# ZOBRAZENÍ DAT

výběr uzpůsobit pro sdělnost a zvýraznění důležitých aspektů

# ZOBRAZENÍ DAT

výběr způsobit pro sdělnost a zvýraznění důležitých aspektů

**NĚKOLIK MÁLO PROMĚNNÝCH:** *záv. kvantitativní vs. nezáv. kvalitativní* – popisná statistika, box ploty, sloupcové grafy, xy graf průměrů objektů; *záv. kvantitativní vs. nezáv. kvantitativní* – korelace, spojnicový graf, bodový xy graf (matematický model), box ploty (rozdělení spojitě proměnné na intervaly); *záv. kvalitativní vs. nezáv. kvalitativní* – frekvenční tabulky

# ZOBRAZENÍ DAT

výběr uzpůsobit pro sdělnost a zvýraznění důležitých aspektů

**NĚKOLIK MÁLO PROMĚNNÝCH:** *záv. kvantitativní vs. nezáv. kvalitativní* – popisná statistika, box ploty, sloupcové grafy, xy graf průměrů objektů; *záv. kvantitativní vs. nezáv. kvantitativní* – korelace, spojnicový graf, bodový xy graf (matematický model), box ploty (rozdělení spojitě proměnné na intervaly); *záv. kvalitativní vs. nezáv. kvalitativní* – frekvenční tabulky

**MNOHO PROMĚNNÝCH:** *hypotézy* – PCA/DCA, NMDS, shlukové analýzy; *vliv faktorů*: RDA/CCA, rozdělení variability, Mantelův test

# INTERPRETACE DAT

porovnávání výsledků, umět číst z grafů

**JEDNOROZMĚRNÉ** – predikce hodnot na základě pozorování; porovnání závislostí mezi skupinami z hodnot a diagramů

**MNOHOROZMĚRNÉ** – shluky interpretovat na základě vnější informace (znalost ekologie druhů, znalost stanovišť); pozor na argumentaci kruhem. Významnost parametrů, zjištění vzájemně korelovaných parametrů (spolu-působení faktorů na variabilitu druhů), rozklad celkové variability.



# REFERENCE

NÁPOVĚDA a MANUÁLY programů; wikipedia.

materiály T HERBENA <http://web.natur.cuni.cz/~herben/biostat.html> *nebo* .../multivar.html

BORCARD D, LEGENDRE P a DRAPEAU P (1992)  
Partialling out the spatial component of ecological variation.

LEGENDRE P a LEGENDRE L (1998) Numerical ecology.

LEPŠ J a ŠMILAUER P (2003) Multivariate analysis of ecological data using CANOCO.

McCUNE B a GRACE JB (2002) Analysis of ecological communities.

MARHOLD K a SUDA J (2001) Analýza multivariačních dat v taxonomii (fenetické metody).

# STATISTICKÉ APLIKACE

komunikace přes MS Excel nebo clipboard (ctrl+c – ctrl+v)

MS EXCEL – zaznamenávání dat, pouze jednorozměrné analýzy

CANOCO – mnohorozměrné ordinační analýzy

eRko – spíše přes příkazový řádek, spousta skriptů; zdarma

PAST – obyčejné grafy, tabulkové výpočty; zdarma

PRIMER – jen mnohorozm., počítá s distancemi/podobnostmi

SIGMA PLOT – pouze kreslení grafů

ZT WIN – Mantelovy testy prostorová autokorelace; zdarma

STATISTICA, ORIGIN, S-PLUS, NCCS – klikací, pěkná grafika grafů

TPS – mnohorozměrné morfometrické analýzy tvaru; zdarma

# ZKUŠEBNÍ DATA

MORFOMETRICKÁ data  
data o SPOLEČENSTVECH  
EXPERIMENTÁLNÍ data

# POPISNÉ TABULKY

Taxon name		Group 1			Group 3	Group 2		
		1	5	6	3	4	7	8
	<i>Adlafia bryophila</i>	X	o	O	O	o	o	+
	<i>Chamaepinnularia australomediocris</i>	O	+	+	O	o	+	+
	<i>Fragilaria capucina</i> morph. 2	+	+	+	o		o	o
	<i>Pinnularia acidicola</i>	+	o	o	+		o	o
	<i>Stauroforma exiguiformis</i>	+	O	o	o		o	o
		+	X	o	o		o	o
		+	X	o	o	o	o	o
		o	O	o	o		o	o
		o	O	o				o
		o	+	o			o	o
		o	+	o				
		o	o	X	o		o	
		o	o	X				
		o		+				
		o			X		o	o
		o			X		o	o
		o	o		O		o	+
		o	o	o	+	X	X	X
		o				O	X	o
		o			o	+	+	o
		o			o		O	+
		+	o	o	o	o	O	o
		o				o	+	o
		o			o	o	+	o
	<i>Psammothidium stauroneioides</i>					o	+	o
	<i>Diademis contenta</i>					o	+	o
						o	o	X
						o	o	+
						o	o	+
						o	o	+
						o	o	o
						o	o	o

**Table 3.** Variations of chlorophyll a (Chla), pheopigment (Pheop) concentrations, the mean density (Dm) and their standard deviation (S.D.) respectively of drifting algae according to the mean current velocity (Vm), the mean depth (Pm) and the hydraulic radius (rh)

		Chl. a (± S.D.)	Pheop. (± S.D.)	Dm (± S.D.)	Vm	Pm	rh
		$\mu\text{g} \cdot \text{l}^{-1}$		$\text{cell} \cdot \text{l}^{-1}$	$\text{m} \cdot \text{s}^{-1}$	m	m
Station 1	Surface	0.46 (± 0.24)	0.84 (± 0.21)	42200 (± 40310)	0.32		
	Half depth	0.50 (± 0.20)	0.64 (± 0.29)	50660 (± 33010)	0.27	0.15	0.024
	Bottom	0.93 (± 0.75)	1.50 (± 0.82)	55820 (± 44540)	0.17		
Station 2	Surface	0.35 (± 0.10)	0.23 (± 0.12)	17710 (± 12180)	0.53		
	Half depth	0.38 (± 0.12)	0.39 (± 0.11)	16430 (± 12660)	0.34	0.21	0.052
	Bottom	0.40 (± 0.26)	0.50 (± 0.18)	16330 (± 14470)	0.22		

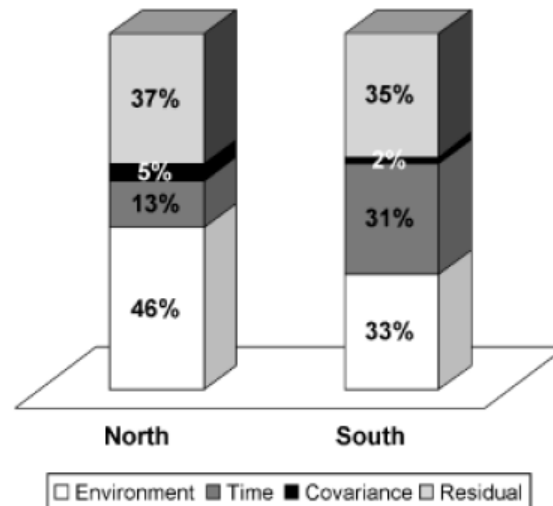
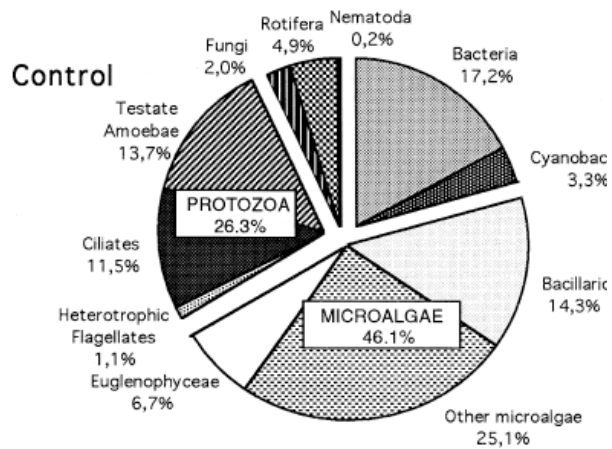
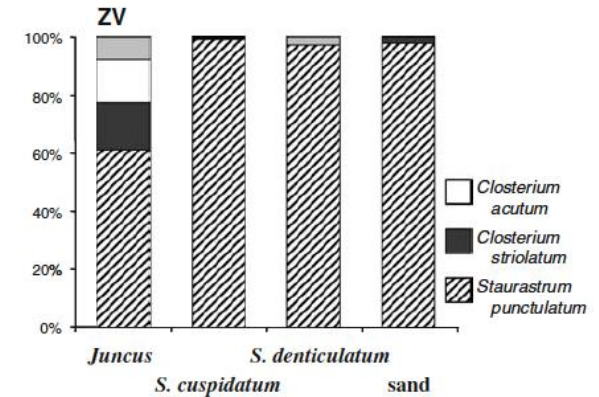
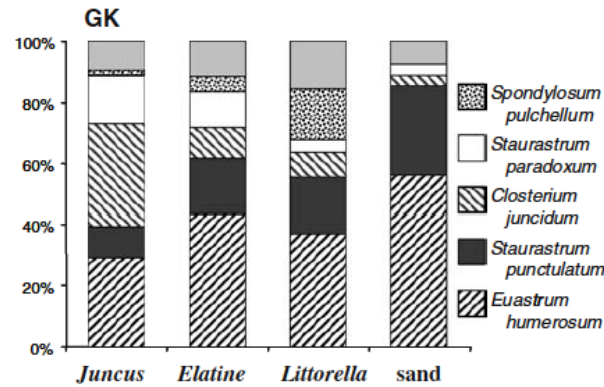
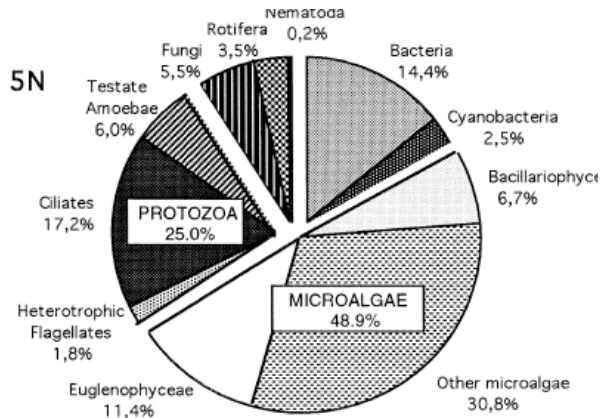
**Table 3** Summary of the main ecological characteristics of the different assemblages

	Group 1			Group 3	Group 2		
	1	5	6	3	4	7	8
Number of samples	62	5	10	15	4	18	18
Mean moisture class	4.1 ± 0.9	3.9 ± 1.2	4.7 ± 0.5	3.3 ± 0.9	3.25 ± 0.29	2.8 ± 1.0	2.3 ± 0.9
Mean water pH <sup>a</sup>	6.8 ± 0.6	6.2	6.9 ± 0.4	6.8 ± 0.6	–	5.8	6.8 ± 0.6
Mean soil pH <sup>a</sup>	6.1 ± 0.6	7	6.3 ± 1.2	5.8 ± 0.4	–	5.98 ± 0.10	5.8 ± 0.5
Mean soil conductivity <sup>a</sup> ( $\mu\text{S}/\text{cm}$ )	0.17 ± 0.08	0.09	0.40 ± 0.20	0.23 ± 0.12	–	0.54 ± 0.21	0.21 ± 0.12
Mean soil salinity <sup>a</sup>	0.10 ± 0.04	0.05	0.23 ± 0.12	0.13 ± 0.7	–	0.31 ± 0.12	0.12 ± 0.08
Mean LOI (%DW) <sup>a</sup>	23 ± 16	1	41 ± 23	22 ± 12	–	14.5 ± 1.0	11 ± 8
Mean no. of taxa	32 ± 10	22 ± 4	21 ± 6	24 ± 6	19 ± 11	30 ± 11	33 ± 13
Mean diversity	2.4 ± 0.5	1.7 ± 0.4	1.7 ± 0.5	1.9 ± 0.5	1.5 ± 0.7	2.4 ± 0.5	2.3 ± 0.7
Mean evenness	0.70 ± 0.10	0.54 ± 0.12	0.55 ± 0.12	0.61 ± 0.12	0.51 ± 0.15	0.70 ± 0.12	0.66 ± 0.11

<sup>a</sup>When measured.



# SLOUPCOVÉ a VÝSEČOVÉ GRAFY



DIATOMS

185

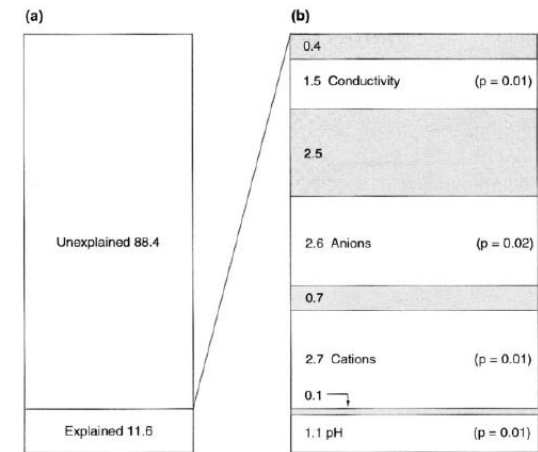
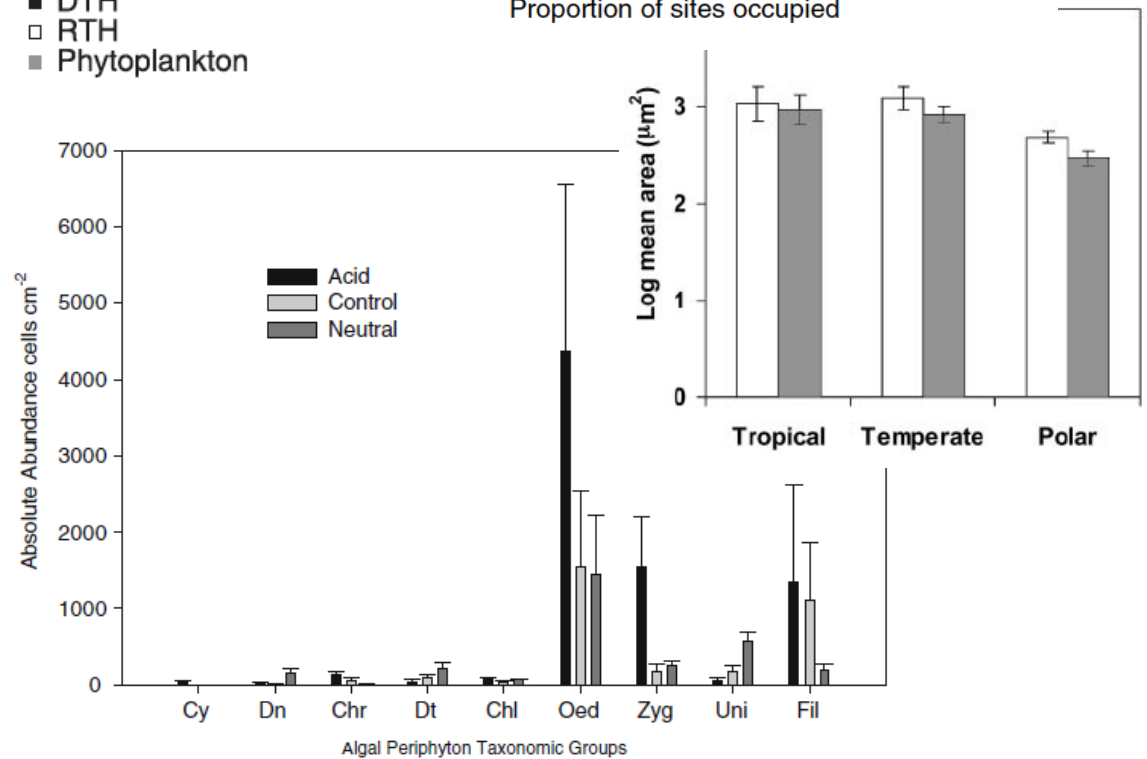
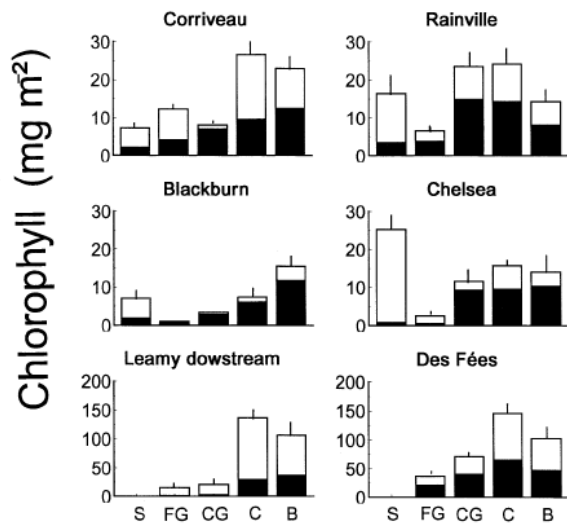
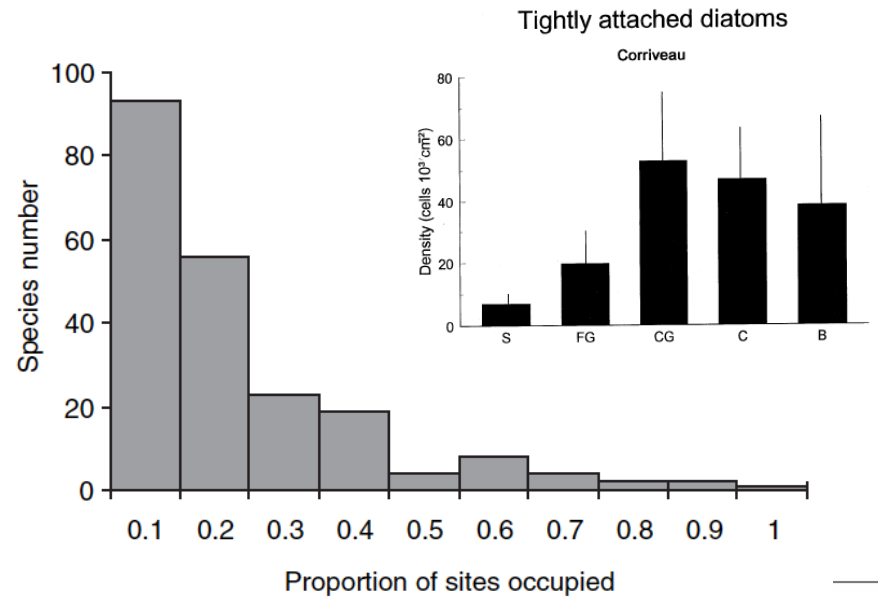
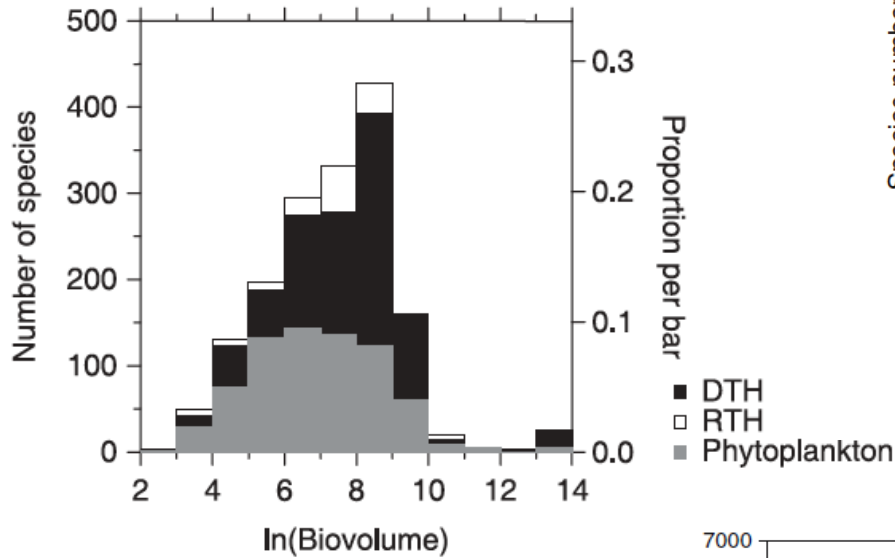


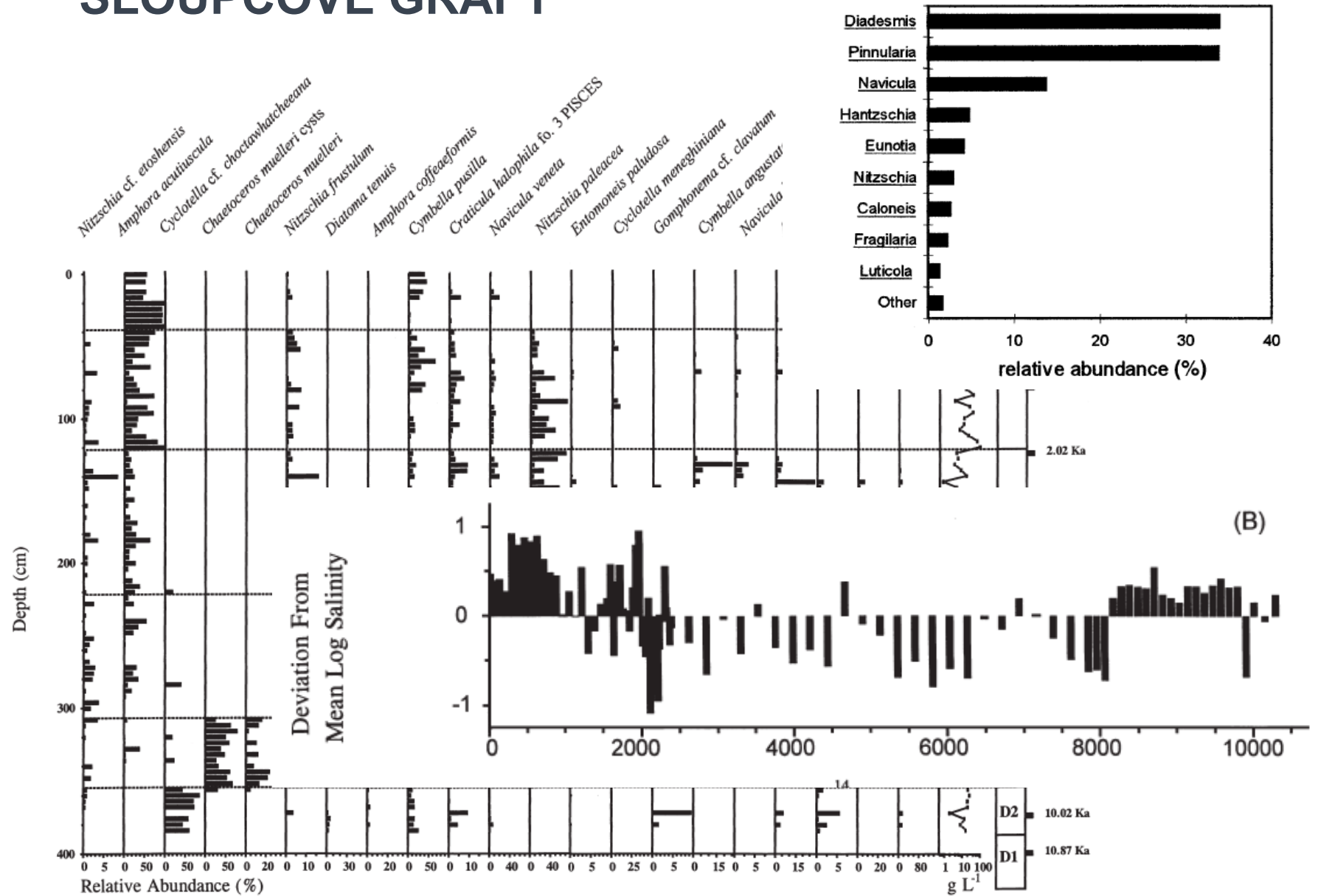
Figure 9. Results of partial CCAs of the training set shown in Figure 8, partitioning the total variance in the diatom data into (a) explained and unexplained portions, and (b) components representing the unique contributions of variables representing the conductivity, pH, cation and anion gradients (open), and correlations between gradients (aded). Significance (p) values are based on 99 random Monte Carlo permutations.

# SLOUPCOVÉ GRAFY



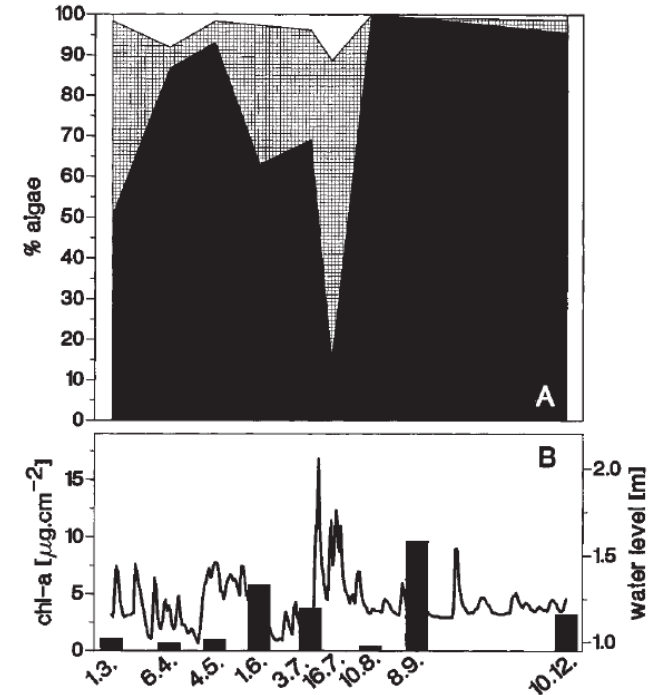
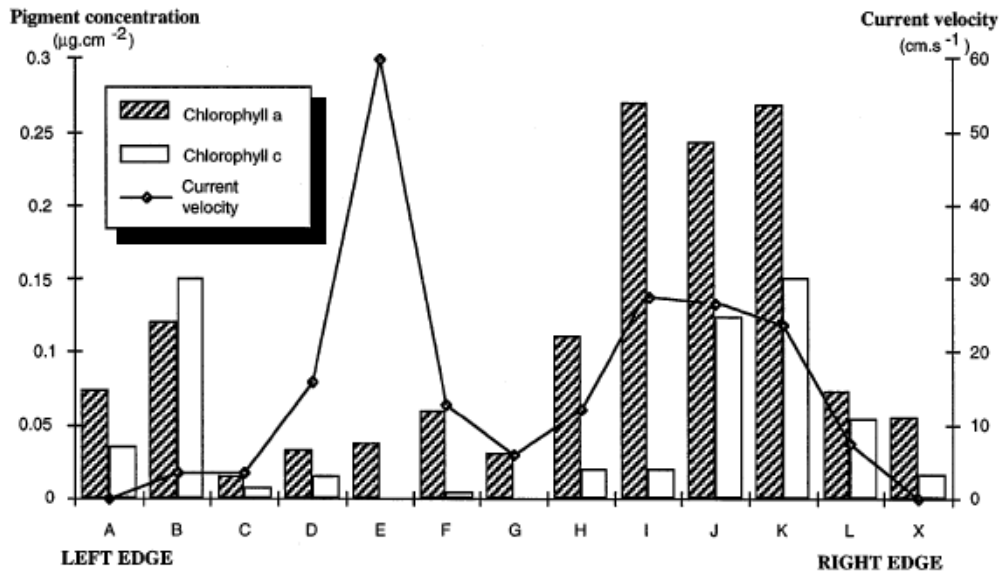
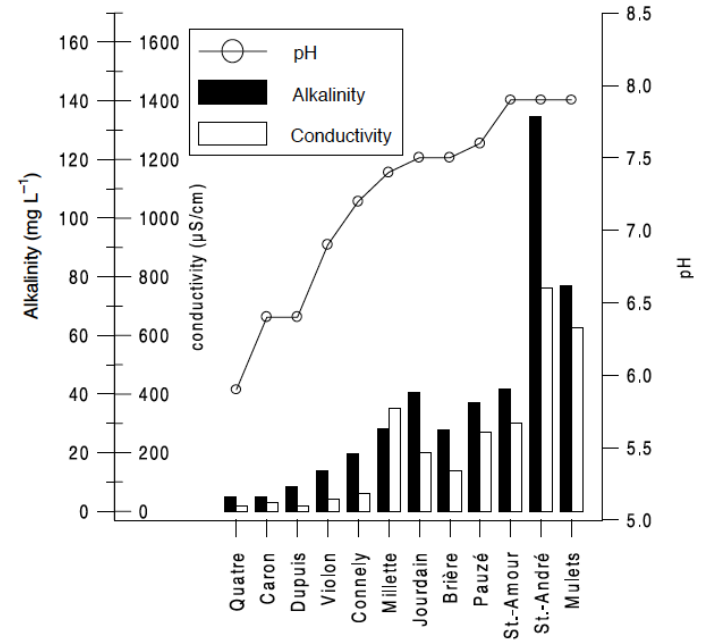
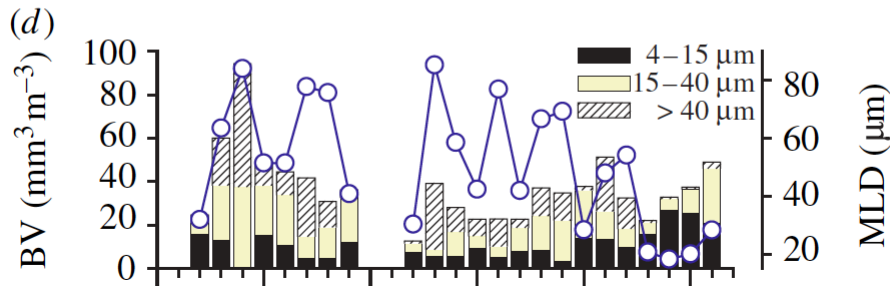
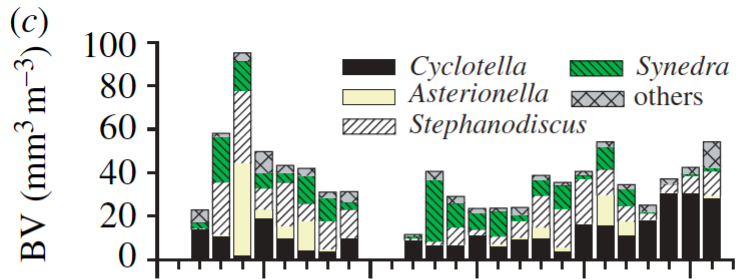


# SLOUPCOVÉ GRAFY





# KOMBINOVANÉ GRAFY



# SPOJNICOVÉ GRAFY

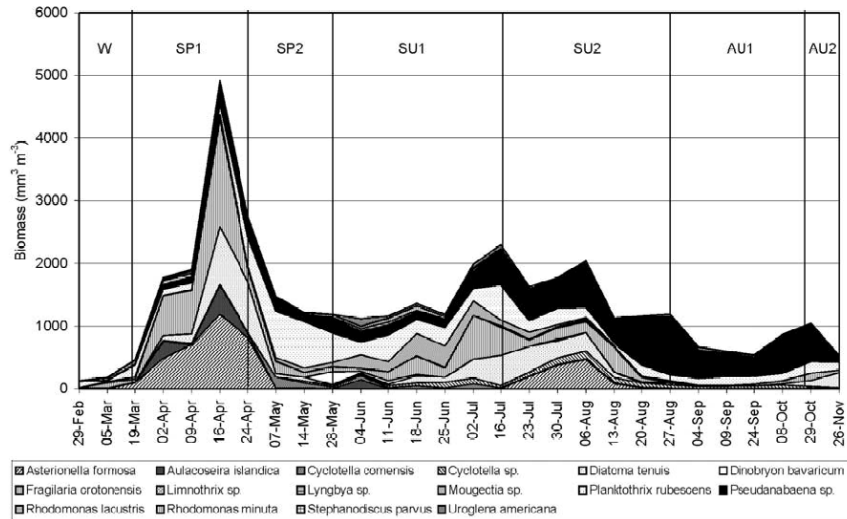
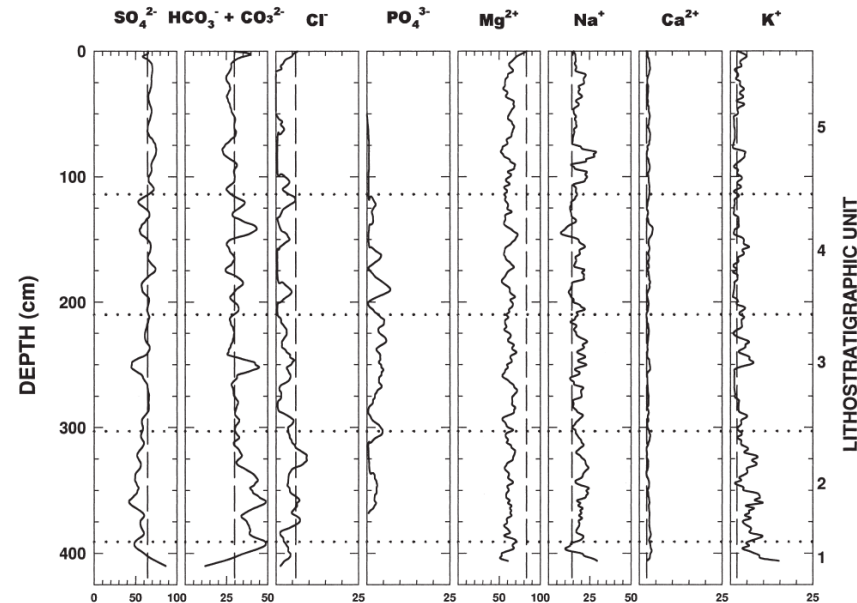
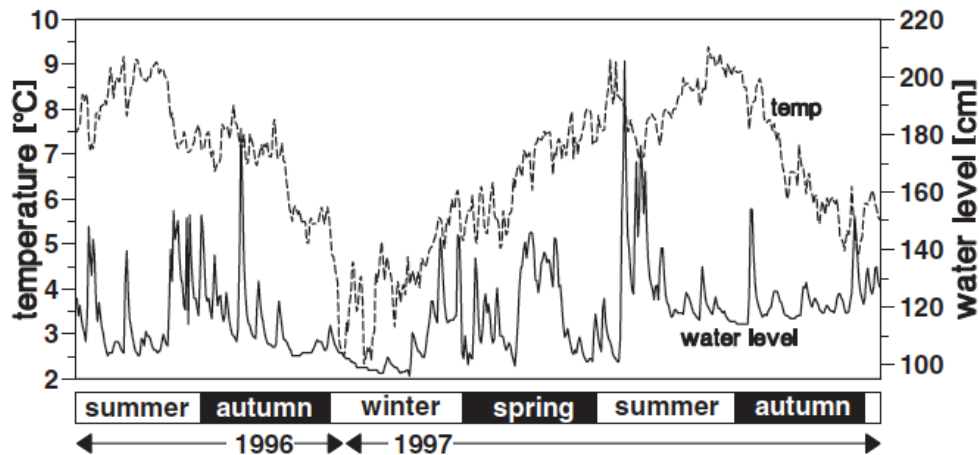
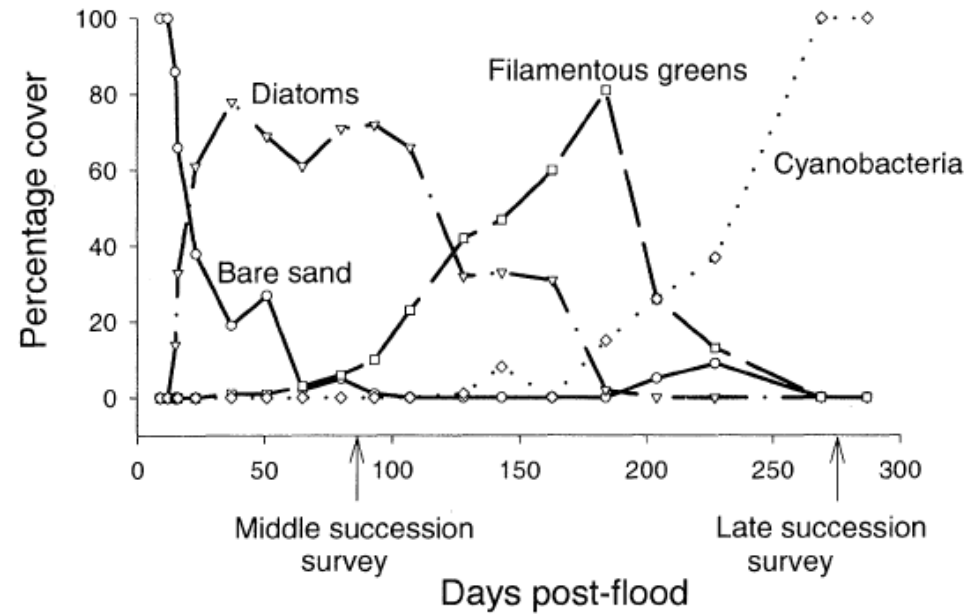


Figure 8. Seasonal changes of the biomass of the dominant species (the species contributing to build up 80% of the total biomass at least once during the year). The groups of samples are divided according to the cluster analysis.



# BOX PLOTY (krabicové diagramy)

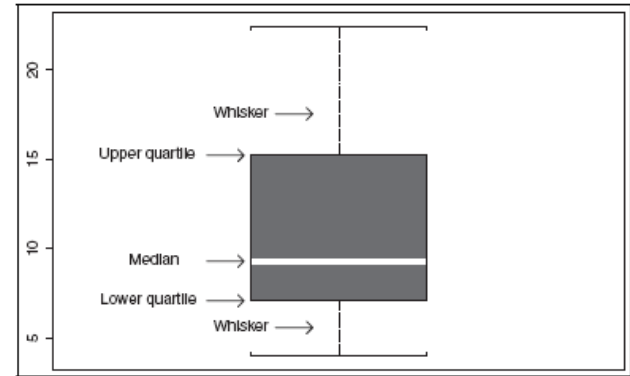
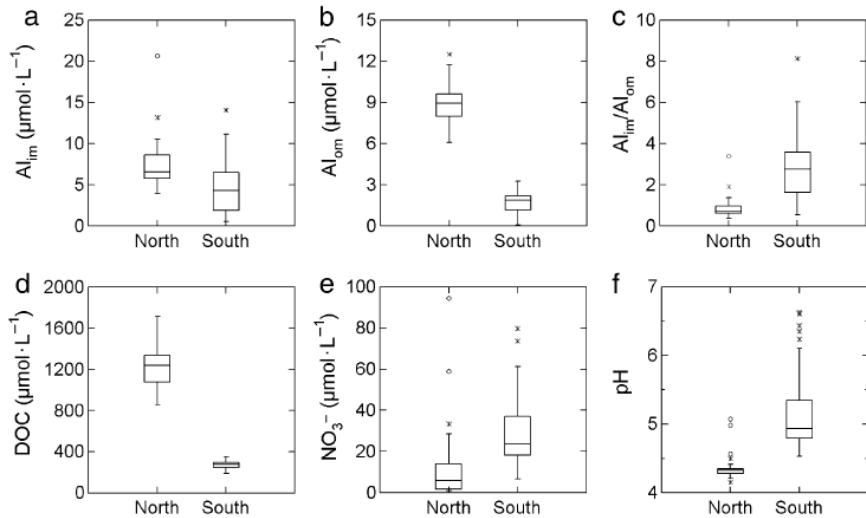


FIGURE 3.2 Example of a boxplot with no outliers.

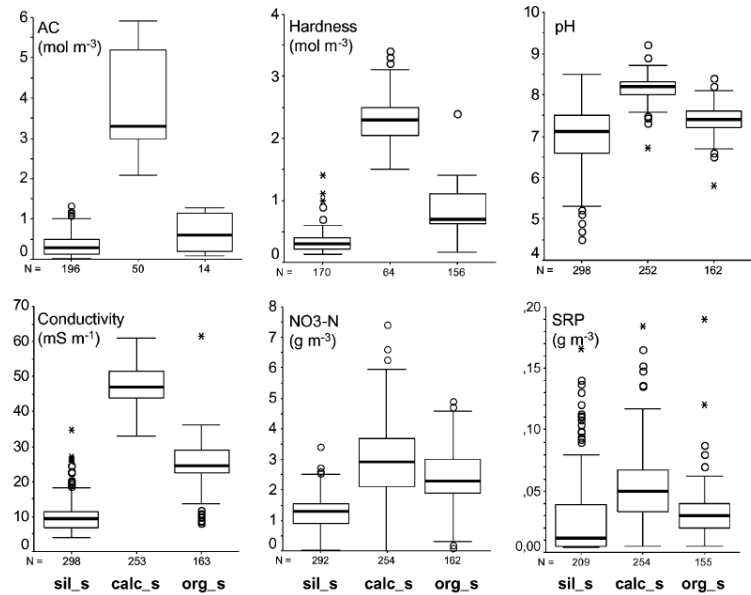
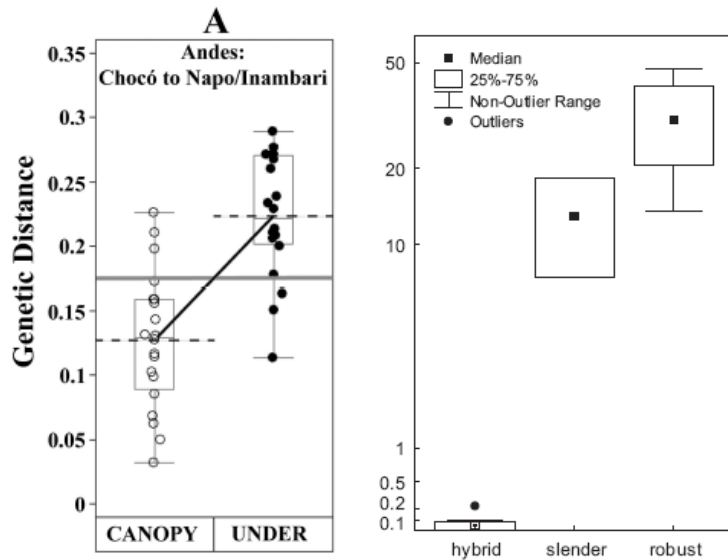
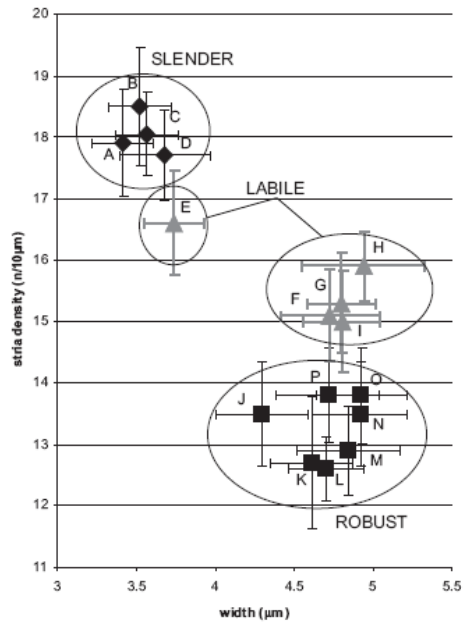
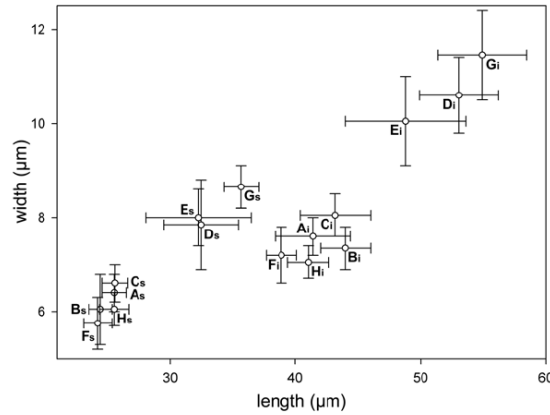


Figure 4. Chemical properties of the sampling sites assigned to the three types: sil\_s = siliceous sites, calc\_s = calcareous sites, org\_s = organic sites. Whiskers show the whole range (0–100%) and boxes represent the main range (25–75%) of the data, the thick black line indicates the median. Circles show extreme values, stars represent outliers.

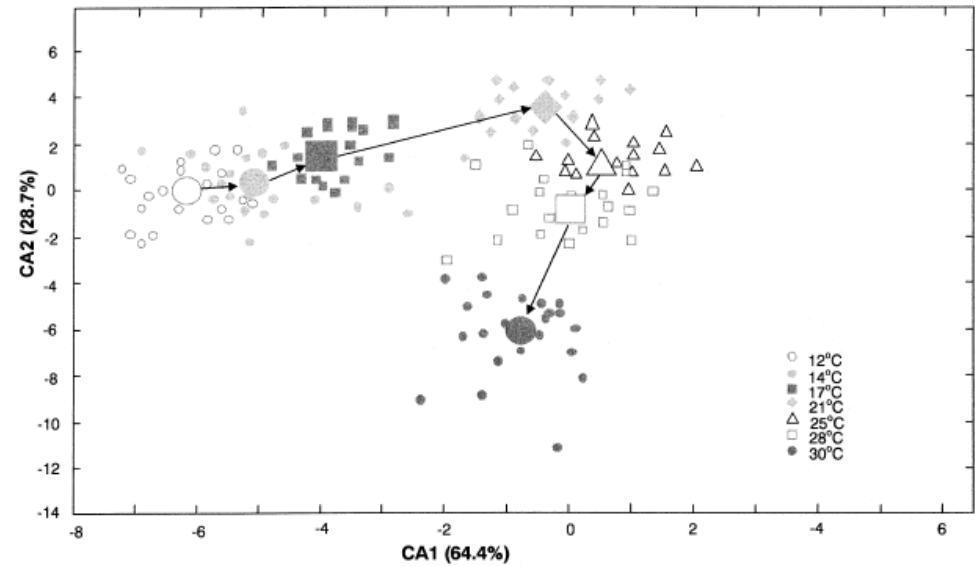
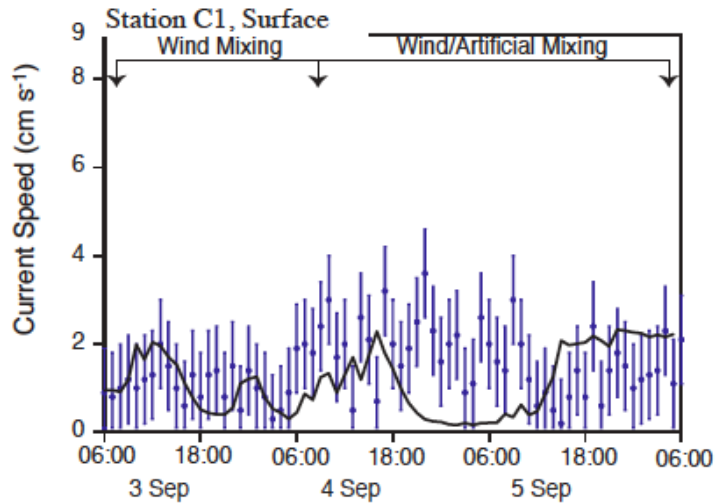
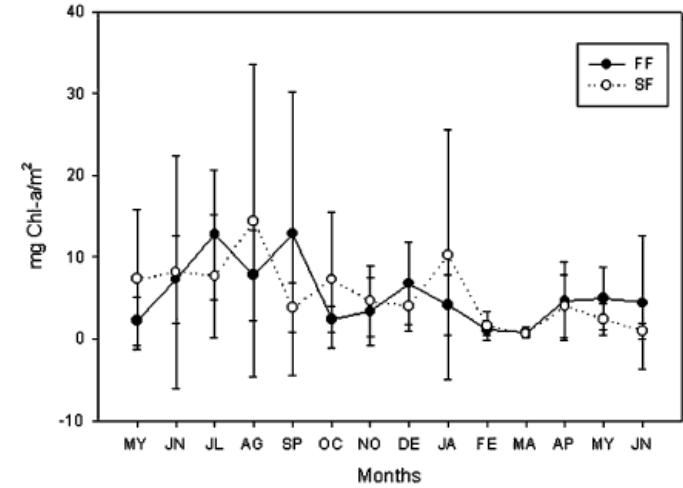
# CENTROIDY/PRŮMĚRNÉ HODNOTY s odchylkami



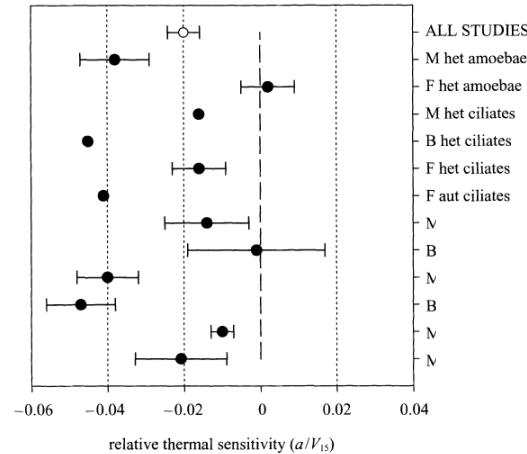
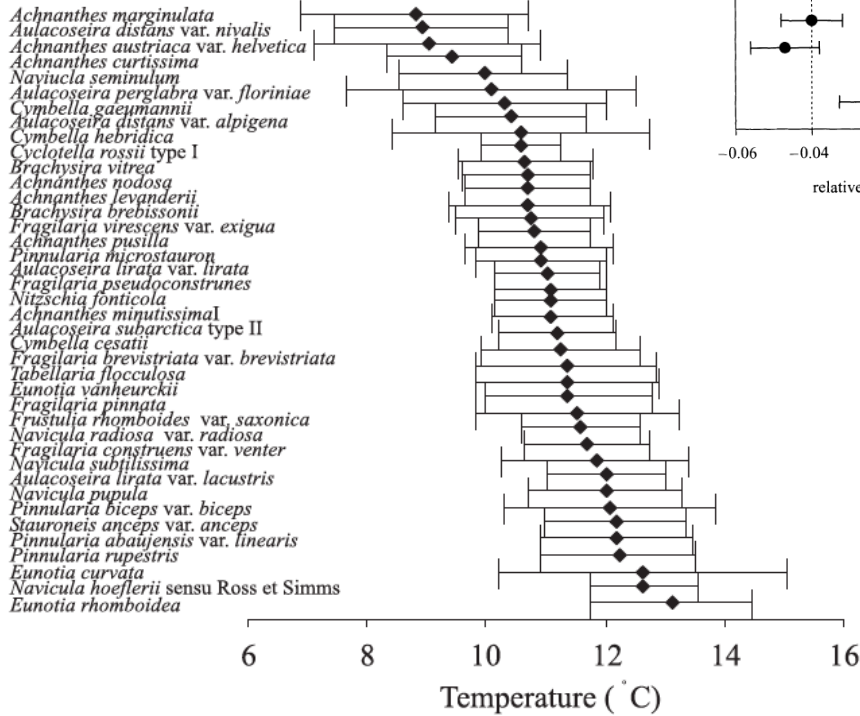
B



Station 2

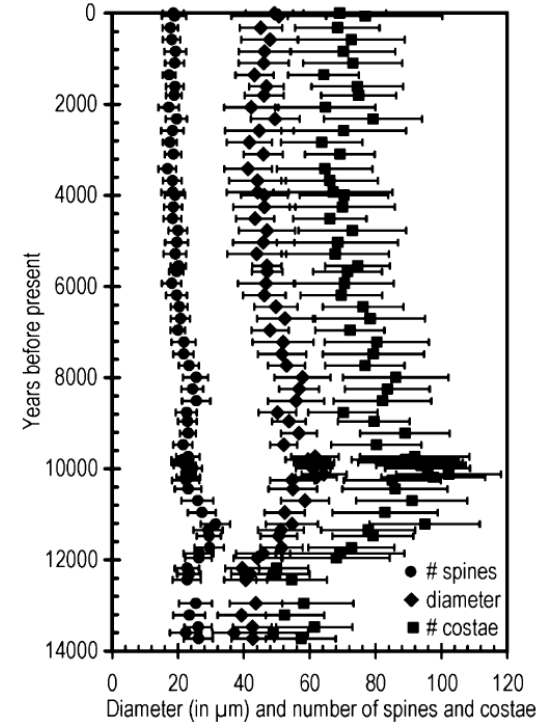


# CENTROIDY/PRŮMĚRNÉ HODNOTY s odchyškami

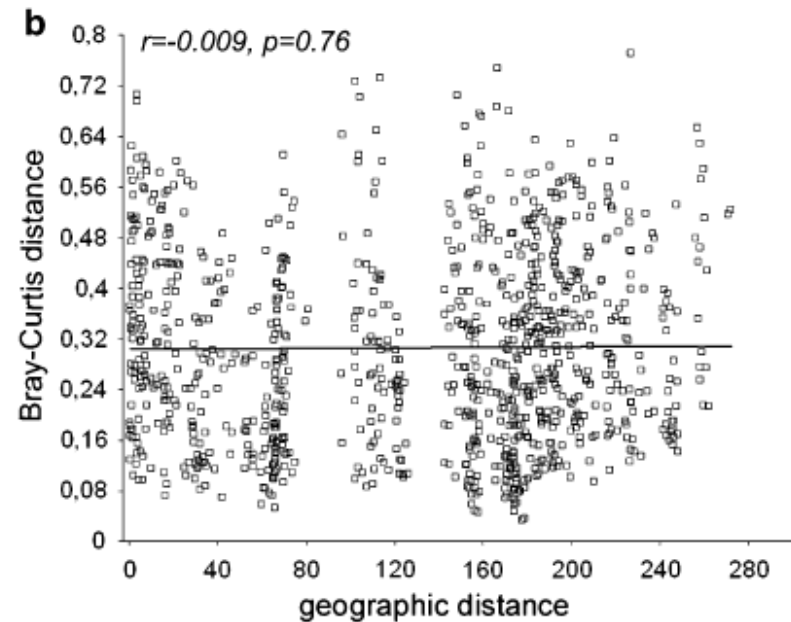
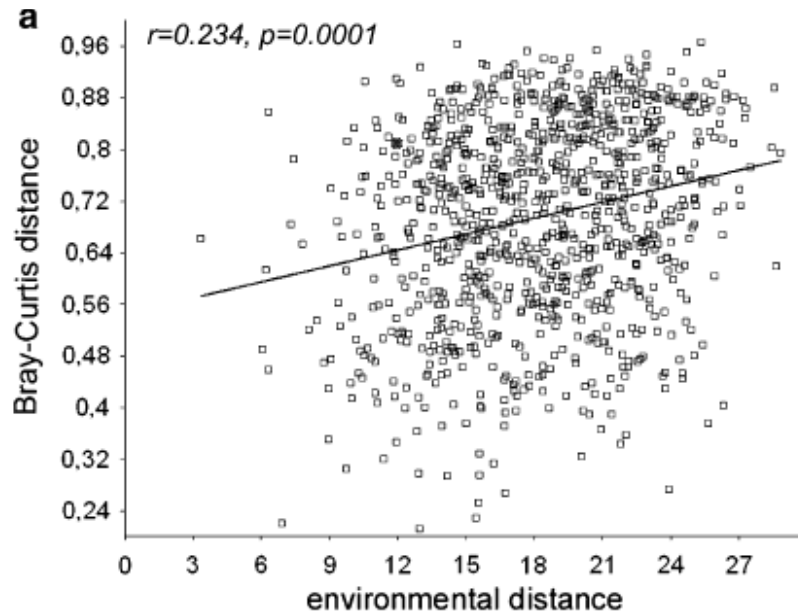
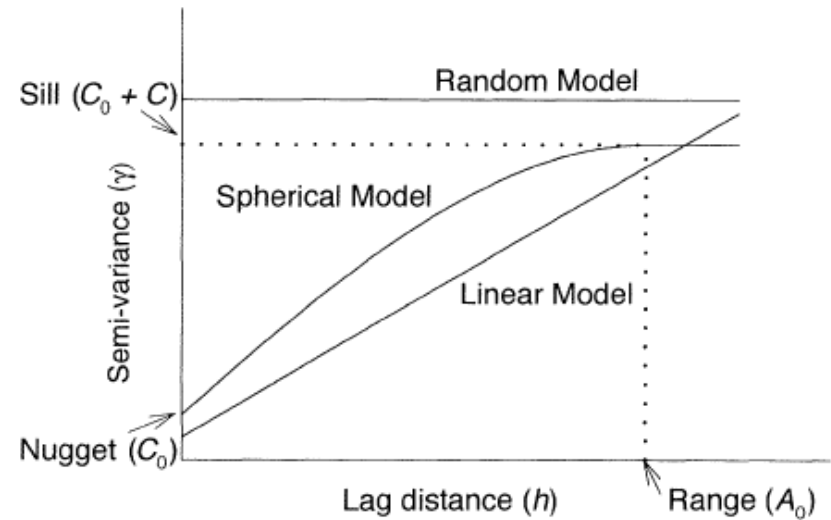
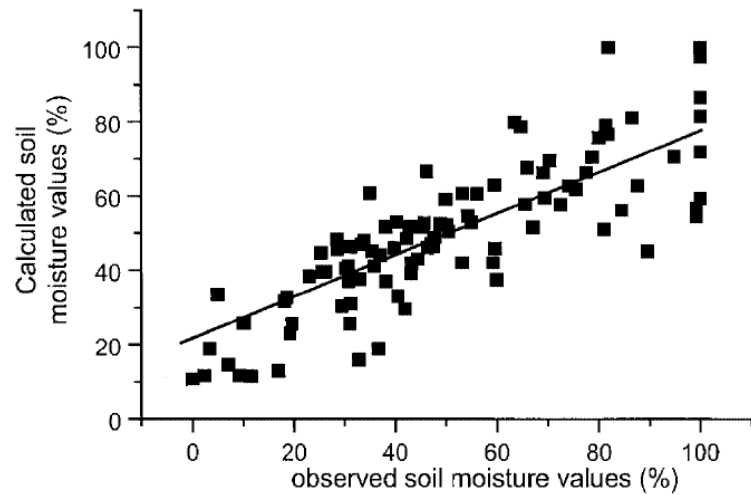


points datasets studies

	points	datasets	studies
ALL STUDIES	455	72	28
M het amoebae	32	9	1
F het amoebae	21	6	1
M het ciliates	4	1	1
B het ciliates	6	1	1
F het ciliates	96	16	3
F aut ciliates	3	1	1

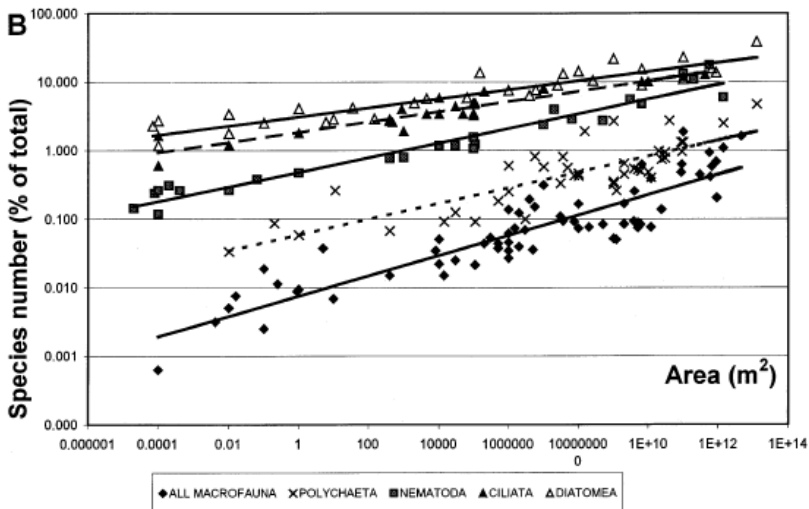
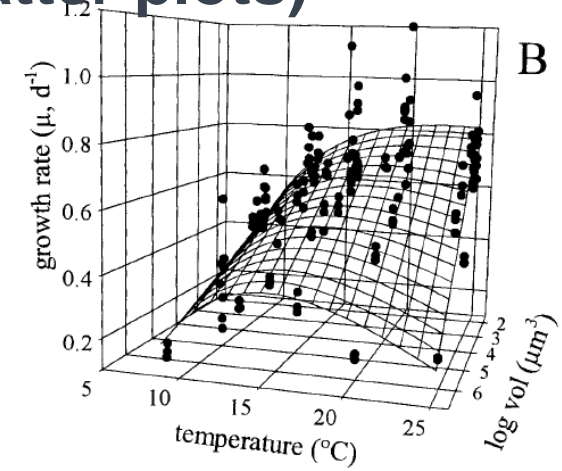
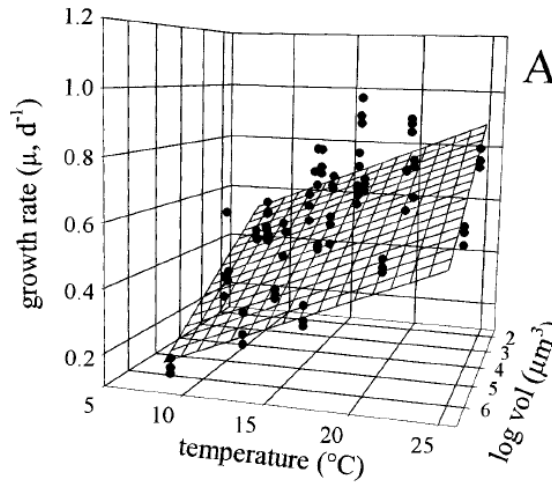
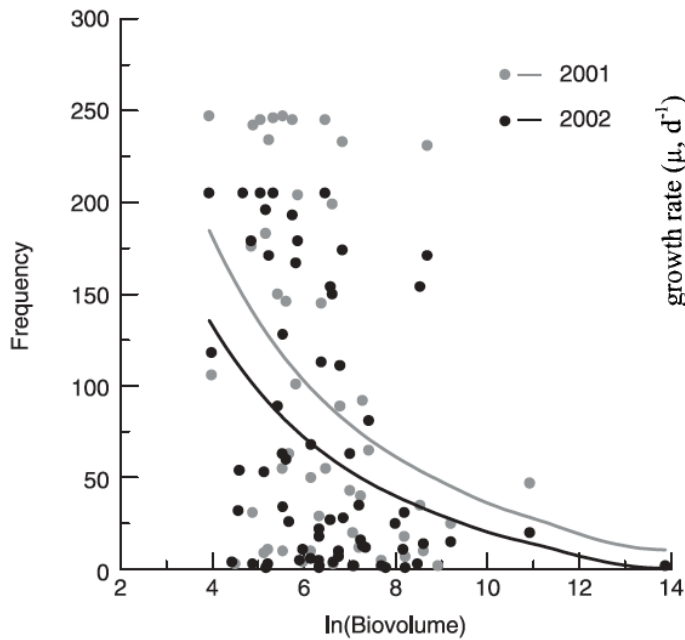


# KORELACE (bodové grafy, xy plots)

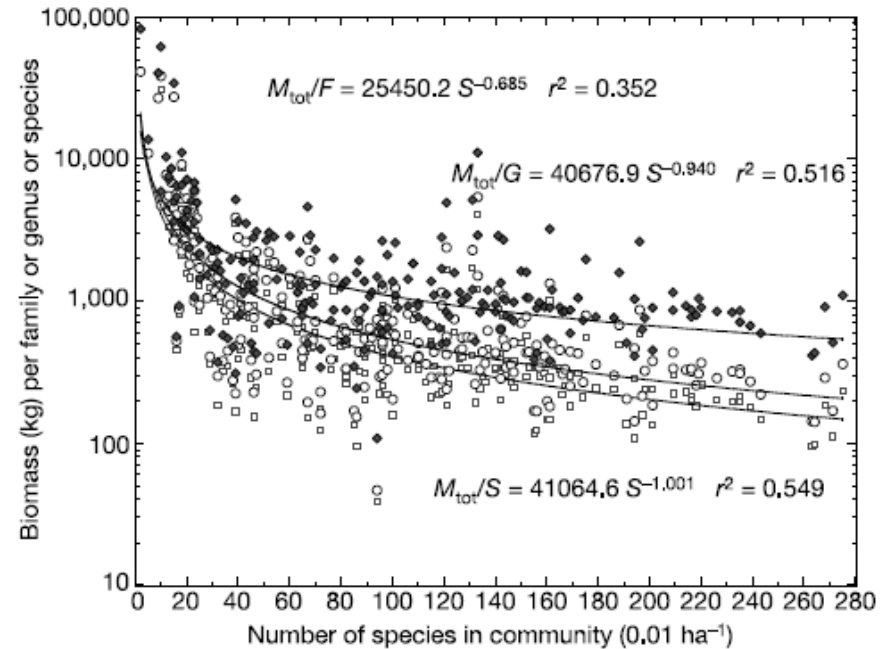




# KORELACE (bodový grafy, xy plots, scatter plots)



Biomass partitioning in global tree communities



# DALŠÍ TYPY GRAFŮ

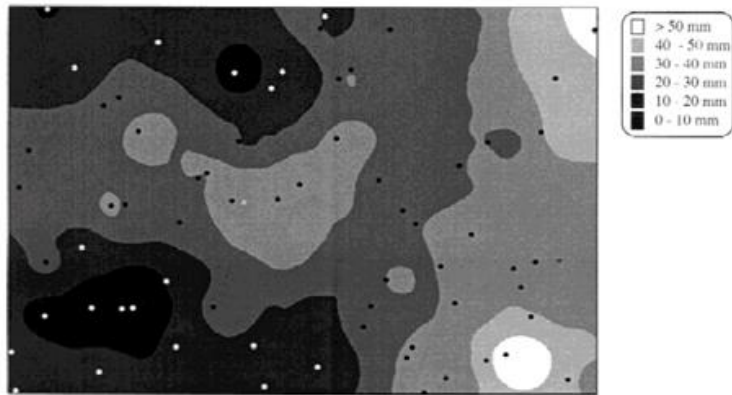


Fig. 6. Interpolated map of the microtopography (altitude in millimeters above an arbitrary level). Gridding parameters as for Fig. 2.

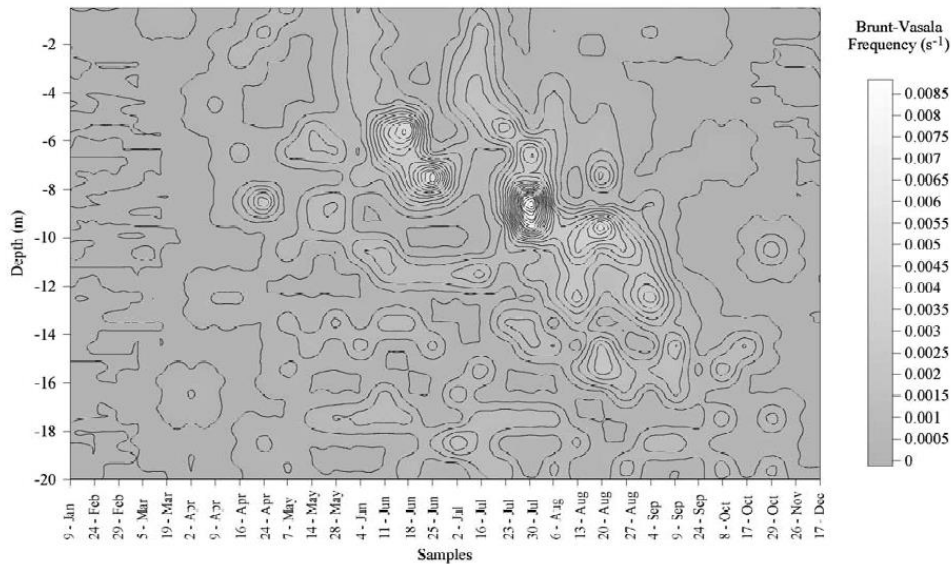
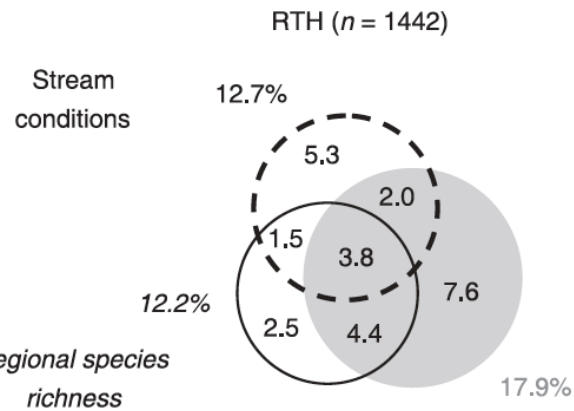
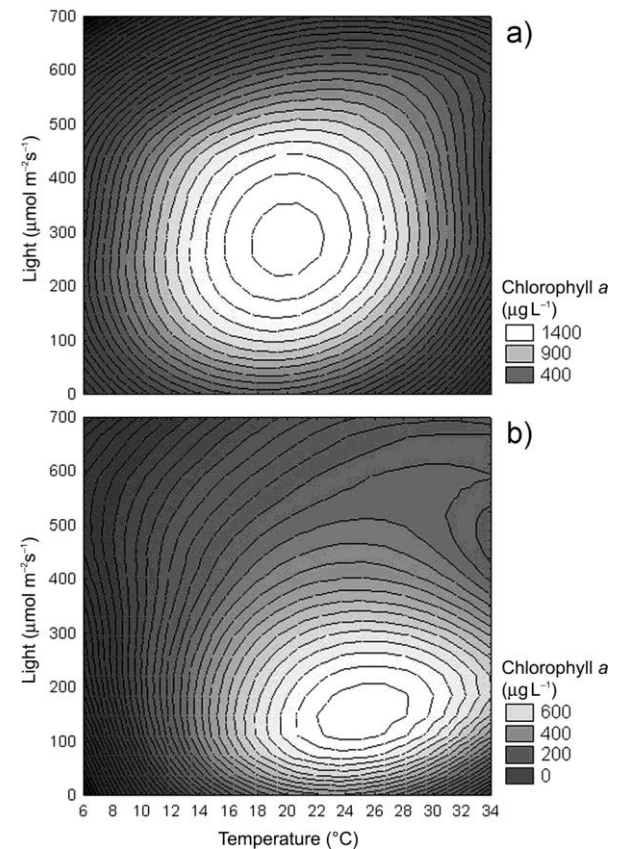


Figure 3. Seasonal changes of the Brunt-Väisälä frequency in the 0–20 m water layer.







# SHLUKOVÉ ANALÝZY (cluster analyses)

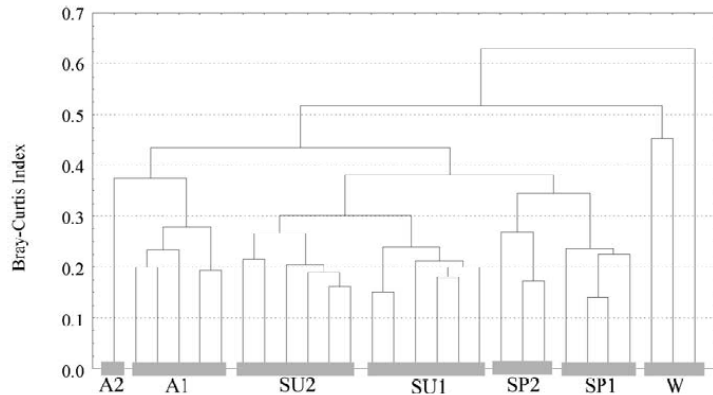
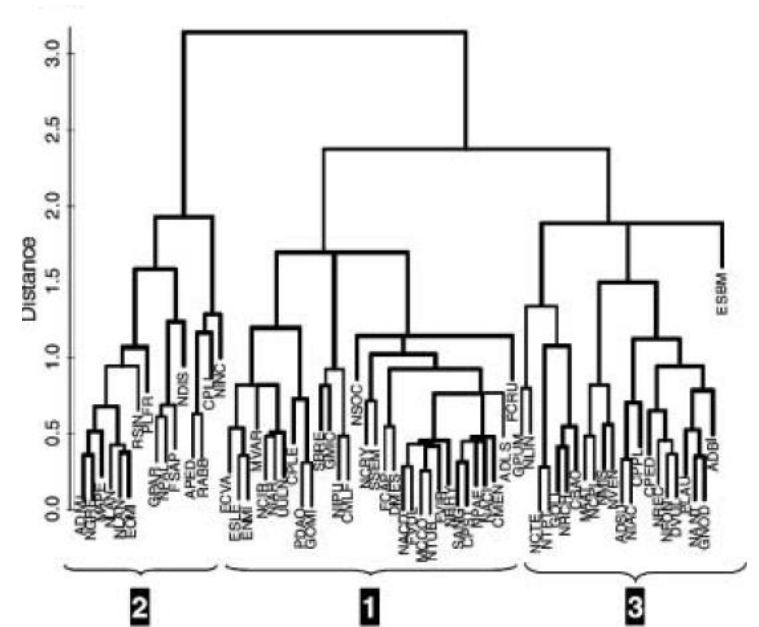
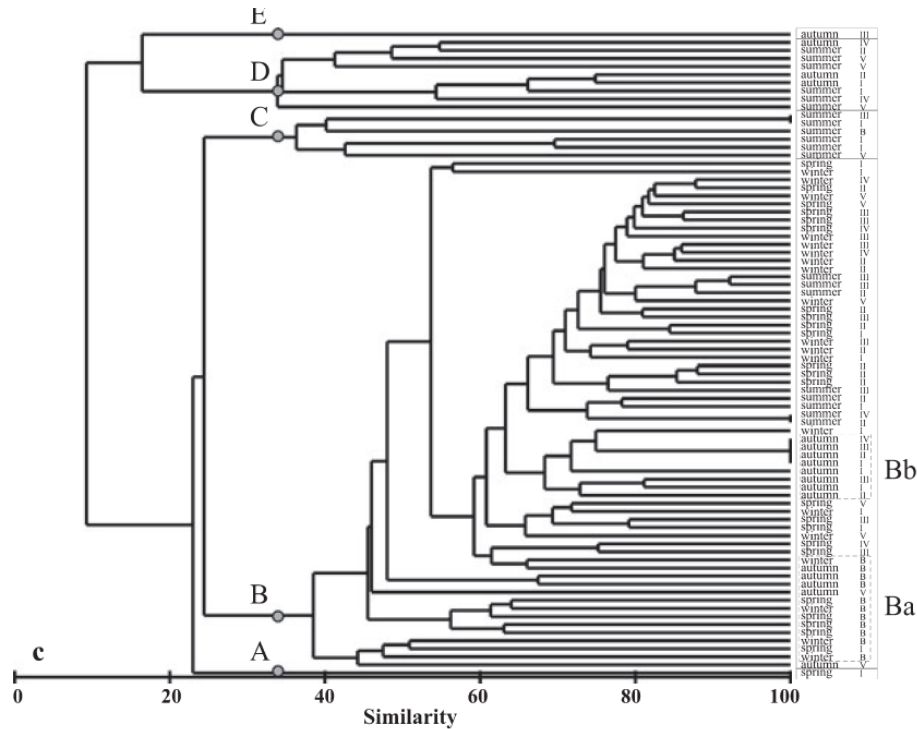
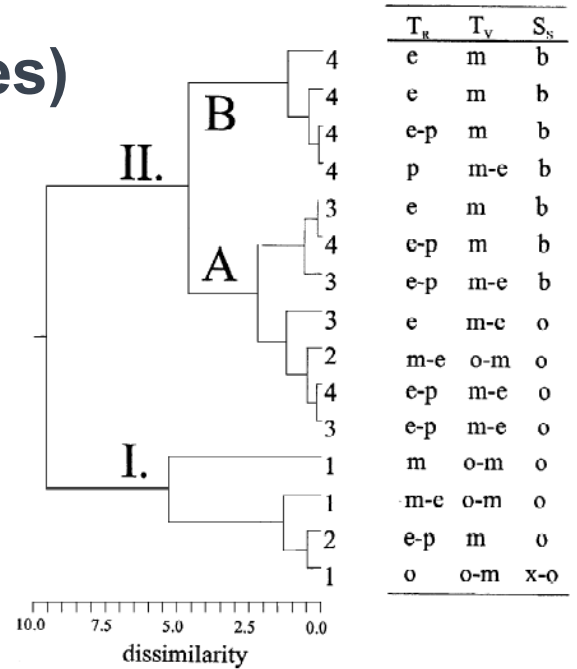
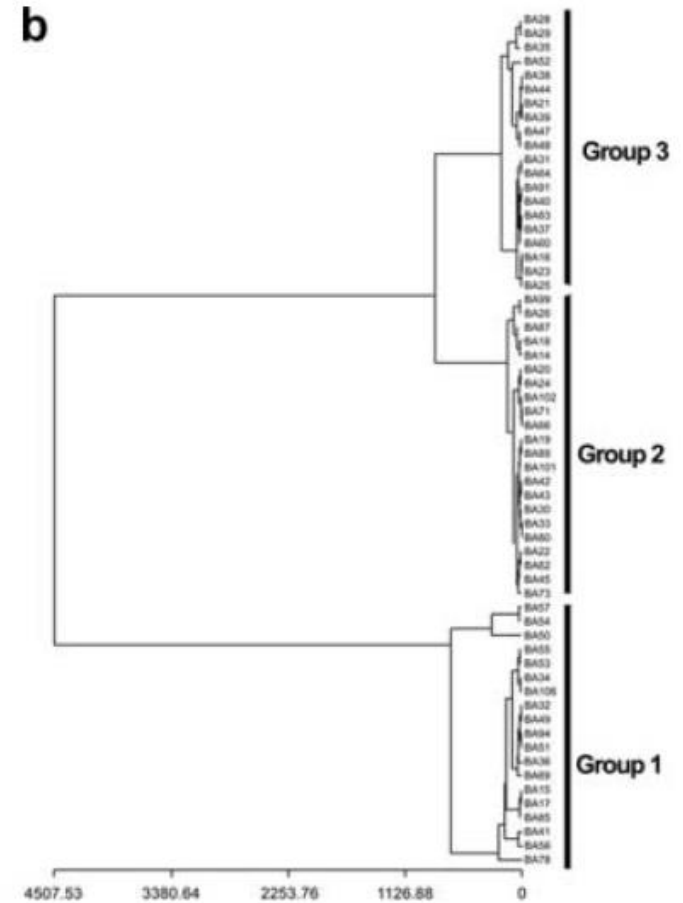
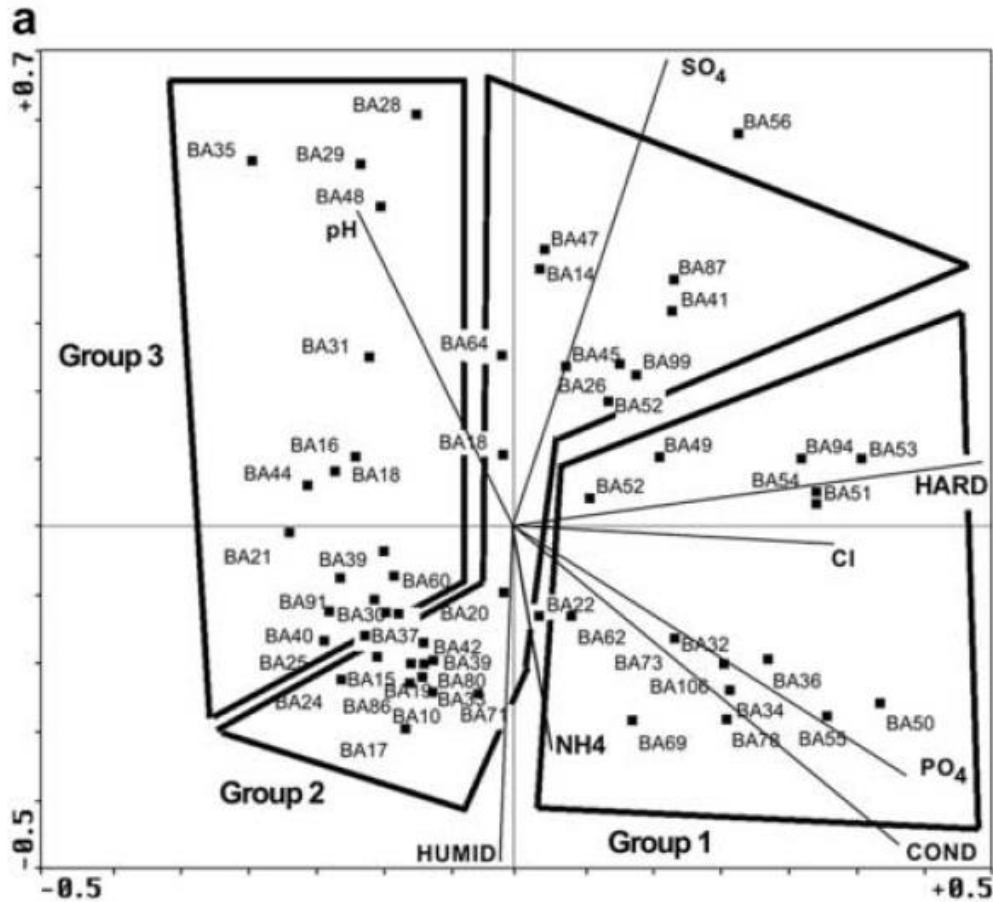


Figure 6. Ordination of the samples in seasonal clusters, according the Bray & Curtis similarity index. W = Winter; SP 1 & 2 = Early Spring & Late Spring; SU 1 & 2 = Early Summer and Late Summer; A1 & 2 = Early Autumn & Late Autumn. The samples are clustered according to the dates reported in Figure 8.



# KOMBINOVANÉ ANALÝZY



# ORDINAČNÍ ANALÝZY

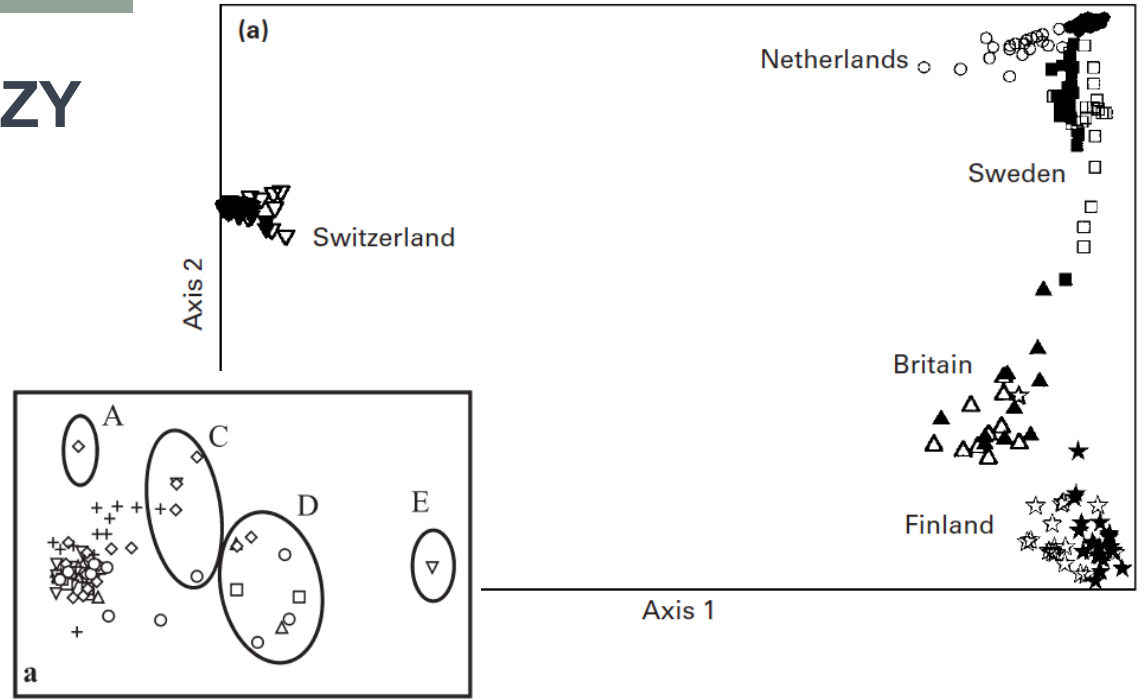
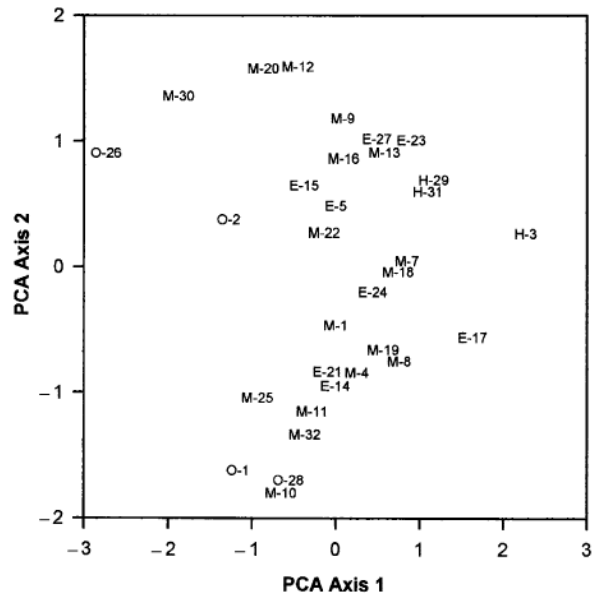
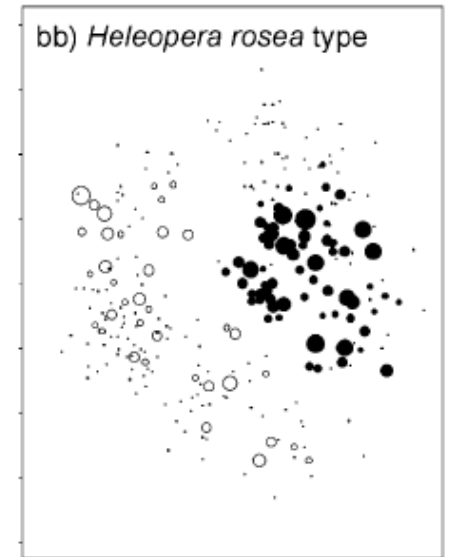
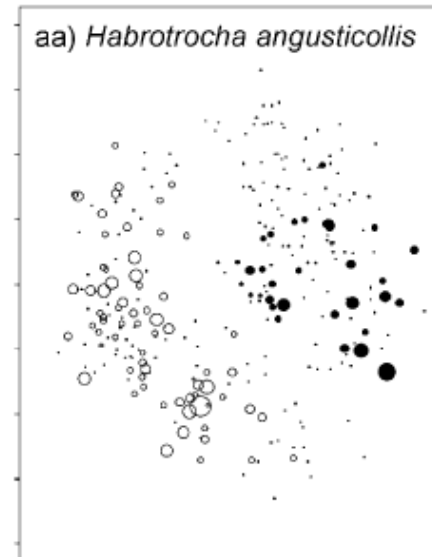
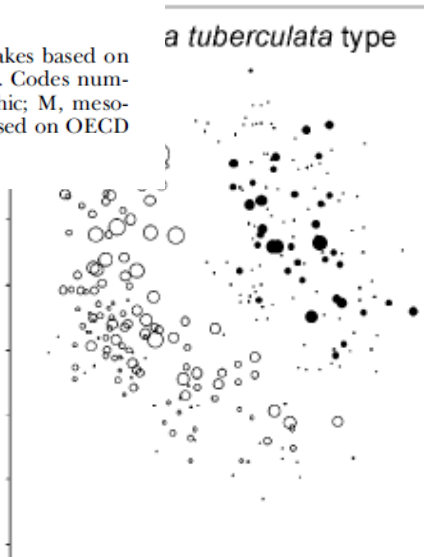
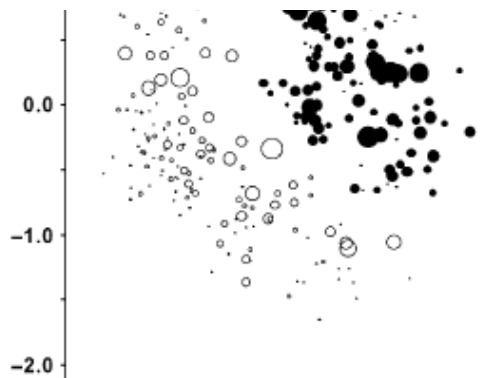


FIG. 2. Principle components ordination of lakes based on the correlation matrix of environmental variables. Codes numbers for lakes are given in Table 1. O, oligotrophic; M, mesotrophic; E, eutrophic; H, hypereutrophic lakes based on OECD classification.



# ORDINAČNÍ ANALÝZY

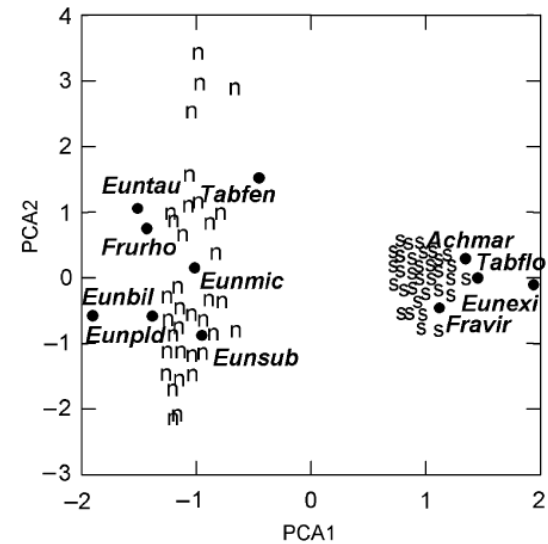
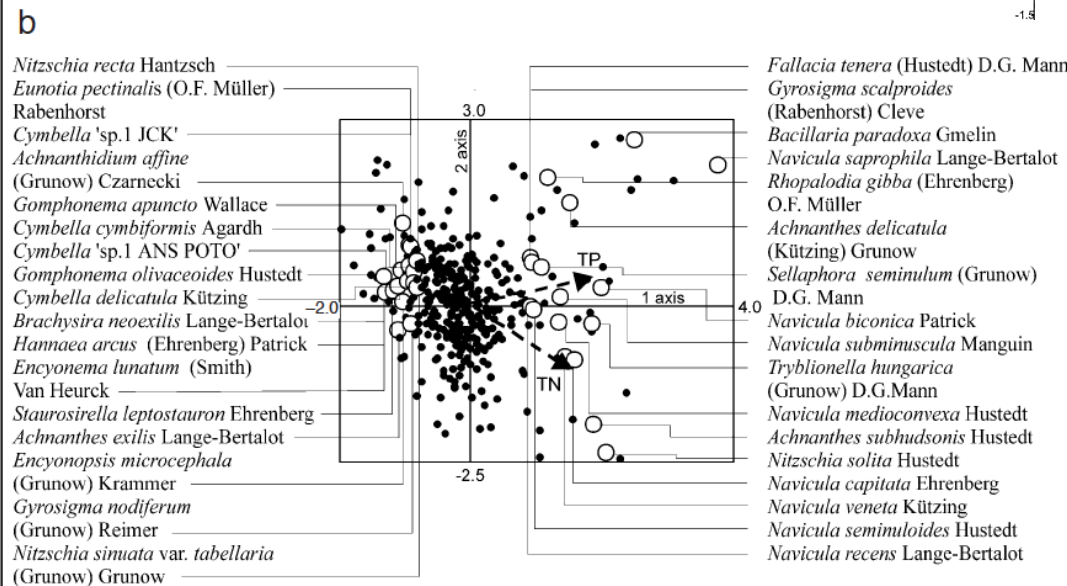
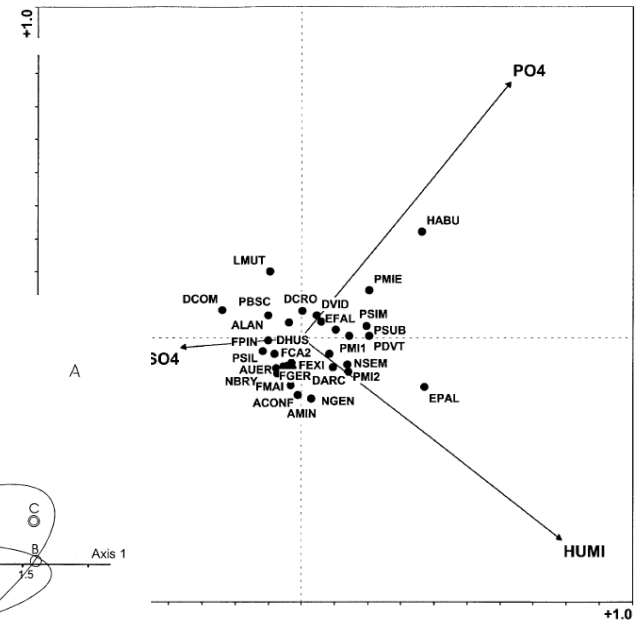
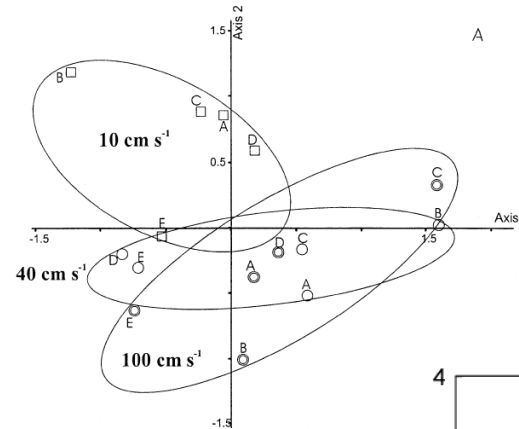
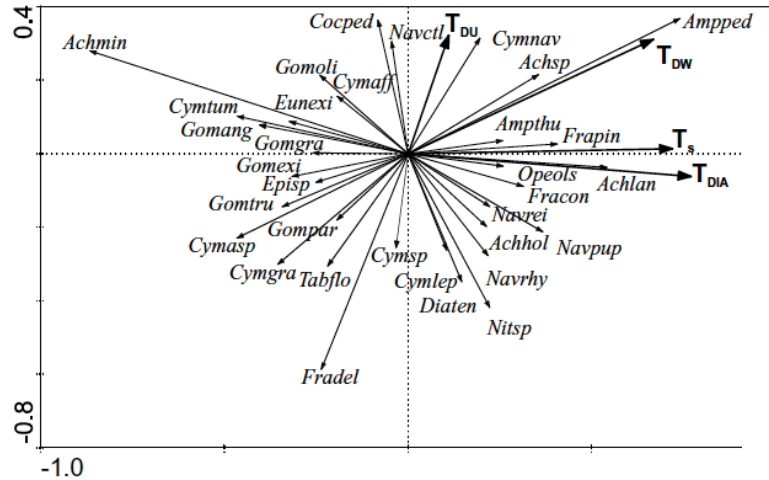


Fig. 4. Canonical Correspondence Analysis ordination diagram showing the principal taxa and their relationship with the environmental variables  
(ACON *Achnanthes confusa*, ALAN *A. lanceolata*, AMIN *A. minutissima*, AUER *A. auri*, DARC *Diademesis arcuata*, DCOM *D. comperi*, DCRO *D. crozetikeruelensis*, DVID *D. ingeae*, DHUS *Diatomella hustedtii*, EFAL *Eunotia fallax*, EPAL *E. paludosa*, FCA2 *Fragilaria capucina*, FEX1 *F. exigua*, FGER *F. germainii*, FMAI *F. maillardii*, FPIN

Fig. 3. Principal component analysis biplot of species scores (multiplied by 2 for clarity) and sample scores from chronically acidified (north = n) and episodically acidified (south = s) tributary.

