

Prostorové autokorelace (a co s nimi)

Ekologická data mají často prostorové a nebo časové uspořádání (tj. sbírána na různých místech a v jiný čas)

Zároveň toto uspořádání může způsobovat že blízko sebe umístěné pozorování budou vzájemně víc podobné, **tj nejsou nezávislé** (teplota vzduch zítra bude víc podobná dnešku než za měsíc, kytky na sousední louce budou víc podobné těm tady než o 100km dál).

S ohledem na tento fakt bychom měli i plánovat sběr dat (jak experimentálních tak korelativních)

Časové autokorelace jsou v podstatě jednodušší případ prostorových autokorelací, kdy je vzájemná podobnost dána jednou dimezí (místo dvěma).

Prostorové autokorelace (a co s nimi)

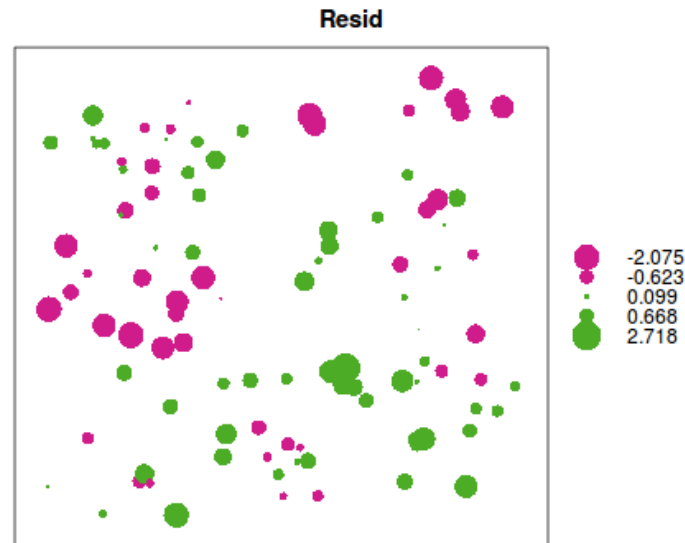
Pokud jsou naše data prostorově strukturované, měli bychom:

- 1. otestovat, zda jsou i prostorově autokorelované.
- 2. v případě, že ano vybrat vhodný model s korelační strukturou, který autokorelaci odstraní.
- 3. následně zkontrolovat, zda opravdu k odstranění autokorelace došlo

Prostorové autokorelace

Detekce prostorové autokorelace

1. Kontrola jestli residuály modelu vykazují systematickou proměnlivost v prostoru



2a. Formální test prostorové autokorelace residuálů (Moran's I..)

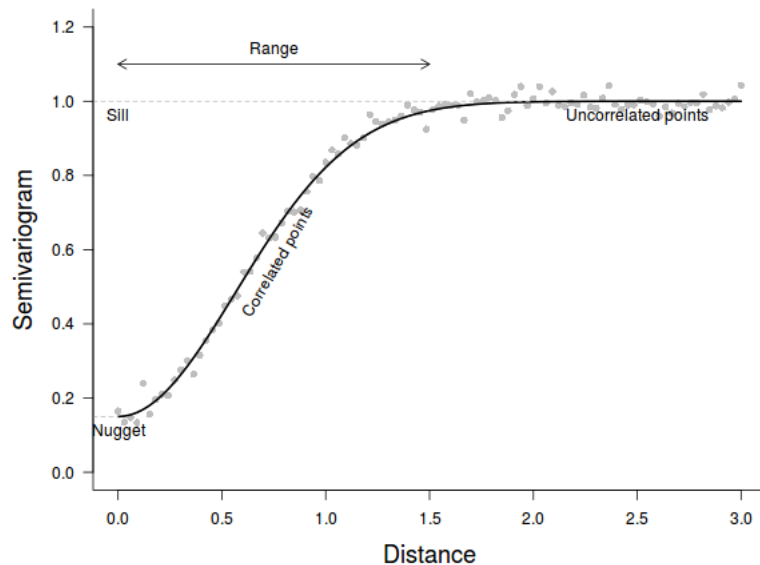
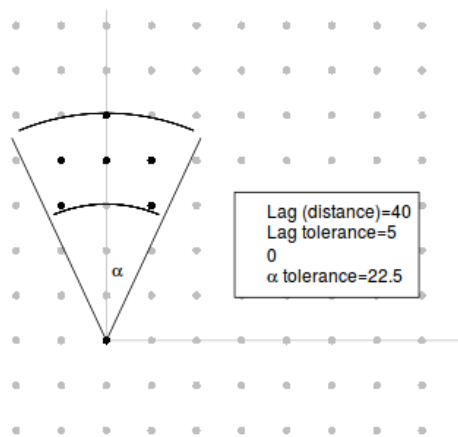
```
library(DHARMA)
testSpatialAutocorrelation(sims, x = dat$x, y = dat$y, plot =
FALSE)
```

Prostorové autokorelace

Detekce prostorové autokorelace

2b. Posouzení residuálů pomocí semi-variogramu

- Variance residuálů pro pozorování v určitých intervalech vzdáleností
- měla by vzrůstat s vzdáleností (semivariance)
- můžeme posoudit i zda je stejná ve všech směrech (izotropní) a nebo ne.



... v případě že prostorovou autokorelaci nedetekujeme, můžeme pokračovat s modely bez ní (tj. ty které už známe)

Prostorové autokorelace

Modely pro prostorově autokorelované data:
Specifikují míru korelace mezi residuály s ohledem na jejich vzdálenost

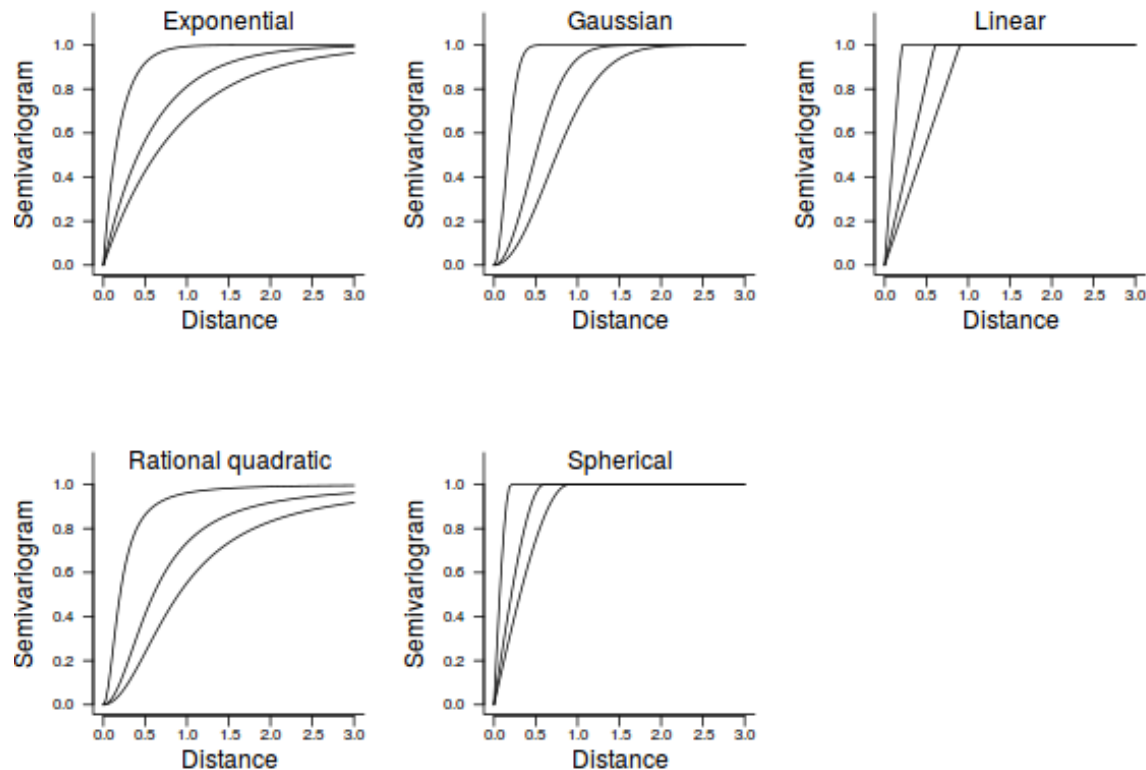
$$y_i = \underbrace{\beta_0 + \beta_1 \times x_i}_{\text{Linearity}} + \varepsilon_i$$

$$\varepsilon_i \sim \underbrace{\mathcal{N}(0, \sigma^2)}_{\text{Normality}}$$

$$\mathbf{V} = \text{cov} = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & \vdots \\ \vdots & \dots & \sigma^2 & \vdots \\ 0 & \dots & \dots & \sigma^2 \end{pmatrix}$$

Homogeneity of variance \leftarrow
 Zero covariance (=independence) \leftarrow

$\text{cor}(\varepsilon) = \begin{pmatrix} 1 & e^{-\delta} & \dots & e^{-\delta D} \\ e^{-\delta} & 1 & \dots & \vdots \\ \vdots & \dots & 1 & \vdots \\ e^{-\delta D} & \dots & \dots & 1 \end{pmatrix}$
 Exponential autoregressive correlation structure



Prostorové autokorelace

Modely pro prostorově autokorelované data:

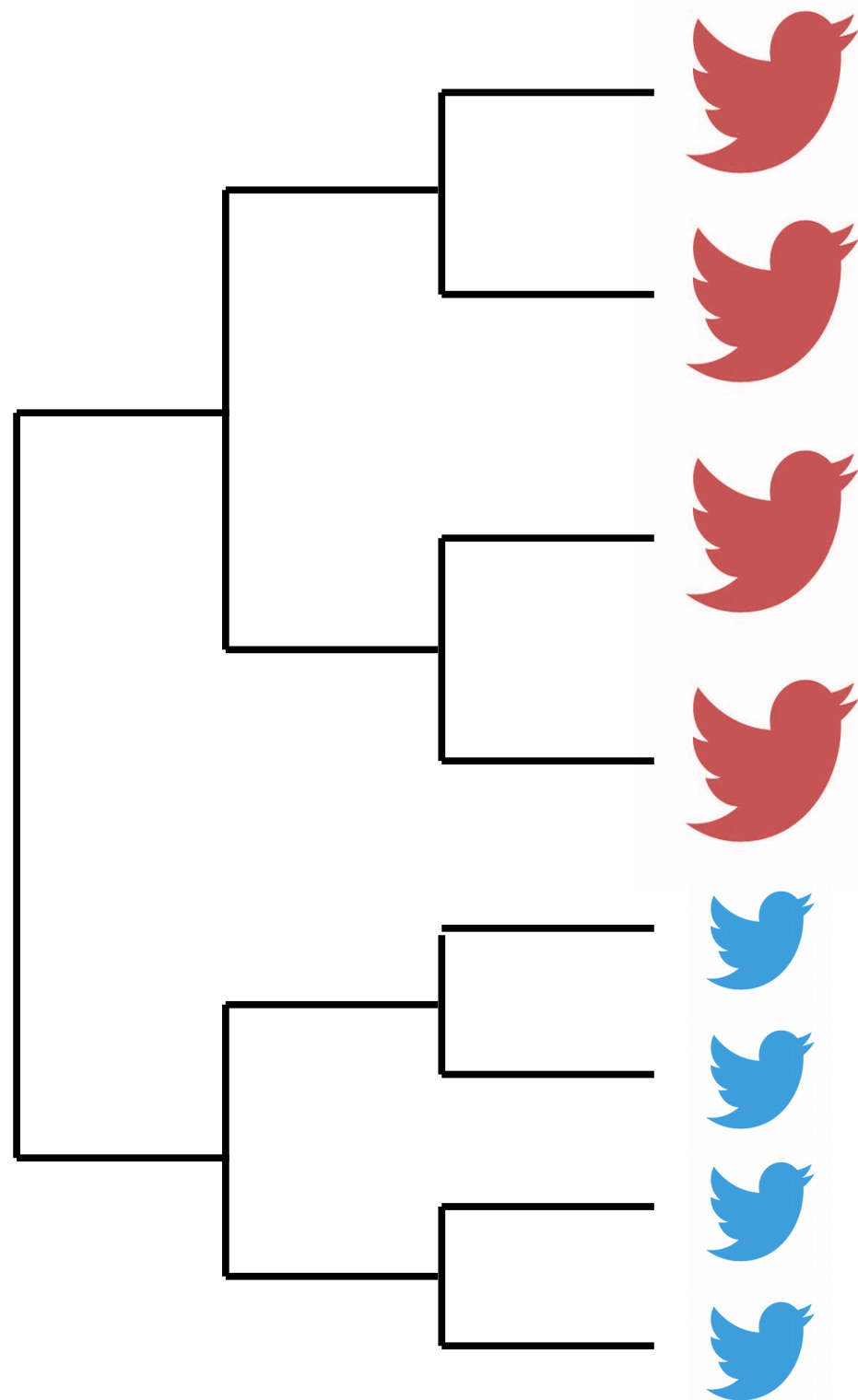
- nlme::gls() - normální rozdělení
- glmmfields - glm pro prostorově uspořádaná data
- nlme::lme() - mixované modely, random efekty + kor. Struktury
 - pouze pro normální rozdělení
- spaMM - mixované modely, pro všechny základní rozdělení
- glmmTMB - mixované modely, všechny možné rozdělení, včetně zero inflated, tweedie...

Celkem dobrý přehled dostupných metod: (Dormann et al. 2007, Ecography)

Souvisí barva s velikostí těla?

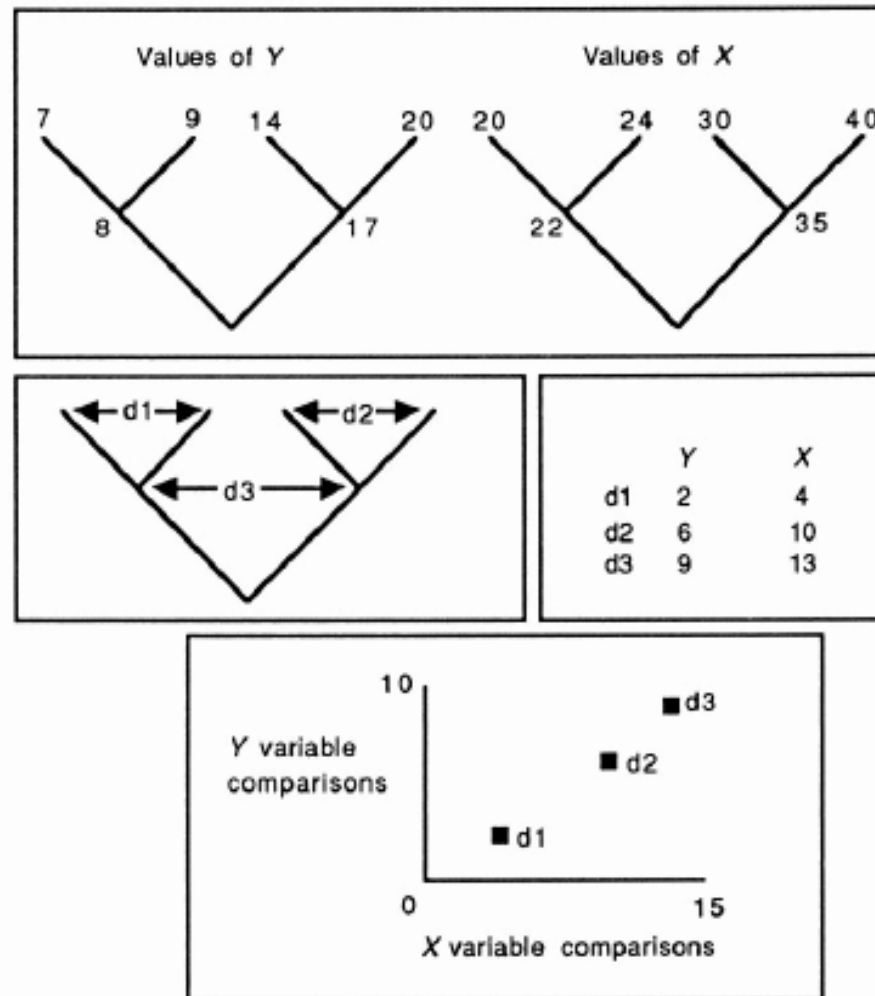


Souvisí barva s velikostí těla?



Phylogenetically independent contrasts

Under a Brownian motion model of evolution, d_1 , d_2 , and d_3 provide independent comparisons. Path length differences are ignored in this illustration.



Fylogenetické, prostorové a časové (auto)korelace pomocí marginálních modelů

Obecný lineární model:

- obsahuje pouze pevné efekty
- pro residuály se předpokládá že (1) jsou nekorelované a (2) jejich rozptyl je konstantní

	A	B	B	D	E
A	1				
B	0	1			
C	0	0	1		
D	0	0	0	1	
E	0	0	0	0	1

Smíšený model (minulá přednáška):

- obsahuje pevné efekty a náhodné efekty (a odhad odpovídajících koeficientů)
- pomocí nich lze specifikovat nezávislost měření v rámci blokového, hierarchického (apod.) uspořádání dat

Fylogenetické, prostorové a časové (auto)korelace pomocí marginálních modelů

Obecný lineární model:

- obsahuje pouze pevné efekty
- pro residuály se předpokládá že (1) jsou nekorelované a (2) jejich rozptyl je konstantní

	A	B	B	D	E
A	1				
B	0	1			
C	0	0	1		
D	0	0	0	1	
E	0	0	0	0	1

Marginální model

- nepracuje s náhodnými efekty
- specifikace heteroskedastity pomocí kovarianční matice
- specifikace závislostí pomocí kovarianční matice

	A	B	C	D	E
A	a				
B	0	a			
C	0	0	b		
D	0	0	0	b	
E	0	0	0	0	b

Rozdílná variance mezi skupinami

	A	B	C	D	E
A	1				
B	a	1			
C	a	a	1		
D	0	0	b	1	
E	0	0	b	b	1

Korelace pro měření uvnitř skupin

Fylogenetické, prostorové a časové (auto)korelace pomocí marginálních modelů

Obecný lineární model:

- obsahuje pouze pevné efekty
- pro residuály se předpokládá že (1) jsou nekorelované a (2) jejich rozptyl je konstantní

	A	B	B	D	E
A	1				
B	0	1			
C	0	0	1		
D	0	0	0	1	
E	0	0	0	0	1

Marginální model

- nepracuje s náhodnými efekty
- specifikace heteroskedastity pomocí kovarianční matice
- specifikace závislosti pomocí kovarianční matice

Tímto způsobem lze efektivně zohlednit pseudoreplikace u blokově uspořádaných data, ale i v případě složitějších typů korelací způsobených fylogenetickými prostorovými a časovými závislostmi

Implementace v R

- normální rozložení – gls (nlme)
- další typy rozložení – gee

Variance/Covariance matrix v PGLS

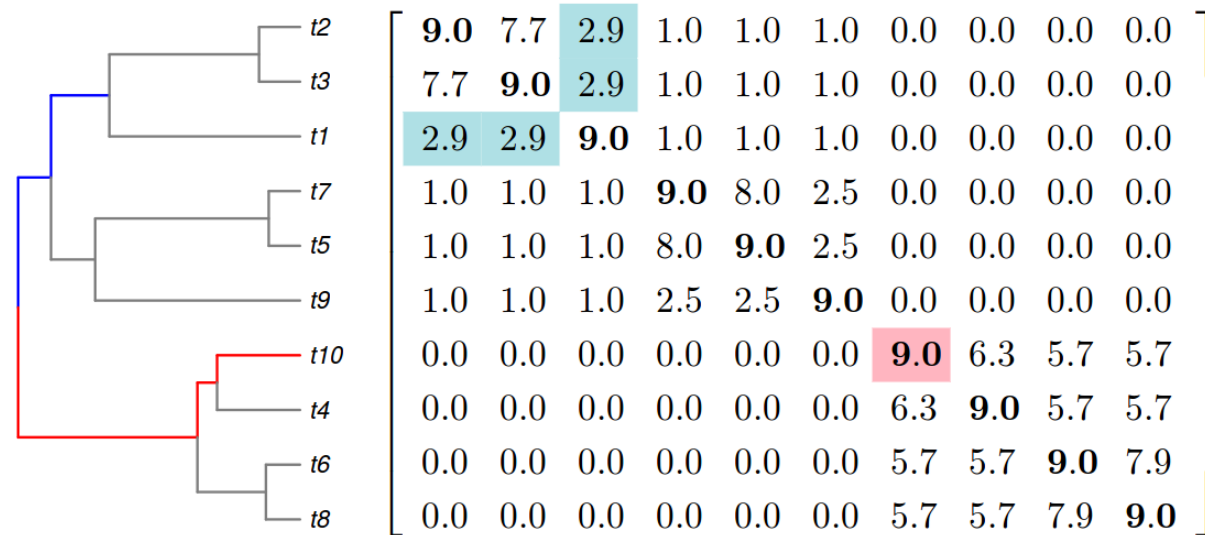


Figure 2: A phylogenetic tree of 10 taxa and the variance covariance matrix (\mathbf{V}) of that phylogeny. The diagonal of the matrix (bold values) shows the path length from each tip to the root (example in red). Off diagonal values show the shared path length for a given pair of tips (example in blue).

Transformace délky větví

- probíhá na základě ML
- cílem je najít takové parametry, pro které výsledná matice nejlépe odpovídá naším datům
- ML hodnotu “Lambda” lze použít jako měřítko fylogenetického signálu v datech

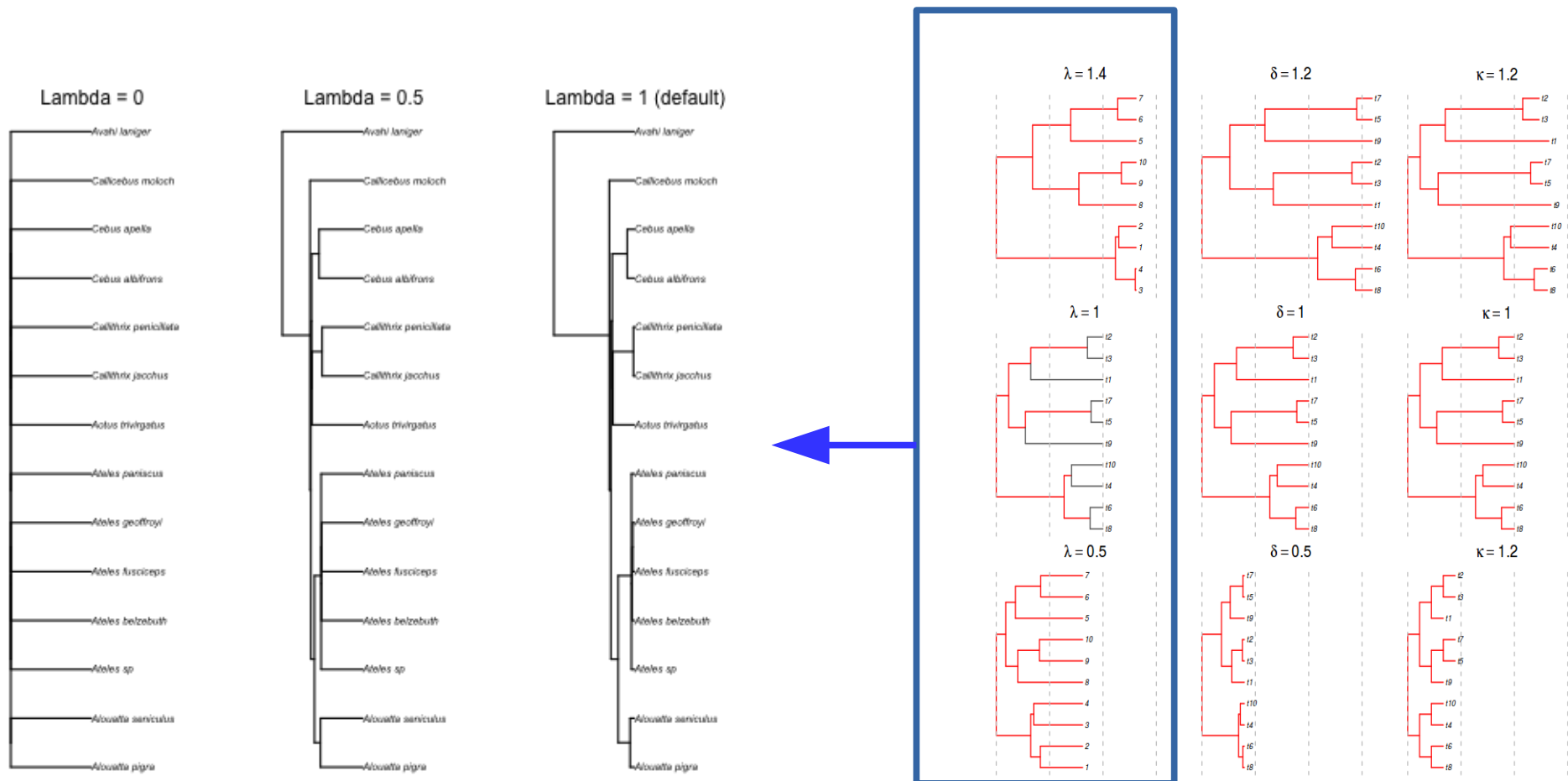


Figure 3: Examples of λ , δ and κ branch length transformations. The branches affected by a given transformation are shown in red.

Transformace délky větví

- probíhá na základě ML
- cílem je najít takové parametry, které (resp. výsledná matice) nejlépe odpovídá naším datům
- ML hodnotu “Lambda” lze použít jako měřítko fylogenetického signálu v datech

Lambda -prodlužuje/zkracuje terminální větve na úkor bazálních

0 ~ nezávislost znaku na fylogenezi

1 ~ evoluce znaku odpovídá fylogenezi

Delta – prodlužuje/zkracuje celý strom

< 1 adaptivní radiace znaku

> 1 akcelarace evoluce v průběhu fylogeneze

Kappa – prodlužuje/zkracuje dlouhé větve na úkor krátkých

>1 akcelarace evoluce v dlouhých větvích

0 bodová evoluce

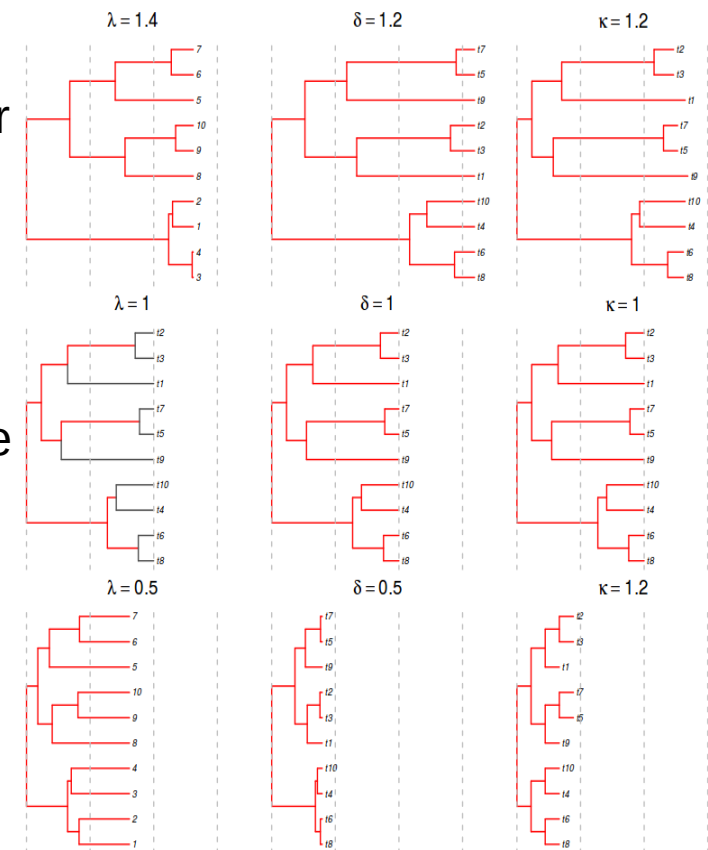


Figure 3: Examples of λ , δ and κ branch length transformations. The branches affected by a given transformation are shown in red.

Další aspekty fylogenetický komparativních metod (kterými se nebudeme zabývat)

Model evoluce znaku:

Brownův pohyb – “náhodná procházka” po možných hodnotách znaku

Ornstein-Uhlenbeck – předpokládá selekční optima, která se mohou (ale nemusí) lišit pro různé druhy

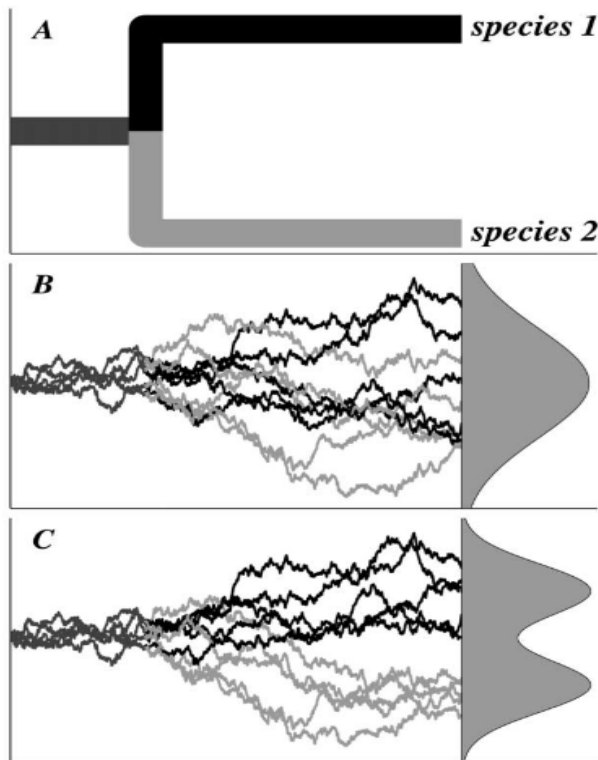


Figure 3: Simulations of Brownian motion (BM) and multiple-optimum Ornstein-Uhlenbeck (OU) processes with branching. Phenotypic value on Y-axis, time along X-axis. A, A single lineage evolves until a speciation event occurs from which two lineages (*light gray, black*) emerge. Lineages evolve according to BM, with light gray and black lineages evolving independently after the speciation event. C, Lineages evolve under an OU model. A single selective regime exists before the speciation event; two distinct selective regimes afterward. Distributions of the evolving quantitative character are shown at right. Brownian motion produces a unimodal trait distribution centered at the initial (ancestral) value, whereas the OU process results in a multimodal trait distribution with peaks near each optimum. An animation of this process is provided in the online edition of the *American Naturalist*.

Další aspekty fylogenetický komparativních metod (kterými se nebudeme zabývat)

Vnitrodruhová variabilita:

Mělo by se zohlednit v případě že je nenulová

Lépe než s pglS se dá specifikovat pomocí MCMCglmm:

<http://www.mpcm-evolution.org/practice/online-practical-material-chapter-11/chapter-11-2-multiple-measurements-model-mcmcglmm>

Literatura a další zdroje:

Pro zobrazování dat ve fylognetickém kontextu:

<https://guangchuangyu.github.io/software/ggtree/>

