Phylogenetic comparative methods: a statisticians perspective

Krzysztof Bartoszek

Linköping University

krzysztof.bartoszek@liu.se

21 X 2025 (Zoom)

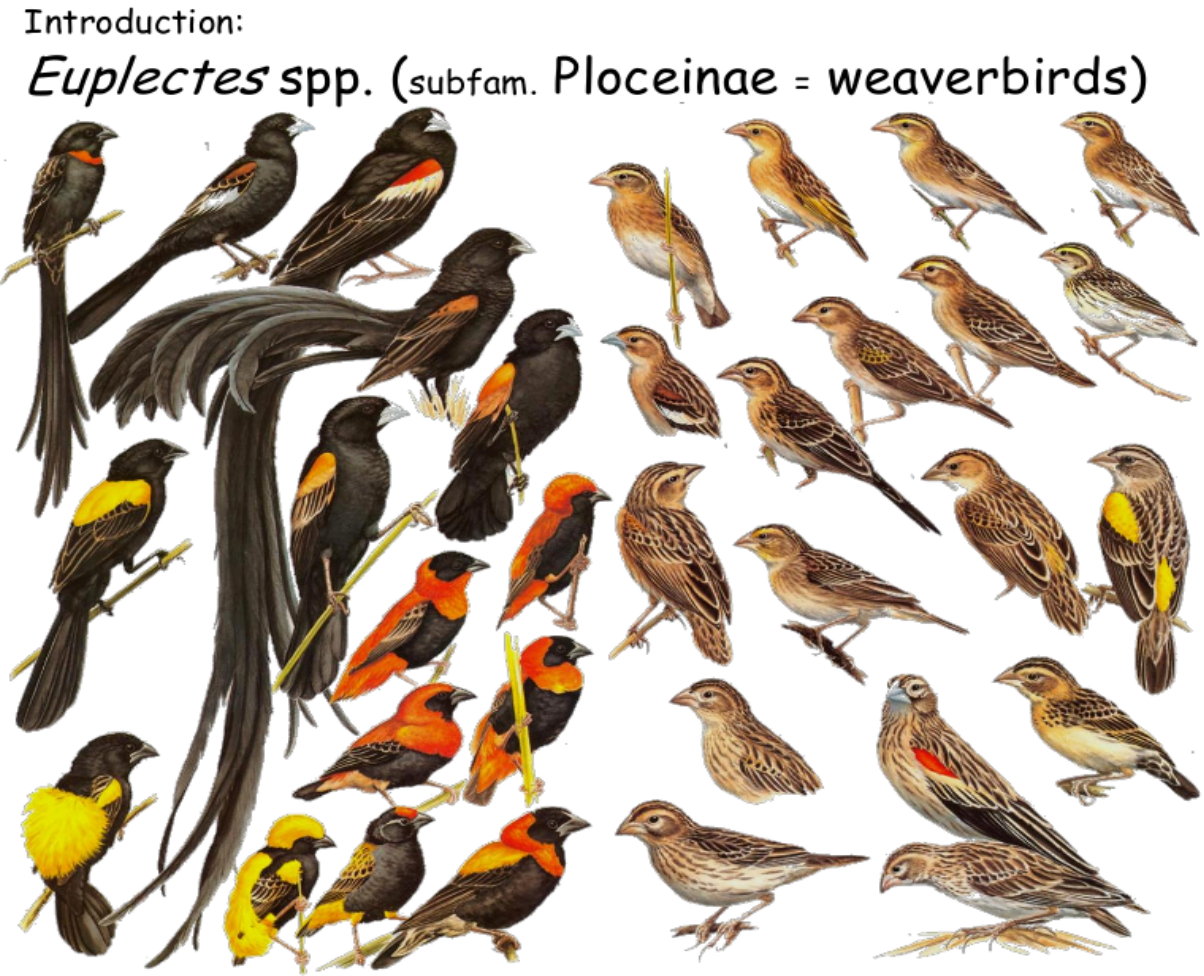

Oslo
Drammen
Fredrikstad
Karlstad
Örebro
Västerås
Uppsala
Stockholm
Linköping University
Interdisciplinary
school for 140 courses
Kristiansand
Gothenburg
LINKÖPING
UNIVERSITY
Gotland
Vendsyssel
Aalborg
Aarhus
Copenhagen
Denmark
Sjælland
Malmö
Odense
Gdynia

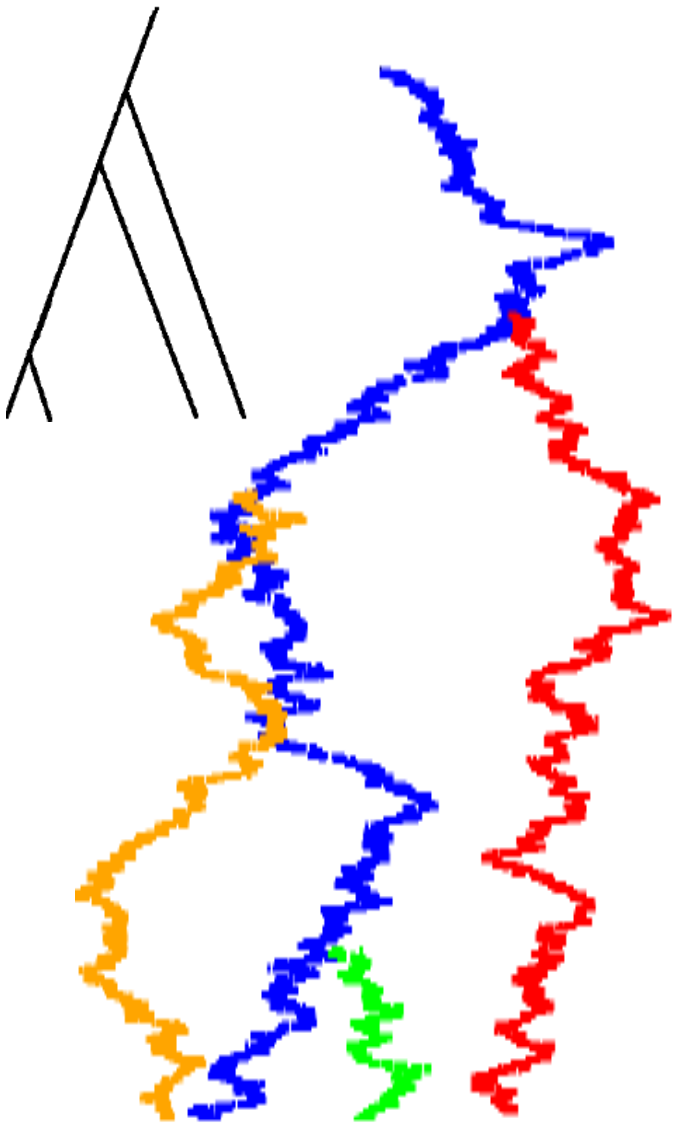
Google Maps

The setup

Species evolve independently after speciation

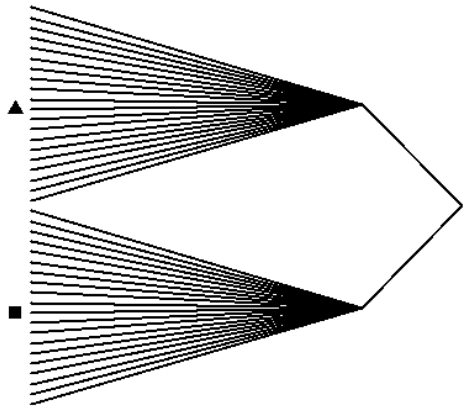
Trait: "average" for species



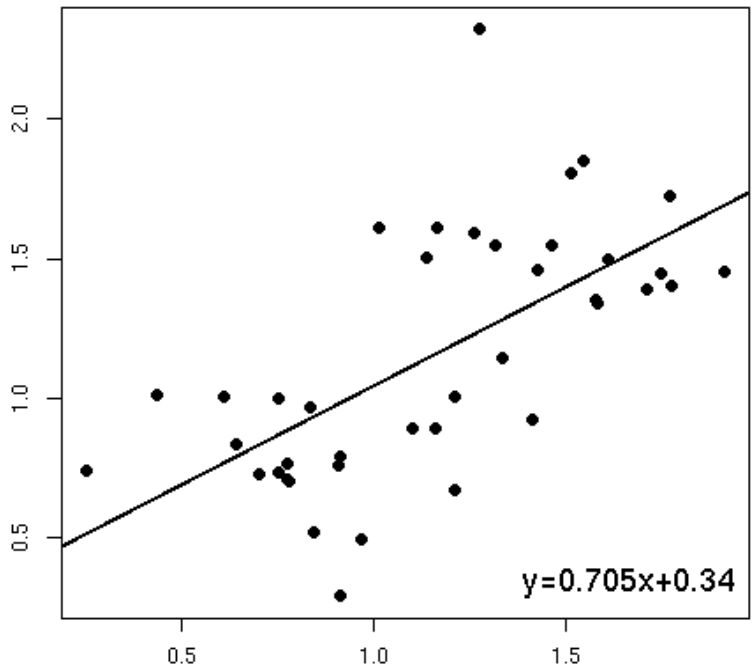


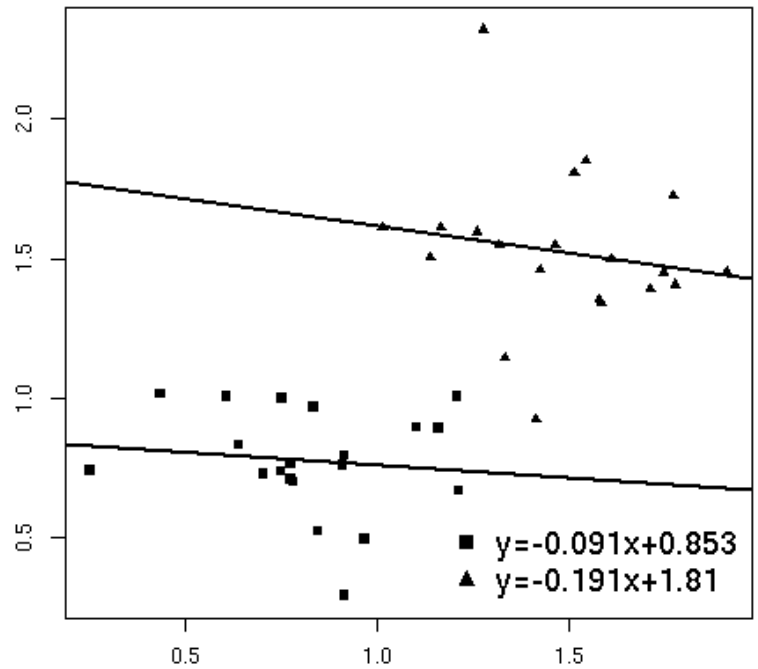
Paradox?

See also J. Felsenstein, 1985. Phylogenies and the comparative method. Am. Nat. 125(1) 1–15



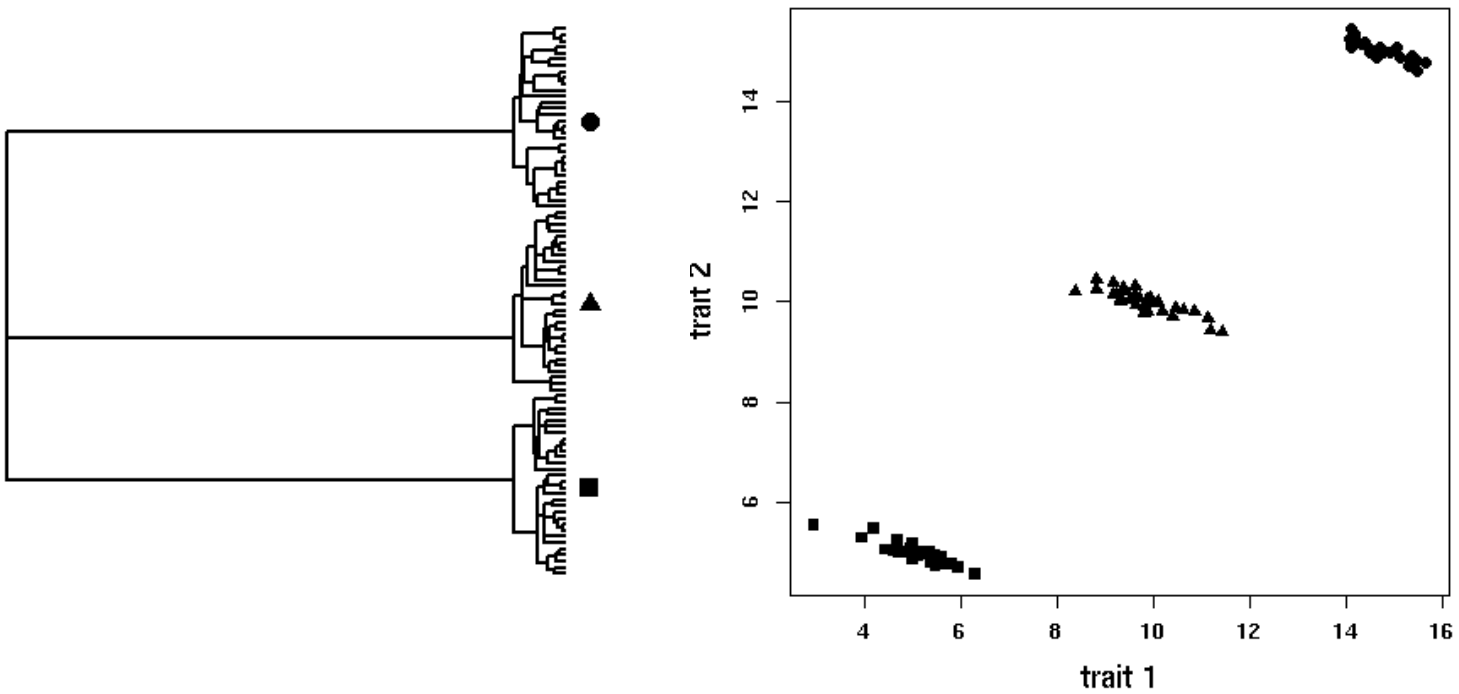
See code: `Script_01_ClusteresPhylogeny.R`





Simpson's paradox

$$\text{Cov}\left[Y_i, Y_j\right](t)$$



See code: `Script_02_SimpsonsParadox.R`

Eigendecomposition

We say that for a square matrix $\mathbf{A} \in \mathbb{R}^{k \times k}$, $\mathbb{R} \ni \lambda$ and $\vec{e}_\lambda \in \mathbb{R}^k$ form an eigenvalue, eigevector pair if they satisfy:

$$\mathbf{A}\vec{e}_\lambda = \lambda \vec{e}_\lambda.$$

We say that for a square matrix $\mathbf{A} \in \mathbb{R}^{k \times k}$, is eigendecomposable if there are k eigenvalue, eigenvector pairs $(\lambda_i, \vec{e}_{\lambda_i})$ such that the vectors $\vec{e}_{\lambda_i}$ are linearly independent
(i.e., $a_1\vec{e}_{\lambda_1} + \ldots + a_k\vec{e}_{\lambda_k} = \vec{0}$ iff $a_1 = \ldots = a_k = 0$).
The eigenvalues may repeat themselves.

Eigendecomposition

Let $\mathbf{A} \in \mathbb{R}^{k \times k}$ be eigendecomposable with $(\lambda_1, \vec{e}_{\lambda_1}), \ldots, (\lambda_k, \vec{e}_{\lambda_k})$ eigenvalue, eigenvector pairs. Denote

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ 0 & 0 & \ddots & \ldots & 0 \\ \vdots & \vdots & \vdots & \ldots & 0 \\ 0 & \ldots & \ldots & \ldots & \lambda_k \end{bmatrix} \quad \mathbf{P} = \begin{bmatrix} | & | & \ldots & | \\ \vec{e}_{\lambda_1} & \vec{e}_{\lambda_2} & \ldots & \vec{e}_{\lambda_k} \\ | & | & \ldots & | \end{bmatrix}$$

$$\text{then we have } \mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$$

Not every matrix is eigendecomposable, example *defective* matrix

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

Ordinary differential equations (ODE)

Derivative: defines a functions change
Define a function by its derivative :

$$f'(x) = g(x)$$

then

$$f(x) = \int g(x)\mathrm{d}x + C$$

fixing $f(x_0) = y_0$ gives us C
$g(x)$ can depend on f **!!!!**

$$f'(x) = g(x, f(x)), \quad f(x_0) = y_0$$

ODE: equation where the unknown is the function f

ODE examples

$$f'(x) = c,\ f(0) = 0, \qquad f(x) = cx$$

$$f'(x) = \alpha f(x),\ f(0) = c_0, \qquad f(x) = c_0 e^{\alpha x}$$

$$f'(x) = -\alpha f(x),\ f(0) = c_0, \qquad f(x) = c_0 e^{-\alpha x}$$

$$f'(t) = -\alpha(f(t) - \theta),\ f(0) = x_0, \qquad f(t) = x_0 e^{-\alpha t} + (1 - e^{-\alpha t})\theta$$

ODE system (multivariate ODE)

Derivatives are taken dimension by dimension.

Examples

$$f : \mathbb{R}^{k_1} \to \mathbb{R}^{k_2}, f(\vec{0}) = \vec{0}, f'(\vec{x}) = \vec{c}, \quad f(\vec{x}) = [\mathrm{diag}(\vec{c})]\,\vec{x}$$

$$f : \mathbb{R} \to \mathbb{R}^k, f(0) = \vec{x_0}, f'(t) = -\mathbf{A}(f(t) - \vec{\theta}),$$
$$f(t) = e^{-\mathbf{A}t}\vec{x_0} + (1 - e^{-\mathbf{A}t})\theta$$

What is $e^{\mathbf{M}}$?

$$e^{\mathbf{M}} = \sum_{i=0}^{\infty} \frac{\mathbf{M}^i}{i!}$$

$M = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$, $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_k)$ (eigendecomposition, if exists)

$$e^{\mathbf{M}} = \mathbf{P}e^{\mathbf{\Lambda}}\mathbf{P}^{-1}$$

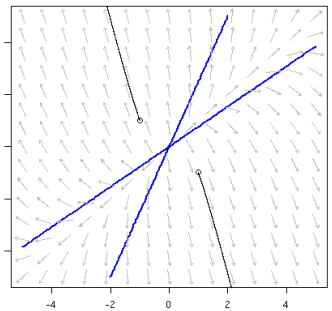
$f'(t) = -\mathbf{A}(f(t))$ 2D

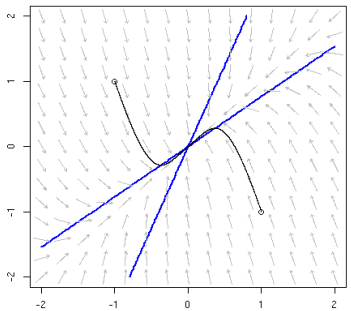
See code: `Script_03_2DODE.R`

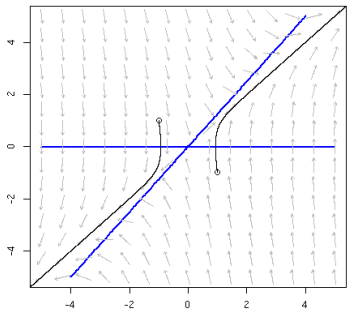
$\lambda_1, \lambda_2 < 0$
unstable fixed point

$\lambda_1, \lambda_2 > 0$
∞stable fixed point

$\lambda_1 \lambda_2 < 0$
unstable saddle point







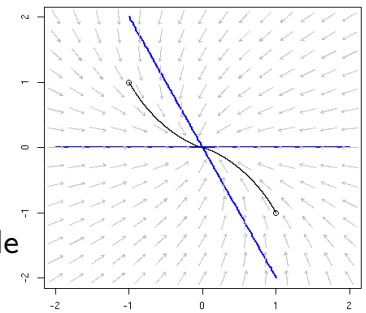
$\mathbb{R} \ni \lambda_1 \neq \lambda_2 \in \mathbb{R}$

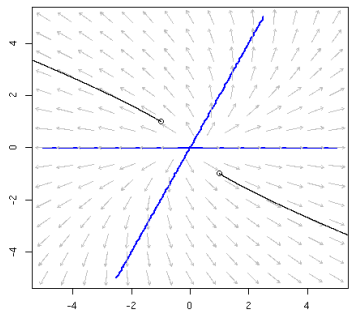
1eigenvector degenarate

2 eigenvectors, star

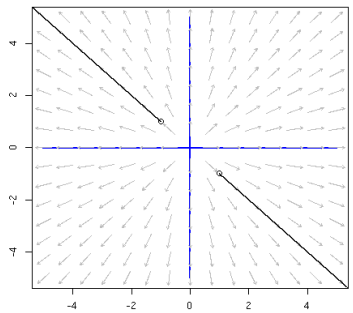
$\lambda_1 = \lambda_2 \in \mathbb{R}$

$\lambda < 0$
unstable

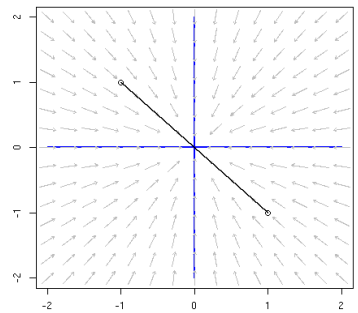




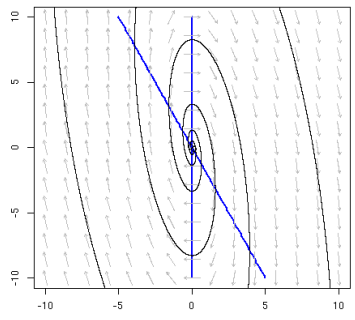
$\lambda > 0$ ∞stable

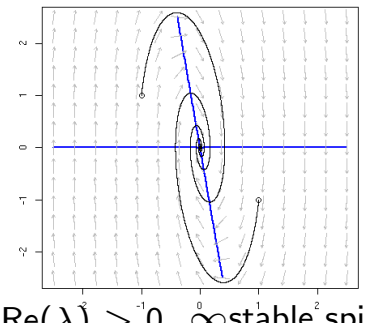


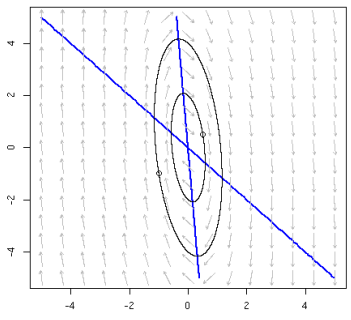
$\lambda < 0$ unstable



$\lambda > 0$
∞stable







$\lambda_1, \lambda_2 \in \mathbb{C}$ $\mathrm{Re}(\lambda) < 0$
unstable spiral

$\mathrm{Re}(\lambda) > 0$ ∞stable spiral

$\mathrm{Re}(\lambda) = 0$
stable centre

ODE plotting

If ODE solvable, then direct graph

Otherwise:

1. choose small time step Δ
2. $f(t_0) = x_0$ initial condition and set $t = t_0$
3. $f(t + \Delta) = f(t) + \Delta g(t)$
4. $t = t + \Delta$
5. Repeat steps 3 and 4 until t reaches t_{max}

multivariate ODE system in the exact same way

Stochastic differential equations (SDE)

Stochastic process: $X(t)$ "a collection of random variables that together form a random function of time"

Intuitive idea: ODE + random noise
But apart from special cases this does not work in this way

Special case: linear SDE (considered here)
Stochastic calculus
Itô formula

Applications:
- Finance
- Biology

Brownian motion

Brownian motion, $B(t)$ is a stochastic process with the following properties

1. (Independence of increments) $B(t) - B(s)$ for $t > s$ is independendent of the past, that is of $B(u)$ for $0 \leq u \leq s$.

2. (Normal increments) $B(t) - B(s)$ is normally distributed with mean 0 and variance $t - s$

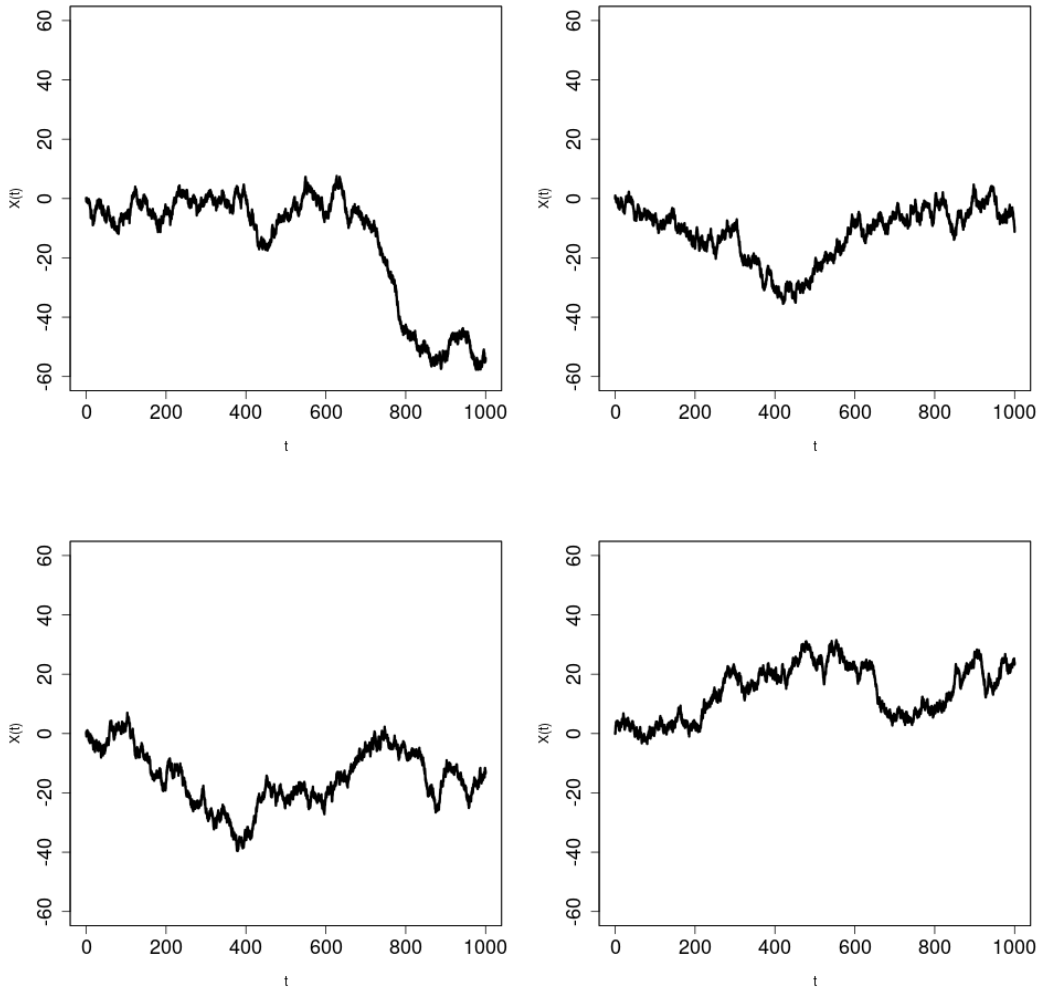
3. (Continuity of paths) $B(t)$ is a continuous function of $t \geq 0$

$$\mathrm{E}\left[B(t)\right] = B(0)$$
$$\mathrm{Var}\left[B(t)\right] = t$$

Brownian motion

$$X_0 = 0,\ \sigma = 1$$

See code: `Script_04_1DBMyuima.R` simulations by `yuima` R package



Diffusion type SDE

$$\mathrm{d}X(t) = \underbrace{\mu(t, X(t))\mathrm{d}t}_{\text{“ODE”}} + \underbrace{\sigma(t, X(t))\mathrm{d}B(t)}_{\text{“noise”}} \qquad X(0) = X_0$$

What is this? It is shorthand for: “white noise”

$$X(t) = \int_0^t \mu(s, X(s))\mathrm{d}s + \int_0^t \sigma(s, X(s))\mathrm{d}B(s)$$

The solution to an SDE is a stochastic process that satisfies the above integral equation.

$\int_0^T X(s)\mathrm{d}B(s)$ is a “special” integral—Itô integral
$X(t)$ such that $\int_0^T \mathrm{E}\left[X(s)^2\right]\mathrm{d}s < \infty$, then $\mathrm{E}\left[\int_0^T X(s)\mathrm{d}B(s)\right] = 0$

In some special cases we know what the distribution of $X(t)$.

SDE examples

$$\mathrm{d}X(t) = \mathrm{d}B(t) \quad X(t) = X_0 + B(t), \quad X(t) \sim \mathcal{N}(X_0, t)$$

$$\mathrm{d}X(t) = \sigma\mathrm{d}B(t) \quad X(t) = X_0 + \sigma B(t), \quad X(t) \sim \mathcal{N}(X_0, \sigma^2 t)$$

$$\mathrm{d}X(t) = \mu\mathrm{d}t + \sigma\mathrm{d}B(t) \quad X(t) = X_0 + \mu t + \sigma B(t), \quad X(t) \sim \mathcal{N}(X_0 + \mu t, \sigma^2 t)$$

$$\mathrm{d}X(t) = \mu\mathrm{d}t + \sigma\mathrm{d}B(t) \quad X(t) = X_0 + \mu t + \sigma B(t), \quad X(t) \sim \mathcal{N}(X_0 + \mu t, \sigma^2 t)$$

linear type SDE:

$$\mathrm{d}X(t) = (c(t) + d(t)X(t))\mathrm{d}t + \sigma(t)\mathrm{d}B(t) \quad \text{denote } h(t) = \int_0^t d(s)\mathrm{d}s$$

$$X(t) = e^{h(t)}X_0 + e^{h(t)}\int_0^t e^{-h(s)}c(s)\mathrm{d}s + e^{h(t)}\int_0^t e^{-h(s)}\sigma(s)\mathrm{d}B(s)$$

$$d(t) \equiv d \quad X(t) = e^{Dt}X_0 + \int_0^t e^{d(t-s)}c(s)\mathrm{d}s + \int_0^t e^{d(t-s)}\sigma(s)\mathrm{d}B(s)$$

Ornstein–Uhlenbeck (OU) process

$$\mathrm{d}X(t) = -\alpha(X(t) - \theta)\mathrm{d}t + \sigma \mathrm{d}B(t), \quad X(0) = X_0$$

$$\mathrm{E}\left[X(t)\right] = e^{-\alpha t}X_0 + (1 - e^{-\alpha t})\theta \xrightarrow{\alpha>0} \theta$$

$$\mathrm{Var}\left[X(t)\right] = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t}) \xrightarrow{\alpha>0} \frac{\sigma^2}{2\alpha}$$

$\alpha > 0$: converges (weakly) to $\mathcal{N}(\theta, \sigma^2/(2\alpha))$

Half–life time to lose half of ancestral

$$\exp(-\alpha t_{0.5}) = 0.5 \text{ then } t_{0.5} = \ln 2/\alpha$$

T.F. Hansen, 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51, 1341-1351

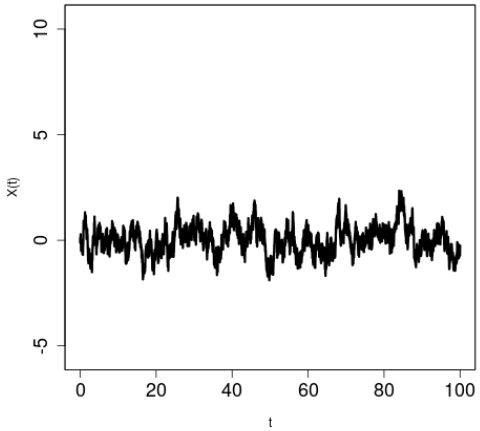
Interest rates model (Vašíček model)

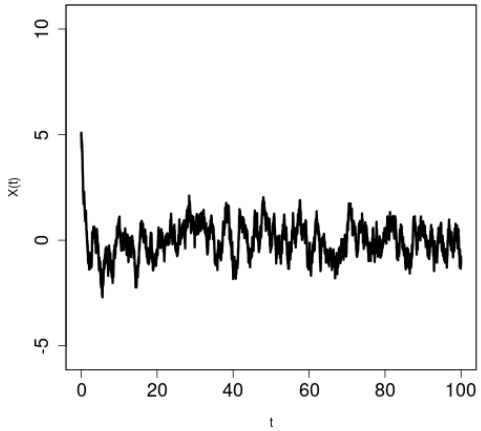
OU trajectories

$$\theta = 0,\ \alpha = 1,\ \sigma = 1$$

See code: `Script_05_1DOUyuima.R` simulations by `yuima` R package

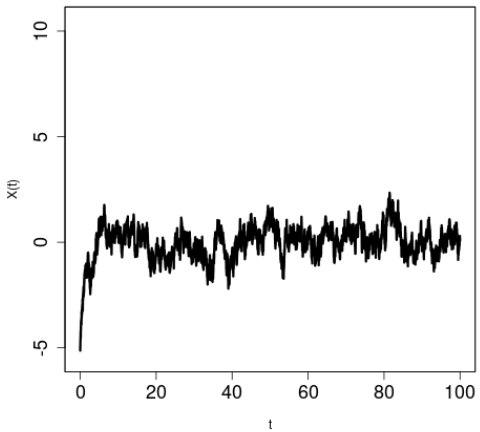
$X_0 = 0$

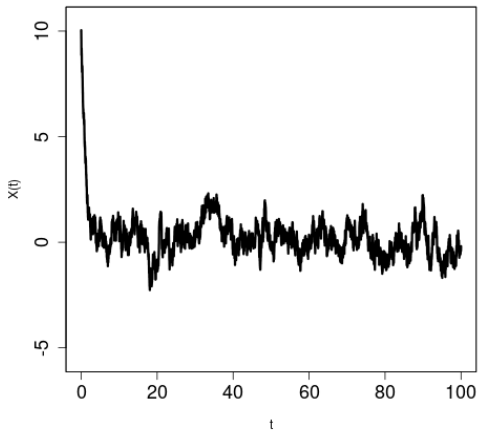




$X_0 = 5$

$X_0 = -5$





$X_0 = 10$

Multivariate SDEs

Same as ODEs:

"derivative"/Itô integral taken dimension by dimension.

Diffusion type SDE

$$\mathrm{d}\vec{X}(t) = \vec{\mu}(t, \vec{X}(t))\mathrm{d}t + \mathbf{\Sigma}(t, \vec{X}(t))\mathrm{d}\vec{B}(t) \qquad \vec{X}(0) = \vec{X}_0$$

$$\vec{X}(t) = \int_0^t \vec{\mu}(s, \vec{X}(s))\mathrm{d}s + \int_0^t \mathbf{\Sigma}(s, \vec{X}(s))\mathrm{d}\vec{B}(s)$$

Multivariate SDE examples

$$\mathrm{d}\vec{X}(t) = \mathrm{d}\vec{B}(t) \quad \vec{X}(t) = \vec{X}_0 + \vec{B}(t), \quad \vec{X}(t) \sim \mathcal{N}(\vec{X}_0, t)$$

$$\mathrm{d}\vec{X}(t) = \mathbf{\Sigma}\mathrm{d}\vec{B}(t) \quad \vec{X}(t) = \vec{X}_0 + \mathbf{\Sigma}\vec{B}(t), \quad \vec{X}(t) \sim \mathcal{N}(\vec{X}_0, \mathbf{\Sigma}\mathbf{\Sigma}^T t)$$

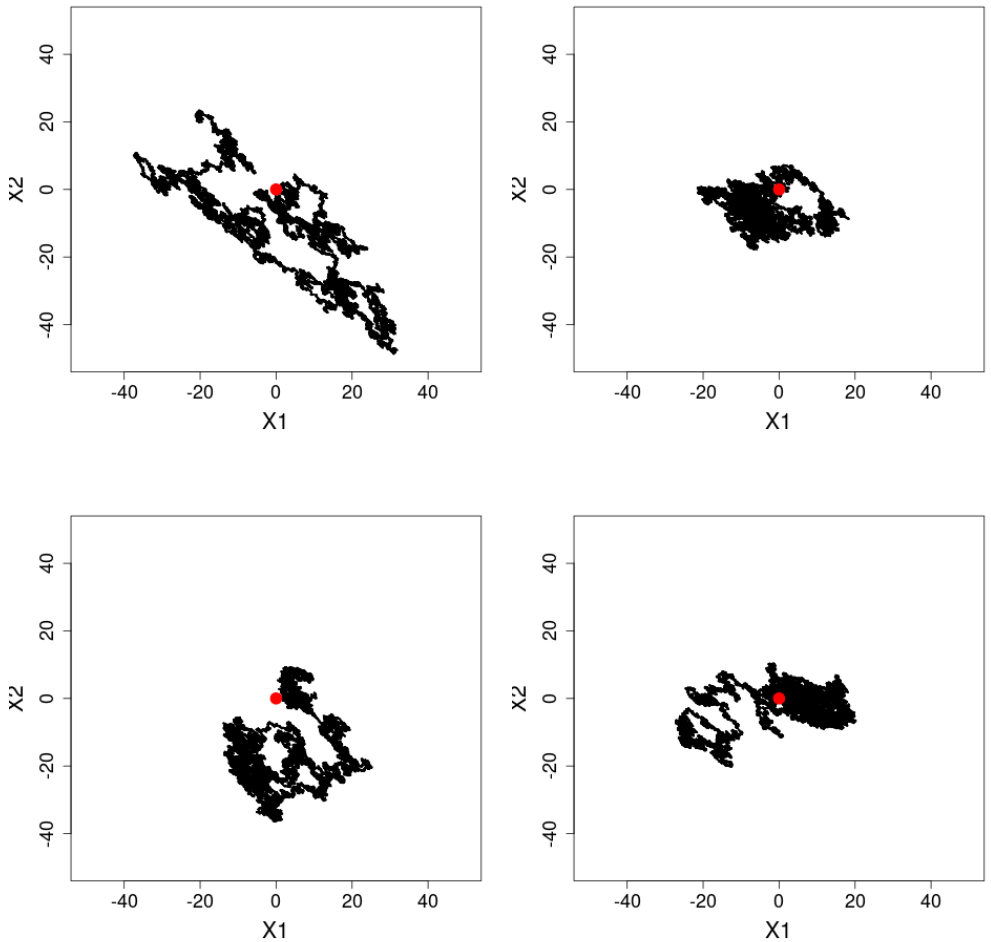
$$\mathrm{d}\vec{X}(t) = \vec{\mu}\mathrm{d}t + \mathbf{\Sigma}\mathrm{d}\vec{B}(t)$$
$$\vec{X}(t) = \vec{X}_0 + \vec{\mu}t + \mathbf{\Sigma}\vec{B}(t), \quad \vec{X}(t) \sim \mathcal{N}(\vec{X}_0 + \vec{\mu}t, \mathbf{\Sigma}\mathbf{\Sigma}^T t)$$

$$\mathrm{d}\vec{X}(t) = \vec{\mu}\mathrm{d}t + \mathbf{\Sigma}\mathrm{d}\vec{B}(t)$$
$$\vec{X}(t) = \vec{X}_0 + \mu t + \sigma\vec{B}(t), \quad \vec{X}(t) \sim \mathcal{N}(\vec{X}_0 + \vec{\mu}t, \mathbf{\Sigma}\mathbf{\Sigma}^T t)$$

2D Brownian motion

$$\vec{X}_0 = (0,0)^T,\ \boldsymbol{\Sigma}\boldsymbol{\Sigma}^T = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}\ \boldsymbol{\Sigma} = \left(\texttt{chol}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T)\right)^T$$

See code: `Script_06_2DBMyuima.R`, simulations by `yuima` R package



Multivariate linear SDEs

$$\mathrm{d}\vec{X}(t) = (\vec{c}(t) + \mathbf{D}(t)\vec{X}(t))\mathrm{d}t + \mathbf{\Sigma}(t)\mathrm{d}\vec{B}(t)$$
$$\vec{X}(t) = \mathbf{\Phi}(t)\vec{X}_0 + \mathbf{\Phi}(t)\int_0^t \mathbf{\Phi}^{-1}(s)\vec{c}(s)\mathrm{d}s + \mathbf{\Phi}(t)\int_0^t \mathbf{\Phi}^{-1}(s)\mathbf{\Sigma}(s)\mathrm{d}\vec{B}(s)$$

where $\mathbf{\Phi}(t)$ is the solution (fundamental matrix) of

$$\frac{\mathrm{d}\mathbf{\Phi}(t)}{\mathrm{d}t} = \mathbf{D}(t)\mathbf{\Phi}(t) \quad \mathbf{\Phi}(0) = \mathbf{I}$$

If $\mathbf{D}(t) \equiv \mathbf{D}$

$$\vec{X}(t) = e^{\mathbf{D}t}\vec{X}_0 + \int_0^t e^{\mathbf{D}(t-s)}\vec{c}(s)\mathrm{d}s + \int_0^t e^{\mathbf{D}(t-s)}\mathbf{\Sigma}(s)\mathrm{d}\vec{B}(s))$$

Multivariate Ornstein–Uhlenbeck process

$$\mathrm{d}\vec{X}(t) = -\mathbf{A}(\vec{X}(t) - \vec{\theta})\mathrm{d}t + \mathbf{\Sigma}\mathrm{d}\vec{B}(t), \quad \vec{X}(0) = \vec{X}_0$$

$$\mathrm{E}\left[\vec{X}(t)\right] = e^{-\mathbf{A}t}\vec{X}_0 + (1 - e^{-\mathbf{A}t})\vec{\theta}$$

$$\mathrm{Var}\left[\vec{X}(t)\right] = \int_0^t e^{\mathbf{A}s}\mathbf{\Sigma}\mathbf{\Sigma}^T e^{\mathbf{A}^T s}\mathrm{d}s$$

If all eigenvalues of $\mathbf{A}$ have positive real part then, converges (weakly) to normal with mean $\vec{\theta}$ and Var $\left[\vec{X}(\infty)\right]$

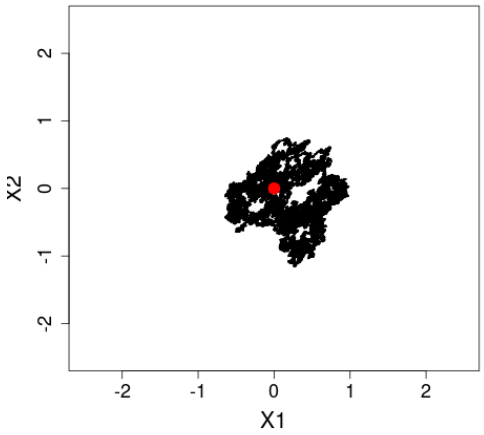
Half–lives can be calculated in eigenvector space

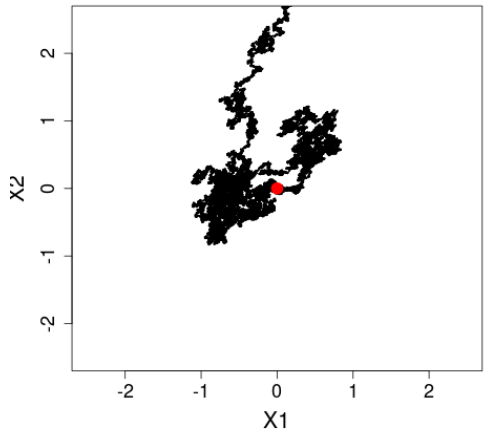
2D OU trajectories

$$\text{OU: } \vec{\theta} = (0,0)^T,\ \mathbf{A} = \begin{bmatrix} 1 & 0.5 \\ -0.5 & 1 \end{bmatrix},\ \mathbf{\Sigma\Sigma}^T = \begin{bmatrix} 0.4 & 0.15 \\ 0.15 & 0.4 \end{bmatrix}$$

See code: `Script_07_2DOUyuima.R`, simulations by `yuima` R package

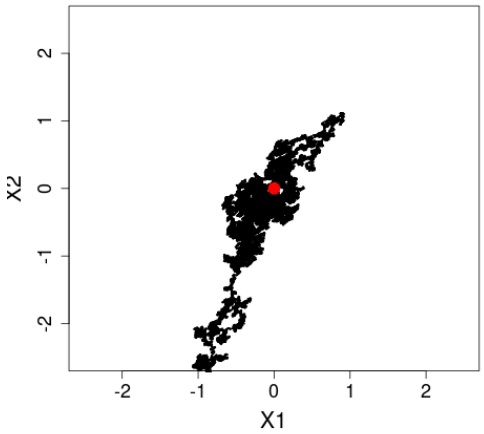
$\vec{X}_0 = (0,0)$

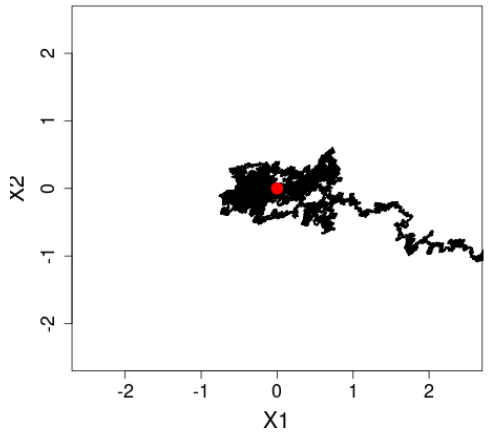




$\vec{X}_0 = (5,5)$

$\vec{X}_0 = (-5,-5)^T$





$\vec{X}_0 = (5,-5)^T$

"Random mean" process—special "singular" case

$$\begin{aligned} \mathrm{d}y(t) &= -\alpha\left(y(t) - (bx(t) + \theta)\right) + \sigma_y \mathrm{d}B_y(t) \quad (1D) \\ \mathrm{d}x(t) &= \sigma_x \mathrm{d}B_x(t) \end{aligned}$$

$$\begin{aligned} \mathrm{d}\vec{Y}(t) &= -\mathbf{D}\left(\vec{(}Y)(t) - (\mathbf{B}\vec{X}(t) + \vec{\theta})\right) + \mathbf{\Sigma}_y \mathrm{d}B_y(t) \quad (mv) \\ \mathrm{d}\vec{X}(t) &= \mathbf{\Sigma}_x \mathrm{d}B_x(t) \end{aligned}$$

No stationary distribution:

$$\mathbf{A} = \begin{bmatrix} \alpha & b \\ 0 & 0 \end{bmatrix} \quad \mathbf{A} = \begin{bmatrix} \mathbf{D} & \mathbf{B} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \text{ have } 0s \text{ as eigenvalues.}$$

Stable "residual processes":
$y(t) - (bx(t) + \theta)$, $\vec{(}Y)(t) - (\mathbf{B}\vec{X}(t) + \vec{\theta})$

Simulating SDEs

$$\mathrm{d}X(t) = \mu(t, X(t))\mathrm{d}t + \sigma(t, X(t))\mathrm{d}B(t) \quad X(0) = X_0$$

Euler–Maruyama method:

1. choose small time step Δ
2. $X(0) = X_0$ initial condition and set $t = t_0$

   If transition density known (e.g. linear—normal), $X(t)|X_0 \sim f(\cdot, t, X_0)$

   3. $X(t + \Delta) = X(t) + \epsilon$, ϵ $f(\cdot, \Delta, X_t)$

   Else

   2'. Set $B(0) = 0$

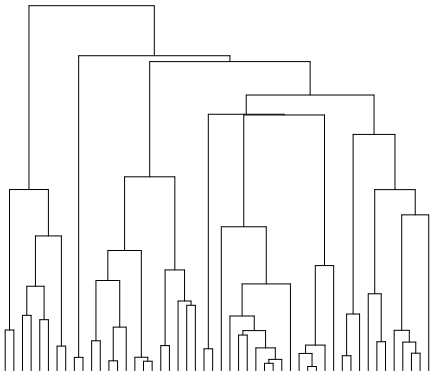
   3'. Draw $\epsilon \sim \mathcal{N}(0, \Delta)$

   3''. $X(t + \Delta) = \mu(\Delta, X(t))\Delta + \sigma(\Delta, X(t))\epsilon$

   4. $t = t + \Delta$

   5. Repeat steps 3/(3',3'') and 4 until t reaches $t_{\max}$

Phylogenetic BM and OU



$n = 50$

BM

OU

See code: `Script_08_BMPmvSLOUCH.R`
simulations by `TreeSim` and `mvSLOUCH` R package

Phylogenetic Brownian motion model

$$X(t)|X(0) \sim \mathcal{N}(X(0), \sigma^2 t)$$

Pairs of tip measurements (X_1, X_2) are NOT independent

$$\text{Cov}[X_1, X_2] = \sigma^2 t_{12} \quad t_{12} \text{ shared path length}$$

PROOF

Law of total covariance:

$$\text{Cov}[Y, Z] = \text{E}[\text{Cov}[Y, Z|U]] + \text{Cov}[\text{E}[Y|U], \text{E}[Z|U]]$$

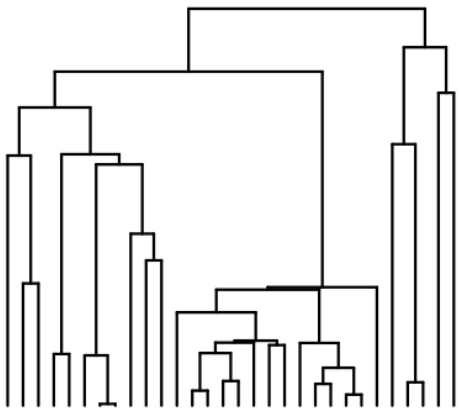
$$\begin{aligned}&\text{Cov}[X_1, X_2] = \text{E}[\text{Cov}[X_1, X_2|X_{\text{anc}_{12}}]] + \text{Cov}[\text{E}[X_1|X_{\text{anc}_{12}}], \text{E}[X_2|X_{\text{anc}_{12}}]] \\ &= 0 + \text{Cov}[X_{\text{anc}_{12}}, X_{\text{anc}_{12}}] = \text{Var}[X_{\text{anc}_{12}}] = \sigma^2 t_{12}\end{aligned}$$

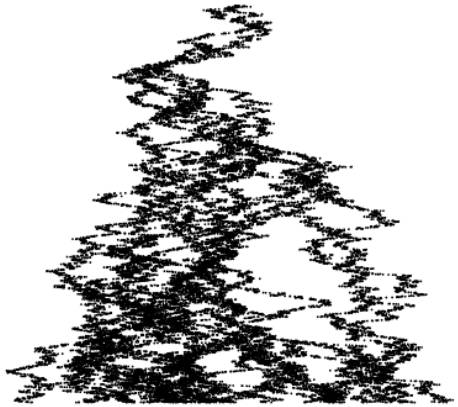
$$\vec{X} = (X_1, \ldots, X_n)^T \sim \mathcal{N}(\vec{1}X(0), \sigma^2 \mathbf{T})$$

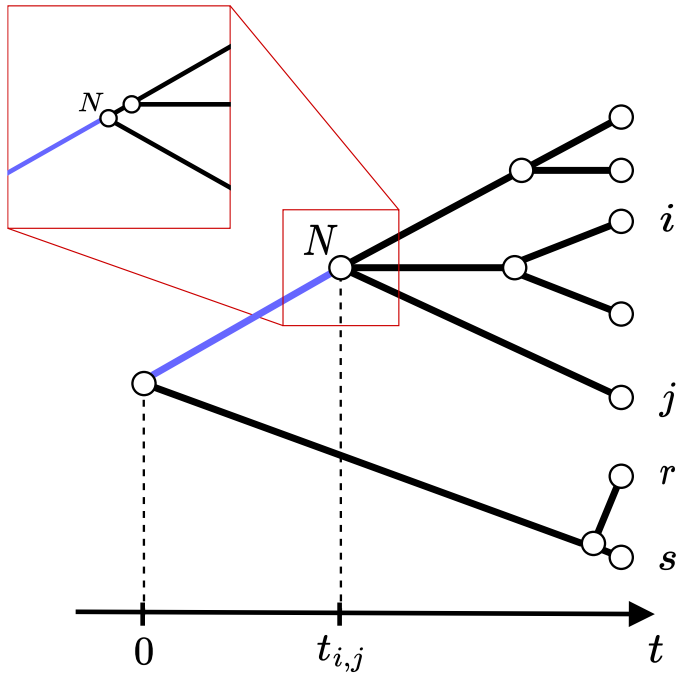
$\mathbf{T}$: matrix of shared branch lengths

Phylogenetic Brownian motion model (Felsenstein 1985)

- $\mathrm{d}\vec{X}_t = \mathbf{\Sigma}_x \mathrm{d}B_t$, $\vec{X}_0 = \vec{X}_0$; $\mathbf{\Sigma} := \mathbf{\Sigma}_x \mathbf{\Sigma}_x^T$
- $\mathrm{E}\left[\vec{X}\right](t) = \vec{X}_0, \quad \mathrm{Var}\left[X\right](t) = \mathbf{\Sigma} t \to \infty$
- $\mathbf{V} := \mathrm{Var}\left[\mathbf{X}\right] = \mathbf{T} \otimes \mathbf{\Sigma}$







See code `Script_09_simBrownianSkirt.R`.

graphic by Bayu Brahmantio

A regression perspective (BM)

$$\vec{X} = \vec{1}X_0 + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{T})$$

GLS estimates:

$$\hat{X_0} = \left(\vec{1}^T \mathbf{T}^{-1} \vec{1}\right)^{-1} \vec{1}^T \mathbf{T}^{-1} \vec{X}$$

$$\hat{\sigma^2} = (\vec{X} - \vec{1}X_0)\mathbf{T}^{-1}(\vec{X} - \vec{1}X_0)^T$$

Independent contrasts: algorithm to calculate the GLS estimates

J. Felsenstein, 1985. Phylogenies and the comparative method. Am. Nat. 125(1) 1–15

OU distribution of phylogenetic sample

$$X(t)|X(0) \sim \mathcal{N}(e^{-\alpha t}X(0) + (1 - e^{-\alpha t})\theta, \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t}))$$

Pairs of tip measurements (X_1, X_2) are NOT independent

$$\text{Cov}\,[X_1, X_2] = \frac{\sigma^2}{2\alpha}(e^{-2\alpha t_{12}} - e^{-2\alpha t}) \quad t_{12} \text{ shared path length}$$

PROOF
Law of total covariance (same approach as BM)

Easy to generalize for piecewise linear θ (notational issues)

Interacting traits — slouch

- Adaptation to a randomly fluctuating optimum
- Allometry $y = ax^b \rightarrow \log y = \log a + b \cdot \log x$
- Hansen, Pienaar, Orzack (2008)
- $$\begin{aligned} \mathrm{d}y(t) &= -\alpha \left(x(t) - \psi(t) - \sum_{l=1}^{k} b_l x_l(t) \right) \mathrm{d}t \\ &\quad +\sigma_y \mathrm{d}B_y(t) \\ \mathrm{d}x_l(t) &= \sigma_l \mathrm{d}B_l(t) \end{aligned}$$

Interacting traits — mvSLOUCH

- Coadaptation of traits
- $$\mathrm{d}\vec{X}(t) = \mathbf{\Sigma}\mathrm{d}\vec{B}_x(t)$$
- $$\mathrm{d}\vec{Y}(t) = -\mathbf{A}\left(\vec{Y}(t) - \vec{\psi}(t)\right)\mathrm{d}t + \mathbf{\Sigma}_y\mathrm{d}\vec{B}_y(t)$$
- $$\begin{aligned}\mathrm{d}\vec{Y}(t) &= -\mathbf{A}\left(\vec{Y}(t) - \vec{\psi}(t) - \mathbf{B}\vec{X}(t)\right)\mathrm{d}t \\ &\quad +\mathbf{\Sigma}_y\mathrm{d}\vec{B}_y(t) \\ \mathrm{d}\vec{X}(t) &= \mathbf{\Sigma}\mathrm{d}\vec{B}_x(t)\end{aligned}$$

Interacting trait: OUOU

$$\mathrm{d}\vec{Y}(t) = -\mathbf{A}\left(\vec{Y}(t) - \vec{\psi}(t)\right)\mathrm{d}t + \mathbf{\Sigma}_y \mathrm{d}\vec{B}_y(t)$$

$$\begin{bmatrix} \mathrm{d}y_1(t) \\ \vdots \\ \mathrm{d}y_k(t) \end{bmatrix} = -\begin{bmatrix} \alpha_1 & -\alpha_1\boldsymbol{\alpha_{12}} & \cdots & -\alpha_1\boldsymbol{\alpha_{1k}} \\ -\alpha_2\boldsymbol{\alpha_{21}} & \alpha_2 & \cdots & -\alpha_2\boldsymbol{\alpha_{2k}} \\ \vdots & \ddots & \ddots & \vdots \\ -\alpha_k\boldsymbol{\alpha_{k1}} & \cdots & -\alpha_k\boldsymbol{\alpha_{k(k-1)}} & \alpha_k \end{bmatrix} \left(\begin{bmatrix} y_1(t) \\ \vdots \\ y_k(t) \end{bmatrix} - \begin{bmatrix} \psi_1(t) \\ \vdots \\ \psi_k(t) \end{bmatrix} \right) \mathrm{d}t + \begin{bmatrix} \sigma_{11} & \sigma 12 & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2k} \\ \vdots & \ddots & \ddots & \vdots \\ \sigma_{k1} & \cdots & \sigma_{k(k-1)} & \sigma_{kk} \end{bmatrix} \begin{bmatrix} W_1(t) \\ \vdots \\ W_k(t) \end{bmatrix}$$

$$\begin{aligned} \mathrm{d}y_1(t) &= -\alpha_1\left(y_1(t) - (\boldsymbol{\alpha_{12}} y_2(t) + \ldots + \boldsymbol{\alpha_{1k}} y_k(t) + \psi_1(t))\right)\mathrm{d}t \\ &\quad + \sigma_{11}\mathrm{d}W_1 + \ldots \sigma_{1k}\mathrm{d}W_k \\ &\vdots \\ \mathrm{d}y_k(t) &= -\alpha_k\left(y_k(t) - (\boldsymbol{\alpha_{k1}} y_1(t) + \ldots + \boldsymbol{\alpha_{kk-1}} y_{k-1}(t) + \psi_k(t))\right)\mathrm{d}t \\ &\quad + \sigma_{k1}\mathrm{d}W_1 + \ldots \sigma_{kk}\mathrm{d}W_k \end{aligned}$$

Interacting traits — mvSLOUCH

- Eigenvalues of $\mathbf{A}$ — λ
- all $\lambda > 0$ — adaptation
- evolutionary regression $\mathrm{E}\left[\vec{Y}|\vec{X}\right](t)$
- optimal regression $\mathbf{B}$
- $\mathrm{E}\left[\vec{Y}|\vec{X}\right](t) \rightarrow \mathbf{B}$ in OUBM mode

Regression description

- *Evolutionary* regression : $\mathrm{E}\left[\vec{Y}|\vec{X}\right](t)$
- *Limiting* regression : $\lim_{t\to\infty} \mathrm{E}\left[\vec{Y}|\vec{X}\right](t)$,
- *Optimal* regression mvOUBM :

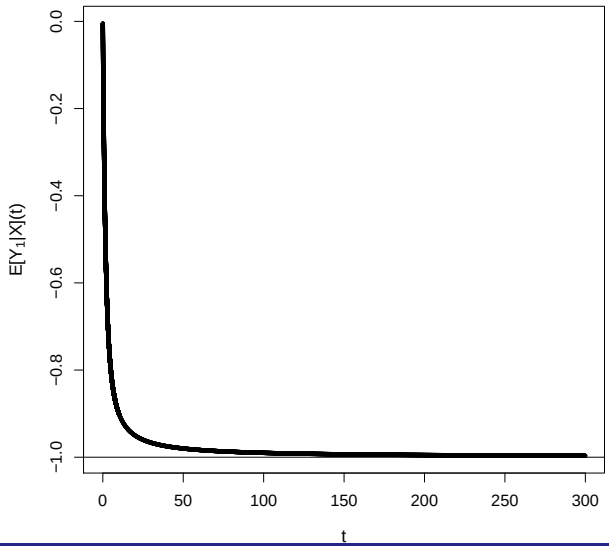
$$\begin{aligned} \mathrm{d}y(t) &= -\alpha(y(t) - (\psi(t) + bx(t)))\mathrm{d}t + \sigma_{yy}\mathrm{d}B_y(t) \\ \mathrm{d}x(t) &= \sigma_{xx}\mathrm{d}B_x(t) \end{aligned}$$

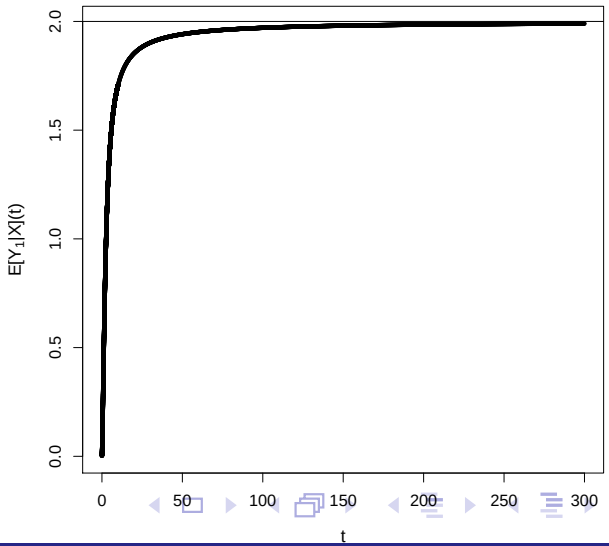
$$\begin{aligned} \mathrm{d}\vec{Y}(t) &= -\mathrm{A}(\vec{Y}(t) - (\vec{\psi}(t) + \mathbf{B}\vec{X}(t)))\mathrm{d}t + \mathbf{\Sigma}_{yy}\mathrm{d}\vec{B}_y(t) \\ \mathrm{d}\vec{X}(t) &= \Sigma_{xx}\mathrm{d}\vec{B}_x(t) \end{aligned}$$

OUBM *Evolutionary* regression

- $\mathrm{E}\left[y|x\right](t) = (1 - \frac{1-e^{-\alpha t}}{\alpha t})bx(t) + \text{intercept}$
- $(1 - \frac{1-e^{-\alpha t}}{\alpha t})b \rightarrow b$
- $\mathrm{E}\left[\vec{Y}|\vec{X}\right](t) = (\mathbf{I} - t^{-1}(\mathbf{I} - e^{-t\mathbf{A}})\mathbf{A}^{-1})\mathbf{B}\vec{X}(t) + \text{intercept}$
- $(\mathbf{I} - t^{-1}(\mathbf{I} - e^{-t\mathbf{A}})\mathbf{A}^{-1})\mathbf{B} \rightarrow \mathbf{B}$

$$\alpha = 1,\ b = -1 \qquad \mathbf{A} = \begin{bmatrix} 1 & -0.5 \\ -0.25 & 2 \end{bmatrix},\ \mathbf{B} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$



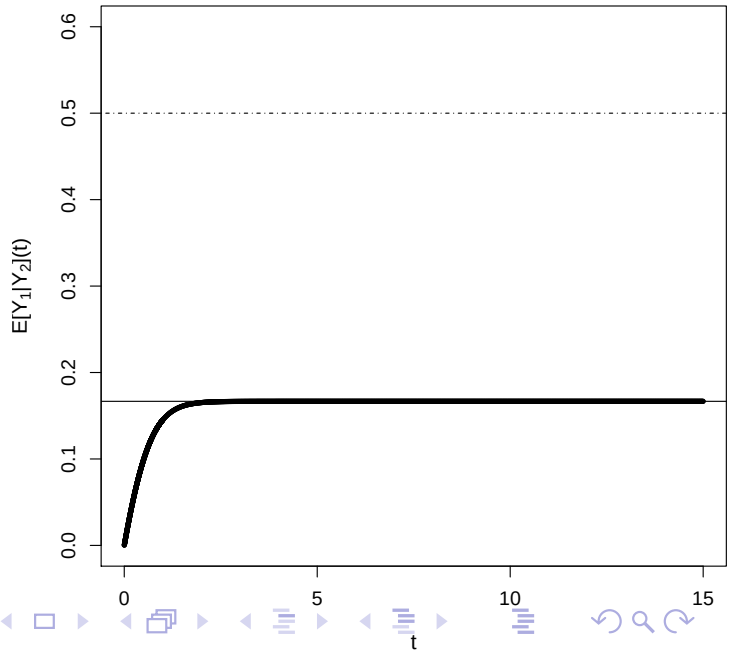


Regression — convergence to optimal always?

$$\mathrm{d}Y(t) = -\alpha_y(Y(t) - (\psi_y + b_1 X(t)))\mathrm{d}t + \sigma_y \mathrm{d}W_y(t) + \sigma_{yx}\mathrm{d}W_x(t)$$
$$\mathrm{d}X(t) = -\alpha_x(X(t) - \psi_x)\mathrm{d}t + \sigma_x \mathrm{d}W_x(t)$$

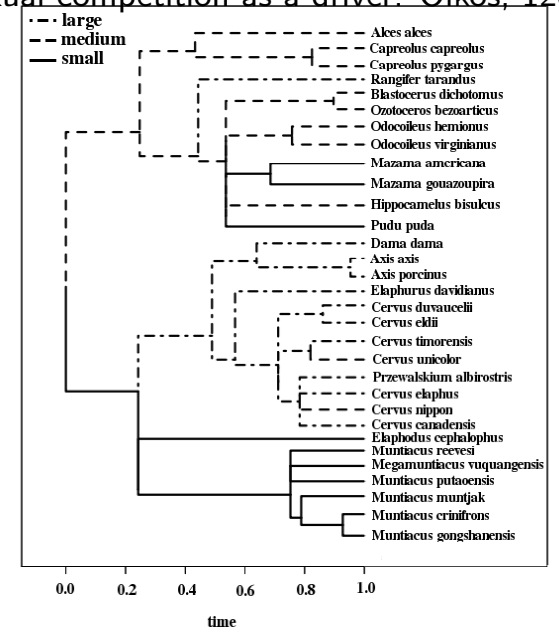
$$\lim_{t\to\infty} \text{coefficient}(\mathsf{E}[Y(t)|X(t)]) \quad = \quad \frac{\alpha_y}{\alpha_x+\alpha_y} b_1 + 2\frac{\alpha_x}{\alpha_x+\alpha_y}\frac{\sigma_{yx}}{\sigma_x}$$

- $\alpha_y = 1$
- $\alpha_x = 2$
- $b_1 = 0.5$
- $\sigma_{yx} = 0$
- $\frac{\alpha_y}{\alpha_x+\alpha_y} b_1 = \frac{1}{6}$



Plard et. al. (2011) — Cervidae

Fig 2A in K. Bartoszek, J. Pienaar, P. Mostad, S. Andersson, and T. F. Hansen. A phylogenetic comparative method for studying multivariate adaptation. J. Theor. Biol., 314:204-215, 2012. phylogeny after F. Plard, C. Bonenfant, and J.M. Gaillard, J.M. Revisiting the allometry of antlers among deer species: male-male sexual competition as a driver. Oikos, 120:601-606, 2011.



USDA photo by Scott Bauer
`https://www.ars.usda.gov/oc/images/photos/may01/k5437-3/`

public domain `https://commons.wikimedia.org/wiki/File:White-tailed_deer.jpg`

- antler length, male and female body size
- breeding group size, mating strategy

Why multivariate models are needed

Plard et. al. (2011) OLS explains the data best,
i.e., *no* phylogenetic effects, but only compared to BM.
Multivariate phylogenetic OU not available at that time.

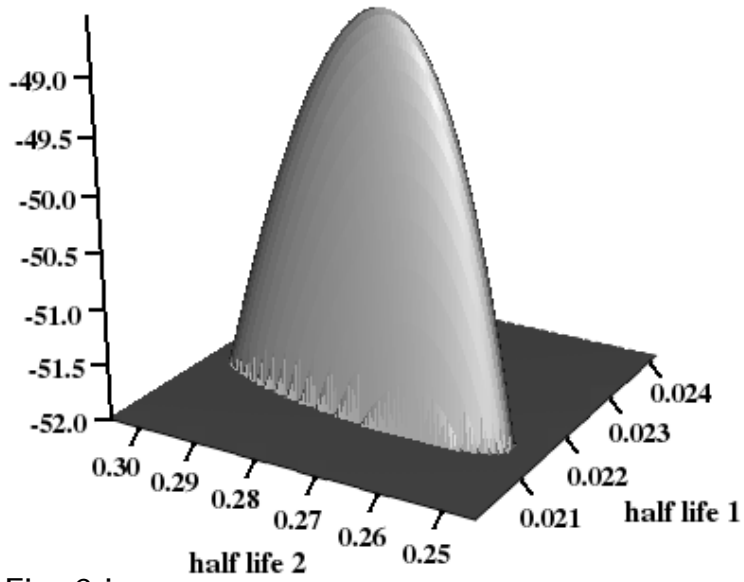


Fig. 3 in

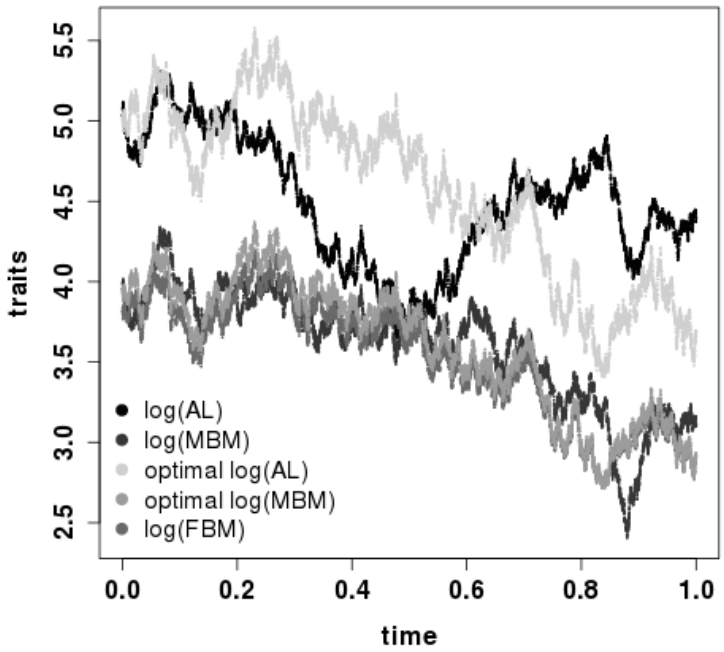
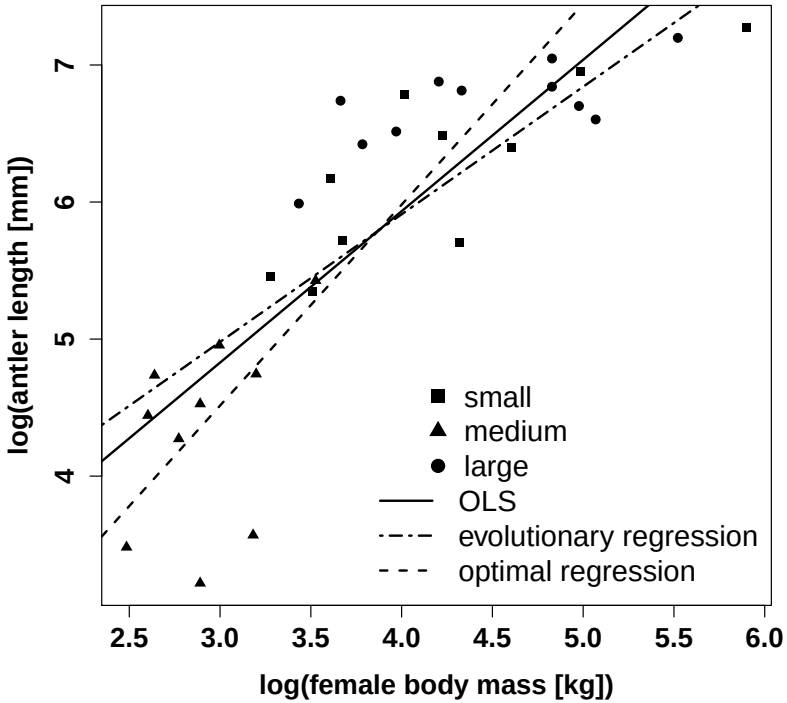


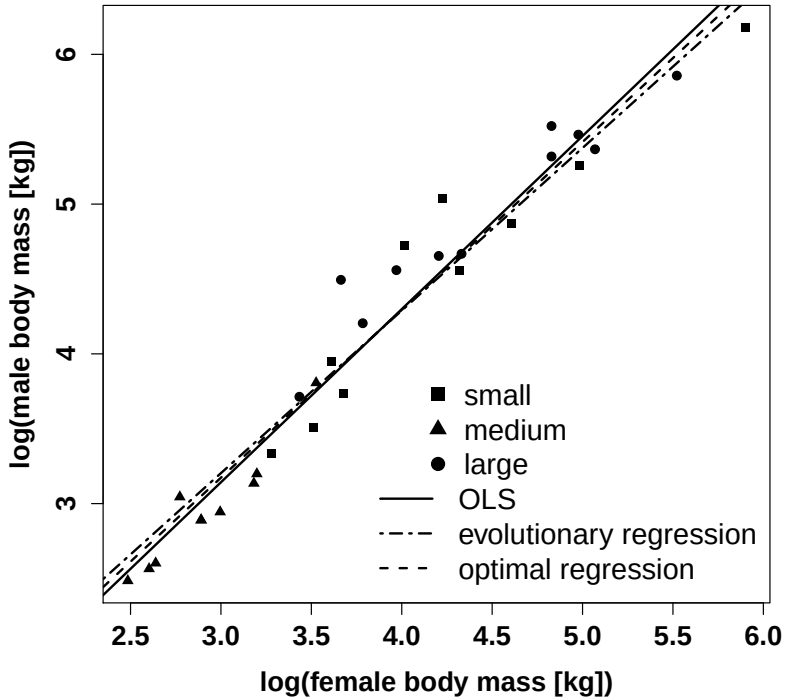
Fig. 1A in

K. Bartoszek, J. Pienaar, P. Mostad, S. Andersson, and T. F. Hansen. A phylogenetic comparative method for studying multivariate adaptation. J. Theor. Biol., 314:204-215, 2012.

mvSLOUCH analysis

Figs. 2C, D in K. Bartoszek, J. Pienaar, P. Mostad, S. Andersson, and T. F. Hansen. A phylogenetic comparative method for studying multivariate adaptation. J. Theor. Biol., 314:204-215, 2012.





- antler length = female body size + breeding group size
- male body size = female body size + breeding group size
- female body size : BM

Maximum likelihood estimation (phylogenetic BM)

$$\vec{\mu} = \vec{1} \otimes \vec{X}_0 \quad \mathbf{V} := \mathrm{Var}\left[\mathbf{X}\right] = \mathbf{T} \otimes \mathbf{\Sigma}$$

$$\log(\mathcal{L}(\vec{\mathbf{X}}|\vec{\mu}, \mathbf{V})) = -\frac{K}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\left(\vec{\mathbf{X}} - \vec{\mu}\right)^T \mathbf{V}^{-1}\left(\vec{\mathbf{X}} - \vec{\mu}\right)$$

$$\begin{aligned}
\hat{\vec{X}}(0) &= \left(\left(\vec{1}_n^T \mathbf{T}^{-1} \vec{1}_n\right)^{-1} \vec{1}_n^T \mathbf{T}^{-1} \otimes \mathbf{I}_k\right) \vec{X}, \\
\hat{\mathbf{\Sigma}} &= n^{-1}\left(\left(\mathbf{X} - \vec{1}_n \vec{X}^T(0)\right)^T \mathbf{T}^{-1}\left(\mathbf{X} - \vec{1}_n \vec{X}^T(0)\right)\right)
\end{aligned}$$

The log–likelihood

$$\mathcal{L}(\vec{\mathbf{X}}|\mathbb{T}, \text{model}) = -\frac{1}{2}\left(K\log(2\pi) + \log|\mathbf{V}| + \left(\left(\vec{\mathbf{X}} - \vec{\mu}\right)\mathbf{V}^{-1}\left(\vec{\mathbf{X}} - \vec{\mu}\right)\right)\right)$$

N tips

K all observed traits in all tip species

$\mathbf{V} \in \mathbb{R}^{K\times K}$

- Construction is a bottleneck: $\mathbf{V} \in O(K^2)$
- Calculating $\mathbf{V}^{-1} \in \Omega(K^{2.373})$
- Calculating $|\mathbf{V}| \in \Omega(K^{2.373})$

a few traits

$<$ 1000 species

PCMBase, mvSLOUCH: Integrating out internal node values

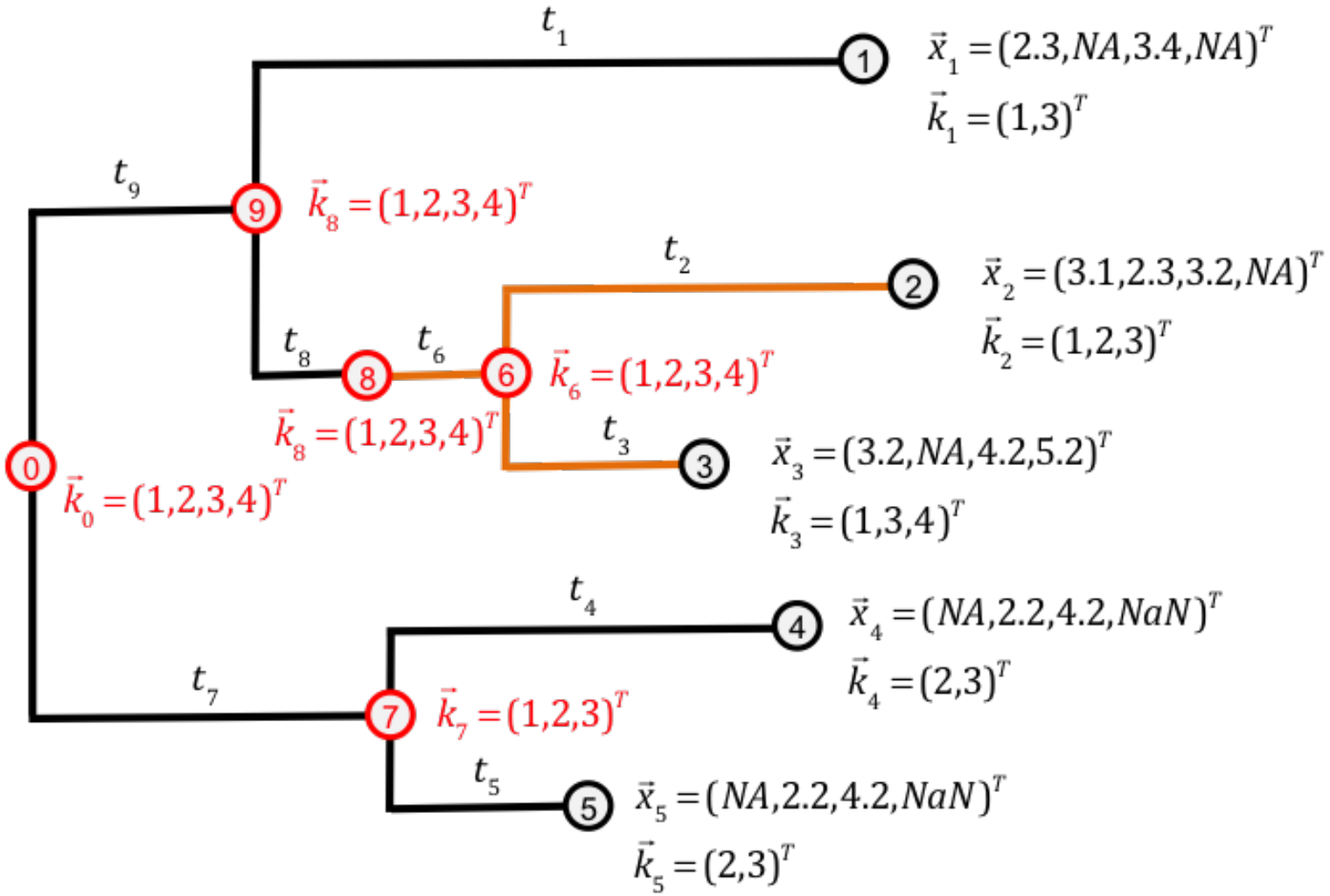


Fig. 1 in V. Mitov, K. Bartoszek, G. Asimomitis, and T. Stadler. Fast likelihood calculation for multivariate Gaussian phylogenetic models with shifts. Theor. Pop. Biol. 131:66-78, 2020.

$\mathcal{G}_{LInv}$ family of models

1. after branching the traits evolve independently in all descending lineages,

2. $\vec{x}(t)$, conditional on $\vec{x}(s)$, $s < t$, is Gaussian with

   2.1 $\mathrm{E}\left[\vec{x}(t)|\vec{x}(s)\right] = \vec{\omega} + \mathbf{\Phi}\vec{x}(s)$

   2.2 $\mathrm{Var}\left[\vec{x}(t)|\vec{x}(s)\right] = \mathbf{V}_{s,t}$

$\vec{\omega}$, $\mathbf{\Phi}$, $\mathbf{V}_{s,t}$ may depend on s and t but **NOT** on $\vec{x}(\cdot)$

Pruning algorithm

Calculating $\mathbf{V}(\mathit{Tree}, \Theta)$ is time–consuming

Faster algorithm:

$X(t) \in \mathbb{R}^k$, state space

i inner node,

$L^i(s)$: likelihood of subtree rooted at i with $X(\text{at node } i) = s$

$$L^i(s) = \prod_{o:\text{daughter of } i} \int_{\mathbb{R}^k} f_{s\to x}(t_o) L^o(x) \mathrm{d}x$$

and for a leaf j

$$L^j(s) = \begin{cases} 1 & \text{if } s \text{ is leaf } j\text{'s state} \\ 0 & \text{otherwise} \end{cases}$$

Provided one can calculate $\int_{\mathbb{R}^k} f_{s\to x}(t_o) L^o(x) \mathrm{d}x$.

Theorem: density representation

$$pdf(\vec{x}_i|\vec{x}_j) = \exp\left[\vec{x}_i^T \mathbf{A}_i \vec{x}_i + \vec{x}_i^T \vec{b}_i + \vec{x}_j^T \mathbf{C}_i \vec{x}_j + \vec{x}_j^T \vec{d}_i + \vec{x}_j^T \mathbf{E}_i \vec{x}_i + f_i\right],$$

$\mathbf{A}_i$, $\vec{b}_i$, $\mathbf{C}_i$, $\vec{d}_i$, $\mathbf{E}_i$, f_i do not depend on $\vec{x}_j$ and

$$\begin{aligned}
\mathbf{A}_i &= -\tfrac{1}{2}\mathbf{V}_i^{-1} \in \mathbb{R}^{k_i \times k_i} \\
\vec{b}_i &= \mathbf{V}_i^{-1}\vec{\omega}_i \in \mathbb{R}^{k_i} \\
\mathbf{C}_i &= -\tfrac{1}{2}\mathbf{\Phi}_i^T \mathbf{V}_i^{-1}\mathbf{\Phi}_i \in \mathbb{R}^{k_j \times k_j} \\
\vec{d}_i &= -\mathbf{\Phi}_i^T \mathbf{V}_i^{-1}\vec{\omega}_i \in \mathbb{R}^{k_j} \\
\mathbf{E}_i &= \mathbf{\Phi}_i^T \mathbf{V}_i^{-1} \in \mathbb{R}^{k_j \times k_i} \\
f_i &= -\tfrac{1}{2}\vec{\omega}_i^T \mathbf{V}_i^{-1}\vec{\omega}_i - \tfrac{k_i}{2}\log(2\pi) - \tfrac{1}{2}\log|\mathbf{V}_i| \in \mathbb{R}.
\end{aligned}$$

Theorem: recursion for the likelihood

$$pdf(\mathbf{X}_j|\vec{x}_j, \mathbb{T}, \Theta) = \exp\left(\vec{x}_j^T \mathbf{L}_j \vec{x}_j + \vec{x}_j^T \vec{m}_j + r_j\right)$$

$$\begin{aligned}
\mathbf{L}_j^{tips} &= \sum \mathbf{C}_i, \\
\vec{m}_j^{tips} &= \sum \vec{d}_i + \mathbf{E}_i \vec{x}_i, \\
r_j^{tips} &= \sum \vec{x}_i^T \mathbf{A}_i \vec{x}_i + \vec{x}_i^T \vec{b}_i + f_i, \\
\mathbf{L}_j^{non-tips} &= \sum \left(\mathbf{C}_i - (1/4)\mathbf{E}_i \left(\mathbf{A}_i + \mathbf{L}_i\right)^{-1} \mathbf{E}_i^T\right), \\
\vec{m}_j^{non-tips} &= \sum \left(\vec{d}_i - (1/2)\mathbf{E}_i \left(\mathbf{A}_i + \mathbf{L}_i\right)^{-1} \left(\vec{b}_i + \vec{m}_i\right)\right), \\
r_j^{non-tips} &= \sum \left(f_i + r_i + \tfrac{k_i}{2}\log(2\pi) - \tfrac{1}{2}\log(|(-2)\left(\mathbf{A}_i + \mathbf{L}_i\right)|)\right. \\
&\quad \left. -(1/4)\left(\vec{b}_i + \vec{m}_i\right)^T \left(\mathbf{A}_i + \mathbf{L}_i\right)^{-1} \left(\vec{b}_i + \vec{m}_i\right)\right)
\end{aligned}$$

Phylogenetic regression: calculating matrix–vector products

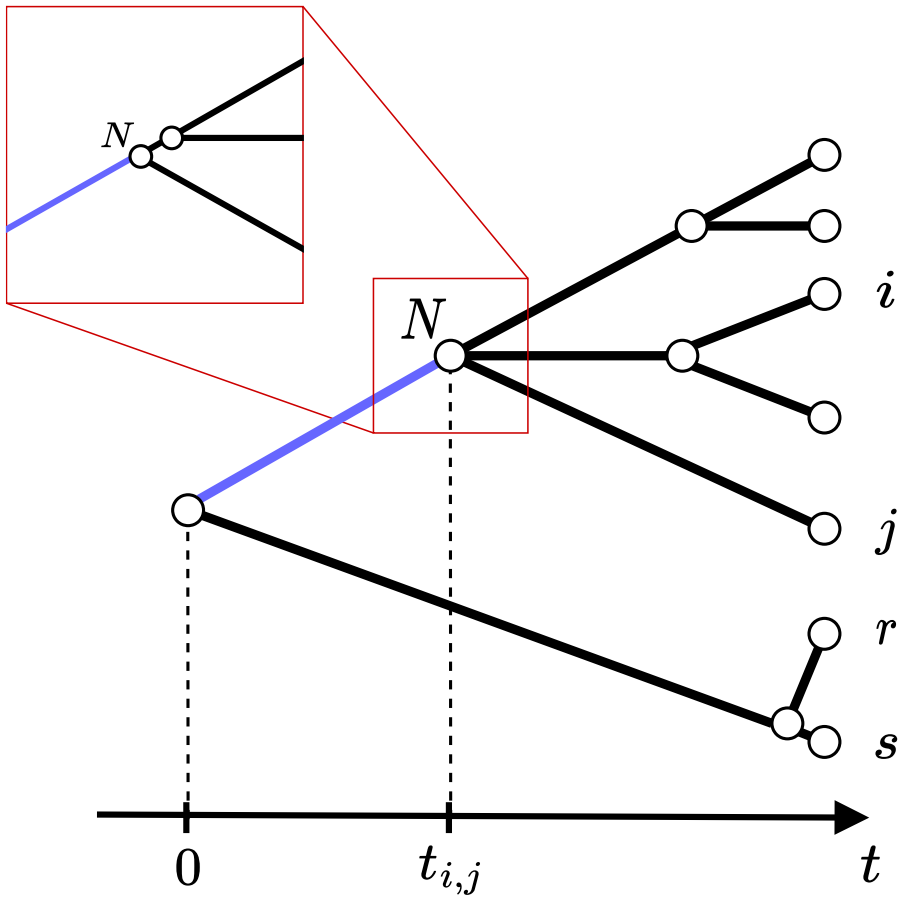
$$\log(\mathcal{L}(\vec{\mathbf{X}}|\vec{\mu},\mathbf{V})) = -\frac{K}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\left(\vec{\mathbf{X}}-\vec{\mu}\right)^T\mathbf{V}^{-1}\left(\vec{\mathbf{X}}-\vec{\mu}\right)$$

Setting $\vec{\mu} = \vec{\mathbf{X}} = \vec{0}$ will give $\log|\mathbf{V}|$ and then notice

$$\vec{u}_1^T\mathbf{V}^{-1}\vec{u}_2 = \frac{1}{2}\left(\vec{u}_1^T\mathbf{V}^{-1}\vec{u}_1 + \vec{u}_2^T\mathbf{V}^{-1}\vec{u}_2 - (\vec{u}_1-\vec{u}_2)^T\mathbf{V}^{-1}(\vec{u}_1-\vec{u}_2)\right)$$

$\vec{\mu} = \vec{0}$, $\vec{\mathbf{X}} = \vec{u}_1,\ \vec{u}_2,\ \vec{u}_1 - \vec{u}_2$

Short branch problem



graphic by Bayu Brahmantio

mvSLOUCH: returns

$$\log(\mathcal{L}(\vec{\mathbf{X}}|\vec{\mu}, \mathbf{V})) = -\infty$$

$$\mathrm{Var}\left[X\right](t) = \mathbf{\Sigma} t$$

$$(\mathrm{Var}\left[X\right](t))^{-1} = \mathbf{\Sigma}^{-1} t^{-1}$$

Short branch problem: solutions

Non–tip branches

1. extend branch to some numerically sensible minimal length
2. **mvSLOUCH, PCMBase, PCMFit**: collapse branch into a polytomy
3. rescale whole tree

Tip branches

1. extend branch to some numerically sensible minimal length
2. keep one branch with average of traits from the two tips
3. keep one branch with the left traits from the two tips
4. keep one branch with the right traits from the two tips

Missing values (normal model)

Multitrait setting
Often measurements on some traits in some species are missing
Removing the whole species wastes the observed traits.

Notation: $\vec{X} = (\vec{X}_1^T, \ldots, \vec{X}_n^T)^T$

Normal distribution framework:
—Remove entries of $\mathrm{E}\left[\vec{X}\right] = \vec{\mu}(Tree, \Theta)$ corresponding to missing values
—Remove rows and columns of $\mathrm{Var}\left[\vec{X}\right] = \mathbf{V}(Tree, \Theta)$ corresponding to missing values

Measurement error (normal model)

The tip measurements are usually averages for a species

Hence natural intra–species variability present

Calculate for each species its variance

Add the variances to $\mathbf{V}(\mathit{Tree}, \Theta)$'s (block) diagonal

Estimating ancestral characters (BM, OU)

Difficult to estimate ancestral state if no fossil data

BM: $\hat{X}_0$ is a linear combination of tips
Var $\left[\hat{X}_0\right] \nrightarrow 0$, as BM's variance $\rightarrow \infty$

OU: estimation is difficult or impossible
wide confidence intervals due to $e^{-\alpha t} X_0$ in mean
if θ constant, then one cannot distinguish

Model selection

Models: Competing phylogenies, evolutionary models e.t.c.

Nested models: R^2, likelihood ratio

$$R^2 = RSS_{\text{model}}/RSS_{\text{under null model}}$$

$$2(\mathcal{L}_1 - \mathcal{L}_0) \sim \chi^2_{K_1 - K_0}$$

Information criteria

The lower the better

Akaike Information Criterion (AIC), corrected (AIC_c)for sample size (n)

$$AIC = -2\mathcal{L} + 2K \qquad AIC_c = AIC + \frac{2K(K+1)}{n-K-1}$$

Bayesian/Schwarz Information Criterion (BIC):

$$BIC = -2\mathcal{L} + K \log n$$

Bayes factor (difficult to compute, BIC approximation for logs)

$$B_{ij} = \frac{P(Data|M_i)}{P(Data|M_j)}$$

We have R competing models, $i = 1, \ldots R$.

$$\Delta AIC_{c_i} = AIC_{c_i} - AIC_{c_{\min}} \qquad \Delta AIC_i = AIC_i - AIC_{\min}$$

$0 < \Delta < 4$ or $0 < \Delta < 7$: plausible

$\Delta > 14$: implausible

$\Delta AIC_i \leq 2$	substantial support (evidence)
$4 \leq \Delta AIC_i \leq 7$	considerably less support
$\Delta AIC_i > 10$	essentially no support

Fig. 2 of K. P. Burnham, D. R. Anderson, K. P. Huyvaert, 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behav. Ecol. Sociobiol. 65:23–35 doi:10.1007/s00265-010-1029-6

p. 271 of K. P. Burnham, D. R. Anderson, 2004. Multimodel Inference: Understanding AIC and BIC in Model Selection. Socio. Meth. Res. 33(2):261–304 doi:10.1007/10.1177/0049124104268644

AIC: contains large scaling constants, e.g., $AIC_1 = 300000$, $AIC_2 = 300020$
only ΔAIC_i: interpretable as strength of evidence (Burnham & Anderson 2004)

Model averaging

We have a large number of competing models

$$\Delta\{A/B\}IC_i = \{A/B\}IC_i - \min\{A/B\}IC$$

$$\text{weight of model } i \; w_i = \frac{\exp(-1/(2\Delta_i)}{\sum_r \exp(-1/(2\Delta_r)}$$

Model–averaged estimate of numerical parameter

$$\hat{\theta} = \frac{\sum_i w_i I_\theta(M_i)\hat{\theta}_i}{\sum_i w_i I_\theta(M_i)}$$

where

$$I_\theta(M_i) = \begin{cases} 1 & \text{if parameter } \theta \text{ belongs to model } M_i \\ 0 & \text{otherwise} \end{cases}$$

Summarizing a model

Likelihood in itself does not say much
Plot support surface in some parameters' directions

Report parametric bootstrap confidence intervals
Calculate regression confidence intervals for Θ_{lin} conditional on $\Theta_{non-lin}$

$R^2 = RSS_{model}/RSS_{null\ \ model}$, but what is null model?

If one can partition traits: predictors/responses:
conditional density of responses on predictors

Half–lives (in eigenspace)

Model selection

Competing models: e.g. BM versus OU

BM: neutral evolution
OU: adaptive evoltion

OU:

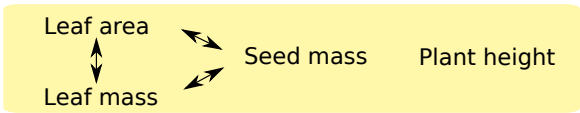
different levels of θ (environments)
assumptions on $\mathbf{A}$

Look at AIC, AIC_c, BIC

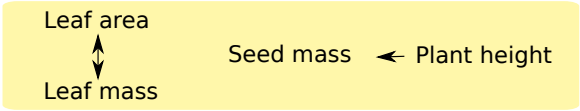
Functional traits in angiosperms

Model 1

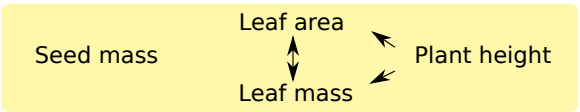
Model 2



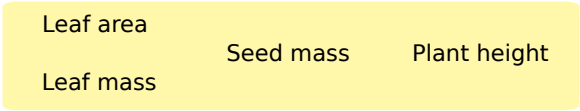
Model 3



Model 4

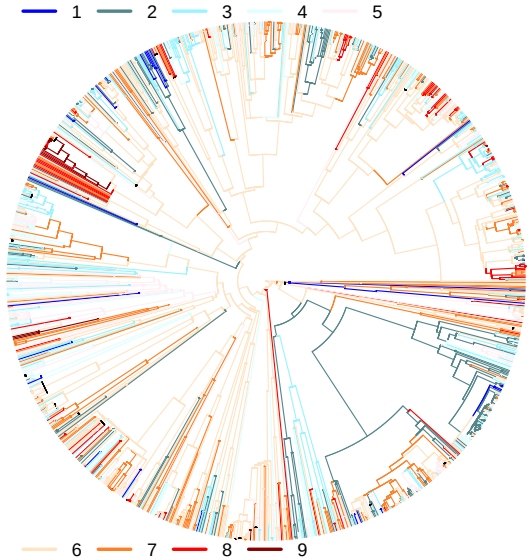


Model 5



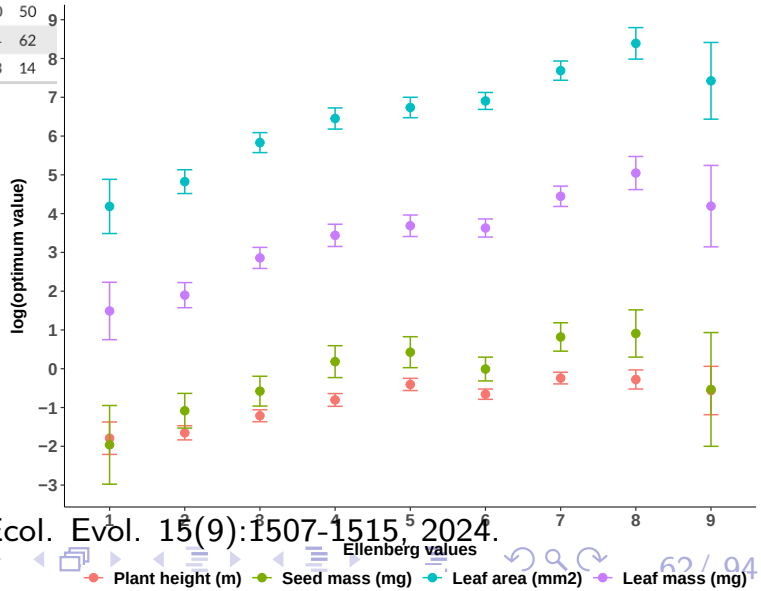
Model 6

Model		A				Model		A			
		PH	SM	LA	LM			PH	SM	LA	LM
1	PH	+	0	0	0	2	PH	+	0	0	0
	SM	?	+	?	?		SM	0	+	?	?
	LA	?	?	+	?		LA	0	?	+	?
	LM	?	?	?	+		LM	0	?	?	+
3	PH	+	0	0	0	4	PH	+	0	0	0
	SM	?	+	0	0		SM	0	+	0	0
	LA	0	0	+	?		LA	?	0	+	?
	LM	0	0	?	+		LM	?	0	?	+
5	PH	+	0	0	0	6	PH	?	?	?	?
	SM	0	+	0	0		SM	?	?	?	?
	LA	0	0	+	0		LA	?	?	?	?
	LM	0	0	0	+		LM	?	?	?	?



	Σ_{yy} diagonal				Σ_{yy} upper-triangular			
Model	LogLik	AIC_c	R^2	df	LogLik	AIC_c	R^2	df
1	−10,598.94	21,305.04	0.052	53	−10,607.68	21,334.80	0.05	59
2	−9968.793	20,038.615	0.076	50	−9747.051	19,607.392	0.054	56
3	−9996.868	20,088.645	0.077	47	−9024.085	18,155.326	0.035	53
4	−10,373.82	20,844.58	0.082	48	−9918.165	19,945.529	0.105	54
5	−9995.955	20,080.709	0.088	44	−9533.795	19,168.618	0.040	50
6	−8156.907	16,427.104	0.088	56	**−8127.739**	**16,381.058**	0.094	62
BM	–	–	–	–	−10,912.55	21,853.19	0.003	14

	Directions (eigenvectors)			
Model 6	$\vec{e}_1$	$\vec{e}_2$	$\vec{e}_3$	$\vec{e}_4$
PH	0.007	0.015	– 0.686	0.184
SM	– 0.046	0.351	– 0.353	0.863
LA	0.122	– 0.668	– 0.367	0.351
LM	0.991	– 0.656	– 0.52	0.313
Half-life	0.0002%	0.033%	0.407%	2.028%
(CI no seed)	(0.107%, 72.392%)	(7.416%, 80.331%)	(26.214%, 192.568%)	(66.732%, 1630.682%)
(CI seed)	(0.00015%, 0.00033%)	(0.0262%, 0.065%)	(0.333%, 1.111%)	(0.489%, 2.613%)



Figs. 2,3, S5 and Tabs. 1,2 in K. Bartoszek et. al. Fast **mvSLOUCH**: Multivariate Ornstein–Uhlenbeck–based models of trait evolution on large phylogenies. Meth. Ecol. Evol. 15(9):1507-1515, 2024.

Some conclusions from simulations

- Usually good identifiability BM v OUOU v OUBM
- Difficult to identify type of **A** (e.g., model 4 v 6)
- Diagonal **A** constraint has mixed effects
- Measurement error makes model identifiability worse
- $\mathbf{\Sigma}_{yy}$, $\vec{\psi}$: consitently estimated
- **A**, half–lives, correlations, regressions: unclear

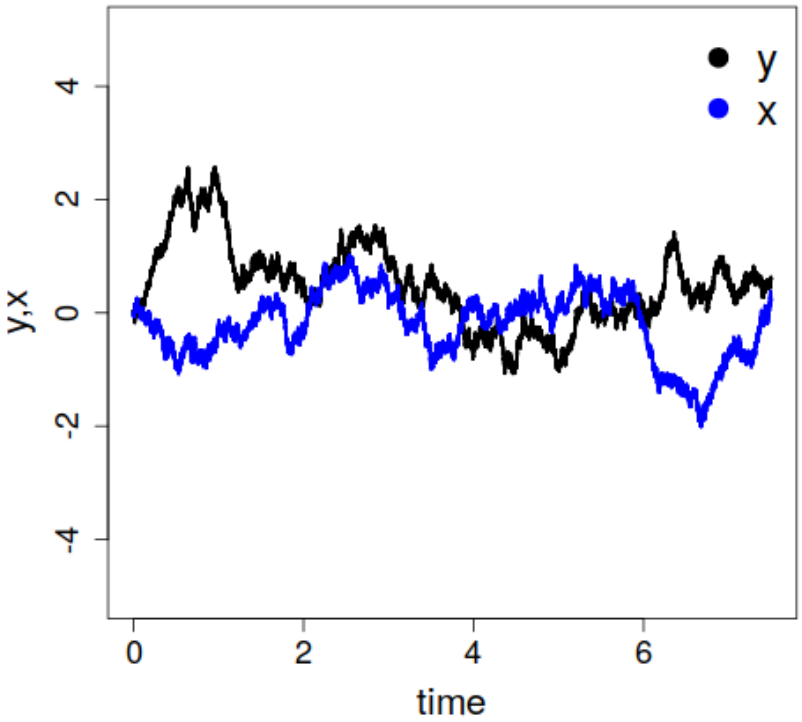
Tree height: 1

Tab. A3 in K. Bartoszek, J. Fuentes-González, V. Mitov, J. Pienaar, M. Piwczyński, R. Puchałka, K. Spalik, and K. L. Voje. Analytical advances alleviate model misspecification in non-Brownian multivariate comparative methods. Evolution 78(3):389-400, 2024.

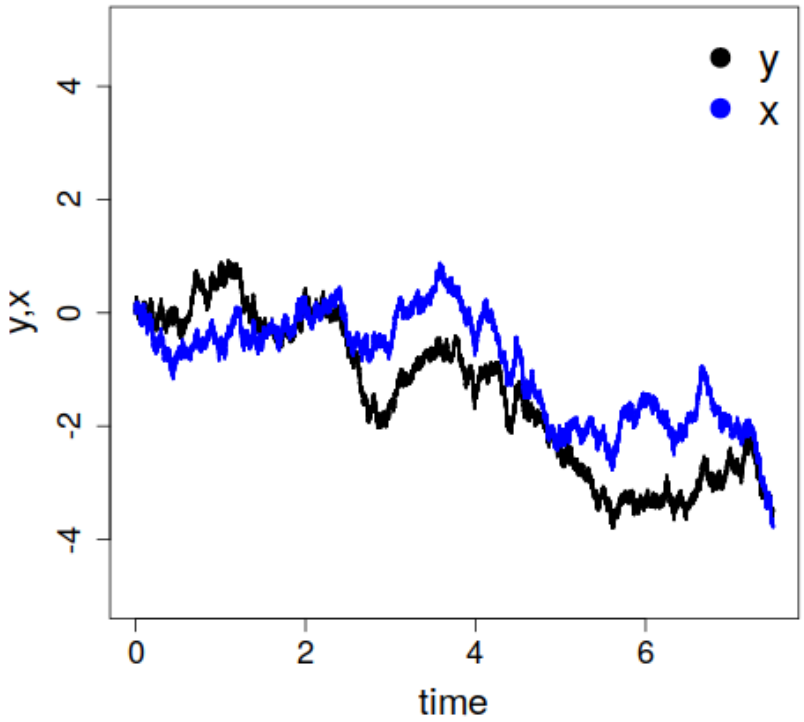
n $k=4$	32	64	128	256	512	1024	2048
required runs	100	100	102	101	104	108	123
# OUBM1ML	53	37	36	37	28	28	33
average $t_{\frac{1}{2}}$	83.426	2.16	158.944	10.546	64.467	29.872	20.737
variance $t_{\frac{1}{2}}$	346334.532	16.109	781932.156	734.702	58073.927	14208.496	923.464
median $t_{\frac{1}{2}}$	0.527	0.867	1.287	1.71	3.208	4.698	7.652
lower quartile $t_{\frac{1}{2}}$	0.388	0.527	0.89	1.38	2.093	3.376	4.546
upper quartile $t_{\frac{1}{2}}$	0.894	1.258	2.044	3.583	6.892	7.642	19.362
$\%t_{\frac{1}{2}} > 1$	18.868%	37.838%	66.667%	89.189%	96.429%	100%	100%
$\%t_{\frac{1}{2}} > 2$	13.208%	16.216%	25%	45.946%	78.571%	92.857%	100%
$\%t_{\frac{1}{2}} > 3$	13.208%	16.216%	19.444%	32.432%	50%	82.143%	96.97%

Half–life "time to lose half of ancestral state", $\mathrm{E}\left[X_{OU}^{1D}(t)\right] = e^{-\alpha t}X_0 + (1-e^{-\alpha t})\theta$
$\exp(-\alpha t_{0.5}) = 0.5 \text{ then } t_{0.5} = \ln 2/\alpha$
Simulated under 4–dimensional standard BM, but $\mathrm{AIC_c}$ preferred OUBM
A: constant diagonal
Increasing dimension improves the identifiablity of BM

$$\mathrm{d}y(t) = -0.01(y(t) - x(t))\mathrm{d}t + \mathrm{d}B_y(t)$$
$$\mathrm{d}x(t) = \mathrm{d}B_x(t)$$



$$\mathrm{d}y(t) = \mathrm{d}B_y(t) + 0.75\mathrm{d}B_x(t)$$
$$\mathrm{d}x(t) = 0.66\mathrm{d}B_x(t)$$



See code `Script_10_simOUBMsmallBM.R`.

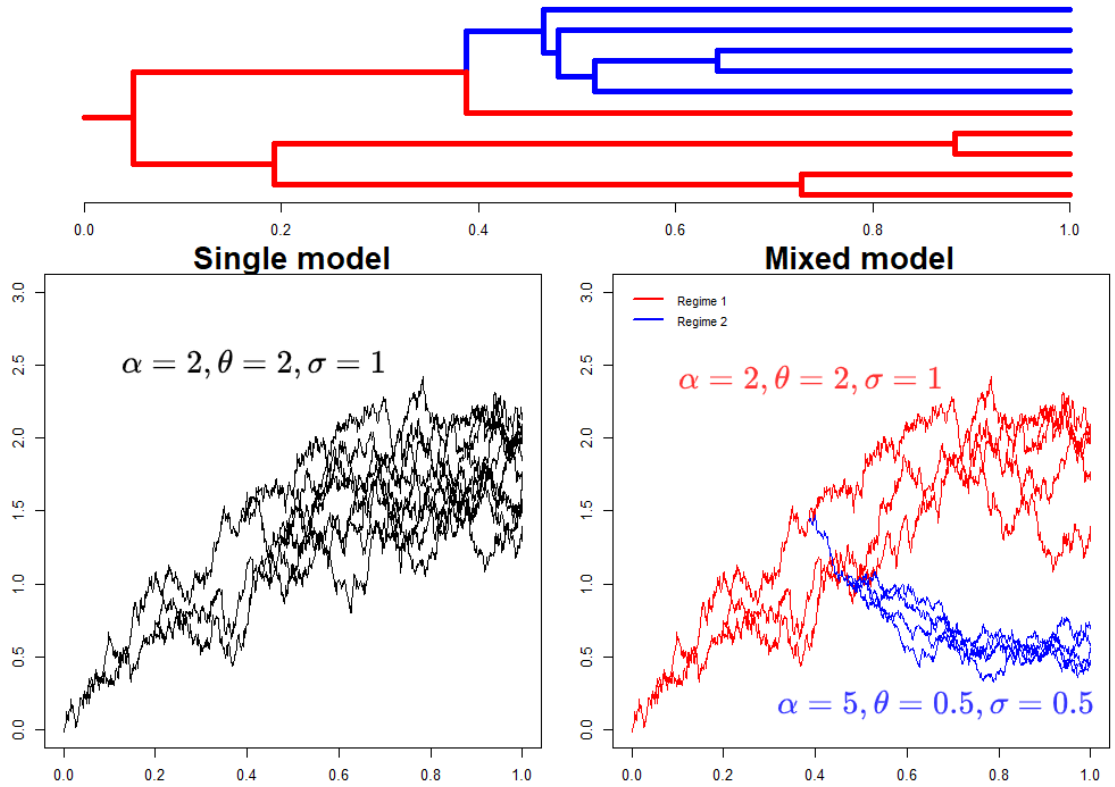
Multiple regimes

- **ouch**, **slouch**, **mvSLOUCH**, **mvMORPH**, **SURFACE**: optima vary over the tree
- **OUwie**: OU parameters can vary (with restrictions) over the tree
- **PCMBase**, **PCMFit**, **glinvci**, **bgphy**: whole models can vary over the tree
- **SURFACE**, **PCMFit**: identify regime placements

MGPM

PCMFit: Mixed Gaussian phylogenetic models (MGPM)

$$\mathrm{d}X(t) = -\alpha(X(t) - \theta)\mathrm{d}t + \sigma \mathrm{d}W(t)$$



graphic by Bayu Brahmantio

Bayesian MGPM

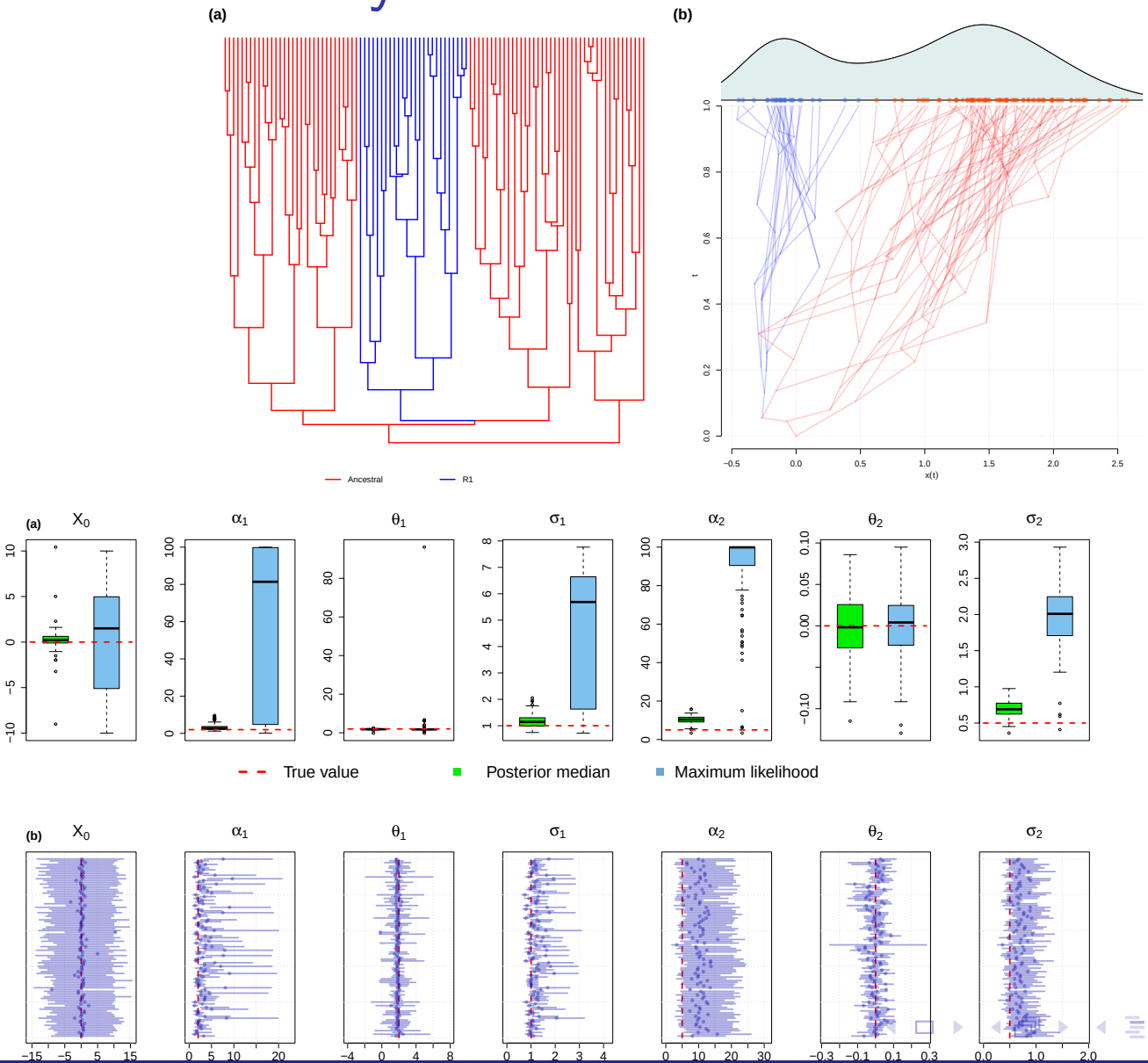
`github.com/bayubeta/bgphy`

likelihood: PCMBase

Bayesian parameter inference:
Laplace's approximation
self–normalized importance sampling
Widely Available Information Criterion

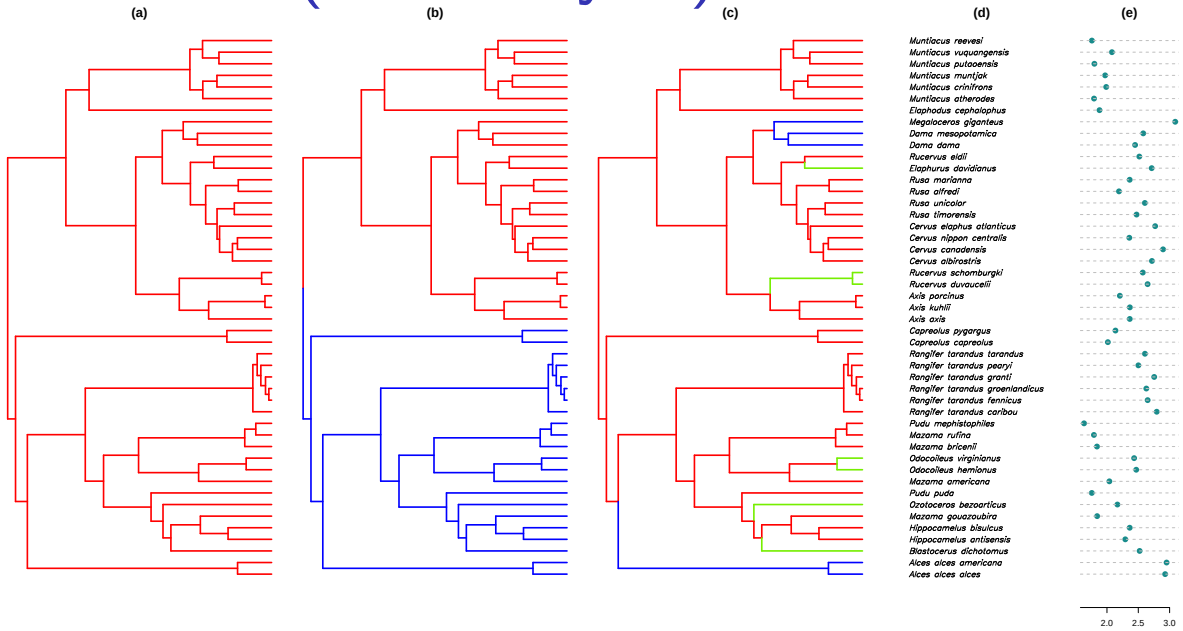
Simulation study

Figs. 2 and 5 in B. Brahmantio, K. Bartoszek, E. Yapar. Bayesian inference of mixed Gaussian phylogenetic models.arXiv:2410.11548, 2024.



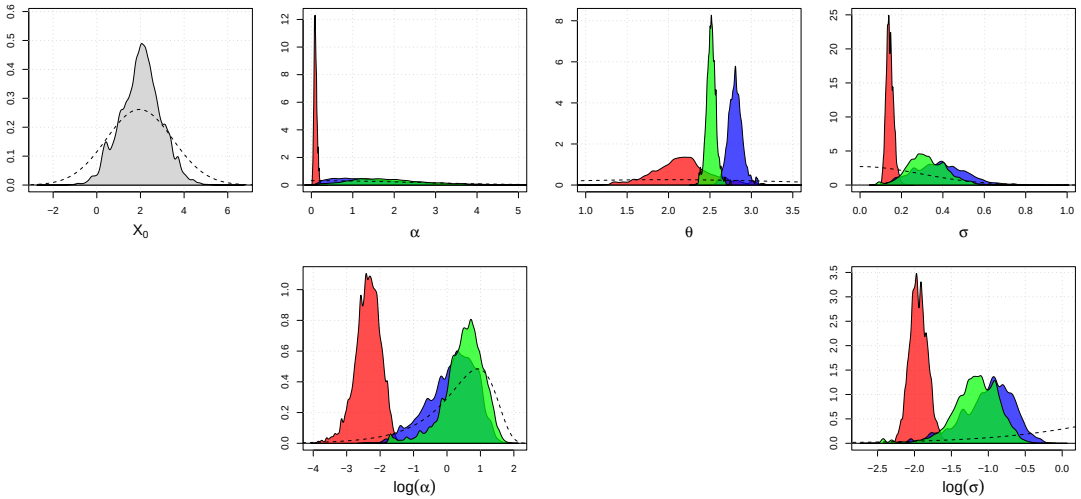
Antlers (1D analysis)

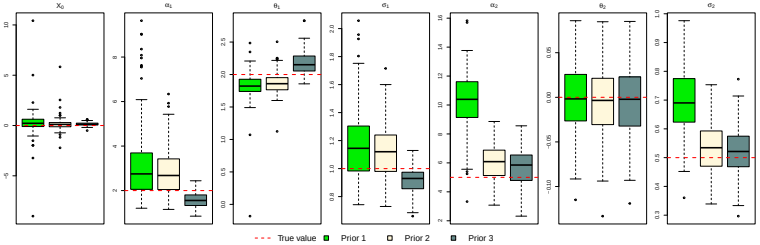
Figs. 4,8,9 in B. Brahmantio, K. Bartoszek, E. Yapar. Bayesian inference of mixed Gaussian phylogenetic models.arXiv:2410.11548, 2024.



b) blue: old world,
red: new world

c) red: main beamed,
blue: palmated,
green: bifurcated





R packages (a primer)

mvSLOUCH
PCMBase; PCMBaseCpp
PCMFit https://github.com/venelin/PCMFit
GLSME
pcmabc
glinvci
github.com/bayubeta/bgphy

slouch ouch
caper ape
geiger OUwie
mvMORPH bayou
surface rphylopars
PhylogeneticEM diversitree
RPANDA

Code examples

mvSLOUCH: vignette in package

PCMFit: tutorials `https://venelin.github.io/PCMFit/`

bgphy: README `https://github.com/bayubeta/bgphy`

PCMBase: tutorials `https://venelin.github.io/PCMBase/`

glinvci: README `https://cran.r-project.org/web/packages/glinvci/readme/README.html`

References: inference packages

mvSLOUCH:
K.B., J.P., P.M., S.A., T.F.H. **A phylogenetic comparative method for studying multivariate adaptation.** *J. Theor. Biol.*314:204-215,2012.
K.B., J.T.G., J.F.-G., V.M., J.P., M.P., R.P., K.S., K.L.V. **Fast mvSLOUCH: Multivariate Ornstein–Uhlenbeck–based models of trait evolution on large phylogenies.** *Meth. Ecol. Evol.* 15(9):1507-1515, 2024.
PCMFit:
V.M., K.B., T.S. **Automatic generation of evolutionary hypotheses using mixed Gaussian phylogenetic models.** *PNAS* 116:16921-16926, 2019.
bgphy:
B.B., K.B., E.Y. **Bayesian inference of mixed Gaussian phylogenetic models.** *arXiv*:2410.11548, 2024.

References: properties

Model **identifiability**, parameter **estimability**:
K.B., J.F.-G., V.M., J.P., M.P., R.P., K.S., K.L.V. **Model Selection Performance in Phylogenetic Comparative Methods under multivariate Ornstein–Uhlenbeck Models of Trait Evolution.** *Syst. Biol.* 72(2):275-293, 2023.
K.B., J.F.-G., V.M., J.P., M.P., R.P., K.S., K.L.V. **Analytical advances alleviate model misspecification in non–Brownian multivariate comparative methods.** *Evolution* 78(3):389-400, 2024.

Very recent **review** of PCMs:
J.P., K.B., B.B., J.L.F, J.F.-G., B.T.K., T.F.H., W.H.C.K, K.L.V. **Phylogenetic Comparative Methods for Studying Adaptation: The Adaptation-Inertia Framework.** *J. Evol. Biol.* voaf113, in press.

References: computational engines

PCMBase:
V.M., K.B., G. Asimomitis, T. Stadler **Fast likelihood calculation for multivariate Gaussian phylogenetic models with shifts.** *Theor. Pop. Biol.* 131:66-78, 2020.

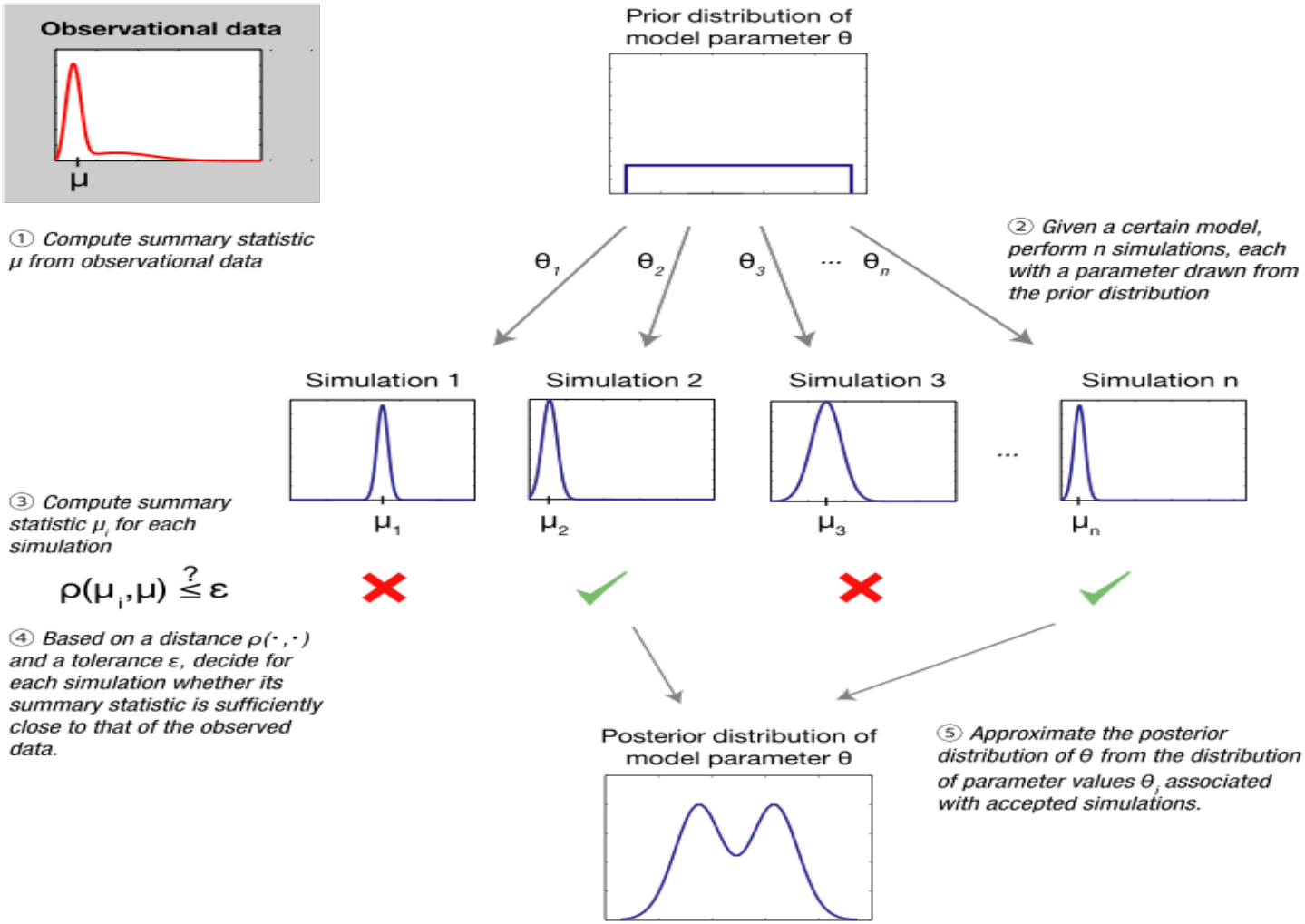
glinvci:
W.H.C.K. **Exact Expressions for the Log-likelihood's Hessian in Multivariate Continuous-Time Continuous-Trait Gaussian Evolution along a Phylogeny.** *arXiv*:2405.07394, 2024.

pcmabc: trait dependent speciation

cran.r-project.org/web/packages/pcmabc/index.html

K. Bartoszek and P. Liò. Modelling trait–dependent speciation with Approximate Bayesian Computation. Acta Physica Polonica B Proceedings Supplement. 12(1):25-47, 2019.

Stochastic model for trait X_t
Birth and death rates functions of X_t



M. Sunnøaker, A. G. Busetto, E. Numminen, J. Corander, M. Foll and C. Dessimoz (2013), Approximate Bayesian Computation. PLoS Comput Biol 9(1): e1002803.
https://en.wikipedia.org/wiki/Approximate_Bayesian_computation CC BY 2.5

`pcmabc`: trait dependent speciation

Stochastic model for trait X_t
Birth and death rates functions of X_t

$$\text{Birth and death rates} \quad \lambda(t, X_t),\ \mu(t, X_t)$$

Estimation: ABC
Distance between phylogenies:
TV distance between exps
`treedist::wRF.dist()`
Distance between trait samples:

$$\sqrt{((\mathrm{Var}_1 - \mathrm{Var}_2)/(\mathrm{Var}_1 + \mathrm{Var}_2))^2 + ((\mathrm{E}_1 - \mathrm{E}_2)/(\mathrm{E}_1 + \mathrm{E}_2))^2}$$

User interface

PCM_ABC()

↑ simulate_phylproc()

↑ simulate_phenotype_**on**_tree()

↑ simulate_sde_**on**_branch()

User interface: PCM_ABC()

PCM_ABC(...,par0,phenotype.**model**,fbirth,fdeath=NULL,...)

- par0=**list**(phenotype.**model**.params=**list**(...), birth.params=**list**(),death.params=**list**())
- phenotype.**model**(**time**,params,X0,**step**): can be a wrapper around simulate_sde_**on**_branch(); evolution independent on each branch,
  returns matrix; rows: dimensions, first row time; columns: samples at time point
- fbirth(**c**(**t**,X),birth.params), fdeath(**c**(**t**,X),birth.params) return a number, **t** is relative to branch start

User interface: simulate_phylproc()

simulate_phylproc (..., simul.params, X0, fbirth , fdeath=NULL, fbirth .params=NULL, fdeath.params=NULL, fsimulphenotype="sde.yuima", ...)
returns: phylogeny + traits' full trajectory/on tips only

- simul.params: input list for fsimulphenotype
- X0: traits' value at root
- fbirth (**c**(**t**,X), birth .params), fdeath(**c**(**t**,X), birth .params)
- fbirth .params=**list** (...) , fdeath.params=**list** (...) : named
- fsimulphenotype="sde.yuima": fsimulphenotype <−simulate_sde_**on**_branch

User interface: simulate_phenotype_**on**_tree()

simulate_phenotype_**on**_tree(phyltree, fsimulphenotype, simul.params, X0, **step**)

```
if (fsimulphenotype == "sde.yuima")) {
    fsimulphenotype <- simulate_sde_on_branch
}
```

User interface: simulate_sde_**on**_branch()

simulate_sde_**on**_branch(branch.**length**, **model**.yuima, X0, **step**)

- **model**.yuima: output of yuima::setModel(), yuima::setYuima(**model**=**model**.yuima,...)
- X0: yuima::simulate (..., xinit =X0,...
- **step**: ns<−branch.**length**/**step** yuima::setSampling(Terminal=branch.**length**,n=numsteps)

Interface with YUIMA

```
simulate_OU_sde<-function(t,params,X0,step){
    A <- c(paste("(-",params$a11,")
    *(x1-(",params$psi1,"))",sep=""))

    S <- matrix( params$s11, 1,1)
    yuima.1d <- yuima::setModel(drift = A,
    diffusion = S, state.variable=c("x1"),
    solve.variable=c("x1") )
    simulate_sde_on_branch(t,yuima.1d,X0,step)
}
fbirth_rate_constrained<-function(x,params,...){
    x<-x[2]; params$b/(1+exp(-x))}
PCM_ABC(...,phenotype.model=simulate_OU_sde,
fbirth=fbirth_rate_constrained,...)}
```

Interface with YUIMA

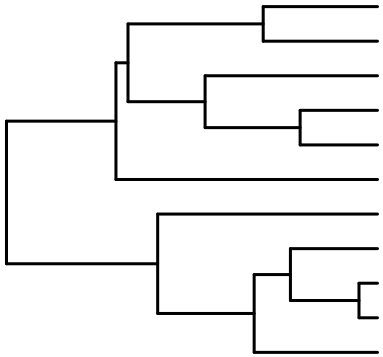
$$\mathrm{d}x_1(t) = -a_{11}(x_1(t) - \psi_1)\mathrm{d}t + s_{11}\mathrm{d}B(t)$$

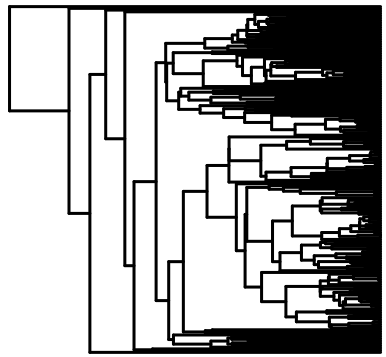
```
    A <- c(paste("(-",params$a11,")
␣␣␣␣*(x1-(",params$psi1,"))",sep=""))

    S <- matrix( params$s11, 1,1)

    model.yuima <- yuima::setModel(drift = A,
    diffusion = S, state.variable=c("x1"),
    solve.variable=c("x1") )
```

Trait dependent speciation (tree height: 1)

$$\mathrm{d}X_t = -3(X_t - \theta)\mathrm{d}t + 0.25\mathrm{d}W_t, \text{ birth rate: } 10 \cdot |\sin(X_t)|$$

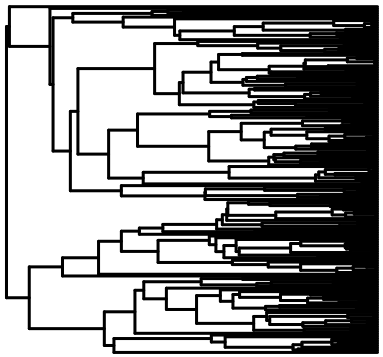
$X_0 = \frac{1}{4}\ \theta = \frac{1}{4}$

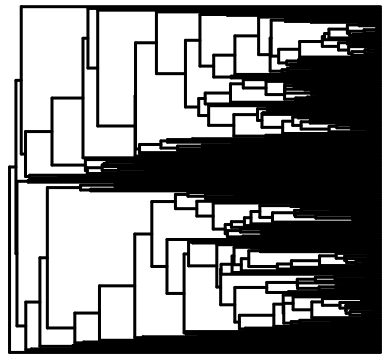




$X_0 = \frac{1}{4}\ \theta = \frac{5}{4}$

$X_0 = \frac{5}{4}\ \theta = \frac{1}{4}$





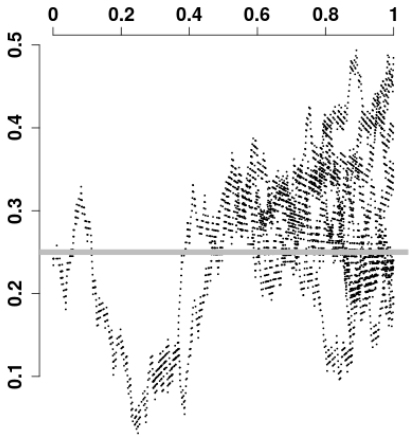
$X_0 = \frac{5}{4}\ \theta = \frac{5}{4}$

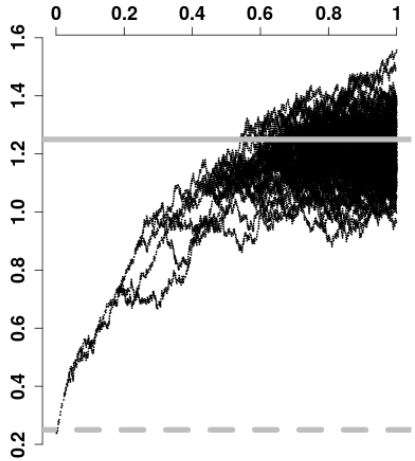
Fig. 2 in K. Bartoszek and P. Liò. Modelling trait-dependent speciation with Approximate Bayesian Computation. Acta Phys. Pol. B Proc. Supp. 12(1):25-47, 2019.

Trait dependent speciation (tree height: 1)

$$\mathrm{d}X_t = -3(X_t - \theta)\mathrm{d}t + 0.25\mathrm{d}W_t, \text{ birth rate: } 10 \cdot |\sin(X_t)|$$

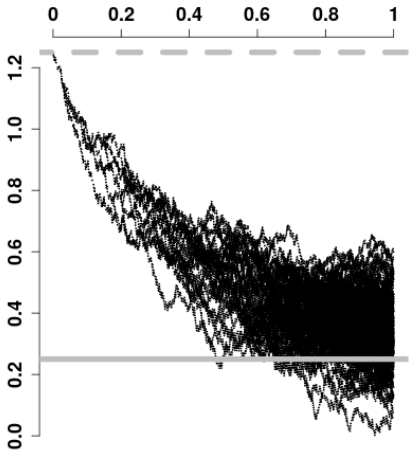
$X_0 = \frac{1}{4}$ $\theta = \frac{1}{4}$

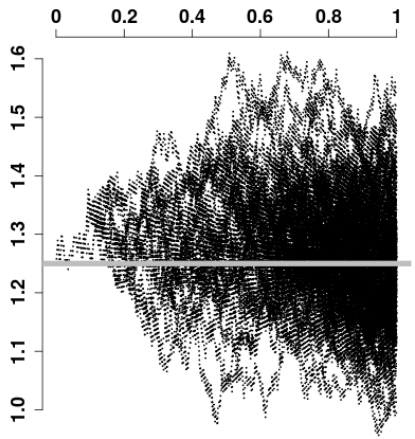




$X_0 = \frac{1}{4}$ $\theta = \frac{5}{4}$

$X_0 = \frac{5}{4}$ $\theta = \frac{1}{4}$





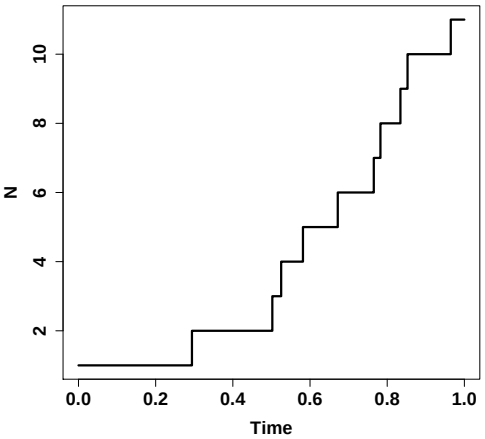
$X_0 = \frac{5}{4}$ $\theta = \frac{5}{4}$

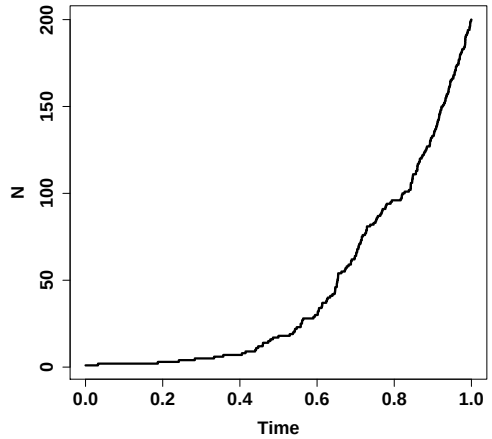
Fig. 2 in K. Bartoszek and P. Liò. Modelling trait-dependent speciation with Approximate Bayesian Computation. Acta Phys. Pol. B Proc. Supp. 12(1):25-47, 2019.

Trait dependent speciation (tree height: 1)

$$\mathrm{d}X_t = -3(X_t - \theta)\mathrm{d}t + 0.25\mathrm{d}W_t, \text{ birth rate: } 10 \cdot |\sin(X_t)|$$

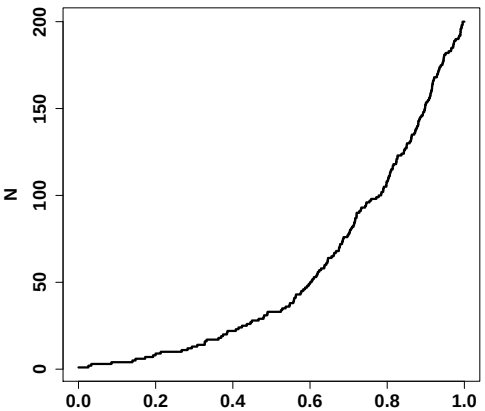
$X_0 = \frac{1}{4}$ $\theta = \frac{1}{4}$

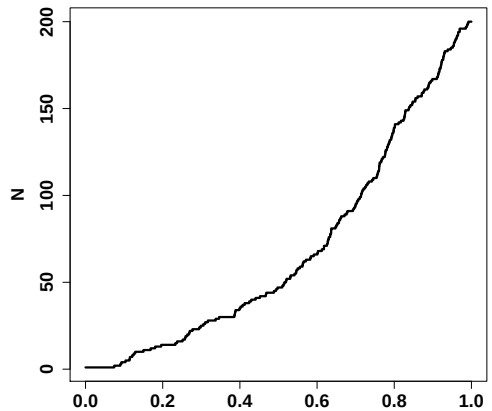




$X_0 = \frac{1}{4}$ $\theta = \frac{5}{4}$

$X_0 = \frac{5}{4}$ $\theta = \frac{1}{4}$





$X_0 = \frac{5}{4}$ $\theta = \frac{5}{4}$

Fig. 3 in K. Bartoszek and P. Liò. Modelling trait-dependent speciation with Approximate Bayesian Computation. Acta Phys. Pol. B Proc. Supp. 12(1):25-47, 2019.

Trait independent speciation (ABC)

```
# simulation+reestimation repeat 170 times,
# tree: 200 tips
PCM_ABC(phyltree,data,par0,simulate_OU_sde,
fbirth,NULL,X0=0,step=0.001,abcsteps=1000,
eps=0.2,tree.fixed=TRUE,
dist_method=c("variancemean",NA))
```

$$\mathrm{d}X_t = -\alpha(X_t - \theta)\mathrm{d}t + \sigma \mathrm{d}W_t = -X_t \mathrm{d}t + \mathrm{d}W_t$$

$$\hat{\alpha} = 2.352 \pm 2.069,\ \hat{\sigma} = 1.319 \pm 0.626,$$
$$v_y = \frac{\sigma^2}{2\alpha},\ \hat{v_y} = 0.521 \pm 0.270,\ \hat{\theta} = 0.027 \pm 0.311$$

Trait independent speciation (ABC)

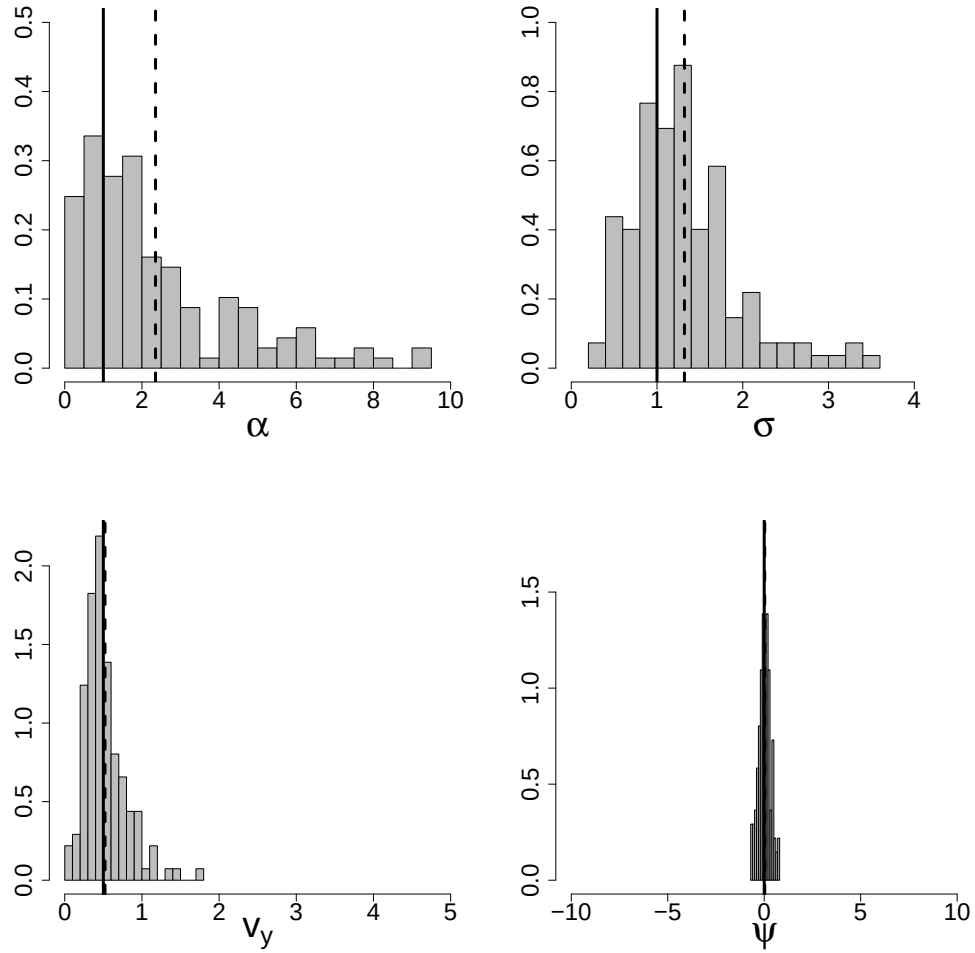


Fig. 4 in K. Bartoszek and P. Liò. Modelling trait-dependent speciation with Approximate Bayesian Computation. Acta Phys. Pol. B Proc. Supp. 12(1):25-47, 2019.

Trait dependent speciation (ABC)

```
# simulate+estimate repeat 170 times, 200 tips
PCM_ABC(phyltree,data,par0,simulate_OU_sde,
fbirth,NULL,X0=0,step=0.001,abcsteps=1000,
eps=0.2,tree.fixed=FALSE,
dist_method=c("variancemean","wRF.dist"))
```

$$\mathrm{d}X_t = -\alpha(X_t - \theta)\mathrm{d}t + \sigma \mathrm{d}W_t = -X_t \mathrm{d}t + \mathrm{d}W_t$$

birth rate: $scale/(1+\exp(-X_t)) = 5/(1+\exp(-X_t))$

$$\hat{\alpha} = 1.928 \pm 1.601, \hat{\sigma} = 1.415 \pm 0.842, \hat{v_y} = 0.731 \pm 0.647,$$
$$\hat{\theta} = 0.589 \pm 4.233,\ \hat{scale} = 6.523 \pm 5.567$$

Trait dependent speciation (ABC)

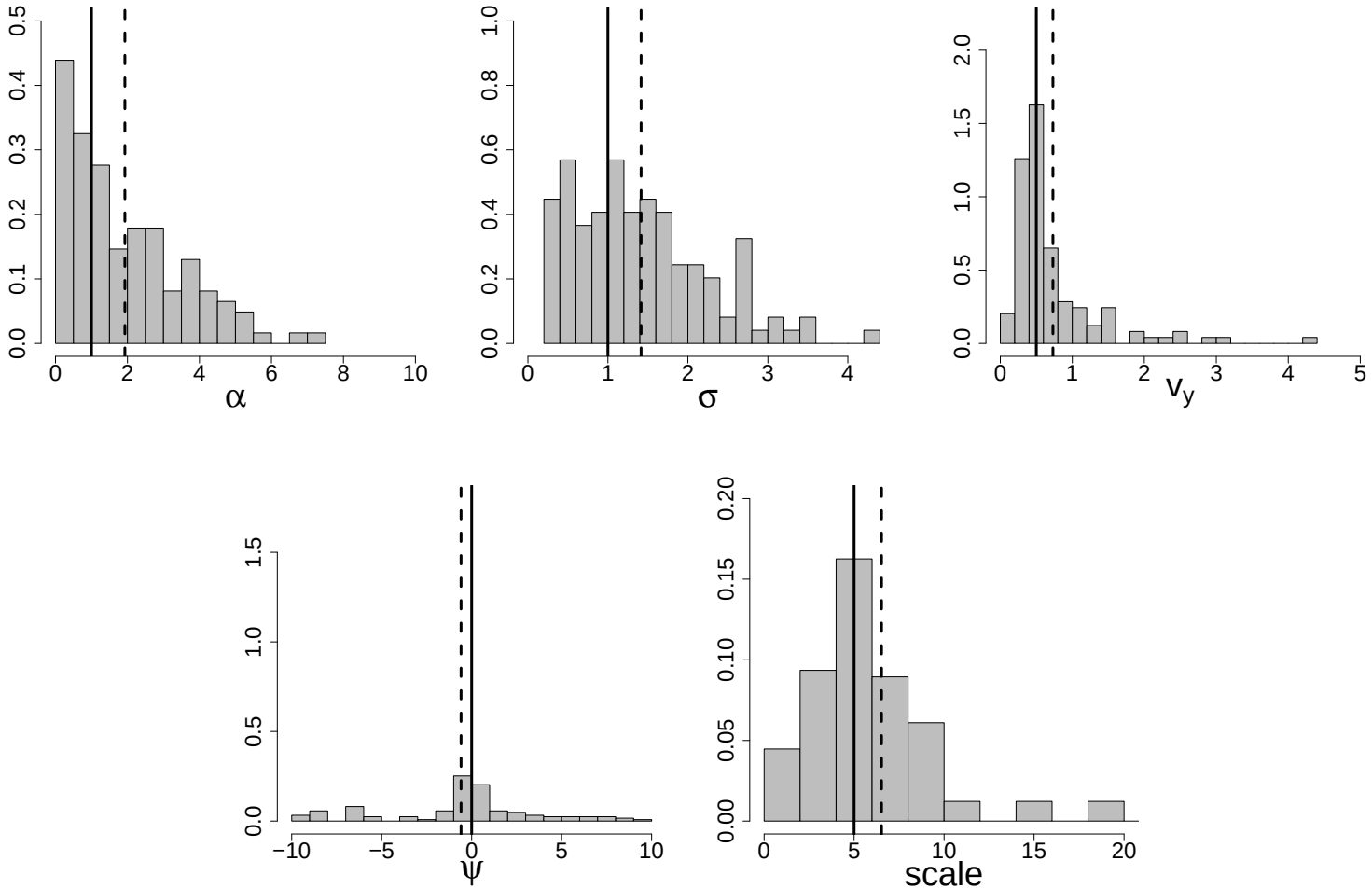


Fig. 5 in K. Bartoszek and P. Liò. Modelling trait-dependent speciation with Approximate Bayesian Computation. Acta Phys. Pol. B Proc. Supp. 12(1):25-47, 2019.

Trait independent speciation (ML)

A=**diag**(1:4); Syy=**diag**(1,4,4); X0=Psi=**matrix**(0,4,1)

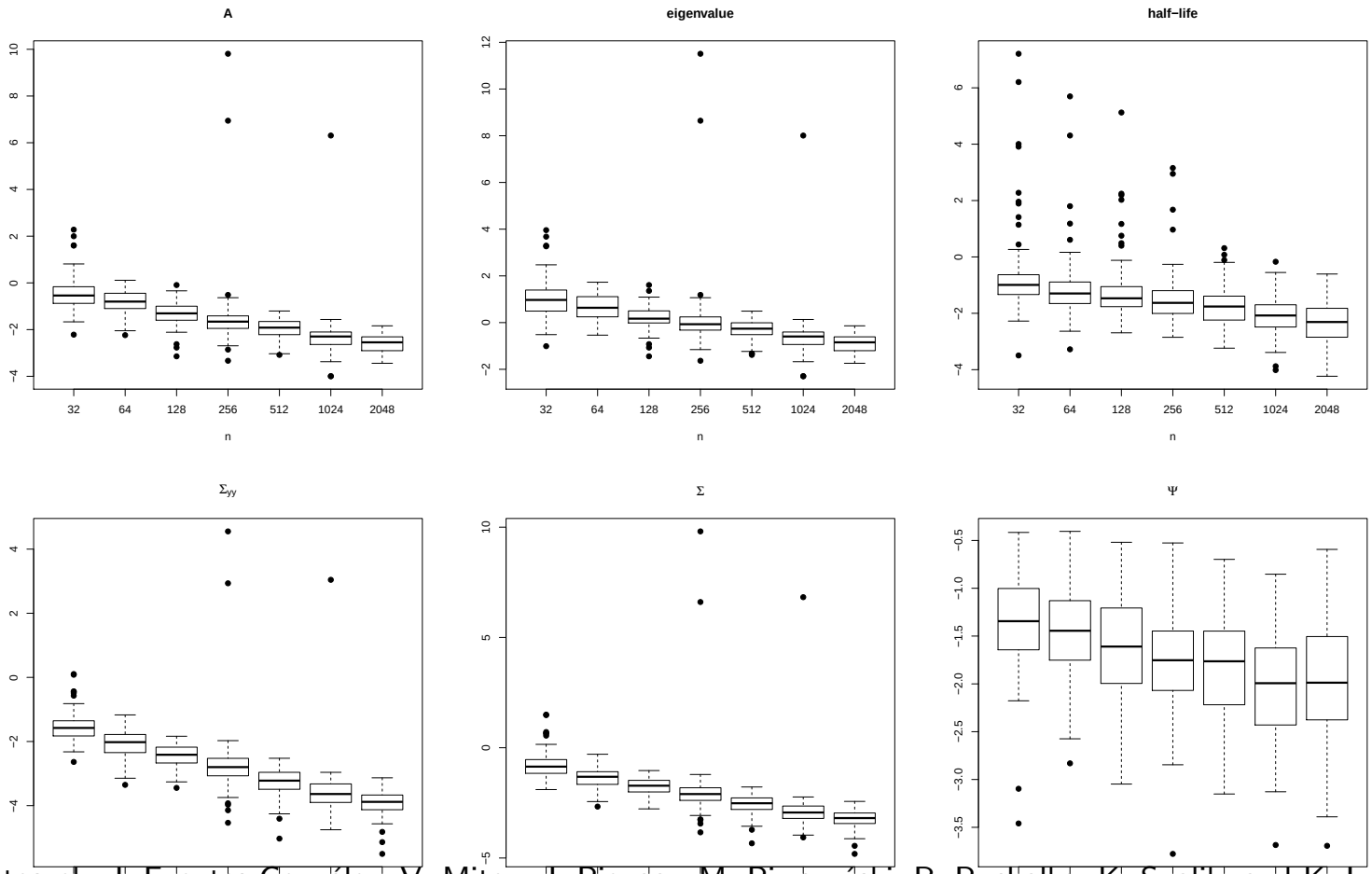


Fig. S2 in K. Bartoszek, J. Fuentes-González, V. Mitov, J. Pienaar, M. Piwczyński, R. Puchałka, K. Spalik and K. L. Voje. Model Selection Performance in Phylogenetic Comparative Methods Under Multivariate Ornstein-Uhlenbeck Models of Trait Evolution. Syst. Biol. 72(2):275-293, 2023.

YUIMA workshop

YUIMA tutorial workshop at Linköping University

17-20 March 2026

YUIMA R package for simulation and inference for SDEs
(also Lévy processes)
lecturers from Japan, Italy, Switzerland
details are being finalized, information will appear on my home page `https://www.ida.liu.se/~krzba67/`

YUIMA: `https://yuimaproject.com/`