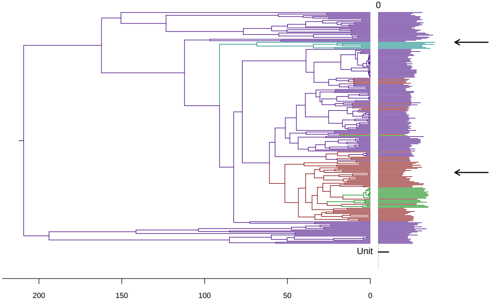
Shifted stochastic processes evolving on trees: application to models of adaptive evolution on phylogenies

Cécile Ané[1,2], Paul Bastide[3,4], Mahendra Mariadassou[4], Stéphane Robin[3]

[1] Department of Statistics, University of Wisconsin–Madison, WI, 53706, USA
[2] Department of Botany, University of Wisconsin–Madison, WI, 53706, USA
[3] INRA - AgroParisTech, UMR518 MIA-Paris, F-75231 Paris Cedex 05, France
[4] INRA, UR1404 Unité MaIAGE, F78352 Jouy-en-Josas, France.

19 November 2015

Introduction



Chelonian phylogenetic tree with habitats.
(Jaffe et al., 2011).

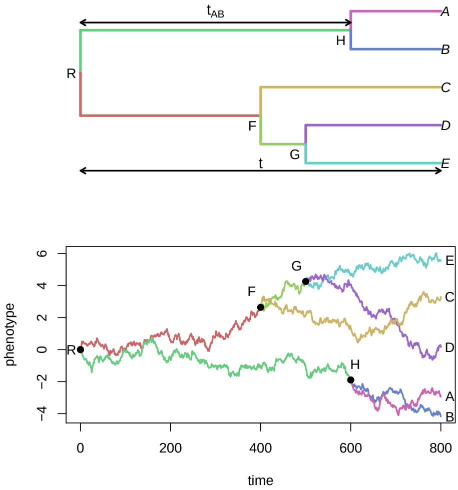
Dermochelys Coriacea

Homopus Areolatus

- A phylogenetic tree for a set of species
- One or several traits measured for each species

Outline

Stochastic Process on a Tree (Felsenstein, 1985)

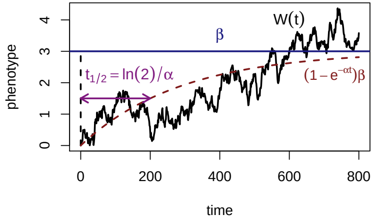


Brownian Motion:

$$\mathbb{V}\text{ar}\left[A \mid R\right] = \sigma^2 t$$

$$\mathbb{C}\text{ov}\left[A; B \mid R\right] = \sigma^2 t_{AB}$$

OU Modeling (Hansen, 1997)



$$dW(t) = \alpha[\beta(t) - W(t)]dt + \sigma dB(t)$$

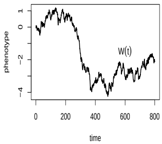
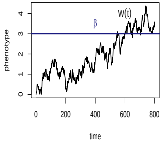
Deterministic part :

- $\beta(t)$: primary optimum, mechanistically defined.
- $\ln(2)/\alpha$: phylogenetic half live.

Stochastic part :

- $W(t)$: actual optimum (trait value).
- $\sigma dB(t)$ Brownian fluctuations.

BM vs OU

	Equation	Stationary State	Variance
	$dW(t) = \sigma dB(t)$	None.	$\sigma_{ij} = \sigma^2 t_{ij}$
	$dW(t) = \sigma dB(t)$ $+\alpha[\beta(t) - W(t)]dt$	$\begin{cases} \mu = \beta_0 \\ \gamma^2 = \dfrac{\sigma^2}{2\alpha} \end{cases}$	$\sigma_{ij} = \gamma^2 e^{-\alpha(t_i+t_j)} \times (e^{2\alpha t_{ij}} - 1)$

Underlying Assumptions

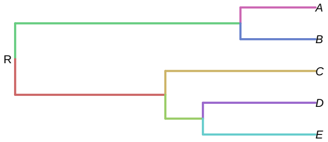
Fixed Tree Model assumes the trait(s) evolve independently from the tree

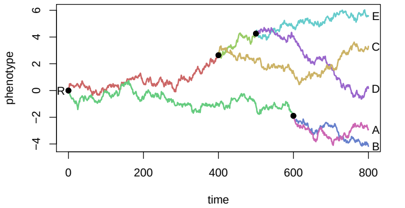
$\mapsto$ No interaction between speciation rate and trait

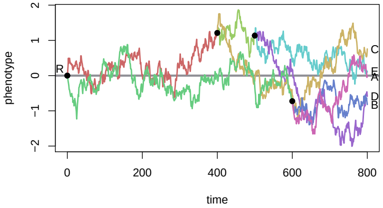
Functionnal Trait OU: stabilizing selection

$\mapsto$ Trait must be linked to the fitness of its bearer

Shifts







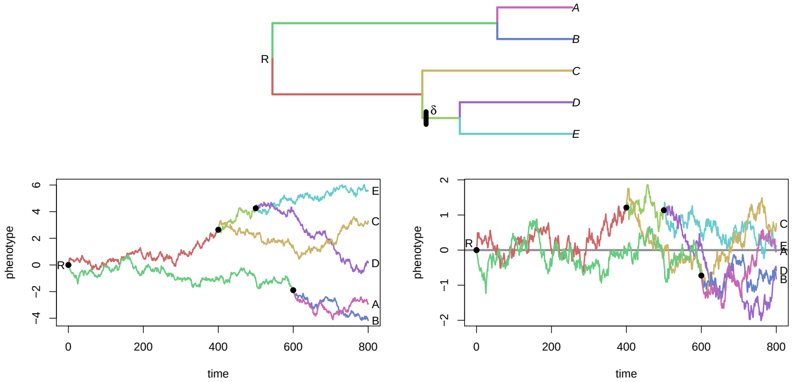
BM Shifts in the **mean**:

$$m_{\text{child}} = m_{\text{parent}} + \delta$$

OU Shifts in the **optimal value**:

$$\beta_{\text{child}} = \beta_{\text{parent}} + \delta$$

Shifts



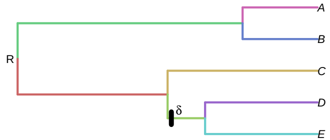
BM Shifts in the **mean**:

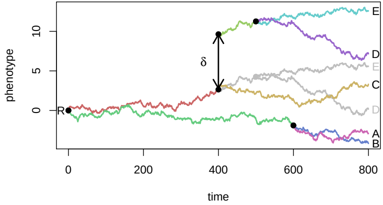
$$m_{\text{child}} = m_{\text{parent}} + \delta$$

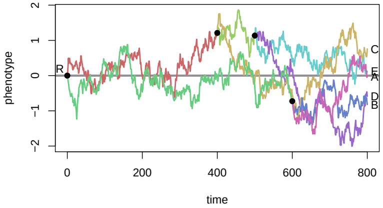
OU Shifts in the **optimal value**:

$$\beta_{\text{child}} = \beta_{\text{parent}} + \delta$$

Shifts







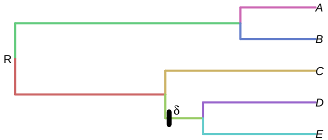
BM Shifts in the **mean**:

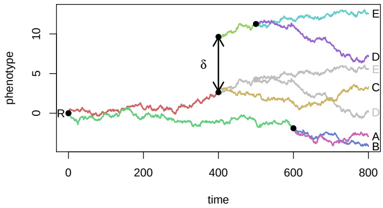
$$m_{\text{child}} = m_{\text{parent}} + \delta$$

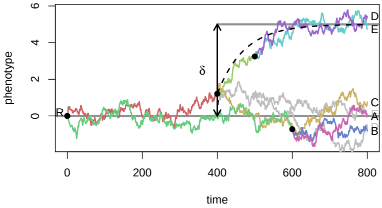
OU Shifts in the **optimal value**:

$$\beta_{\text{child}} = \beta_{\text{parent}} + \delta$$

Shifts







BM Shifts in the **mean**:

$$m_{\text{child}} = m_{\text{parent}} + \delta$$

OU Shifts in the **optimal value**:

$$\beta_{\text{child}} = \beta_{\text{parent}} + \delta$$

Outline

BM vs OU - bis

If the tree is ultrametric and the root is fixed, then:

$$\text{OU} \iff \text{BM on a re-scaled tree with } t' = e^{-2\alpha h}(e^{2\alpha t} - 1)$$

OU: Non-identifiability of μ and β_0

Simple process on a fixed tree, no shifts, α fixed.

BM $\iff$ OU

OU: Non-identifiability of μ and β_0

Simple process on a fixed tree, no shifts, α fixed.

BM : 2 parameters $\iff$ OU

- σ^2 variance
- μ ancestral state

OU: Non-identifiability of μ and β_0

Simple process on a fixed tree, no shifts, α fixed.

BM : 2 parameters $\iff$ OU : 3 parameters

BM:
- σ^2 variance
- μ ancestral state

OU:
- σ^2 variance
- μ ancestral state
- β_0 optimal value

OU: Non-identifiability of μ and β_0

Simple process on a fixed tree, no shifts, α fixed.

BM : 2 parameters

- σ^2 variance
- μ ancestral state

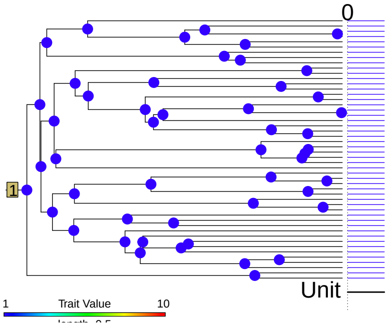
$\iff$ OU : 3 parameters

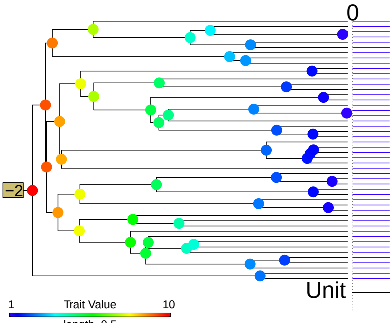
- σ^2 variance
- μ ancestral state
- β_0 optimal value

$$\lambda = \mu e^{-\alpha h} + \beta_0(1 - e^{-\alpha h})$$

OUfun: Non-identifiability of μ and β_0

Only $\lambda = \mu e^{-\alpha h} + \beta_0(1 - e^{-\alpha h})$ is identifiable





OUfun, with:

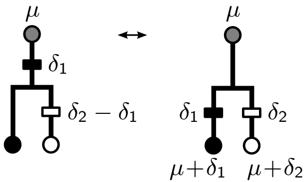
$\lambda = \beta_0 = \mu = 1$ and $\ln(2)/\alpha = 0.5$

Same OUfun, with:

$\lambda = 1,\ \beta_0 = -2,\ \mu = 10$

Equivalencies

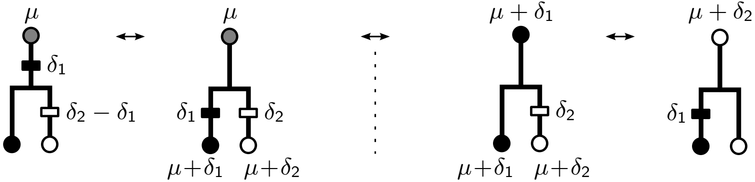
- K fixed, several equivalent solutions.



- Problem of over-parametrization: parsimonious configurations.

Equivalencies

- K fixed, several equivalent solutions.

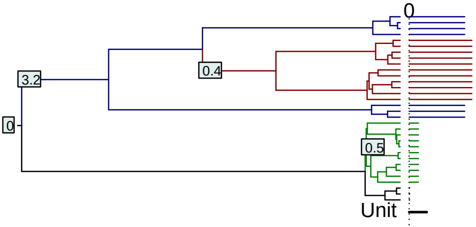


- Problem of over-parametrization: parsimonious configurations.

Process Induced Tip Coloring

Definition (Tips Coloring)

Two tips have the same color if they have the same mean under the process studied.



$$\color{green}{BM} \quad m_Y = T\Delta^{BM}$$

Parsimonious Solution : Definition

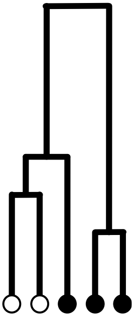
Definition (Parsimonious Allocation)

A coloring of the tips being given, a *parsimonious* allocation of the shifts is such that it has a minimum number of shifts.

Parsimonious Solution : Definition

Definition (Parsimonious Allocation)

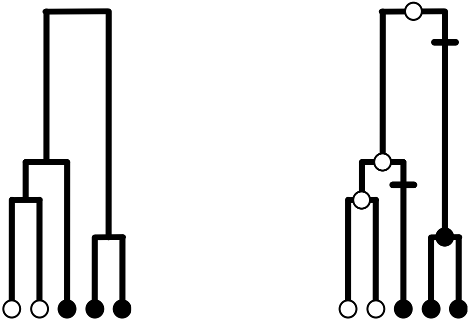
A coloring of the tips being given, a *parsimonious* allocation of the shifts is such that it has a minimum number of shifts.



Parsimonious Solution : Definition

Definition (Parsimonious Allocation)

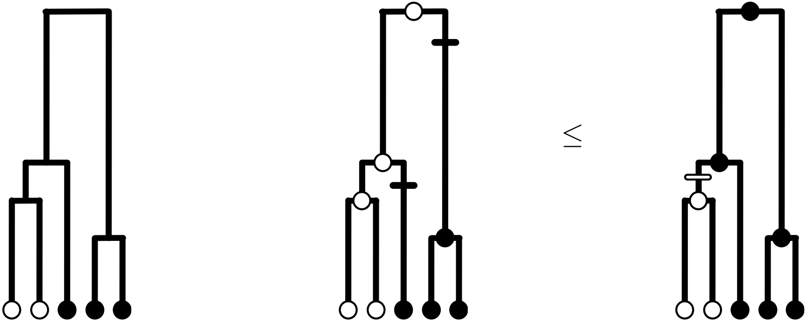
A coloring of the tips being given, a *parsimonious* allocation of the shifts is such that it has a minimum number of shifts.



Parsimonious Solution : Definition

Definition (Parsimonious Allocation)

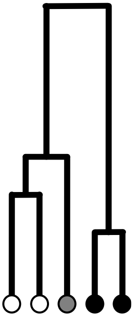
A coloring of the tips being given, a *parsimonious* allocation of the shifts is such that it has a minimum number of shifts.



Parsimonious Solution : Definition

Definition (Parsimonious Allocation)

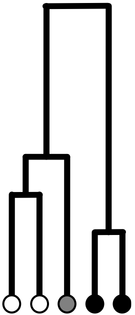
A coloring of the tips being given, a *parsimonious* allocation of the shifts is such that it has a minimum number of shifts.

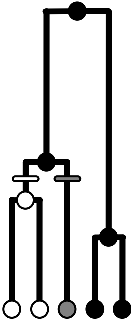


Parsimonious Solution : Definition

Definition (Parsimonious Allocation)

A coloring of the tips being given, a *parsimonious* allocation of the shifts is such that it has a minimum number of shifts.

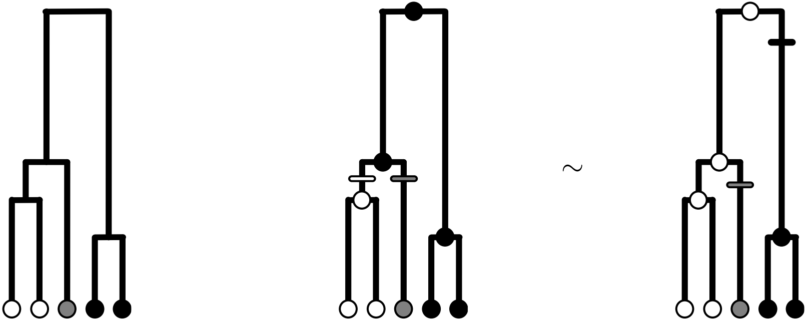




Parsimonious Solution : Definition

Definition (Parsimonious Allocation)

A coloring of the tips being given, a *parsimonious* allocation of the shifts is such that it has a minimum number of shifts.



Equivalent Parsimonious Allocations

Definition (Equivalency)

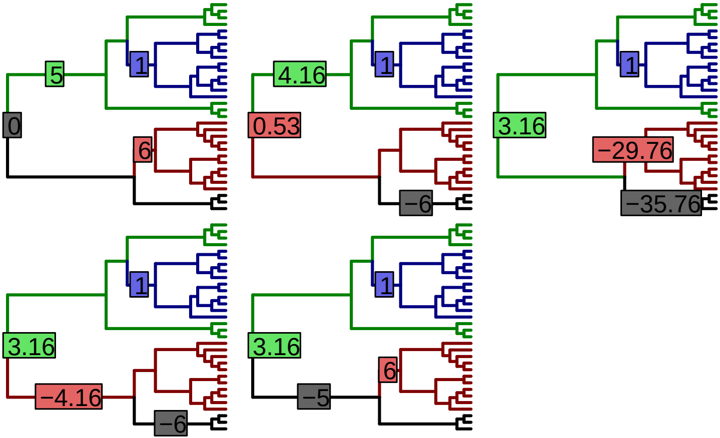
Two allocations are said to be *equivalent* (noted $\sim$) if they are both parsimonious and give the same colors at the tips.

Find one solution: Several existing Dynamic Programming algorithms (Fitch, Sankoff, see Felsenstein, 2004).

Enumerate all solutions: New recursive algorithm, adapted from previous ones (and implemented in R).

Algorithm
Colors/Model

Equivalent Parsimonious Solutions for an OU Model.

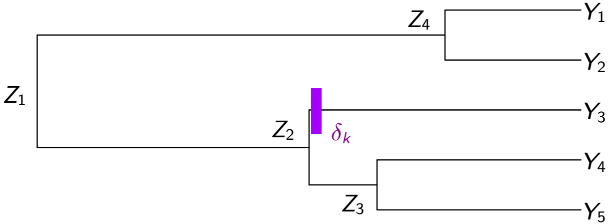


Equivalent allocations and values of the shifts.

BM

Outline

EM Algorithm: K fixed



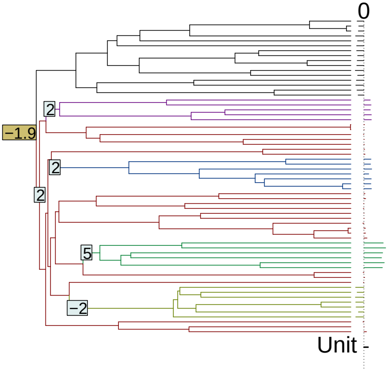
EM Algorithm: Recursive “Expectation - Maximization” for Likelihood Maximization

E step: Given current parameters, compute estimates of ancestral states Z

M step: Given these estimates, re-compute parameters

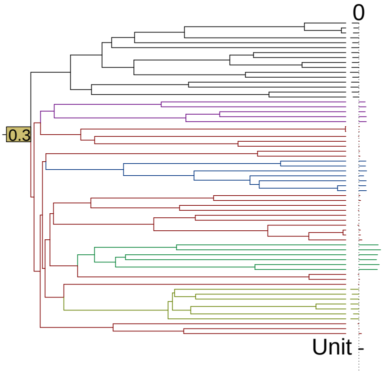
Details

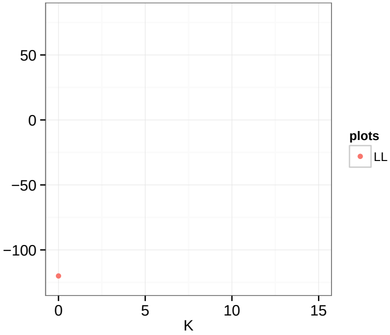
Model Selection on K



Simulated OUsun (α =3, γ^2 =0.1)

Model Selection on K

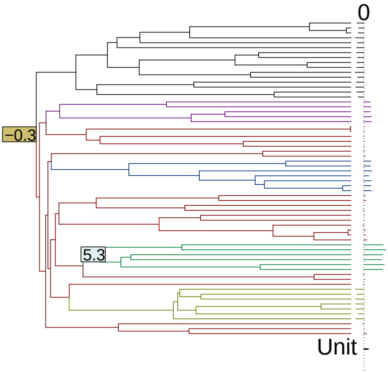


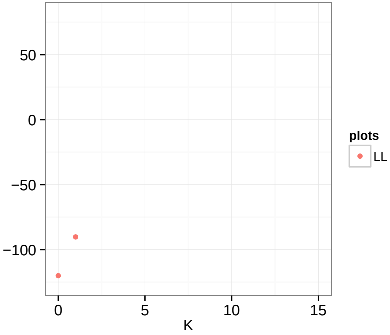


$$\hat{Y}_K = EM(K)$$

$$LL(\hat{Y}_K)$$

Model Selection on K

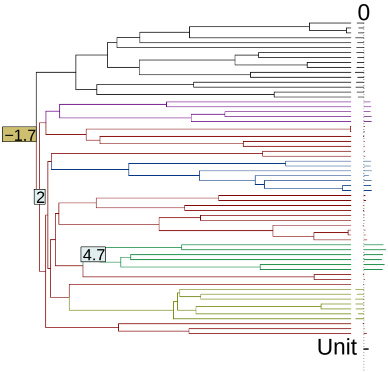


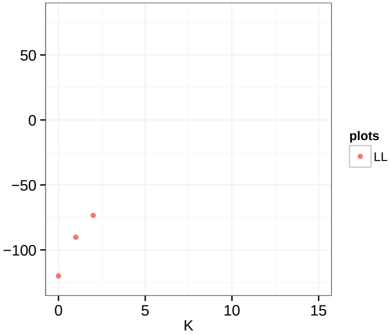


$$\hat{Y}_K = EM(K)$$

$$LL(\hat{Y}_K)$$

Model Selection on K

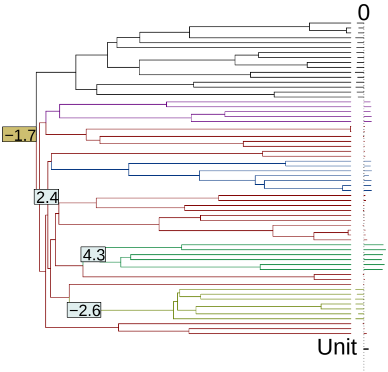


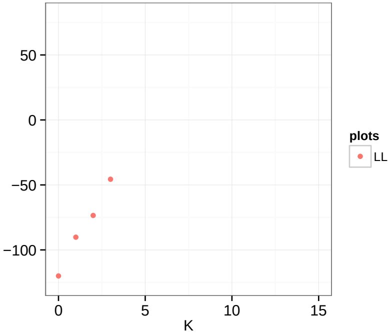


$$\hat{Y}_K = EM(K)$$

$$LL(\hat{Y}_K)$$

Model Selection on K

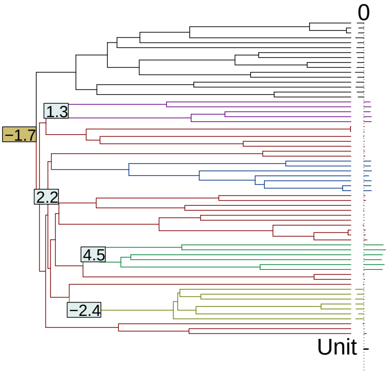


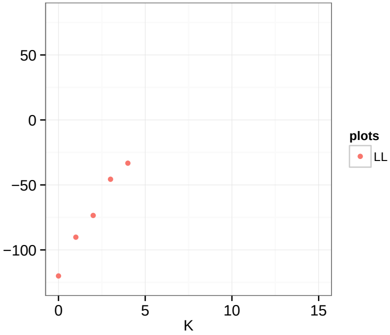


$$\hat{Y}_K = EM(K)$$

$$LL(\hat{Y}_K)$$

Model Selection on K

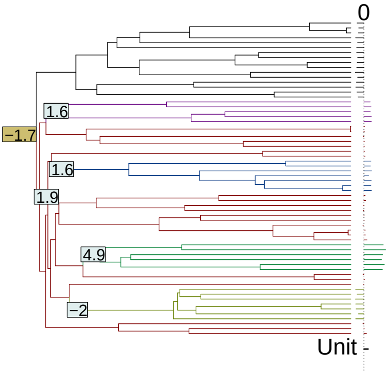


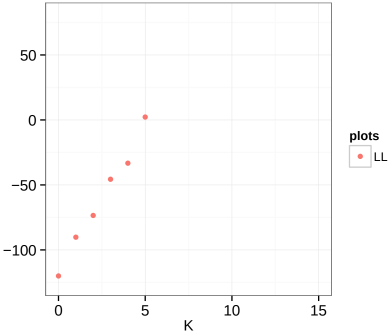


$$\hat{Y}_K = EM(K)$$

$$LL(\hat{Y}_K)$$

Model Selection on K

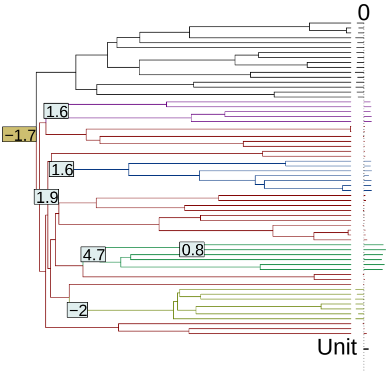


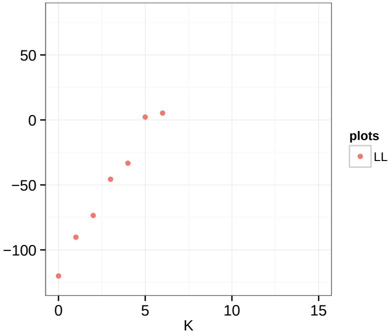


$$\hat{Y}_K = EM(K)$$

$$LL(\hat{Y}_K)$$

Model Selection on K

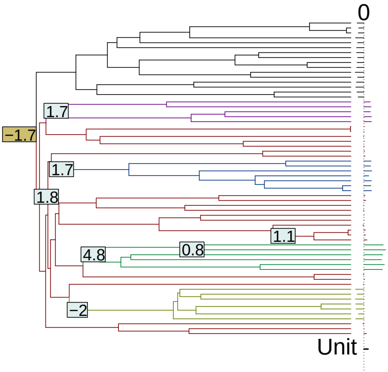


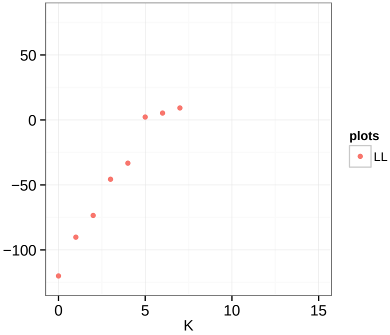


$$\hat{Y}_K = EM(K)$$

$$LL(\hat{Y}_K)$$

Model Selection on K

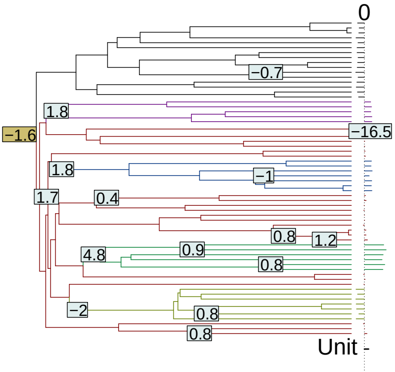


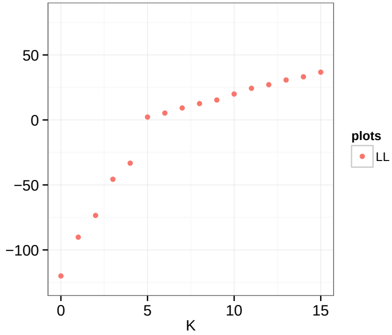


$$\hat{Y}_K = EM(K)$$

$$LL(\hat{Y}_K)$$

Model Selection on K



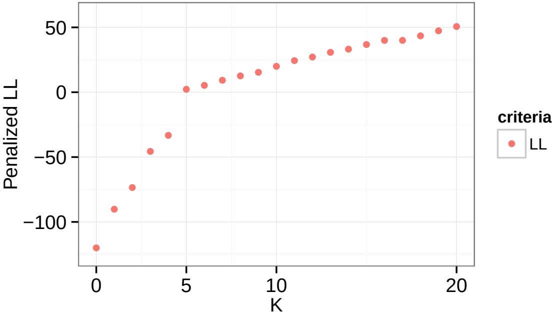


$$\hat{Y}_K = EM(K)$$

$$LL(\hat{Y}_K)$$

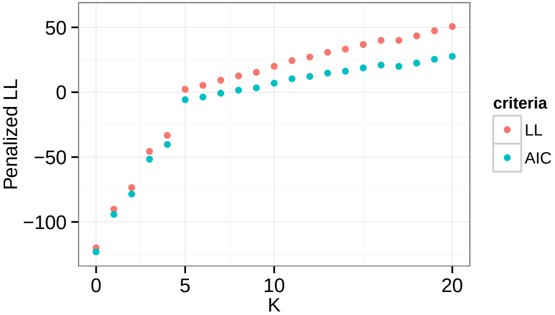
Model Selection: Penalized Likelihood

Idea $\hat{K} = LL(\hat{Y}_K) - \text{pen}'(K)$



Model Selection: Penalized Likelihood

Idea $\hat{K} = LL(\hat{Y}_K) - \text{pen}'(K)$

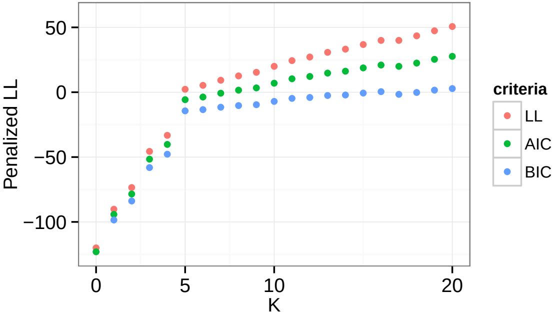


Penalties:

AIC $K + 3$

Model Selection: Penalized Likelihood

Idea $\hat{K} = LL(\hat{Y}_K) - \text{pen}'(K)$



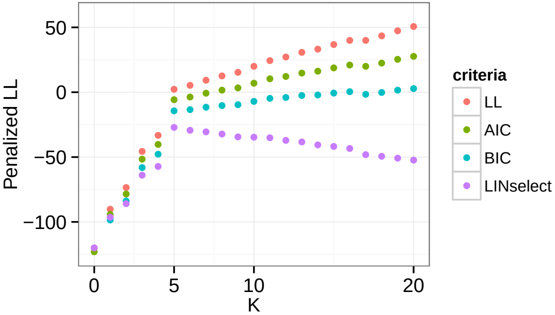
Penalties:

AIC $K + 3$

BIC $\frac{1}{2}(K + 3)\log(n)$

Model Selection: Penalized Likelihood

Idea $\hat{K} = LL(\hat{Y}_K) - \text{pen}'(K)$



Penalties:

AIC $K + 3$

BIC $\frac{1}{2}(K + 3)\log(n)$

LINselect $\text{pen}(n, K, N_K^{\mathcal{T}})$

Model Selection: How to choose the Penalty ?

$$\mathrm{pen}(n, K, N_K^{\mathcal{T}})$$

$N_K^{\mathcal{T}}$: Number of *different* models with K shifts

$\mapsto$ Two equivalent models count for only one !

Under the no-homoplasy hypothesis:

- $N_K^{\mathcal{T}} \leq \binom{m+n-1}{K} = \binom{\#\text{ of edges}}{\#\text{ of shifts}}$
- A recursive algorithm can compute $N_K^{\mathcal{T}}$ (implemented in R).

$\mapsto$ Generally dependent on the topology of the tree.

- Binary tree: $N_K^{\mathcal{T}} = \binom{2n-2-K}{K} = \binom{\#\text{ of edges}-\#\text{ of shifts}}{\#\text{ of shifts}}$

No Homoplasy Algorithm Formal Description

Model Selection: Proposed Penalty (LINselect)

$$\text{pen}(n, K, N_K^{\mathcal{T}})$$

Based on Baraud, Giraud, and Huet (2009)

- Non-asymptotic bound.
- Unknown variance.
- No constant to be calibrated.

Guarantee "Oracle Inequality"

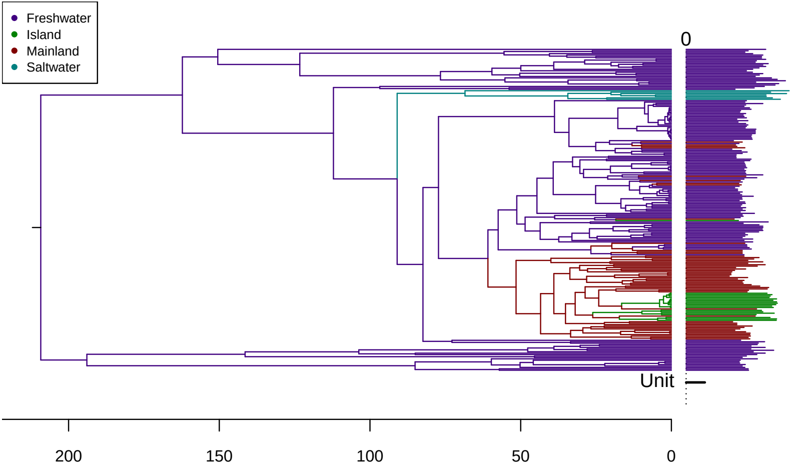
Novelties

- Non iid variance.
- Penalty depends on the tree topology (through $N_K^{\mathcal{T}}$).

Details

Outline

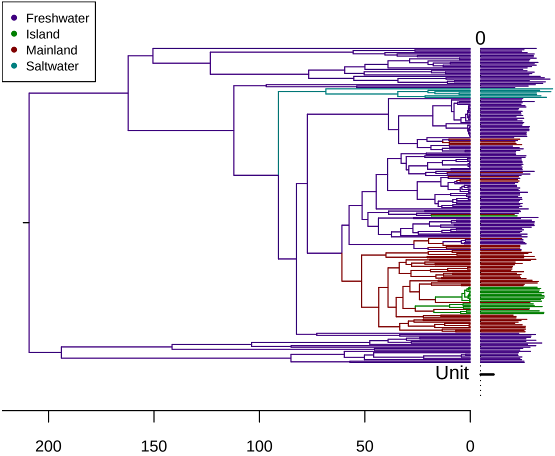
Data



Colors: habitats.

Fixed Regimes

(Jaffe et al., 2011)



Colors: habitats.

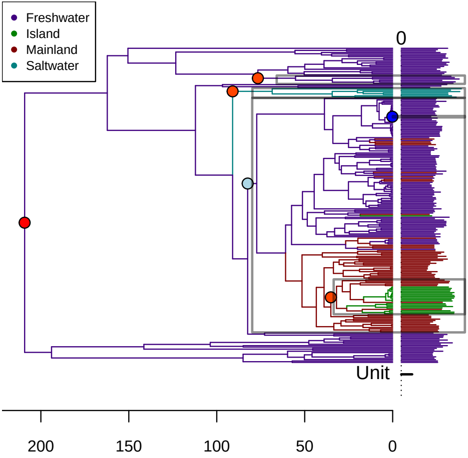
	Habitat
No. of shifts	16
No. of regimes	4
lnL	-133.86
$\ln 2/\alpha$ (%)	7.44
γ^2	0.33
CPU time (min)	65.25

Automatic detection of shifts

```r
## Grid on alpha
alpha_grid <- 1:10/100

## Inference
res <- PhyloEM(phylo = tree,
               Y_data = data,
               process = "OU",
               K_max = 20,
               alpha_known = TRUE,
               alpha = alpha_grid,
               random.root = TRUE,
               methods.segmentation = "lasso")
```

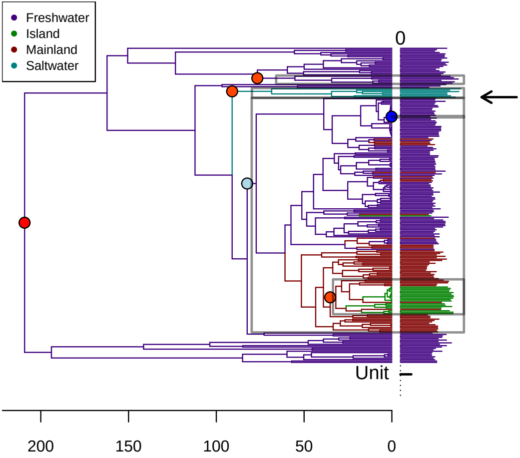
Automatic Detection of Shifts



Colors: habitats.
Boxes: selected EM regimes.

	Habitat	EM
No. of shifts	16	5
No. of regimes	4	6
lnL	-133.86	-97.59
$\ln 2/\alpha$ (%)	7.44	5.43
γ^2	0.33	0.22
CPU time (min)	65.25	134.49

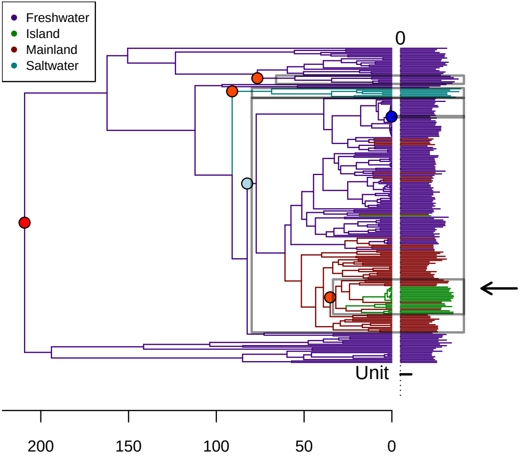
Automatic Detection of Shifts



Colors: habitats.
Boxes: selected EM regimes.

Chelonia mydas

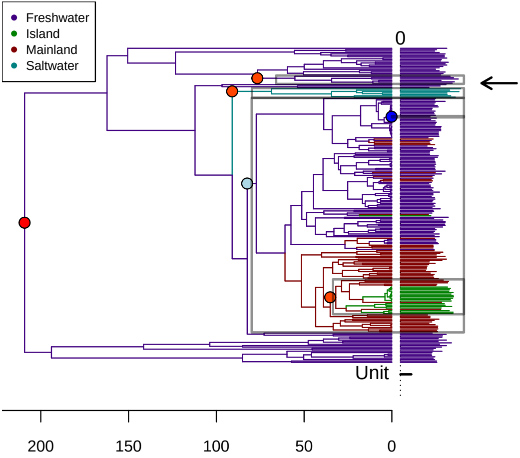
Automatic Detection of Shifts



Colors: habitats.
Boxes: selected EM regimes.

Geochelone nigra abingdoni

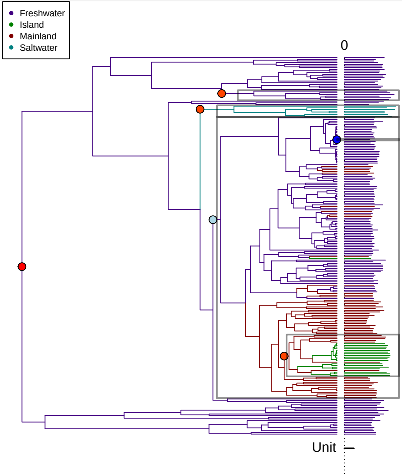
Automatic Detection of Shifts



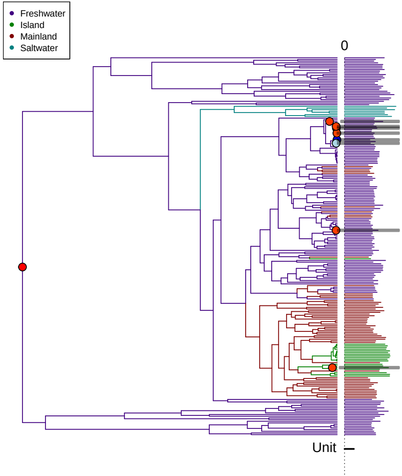
Colors: habitats.
Boxes: selected EM regimes.

Chitra indica

Comparison with BM

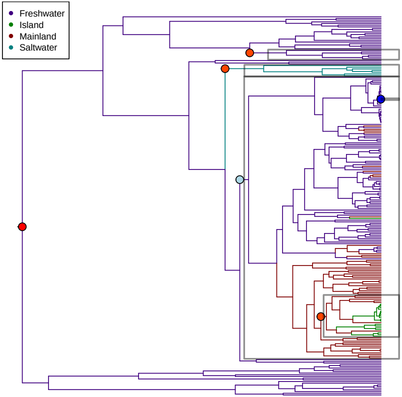


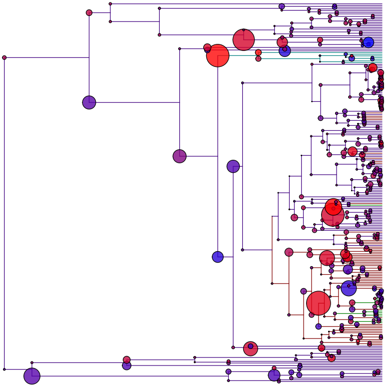
OU: 5 shifts selected



BM: 8 shifts selected

Comparison with Bayou

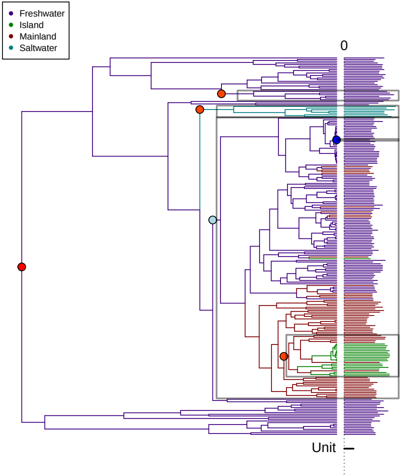




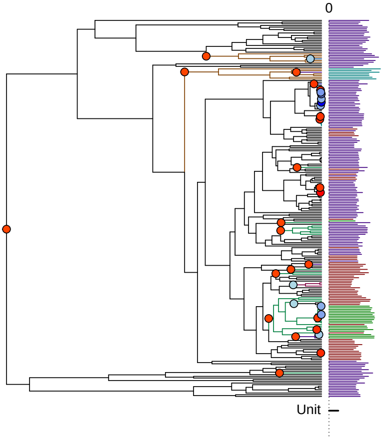
Colors: habitats.
Boxes: selected EM regimes.

Colors: habitats.
Circles: posterior probability of shift.

Comparison with SURFACE



Colors: habitats.
Boxes: selected EM regimes.



Colors at tips: habitats.
Colors of edges: Surface Regimes

Summary

	EM	Habitat	bayou	Surface
No. of shifts	5	16	17	33
No. of regimes	6	4	18	13
lnL	-97.59	-133.86	-91.54	30.38
MlnL	NaN	NaN	-149.09	NaN
$\ln 2/\alpha$ (%)	5.43	7.44	1.90	40.28
γ^2	0.22	0.33	0.16	0.21
CPU time (min)	134.49	65.25	136.81	634.16

Outline

BM Model

Data *n vectors* of p traits at the tips: $\mathbf{Y}_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{ip} \end{pmatrix}$

Model $d\mathbf{W}(t) = \mathbf{\Sigma} d\mathbf{B}_t$

Rate matrix $\mathbf{R} = \mathbf{\Sigma}\mathbf{\Sigma}^T = \begin{pmatrix} R_{11} & \cdots & R_{1p} \\ \vdots & \ddots & \vdots \\ R_{p1} & \cdots & R_{pp} \end{pmatrix}$

Covariances $\mathbb{C}\text{ov}\left[Y_{il}; Y_{jq}\right] = t_{ij} R_{lq}$ for i, j tips, and l, q characters

Shifts K shifts $\boldsymbol{\delta}_1, \cdots, \boldsymbol{\delta}_K$ *vectors* size p

$\mapsto$ All the characters shift at the same time

OU Model: General Case

Data *n vectors* of p traits at the tips: $\mathbf{Y}_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{ip} \end{pmatrix}$

SDE $\mathbf{A}$ $(p \times p)$ "selection strength"

$$d\mathbf{W}(t) = -\mathbf{A}(\mathbf{W}(t) - \beta(t))dt + \mathbf{\Sigma} d\mathbf{B}_t$$

Covariances Depends on $\mathbf{R} = \mathbf{\Sigma}\mathbf{\Sigma}^T$ and $\mathbf{A}$ in general.

Shifts K shifts $\boldsymbol{\delta}_1, \cdots, \boldsymbol{\delta}_K$ *vectors* size p

$\mapsto$ On the optimal values

Intractable

OU Model: Scalar Case

Hyp $\mathbf{A} = \alpha \mathbf{I}_p = \begin{pmatrix} \alpha & 0 & \cdots & 0 \\ 0 & \alpha & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \alpha \end{pmatrix}$ is "scalar"

Correlations Depends on $\mathbf{R}$ and α:

$$\mathbb{C}\mathrm{ov}\left[Y_{il}; Y_{jq}\right] = \frac{1}{2\alpha}\left(e^{2\alpha t_{ij}} - 1\right) e^{-\alpha(t_j + t_j)} R_{lq}$$

Shifts K shifts $\boldsymbol{\delta}_1, \cdots, \boldsymbol{\delta}_K$ *vectors* size p

$\mapsto$ On the optimal values

Equivalent to a re-scaled BM

Statistical Inference

EM Maximum Likelihood solution when K is fixed

$\mapsto$ Can deal with missing data.

Model Selection Use the “Slope Heuristic” on the likelihood

Simulated Example

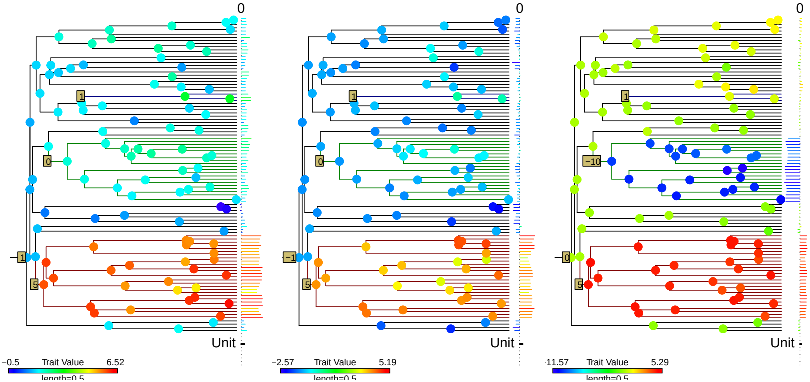


Figure: *Simulated BM Process with 3 shifts.*

Simulated Example

```
res <- PhyloEM(phylo = tree,
               Y_data = Y_data,
               process = "BM",
               K_max = 10,
               random.root = FALSE,
               progress.bar = FALSE)
```

$$\mathbf{R} = \begin{pmatrix} 0.5 & 0.2 & 0.2 \\ 0.2 & 0.5 & 0.2 \\ 0.2 & 0.2 & 0.5 \end{pmatrix} \qquad \hat{\mathbf{R}} = \begin{pmatrix} 0.45 & 0.17 & 0.15 \\ 0.17 & 0.43 & 0.22 \\ 0.15 & 0.22 & 0.48 \end{pmatrix}$$

Simulated Example

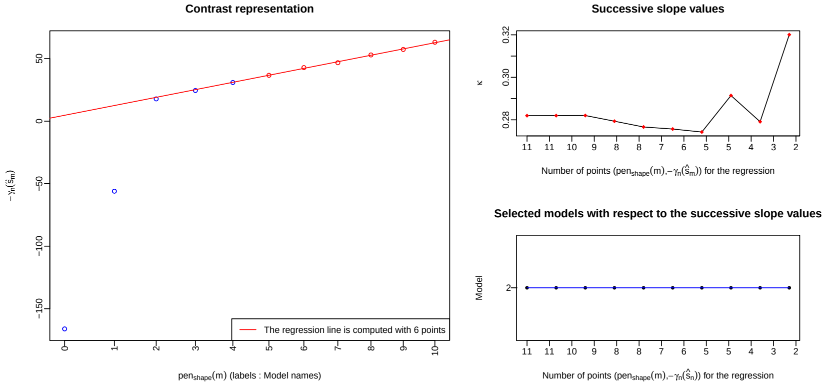


Figure: capushe *output for penalized log-likelihood.*

Simulated Example

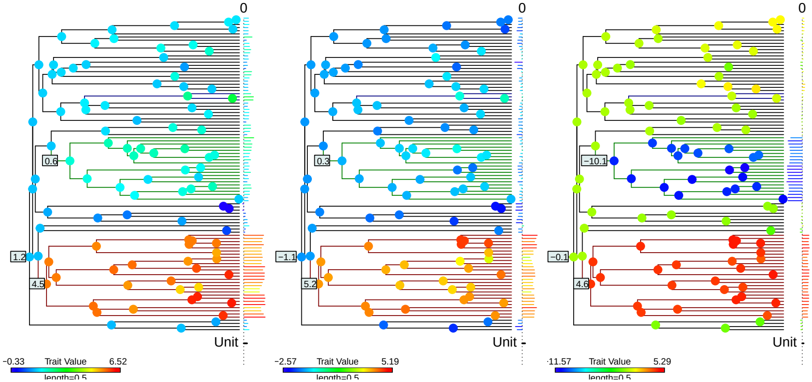


Figure: Reconstructed BM Process. Only 2 shifts are recovered.

Conclusion and Perspectives

A general inference framework for trait evolution models.

Conclusions

- Some problems of identifiability arise.
- Univariate Case: EM & Model selection for Maximum Likelihood
- Multivariate: BM, OU scalar

R codes Available on GitHub:

`https://github.com/pbastide/Phylogenetic-EM`

Perspectives

- Multivariate: reasonable assumptions on selection strength matrix $\mathbf{A}$?
- Deal with uncertainty (tree, data).
- Use fossil records.

Bibliography

Y. Baraud, C. Giraud, and S. Huet. Gaussian model selection with an unknown variance. *Annals of Statistics*, 37 (2):630–672, Apr. 2009.

J. Felsenstein. Phylogenies and the Comparative Method. *The American Naturalist*, 125(1):pp. 1–15, Jan. 1985. ISSN 00030147.

J. Felsenstein. *Inferring Phylogenies*. Sinauer Associates, Suderland, USA, 2004.

T. F. Hansen. Stabilizing selection and the comparative analysis of adaptation. *Evolution*, 51(5):1341–1351, oct 1997.

A. L. Jaffe, G. J. Slater, and M. E. Alfaro. The evolution of island gigantism and body size variation in tortoises and turtles. *Biology letters*, 2011.

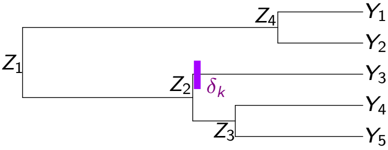
J. C. Uyeda and L. J. Harmon. A Novel Bayesian Method for Inferring and Interpreting the Dynamics of Adaptive Landscapes from Phylogenetic Comparative Data. *Syst. Biol.*, July 2014. doi: 10.1093/sysbio.

Photo Credits :

- "Parrot-beaked Tortoise Homopus areolatus CapeTown 8" by Abu Shawka - Own work. Licensed under CC0 via Wikimedia Commons
- "Leatherback sea turtle Tinglar, USVI (5839996547)" by U.S. Fish and Wildlife Service Southeast Region - Leatherback sea turtle/ Tinglar, USVIUploaded by AlbertHerring. Licensed under CC BY 2.0 via Wikimedia Commons
- "Hawaii turtle 2" by Brocken Inaglory. Licensed under CC BY-SA 3.0 via Wikimedia Commons
- "Dudhwalive chitra" by Krishna Kumar Mishra — Own work. Licensed under CC BY 3.0 via Wikimedia Commons
- "Lonesome George in profile" by Mike Weston - Flickr: Lonesome George 2. Licensed under CC BY 2.0 via Wikimedia Commons
- "Florida Box Turtle Digon3a", "Jonathan Zander (Digon3)" derivative work: Materialscientist (talk). Licensed under CC BY-SA 3.0 via Wikimedia Commons

Appendices

EM Algorithm: K fixed



$$X_j | X_{\mathrm{pa}(j)} \sim \mathcal{N}\left(q_j X_{\mathrm{pa}(j)} + r_j + s_j \sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k, \sigma_j^2 \right)$$

$$\log p_\theta(Y) = \mathbb{E}_\theta[\log p_\theta(Z, Y) \mid Y] - \mathbb{E}_\theta[\log p_\theta(Z) \mid Y]$$

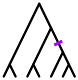
EM Algorithm: Maximize $\mathbb{E}_\theta[\log p_\theta(Z, Y) \mid Y]$

E step: Given θ^h, compute $p_{\theta^h}(Z \mid Y)$

M step: $\theta^{h+1} = \mathrm{argmax}_\theta\, \mathbb{E}_{\theta^h}[\log p_\theta(Z, Y) \mid Y]$

Likelihood

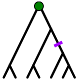
$$\begin{cases} X_1 \sim \mathcal{N}(\mu, \gamma^2) \\ \forall j > 1, \quad X_j | X_{\mathrm{pa}(j)} \sim \mathcal{N}\left(q_j X_{\mathrm{pa}(j)} + r_j + s_j \sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k, \sigma_j^2\right) \end{cases}$$



$$p_\theta(X) = p_\theta(Z_1) \prod_{1<j\leq m} p_\theta(Z_j|Z_{\mathrm{pa}(j)}) \prod_{1\leq i\leq n} p_\theta(Y_i|Z_{\mathrm{pa}(i')})$$

Likelihood

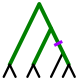
$$\begin{cases} X_1 \sim \mathcal{N}(\mu, \gamma^2) \\ \forall j > 1, \quad X_j | X_{\text{pa}(j)} \sim \mathcal{N}\left(q_j X_{\text{pa}(j)} + r_j + s_j \sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k, \sigma_j^2\right) \end{cases}$$



$$p_\theta(X) = p_\theta(Z_1) \prod_{1<j\leq m} p_\theta(Z_j|Z_{\text{pa}(j)}) \prod_{1\leq i\leq n} p_\theta(Y_i|Z_{\text{pa}(i')})$$

Likelihood

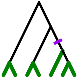
$$\begin{cases} X_1 \sim \mathcal{N}(\mu, \gamma^2) \\ \forall j > 1, \quad X_j | X_{\text{pa}(j)} \sim \mathcal{N}\left(q_j X_{\text{pa}(j)} + r_j + s_j \sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k, \sigma_j^2\right) \end{cases}$$



$$p_\theta(X) = p_\theta(Z_1) \prod_{1<j\leq m} p_\theta(Z_j | Z_{\text{pa}(j)}) \prod_{1\leq i\leq n} p_\theta(Y_i | Z_{\text{pa}(i')})$$

Likelihood

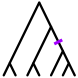
$$\begin{cases} X_1 \sim \mathcal{N}(\mu, \gamma^2) \\ \forall j > 1, \quad X_j | X_{\text{pa}(j)} \sim \mathcal{N}\left(q_j X_{\text{pa}(j)} + r_j + s_j \sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k, \sigma_j^2\right) \end{cases}$$



$$p_\theta(X) = p_\theta(Z_1) \prod_{1<j\leq m} p_\theta(Z_j | Z_{\text{pa}(j)}) \prod_{1\leq i\leq n} p_\theta(Y_i | Z_{\text{pa}(i')})$$

Likelihood

$$\begin{cases} X_1 \sim \mathcal{N}(\mu, \gamma^2) \\ \forall j > 1, \quad X_j | X_{\text{pa}(j)} \sim \mathcal{N}\left(q_j X_{\text{pa}(j)} + r_j + s_j \sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k, \sigma_j^2 \right) \end{cases}$$



$$p_\theta(X) = p_\theta(Z_1) \prod_{1<j\leq m} p_\theta(Z_j | Z_{\text{pa}(j)}) \prod_{1\leq i\leq n} p_\theta(Y_i | Z_{\text{pa}(i')})$$

$$\mathbb{E}\left[\log p_\theta(X) \mid Y\right] = -\sum_{j=2}^{m+n} C_j(\alpha, \tau, \delta) + \mathcal{F}\left(\theta, \mathbb{V}\text{ar}\left[Z_j \mid Y\right]_j, \mathbb{C}\text{ov}\left[Z_j; Z_{\text{pa}(j)} \mid Y\right]_j\right)$$

$$C_j(\alpha, \tau, \delta) = \sigma_j^{-2} \left(\mathbb{E}\left[X_j \mid Y\right] - q_j \mathbb{E}\left[X_{\text{pa}(j)} \mid Y\right] - r_j - s_j \sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k \right)^2$$

E step

Compute the following quantities:

$$\mathbb{E}^{(h)}[Z_j \mid Y], \mathbb{V}\mathrm{ar}^{(h)}\left[Z_j \mid Y\right], \mathbb{C}\mathrm{ov}^{(h)}\left[Z_j, Z_{\mathrm{pa}(j)} \mid Y\right]$$

- Using Gaussian properties. Need to invert matrices: complexity in $O(n^3)$.
- Using Gaussian properties **and** the tree structure: "Upward-Downward" algorithm. Complexity in $O(n)$.

M Step

Maximize:

$$\mathbb{E}\left[\log p_\theta(X) \mid Y\right] = -\sum_{j=2}^{m+n} C_j(\alpha, \tau, \delta) + \mathcal{F}^{(h)}\left(\mu, \gamma^2, \sigma^2, \alpha\right)$$

- μ, γ^2, σ^2: simple maximization
- τ, δ: discrete location of K shifts
  - Exact and fast for the BM
  - Heuristic for the OU: GEM
- α: numerical maximization

Initialization

The shifts (τ, δ) : Lasso regression.

$$\hat{\Delta} = \underset{\Delta}{\operatorname{argmin}} \left\{ \| Y - R\Delta \|^2_{\Sigma_{YY}} + \lambda \, |\Delta_{-1}|_1 \right\}$$

- Initialize Σ^2_{YY} with some default parameters, then estimate Δ with a Gauss Lasso procedure, using a Cholesky decomposition.

+

- λ chosen to get K shifts.

The selection strength α : Initialization using couples of tips.

Back

Cholesky Decomposition

The problem is:

$$\hat{\Delta} = \underset{\Delta}{\operatorname{argmin}} \left\{ \| Y - R\Delta \|^2_{\Sigma_{YY}} + \lambda \, |\Delta_{-1}|_1 \right\}$$

Cholesky decomposition of Σ_{YY}:

$$\Sigma_{YY} = LL^T \ , \ L \text{ a lower triangular matrix}$$

Then:

$$\| Y - R\Delta \|^2_{\Sigma_{YY}} = \left\| L^{-1} Y - L^{-1} R \Delta \right\|^2$$

And if $Y' = L^{-1} Y$ and $R' = L^{-1} R$, the problem becomes:

$$\hat{\Delta} = \underset{\Delta}{\operatorname{argmin}} \left\{ \left\| Y' - R'\Delta \right\|^2 + \lambda \, |\Delta_{-1}|_1 \right\}$$

Gauss Lasso

Let $\hat{m}_\lambda$ be the set of selected variables (including the root). Then:

$$\hat{\Delta}^{\text{Gauss}} = \Pi_{\hat{F}_\lambda}(Y') \text{ with } \hat{F}_\lambda = \text{Span}\{R'_j : j \in \hat{m}_\lambda\}$$

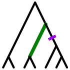
back

M Step: Segmentation

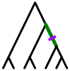
$$C_j(\alpha, \tau, \delta) = \sigma_j^{-2}\left(\mathbb{E}\left[X_j \mid Y\right] - q_j\mathbb{E}\left[X_{\text{pa}(j)} \mid Y\right] - r_j - s_j\sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k\right)^2$$

BM : $r_j = 0$, each cost is independent.

$$C_j^0(\alpha) = \sigma_j^{-2}\left(\mathbb{E}\left[X_j \mid Y\right] - q_j\mathbb{E}\left[X_{\text{pa}(j)} \mid Y\right]\right)^2$$



$$C_j^1(\alpha, \tau, \delta) = \sigma_j^{-2}\left(\mathbb{E}\left[X_j \mid Y\right] - q_j\mathbb{E}\left[X_{\text{pa}(j)} \mid Y\right] - s_j\sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k\right)^2$$



Algorithm:

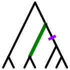
1. Find the K branches $j_1, \ldots, j_K$ with largest C_j^0;
2. Allocate one change point in the first K branches;
3. For each of these branches, set $\delta_{j_k}^{(h+1)}$ so that $C_j^1(\tau, \delta) = 0$

M Step: Segmentation

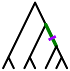
$$C_j(\alpha,\tau,\delta) = \sigma_j^{-2}\left(\mathbb{E}\left[X_j \mid Y\right] - q_j\mathbb{E}\left[X_{\mathrm{pa}(j)} \mid Y\right] - r_j - s_j\sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k\right)^2$$

BM : $r_j = 0$, each cost is independent.

$$C_j^0(\alpha) = \sigma_j^{-2}\left(\mathbb{E}\left[X_j \mid Y\right] - q_j\mathbb{E}\left[X_{\mathrm{pa}(j)} \mid Y\right]\right)^2$$



$$C_j^1(\alpha,\tau,\delta) = \sigma_j^{-2}\left(\mathbb{E}\left[X_j \mid Y\right] - q_j\mathbb{E}\left[X_{\mathrm{pa}(j)} \mid Y\right] - s_j\sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k\right)^2$$



Algorithm:

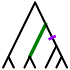
1. Find the K branches $j_1, \dots, j_K$ with largest C_j^0;
2. Allocate one change point in the first K branches;
3. For each of these branches, set $\delta_{j_k}^{(h+1)}$ so that $C_j^1(\tau,\delta) = 0$

M Step: Segmentation

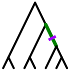
$$C_j(\alpha, \tau, \delta) = \sigma_j^{-2} \left(\mathbb{E}\left[X_j \mid Y\right] - q_j \mathbb{E}\left[X_{\mathrm{pa}(j)} \mid Y\right] - r_j - s_j \sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k \right)^2$$

BM : $r_j = 0$, each cost is independent.

$$C_j^0(\alpha) = \sigma_j^{-2} \left(\mathbb{E}\left[X_j \mid Y\right] - q_j \mathbb{E}\left[X_{\mathrm{pa}(j)} \mid Y\right] \right)^2$$



$$C_j^1(\alpha, \tau, \delta) = \sigma_j^{-2} \left(\mathbb{E}\left[X_j \mid Y\right] - q_j \mathbb{E}\left[X_{\mathrm{pa}(j)} \mid Y\right] - s_j \sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k \right)^2$$



Algorithm:

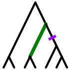
1. Find the K branches $j_1, \dots, j_K$ with largest C_j^0;
2. Allocate one change point in the first K branches;
3. For each of these branches, set $\delta_{j_k}^{(h+1)}$ so that $C_j^1(\tau, \delta) = 0$

M Step: Segmentation

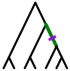
$$C_j(\alpha, \tau, \delta) = \sigma_j^{-2} \left(\mathbb{E}[X_j \mid Y] - q_j \mathbb{E}\left[X_{\text{pa}(j)} \mid Y\right] - r_j - s_j \sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k \right)^2$$

BM : $r_j = 0$, each cost is independent.

$$C_j^0(\alpha) = \sigma_j^{-2} \left(\mathbb{E}[X_j \mid Y] - q_j \mathbb{E}\left[X_{\text{pa}(j)} \mid Y\right] \right)^2$$



$$C_j^1(\alpha, \tau, \delta) = \sigma_j^{-2} \left(\mathbb{E}[X_j \mid Y] - q_j \mathbb{E}\left[X_{\text{pa}(j)} \mid Y\right] - s_j \sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k \right)^2$$



Algorithm:

1. Find the K branches $j_1, \ldots, j_K$ with largest C_j^0;
2. Allocate one change point in the first K branches;
3. For each of these branches, set $\delta_{j_k}^{(h+1)}$ so that $C_j^1(\tau, \delta) = 0$

M Step: Segmentation

$$C_j(\alpha, \tau, \delta) = \sigma_j^{-2} \left(\mathbb{E}\left[X_j \mid Y\right] - q_j \mathbb{E}\left[X_{\text{pa}(j)} \mid Y\right] - r_j - s_j \sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k \right)^2$$

OU : $r_j = \beta^{\text{pa}(j)}$, a cost depends on all its parents.

- Exact minimization: too costly.
- Need of an heuristic.
- Idea: rewrite as a least square:

$$\|D - AU\Delta\|^2$$

  with D a vector of size $n+m$, A a diagonal matrix of size $n+m$, Δ the vector of shifts and U the incidence matrix of the tree.
- Then use Stepwise selection or LASSO.

back

Goal and Notations

Data A process on a tree with the following structure:

$$\forall j > 1, \quad X_j | X_{\text{pa}(j)} \sim \mathcal{N}\left(m_j(X_{\text{pa}(j)}) = q_j X_{\text{pa}(j)} + r_j, \sigma_j^2\right)$$

$$\text{BM:} \begin{cases} q_j = 1 \\ r_j = \sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k \\ \sigma_j^2 = \ell_j \sigma^2 \end{cases} \qquad \text{OU:} \begin{cases} q_j = e^{-\alpha \ell_j} \\ r_j = \beta^{\text{pa}(j)}(1 - e^{-\alpha \ell_j}) + \sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k \left(1 - e^{-\alpha(1-\nu_k)\ell_j}\right) \\ \sigma_j^2 = \dfrac{\sigma^2}{2\alpha}(1 - e^{-2\alpha \ell_j}) \end{cases}$$

Goal Compute the following quantities, at every node j:

$$\mathbb{V}\text{ar}^{(h)}[Z_j \mid Y], \mathbb{C}\text{ov}^{(h)}\left[Z_j, Z_{\text{pa}(j)} \mid Y\right], \mathbb{E}^{(h)}[Z_j \mid Y]$$

Upward

Goal Compute for a vector of tips, given their common ancestor:

$$f_{\mathbf{Y}^j|X_j}(\mathbf{Y}^j; a) = A_j(\mathbf{Y}^j)\Phi_{M_j(\mathbf{Y}^j), S_j^2(\mathbf{Y}^j)}(a)$$

Initialization For tips: $f_{Y_i|Y_i}(Y_i; a) = \Phi_{Y_i,0}(a)$

Propagation

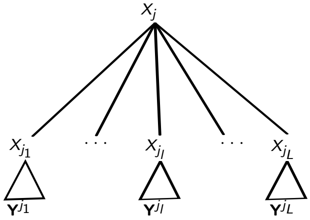
$$f_{\mathbf{Y}^j|X_j}(\mathbf{Y}^j; a) = \prod_{l=1}^{L} f_{\mathbf{Y}^{j_l}|X_j}(\mathbf{Y}^{j_l}; a)$$

$$f_{\mathbf{Y}^{j_l}|X_j}(\mathbf{Y}^{j_l}; a) = \int_{\mathbb{R}} f_{\mathbf{Y}^{j_l}|X_{j_l}}(\mathbf{Y}^{j_l}; b) f_{X_{j_l}|X_j}(b; a) db$$

Root Node and Likelihood At the root:

$$f_{X_1|\mathbf{Y}}(a; \mathbf{Y}) \propto f_{\mathbf{Y}|X_1}(\mathbf{Y}; a) f_{X_1}(a)$$

$$\begin{cases} \mathbb{V}\mathrm{ar}\left[X_1 \mid \mathbf{Y}\right] = \left(\dfrac{1}{\gamma^2} + \dfrac{1}{S_1^2(\mathbf{Y})}\right)^{-1} \\ \mathbb{E}\left[X_1 \mid \mathbf{Y}\right] = \mathbb{V}\mathrm{ar}\left[X_1 \mid \mathbf{Y}\right]\left(\dfrac{\mu}{\gamma^2} + \dfrac{M_1(\mathbf{Y})}{S_1^2(\mathbf{Y})}\right) \end{cases}$$

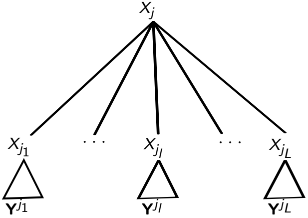


Downward

Compute $E_j = \mathbb{E}\left[X_j \mid \mathbf{Y}\right]$, $V_j^2 = \mathbb{V}\mathrm{ar}\left[X_j \mid \mathbf{Y}\right]$, $C_{j,\mathrm{pa}(j)}^2 = \mathbb{C}\mathrm{ov}\left[X_j; X_{\mathrm{pa}(j)} \mid \mathbf{Y}\right]$

Initialization Last step of Upward.

Propagation



$$f_{X_{\mathrm{pa}(j)},X_j|\mathbf{Y}}(a,b;\mathbf{Y}) = f_{X_{\mathrm{pa}(j)}|\mathbf{Y}}(a;\mathbf{Y}) f_{X_j|X_{\mathrm{pa}(j)},\mathbf{Y}}(b;a,\mathbf{Y})$$

$$f_{X_j|X_{\mathrm{pa}(j)},\mathbf{Y}}(b;a,\mathbf{Y}) = f_{X_j|X_{\mathrm{pa}(j)},\mathbf{Y}^j}(b;a,\mathbf{Y}^j)$$

$$\propto f_{X_j|X_{\mathrm{pa}(j)}}(b;a) f_{\mathbf{Y}^j|X_j}(\mathbf{Y}^j;b)$$

Formulas

Upward

$$\begin{cases} S_j^2(\mathbf{Y}^j) = \left(\sum_{l=1}^{L} \frac{q_{j_l}^2}{S_{j_l}^2(\mathbf{Y}^{j_l}) + \sigma_{j_l}^2} \right)^{-1} \\ M_j(\mathbf{Y}^j) = S_j^2(\mathbf{Y}^j) \sum_{l=1}^{L} q_{j_l} \frac{M_{j_l}(\mathbf{Y}^{j_l}) - r_{j_l}}{S_{j_l}^2(\mathbf{Y}^{j_l}) + \sigma_{j_l}^2} \end{cases}$$

Downward

$$\begin{cases} C_{j,\mathrm{pa}(j)}^2 = q_j \frac{S_j^2(\mathbf{Y}^j)}{S_j^2(\mathbf{Y}^j) + \sigma_j^2} V_{\mathrm{pa}(j)}^2 \\ E_j = \frac{S_j^2(\mathbf{Y}^j)(q_j E_{\mathrm{pa}(j)} + r_j) + \sigma_j^2 M_j(\mathbf{Y}^j)}{S_j^2(\mathbf{Y}^j) + \sigma_j^2} \\ V_j^2 = \frac{S_j^2(\mathbf{Y}^j)}{S_j^2(\mathbf{Y}^j) + \sigma_j^2} \left(\sigma_j^2 + p_j^2 \frac{S_j^2(\mathbf{Y}^j)}{S_j^2(\mathbf{Y}^j) + \sigma_j^2} V_{\mathrm{pa}(j)}^2 \right) \end{cases}$$

back

Model Selection on K

Assumption α fixed : design and structure of covariance fixed.

$$Y = TW(\alpha)\Delta + \gamma E = s + \gamma E \quad E \sim \mathcal{N}(0, V(\alpha))$$

Models $\mathcal{S} = \left\{ S_\eta = \mathrm{Span}(T_i, i \in \eta),\ \eta \in \mathcal{M} = \bigcup_{K=0}^{p-1} \mathcal{S}_K^{PI} \right\}$

$$\dim(S_\eta) = |\eta| = K_\eta + 1$$

Oracle $\inf_{\eta \in \mathcal{M}} \|s - s_\eta\|_V^2$ where $s_\eta = \mathrm{Proj}_{S_\eta}^V(s) = \underset{a \in S_\eta}{\mathrm{argmin}} \|s - a\|_V^2$

Estimators $\hat{s}_\eta = \mathrm{Proj}_{S_\eta}^V(Y)$, $\hat{s}_K = \underset{\eta \in \mathcal{S}, |\eta| = K+1}{\mathrm{argmin}} \|Y - \hat{s}_\eta\|_V^2$

Model Selection on K

Assumption α fixed : design and structure of covariance fixed.

$$Y = TW(\alpha)\Delta + \gamma E = s + \gamma E \quad E \sim \mathcal{N}(0, V(\alpha))$$

Models $$\mathcal{S} = \left\{ S_\eta = \text{Span}(T_i, i \in \eta),\ \eta \in \mathcal{M} = \bigcup_{K=0}^{p-1} \mathcal{S}_K^{PI} \right\}$$

$$\dim(S_\eta) = |\eta| = K_\eta + 1$$

Oracle $$\inf_{\eta \in \mathcal{M}} \|s - s_\eta\|_V^2 \text{ where } s_\eta = \text{Proj}_{S_\eta}^V(s) = \underset{a \in S_\eta}{\text{argmin}} \|s - a\|_V^2$$

Estimators $$\hat{s}_\eta = \text{Proj}_{S_\eta}^V(Y)\ ,\ \hat{s}_K = \underset{\eta \in \mathcal{S}, |\eta| = K+1}{\text{argmin}} \|Y - \hat{s}_\eta\|_V^2$$

Model Selection on K

Assumption α fixed : design and structure of covariance fixed.

$$Y = TW(\alpha)\Delta + \gamma E = s + \gamma E \quad E \sim \mathcal{N}(0, V(\alpha))$$

Models $\mathcal{S} = \left\{ S_\eta = \text{Span}(T_i, i \in \eta),\ \eta \in \mathcal{M} = \bigcup_{K=0}^{p-1} \mathcal{S}_K^{PI} \right\}$

$$\dim(S_\eta) = |\eta| = K_\eta + 1$$

Oracle $\inf_{\eta \in \mathcal{M}} \|s - s_\eta\|_V^2$ where $s_\eta = \text{Proj}_{S_\eta}^V(s) = \underset{a \in S_\eta}{\text{argmin}} \|s - a\|_V^2$

Estimators $\hat{s}_\eta = \text{Proj}_{S_\eta}^V(Y)$, $\hat{s}_K = \underset{\eta \in \mathcal{S}, |\eta| = K+1}{\text{argmin}} \|Y - \hat{s}_\eta\|_V^2$

Model Selection on K

Assumption α fixed : design and structure of covariance fixed.

$$Y = TW(\alpha)\Delta + \gamma E = s + \gamma E \quad E \sim \mathcal{N}(0, V(\alpha))$$

Models $\mathcal{S} = \left\{ S_\eta = \text{Span}(T_i, i \in \eta),\ \eta \in \mathcal{M} = \bigcup_{K=0}^{p-1} \mathcal{S}_K^{PI} \right\}$

$$\dim(S_\eta) = |\eta| = K_\eta + 1$$

Oracle $\inf_{\eta \in \mathcal{M}} \|s - s_\eta\|_V^2$ where $s_\eta = \text{Proj}_{S_\eta}^V(s) = \underset{a \in S_\eta}{\text{argmin}} \|s - a\|_V^2$

Estimators $\hat{s}_\eta = \text{Proj}_{S_\eta}^V(Y)$, $\hat{s}_K = \underset{\eta \in \mathcal{S}, |\eta| = K+1}{\text{argmin}} \|Y - \hat{s}_\eta\|_V^2$

Model Selection on K

Definition (Baraud et al. (2009))

Let D, $N > 0$, and $X_D \sim \chi^2(D)$, $X_N \sim \chi^2(N)$, $X_D \perp X_N$.

$$\text{Dkhi}[D, N, x] = \frac{1}{\mathbb{E}[X_D]} \mathbb{E}\left[\left(X_D - x\frac{X_N}{N}\right)_+\right], \quad \forall x > 0$$

$$\text{Dkhi}[D, N, \text{EDkhi}[D, N, q]] = q, \quad \forall 0 < q \leq 1$$

Proposition

Proposition (Form of the Penalty and guarantees (α known))

Under our setting: $Y = R\Delta + \gamma E$ *with* $E \sim \mathcal{N}(0, V)$*, define the penalty:*

$$\mathrm{pen}(K) = A\frac{n-K-1}{n-K-2}\,\mathrm{EDkhi}[K+2, n-K-2, e^{-L_K}]$$

$$\textit{with } L_K = \log\left|\mathcal{S}_K^{PI}\right| + 2\log(K+2)$$

If $\kappa < 1$*, and* $p \leq \min\left(\frac{\kappa n}{2+\log(2)+\log(n)}, n-7\right)$*, we get:*

$$\mathbb{E}\left[\frac{\left\|s-\hat{s}_{\hat{K}}\right\|_V^2}{\gamma^2}\right] \leq C(A,\kappa)\inf_{\eta\in\mathcal{M}}\left\{\frac{\left\|s-s_\eta\right\|_V^2}{\gamma^2} + (K_\eta+2)(3+\log(n))\right\}$$

with $C(A,\kappa)$ *a constant depending on* A *and* κ *only.*

Based on Baraud et al. (2009) + Back

Model Selection with Unknown Variance

Theorem (Baraud et al. (2009))

Under the following setting:

$$Y' = s' + \gamma E' \quad \textit{with} \quad E' \sim \mathcal{N}(0, I_n) \quad \textit{and} \quad \mathcal{S}' = \{S'_\eta, \eta \in \mathcal{M}\}$$

If $D_\eta = \dim(S'_\eta)$, $N_\eta = n - D_\eta \geq 7$, $\max(L_\eta, D_\eta) \leq \kappa n$, *with* $\kappa < 1$, *and:*

$$\Omega' = \sum_{\eta \in \mathcal{M}} (D_\eta + 1) e^{-L_\eta} < +\infty$$

$$\textit{If:} \quad \hat{\eta} = \underset{\eta \in \mathcal{M}}{\operatorname{argmin}} \left\| Y' - \hat{s}'_\eta \right\|^2 \left(1 + \frac{\operatorname{pen}(\eta)}{N_\eta}\right)$$

$$\textit{with:} \quad \operatorname{pen}(\eta) = \operatorname{pen}_{A,\mathcal{L}}(\eta) = A \frac{N_\eta}{N_\eta - 1} \operatorname{EDkhi}[D_\eta + 1, N_\eta - 1, e^{-L_\eta}] \quad , \quad A > 1$$

$$\textit{Then:} \quad \mathbb{E}\left[\frac{\left\| s' - \hat{s}'_{\hat{\eta}} \right\|^2}{\gamma^2}\right] \leq C(A, \kappa) \left[\inf_{\eta \in \mathcal{M}} \left\{ \frac{\left\| s' - s'_\eta \right\|^2}{\gamma^2} + \max(L_\eta, D_\eta) \right\} + \Omega' \right]$$

IID Framework ($\alpha = 0$)

$$\text{Assume} \quad K_\eta = D_\eta - 1 \leq p - 1 \leq n - 8, \quad \forall \eta \in \mathcal{M}$$

Then:

$$\begin{aligned}
\Omega' &= \sum_{\eta \in \mathcal{M}} (D_\eta + 1) e^{-L_\eta} = \sum_{\eta \in \mathcal{M}} (K_\eta + 2) e^{-L_\eta} \\
&= \sum_{K=0}^{p-1} \left| \mathcal{S}_K^{PI} \right| (K+2) e^{-L_K} = \sum_{K=0}^{p-1} \left| \mathcal{S}_K^{PI} \right| (K+2) e^{-(\log \left| \mathcal{S}_K^{PI} \right| + 2\log(K+2))} \\
&= \sum_{K=0}^{p-1} \frac{1}{K+2} \leq \log(p) \leq \log(n)
\end{aligned}$$

And:

$$L_K \leq \log \binom{n+m-1}{K} + 2\log(K+2) \leq K\log(n+m-1) + 2(K+1) \leq p(2 + \log(2n-2))$$

Hence, if $p \leq \min\left(\frac{\kappa n}{2+\log(2)+\log(n)}, n-7\right)$, then $\max(L_\eta, D_\eta) \leq \kappa n$ for any $\eta \in \mathcal{M}$.

Non-IID Framework ($\alpha \neq 0$)

Cholesky decomposition: $V = LL^T \quad Y' = L^{-1}Y \quad s' = L^{-1}s \quad E' = L^{-1}E$

$$Y' = s' + \gamma E', \text{ with: } E' \sim \mathcal{N}(0, I_n)$$

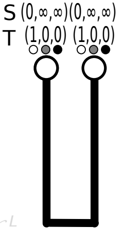
$$S'_\eta = L^{-1}S_\eta, \quad \hat{s}'_\eta = \text{Proj}_{S'_\eta} Y' = \underset{a' \in S'_\eta}{\text{argmin}} \left\| Y - La' \right\|_V^2 = L^{-1}\hat{s}_\eta$$

$$\|s - \hat{s}_{\hat{\eta}}\|_V^2 = \left\|s' - \hat{s}'_{\hat{\eta}}\right\|^2, \quad \|Y - \hat{s}_\eta\|_V^2 = \left\|Y' - \hat{s}'_\eta\right\|^2$$

$$\text{Crit}_{MC}(\eta) = \left\|Y' - \hat{s}'_\eta\right\|^2 \left(1 + \frac{\text{pen}_{A,\mathcal{L}}(\eta)}{N_\eta}\right) = \|Y - \hat{s}_\eta\|_V^2 \left(1 + \frac{\text{pen}_{A,\mathcal{L}}(\eta)}{N_\eta}\right)$$

back

Cardinal of Equivalence Classes



Initialization For tips

Propagation

$$\mathcal{K}_k^l = \underset{1 \leq p \leq K}{\operatorname{argmin}} \{S_{i_l}(p) + \mathbb{I}\{p \neq k\}\}$$

$$S_i(k) = \sum_{l=1}^{L} S_{i_l}(p_l) + \mathbb{I}\{p_l \neq k\} \ , \ \forall (p_1, \ldots p_L) \in \mathcal{K}_k^1 \times \ldots \times \mathcal{K}_k^L$$

$$T_i(k) = \sum_{(p_1, \ldots p_L) \in \mathcal{K}_k^1 \times \ldots \times \mathcal{K}_k^L} \prod_{l=1}^{L} T_{i_l}(p_l) = \prod_{l=1}^{L} \sum_{p_l \in \mathcal{K}_k^l} T_{i_l}(p_l)$$

Termination Sum on the root vector

back

Cardinal of Equivalence Classes

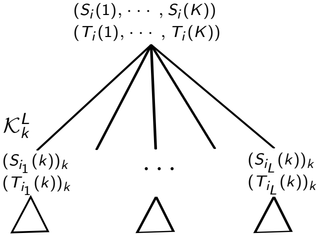
Initialization For tips

Propagation

$$\mathcal{K}_k^l = \underset{1 \leq p \leq K}{\operatorname{argmin}} \left\{ S_{i_l}(p) + \mathbb{I}\{p \neq k\} \right\}$$

$$S_i(k) = \sum_{l=1}^{L} S_{i_l}(p_l) + \mathbb{I}\{p_l \neq k\} \ , \ \forall (p_1, \dots p_L) \in \mathcal{K}_k^1 \times \dots \times \mathcal{K}_k^L$$

$$T_i(k) = \sum_{(p_1, \dots p_L) \in \mathcal{K}_k^1 \times \dots \times \mathcal{K}_k^L} \prod_{l=1}^{L} T_{i_l}(p_l) = \prod_{l=1}^{L} \sum_{p_l \in \mathcal{K}_k^l} T_{i_l}(p_l)$$



Termination Sum on the root vector

back

Cardinal of Equivalence Classes

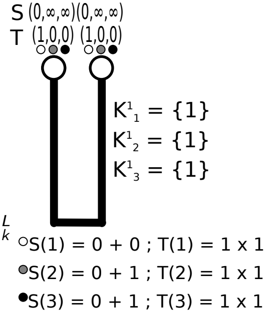
Initialization For tips

Propagation

$$\mathcal{K}_k^l = \underset{1 \leq p \leq K}{\text{argmin}} \left\{ S_{i_l}(p) + \mathbb{I}\{p \neq k\} \right\}$$

$$S_i(k) = \sum_{l=1}^{L} S_{i_l}(p_l) + \mathbb{I}\{p_l \neq k\} \ , \ \forall (p_1, \dots p_L) \in \mathcal{K}_k^1 \times \dots \times \mathcal{K}_k^L$$

$$T_i(k) = \sum_{(p_1, \dots p_L) \in \mathcal{K}_k^1 \times \dots \times \mathcal{K}_k^L} \prod_{l=1}^{L} T_{i_l}(p_l) = \prod_{l=1}^{L} \sum_{p_l \in \mathcal{K}_k^l} T_{i_l}(p_l)$$



Termination Sum on the root vector

back

Cardinal of Equivalence Classes

Initialization For tips

Propagation

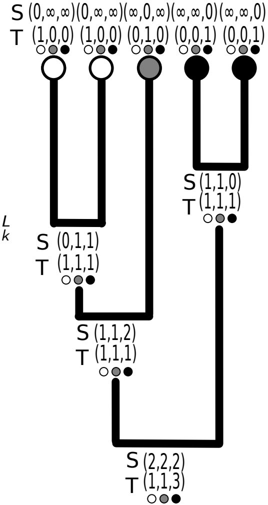
$$\mathcal{K}_k^l = \underset{1 \le p \le K}{\operatorname{argmin}} \left\{ S_{i_l}(p) + \mathbb{I}\{p \neq k\} \right\}$$

$$S_i(k) = \sum_{l=1}^{L} S_{i_l}(p_l) + \mathbb{I}\{p_l \neq k\} \ , \ \forall (p_1, \dots p_L) \in \mathcal{K}_k^1 \times \dots \times \mathcal{K}_k^L$$

$$T_i(k) = \sum_{(p_1,\dots p_L) \in \mathcal{K}_k^1 \times \dots \times \mathcal{K}_k^L} \prod_{l=1}^{L} T_{i_l}(p_l) = \prod_{l=1}^{L} \sum_{p_l \in \mathcal{K}_k^l} T_{i_l}(p_l)$$

Termination Sum on the root vector

back



Surjection :

$$\phi : \mathcal{S}_K^P \to \mathcal{C}_{K+1}$$

$$\mathcal{S}_K^P = \{\text{Parsimonious allocations of } K \text{ shifts}\}$$
$$\mathcal{C}_{K+1} = \{\text{Tree compatible clustering of tips in } K+1 \text{ groups}\}$$

Equivalence Relation :

$$\forall s_1, s_1 \in \mathcal{S}_K^P, s_1 \sim s_2 \iff \phi(s_1) = \phi(s_2)$$

Quotient Set :

$$\mathcal{S}_K^{PI} = \mathcal{S}_K^P / \sim \quad \text{gives} \quad \mathcal{S}_K^{PI} \underset{\sim}{\to} \mathcal{C}_{K+1}$$

$$\mathcal{S}_K^{PI} = \{\text{Parsimonious } \textit{identifiable} \text{ allocations of } K \text{ shifts}\}$$

Back

Surjection :

$$\phi : \mathcal{S}_K^P \to \mathcal{C}_{K+1}$$

$$\mathcal{S}_K^P = \{\text{Parsimonious allocations of } K \text{ shifts}\}$$

$$\mathcal{C}_{K+1} = \{\text{Tree compatible clustering of tips in } K+1 \text{ groups}\}$$

Equivalence Relation :

$$\forall s_1, s_1 \in \mathcal{S}_K^P, s_1 \sim s_2 \iff \phi(s_1) = \phi(s_2)$$

Quotient Set :

$$\mathcal{S}_K^{PI} = \mathcal{S}_K^P / \sim \quad \text{gives} \quad \mathcal{S}_K^{PI} \underset{\sim}{\to} \mathcal{C}_{K+1}$$

$$\mathcal{S}_K^{PI} = \{\text{Parsimonious } \textit{identifiable} \text{ allocations of } K \text{ shifts}\}$$

Back

Surjection :

$$\phi : \mathcal{S}_K^P \to \mathcal{C}_{K+1}$$

$$\mathcal{S}_K^P = \{\text{Parsimonious allocations of } K \text{ shifts}\}$$
$$\mathcal{C}_{K+1} = \{\text{Tree compatible clustering of tips in } K+1 \text{ groups}\}$$

Equivalence Relation :

$$\forall s_1, s_1 \in \mathcal{S}_K^P, s_1 \sim s_2 \iff \phi(s_1) = \phi(s_2)$$

Quotient Set :

$$\mathcal{S}_K^{PI} = \mathcal{S}_K^P / \sim \quad \text{gives} \quad \mathcal{S}_K^{PI} \underset{\sim}{\to} \mathcal{C}_{K+1}$$

$$\mathcal{S}_K^{PI} = \{\text{Parsimonious } \textit{identifiable} \text{ allocations of } K \text{ shifts}\}$$

Back

Linking Shifts and Clustering

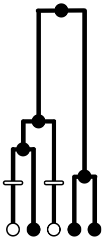
Assumption "No Homoplasy": 1 shift = 1 new color

Proposition "K shifts $\iff$ $K+1$ clusters"

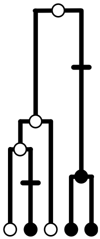
back

Linking Shifts and Clustering

Assumption "No Homoplasy": 1 shift = 1 new color



$\sim$



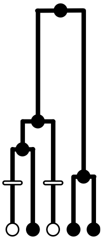
The No Homoplasy hypothesis is not respected.

Proposition "K shifts $\iff$ $K+1$ clusters"

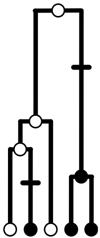
back

Linking Shifts and Clustering

Assumption "No Homoplasy": 1 shift = 1 new color



$\sim$



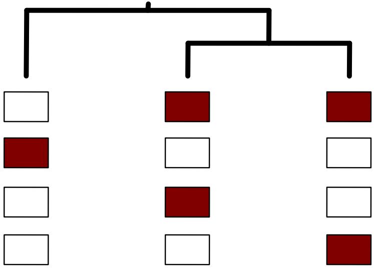
The No Homoplasy hypothesis is not respected.

Proposition "K shifts $\iff$ $K + 1$ clusters"

back

Definitions

- $\mathcal{T}$ a rooted tree with n tips
- $N_K^{(\mathcal{T})} = |\mathcal{C}_K|$ the number of possible partitions of the tips in K clusters
- $A_K^{(\mathcal{T})}$ the number of possible *marked* partitions



Partitions in two groups for a binary tree with 3 tips

Difference between $N_2^{(\mathcal{T}_3)}$ and $A_2^{(\mathcal{T}_3)}$:

- $N_2^{(\mathcal{T}_3)} = 3$: partitions 1 and 2 are equivalent
- $A_2^{(\mathcal{T}_3)} = 4$: one marked color ("white = ancestral state")

General Formula (Binary Case)

If $\mathcal{T}$ is a binary tree, consider T_ℓ and $\mathcal{T}_r$ the left and right sub-trees of $\mathcal{T}$. Then:

$$\begin{cases} N_K^{(\mathcal{T})} = \sum_{k_1+k_2=K} N_{k_1}^{(\mathcal{T}_\ell)} N_{k_2}^{(\mathcal{T}_r)} + \sum_{k_1+k_2=K+1} A_{k_1}^{(\mathcal{T}_\ell)} A_{k_2}^{(\mathcal{T}_r)} \\ A_K^{(\mathcal{T})} = \sum_{k_1+k_2=K} A_{k_1}^{(\mathcal{T}_\ell)} N_{k_2}^{(\mathcal{T}_r)} + N_{k_1}^{(\mathcal{T}_\ell)} A_{k_2}^{(\mathcal{T}_r)} + \sum_{k_1+k_2=K+1} A_{k_1}^{(\mathcal{T}_\ell)} A_{k_2}^{(\mathcal{T}_r)} \end{cases}$$

We get:

$$N_{K+1}^{(\mathcal{T})} = N_{K+1}^{(n)} = \binom{2n-2-K}{K} \quad \text{and} \quad A_{K+1}^{(\mathcal{T})} = A_{K+1}^{(n)} = \binom{2n-1-K}{K}$$

Recursion Formula (General Case)

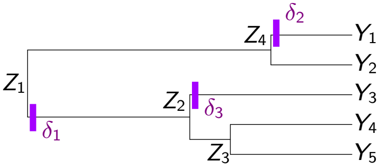
If we are at a node defining a tree $\mathcal{T}$ that has p daughters, with sub-trees $\mathcal{T}_1, \dots, \mathcal{T}_p$, then we get the following recursion formulas:

$$\begin{cases} N_K^{(\mathcal{T})} = \displaystyle\sum_{\substack{k_1+\dots+k_p=K \\ k_1,\dots,k_p \geq 1}} \prod_{i=1}^{p} N_{k_i}^{(\mathcal{T}_i)} + \sum_{\substack{I \subset [\![1,p]\!] \\ |I| \geq 2}} \sum_{\substack{k_1+\dots+k_p=K+|I|-1 \\ k_1,\dots,k_p \geq 1}} \prod_{i \in I} A_{k_i}^{(\mathcal{T}_i)} \prod_{i \notin I} N_{k_i}^{(\mathcal{T}_i)} \\ A_K^{(\mathcal{T})} = \displaystyle\sum_{\substack{I \subset [\![1,p]\!] \\ |I| \geq 1}} \sum_{\substack{k_1+\dots+k_p=K+|I|-1 \\ k_1,\dots,k_p \geq 1}} \prod_{i \in I} A_{k_i}^{(\mathcal{T}_i)} \prod_{i \notin I} N_{k_i}^{(\mathcal{T}_i)} \end{cases}$$

No general formula. The result depends on the topology of the tree.

back

Linear Regression Model

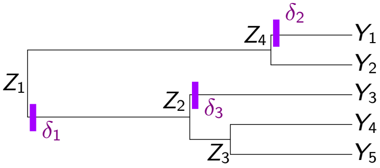


$$\Delta = \begin{pmatrix} \mu \\ \delta_1 \\ 0 \\ 0 \\ \delta_2 \\ 0 \\ \delta_3 \\ 0 \\ 0 \end{pmatrix} \qquad T\Delta = \begin{pmatrix} \mu + \delta_2 \\ \mu \\ \mu + \delta_1 + \delta_3 \\ \mu + \delta_1 \\ \mu + \delta_1 \end{pmatrix}$$

$$T = \begin{array}{c} \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{array} \begin{array}{c} \begin{array}{ccccccccc} Z_1 & Z_2 & Z_3 & Z_4 & Y_1 & Y_2 & Y_3 & Y_4 & Y_5 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

$$BM : \quad Y = T\Delta^{BM} + E^{BM}$$

Linear Regression Model



$$\Delta = \begin{pmatrix} \lambda \\ \delta_1 \\ 0 \\ 0 \\ \delta_2 \\ 0 \\ \delta_3 \\ 0 \\ 0 \end{pmatrix} \qquad TW(\alpha)\Delta = \begin{pmatrix} \lambda + w_5\delta_2 \\ \lambda \\ \lambda + w_2\delta_1 + w_7\delta_3 \\ \lambda + w_2\delta_1 \\ \lambda + w_2\delta_1 \end{pmatrix}$$

$$W(\alpha) = \mathrm{Diag}(1 - e^{-\alpha(h - t_{\mathrm{pa}(i)})}, 1 \leq i \leq m+n)$$

$$\lambda = \mu e^{-\alpha h} + \beta_0(1 - e^{-\alpha h})$$

$$BM: \quad Y = T\Delta^{BM} + E^{BM}$$

$$OU: \quad Y = TW(\alpha)\Delta^{OU} + E^{OU}$$

OUfun model and equivalence with BM

Root Fixed on an Ultrametric tree, shifts at Nodes.

Expectations

$$\mathbb{E}\left[Y \mid X_1 = \mu\right] = T \underbrace{W(\alpha)\Delta^{OU}}_{\Delta^{BM}}$$

Rq: $\mu^{BM} = \lambda^{OU} = \mu e^{-\alpha h} + \beta_0(1 - e^{-\alpha h})$

Variance

$$\mathbb{C}\text{ov}\left[Y_i ; Y_j \mid X_1 = \mu\right] = \sigma^2 \times \underbrace{\frac{1}{2\alpha} e^{-2\alpha h}(e^{2\alpha t_{ij} - 1})}_{t'_{ij}}$$

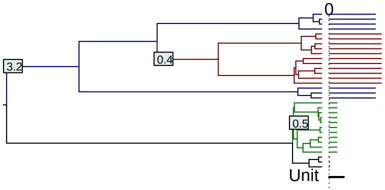
OUfun $\iff$ BM on a re-scaled tree with $t' = e^{-2\alpha h}(e^{2\alpha t} - 1)$

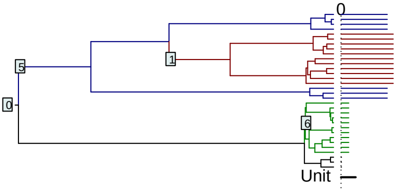
back

Coloring and Process

Definition (Tips Coloring)

Two tips have the same color if they have the same mean under the process studied.



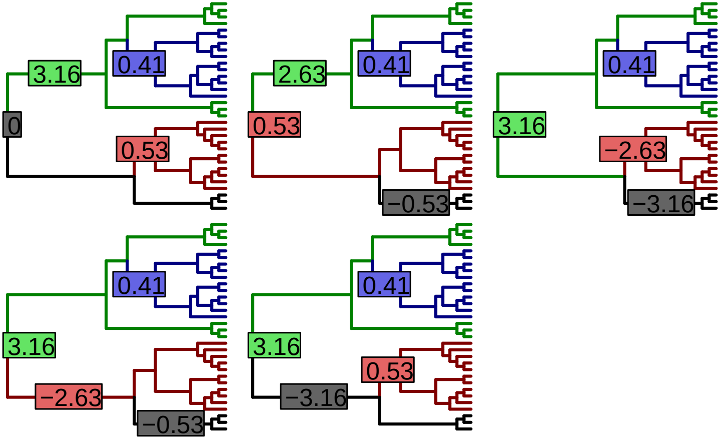


$BM \quad m_Y = T\Delta^{BM}$

$OU \quad m_Y = T\underbrace{W(\alpha)\Delta^{OU}}_{\Delta^{BM}}$

back

Equivalent Parsimonious Solutions for a BM Model.



Equivalent allocations and values of the shifts - BM.

back

Simulations Design

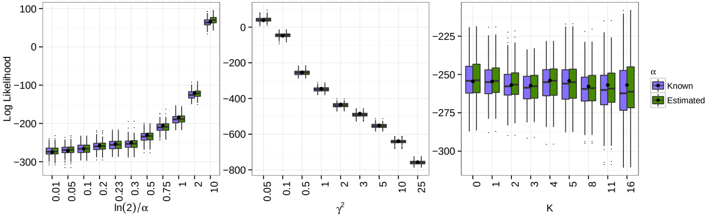
(Uyeda and Harmon, 2014)

- Topology of the tree fixed (unit height, $\lambda = 0.1$, with 64, 128, 256 taxa).
- Initial optimal value fixed: $\beta_0 = 0$
- One "base" scenario $\alpha_b = 3$, $\gamma_b^2 = 0.5$, $K_b = 5$.
- $\alpha \in \log(2)/\{0.01, 0.05, 0.1, 0.2, 0.23, 0.3, 0.5, 0.75, 1, 2, 10\}$.
- $\gamma^2 \in \{0.3, 0.6, 3, 6, 12, 18, 30, 60, 150\}/(2\alpha_b)$.
- $K \in \{0, 1, 2, 3, 4, 5, 8, 11, 16\}$.
- Shifts values $\sim \frac{1}{2}\mathcal{N}(4, 1) + \frac{1}{2}\mathcal{N}(-4, 1)$
- Shifts randomly placed at regular intervals separated by 0.1 unit length.
- $n = 200$ repetitions : 16200 configurations.

CPU time on cluster MIGALE (Jouy-en-Josas):

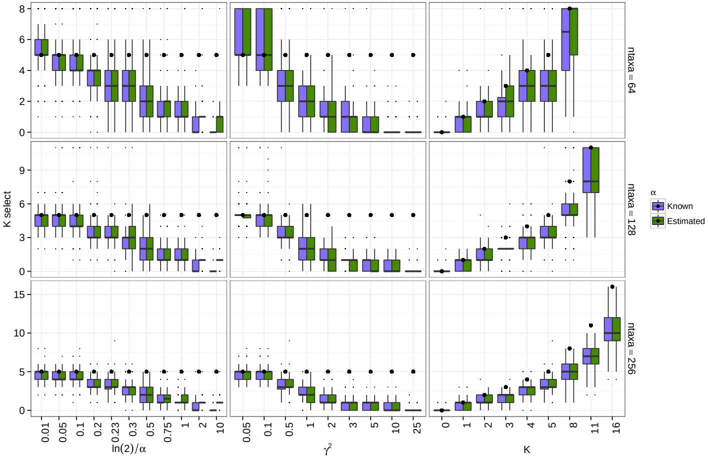
- α known: 66 days (6 minutes per estimation).
- α unknown: 570 days (52 minutes per estimation).

Log-Likelihood

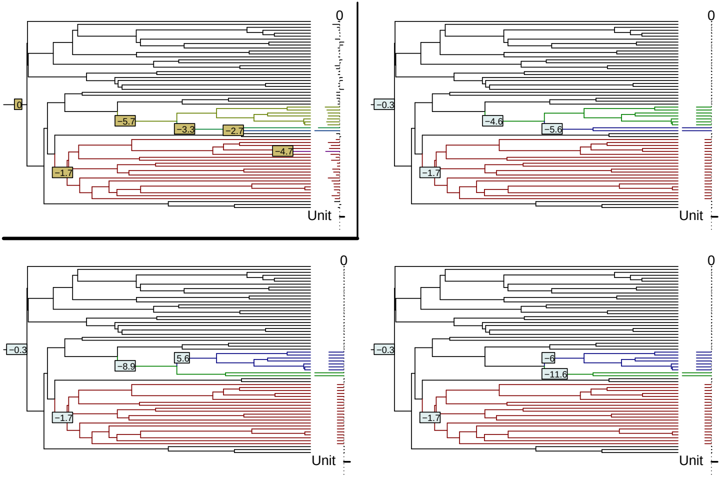


Log likelihood for a tree with 256 tips. Solid black dots are the median of the log likelihood for the true parameters.

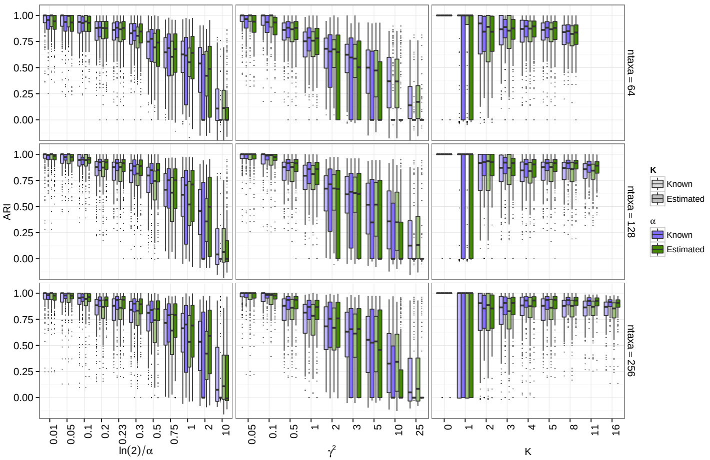
Number of Shifts



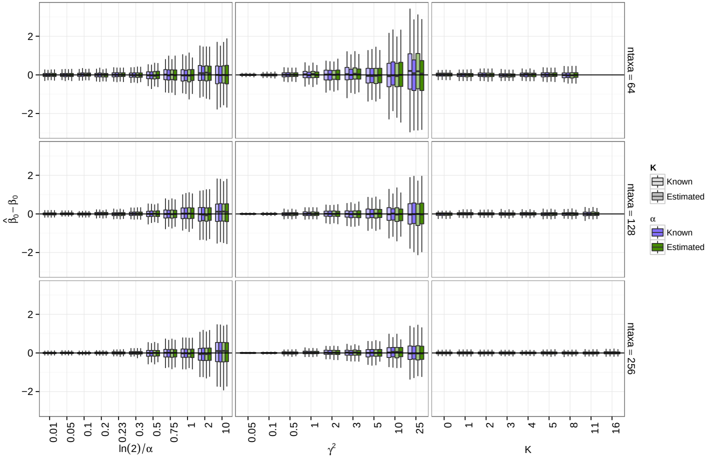
One Example


0
−5.7
−3.3
−2.7
−4.7
−1.7
Unit
0
−0.3
−4.6
−5.6
−1.7
Unit
0
−0.3
−8.9
5.6
−1.7
Unit
0
−0.3
−6
−11.6
−1.7
Unit

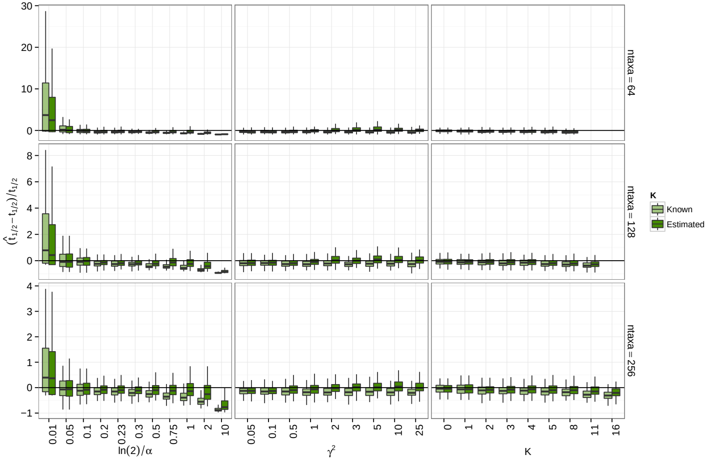
Adjusted Rand Index



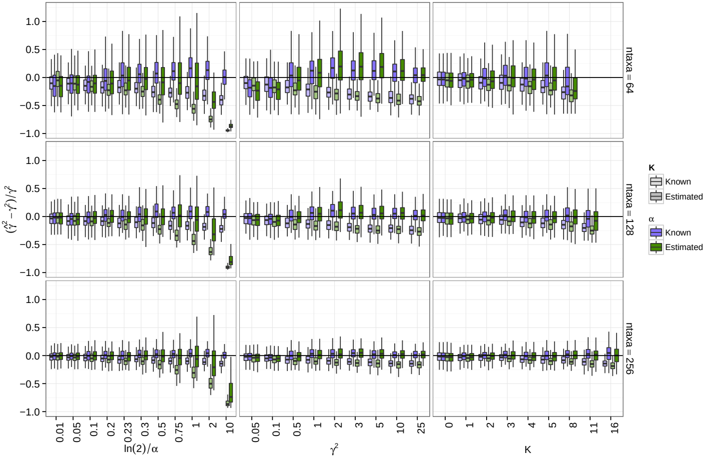
Parameters


2
0
−2
$\hat{\beta}_0 - \beta_0$
ntaxa = 64
ntaxa = 128
ntaxa = 256
K
Known
Estimated
α
Known
Estimated
0.01
0.05
0.1
0.2
0.23
0.3
0.5
0.75
1
2
10
ln(2)/α
0.05
0.1
0.5
1
2
3
5
10
25
γ^2
0
1
2
3
4
5
8
11
16
K

Parameters


$(\hat{t}_{1/2}-t_{1/2})/t_{1/2}$
ntaxa = 64
ntaxa = 128
ntaxa = 256
K
Known
Estimated
ln(2)/α
γ^2
K
0.01
0.05
0.1
0.2
0.23
0.3
0.5
0.75
1
2
10
0.05
0.1
0.5
1
2
3
5
10
25
0
1
2
3
4
5
8
11
16

Parameters


ntaxa = 64
ntaxa = 128
ntaxa = 256
K
Known
Estimated
α
Known
Estimated
ln(2)/α
γ²
K