

Package: extendedSurface (via r-universe)

August 12, 2024

Type Package

Title Fit Complex Multi-OU Models by Extending SURFACE and Providing an Interface with OUwie

Version 0.1.0

Description Multi Ornstein Uhlenbeck (OUM) models are routinely employed to describe the phenotypic evolution of clades across a macroevolutionary adaptive landscape. Multiple implementations exist, including those that require an a priori assignment of lineages to adaptive regimes (e.g., OUwie), and those that infer the location of such regimes shifts (e.g., SURFACE). However, SURFACE has been found to favor overly complex models, which is likely a consequence of fixing a single rate of evolution and force of attraction for the entire tree. Although these parameters can be optimized for each regime by using the SURFACE output as the set of regime shifts required by OUwie, such complex models often fail to be optimized. extendedSurface provides a solution by continuing the backwards phase of SURFACE, merging regimes and generating simpler (yet suboptimal) multiOU models. These are then fed to OUwie, where parameters for each regime can be independently optimized. Using simpler OUM models as starting points improves the probability of successful model fitting. The resulting models are more realistic, and have been found to be favored over others using empirical comparative datasets. The package allows the user to decide which models should be fit (OUMA, OUMV, OUMVA, with and without stationary root state), and plots the fit of these against those explored by SURFACE to easily compare model fit. Additionally, several approaches to record and plot the evolutionary dynamics of discrete traits through time are also implemented.

URL <http://github.com/mongiardino/extendedSurface>

BugReports <https://github.com/mongiardino/extendedSurface/issues>

License GPL (>= 2)

Encoding UTF-8

LazyData true

Depends R (>= 2.10)

Imports OUwie (>= 2.5), surface (>= 0.4.1), phytools (>= 0.6.99), ape (>= 5.3), phangorn (>= 2.5.5), magrittr (>= 1.5), dplyr (>= 1.0.2), ggplot2 (>= 3.2.1), cowplot (>= 1.0.0)

RoxygenNote 7.1.0

Repository <https://phylotastic.r-universe.dev>

RemoteUrl <https://github.com/mongiardino/extendedSurface>

RemoteRef HEAD

RemoteSha de922e5de538cb2cc0ad28ab5ce48ff964191444

Contents

echinoid_data	2
extended_surfaceAICPlot	3
plot_ttt	4
surfaceExtended	5
transitions_through_time	8

Index **10**

echinoid_data	<i>Comparative dataset used to explore the macroevolution of echinoid body size.</i>
---------------	--

Description

A list containing a data.set of body sizes, a vector of measurement errors, a time-calibrated phylogeny, and the results of running forward and backward phases of SURFACE on this dataset.

Usage

```
echinoid_data
```

Format

A list with all the data to replicate the macroevolutionary analysis in Mongiardino Koch & Thompson (2020):

size A data.frame including body sizes for all terminals and terminal names as row names

error A named vector with measurement errors, in the same order as \$size

tree A time-calibrated phylogeny of all taxa with size data

fwd_surface The result of running surfaceForward on this dataset

bwd_surface The result of running surfaceBackward on this dataset

References

Mongiardino Koch N. 2021. Exploring adaptive landscapes across deep time: A case study using echinoid body size. *Evolution*, <https://doi.org/10.1111/evo.14219>.

extended_surfaceAICPlot

Plot AICc values of multi-OU models

Description

Explore model fit by visually comparing the AICc values of the results obtained using `surfaceExtended`. The fit of OUM models obtained using forward, backwards and extended phases of SURFACE, as well as that of models explored with OUwie, are plotted against the number of regimes.

Usage

```
extended_surfaceAICPlot(ext_surface, summary, fwd_surface = NA)
```

Arguments

<code>ext_surface</code>	List of models obtained using the extended phase of SURFACE. This is always the first element in the list returned by <code>surfaceExtended</code> .
<code>summary</code>	A <code>data.frame</code> with AICc values of OUwie models. This is always the second element in the list returned by <code>surfaceExtended</code> .
<code>fwd_surface</code>	Optional. List of models obtained using the forward phase of SURFACE. If provided these will also be included in the plot.

Value

Plot of AICc values against number of regimes.

See Also

[surfaceExtended](#)

plot_ttt

*Plot trait dynamics through time***Description**

This function plots evolutionary dynamics of discrete traits as summarized using [transitions_through_time](#). Two types of plots can be generated: the number of different states active through time (where identical states that have different origins are not counted as the same one), and the rates at which these states are originating (birth rate), becoming extinct (death rate), or accumulating (diversification rate) through time. Depending on the character being investigated, these plots might or might not be meaningful.

Usage

```
plot_ttt(
  ttt,
  interval = NA,
  window_size = NA,
  CI = 80,
  trim = T,
  graphs = "both",
  rates = "diversification",
  k = NA
)
```

Arguments

ttt	Data.frame output by transitions_through_time .
interval	Numeric value that determines the temporal resolution (i.e., the step size in Ma at which the number of active regimes is recorded). Defaults to slightly less than 100 given the time spanned by the phylogeny.
window_size	Numeric value that sets the width of the window used to smoot rate estimates. Defaults to a width that includes approx. 10 intervals (see above).
CI	Numeric value that sets the confidence interval (expressed as percentage). Determines the amount of results that are discarded before plotting (default = 80).
trim	Whether to trim a few values at the beginning and end of plot that contain fewer intervals and can be noisier. Default is TRUE.
graphs	Which graphs to plot. Options include 'active_regimes', 'rate_ttt', and 'both'.
rates	Which rates to plot. Options include any combination of 'birth', 'death', and 'diversification'. Defaults to only the latter.
k	The value of k used for gam regression. If not specified this is automatically determined (see more details in gam).

Details

By default, this is used by `transitions_through_time` to plot results. However, the object returned by that function can also be used here with more control on the plotting options. These include the intervals (in Ma) at which the number of states are recorded, the size of the window used to smooth rates, the type of plot generated, and the type of rate to plot (see Arguments).

Trends are depicted using GAM regressions (see `gam`). Depending on the combination of the size of the smoothing window and the number of smoothing functions used, nonsensical results can be obtained. Some tuning might be necessary to correctly depict trends in the data.

Value

A plot including different visual summaries of the evolutionary dynamics of discrete traits through time.

Author(s)

Nicolás Mongiardino Koch

References

Mongiardino Koch N. 2021. Exploring adaptive landscapes across deep time: A case study using echinoid body size. *Evolution*, <https://doi.org/10.1111/evo.14219>.

See Also

`transitions_through_time`

surfaceExtended

Interface between SURFACE and OUwie to fit complex multi-OU models

Description

surfaceExtended uses the output of a SURFACE run to attempt to fit more complex Ornstein-Uhlenbeck models using OUwie. These models can incorporate differences between regimes in the rate of evolution (σ^2), strength of attraction to trait optima (α), as well as optimize the state at the root of the tree (θ_0). Optionally, the backwards phase of SURFACE is then extended by further merging regimes and attempting to use these simpler models as successive inputs to OUwie. Only one morphological trait can be employed. The function keeps track of successful instances of model fitting, and returns the models obtained through both SURFACE and OUwie, as well as an overall summary.

Usage

```

surfaceExtended(
  bwd_surface,
  data,
  tree,
  error = NA,
  models = c("OUMVA", "OUMVAZ"),
  limit = 2,
  plot = T,
  fwd_surface = NA
)

```

Arguments

bwd_surface	List of models obtained using the backwards phase of SURFACE.
data	The morphological character under investigation. Needs to be the same data.frame used to run SURFACE.
tree	Phylogenetic tree in 'phylo' format. Needs to be the same used to run SURFACE.
error	Optional. Measurement errors to be incorporated in the process of model fitting. Can be a data.frame or vector, but it is assumed the order matches that of data.
models	Character vector specifying the models to be explored (see Details). Explores 'OUMVA' and 'OUMVAZ' models by default.
limit	Minimum number of regimes to explore. The default is 2.
plot	A logical indicating whether to plot the AICc of models output by SURFACE and OUwie. Default is TRUE.
fwd_surface	Optional. List of models obtained using the backwards phase of SURFACE. Only used to produce a more thorough comparison of models when plot=TRUE.

Details

Paleontological data has been shown to improve the accuracy of models describing morphological evolution using OU models (Ho & Ané 2014). Nonetheless, many of the methods to fit multi-OU models that do not require users to specify the number and location of regime shifts work only on ultrametric trees. One that does not, SURFACE (Ingram & Mahler 2013), tends to favor overly complex models (Khabbazian et al. 2016), likely a consequence of assuming that regimes share a common σ^2 and α parameters (Mongiardino Koch & Thompson 2020). Relaxing this assumption is not straightforward, as estimating these parameters for models with multiple regimes is often unfeasible (Benson et al. 2017).

surfaceExtended employs the optimal model found by the SURFACE algorithm and uses OUwie (Beaulieu et al. 2012) to attempt to fit multi-OU models in which rates of evolution and strengths of selection vary between regimes. The function then extends the backwards phase of SURFACE to merge independent regimes and find simpler multi-OU models. This is done in a stepwise fashion, and every time two regimes are merged, the result is used as input for OUwie.

The user can specify which parameters to estimate for each regime with OUwie, including different rates of evolution (models = 'OUMV'), different strengths of selection (models = 'OUMA'), or

both (`models = 'OUMVA'`). The state at the root of the tree can be further considered an independent parameter by adding 'Z' at the end of the model's name (e.g., `models = 'OUMAZ'`), although this can destabilize parameter estimates (see [OUwie](#) for more details). Multiple models can be explored simultaneously by providing a vector with their names, or using `models = 'all'` or `models = 'all_noZ'`. In the latter, only models assuming the root value is distributed according to the stationary distribution of the ancestral OU process are optimized. The minimum number of regimes to be explored is determined by the `limit` parameter. By default, the results will be plotted using [extended_surfaceAICPlot](#).

Can be very time-consuming depending on the size of the phylogeny and the complexity (number of regimes) of the starting model.

Value

A list with the following elements:

\$ext_surface A list containing all the models explored by extending the backwards phase of SURFACE, identical to the one returned by [surfaceBackward](#).

\$summary A data.frame including information on all the models explored using OUwie, including whether model fit was successful, and if so the AICc value.

Additionally, if model fit was successful, the list will also include the best option found for each of the models specified with `models`. If multiple models were explored, the best option for each will be returned. For example, if OUMVA and OUMVAZ models were explored (as is the default), and model fitting was successful, the returned list will also contain two more elements, 'best_OUMVA' and 'best_OUMVAZ'. Note that these might differ in the number of regimes they contain.

Author(s)

Nicolás Mongiardino Koch

References

Beaulieu J.M., Jhuweng D.-C., Boettiger C., O'Meara B.C. 2012. Modeling stabilizing selection: expanding the Ornstein–Uhlenbeck model of adaptive evolution. *Evolution*, 66:2369–2383. Benson R.B.J., Hunt G., Carrano M.T., Campione N. (2018), Cope's rule and the adaptive landscape of dinosaur body size evolution. *Palaeontology*, 61:13-48. Ho L.S.T, Ané C. 2014. Intrinsic inference difficulties for trait evolution with Ornstein-Uhlenbeck models. *Methods in Ecology & Evolution*, 5:1133–1146. Ingram T., Mahler D.L. 2013. SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. *Methods in Ecology & Evolution*, 4:416–425. Khabbazian M., Kriebel R., Rohe K., Ané, C. 2016. Fast and accurate detection of evolutionary shifts in Ornstein-Uhlenbeck models. *Methods in Ecology & Evolution*, 7:811–824. Mongiardino Koch N. 2021. Exploring adaptive landscapes across deep time: A case study using echinoid body size. *Evolution*, <https://doi.org/10.1111/evo.14219>.

See Also

For details on how these models are fit visit [surfaceBackward](#) and [OUwie](#). Plots of AIC values can be obtained with [extended_surfaceAICPlot](#)

Examples

```
## Not run:
data(echinoid_data)
OUmodels <- surfaceExtended(bwd_surface = echinoid_data$bwd_surface, data =
  echinoid_data$size, tree = echinoid_data$tree, error = echinoid_data$error,
  models = 'OUMVAZ', limit = 4, plot = T, fwd_surface =
  echinoid_data$fwd_surface)

## End(Not run)
```

```
transitions_through_time
```

Summarize trait dynamics through time

Description

This function provides a way of summarizing the temporal dynamics of traits by registering the times when there are transitions between character states. First, a number of replicates of stochastic character mappings are created (see [make.simmap](#) for more details), and the phylogeny is traversed in order to register the moments in time when there are transitions between states. For every time a transition occurs, the function will follow all descendants and also register the point in time where that instance of a given state goes extinct (either by all descendants transitioning to a different state or by all tips going extinct).

Usage

```
transitions_through_time(tree, char, repl = 100, model = "ER", plot = T)
```

Arguments

<code>tree</code>	Phylogenetic tree in 'phylo' format.
<code>char</code>	Named vector including character states for all tips. Names need to correspond to tips in the phylogeny.
<code>repl</code>	Number of replicates of stochastic character mapping (default = 100)
<code>model</code>	A character specifying the model of evolution used to reconstruct the evolutionary history of the character. Options include 'ER' (default), 'SYM' and 'ARD', see ace .
<code>plot</code>	A logical indicating whether to plot the results. Default is TRUE.

Details

The function only requires a tree and a character, and allows the user to modify the number of replicates, the model used for character mapping, as well as whether to return a plot summarizing the results. More control on what and how to plot these results is attained by providing the output to the function [plot_ttt](#).

Value

A data.frame including the times of birth and death of each instance of a character across replicates. This object can be passed to [plot_ttt](#) to obtain different visualizations.

Author(s)

Nicolás Mongiardino Koch

References

Mongiardino Koch N. 2021. Exploring adaptive landscapes across deep time: A case study using echinoid body size. *Evolution*, <https://doi.org/10.1111/evo.14219>.

See Also

For details on stochastic character mapping visit [make.simmap](#). Plots can be explored with [plot_ttt](#)

Index

* datasets

echinoid_data, 2

ace, 8

echinoid_data, 2

extended_surfaceAICPlot, 3, 7

gam, 4, 5

make.simmap, 8, 9

OUwie, 7

plot_ttt, 4, 8, 9

surfaceBackward, 7

surfaceExtended, 3, 5

transitions_through_time, 4, 5, 8