Package ‘phylosem’

July 23, 2025

Type Package

Title Phylogenetic Structural Equation Model

Version 1.1.4

Date 2024-04-02

Imports sem, ape, phylobase, phylopath, methods

Depends TMB, R (>= 4.0.0),

Suggests semPlot, TreeTools, Rphylopars, phylolm, fishtree, phyr, knitr, rmarkdown, ggplot2, testthat, phylosignal, adephylo

Enhances phytools, DiagrammeR

LinkingTo RcppEigen, TMB

Description

Applies phylogenetic comparative methods (PCM) and phylogenetic trait imputation using structural equation models (SEM), extending methods from Thorson et al. (2023) <doi:10.1111/2041-210X.14076>. This implementation includes a minimal set of features, to allow users to easily read all of the documentation and source code. PCM using SEM includes phylogenetic linear models and structural equation models as nested submodels, but also allows imputation of missing values. Features and comparison with other packages are described in Thorson and van der Bijl (2023) <doi:10.1111/jeb.14234>.

License GPL-3

Encoding UTF-8

RoxygenNote 7.3.1

VignetteBuilder knitr

URL `https://james-thorson-noaa.github.io/phylosem/`

BugReports `https://github.com/James-Thorson-NOAA/phylosem/issues`

NeedsCompilation yes

Author James Thorson [aut, cre] (ORCID: <`https://orcid.org/0000-0001-7415-1010`>), Wouter van der Bijl [ctb] (ORCID: <`https://orcid.org/0000-0002-7366-1868`>)

Maintainer James Thorson <`James.Thorson@noaa.gov`>

Repository CRAN

Date/Publication 2024-04-02 21:15:03 UTC

Contents

`as_fitted_DAG` *Convert phylosem to phylopath output*

Description

Convert output from package phylosem to phylopath

Usage

```
as_fitted_DAG(object)
```

Arguments

`object`	Output from `phylosem`

Value

Convert output to format supplied by `est_DAG`

`as_phylo4d` *Convert phylosem to phylo4d*

Description

Convert output from package phylosem to phylo4d object from package phylobase

Usage

```
as_phylo4d(object, what = c("Estimate", "Std. Error"))
```

Arguments

`object`	Output from `phylosem`
`what`	Select what to convert (Estimate / Std. Error).

Details

This package is intended to for use in using plots assocaited with package sem, e.g., using package plotSEM `semPlot::semPlotModel`

Value

phylosem output to converted format supplied by `phylo4d`

`as_sem` *Convert phylosem to sem output*

Description

Convert output from package phylosem to output from package sem

Usage

```
as_sem(object)
```

Arguments

`object`	Output from `phylosem`

Value

Output converted to format supplied by `sem`

`average` *Choose model*

Description

Choose model

Usage

```
average(x, cut_off, avg_method)
```

Arguments

`x`	output from `compare_phylosem`
`cut_off`	threshold where any model with delta-AIC greater than this value is excluded from average
`avg_method`	see `average_DAGs`

Value

Returns an AIC-weighted average of fitted models from `compare_phylosem` after conversion to format from [phylopath::est_DAG]

`average.compare_phylosem`
Choose model

Description

Choose model

Usage

```
## S3 method for class 'compare_phylosem'
average(x, cut_off = 2, avg_method = "conditional")
```

Arguments

`x`	output from `compare_phylosem`
`cut_off`	threshold where any model with delta-AIC greater than this value is excluded from average
`avg_method`	see `average_DAGs`

Value

Returns an AIC-weighted average of fitted models from `compare_phylosem` after conversion to format from `est_DAG`

`best` *Extract best fitted model*

Description

Extract best fitted model

Usage

```
best(x)
```

Arguments

`x`	output from `compare_phylosem`

Value

Returns best model from those fitted using `compare_phylosem`

`best.compare_phylosem` *Extract best fitted model*

Description

Extract best fitted model

Usage

```
## S3 method for class 'compare_phylosem'
best(x)
```

Arguments

`x`	output from `compare_phylosem`

Value

Returns best model from those fitted using `compare_phylosem`

`choice` *Choose model*

Description

Choose model

Usage

```
choice(x, choice)
```

Arguments

`x`	output from `compare_phylosem`
`choice`	Integer indicating model to extract

Value

Returns chosen model from those fitted using `compare_phylosem`

`choice.compare_phylosem` *Choose model*

Description

Choose model

Usage

```
## S3 method for class 'compare_phylosem'
choice(x, choice)
```

Arguments

`x`	output from `compare_phylosem`
`choice`	Integer indicating model to extract

Value

Returns chosen model from those fitted using `compare_phylosem`

`coef.phylosem` *Extract path coefficients*

Description

Extract path coefficients.

Usage

```
## S3 method for class 'phylosem'
coef(object, standardized = FALSE, ...)
```

Arguments

`object`	Output from `phylosem`
`standardized`	Whether to standardize regression coefficients
`...`	Not used

Value

Data-frame listing all path coefficients, their parameter index and estimated values

---

`compare_phylosem` *Compare phylogenetic structural equation models*

Description

Fits several phylogenetic structural equation model for further comparison

Usage

```
compare_phylosem(
  sem_set,
  tree,
  data,
  family = rep("fixed", ncol(data)),
  covs,
  estimate_ou = FALSE,
  estimate_lambda = FALSE,
  estimate_kappa = FALSE,
  control = phylosem_control(),
  ...
)
```

Arguments

`sem_set`	A named list of structural equation model specifications, where each element will be passed as argument `sem` to `phylosem`
`tree`	phylogenetic structure, using class `as.phylo`
`data`	data-frame providing variables being modeled. Missing values are inputted as NA. If an SEM includes a latent variable (i.e., variable with no available measurements) then it still must be inputted as a column of `data` with entirely NA values.
`family`	Character-vector listing the distribution used for each column of `data`, where each element must be `fixed`, `normal`, `binomial`, or `poisson`. `family="fixed"` is default behavior and assumes that a given variable is measured exactly. Other options correspond to different specifications of measurement error.
`covs`	optional: a character vector of one or more elements, with each element giving a string of variable names, separated by commas. Variances and covariances among all variables in each such string are added to the model. For confirmatory factor analysis models specified via `cfa`, `covs` defaults to all of the factors in the model, thus specifying all variances and covariances among these factors. *Warning*: `covs="x1, x2"` and `covs=c("x1", "x2")` are *not* equivalent: `covs="x1, x2"` specifies the variance of `x1`, the variance of `x2`, *and* their covariance, while `covs=c("x1", "x2")` specifies the variance of `x1` and the variance of `x2` *but not* their covariance.
`estimate_ou`	Boolean indicating whether to estimate an autoregressive (Ornstein-Uhlenbeck) process using additional parameter `lnalpha`, corresponding to the `model="OUrandomRoot"` parameterization from **phylolm** as listed in doi:10.1093/sysbio/syu005
`estimate_lambda`	Boolean indicating whether to estimate additional branch lengths for phylogenetic tips (a.k.a. the Pagel-lambda term) using additional parameter `logitlambda`
`estimate_kappa`	Boolean indicating whether to estimate a nonlinear scaling of branch lengths (a.k.a. the Pagel-kappa term) using additional parameter `lnkappa`
`control`	Output from `phylosem_control`, used to define user settings, and see documentation for that function for details.
`...`	Additional arguments passed to `phylosem`

Value

An object (list) of class ‘compare_phylosem‘, containing a list of output from `phylosem`

`list_parameters` *List fixed and random effects*

Description

`list_parameters` lists all fixed and random effects

Usage

```
list_parameters(Obj, verbose = TRUE)
```

Arguments

`Obj`	Compiled TMB object
`verbose`	Boolean, whether to print messages to terminal

Value

Tagged-list of fixed and random effects, returned invisibly and printed to screen

---

`Mlifehist_ver1_0` *Fisheries natural mortality example*

Description

Data used to demonstrate phylogenetic comparative methods for fisheries science. Specifically a copy of the Then et al. database doi:10.1093/icesjms/fsu136 using file "Mlifehist_ver1.0.csv" accessed from `https://www.vims.edu/research/departments/fisheries/programs/mort_db/`

Usage

```
data(Mlifehist_ver1_0)
```

`parse_path` *Parse path*

Description

`parse_path` is copied from `sem::parse.path`

Usage

```
parse_path(path)
```

Arguments

`path`	text to parse

Details

Copied with permission from John Fox under licence GPL (>= 2)

Value

Tagged-list defining variables and direction for a specified path coefficient

---

`phylosem` *Fit phylogenetic structural equation model*

Description

Fits a phylogenetic structural equation model

Usage

```
phylosem(
  sem,
  tree,
  data,
  family = rep("fixed", ncol(data)),
  covs = colnames(data),
  estimate_ou = FALSE,
  estimate_lambda = FALSE,
  estimate_kappa = FALSE,
  data_labels = rownames(data),
  tmb_inputs = NULL,
  control = phylosem_control()
)
```

Arguments

`sem`	structural equation model structure, passed to either `specifyModel` or `specifyEquations` and then parsed to control the set of path coefficients and variance-covariance parameters
`tree`	phylogenetic structure, using class `as.phylo`
`data`	data-frame providing variables being modeled. Missing values are inputted as NA. If an SEM includes a latent variable (i.e., variable with no available measurements) then it still must be inputted as a column of `data` with entirely NA values.
`family`	Character-vector listing the distribution used for each column of `data`, where each element must be `fixed`, `normal`, `binomial`, or `poisson`. `family="fixed"` is default behavior and assumes that a given variable is measured exactly. Other options correspond to different specifications of measurement error.
`covs`	optional: a character vector of one or more elements, with each element giving a string of variable names, separated by commas. Variances and covariances among all variables in each such string are added to the model. For confirmatory factor analysis models specified via `cfa`, `covs` defaults to all of the factors in the model, thus specifying all variances and covariances among these factors. *Warning*: `covs="x1, x2"` and `covs=c("x1", "x2")` are *not* equivalent: `covs="x1, x2"` specifies the variance of `x1`, the variance of `x2`, *and* their covariance, while `covs=c("x1", "x2")` specifies the variance of `x1` and the variance of `x2` *but not* their covariance.

`estimate_ou`	Boolean indicating whether to estimate an autoregressive (Ornstein-Uhlenbeck) process using additional parameter `lnalpha`, corresponding to the `model="OUrandomRoot"` parameterization from **phylolm** as listed in doi:10.1093/sysbio/syu005
`estimate_lambda`	Boolean indicating whether to estimate additional branch lengths for phylogenetic tips (a.k.a. the Pagel-lambda term) using additional parameter `logitlambda`
`estimate_kappa`	Boolean indicating whether to estimate a nonlinear scaling of branch lengths (a.k.a. the Pagel-kappa term) using additional parameter `lnkappa`
`data_labels`	For each row of `data`, listing the corresponding name from `tree$tip.label`. Default pulls `data_labels` from `rownames(data)`
`tmb_inputs`	optional tagged list that overrides the default constructor for TMB inputs (use at your own risk)
`control`	Output from `phylosem_control`, used to define user settings, and see documentation for that function for details.

Details

Note that parameters `logitlambda`, `lnkappa`, and `lnalpha` if estimated are each estimated as having a single value that applies to all modeled variables. This differs from default behavior in **phylolm**, where these parameters only apply to the "response" and not "predictor" variables. This also differs from default behavior in **phylopath**, where a different value is estimated in each call to **phylolm** during the d-separation estimate of path coefficients. However, it is consistent with default behavior in **Rphylopars**, and estimates should be comparable in that case. These additional parameters are estimated with unbounded support, which differs somewhat from default bounded estimates in **phylolm**, although parameters should match if overriding **phylolm** defaults to use unbounded support. Finally, `phylosem` allows these three parameters to be estimated in any combination, which is expanded functionality relative to the single-option functionality in **phylolm**.

Also note that **phylopath** by default uses standardized coefficients. To achieve matching parameter estimates between **phylosem** and **phylopath**, standardize each variable to have a standard deviation of 1.0 prior to fitting with **phylosem**.

Value

An object (list) of class ‘phylosem‘. Elements include:

data Copy of argument `data`

SEM_model SEM model parsed from `sem` using `specifyModel` or `specifyEquations`

obj TMB object from `MakeADFun`

tree Copy of argument `tree`

tmb_inputs The list of inputs passed to `MakeADFun`

opt The output from `nlminb`

sdrep The output from `sdreport`

report The output from `obj$report()`

parhat The output from `obj$env$parList()` containing maximum likelihood estimates and empirical Bayes predictions

References

Introducing the package, its features, and comparison with other software (to cite when using phylosem):

Thorson, J. T., & van der Bijl, W. (In press). phylosem: A fast and simple R package for phylogenetic inference and trait imputation using phylogenetic structural equation models. Journal of Evolutionary Biology. doi:10.1111/jeb.14234

Statistical methods for phylogenetic structural equation models

Thorson, J. T., Maureaud, A. A., Frelat, R., Merigot, B., Bigman, J. S., Friedman, S. T., Palomares, M. L. D., Pinsky, M. L., Price, S. A., & Wainwright, P. (2023). Identifying direct and indirect associations among traits by merging phylogenetic comparative methods and structural equation models. Methods in Ecology and Evolution, 14(5), 1259-1275. doi:10.1111/2041210X.14076

Earlier development of computational methods, originally used for phlogenetic factor analysis:

Thorson, J. T. (2020). Predicting recruitment density dependence and intrinsic growth rate for all fishes worldwide using a data-integrated life-history model. Fish and Fisheries, 21(2), 237-251. doi:10.1111/faf.12427

Thorson, J. T., Munch, S. B., Cope, J. M., & Gao, J. (2017). Predicting life history parameters for all fishes worldwide. Ecological Applications, 27(8), 2262-2276. doi:10.1002/eap.1606

Earlier development of phylogenetic path analysis:

van der Bijl, W. (2018). phylopath: Easy phylogenetic path analysis in R. PeerJ, 6, e4718. doi:10.7717/peerj.4718

von Hardenberg, A., & Gonzalez-Voyer, A. (2013). Disentangling evolutionary cause-effect relationships with phylogenetic confirmatory path analysis. Evolution; International Journal of Organic Evolution, 67(2), 378-387. doi:10.1111/j.15585646.2012.01790.x

Interface involving SEM ‘arrow notation‘ is repurposed from:

Fox, J., Nie, Z., & Byrnes, J. (2020). Sem: Structural equation models. R package version 3.1-11. `https://CRAN.R-project.org/package=sem`

Coercing output to phylo4d depends upon:

Bolker, B., Butler, M., Cowan, P., de Vienne, D., Eddelbuettel, D., Holder, M., Jombart, T., Kembel, S., Michonneau, F., & Orme, B. (2015). phylobase: Base package for phylogenetic structures and comparative data. R Package Version 0.8.0. `https://CRAN.R-project.org/package=phylobase`

Laplace approximation for parameter estimation depends upon:

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. M. (2016). TMB: Automatic differentiation and Laplace approximation. Journal of Statistical Software, 70(5), 1-21. doi:10.18637/jss.v070.i05

Examples

```
# Load data set
data(rhino, rhino_tree, package=”phylopath”)

# Run phylosem
model = ”
  DD -> RS, p1
  BM -> LS, p2
```

```
  BM -> NL, p3
  NL -> DD, p4
"
psem = phylosem( sem = model,
          data = rhino[,c("BM","NL","DD","RS","LS")],
          tree = rhino_tree )

# Convert and plot using phylopath
library(phylopath)
my_fitted_DAG = as_fitted_DAG(psem)
coef_plot( my_fitted_DAG )
plot( my_fitted_DAG )

# Convert to phylo4d to extract estimated traits and Standard errors
# for all ancestors and tips in the tree.
# In this rhino example, note that species are labeled s1-s100
# and ancestral nodes are not named.
(traits_est = as_phylo4d(psem))
(traits_SE = as_phylo4d(psem, what="Std. Error"))

# Convert to sem and plot
library(sem)
my_sem = as_sem(psem)
pathDiagram( model = my_sem,
                  style = "traditional",
                  edge.labels = "values" )
effects( my_sem )

# Plot using semPlot
if( require(semPlot) ){
  myplot = semPlotModel( my_sem )
  semPaths( my_sem,
                   nodeLabels = myplot@Vars$name )
}
```

`phylosem_control` *Detailed control for phylosem structure*

Description

Define a list of control parameters. Note that the format of this input is likely to change more rapidly than that of `phylosem`

Usage

```
phylosem_control(
  nlminb_loops = 1,
  newton_loops = 1,
  trace = 0,
```

```
  eval.max = 1000,
  iter.max = 1000,
  getsd = TRUE,
  quiet = FALSE,
  run_model = TRUE,
  getJointPrecision = FALSE
)
```

Arguments

`nlminb_loops`	Integer number of times to call `nlminb`.
`newton_loops`	Integer number of Newton steps to do after running `nlminb`.
`trace`	Parameter values are printed every ‘trace‘ iteration for the outer optimizer. Passed to ‘control‘ in `nlminb`.
`eval.max`	Maximum number of evaluations of the objective function allowed. Passed to ‘control‘ in `nlminb`.
`iter.max`	Maximum number of iterations allowed. Passed to ‘control‘ in `nlminb`.
`getsd`	Boolean indicating whether to call `sdreport`
`quiet`	Boolean indicating whether to run model printing messages to terminal or not;
`run_model`	Boolean indicating whether to estimate parameters (the default), or instead to return the model inputs and compiled TMB object without running;
`getJointPrecision`	whether to get the joint precision matrix. Passed to `sdreport`.

Value

An S3 object of class "phylosem_control" that specifies detailed model settings, allowing user specification while also specifying default values

---

`print.phylosem` *Print parameter estimates and standard errors.*

Description

Print parameter estimates

Usage

```
## S3 method for class 'phylosem'
print(x, ...)
```

Arguments

`x`	Output from `phylosem`
`...`	Not used

Value

prints (and invisibly returns) output from `nlminb`

`summary.phylosem` *summarize phylosem*

Description

Summarize phylosem output from phylosem, including calculating intercepts at the tree root

Usage

```
## S3 method for class 'phylosem'
summary(object, ...)
```

Arguments

`object`	Output from `phylosem`
`...`	Not used

Value

Data-frame containing all estimated intercepts, path coefficients, and variance-covariance parameters as well as their standard errors

---

`vcov.phylosem` *Extract Variance-Covariance Matrix*

Description

extract the covariance of fixed effects, or both fixed and random effects.

Usage

```
## S3 method for class 'phylosem'
vcov(object, which = c("fixed", "random", "both"), ...)
```

Arguments

`object`	output from `phylosem`
`which`	whether to extract the covariance among fixed effects, random effects, or both
`...`	ignored, for method compatibility

Index