

Package: do3PCA (via r-universe)

October 4, 2026

Title Probabilistic Phylogenetic Principal Component Analysis

Version 1.0.2

Description Estimates probabilistic phylogenetic Principal Component Analysis (PCA) and non-phylogenetic probabilistic PCA. Provides methods to implement alternative models of trait evolution including Brownian motion (BM), Ornstein-Uhlenbeck (OU), Early Burst (EB), and Pagel's lambda. Also provides flexible biplot functions.

License GPL (>= 2.0)

Encoding UTF-8

Roxygen list(markdown = TRUE)

Depends R (>= 3.5.0), Rdimtools

Imports ape, phytools, matrixcalc, mclust, nloptr, ratematrix, stats, utils, mcmc, Matrix

RoxygenNote 7.3.2

Suggests knitr, rmarkdown

VignetteBuilder knitr

Config/pak/sysreqs
cmake libgmp-dev libgmp3-dev make libxml2-dev pkg-config libx11-dev libclang-dev

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

Date/Publication 2025-09-05 14:21:05 UTC

RemoteUrl <https://github.com/caetanods/do3pca>

RemoteRef HEAD

RemoteSha dc0eece3ed6bc20d44a3c2ac0020a4a18fd0bec7

Contents

BrokenStick	2
doBiplot	3
loglik	4

MCMCphylProbPCA	5
MCMCProbPCA	7
modelChoice	8
phylProbPCA	9
phylProbPCAJoint	10
picProbPCA	11
ProbPCA	13
traitModelAIC	14
Index	15

BrokenStick	<i>Broken stick test for number of principal components</i>
-------------	---

Description

Function generates the expected distribution of variance across the principal components and compares it with the observed vector of explained variance (if provided).

Usage

```
BrokenStick(n_dim, var_vec = NULL, tol = 0, plot = TRUE)
```

Arguments

- n_dim the number of variables of the data (required).
- var_vec the vector of explained variance of the PCA (optional).
- tol a tolerance value that is used to check if the observed explained variance is larger than the expected under the Broken Stick model.
- plot if a plot should be produced.

Value

Function will return the vector of explained variance for each PC. Will also return the number of PCs to keep and optionally a plot of the results if 'var_vec' is provided.

doBiplot

*Make biplot for any type of PCA***Description**

Function to make biplots for any kind of PCA. It accepts the outputs from standard PCA (princomp and prcomp). It also works with the "phylProbPCA" and "ProbPCA" functions. It provides more options to the plot than the standard "stats::biplot".

Usage

```
doBiplot(x, choices = 1L:2L, scale = 1, pc.biplot = FALSE, col, ...)
```

Arguments

x	output from PCA analysis.
choices	numeric vector of length 2. Use to choose which of the PC axes to plot. Default plots first and second axes: "choices = c(1,2)".
scale	numeric value between 0 and 1. Same as in "stats::biplot.princomp". See ?biplot.princomp for more information.
pc.biplot	logical. If TRUE it will produce a "principal component biplot" (sensu Gabriel, 1971). Same as in "stats::biplot.princomp". See ?biplot.princomp for more information.
col	character vector of length 3 with the colors of the biplot. First color is used for the score points (or sample names), second color for arrows and variable names, and third color for the right and top-side ticks (plot axes).
...	extra parameters for the function. Same as "stats::biplot".

Details

Function has the same options as "stats::biplot", with the addition of the following arguments. "plot_dimnames" controls if the names of the samples (species) will be plotted. "add_points" controls if the score points will be plotted. "add_margin" is a numeric value that expands the area of the plot. You can use this to make sure the names of variables and samples (species) fit the plot.

Value

makes a biplot of the PCA results.

Examples

```
phy <- ratematrix::anoles$phy[[1]]
dt <- as.matrix( ratematrix::anoles$data[,1:3] )

## Using probabilistic phylogenetic PCA
phylppca <- phylProbPCA(phy = phy, x = dt, ret_dim = 2)
```

```

doBiplot(x = phylppca, add_margin = 0.3)

## Using standard phylogenetic PCA
phylpca <- phytools::phyl.pca(tree = phy, Y = dt)
doBiplot(x = phylpca, add_margin = 0.3)

## Using probabilistic PCA
ppca <- ProbPCA(x = dt)
doBiplot(x = ppca, add_margin = 0.3)

## Using standard PCA
pca1 <- princomp(x = dt)
doBiplot(x = pca1, add_margin = 0.1)

## Using standard PCA
pca2 <- prcomp(x = dt)
doBiplot(x = pca2, add_margin = 0.1)

```

loglik

Probabilistic PCA likelihood function

Description

Likelihood function for the standard probabilistic PCA or the phylogenetic probabilistic PCA. Note that one just need to change between the covariance of the data and the R matrix (multivariate Brownian Motion covariance matrix). The likelihood function only uses the eigenvalues of the covariance matrix (see parameter g).

Usage

```
loglik(L, sigma, N, d, g)
```

Arguments

L	vector with length equal to the number of retained dimensions.
sigma	variance. This is the average variance lost per discarded dimension.
N	sample size. Number of observations.
d	integer. Number of dimensions in the data.
g	eigenvalues of the covariance matrix.

Value

returns the $-1 * \log$ -likelihood of the data.

References

- Revell, L. J. 2009. Size-Correction and Principal Components for Interspecific Comparative Studies. *Evolution* 63:3258–3268. doi: 10.1111/j.1558-5646.2009.00804.x
- Revell, L. J. 2024. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ* 12:e16505. doi: 10.7717/peerj.16505
- Tipping, M. E., and C. M. Bishop. 1999. Probabilistic Principal Component Analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 61(3):611–622. doi: 10.1111/1467-9868.00196
- Zhang, A., Chan, K.L., Kwok, J.T., and D-Y Yeung. 2004. Proceedings of the AAAI Conference on Artificial Intelligence 19 Book One, Volume. <https://aaai.org/papers/00372-aaai04-060-bayesian-inference-on-principal-component-analysis-using-reversible-jump-markov-chain-monte-carlo/>

Examples

```
phy <- ratematrix::anoles$phy[[1]]
dt <- as.matrix( ratematrix::anoles$data[,1:3] )
L_start <- rgamma(n = 2, shape = 2, rate = 1)
sigma_start <- rgamma(n = 1, shape = 2, rate = 1)
R <- do3PCA:::fastR(x = dt, phy = phy)
g <- eigen(R)$values
loglik(L = L_start, sigma = sigma_start, N = nrow(dt), d = 2, g = g)
```

MCMCphylProbPCA

MCMC for Probabilistic Phylogenetic PCA

Description

Function to perform probabilistic phylogenetic PCA using Bayesian Markov-Chain Monte Carlo.

Usage

```
MCMCphylProbPCA(
  phy,
  x,
  ret_dim = 2,
  scale = TRUE,
  fixed_R = NULL,
  gamma_shape = 1,
  gamma_rate = 2,
  gen = 10^6,
  step_scale = 0.08,
  burn = 0.2,
  post_score_samples = 100
)
```

Arguments

phy	phylogeny in "phylo" format.
x	a matrix with traits in columns and species values in rows. Rownames must match the tip labels of phylogeny.
ret_dim	number of dimensions (PC axes) to be kept by the model.
scale	if data should be rescaled to have zero mean and unit variance using the function <code>scale()</code> . Default is TRUE.
fixed_R	an evolutionary rate matrix to be used for the computation.
gamma_shape	The shape parameter for the gamma prior.
gamma_rate	The rate parameter for the gamma prior.
gen	The number generations for the MCMC.
step_scale	The scaler for the mulplier proposal of the MCMC.
burn	The proportion of the burn-in to the excluded before computing the average posterior parameter value.
post_score_samples	The number of posterior samples used to compute the posterior of scores.

Details

This function uses the "mcmc" R package to conduct a Metropolis-Hastings search. The output "mcmc" can be used for further analyses of the MCMC chain using the "mcmc" package.

Value

returns a list with "mean_scores", "percent_var_pcs", "var_loadings", and "stats". Also return "mcmc" which is the output from the "mcmc" R package.

References

- Revell, L. J. 2009. Size-Correction and Principal Components for Interspecific Comparative Studies. *Evolution* 63:3258–3268. doi: 10.1111/j.1558-5646.2009.00804.x
- Revell, L. J. 2024. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ* 12:e16505. doi: 10.7717/peerj.16505
- Tipping, M. E., and C. M. Bishop. 1999. Probabilistic Principal Component Analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 61(3):611–622. doi: 10.1111/1467-9868.00196
- Zhang, A., Chan, K.L., Kwok, J.T., and D-Y Yeung. 2004. Proceedings of the AAAI Conference on Artificial Intelligence 19 Book One, Volume. <https://aaai.org/papers/00372-aaai04-060-bayesian-inference-on-principal-component-analysis-using-reversible-jump-markov-chain-monte-carlo/>

Examples

```

phy <- ratematrix::anoles$phy[[1]]
dt <- as.matrix( ratematrix::anoles$data[,1:3] )
## Do a VERY small chain to test:
mcmc3PCA <- MCMCphylProbPCA(phy = phy, x = dt, gen = 500)

```

MCMCProbPCA

*MCMC for Probabilistic PCA***Description**

Function to perform probabilistic PCA using Bayesian Markov-Chain Monte Carlo.

Usage

```

MCMCProbPCA(
  x,
  ret_dim = 2,
  scale = TRUE,
  gamma_shape = 1,
  gamma_rate = 2,
  gen = 10^6,
  step_scale = 0.08,
  burn = 0.2
)

```

Arguments

<code>x</code>	a matrix with traits in columns and species values in rows.
<code>ret_dim</code>	number of dimensions (PC axes) to be kept by the model.
<code>scale</code>	if data should be rescaled to have zero mean and unit variance using the function <code>scale()</code> . Default is <code>TRUE</code> .
<code>gamma_shape</code>	The shape parameter for the gamma prior.
<code>gamma_rate</code>	The rate parameter for the gamma prior.
<code>gen</code>	The number generations for the MCMC.
<code>step_scale</code>	The scaler for the mulplier proposal of the MCMC.
<code>burn</code>	The proportion of the burn-in to the excluded before computing the average posterior parameter value.

Details

This function uses the "mcmc" R package to conduct a Metropolis-Hastings search. The output "mcmc" can be used for further analyses of the MCMC chain using the "mcmc" package.

Value

returns a list with "mean_scores", "percent_var_pcs", "var_loadings", and "stats". Also return "mcmc" which is the output from the "mcmc" R package.

References

Revell, L. J. 2009. Size-Correction and Principal Components for Interspecific Comparative Studies. *Evolution* 63:3258–3268. doi: 10.1111/j.1558-5646.2009.00804.x

Revell, L. J. 2024. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ* 12:e16505. doi: 10.7717/peerj.16505

Tipping, M. E., and C. M. Bishop. 1999. Probabilistic Principal Component Analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 61(3):611–622. doi: 10.1111/1467-9868.00196

Zhang, A., Chan, K.L., Kwok, J.T., and D-Y Yeung. 2004. Proceedings of the AAAI Conference on Artificial Intelligence 19 Book One, Volume. <https://aaai.org/papers/00372-aaai04-060-bayesian-inference-on-principal-component-analysis-using-reversible-jump-markov-chain-monte-carlo/>

Examples

```
dt <- as.matrix( ratematrix::anoles$data[,1:3] )
## Do a VERY small chain to test:
mcmcPPCA <- MCMCProbPCA(x = dt, gen = 100)
```

modelChoice

Automatic choice of number of principal components

Description

Automatic choice of number of principal components

Usage

```
modelChoice(x = NULL, phy = NULL, cov = NULL, N = NULL)
```

Arguments

x	data. Columns are variables/features. Rows are observations/species.
phy	phylogenetic tree. Number of tips must match number of rows in x.
cov	a covariance matrix. Either the covariance of the data or an evolutionary ratematrix (R).
N	number of observations/species in the tree.

Value

a list with the best fitting number of principal components (k) and a vector with the log probability of the data given the number of principal components (k).

phylProbPCA

Probabilistic Phylogenetic PCA

Description

Function to perform probabilistic phylogenetic PCA. Allows for fit of alternative models of trait evolution using branch length transformation.

Usage

```
phylProbPCA(phy, x, ret_dim = 2, model = "BM", scale = TRUE, quiet = FALSE)
```

Arguments

phy	phylogeny in "phylo" format.
x	a matrix with traits in columns and species values in rows. Rownames must match the tip labels of phylogeny.
ret_dim	number of dimensions (PC axes) to be kept by the model.
model	choice of model of trait evolution. One of "BM", "lambda", "OU", or "EB".
scale	if data should be rescaled to have zero mean and unit variance using the function scale(). Default is TRUE.
quiet	if function should suppress output to the console while running

Details

The function can be used to estimate the probabilistic phylogenetic PCA (3PCA) using distinct models of trait evolution. Models are implemented using branch length transformation. Model fitting happens in two steps. First the maximum likelihood of the evolutionary covariance matrix (R) and the parameter of the model is estimated. Then the 3PCA model is estimated using the phylogenetic tree with branch lengths transformed following the MLE for the parameter of each trait evolution model.

The function returns a list with the following elements. scores: the scores of the principal components; e_values: eigenvalues; e_vectors: eigenvectors or the projection; model_fit: information about the trait evolution model; loadings: the loadings of the PCs; varnames: the names of the variables; sig: the MLE of the error; mle.W: the MLE of the W matrix; Function also returns AIC, AICc, and BIC for the model.

Value

returns a list of class "phylPPCA". See "Details" for more information.

References

- Revell, L. J. 2009. Size-Correction and Principal Components for Interspecific Comparative Studies. *Evolution* 63:3258–3268. doi: 10.1111/j.1558-5646.2009.00804.x
- Revell, L. J. 2024. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ* 12:e16505. doi: 10.7717/peerj.16505
- Tipping, M. E., and C. M. Bishop. 1999. Probabilistic Principal Component Analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 61(3):611–622. doi: 10.1111/1467-9868.00196

Examples

```
phy <- ratematrix::anoles$phy[[1]]
dt <- as.matrix( ratematrix::anoles$data[,1:3] )
ppca <- phylProbPCA(phy = phy, x = dt, ret_dim = 2)
doBiplot(x = ppca, add_margin = 0.3)
```

phylProbPCAJoint	<i>Probabilistic Phylogenetic PCA (Joint)</i>
------------------	---

Description

Function to perform probabilistic phylogenetic PCA. Allows for fit of alternative models of trait evolution using branch length transformation.

Usage

```
phylProbPCAJoint(
  phy,
  x,
  ret_dim = 2,
  model = "BM",
  scale = TRUE,
  quiet = FALSE
)
```

Arguments

phy	phylogeny in "phylo" format.
x	a matrix with traits in columns and species values in rows. Rownames must match the tip labels of phylogeny.
ret_dim	number of dimensions (PC axes) to be kept by the model.
model	choice of model of trait evolution. One of "BM", "lambda", "OU", or "EB".
scale	if data should be rescaled to have zero mean and unit variance using the function scale(). Default is TRUE.
quiet	if function should suppress output to the console while running

Details

The function can be used to estimate the probabilistic phylogenetic PCA (3PCA) using distinct models of trait evolution. Models are implemented using branch length transformation. Model fitting happens in two steps. First the maximum likelihood of the evolutionary covariance matrix (R) and the parameter of the model is estimated. Then the 3PCA model is estimated using the phylogenetic tree with branch lengths transformed following the MLE for the parameter of each trait evolution model.

The function returns a list with the following elements. scores: the scores of the principal components; e_values: eigenvalues; e_vectors: eigenvectors or the projection; model_fit: information about the trait evolution model; loadings: the loadings of the PCs; varnames: the names of the variables; sig: the MLE of the error; mle.W: the MLE of the W matrix; Function also returns AIC, AICc, and BIC for the model.

Value

returns a list of class "phylPPCA". See "Details" for more information.

References

- Revell, L. J. 2009. Size-Correction and Principal Components for Interspecific Comparative Studies. *Evolution* 63:3258–3268. doi: 10.1111/j.1558-5646.2009.00804.x
- Revell, L. J. 2024. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ* 12:e16505. doi: 10.7717/peerj.16505
- Tipping, M. E., and C. M. Bishop. 1999. Probabilistic Principal Component Analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 61(3):611–622. doi: 10.1111/1467-9868.00196

Examples

```
phy <- ratematrix::anoles$phy[[1]]
dt <- as.matrix( ratematrix::anoles$data[,1:3] )
ppca <- phylProbPCA(phy = phy, x = dt, ret_dim = 2)
doBiplot(x = ppca, add_margin = 0.3)
```

picProbPCA

Probabilistic PCA of independent contrasts (PIC-PCA)

Description

Function to perform probabilistic PCA of independent contrasts. Allows for fit of alternative models of trait evolution using branch length transformation.

Usage

```
picProbPCA(phy, x, ret_dim = 2, model = "BM", scale = TRUE, quiet = FALSE)
```

Arguments

phy	phylogeny in "phylo" format.
x	a matrix with traits in columns and species values in rows. Rownames must match the tip labels of phylogeny.
ret_dim	number of dimensions (PC axes) to be kept by the model.
model	choice of model of trait evolution. One of "BM", "lambda", "OU", or "EB".
scale	if data should be rescaled to have zero mean and unit variance using the function <code>scale()</code> . Default is TRUE.
quiet	if function should suppress output to the console while running

Details

The function can be used to estimate PIC-PCA using distinct models of trait evolution. Models are implemented using branch length transformation. Model fitting happens in two steps. First the maximum likelihood of the evolutionary covariance matrix (R) and the parameter of the model is estimated. Then the PIC-PCA model is estimated using the phylogenetic tree with branch lengths transformed following the MLE for the parameter of each trait evolution model.

The function returns a list with the following elements. `scores`: the scores of the principal components; `e_values`: eigenvalues; `e_vectors`: eigenvectors or the projection; `model_fit`: information about the trait evolution model; `loadings`: the loadings of the PCs; `varnames`: the names of the variables; `sig`: the MLE of the error; `mle.W`: the MLE of the W matrix; Function also returns AIC, AICc, and BIC for the model.

Value

returns a list of class "phylPPCA". See "Details" for more information.

References

- Revell, L. J. 2009. Size-Correction and Principal Components for Interspecific Comparative Studies. *Evolution* 63:3258–3268. doi: 10.1111/j.1558-5646.2009.00804.x
- Revell, L. J. 2024. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ* 12:e16505. doi: 10.7717/peerj.16505
- Garland Jr, T., Harvey, P.H. and Ives, A.R., 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Systematic biology*, 41(1), pp.18-32.
- Tipping, M. E., and C. M. Bishop. 1999. Probabilistic Principal Component Analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 61(3):611–622. doi: 10.1111/1467-9868.00196

Examples

```
phy <- ratematrix::anoles$phy[[1]]
dt <- as.matrix( ratematrix::anoles$data[,1:3] )
ppca <- picProbPCA(phy = phy, x = dt, ret_dim = 2)
doBiplot(x = ppca, add_margin = 0.3)
```

ProbPCA

Probabilistic PCA

Description

Function to perform (non-phylogenetic) probabilistic PCA.

Usage

```
ProbPCA(x, ret_dim = 2, scale = TRUE)
```

Arguments

x	a matrix with traits in columns and observations in rows.
ret_dim	number of dimensions (PC axes) to be kept by the model.
scale	if data should be rescaled to have zero mean and unit variance using the function <code>scale()</code> . Default is TRUE.

Details

Optimization is performed using `nloptr`.

Value

returns a list.

References

Tipping, M. E., and C. M. Bishop. 1999. Probabilistic Principal Component Analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 61(3):611–622. doi: 10.1111/1467-9868.00196

Examples

```
dt <- as.matrix( ratematrix::anoles$data[,1:3] )
ppca <- ProbPCA(x = dt, ret_dim = 2)
```

`traitModelAIC`*Compute the MLE and AIC scores for a model of trait evolution*

Description

Function will compute the parameter values and the AIC for a model of trait evolution. Models are estimated using a single parameter to model all traits in the multivariate data (x). The models are fit with the product of the likelihood of each independent trait fitted to a tree with the branch lengths transformed under the model M and a single model parameter p . The model parameter p represents an average of the process across all traits. The parameters for the models are: 1) λ for Pagel's λ , 2) α for OU (with θ as the root value), 3) b for the EB, and 4) σ for Brownian motion. Please note that fitting BM does not require transformation of branch lengths.

Usage

```
traitModelAIC(x, phy, model)
```

Arguments

<code>x</code>	a dataset with traits in columns and species on the rows. Please make sure that dataset matches the phylogeny.
<code>phy</code>	a phylogenetic tree with the same species as in <code>x</code> .
<code>model</code>	the name of a model of trait evolution. The options are "BM", "lambda", "OU", or "EB".

Value

a list with the elements: `model` (the model name), `lik` (the log-likelihood), `AIC` (the Akaike Information Criterion), `log_par` (the log of the parameter value), `nlopt_out` (the output object from nloptr optimizer). If the model is "BM", then the output has `R` (the evolutionary R matrix).

Index

BrokenStick, [2](#)

doBiplot, [3](#)

loglik, [4](#)

MCMCphylProbPCA, [5](#)

MCMCProbPCA, [7](#)

modelChoice, [8](#)

phylProbPCA, [9](#)

phylProbPCAJoint, [10](#)

picProbPCA, [11](#)

ProbPCA, [13](#)

traitModelAIC, [14](#)