Package 'EpiFusionUtilities'

December 12, 2024

Title Useful functions for preparing data for and parsing the output of EpiFusion models

Version 0.0.1

Description What the package does (one paragraph).

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.1

Suggests knitr, rmarkdown

VignetteBuilder rmarkdown

Imports stringr,
   HDInterval,
   ape,
   stableGR,
   dplyr,
   castor,
   xml2,
   tidyverse,
   tidyr,
   ggplot2,
   ggpubr

License GPL (>= 3) + file LICENSE

R topics documented:

1

Index **12**

extract_posterior_epifusion

Parse results from raw EpiFusion object into a R object with infection, Rt, cumulative infection and parameter posteriors

Description

This function discards the burn-in from the MCMC samples enclosed in a raw EpiFusion object. The posterior samples, and Rhat and ESS statistics for the parameters of the analysis, are stored in a list. Mean values, and HPD intervals for each trajectory type (infections, Rt, cumulative infections and, if applicable, modelled cases) and parameter are provided in the result. To save the posterior samples of each, set include_samples to TRUE (although doing this may made your object very large)

Usage

```
extract_posterior_epifusion(
  raw_epifusion,
  burn_in,
  discard_chains = NA,
  include_samples = FALSE
)
```

Arguments

raw_epifusion   raw EpiFusion object

burn_in   proportion (as a decimal) of each chain to discard as burn in

discard_chains   (optional) integer vector of chain IDs to discard in the posterior if there were non convergent chains

include_samples

(optional) boolean specifying whether to save the posterior samples in the output object (as opposed to just means and HPDs)

Value

a list of R compatible EpiFusion output objects with burnin discarded

---

```
generate_epifusion_XML
```
Generate an EpiFusion XML file

Description

This function takes a phylogenetic tree and/or case incidence data, in addition to some other information (analysis index date etc) and creates a parameter file for EpiFusion in eXtensible Markup Language.

Usage

```
generate_epifusion_XML(
  tree = NA,
  case_incidence = NA,
  index_date,
  loggers = NA,
  analysis = NA,
  model = NA,
  parameters = NA,
  priors = NA,
  xml_filepath = "input.xml"
)
```

Arguments

tree	phylogenetic tree (in S3 Phylo Object format) with branch lengths in terms of days, or path to a file with the formatted tree(s)
case_incidence	a data frame with case incidence and their date of occurrence. The data frame should consist of two columns: 'Date', a column with the observation dates, and 'Cases', a numeric column with reported cases on the corresponding day.
index_date	the date, in format, that you want to start modelling the outbreak from. Should be some time before
loggers	list of logging information (frequency, where to log the results files)
analysis	(optional) list of analysis instructions (e.g. how you wish to fit beta), if you wish to deviate from the default
model	(optional) list of model instructions (e.g. epidemiological observation model), if you wish to deviate from the default
parameters	(optional) list of parameters (e.g. how many MCMC steps), if you wish to deviate from the default
priors	(optional) list of prior distributions for the pMCMC
xml_filepath	filepath to write the complete XML file

---

generate_XML_chunk   *Generate an XML chunk from a list*

Description

Generate an XML chunk from a list

Usage

```
generate_XML_chunk(plist, node)
```

Arguments

plist	...
node	...

---

load_acceptance   *Load acceptance rates from EpiFusion folder into an R object*

Description

This function loads the acceptance rates for each chain of an EpiFusion analysis into a list of vectors.

Usage

```
load_acceptance(folderpath)
```

Arguments

folderpath	filepath to folder

Value

a list of vectors of acceptance rates of each chain

```
load_beta_trajectories
```
Load beta trajectories from EpiFusion output folder into a list of dataframes

---

Description

This function loads the beta trajectories of an EpiFusion analysis into a list of dataframes, with one dataframe for each chain in the analysis. The data frames are structured with ncol = number of days in the analysis, and nrow = number of MCMC samples.

Usage

```
load_beta_trajectories(folderpath)
```

Arguments

folderpath   filepath to folder

Value

a list of beta trajectory dataframes

```
load_cumulativeinfection_trajectories
```
Load cumulative infections trajectories from EpiFusion output folder into a list of dataframes

---

Description

This function loads the cumulative infection trajectories of an EpiFusion analysis into a list of dataframes, with one dataframe for each chain in the analysis. The data frames are structured with ncol = number of days in the analysis, and nrow = number of MCMC samples.

Usage

```
load_cumulativeinfection_trajectories(folderpath)
```

Arguments

folderpath   filepath to folder

Value

a list of cumulative trajectory dataframes

load_fitted_epi_cases   *Load trajectories of fitted epidemiological cases from EpiFusion output folder into a list of dataframes*

Description

This function loads the trajectories of fitted epidemiological cases from an EpiFusion analysis into a list of dataframes, with one dataframe for each chain in the analysis. The data frames are structured with ncol = number of epidemiological case data points, and nrow = number of MCMC samples.

Usage

```
load_fitted_epi_cases(folderpath)
```

Arguments

folderpath   filepath to folder

Value

a list of dataframes with fitted epidemiological case trajectories

load_infection_trajectories

Load infection trajectories from EpiFusion folder into a list of dataframes

Description

This function loads the infection trajectories of an EpiFusion analysis into a list of dataframes, with one dataframe for each chain in the analysis. The data frames are structured with ncol = number of days in the analysis, and nrow = number of MCMC samples.

Usage

```
load_infection_trajectories(folderpath)
```

Arguments

folderpath   filepath to folder

Value

a list of infection trajectory dataframes

| load_likelihoods | *Load likelihood trace from each chain of an EpiFusion analysis* |

Description

This function loads the likelihood traces of each chain of an EpiFusion analysis into a list of vectors, with one vector for each chain

Usage

```
load_likelihoods(folderpath)
```

Arguments

folderpath   filepath to folder

Value

a list of numeric vectors of likelihoods from an EpiFusion analysis

---

load_parameter_samples

Load parameter samples from EpiFusion analysis

Description

This function loads the parameter samples from an EpiFusion analysis into an R list, with one data frame per chain, with one column per parameter and one row per MCMC sample

Usage

```
load_parameter_samples(folderpath)
```

Arguments

folderpath   filepath to folder

Value

a list of data.frames with parameter samples per MCMC chain

---

load_raw_epifusion   *Load results from EpiFusion folder into an R object*

Description

This function loads the result of an EpiFusion folder into an R list, with the key characteristics and results of the analysis included.

Usage

```
load_raw_epifusion(folderpath)
```

Arguments

folderpath   filepath to EpiFusion output folder

Value

a list of R compatible EpiFusion output objects

---

load_rt_trajectories   *Load Rt trajectories from EpiFusion folder into a list of dataframes*

Description

This function loads the Rt trajectories of an EpiFusion analysis into a list of dataframes, with one dataframe for each chain in the analysis. The data frames are structured with ncol = number of days in the analysis, and nrow = number of MCMC samples. The Rt is calculated using the sampled beta trajectory and gamma value from each MCMC step.

Usage

```
load_rt_trajectories(folderpath)
```

Arguments

folderpath   filepath to folder

Value

a list of Rt trajectory dataframes

---

lshtm_theme *LSHTM Plotting Theme*

Description

Th

Usage

```
lshtm_theme()
```

---

plot_likelihood_trace *Plots the likelihood trace of the chains of an EpiFusion analysis*

Description

This function plots the likelihood trace of the chains of an EpiFusion analysis to assess convergence and help to select the burn-in proportion.

Usage

```
plot_likelihood_trace(raw_epifusion)
```

Arguments

raw_epifusion   raw epifusion object

---

plot_parameter_trace *Plot the parameter traces from a raw EpiFusion object*

Description

T

Usage

```
plot_parameter_trace(raw_epifusion)
```

Arguments

raw_epifusion   raw epifusion object

---

plot_trajectories   *Plot the trajectory posteriors from a parsed EpiFusion trajectory table*

Description

Plots the trajectory posteriors from an EpiFusion trajectory table. You can specify the trajectory type (infection, Rt, cumulative infection) using the type parameter, or the function will plot each type by default. You may also specify any number of colours using the plot_colours parameter.

Usage

```
plot_trajectories(trajectory_table, type = NA, plot_colours = NA)
```

Arguments

trajectory_table

a trajectory table extracted from an EpiFusion posterior object

type   (optional) "infection" "rt" or "cumulativeinfection"

plot_colours   (optional) a vector of colours to be used in the plots

---

prepare_epifusion_tree

Prepare a tree (or trees) for EpiFusion

Description

This function prepares a tree or tree posterior, given as an S3 phylo object, for EpiFusion analysis by adding 'time during outbreak' to the node and leaf labels. You must provide an index date (in date form) which is day 0 of your analysis, i.e. the earliest point from which you will model trajectories.

Usage

```
prepare_epifusion_tree(
  tree,
  index_date,
  final_sequence_date,
  treefile_path = "processedtree.tree"
)
```

Arguments

tree   phylogenetic tree or tree posterior with branch lengths in terms of years (in S3
                Phylo Object format)

index_date   analysis index date (day 0 of your analysis)

final_sequence_date

                date of sampling of the final sequence in the tree or tree posterior

treefile_path   path to the file where the processed tree should be stored (default processedtree.tree)

Value

a tree (in S3 Phylo Object format) with the node and leaf labels in terms of time from the index date. If a tree posterior is passed to the function, nothing is returned to the function and the trees are simply written to the `treefile_path`

---

run_epifusion	*Run EpiFusion from R*

Description

This function runs EpiFusion from R, assuming you have java installed and callable from your system using the `java` command.

Usage

```
run_epifusion(xml_filepath, output_folder_name = NA)
```

Arguments

xml_filepath   path to the xml input file

output_folder_name

                the name of the folder you'd like created for the output

---

trajectory_table	*Extract a trajectory table from an EpiFusion posterior R object*

Description

This function extracts a table with the mean and HPD intervals of infection, Rt and cumulative infection trajectories from an EpiFusion posterior R object. You can provide your analysis index date if you want to have your 'Time' column in terms of dates.

Usage

```
trajectory_table(epifusion_posterior, start_date = NA)
```

Arguments

epifusion_posterior

                an extracted EpiFusion posterior object made with the extract_posterior_epifusion() function

start_date   (optional) date you'd like the time series to count from, as a date object or as a string with format YYYY-MM-DD

Value

a table with a column for Time and a column for Mean, Upper and Lower 95%, 88% and 66% HPDs for inferred Infections and Rt

Index