Package 'staRdom'

June 21, 2018

Type Package

Title PARAFAC Analysis of EEMs from DOM

Version 1.0.8

Date 2018-06-07

Depends R (>= 3.3), dplyr (>= 0.7.4), tidyr (>= 0.7.1), ggplot2 (>= 2.2.1), eemR (>= 0.1.5)

Description This is a user-friendly way to run a parallel factor (PARAFAC) analysis (Harshman, 1971) <doi:10.1121/1.1977523> on excitation emission matrix (EEM) data from dissolved organic matter (DOM) samples (Murphy et al., 2013) <doi:10.1039/c3ay41160e>. The analysis includes profound methods for model validation. Some additional functions allow the calculation of absorbance slope parameters and create beautiful plots.

License AGPL

Encoding UTF-8

LazyData true

Imports stringr (>= 1.2.0), pracma (>= 2.1.1), tools (>= 3.3), readr (>= 1.1.1), zoo (>= 1.8), tibble (>= 1.3), multiway (>= 1.0-5), plotly (>= 4.7), parallel (>= 3.4), GGally (>= 1.3), graphics (>= 3.3), doParallel (>= 1.0.11), drc (>= 3.0-1), foreach (>= 1.4.4), data.table (>= 1.11.2), matrixStats (>= 0.53.1)

Suggests xlsx, knitr, kableExtra, rmarkdown, knitcitations

RoxygenNote 6.0.1

VignetteBuilder knitr

URL https://github.com/MatthiasPucher/staRdom

BugReports https://github.com/MatthiasPucher/staRdom/issues

NeedsCompilation no

Author Matthias Pucher [aut, cre],
Daniel Graeber [aut, ctb],
Stefan Preiner [ctb],
Renata Pinto [ctb],
Nora Maria Zechmeister [ctb],
Christoph Zechmeister [ctb]

Maintainer Matthias Pucher <matthias.pucher@wcl.ac.at>

Repository CRAN

Date/Publication 2018-06-21 13:59:05 UTC

R topics documented:

`absorbance_read` *Reading absorbance data from txt and csv files.*

Description

Reading absorbance data from txt and csv files.

Usage

```
absorbance_read(absorbance_path, order = TRUE, recursive = TRUE,
  dec = NULL, ...)
```

Arguments

absorbance_path	directory containing absorbance data files or path to single file. See details for format of absorbance data.
order	logical, data is ordered according to wavelength
recursive	read files recursive, include subfolders
dec	optional, either you set a decimal separator or the table is tested for . and ,
...	additional arguments that are passed on to fread.

Details

If absorbance_path is a directory, contained files that end on "csv" or "txt" are passed on to read.table. If the path goes to a file, this file is passed on. Tables can either contain data from one sample or from several samples in columns. The column header containig the wavelength must be either "wavelength" or "Wavelength". A multi-sample file must have sample names as column names. A single-sample file can have sample name as column name or sample name as file name and "Abs." as column name. All tables are combined to one with one wavelength column and one column for each sample containing the absorbance data.

Value

A data frame containing absorbance data. An attribute "location" contains the filenames where each sample was taken from.

See Also

fread

Examples

```
absorbance_path <- system.file("extdata", "absorbance_eemR", package = "staRdom")
absorbance_read(absorbance_path)
```

---

abs_data *raw data from absorbance measurements, 5cm pathlength*

Description

raw data from absorbance measurements, 5cm pathlength

Usage

```
abs_data
```

Format

list of parafacs

abs_fit_slope *Fit absorbance data to exponential curve. drm is used for the fitting process.*

Description

Fit absorbance data to exponential curve. drm is used for the fitting process.

Usage

```
abs_fit_slope(wl, abs, lim, l_ref = 350, control = drmc(errorm = FALSE,
  noMessage = TRUE), ...)
```

Arguments

wl	vector containing wavelengths
abs	vector containing absorption in m^-1
lim	vector containing lower and upper limits for wavelengths to use
l_ref	numerical. reference wavelength, default is 350, if set to NA l_ref is fitted
control	control parameters for drm, see drmc
...	parameters that are passed on to drm

Value

numeric exponential slope coefficient

See Also

drm

Examples

```
data(abs_data)
abs_fit_slope(abs_data$wavelength,abs_data$sample1,lim=c(350,400),l_ref=350)
```

abs_parms *Calculating slopes and slope ratios of a data frame of absorbance data.*

Description

Calculating slopes and slope ratios of a data frame of absorbance data.

Usage

```
abs_parms(abs_data, cuvle, limits = list(c(275, 295), c(350, 400), c(300,
  700)), l_ref = list(350, 350, 350), S = TRUE, lref = FALSE, p = FALSE,
  model = FALSE, cores = parallel::detectCores()/2)
```

Arguments

abs_data	data frame containing absorbance data.
cuvle	length of used cuvette in cm
limits	list with vectors containig upper and lower bounds of wavelengeth ranges to be fitted
l_ref	list with reference wavelengths, same length as limits
S	logical, include slope parameter in the table
lref	logical, include reference wavelength in the table
p	logical, include ps of the coefficients in the table
model	logical, include complete model in data frame
cores	number of cores to be used for parallel processing

Details

The absorbance data is a data frame with the first column called "wavelength" containg the wavelength. Each other column contains the data from one sample. You can use absorbance_read to read in appropriate data.

The following spectral parameters are calculated:

- $S_275-295$ slope between 275 and 295 nm calculated with nonlinear regression
- $S_350-400$ slope between 350 and 400 nm calculated with nonlinear regression
- $S_300-700$ slope between 275 and 295 nm calculated with nonlinear regression
- SR slope ratio, calculated by $S_275-295$/$S_350-400$
- E2:E3 ratio a_250/a_365
- E4:E6 ratio a_465/a_665
- a_254 absorbance at 254 nm
- a_300 absorbance at 300 nm

Depending on available wavelength range, values might be NA. Additionally other wavelength limits can be defined. The slope ratio might fail in this case. For further details please refer to Helm et al. (2008).

Value

A data frame containing the adsorption slopes and slope ratios in column, one line for each sample.

References

Helms, J., Kieber, D., Mopper, K. 2008. Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. Limnol. Oceanogr., 53(3), 955–969 http://onlinelibrary.wiley.com/doi/10.4319/lo.2008.53.3.0955/pdf

Examples

```
data(abs_data)
abs_parms(abs_data[,1:5],5)
abs_parms(abs_data[,1:5],5,l_ref=list(NA,NA,NA), lref=TRUE) # fit lref as well
```

`as.data.frame.eem` *Converting EEM data from class eem to data.frame.*

Description

Converting EEM data from class eem to data.frame.

Usage

```
## S3 method for class 'eem'
as.data.frame(x, row.names = NULL, optional = FALSE, ...,
  gather = TRUE)
```

Arguments

`x`	blabla
`row.names`	asfas
`optional`	ignored
`...`	ignored
`gather`	logical, says whether data.frame is returned with excitation wavelength as column names or as values of a column. If the data is gathered, the sample name is added as value in a calumn

Value

A data frame containing the EEM data.

Examples

```
data(eem_list)
as.data.frame(eem_list[[1]])
as.data.frame(eem_list[[1]],gather=FALSE)
```

`A_missing` *Calculate the amount of each component for samples not involved in model building*

Description

Samples from an eemlist that were not used in the modelling process are added as entries in the A-modes. Values are calculated using fixed B and C modes in the PARAFAC algorithm.

Usage

```
A_missing(eem_list, pfmodel, cores = parallel::detectCores(logical = FALSE)/2,
  ...)
```

Arguments

`eem_list`	object of class eemlist with sample data
`pfmodel`	object of class parafac
`cores`	number of cores to use for parallel processing
`...`	additional arguments passed to eem_parafac

Value

object of class parafac

Examples

```
data(eem_list)
data(pfres_comps2)

A_missing(eem_list,pfres_comps2[[3]])
```

blank *blank samples*

Description

blank samples

Usage

```
blank
```

Format

eemlist

eem2array *Data from an eemlist is transformed into an array*

Description

Data matrices from EEM are combined to an array that is needed for a PARAFAC analysis.

Usage

```
eem2array(eem_list)
```

Arguments

eem_list	object of class eemlist

Value

object of class array

Examples

```
data(eem_list)

eem2array(eem_list)
```

`eempf4analysis` *Create table of PARAFAC components and (optionally) EEM peaks and indices as well as absorbance slope parameters.*

Description

Please refer to `eem_biological_index`, `eem_coble_peaks`, `eem_fluorescence_index`, `eem_biological_index` and `abs_parms` for details on the certain values

Usage

```
eempf4analysis(pfmodel, eem_list = NULL, absorbance = NULL, cuvl = NULL,
  n = 4, export = NULL, ...)
```

Arguments

`pfmodel`	PARAFAC model where loadings of the components are extracted
`eem_list`	optional eemlist used for peak and indices calculation
`absorbance`	optional absorbance table used for absorbance slope parameter calculation
`cuvl`	optional cuvette length of absorbance data in cm
`n`	optional size of moving window in nm for data smoothing in advance of peak picking
`export`	optional file path of csv or txt table where data is exported
`...`	additional parameters passed to `write.table`

Value

data frame

Examples

```
data(eem_list)
data(pfres_comps2)
data(abs_data)

results <- eempf4analysis(pfmodel = pfres_comps2[[2]],
                          eem_list = eem_list, absorbance = abs_data,
                          cuvl = 5, n = 4)
```

`eempf_compare` *Plot a set of PARAFAC models to compare the single components*

Description

Three plots are returned:

1. plot of numer of components vs. model fit
2. plot of different components as colour maps
3. plot of different components as peak lines

The plots are intended to help with a suitable number of components.

Usage

```
eempf_compare(pfres)
```

Arguments

`pfres`	list of several objects of class parafac

Value

3 objects of class ggplot

See Also

`eempf_fits`, `eempf_plot_comps`

Examples

```
data(pfres_comps1)

eempf_compare(pfres_comps)
```

eempf_comps3D *3D plots of PARAFAC components*

Description

Interactive 3D plots are created using plotly.

Usage

```
eempf_comps3D(pfmodel, which = NULL)
```

Arguments

pfmodel	object of class parafac
which	optional, if numeric selects certain component

Value

plotly plot

Examples

```
## Not run:
data(pfres_comps1)

eempf_comps3D(pfres_comps[[3]])

## End(Not run)
```

eempf_comp_load_plot *Plot components from a PARAFAC model*

Description

Additionally a bar plot with the amounts of each component in each sample is produced.

Usage

```
eempf_comp_load_plot(pfmodel)
```

Arguments

pfmodel	object of class parafac

Value

ggplot

See Also

ggeem, eempf_load_plot

Examples

```
data(pfres_comps1)

eempf_comp_load_plot(pfres_comps[[2]])
```

`eempf_comp_mat` *Extract EEM matrix for single components determined in the PARAFAC analysis*

Description

The components of a PARAFAC analysis are extracted as a data frame

Usage

```
eempf_comp_mat(pfmodel, gather = TRUE)
```

Arguments

`pfmodel`	object of class parafac
`gather`	logical value whether excitation wavelengths are a column, otherwise excitation wavelengths are column names

Value

a list of class data frames

Examples

```
data(pfres_comps1)

eempf_comp_mat(pfres_comps[[3]])
```

eempf_corcondia *Calculate the core consistancy of an EEM PARAFAC model*

Description

This is basically a wrapper for corcondia that deals with the normalisation of the original data., Other than corcondia, the default dicisor = "core".

Usage

```
eempf_corcondia(pfmodel, eem_list, divisor = "core")
```

Arguments

pfmodel	PARAFAC model
eem_list	eemlist
divisor	divisor, please refer to corcondia

Value

numeric

Examples

```
## Not run:
# due to data limitation in package, example does not work with that data!

eempf_corecondia(pfmodel,eem_list)


## End(Not run)
```

eempf_corplot *Plot correlations of components in samples*

Description

A pair plot showing correlations between samples is created.

Usage

```
eempf_corplot(pfmodel, normalisation = FALSE, lower = list(continuous =
  "smooth"), mapping = aes(alpha = 0.2), ...)
```

Arguments

pfmodel	object of class parafac
normalisation	logical, whether normalisation is undone or not
lower	style of lower plots, see ggpairs
mapping	aesthetic mapping, see ggpairs
...	passed on to ggpairs

Value

object of class ggplot

See Also

ggpairs

Examples

```
data(pfres_comps1)
eempf_corplot(pfres_comps[[1]])
```

---

eempf_cortable *Calculating correlations between the component loadings in all samples (C-Modes).*

Description

Calculating correlations between the component loadings in all samples (C-Modes).

Usage

```
eempf_cortable(pfmodel, normalisation = FALSE, method = "pearson", ...)
```

Arguments

pfmodel	results from a PARAFAC analysis, class parafac
normalisation	logical, whether normalisation is undone or not
method	method of correlation, passed to cor
...	passed on to cor

Value

matrix

Examples

```
data(pfres_comps1)
eempf_cortable(pfres_comps[[2]])
```

`eempf_eemqual` *Calculating EEMqual which is an indicator of a PARAFAC model's quality*

Description

Calculating EEMqual which is an indicator of a PARAFAC model's quality

Usage

```
eempf_eemqual(pfmodel, eem_list, splithalf = NULL, ...)
```

Arguments

`pfmodel`	PARAFAC model
`eem_list`	EEM data as eemlist
`splithalf`	optionally, you can supplie available splithalf results from model to decrease computation time
`...`	additional arguments passed to splithalf

Value

data frame containing fit, corcondia, product of best TCCs from splithalf analysis, eemqual and splithalf models

References

Rasmus Bro, Maider Vidal, EEMizer: Automated modeling of fluorescence EEM data, Chemometrics and Intelligent Laboratory Systems, Volume 106, Issue 1, 2011, Pages 86-92, ISSN 0169-7439

Examples

```
data(eem_list)
data(pfres_comps2)

pfmodel <- pfres_comps2[[2]]
eempf_eemqual(eem_list,pfmodel)
```

eempf_export — *Create one table containing the PARAFAC models factors and optionally exporting it to csv or txt*

Description

Create one table containing the PARAFAC models factors and optionally exporting it to csv or txt

Usage

```
eempf_export(pfmodel, export = NULL, ...)
```

Arguments

pfmodel	PARAFAC model
export	file path to export table
...	additional parameters passed to write.table

Value

data frame

Examples

```
data(pfres_comps2)

factor_table <- eempf_export(pfres_comps2[[2]])
```

eempf_fits — *Fits vs. components of PARAFAC models are plotted*

Description

Fits vs. components of PARAFAC models are plotted

Usage

```
eempf_fits(pfres)
```

Arguments

pfres	list of objects of class parafac

Value

object of class ggplot

Examples

```
data(pfres_comps1)

eempf_fits(pfres_comps)
```

`eempf_leverage` *Calculate the leverage of each emission and excitation wavelength and each sample from a single PARAFAC model*

Description

Calculate the leverage of each emission and excitation wavelength and each sample from a single PARAFAC model

Usage

```
eempf_leverage(pfmodel)
```

Arguments

`pfmodel`	object of class parafac

Value

list of 3 named vectors (emission, excitation wavelengths and samples)

Examples

```
data(pfres_comps1)

eempf_leverage(pfres_comps[[2]])
```

`eempf_leverage_data` *Combine leverages into one data frame and add optional labels.*

Description

Combine leverages into one data frame and add optional labels.

Usage

```
eempf_leverage_data(cpl, qlabel = 0.1)
```

Arguments

`cpl`	leverage, outpout from `eempf_leverage`
`qlabel`	optional, quantile of which labels are shown (1 = all, 0 = no labels)

Value

data frame

Examples

```
data(pfres_comps1)

leverage <- eempf_leverage(pfres_comps[[2]])
lev_data <- eempf_leverage_data(leverage)
```

`eempf_leverage_ident` *Plot leverage of emission wavelengths, excitation wavelengths and samples.*

Description

Plot is interactive where you can select values with your mouse. A list of vectors is returned to remove this outliers in a further step from your samples. The labels to be shown can be selected by adding the quatile of samples with highest leverages to be labeled.

Usage

```
eempf_leverage_ident(cpl, qlabel = 0.1)
```

Arguments

`cpl`	leverage, outpout from `eempf_leverage`
`qlabel`	optional, quantile of which labels are shown (1 = all, 0 = no labels)

Value

list of three vectors containing the names of selected samples

See Also

`eempf_leverage_plot`

Examples

```
data(pfres_comps1)

leverage <- eempf_leverage(pfres_comps[[2]])
outliers <- eempf_leverage_ident(leverage)
```

eempf_leverage_plot *Plot leverage of emission wavelengths, excitation wavelengths and samples.*

Description

The labels to be shown can be selected by adding the quatile of samples with highest leverages to be labeled.

Usage

```
eempf_leverage_plot(cpl, qlabel = 0.1)
```

Arguments

cpl	leverage, outpout from eempf_leverage
qlabel	optional, quantile of which labels are shown (1 = all, 0 = no labels)

Value

ggplot

See Also

eempf_leverage_ident

Examples

```
data(pfres_comps1)

leverage <- eempf_leverage(pfres_comps[[3]])
eempf_leverage_plot(leverage)
```

eempf_load_plot *Plot amount of each component in each sample as bar plot*

Description

Plot amount of each component in each sample as bar plot

Usage

```
eempf_load_plot(pfmodel)
```

Arguments

pfmodel	parafac model

Value

ggplot

Examples

```
data(pfres_comps1)

eempf_load_plot(pfres_comps[[2]])
```

`eempf_mleverage` *Calculate the leverage of each emission and excitation wavelength and each sample from a list of PARAFAC models*

Description

Calculate the leverage of each emission and excitation wavelength and each sample from a list of PARAFAC models

Usage

```
eempf_mleverage(pfres_comps, ecdf = FALSE, stats = FALSE)
```

Arguments

`pfres_comps`	object of class parafac
`ecdf`	logical, transforme leverages to according empirical quantiles (`ecdf`)
`stats`	logical, whether means and standard deviations are calculated from leverages

Value

data frame containing leverages of wavelengths and samples for each model

Examples

```
data(pfres_comps1)

eempf_mleverage(pfres_comps)
```

eempf_openfluor *Write out PARAFAC components to submit to openfluor.org.*

Description

openfluor.org offers the possibility to compare your results to others, that were uploaded to the database. This functions writes out a txt containing the header lines and your components. Please open the file in an editor and fill in further information that cannot be covered by this function.

Usage

```
eempf_openfluor(pfmodel, file)
```

Arguments

pfmodel	PARAFAC model
file	string, path to outputfile. The directory must exist, the file will be created or overwritten if already present.

Value

txt file

Examples

```
data(pfres_comps2)
eempf_openfluor(pfres_comps2[[2]],file.path(tempdir(),"openfluor_example.txt"))
```

eempf_plot_comps *Plot all components of PARAFAC models*

Description

The components can be plottet in two ways: either as a colour map or as two lines (emission, excitation wavelengths) intersecting at the component maximum.

Usage

```
eempf_plot_comps(pfres, type = 1)
```

Arguments

pfres	list of PARAFAC models
type	1 for a colour map and 2 for peak lines

Value

object of class ggplot

Examples

```
data(pfres_comps1)

eempf_plot_comps(pfres_comps)
eempf_plot_comps(pfres_comps,type=2)
```

`eempf_report` *Create a html report of a PARAFAC analysis*

Description

Create a html report of a PARAFAC analysis

Usage

```
eempf_report(pfmodel, export, eem_list = NULL, absorbance = NULL,
  meta = NULL, metacolumns = NULL, splithalf = FALSE, shmodel = NULL,
  performance = FALSE, residuals = FALSE, spp = 5, ...)
```

Arguments

`pfmodel`	PARAFAC model
`export`	path to exported html file
`eem_list`	optional EEM data
`absorbance`	optional absorbance data
`meta`	optional meta data table
`metacolumns`	optional column names of metadata table
`splithalf`	optional logical, states whether split-half analysis should be included
`shmodel`	optional results from split-half analysis. If this data is not supplied but EEM data is available the split-half analysis is calculated on the creation of the report. Calculating the split-half analysis takes some time!
`performance`	calculating model performance: `eempf_eemqual`
`residuals`	logical, whether residuals are plotted in the report
`spp`	plots per page for loadgins and residuals plot
`...`	arguments to or from other functions

Value

TRUE if report was created

Examples

```
## no test yet
```

`eempf_rescaleBC` *Rescale B and C modes of PARAFAC model*

Description

B and C modes (emission and excitation wavelengths) are rescaled to RMS of value newscale. This is compensated in A mode (sample loadings).

Usage

```
eempf_rescaleBC(pfmodel, newscale = 1)
```

Arguments

`pfmodel`	object of class parafac
`newscale`	Desired root mean-square for each column of rescaled mode. Can input a scalar or a vector with length equal to the number of factors for the given mode. If newscale = "Fmax", each component will be scaled so the maximum of each component is 1.

Value

object of class parafac

See Also

`rescale`

Examples

```
data(pfres_comps1)

new_pf <- eempf_rescaleBC(pfres_comps[[2]])
```

eempf_residuals *Calculate residuals of EEM data according to a certain model*

Description

Calculate residuals of EEM data according to a certain model

Usage

```
eempf_residuals(pfmodel, eem_list, select = NULL,
  cores = parallel::detectCores(logical = FALSE)/2)
```

Arguments

pfmodel	PARAFAC model of class parafac
eem_list	eemlist containing EEM data
select	character vector containing the names of the desired samples
cores	number of cores to use for parallel processing

Value

data frame with EEM residuals

Examples

```
data(eem_list)
data(pfres_comps2)

eempf_residuals(pfres_comps2[[2]],eem_list)
```

eempf_residuals_plot *Plot samples by means of whole sample, each single component and residuum*

Description

A raster of plots is created. Each column shows one sample. The top n rows show the n components from the model according their occurance in the certain samples. The second last row shows the residual, not covered by any component in the model and the last row shows the whole sample.

Usage

```
eempf_residuals_plot(pfmodel, eem_list, res_data = NULL, spp = 5,
  select = NULL, residuals_only = FALSE,
  cores = parallel::detectCores(logical = FALSE)/2)
```

Arguments

pfmodel	object of class parafac containing the generated model
eem_list	object of class eemlist with all the samples that should be plotted
res_data	optional, data of sample residuals related to the model, output from eempf_residuals
spp	optional, samples per plot
select	optional, character vector of samples you want to plot
residuals_only	plot only residuals
cores	number of cores to use for parallel processing

Details

eem_list may contain samples not used for modelling. Calculation is done by A_missing. This especially interesting if outliers are excluded prior modelling and should be evaluated again afterwards.

Value

several ggplot objects

Examples

```
data(eem_list)
data(pfres_comps1)

eempf_residuals_plot(pfres_comps[[3]],eem_list)
```

---

eem_absdil *Multiply absorbance data according to the dilution and remove absorbance from samples where undiluted data is used.*

Description

According to dilution data absorbance is either multiplied by the according factor or the undiluted absorbance data is deleted. You can either specify the cor_data data table coming from eem_dilcorr or supply an eemlist, and the dilution data to created on the fly.

Usage

```
eem_absdil(abs_data, eem_list = NULL, dilution = NULL, cor_data = NULL,
  auto = TRUE, verbose = FALSE)
```

Arguments

abs_data	absorbance data
eem_list	optional eemlist
dilution	optional dilution data as data frame
cor_data	optional output from eem_dilcorr as data frame
auto	optional, see eem_dilcorr
verbose	optional, see eem_dilcorr

Value

data frame

Examples

```
data(eem_list)
data(abs_data)

abs_data <- abs_data[1:31]

dilution <- data.frame(dilution = c(rep(1,10),rep(5,10),rep(10,10)))
rownames(dilution) <- eem_names(eem_list)

abs_data2 <- eem_absdil(abs_data,eem_list,dilution)
```

---

eem_checkdata *Check your EEM, absorption and metadata before processing*

Description

The function tries to lead you to possible problems in your data.

Usage

```
eem_checkdata(eem_list, absorbance, metadata = NULL, metacolumns = NULL,
  error = TRUE)
```

Arguments

eem_list	eemlist continaing EEM data.
absorbance	data.frame containing absorbance data.
metadata	optional data.frame containing metadata.
metacolumns	character vector of columns that are checkt for complete data sets
error	logical, whether a problem should cause an error or not.

Details

The returned list contains character vectors with sample names where possible problems were found: problem (logical, whether a severe problem was found), nas (sample names with NAs in EEM data), missing_correction (correction of EEM samples was not done or not done successfully),eem_no_abs (EEM samples with no absorbance data), abs_no_eem (samples with present absorbance but no EEM data), duplse (duplicate sample names in EEM data), duplsa (duplicate sample names in absorbance data), invalid_eem (invalid EEM sample name), invalid_abs (invalid absorbance sample name), range_mismatch (wavelength ranges of EEM and absorbance data are mismatching), metadupls (duplicate sample names in metadata), metamissing (EEM samples where metadata is missing), metaadd (samples in metadata without EEM data)

Value

writes out possible porblems to command line, additionally list with sample names where possible problems were found, see details.

Examples

```
data(eem_list)
data(abs_data)

eem_checkdata(eem_list,abs_data)
```

---

eem_corrections *Return names of samples where certain corrections are missing.*

Description

Return names of samples where certain corrections are missing.

Usage

```
eem_corrections(eem_list)
```

Arguments

eem_list	eemlist to be checked

Value

prints out sample names

Examples

```
data(eem_list)

eem_corrections(eem_list)
```

`eem_dilcorr` *Create table how samples should be corrected because of dilution*

Description

Due to dilution absorbance spectra need to be multiplied by the dilution factor and names of EEM samples can be adjusted to be similar to their undiluted absorbance sample. The table contains information about these two steps. Undiluted samples are suggested by finding absorbance samples match the beginning of EEM sample name (see details).

Usage

```
eem_dilcorr(eem_list, abs_data, dilution, auto = FALSE, verbose = TRUE)
```

Arguments

`eem_list`	eemlist
`abs_data`	absorbance data as data frame
`dilution`	dilution data as data frame with rownames
`auto`	way how to deal with dilution is chosen automatically. See details.
`verbose`	print out more information

Details

If you choose an automatic analysis EEMs are renamed if there is only one matching undiluted absorbance sample. Matching samples is done by comparing the beginning of the sample name (e.g. "sample3_1to10" fits "sample3").

Value

data frame

Examples

```
data(eem_list)
data(abs_data)

abs_data <- abs_data[1:31]

dilution <- data.frame(dilution = c(rep(1,10),rep(5,10),rep(10,10)))
rownames(dilution) <- eem_names(eem_list)

dilcorr <- eem_dilcorr(eem_list,abs_data,dilution, auto = TRUE, verbose = FALSE)
```

eem_dilution *Modifying fluorescence data according to dilution.*

Description

If samples were diluted before measuring, a dilution factor has to be added to the measured data. This function can do that by either multilpying each sample with the same value or using a data frame with different values for each sample.

Usage

```
eem_dilution(data, dilution = 1)
```

Arguments

data	fluorescence data with class eemlist
dilution	dilution factor(s), either numeric value or data frame. Row names of data frame have to be similar to sample names in eemlist.

Value

fluorescence data with class eemlist

Examples

```
data(eem_list)

eem_list <- eem_dilution(eem_list,dilution=5)
```

eem_duplicates *Check for duplicate sample names*

Description

Check for duplicate sample names

Usage

```
eem_duplicates(data)

## Default S3 method:
eem_duplicates(data)

## S3 method for class 'eemlist'
eem_duplicates(data)

## S3 method for class 'data.frame'
eem_duplicates(data)
```

Arguments

data	eemlist or data.frame containing absorbance data

Value

named character vector with duplicate sample names

Examples

```
### check
```

---

eem_easy — *Opens an R markdown template for an easy and userfriendly analysis of EEM data.*

Description

In your default editor (e.g. RStudio), a Rmd file is opened. It consists of bloacks gathering the parameters and information needed and continues with a series of data corrections, peak picking and plots. Finally you get a report of your analysis, a table with the peaks and optional pngs of your fluorescence data. To continue working and keeping your settings, the file can be sa ved anywhere and reused anytime.

Usage

```
eem_easy()
```

Details

Function does not work well in Windows. You might try file.edit(system.file("EEM_simple_analysis.Rmd", package = "staRdom"))

Value

A pdf report, a peak picking table and optional plots.

Examples

```
## Not run:
#
eem_easy()

# this function fails very often, so you might use that:
file.edit(system.file("EEM_simple_analysis.Rmd", package = "staRdom"))

## End(Not run)
```

eem_eemdil *Correct names of EEM samples to match undiluted absorbance data.*

Description

Correct names of EEM samples to match undiluted absorbance data.

Usage

```
eem_eemdil(eem_list, abs_data = NULL, dilution = NULL, cor_data = NULL,
  auto = TRUE, verbose = FALSE)
```

Arguments

eem_list	eemlist
abs_data	optinal absorbance data as data frame
dilution	optinal dilution data as data frame
cor_data	optional output from eem_dilcorr as data frame
auto	optional, see eem_dilcorr
verbose	optional, see eem_dilcorr

Value

eemlist

Examples

```
data(eem_list)
data(abs_data)

dilution <- data.frame(dilution = c(rep(1,10),rep(5,10),rep(10,10)))
rownames(dilution) <- eem_names(eem_list)

eem_list2 <- eem_eemdil(eem_list,abs_data,dilution)
```

`eem_exclude` *Exclude complete wavelengths or samples form data set*

Description

Outliers in all modes should be avoided. With this functions excitation or emission wavelengths as well as samples can be removed completely from your sample set.

Usage

```
eem_exclude(eem_list, exclude = list, verbose = FALSE)
```

Arguments

`eem_list`	object of class eemlist
`exclude`	list of three vectors, see details
`verbose`	states whether additional information is given in the command line

Details

The argument exclude is a named list of three vectors. The names must be "ex", "em" and "sample". Each element contains a vector of wavelengths or sample names that are to be excluded from the data set.

Value

object of class eemlist

Examples

```
data(eem_list)

exclude <- list("ex" = c(280,285,290,295),
"em" = c(),
"sample" = c("sample3","sample5","sample14")
)

eem_list_ex <- eem_exclude(eem_list, exclude)
```

eem_getextreme — *Determines the the biggest range of EEM spectrum where data is available from each sample.*

Description

Determines the the biggest range of EEM spectrum where data is available from each sample.

Usage

```
eem_getextreme(data)
```

Arguments

data	eemlist

Value

list of numeric vector containing the biggest available range

Examples

```
data(eem_list)
eem_getextreme(eem_list)

eem_list <- eem_range(eem_list,ex = c(250,Inf),em = c(280,500))
eem_getextreme(eem_list)
```

eem_ife_correction — *Wrapper function to allow eem_inner_filter_effect (eemR) handling different cuvette lengths.*

Description

Calls eem_inner_filter_effect for each sample to use different cuvette lengths.

Usage

```
eem_ife_correction(data, abs_data, cuvl)
```

Arguments

data	fluorescence data of class eemlist
abs_data	absorbance data
cuvl	length of cuvette of absorption measurment in cm. Either a number or a data frame. Row names of data frame have to be similar to sample names in data

Value

fluorescence data of class eemlist

Examples

```
data(eem_list)
data(abs_data)

eem_ife_correction(eem_list,abs_data,5)
```

`eem_import_dir` *Load all eemlist obects saved in different Rdata files in folder*

Description

Reads Rdata files with one eemlist each. eemlists are combined into one and returned.

Usage

```
eem_import_dir(dir)
```

Arguments

`dir`	folder where RData files are saved

Value

eemlist

Examples

```
## Not run:
# due to package size issues no example data is provided for this function
eem_import_dir("C:/some_folder/with_EEMS/only_Rdata_files")

## End(Not run)
```

eem_interp *Missing values are interpolated within EEM data*

Description

Missing EEM data can be interpolated. Usually it is the result of removing scatter. It is done along each excitation wavelength. This step is recommended if you aim for a PARAFAC analysis. Interpolation is done with hermitean interpolation polynomials using pchip.

Usage

```
eem_interp(data, cores = detectCores(logical = FALSE)/2)
```

Arguments

data	object of class eemlist with spectra containing missing values
cores	specify number of cores for parallel computation

Value

object of class eemlist with interpoleted spectra.

References

Elcoroaristizabal, S., Bro, R., García, J., Alonso, L. 2015. PARAFAC models of fluorescence data with scattering: A comparative study. Chemometrics and Intelligent Laboratory Systems, 142, 124-130 https://doi.org/10.1016/j.chemolab.2015.01.017

See Also

pchip

Examples

```
data(eem_list)

remove_scatter <- c()
remove_scatter["raman1"] = TRUE
remove_scatter["raman2"] = TRUE
remove_scatter["rayleigh1"] = TRUE
remove_scatter["rayleigh2"] = TRUE
remove_scatter_width = c(15,10,16,12)

eem_list <- eem_rem_scat(eem_list,remove_scatter,remove_scatter_width)

eem_list <- eem_interp(eem_list)
```

eem_is.na *Check for NAs in EEM data*

Description

Check for NAs in EEM data

Usage

```
eem_is.na(eem_list)
```

Arguments

eem_list	eemlist to check

Value

named character vector with sample names where EEM data contains NAs

Examples

```
### check
```

eem_list *Fluorescence data of 52 corrected samples.*

Description

Fluorescence data of 52 corrected samples.

Usage

```
eem_list
```

Format

eemlist

eem_metatemplate *Create table that contains sample names and locations of files.*

Description

You can use this table as an overview of your files and/or as a template for creating a metadata table.

Usage

```
eem_metatemplate(eem_list = NULL, absorbance = NULL)
```

Arguments

eem_list	eemlist
absorbance	data frame with absorbance data

Value

data frame

Examples

```
data(eem_list)
data(abs_data)

eem_metatemplate(eem_list,abs_data)
```

eem_name_replace *Replace matched patterns in sample names*

Description

Sample names in eemlist can be altered.

Usage

```
eem_name_replace(eem_list, pattern, replacement)
```

Arguments

eem_list	data of class eemlist
pattern	character vector containing pattern to look for.
replacement	character vector of replacements. Has to have the same length as pattern

Details

str_replace_all from package stringr is used for the replacement. Please read the corresponding help for further options.

Value

An eemlist.

See Also

str_replace_all

Examples

```
eem_names(eem_list)

eem_list <- eem_name_replace(eem_list,"sample","Sample")
eem_names(eem_list)
```

---

eem_overview_plot *Plot fluorescence data from several samples split into several plots.*

Description

Plot fluorescence data from several samples split into several plots.

Usage

```
eem_overview_plot(data, spp = 8)
```

Arguments

data	fluorescence data of class eemlist
spp	number of samples per plot

Value

list of ggplots

Examples

```
data(eem_list)
eem_overview_plot(eem_list,spp=3)
```

`eem_parafac` *Runs a PARAFAC analysis on EEM data*

Description

One or more PARAFAC models can be calculated depending on the number of components. The idea is to compare the different models to get the most suitable. B-mode is emmission wavelengths, C-mode is excitation wavelengths and, A-mode is the loadings of the samples. The calculation is done with `parafac`, please see details there.

Usage

```
eem_parafac(eem_list, comps, maxit = 500, normalise = TRUE,
  const = c("nonneg", "nonneg", "nonneg"), nstart = 10, ctol = 10^-4,
  cores = parallel::detectCores()/2, verbose = FALSE, ...)
```

Arguments

`eem_list`	object of class `eem`
`comps`	vector containing the desired numbers of components. For each of these numbers one model is calculated
`maxit`	maximum iterations for PARAFAC algorithm
`normalise`	state whether EEM data should be normalised in advance
`const`	constraints of PARAFAC analysis. Default is non-negative ("nonneg"), alternatively smooth and non-negative ("smonon") might be interesting for an EEM analysis.
`nstart`	number of random starts
`ctol`	Convergence tolerance (R^2 change)
`cores`	number of parallel calculations (e.g. number of physical cores in CPU)
`verbose`	print infos
`...`	additional parameters that are passed on to `parafac`

Value

object of class parafac

See Also

`parafac`

Examples

```
data(eem_list)

dim_min <- 3 # minimum number of components
dim_max <- 7 # maximum number of components
nstart <- 10 # random starts for PARAFAC analysis, models built simulanuously, best selected
cores <- parallel::detectCores()/2 # use all cores but do not use all threads
maxit = 500
ctol <- 10^-4 # tolerance for parafac

pfres_comps <- eem_parafac(eem_list,comps=seq(dim_min,dim_max),
    normalise = TRUE,maxit=10000,nstart=nstart,ctol=ctol,cores=cores)
```

---

eem_raman_area *Calculate raman area of EEM samples*

Description

Calculate raman area of EEM samples

Usage

```
eem_raman_area(eem_list, blanks_only = TRUE, average = FALSE)
```

Arguments

eem_list	An object of class eemlist.
blanks_only	logical. States whether all samples or just blanks will be used.
average	logical. States whether samples will be averaged before calculating the raman area.

Details

Code based on eem_raman_normalisation.

Value

data frame containing sample names, locations and raman areas

Examples

```
data(blank)
eem_raman_area(blank)
```

`eem_raman_normalisation2` *Wrapper function to eem_raman_normalisation (eemR).*

Description

Usually Raman normalisation is done with fluorescence data from a blank sample. Sometimes you already know a value for the Raman area. This function can do both.

Usage

```
eem_raman_normalisation2(data, blank = "blank")
```

Arguments

`data`	fluorescence data of class eemlist
`blank`	defines how Raman normalisation is done (see 'Details')

Details

Possible values for blank:

"blank": normalisation is done with a blank sample. Please refer to `eem_raman_normalisation`.

numeric: normalisation is done with one value for all samples.

data frame: normalisation is done with different values for different samples. Values are taken from a data.frame with sample names as rownames and one column containing the raman area values.

Value

fluorescence data of class eemlist

Examples

```
data(eem_list)
# correction by blank
eems_bl <- eem_raman_normalisation2(eem_list,blank="blank")

# correction by value
eems_num <- eem_raman_normalisation2(eem_list,blank=168)
```

eem_range *Cut EEM data matching a given wavelength range*

Description

Cut EEM data matching a given wavelength range

Usage

```
eem_range(data, ex = c(0, Inf), em = c(0, Inf))
```

Arguments

data	EEM data as eemlist
ex	optional desired range of excitation wavelength
em	optional desired range of emission wavelength

Value

An eemlist of reduced spectra size.

Examples

```
data(eem_list)
eem_range(eem_list,ex = c(250,Inf),em = c(280,500))
```

eem_red2smallest *Reduce wavelength range of EEM spectra to widest available in the whole sample set.*

Description

Reduce wavelength range of EEM spectra to widest available in the whole sample set.

Usage

```
eem_red2smallest(data, verbose = FALSE)
```

Arguments

data	data of EEM samples as eemlist
verbose	states whether additional information is given in the command line

Details

This step is neccessary to perform a PARAFAC analysis which can only be calculated with spectra of similar range.

Value

eemlist with reduced spectral width

Examples

```
data(eem_list)
eem_red2smallest(eem_list)
```

`eem_rem_scat` *Remove Raman and Rayleigh scattering in fluorescence data*

Description

Wrapper function to remove several scatterings in one step using `eem_remove_scattering`.

Usage

```
eem_rem_scat(data, remove_scatter, remove_scatter_width = 10,
  interpolation = FALSE, cores = parallel::detectCores()/2)
```

Arguments

`data`	object of class eemlist
`remove_scatter`	named logical vector. The names of the vector must be out of "raman1", "raman2", "rayleigh1" and "rayleigh2". Set `TRUE` if certain scattering should be removed.
`remove_scatter_width`	numeric vector containing width of scattering to remove. If there is only one element in this vector, each this is the width of each removed scattering. If there are 4 values, differnt widths are used ordered by "raman1", "raman2", "rayleigh1" and "rayleigh2".
`interpolation`	logical, optionally states whether interpolation is done right away
`cores`	optional, CPU cores to use for interpolation

Value

eemlist

Examples

```
data(eem_list)

remove_scatter <- c()
remove_scatter["raman1"] = TRUE
remove_scatter["raman2"] = TRUE
remove_scatter["rayleigh1"] = TRUE
remove_scatter["rayleigh2"] = TRUE
remove_scatter_width = c(15,10,16,12)

eem_rem_scat(eem_list,remove_scatter,remove_scatter_width)
```

---

eem_scale_ext *Determine the range of fluorescence values in a set of samples*

Description

Determine the range of fluorescence values in a set of samples

Usage

```
eem_scale_ext(data)
```

Arguments

data	eemlist containing the EEM data

Value

numeric vector

Examples

```
data(eem_list)
eem_scale_ext(eem_list)
```

---

eem_smooth *Smooth fluorescence data by calculating rolling mean along excitation wavelengths.*

Description

Smooth fluorescence data by calculating rolling mean along excitation wavelengths.

Usage

```
eem_smooth(data, n = 4)
```

Arguments

data	fluorescence data of class eemlist
n	width of rolling mean window in nm

Value

eemlist with smoothed data

Examples

```
data(eem_list)

eem_list <- eem_smooth(eem_list,n=4)
```

---

ggeem *EEM spectra plotted with ggplot2*

Description

ggeem creates nice plots from EEM spectra of class ggplot. Plots can be modified as any ggplot by adding layers and/or elements with "+".

Usage

```
ggeem(data, fill_max = FALSE, ...)

## Default S3 method:
ggeem(data, fill_max = FALSE, ...)

## S3 method for class 'eemlist'
ggeem(data, fill_max = FALSE, ...)

## S3 method for class 'eem'
ggeem(data, fill_max = FALSE, ...)

## S3 method for class 'parafac'
ggeem(data, fill_max = FALSE, ...)

## S3 method for class 'data.frame'
ggeem(data, fill_max = FALSE, ...)
```

Arguments

data	eem, eemlist, parafac or data.frame. The details are given under 'Details'
fill_max	FALSE
...	parameters passed on to ggplot

Details

The data can be of different sources: eem: a single EEM pectrum is plotted eemlist: all spectra of the samples are plotted in one facet plot data.frame: a data.frame containing EEM data. Can be created by e.g. `as.data.frame.eem`

Value

a ggplot object

Examples

```
## plotting one distinct sample
data(eem_list)
eem <- eem_extract(eem_list,c("sample6","sample7"),keep=TRUE)
ggeem(eem)
```

---

list_join *Full join of a list of data frames.*

Description

Full join of a list of data frames.

Usage

```
list_join(df_list, by)
```

Arguments

df_list	list of data frames to by joined
by	character vector containing information how to join data frames. Format to be according to by in full_join. Each data frame has to contain the column(s) used for joining.

Value

The joint data frame.

See Also

full_join

Examples

```
a <- data.frame(what=letters[1:5],a=c(1:5))
b <- data.frame(what=letters[1:5],b=c(7:11))
c <- data.frame(what=letters[1:5],c=c(20:24))

df_list <- list(a,b,c)

list_join(df_list,by="what")
```

`maxlines` *Extract data from emission and excitation wavelengths of the components of a PARAFAC model*

Description

Data of wavelengths is returned. For each component the lines intersecting at the component maxima are returned.

Usage

```
maxlines(pfmodel)
```

Arguments

`pfmodel`	object of class parafac

Value

data frame

Examples

```
data(pfres_comps1)

ml <- maxlines(pfres_comps[[1]])
```

`norm2A` *Compensate for normalisation in C-modes*

Description

Factors used for normalisation are saved separately in the PARAFAC models. With this function, the normalisation factors are combined with the A-modes of the model and removed as a separate vector. This means former normalisation is accounted for in the amount of each component in each sample. If no normalisation was done, the original model is returned without warning.

Usage

```
norm2A(pfmodel)
```

Arguments

pfmodel	object of class parafac

Value

object of class parafac

Examples

```
data(pfres_comps1)

pfres_comps[[2]][[3]] <- norm2A(pfres_comps[[2]][[3]])
```

---

norm_array *Normalise 3-dimensional array in first and second dimension*

Description

Normalise 3-dimensional array in first and second dimension

Usage

```
norm_array(eem_array)
```

Arguments

eem_array	3-dimensional array

Value

array

Examples

```
data(eem_list)

a <- eem2array(eem_list)
an <- norm_array(a)
```

pfres_comps *PARAFAC model*

Description

PARAFAC model

Usage

```
pfres_comps
```

Format

list of parafacs

pfres_comps2 *Absorption data of 7 samples*

Description

Absorption data of 7 samples

Usage

```
pfres_comps2
```

Format

list of parafacs

sh *result from PARAFAC split-half analysis*

Description

result from PARAFAC split-half analysis

Usage

```
sh
```

Format

list of parafacs

splithalf *Running a Split-Half analysis on a PARAFAC model*

Description

The samples are split into four subsamples: A,B,C,D. Subsamples are then combined and compared: AB vs. CD, AC vs. BD, AD vs. BC. The results show graphs from the components of each of the 6 models.

Usage

```
splithalf(eem_list, comps, splits = NA, rand = FALSE, normalise = TRUE,
  nstart = 10, cores = parallel::detectCores()/2, maxit = 500,
  ctol = 10^(-5), ..., verbose = FALSE)
```

Arguments

eem_list	eemlist containing sample data
comps	number of desired components
splits	optional, list of 4 numerical vectors containing the sample numbers for A,B,C and D sample subsets
rand	logical, splits are randomised
normalise	state whether EEM data should be normalised in advance
nstart	number of random starts
cores	number of parallel calculations (e.g. number of physical cores in CPU)
maxit	maximum iterations for PARAFAC algorithm
ctol	Convergence tolerance (R^2 change)
...	additional parameters that are passed on to parafac
verbose	states whether you want additional information during calculation

Details

Split data sets can be split suboptimal and cause low TCCs. Therefore, subsamples are recombined in 3 different ways and a TCC close to 1 in only one split combination per component is already a positive result. Check the split sets to check for sample independency.

Value

data frame containing components of the splithalf models

See Also

splithalf_plot, splithalf_tcc

Examples

```
data(eem_list)

splithalf(eem_list,6,nstart=2)
```

`splithalf_plot` *Plot results from a splithalf analysis*

Description

Graphs of all components of all models are plotted to be compared.

Usage

```
splithalf_plot(fits)
```

Arguments

`fits`	list of components data

Value

ggplot

See Also

`splithalf`

Examples

```
data(sh)

splithalf_plot(sh)
```

splithalf_splits — *Extracting a list of sample names in each subsample from a splithalf analysis*

Description

Extracting a list of sample names in each subsample from a splithalf analysis

Usage

```
splithalf_splits(fits)
```

Arguments

fits	list of parafac models (from a splithalf analysis)

Value

data frame containing TCC values

Examples

```
data(sh)
splithalf_splits(sh)
```

splithalf_tcc — *Extracting TCC values from a splithalf analysis*

Description

Extracting TCC values from a splithalf analysis

Usage

```
splithalf_tcc(fits)
```

Arguments

fits	list of parafac models (from a splithalf analysis)

Value

data frame containing TCC values

Examples

```
data(sh)

splithalf_tcc(sh)
```

tcc *Caluclate Tucker's Congruence Coefficient of PARAFAC components*

Description

Componets must be passed as peak lines `maxlines`

Usage

```
tcc(maxl_table, na.action = "na.omit")
```

Arguments

`maxl_table`	data frame containing the peak lines of components
`na.action`	if "na.omit" NA are deleted from prior the test

Value

data.frame containing the TCCs

Examples

```
data(pfres_comps1)

ml <- maxlines(pfres_comps[[2]])

tcc(ml)
```

tcc_find_pairs *Reorders components of different PARAFAC models according to best fit (TCC)*

Description

When running a splithalf analysis similar components are not necessarily on the same position. This function looks for best fits with Tucker's Congruence Coefficients and returns a list of models with reordered components.

Usage

```
tcc_find_pairs(fits)
```

Arguments

`fits`	list of parafac models

Value

list of parafac models

See Also

splithalf

Examples

```
data(eem_list)

# function currently only used from within splithalf
splithalf(eem_list,6,nstart=2)
```

Index