Package: kwb.waterParcel (via r-universe)

November 20, 2024

Title Lagrangian Sampling Interpretation

Version 0.0.0.9000

Description Conservative tracers within the water can be used to interpret the results of a sampling compaign following one water parcel (Lagrangian Sampling).

License MIT + file LICENSE

URL https://github.com/KWB-R/kwb.waterParcel

BugReports https://github.com/KWB-R/kwb.waterParcel/issues

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.1

Suggests covr

Imports tidyr, magrittr

Config/pak/sysreqs libicu-dev

Repository https://kwb-r.r-universe.dev

RemoteUrl https://github.com/KWB-R/kwb.waterParcel

RemoteRef HEAD

RemoteSha 85ce0bcf7407003e418ac2139aef6bffaabc7fde

Contents

check_for_units *Check for unique units per parameter*

Description

Check for unique units per parameter

Usage

```
check_for_units(df_in)
```

Arguments

df_in	Data frame loaded with load_lab_data()

Value

Returns either a warning or a confirmation

discard_na_samples *Removes all samples without complete tracer data*

Description

Removes all samples without complete tracer data

Usage

```
discard_na_samples(MCS_input)
```

Arguments

MCS_input	A list of tracer concentrations per sample

Value

A list of tracer concentrations per sample

`get_all_sample_concentrations` *Random concentrations in all samples based on measured data and assumed deviation*

Description

Random concentrations in all samples based on measured data and assumed deviation

Usage

```
get_all_sample_concentrations(
  input_table,
  predictor_samples,
  tracer_names,
  MCS_runs,
  rel_deviation
)
```

Arguments

`input_table`	Data frame reshaped by function `reshape_table()`
`predictor_samples`	A character vector of names of samples used for predicting the substance concentration of the remaining samples.
`tracer_names`	A vector of character strings defining the tracers used
`MCS_runs`	The number of random samples drawn (Monte-Carlo-Simulation runs)
`rel_deviation`	Either one numeric value between 0 and 1 defining the standard deviation relative to the measured value for all tracers or a vector of the same length as 'tracer_names' defining a deviation for each tracer

Value

A List of data frames with columns corresponding to the tracers and rows correspoding to the Monte Carlo runs. The predictor samples are the first three entries and are named "A", "B" and "C"

`get_concentration_ranges` *Find minimum and maximum of the tracers in all samples (including MCS)*

Description

Find minimum and maximum of the tracers in all samples (including MCS)

Usage

```
get_concentration_ranges(MCS_input)
```

Arguments

MCS_input	A list of sample Data frames created by get_all_sample_concentrations()

---

get_varrying_sample_concentration *Random concentrations in sample based on measured data and assumed deviation*

Description

Random concentrations in sample based on measured data and assumed deviation

Usage

```
get_varrying_sample_concentration(
  input_table,
  sample_name,
  tracer_names,
  MCS_runs,
  rel_deviation
)
```

Arguments

input_table	Data frame reshaped by function reshape_table()
sample_name	A character string defining the sample
tracer_names	A vector of character strings defining the tracers used
MCS_runs	The number of random samples drawn (Monte-Carlo-Simulation runs)
rel_deviation	Either one numeric value between 0 and 1 defining the standard deviation relative to the measured value for all tracers or a vector of the same length as 'tracer_names' defining a deviation for each tracer

Value

A data frame, columns corresponding to the tracers and rows correspoding to the runs.

`load_lab_data` *Loads CSV-Data table*

Description

The CSV table must contain one row per measurement and standard column names as defined in the details section

Usage

```
load_lab_data(path, file)
```

Arguments

`path`	The filepath
`file`	The filename including the ".csv" ending

Details

Required table column names:

- sample_name: unique ID of the sample (at least unique per sampling campaign)
- sampling_campaign: unique ID of the sampling campaign
- tBeg: starting time of the sample (format: "YYYY-mm-dd HH:MM:SS")
- tEnd: ending time of the sample (format: "YYYY-mm-dd HH:MM:SS")
- operator: definition of LOD-relation (either "<", ">" or empty)
- value: measured value
- parameter: unique name of the measured parameter (this parameter is used column name after reshaping the table, so it should not start with a number or contain any special symbols or spaces)
- unit: the unit of the measured value

Value

A data frame corresding to the csv file, with date columns specified as POSIXct

normalize_MCS_input *Transforms tracer concentrations into normalized values between 0 and 1*

Description

Transforms tracer concentrations into normalized values between 0 and 1

Usage

```
normalize_MCS_input(MCS_input)
```

Arguments

MCS_input	A list of sample Data frames created by get_all_sample_concentrations()

Details

Normalization per tracer by equation:
(c_i - min(c)) / (max(c) - min(c))
This results in values between 0 (equals the overall minimal value of the tracer within all samples) and 1 (equals the overall maximal value of the tracer within all samples).

Value

Same structured list as the input with normalized values instead of absolute concentrations

reshape_table *Reshape Data to Parameter-as-Columns-Structure*

Description

This function keeps the columns "sample_name", "sampling_campaign", "tBeg", "tEnd" and adds one column per parameter. The operators are taken into account.

Usage

```
reshape_table(df_in, sampling_campaigns, bLOD = "na", aLOD = "na")
```

Arguments

df_in	Data frame loaded with load_lab_data()
sampling_campaigns	One or more sampling Campaign IDs to be considered
bLOD	Character string defining the handling of operator "<" (below limit of detection). Either "half" for dividing the value by 2, "zero" or "na" for setting the value to 0 or NA, respectively.
aLOD	Character string defining the handling of operator ">" (above limit of detection). Either "double" for multiplying the value by 2, "use_limit" for using the limit value or "na" for setting the value to NA.

Value

A data frame of all data per sample in a row, arranged by the starting time of the sample.

Index