Package 'TREXr'

March 24, 2021

Title Tree Sap Flow Extractor

Version 1.0.0

Description Performs data assimilation, processing and analyses on sap flow data obtained with the thermal dissipation method (TDM). The package includes functions for gap filling time-series data, detecting outliers, calculating data-processing uncertainties and generating uniform data output and visualisation. The package is designed to deal with large quantities of data and to apply commonly used data-processing methods. The functions have been validated on data collected from different tree species across the northern hemisphere (Peters et al. 2018 <doi:10.1111/nph.15241>).

License MIT + file LICENSE

LazyData true

URL https://the-hull.github.io/TREX/

BugReports https://github.com/the-Hull/TREX/issues

Encoding UTF-8

Depends R (>= 3.5.0)

Imports chron, zoo, solaR, stats, graphics, lubridate, utils, grDevices, lhs, doParallel, foreach, tibble, dplyr, sensitivity, randtoolbox, boot, doSNOW, msm, parallel, magrittr, tools

RoxygenNote 7.1.1

Suggests shiny, plotly

NeedsCompilation no

Author Richard Peters [cre, aut, cph] (<https://orcid.org/0000-0002-7441-1297>), Christoforos Pappas [aut, ctb] (<https://orcid.org/0000-0001-5721-557X>), Alexander Hurley [aut, cph] (<https://orcid.org/0000-0002-9641-2805>), Rafael Poyatos [ctb], Victor Flo [ctb], Roman Zweifel [ctb], Willem Goossens [ctb], Kathy Steppe [ths, etr, fnd]

Maintainer Richard Peters <richardlouispeters3@hotmail.com>

Repository CRAN

Date/Publication 2021-03-24 10:10:02 UTC

R topics documented:

`agg.data` *Aggregation of time-series data*

Description

Aggregation of time-series data and start/end time selection. This function provides the option to select the temporal step size for aggregation of a single time series origination from an `is.trex`-compliant object. Additionally, the user can define the start and end time of the series and select the function used for aggregation.

Usage

```
agg.data(input,
 time.agg = 60*24,
 start = "2012-07-28 00:00",
 end = "2012-10-29 00:00",
 FUN = "mean",
 unit = 60,
 na.rm = TRUE,
 df = FALSE)
```

Arguments

input	An is.trex-compliant time series from tdm_cal.sfd outputs (e.g., X$sfd.mw$sfd).
time.agg	Numeric, the aggregation time in minutes (default = 60).
start	Character string, the start time for the series. Format has to be provided in "UTC" (e.g. "2012-05-28 00:00" or Year-Month-Day Hour:Minute). Starting time should not be earlier than the start of the series. If not provided the entire series is considered.
end	character string, the end time for the series. Format has to be provided in "UTC" (e.g. "2012-06-28 00:50" or Year-Month-Day Hour:Minute). Starting time should be earlier than the end time and the end time should not exceed the length of the series. If not provided the entire series is considered.
FUN	Quoted function name to compute the summary statistics which can be applied to all data subsets (see aggregate; including "sum", "mean", "median", "sd", "se", "min", "max").
unit	Numeric, the minutes in which a velocity unity is provided (e.g., $cm^3cm^{-2}h^{-1} = 60$) for summation (FUN = "sum"; default = 60).
na.rm	Logical; if TRUE missing values are removed (default = TRUE).
df	Logical; if TRUE, output is provided in a data.frame format with a timestamp and a value column. If FALSE, output is provided as a zoo vector object (default = FALSE).

Details

Time series are often derived at variable resolutions. This function provides the option to aggregate (homogenize) time steps with standard FUN statistics. When applying this function to calculate summed sap flow values (e.g., $cm^3cm^{-2}d^{-1}$) one needs to include the velocity unit, as the summation is dependent upon the minimum timestep of the time series (e.g., $cm^3cm^{-2}h^{-1}$, unit = 60).

Value

A zoo object or data.frame in the appropriate format for other functionalities.

Examples

```
#aggregate SFD values to mean hourly and daily sums

raw   <- example.data(type="doy")

input <- is.trex(raw,tz="GMT",time.format="%H:%M",
                  solar.time=TRUE,long.deg=7.7459,ref.add=FALSE,df=FALSE)

input[which(input<0.4)]<-NA

k.input<-tdm_dt.max(dt.steps(input,time.int=15,
               max.gap=180,decimals=10),methods=c("mw"))
```

```
sfd.input<-tdm_cal.sfd(k.input,make.plot=FALSE,
                  df=FALSE,wood="Coniferous")$sfd.mw$sfd

# means
output.1hmean <- agg.data(sfd.input,
                        time.agg=60,
                        start="2012-07-28 00:00",
                        end="2012-08-29 00:00",
                        FUN="mean",
                        na.rm=TRUE,
                        df=FALSE)
output.6hmean <- agg.data(sfd.input,
                          time.agg=60*6,
                          start="2012-07-28 00:00",
                          end="2012-08-29 00:00",
                          FUN="mean",
                          na.rm=TRUE,
                          df=FALSE)
plot(output.1hmean,col="cyan")
lines(output.6hmean,col="black")

# daily sums
output.dsum<-agg.data(sfd.input,
                      time.agg=60*24,
                      start="2012-07-28 00:00",
                      end="2012-10-29 00:00",
                      FUN="sum",
                      unit=60,
                      na.rm=TRUE,
                      df=FALSE)
plot(output.dsum)
points(output.dsum,pch=16)
```

`cal.data` *Calibration Measurements*

Description

Returns raw calibration experiment data obtained from literature, with K values combined with gravimetrically determined sap flux density, as detailed in Flo *et al.* (2019). The `data.frame` contains 22 studies with 37 different species. The data is used within the `tdm_cal.sfd` function to calculate sap flux density. Description on the genus, species, calibration material, wood porosity and diameter of the stem is provided in Flo *et al.* (2019). The presented data is open for public use.

Usage

```
cal.data
```

Format

Provides a data.frame with 4024 rows and 10 columns.

Study Study from which the data originates (see Flo et al. 2019) (`character`)

Method Heat-based sap flow measurement method (TD = Thermal Dissipation) (`character`)

Genus Monitored genus (`character`)

Species Monitored species (`character`)

Calibration.material Description on the calibration method that was used, including stem segment, whole plant and whole plant without roots (`character`)

Wood.porosity Wood structure type of the examined species, including coniferous, diffuse-porous, ring-porous and monocots (`character`)

Diameter Diameter at breast height of the calibration subject (in cm) (`numeric`)

k Proportional difference between ΔT and ΔT_{max} measured by the thermal dissipation probes (unitless; `numeric`)

SFD Sap flux density measured gravimetrically (in $cm^3 cm^{-2} h^{-1}$; `numeric`)

Granier Sap flux density calculated according to Granier et al. 1985 using k (in $cm^3 cm^{-2} h^{-1}$; using $43.84k^{1.231}$) (`numeric`)

Details

Currently included studies are given below; individual labels (quoted) can be applied in `tdm_cal.sfd` (argument "study"):

- Braun and Schmid 1999
- Cabibel et al. 1991
- Cain 2009
- Chan 2015
- Fuchs et al. 2017
- Granier 1985
- Gutierrez & Santiago 2005
- Herbst et al. 2007
- Liu et al. 2008
- Lu 2002
- Lu and Chacko 1998
- Oliveira et al. 2006
- Paudel et al. 2013
- Rubilar et al. 2016
- Schmidt-walter et al. 2014
- Sperling et al. 2012
- Sugiura et al. 2009
- Sun et al. 2012

- Vellame et al. 2009
- Hubbard et al. 2010
- Peters et al. 2017
- Steppe et al. 2010

References

Flo V, Martinez-Vilalta J, Steppe K, Schuldt B, Poyatos, R. 2019. A synthesis of bias and uncertainty in sap flow methods. Agricultural and Forest Meteorology 271:362-374. doi: 10.1016/j.agrformet.2019.03.012

Granier A. 1985. Une nouvelle methode pour la measure du flux de seve brute dans le tronc des arbres. Annales des Sciences Forestieres 42:193–200. doi: 10.1051/forest:19850204

`dt.steps` *Determine temporal resolution*

Description

Performs minimum time step standardization, gap filling and start/end time selection. This function provides the option to select the minimum temporal step size of an `is.trex` object. Additionally, the user can define the start and end time of the series and select the minimum size under which gaps should be filled, using linear interpolation.

Time series have different temporal resolutions. This function provides the option to standardize the minimum time step by either performing a linear interpolation when the requested time step is smaller than the minimum time step of the series or average values when the requested time step is larger than the minimum time step of the series. Before this process, the entire time series is converted to a one-minute time step by applying a linear interpolation (excluding gap $periods >$ `max.gap`).

Usage

```
dt.steps(input, start,
       end, time.int = 10, max.gap = 60,
       decimals = 10, df = FALSE)
```

Arguments

`input`	An `is.trex`-compliant (output) object
`start`	Character string providing the start time for the series. Format has to be provided in "UTC" (e.g., "2012-05-28 00:00" or Year-Month-Day Hour:Minute). Starting time should not be earlier than the start of the series.
`end`	Character string providing the start time for the series. Format has to be provided in "UTC" (e.g., "2012-06-28 00:50" or Year-Month-Day Hour:Minute). End time should be earlier than the end time and not later than that of the series.

time.int	Numeric value providing the number of minutes for the minimum time step. When time.int is smaller than the minimum time step of the series, a linear interpolation is applied. If time.int is larger than the minimum time step of the series values are averaged to the chosen value of time.int (after performing a linear interpolation to obtain a one-minute resolution).
max.gap	Numeric value providing the maximum size of a gap in minutes, which can be filled by performing a linear interpolation.
decimals	Integer value defining the number of decimals of the output (default = 10).
df	Logical; if TRUE, output is provided in a data.frame format with a timestamp and a value (ΔT or ΔV) column. If FALSE, output is provided as a zoo object (default = FALSE).

Value

A zoo object or data.frame in the appropriate format for further processing.

Examples

```
input <- is.trex(example.data(type="doy"),
           tz="GMT",time.format="%H:%M", solar.time=TRUE,
           long.deg=7.7459,ref.add=FALSE)
in.ts <- dt.steps(input=input,start='2012-06-28 00:00',end='2012-07-28 00:00',
                  time.int=60,max.gap=120,decimals=6,df=FALSE)
plot(in.ts)
head(in.ts)
```

---

example.data *Generate example TDM input data*

Description

This function returns a data.frame containing standard TDM (thermal dissipation method) measurements provided in two different formats. The data is obtained from tdm.data where ΔV measurements are given for Norway spruce (*Picea abies* Karts.) growing in a valley in the Swiss Alps. See tdm.data for additional details.

Usage

```
example.data(type = "timestamp")
```

Arguments

type	Character string, indicating whether the example data should be displayed with a timestamp (default = "timestamp") or separate year, day of day ("doy").

Details

This dataset can be applied for testing the functions provided in TREXr.

Value

A data.frame containing TDM measurements according to a specific type.

Examples

```
# get example data
input_data <- example.data(type = "timestamp")
input_data <- example.data(type = "doy")
head(input_data)
```

---

gap.fill *Gap filling by linear interpolation*

Description

Fills gaps by linear interpolation between observations. This function provides the option to define the minimum window under which gaps should be filled, using linear interpolation.

Usage

```
gap.fill(input, max.gap = 60, decimals = 10, df = FALSE)
```

Arguments

input	an is.trex-compliant object.
max.gap	Numeric value providing the maximum size of a gap in minutes, which can be filled by performing a linear interpolation.
decimals	Integer value defining the number of decimals of the output (default = 10).
df	Logical; if TRUE, output is provided in a data.frame format with a timestamp and a value (ΔT or ΔV) column. If FALSE, output is provided as a zoo object (default = FALSE).

Value

A zoo object or data.frame in the appropriate format for further processing.

Examples

```
# fill two hour gaps
raw   <- example.data(type = "doy")
input <-
  is.trex(
    raw,
    tz = "GMT",
    time.format = "%H:%M",
    solar.time = TRUE,
    long.deg = 7.7459,
    ref.add = FALSE,
    df = FALSE)

# create gaps in data
input[which(input < 0.4 | input > 0.82)] <- NA
fill_120 <- gap.fill(
  input = input,
  max.gap = 120,
  decimals = 10,
  df = FALSE)
fill_15 <- gap.fill(
  input = input,
  max.gap = 15,
  decimals = 10,
  df = FALSE)
```

`is.trex` *Testing and preparing input data*

Description

Tests if the structure of the input matches the requirements of `TREXr` functions and specifies the time zone. The input has to be presented in one of two different `data.frame` formats. i) Timestamp format: including a 1) timestamp of the measurements column (`character`), and 2) value of ΔV (or ΔT; [`as.numeric`]). ii) DOY format: including a 1) year of measurements column `as.integer`, 2) day of the year (DOY) of measurement (`as.integer`), 3) hour of the measurement (`character`), and 4) value of ΔV (or ΔT; [`as.numeric`]). TREXr functions are applied on time series obtained from a set of thermal dissipation probes. This includes the option where the thermal dissipation method (TDM) is used with only a reference and heating probe, or when including addition reference probes (see `ref.add`). These reference probe measurements can be added to the DOY or timestamp format in ΔV (or ΔT) (`as.numeric`) labelled `ref1`, `ref2`, etc. (depending on the number of reference probes). For this function the following column names have to be present within the `data.frame`: "timestamp" or "year" & "doy" & "hour" = indicators of time and "value" = TDM measurements (option "ref1", "ref2, ..., refn = reference probes). After specifying the time zone (`tz`), one can select whether to standardize the temporal series to solar time (see `solar.time`) by providing the longitude in decimal degrees at which the measurements were

obtained (in long.deg). All timestamps within the function are rounded to minute resolution and output can be either provided in a zoo format (df = FALSE) or data.frame (df = TRUE; default is FALSE). **Note, that the output time series is always given in UTC time zone.**

Usage

```
is.trex(data, tz = 'UTC', tz.force = FALSE, time.format = '%m/%d/%y %H:%M:%S',
  solar.time = TRUE, long.deg = 7.7459,
   ref.add = FALSE, df = FALSE)
```

Arguments

data	A data.frame in either timestamp format or doy format.
tz	Character string, indicates the time zone in which the measurements have been recorded.
tz.force	Logical; if TRUE, the time zone is forced to "UTC" or re-labelled yet not shifted in time. "UTC" is required for internal processing (default = FALSE).
time.format	Character string, indicates the format of the timestamp.
solar.time	Logical; if TRUE, time is converted to solar time, depending upon the location where the measurements have been taken. If FALSE, the output is provided in "UTC" (default = FALSE).
long.deg	Numeric, longitude in decimal degrees East to perform the solar time conversion. Only required when solar.time=TRUE.
ref.add	Logical; if TRUE, additional probes provided within data are considered. The ΔT values are then corrected by subtracting the ΔT measured between the reference probes from the ΔT measured between the heated and unheated probe (default = FALSE).
df	Logical; if TRUE, output is provided in a data.frame format with a timestamp and a value (ΔT or ΔV) column. If FALSE, output is provided as a zoo object (default = FALSE).

Details

To prevent errors occurring in subsequent TREXr functions, it is advised to run this function for checking the data structure and preparing it for further analyses. For the specific time zone see https://en.wikipedia.org/wiki/List_of_tz_database_time_zones or for formatting see OlsonNames(). The format of the timestamp has to be provided according to https://www.stat.berkeley.edu/~s133/dates.html. For the method behind the solar time conversion, see the solar package (https://CRAN.R-project.org/package=solaR). The longitude has to be provided in positive decimal degrees for study sites East from the Greenwich meridian and negative for sites to the West.

Value

A zoo object or data.frame in the appropriate format for other functionalities.

Examples

```
#validating and structuring example data
raw   <- example.data(type="doy")
input <- is.trex(raw,tz="GMT",time.format="%H:%M",
    solar.time=TRUE,long.deg=7.7459,
    ref.add=FALSE,df=FALSE)
head(raw)
str(input)
head(input)
plot(input)
```

`out.data` *Generating TDM output*

Description

Generates relevant outputs from the sap flux density (SFD) values. This function provides both F_d (SFD expressed in $mmol m^{-2} s^{-1}$) and crown conductance values (G_C; an analogue to stomatal conductance) in an easily exportable format. Additionally, the function can perform environmental filtering on F_d and G_C and model G_C sensitivity to vapour pressure deficit (VPD). The user can choose between in- (`method = "env.filt"`) or excluding (`method = "stat"`) environmental filtering on the G_C and adjust the filter threshold manually.

Usage

```
out.data(
  input,
  vpd.input,
  sr.input,
  prec.input,
  peak.hours = c(10:14),
  low.sr = 150,
  peak.sr = 300,
  vpd.cutoff = 0.5,
  prec.lim = 1,
  method = "env.filt",
  max.quant = 1,
  make.plot = TRUE
)
```

Arguments

`input`	An `is.trex`-compliant time series from `tdm_cal.sfd` outputs (e.g., `X$sfd.mw$sfd`)
`vpd.input`	An `is.trex`-compliant object an individual series of VPD in kPa (see vpd). The extent and temporal resolution should be equal to input. Use `dt.steps` to correct if needed.

`sr.input` An `is.trex`-compliant object of an individual series of solar irradiance (e.g. either PAR or global radiation; see `sr`). The extent and temporal resolution should be equal to input. Use `dt.steps` to correct if needed. This data is only needed when applying the "env.filt" method.

`prec.input` An `is.trex`-compliant object of daily precipitation in mm d-1 (see `preci`). The extent should be equal to input with a daily temporal resolution. Use `dt.steps` to correct if needed. This data is only needed when applying the "env.filt" method.

`peak.hours` Numeric vector with hours which should be considered as peak-of-the-day hours (default = `c(10:14)`). This variable is only needed when the "stat" method is selected.

`low.sr` Numeric threshold value in the unit of the sr.input time-series (e.g., W m-2) to exclude cloudy days which impact G_C (default = 150 W m-2). This variable is only needed when the "env.filt" method is selected.

`peak.sr` Numeric threshold value in the unit of the sr.input time-series (e.g., W m-2) to exclude hours which are not considered as peak-of-the-day hours (default = 300 W m-2). This variable is only needed when the "env.filt" method is selected.

`vpd.cutoff` Numeric threshold value in kPa for peak-of-the-day mean VPD to eliminate unrealistic and extremely high values of G_C due to low VPD values or high values of G_C (default = 0.5 kPa).

`prec.lim` Numeric threshold value in mm d-1 for daily precipitation to remove rainy days (default = 1 mm d-1). This variable is only needed when "env.filt" method is selected.

`method` Character string indicating whether precipitation and solar irradiance data should be used to determined peak-of-the-day G_C values and filter the daily G_C values ("env.filt") or not ("stat"; default). When `"env.filt"` is selected, `input`, `vpd.input`, `sr.input`, `prec.input`, `peak.sr`, `low.sr`, `vpd.cutoff` and `prec.lim` have to be provided. When `"stat"` is selected only `input`, `vpd.input` and `peak.hours`.

`max.quant` Numeric, defining the quantile of the G_C data which should be considered as GC.max (default = 1).

`make.plot` Logical; if `TRUE`, a plot is generated presenting the response of G_C to VPD.

Details

Various relevant outputs can be derived from the SFD data. This function provides the option to recalculate SFD to F_d (expressed in mmol m-2 s-1) and crown conductance (according to Pappas *et al.* 2018). G_C is estimated per unit sapwood area, where $G_C = F_d/VPD$ (in kPa), assuming that i) the stem hydraulic capacitance between the height of sensor and the leaves is negligible, and ii) that the canopy is well coupled to the atmosphere. In order to reduce the effect of stem hydraulic capacitance, peak-of-the-day G_C are solely considered for calculating daily average G_C. Peak-of-the-day conditions are defined by `peak.hours` or `peak.sr`. Moreover, to analyse the relationship between G_C and environmental measurements (e.g., VPD), the daily mean peak-of-the-day G_C values can be restricted to: i) non-cloudy days (see `low.sr`), to reduce the impact of low irradiance on G_C, ii) non-rainy days (see `prec.lim`), as wet leaves are not well coupled to the atmosphere, and iii) daily mean peak-of-the-day G_C great then a threshold (see `vpd.cutoff`), to eliminate unrealistically high

G_C values due to low F_d or VPD values (when method = "env.filt"). Moreover, the sensitivity of the daily mean peak-of-the-day G_C to VPD is modelled by fitting the following model:

$$G_C = \alpha + \beta VPD^{-0.5}$$

Besides using the raw daily mean peak-of-the-day G_C values, the function also applies a normalization where daily mean peak-of-the-day G_C is standardized to the maximum conductance (GC.max; see max.quant).

Value

A named list of data.frame objects, containing the following items:

raw A data.frame containing the input data and filtered values. Columns include the timestamp [,"timestamp"] (e.g., "2012-01-01 00:00:00"), year of the data [,"year"], day of year [,"doy"], input solar radiance data [,"sr"], daily average radiance data [,"sr"], input vapour pressure deficit data [,"vpd"], isolated peak-of-the-day vapour pressure deficit values [,"vpd.filt"], input daily precipitation [,"prec.day"], sap flux density expressed in mmol m-2 s-1 [,"fd"], crown conductance expressed in mmol m-2 s-1 kPa-1 [,"gc"], and the filtered crown conductance [,"gc.filt"]

peak.mean A data.frame containing the daily mean crown conductance values. Columns include the timestamp [,"timestamp"] (e.g., "2012-01-01"), peak-of-the-day vapour pressure deficit [,"vpd.filt"], the filtered crown conductance mmol m-2 s-1 kPa-1 [,"gc.filt"], and the normalized crown conductance according to the maximum crown conductance [,"gc.norm"].

sum.mod A model summary object (see summary()) of the model between VPD and G_C.

sum.mod.norm A model summary object (see summary()) of the model between VPD and $G_C/GC.max$.

Examples

```
#Gc response function
#Gc response function
raw   <- is.trex(example.data(type="doy"), tz="GMT",
                 time.format="%H:%M", solar.time=TRUE,
                 long.deg=7.7459, ref.add=FALSE)

input <- dt.steps(input=raw, start="2013-05-01 00:00", end="2013-11-01 00:00",
                  time.int=15, max.gap=60, decimals=10, df=FALSE)

input[which(input<0.2)]<- NA
input <- tdm_dt.max(input, methods=c("dr"), det.pd=TRUE, interpolate=FALSE,
                max.days=10, df=FALSE)

output.data<- tdm_cal.sfd(input,make.plot=TRUE,df=FALSE,wood="Coniferous", decimals = 6)

input<- output.data$sfd.dr$sfd

output<- out.data(input=input, vpd.input=vpd, sr.input=sr, prec.input=preci,
                 low.sr = 150, peak.sr=300, vpd.cutoff= 0.5, prec.lim=1,
                 method="env.filt", max.quant=0.99, make.plot=TRUE)
```

```
head(output)
```

`outlier` *Data cleaning and outlier detection*

Description

This function launches a `Shiny` application that (1) visualizes raw and outlier-free time series interactively (using `plotly`), (2) highlights automatically detected outliers, (3) allows the user to revise the automatically detected outliers and manually include data points, and (4) exports the original data, the automatically selected outliers, the manually selected outliers, and the outlier-free time series in an `is.trex`-compliant object that can be further processed.

Usage

```
outlier()
```

Details

Note, that due to the interactive nature of the application, the reactive graphs can become rather slow in updating. We hence suggest breaking long-time series into smaller chunks that do not strain the available memory too much. Trial and error is useful here, but we generally suggest working on a maximum of up to one year at a time. Once the application is launched, the user can load an `.RData` file where a `data.frame` with a imestamp and sensor data (multiple sensor columns are supported). The timestamp in this `data.frame` should be of class `POSIXct`. Users can select the x and y axes of the interactive time series plots. In addition, the user can provide the units of the imported data (e.g., degrees C or mV for ΔT or ΔV, respectively). A parameter (alpha) for automatic outlier detection can be supplied. More specifically, the automatic identification of outliers is based on a two-step procedure: i) the Tukey's method (Tukey, 1977) is applied to detect statistical outliers as values falling outside the range $[q_{0.25} - alpha * IQR, q_{0.75} + alpha * IQR]$, where IQR is the interquartile range ($q_{0.75} - q_{0.25}$) with $q_{0.25}$ denoting the 25% lower quartile and $q_{0.75}$ the 75% upper quartile, and alpha is a user-defined parameter (default value `alpha = 3`; although visual inspection through the interactive plots allows for adjusting alpha and optimizing the automatic detection of outliers), and ii) the lag-1 differences of the raw data are calculated and data points with lag-1 differences greater than the mean of the raw input time series, are excluded. The raw input data from the provided `.RData` file are depicted with black points in the first plot titled 'Raw and automatic detection' while the automatically detected outliers are also highlighted in this plot in red. The user can adjust the parameter `alpha` and visually inspect the automatically detected outliers in order to achieve the optimal automatic outlier selection. This plot allows also interactivity (by hovering the mouse in the upper right corner the available interactive tools appear, e.g., zoom in/out). Also, the lower subpanel of this plot provides a better overview of the temporal extent of the data and allows the user to select narrower time window for a more thorough data inspection.

Once the user is satisfied with the automatically selected data points, one can proceed to the manual outlier selection. The second interactive plot (titled 'Filtered and manual selection') presents the raw data after removing the automatically detected outliers of the previous step, and allows the user to manually select (point, rectangular, and lasso selections are allowed) data points. The first selection identifies points to be removed (outliers), and their color changes to red. If a point is selected for a second time, this will undo its classification as outlier and its color is set back to black (i.e., not an outlier). The red-color data points correspond to the selected outliers to be removed from the data, in addition to those identified in the automated detection.

Value

The function does not return a value, but allows the user to save a `list` containing the raw and outlier-free data, as well as the automatically and manually selected outliers in separate items. Once the user is satisfied with the selected outliers, the 'Download Cleaned Time Series' button will allow to export this `list` as a `".Rds"` file. This file can be subsequently assigned to an object using `readRDS`. The list contained in this file is called `trex_outlier_output` and has four `data.frames`, namely `series_input` with the raw data, `select_auto` with the automatically selected outliers, `select_manual` with the manually selected outliers, and `series_cleaned` with the outlier-free time series. Each of these data frames has a column with the timestamp and a column for the sensor values.

Examples

```
# find example file path
system.file("exdata", "example.RData", package = "TREXr", mustWork = TRUE)
# either copy-paste this into the navigation bar of the file selection window
# or navigate here manually for selection

# launch shiny application
outlier()

# after saving the output, run e.g.:

my_cleaned_data <- readRDS("./cleaned_file.Rds")

## With full workflow:

# get an example time series
raw   <- example.data(type="doy")
input <- is.trex(raw, tz="GMT", time.format="%H:%M",
                 solar.time=TRUE, long.deg=7.7459, ref.add=FALSE, df=FALSE)

# clip a period of interest
input<-dt.steps(input,time.int=60,start="2014-02-01 00:00",
                end="2014-05-01 00:00",max.gap=180,decimals=15)

# organise a data.frame
input_df  = data.frame(date = zoo::index(input), data = zoo::coredata(input))

# save the RData file to e.g. a temp file, or your project root directory
```

```
#temp_file_path <- tempfile()
# save(input_df, file=temp_file_path)

# project_root_path <- "."
# save(input_df, file=project_root_path)


# call the oulier function and navigate to where the "test.RData" is stored
outlier()
```

`preci` *Daily precipitation (raw)*

Description

Returns an example dataset of daily precipitation data in $mm\ d^{-1}$ from 2012-2015 originating from weather stations surrounding the Loetschental in the Swiss Alps. The data was obtained from the nine nearest weather stations (6-to 43-km distance to the site, including Adelboden, Blatten, Grächen, Montana, Jungfraujoch, Sion, Ulrichen, Visp, and Zermatt; Federal Office of Meteorology and Climatology MeteoSwiss).

Usage

```
preci
```

Format

Provides an `is.trex`-compliant object with 1415 rows and 1 column.

index Date of the measurements in solar time ("yyyy-mm-dd") (`character`)

value Daily precipitation ($mm\ d^{-1}$) from local weather stations (`numeric`)

References

Peters RL, Speich M, Pappas C, Kahmen A, von Arx G, Graf Pannatier E, Steppe K, Treydte K, Stritih A, Fonti P. 2018. Contrasting stomatal sensitivity to temperature and soil drought in mature alpine conifers. Plant, Cell & Environment 42:1674-1689 doi: 10.1111/pce.13500

sr *Solar radiation measurements (raw)*

Description

Returns an example dataset of solar irradiance monitoring from 2012-2015 at 1300 m a.s.l. in the Swiss Alps (Loetschental, Switzerland; Peters et al. 2019). Solar irradiance (Wm^{-2}) was measured with 15-min resolution using a microstation (Onset,USA, H21-002 Micro Station) and pyranometer (Onset, USA, S-LIB-M003) positioned in an open field.

Usage

```
sr
```

Format

Provides an `is.trex`-compliant object with 135840 rows and 1 column.

index Date of the measurements in solar time ("yyyy-mm-dd") (`character`)

value W m-2 values obtained from the site-specific monitoring (`numeric`)

References

Peters RL, Speich M, Pappas C, Kahmen A, von Arx G, Graf Pannatier E, Steppe K, Treydte K, Stritih A, Fonti P. 2018. Contrasting stomatal sensitivity to temperature and soil drought in mature alpine conifers. Plant, Cell & Environment 42:1674-1689 doi: 10.1111/pce.13500

tdm.data *Sap flow measurements*

Description

Returns an example thermal dissipation probe (TDM) dataset with time stamp (n = 1) and doy-columns (n = 3), a value and a species column. TDM ΔV measurements are provided at a 15-minute resolution from 2012-2015 from a Norway spruce (*Picea abies* (L.) Karts.) growing at 1300 m a.s.l. in the Swiss Alps (Loetschental, Switzerland; see Peters *et al.* 2019). The presented data is open for public use.

Usage

```
tdm.data
```

Format

Provides a data.frame with 11,6466 rows and 6 columns.

timestamp Date and time of the measurements (`character`)

year Year of measurements (`integer`)

doy Day of year (`integer`)

hour Hour of the measurements (`character`)

value ΔV values obtained from TDM measurements (`numeric`)

species Monitored species (`character`)

`tdm_cal.sfd` *Calculate sap flux density*

Description

The acquired K values are calculated to sap flux density (SFD in $cm^3 cm^{-2} h^{-1}$). As many calibration curves exist (see Peters *et al.* 2018; Flo *et al.* 2019), the function provides the option to calculate SFD using calibration experiment data from the meta-analyses by Flo *et al.* (2019; see `cal.data`). Additionally, raw calibration data can be provided or parameters a and b for a specific calibration function (aK^b) can be provided. The algorithm determines for each calibration experiment dataset the calibration curve ($SFD = aK^b$) and calculates SFD from either the mean of all curves and the 95% confidence interval of either all curves, or bootstrapped resampled uncertainty around the raw calibration experiment data when one calibration dataset is selected.

Usage

```
tdm_cal.sfd(
  input,
  genus,
  species,
  study,
  wood,
  calib,
  a,
  b,
  decimals,
  make.plot = TRUE,
  df = FALSE
)
```

Arguments

`input`	An `is.trex`-compliant object (`zoo` vector, `data.frame`) of K values containing a timestamp and a value column.

genus	Optional, character vector specifying genus-specific calibration data that should be used (e.g., c("Picea","Larix")). See cal.data for the specific labels (default = Use all).
species	Optional, character vector of species specific calibration data that should be used, e.g. c("Picea abies"). See cal.data for the specific labels (default = Use all).
study	Optional character vector of study specific calibration data that should be used (e.g., c("Peters et al. 2018")). See cal.data for the specific labels (default= Use all).
wood	Optional, character vector of wood type specific calibration data that should be used (one of c("Diffuse-porous","Ring-porous","Coniferous")). See cal.data for the specific labels (default= Use all).
calib	Optional data.frame containing raw calibration experiment values. Required columns include: [,1] K = K values measured with the probe (numeric), and [,2] SFD = Gravimetrically measured sap flux density ($cm^3cm^{-2}h^{-1}$)) (numeric). If not provided, literature values are used.
a	Optional, numeric value for the calibration curve ($SFD = aK^b$). No uncertainty can be calculated when this value is provided.
b	Optional, numeric value for the calibration curve ($SFD = aK^b$). No uncertainty can be calculated when this value is provided.
decimals	Integer, the number of decimals of the output (default = 6).
make.plot	Logical; if TRUE, a plot is generated showing the calibration curve with $Kvssapfluxdensity$ ($cm^3cm^{-2}h^{-1}$)).
df	Logical; If TRUE, output is provided in a data.frame format with a timestamp and a value column. If FALSE, output is provided as a zoo vector object (default = FALSE).

Details

The function fits a calibration curve ($SFD = aK^b$) through all selected raw calibration data. If multiple studies are provided, multiple calibration curves are fitted. In case a single calibration dataset is provided a bootstrap resampling is applied (n = 100) to determined the mean and 95% confidence interval of the fit. When multiple calibration curves are requested the mean and 95% confidence interval is determined on the fitted functions. The mean and confidence interval are used to calculate SFD from K.

Value

A list containing either a zoo object or data.frame in the appropriate format for other functionalities (see tdm_dt.max output specifications), as well as all SFD values for each method are provided and added to the is.trex-compliant object (e.g., [['sfd.pd']], [['sfd.mw']]) if this format was provided as an input, and, finally, a data.frame is provided with the mean and 95% confidence interval of the applied calibration functions (see [['model.ens']]). If an individual time series is provided for input with K values an alternative output is provided:

input K values provided as input.

sfd.input SFD values calculated for the input according to the mean of the calibration function.

model.ens A `data.frame` providing the mean and 95% confidence interval of the applied calibration function.

out.param A `data.frame` with the coefficients of calibration function.

References

Peters RL, Fonti P, Frank DC, Poyatos R, Pappas C, Kahmen A, Carraro V, Prendin AL, Schneider L, Baltzer JL, Baron-Gafford GA, Dietrich L, Heinrich I, Minor RL, Sonnentag O, Matheny AM, Wightman MG, Steppe K. 2018. Quantification of uncertainties in conifer sap flow measured with the thermal dissipation method. New Phytologist 219:1283-1299 <doi: 10.1111/nph.15241>

Flo V, Martinez-Vilalta J, Steppe K, Schuldt B, Poyatos, R. 2019. A synthesis of bias and uncertainty in sap flow methods. Agricultural and Forest Meteorology 271:362-374 doi: 10.1016/j.agrformet.2019.03.012

Examples

```
#calculating sap flux density

raw   <-is.trex(example.data(type="doy"),
    tz="GMT",time.format="%H:%M",
    solar.time=TRUE,long.deg=7.7459,
    ref.add=FALSE)

input <-dt.steps(input=raw,start="2014-05-08 00:00",
end="2014-07-25 00:50",
     time.int=15,max.gap=60,decimals=10,df=FALSE)

input[which(input<0.2)]<-NA

input <-tdm_dt.max(input, methods=c("pd","mw","dr"),
     det.pd=TRUE,interpolate=FALSE,max.days=10,df=FALSE)

output.data<-tdm_cal.sfd(input,make.plot=TRUE,df=FALSE,
wood="Coniferous", decimals = 6)


str(output.data)
plot(output.data$sfd.pd$sfd,ylim=c(0,10))
lines(output.data$sfd.pd$q025,lty=1,col="grey")
lines(output.data$sfd.pd$q975,lty=1,col="grey")
lines(output.data$sfd.pd$sfd)

output.data$out.param
```

tdm_damp *Signal dampening correction*

Description

When long-term K time series (~3 years) are provided, one can perform a signal dampening correction (when sensors were not re-installed; see Peters *et al.* 2018). Applying the signal dampening correction requires visually inspecting the correction curve (see `make.plot = TRUE`). The correction curve is constructed with the day since installation and the day of year (DOY) to account for seasonal changes in K values. The function returns corrected K values and the applied correction curve.

Usage

```
tdm_damp(input, k.threshold = 0.05, make.plot = TRUE, df = FALSE)
```

Arguments

`input`	An `is.trex`-compliant object (`zoo` vector, `data.frame`) of K values containing a timestamp and a value column.
`k.threshold`	Numeric, the threshold below which daily maximum K values should not be considered (default = 0.05).
`make.plot`	Logical; if `TRUE`, a plot is generated presenting the correction curve and the K time series.
`df`	Logical; If `TRUE`, output is provided in a `data.frame` format with a timestamp and a value column. If `FALSE`, output is provided as a `zoo` vector object (default = FALSE).

@details The function fits a correction curve for signal dampening (e.g., due to wounding) according to Peters *et al.* (2018). A sensor specific function is fitted to daily maximum K values (considering a minimum cut-off threshold; see `k.threshold`). Dependent variables for the function include seasonality (DOY) and days since installation (t). First, seasonal effects are removed by correcting the K series (residuals; $Kresid$) to a second-order polynomial with DOY. These residuals are then used within a non-linear model:

$$K_{resid} = (a + b * t)/(1 + c * t + d * t^2)$$

The fitted parameters for t (with a, b, c and d) are used to correct K and scale it to the maximum within the first year of installation. **Note, that the stability of the fit has to be visually inspected before using the output data**.

Value

A `zoo` object or `data.frame` in the appropriate format for other functionalities. See `tdm_dt.max` output specifications. All K values for each method are provided when an `is.trex`-object was used as input. If an individual time series was provided for input with K values an alternative output is given:

k.cor corrected K values according to the correction curve.

k K values provided as input.

damp.mod `data.frame` with the coefficients of the correction curve.

References

Peters RL, Fonti P, Frank DC, Poyatos R, Pappas C, Kahmen A, Carraro V, Prendin AL, Schneider L, Baltzer JL, Baron-Gafford GA, Dietrich L, Heinrich I, Minor RL, Sonnentag O, Matheny AM, Wightman MG, Steppe K. 2018. Quantification of uncertainties in conifer sap flow measured with the thermal dissipation method. New Phytologist 219:1283-1299 doi: 10.1111/nph.15241

Examples

```
 #correct for dampening of the signal
raw   <-
  is.trex(
    example.data(type = "doy"),
    tz = "GMT",
    time.format = "%H:%M",
    solar.time = TRUE,
    long.deg = 7.7459,
    ref.add = FALSE
  )
input <-
  dt.steps(
    input = raw,
    time.int = 15,
    max.gap = 60,
    decimals = 6,
    df = FALSE
  )
input[which(input < 0.2)] <- NA
input <-
  tdm_dt.max(
    input,
    methods = c("pd", "mw", "dr"),
    det.pd = TRUE,
    interpolate = FALSE,
    max.days = 10,
    df = FALSE
  )
output.data <- tdm_damp(input,
                    k.threshold = 0.05,
                    make.plot = TRUE,
                    df = FALSE)
str(output.data)
head(output.data[["k.dr"]])
plot(output.data[["k.dr"]], ylab = expression(italic("K")))
```

tdm_dt.max *Calculate zero-flow conditions*

Description

Determine zero flow conditions (ΔT_{max}; or ΔV_{max}) according to four methods; namely, 1) predawn (pd), 2) moving-window (mw), 3) double regression (dr), and 4) environmental-dependent (ed) as applied in Peters *et al.* 2018. The function can provide (ΔT_{max} values and subsequent *K* values for all methods. All outputs are provided in a list including the input data and calculated outputs.

Usage

```
tdm_dt.max(input, methods = c("pd","mw","dr","ed"),
zero.end = 8*60,
zero.start =  1*60,
interpolate = FALSE, det.pd = TRUE,
max.days = 7,
ed.window = 2*60,
vpd.input,
sr.input,
sel.max,
criteria = c(sr = 30, vpd = 0.1, cv = 0.5),
df = FALSE)
```

Arguments

input	An is.trex-compliant object of K values containing a timestamp and a value column.
methods	Character vector of the requested ΔT_{max} methods. Options include "pd" (predawn), "mw" (moving-window), "dr" (double regression), and "ed" (environmental-dependent; default= c("pd","mw","dr")).
zero.end	Numeric, optionally defines the end of the predawn period. Values should be in minutes (e.g. predawn conditions until 08:00 = 8 * 60). When not provided, the algorithm will automatically analyse the cyclic behaviour of the data and define the day length.
zero.start	Numeric, optionally defines the beginning of the predawn period. Values should be in minutes (e.g., 01:00 = 1*60).
interpolate	Logical: if TRUE, detected ΔT_{max} values are linearly interpolated. If FALSE, constant ΔT_{max} values will be selected daily (default = FALSE).
det.pd	Logical; if TRUE and no zero.end and zero.start values are provided, predawn ΔT_{max} will be determined based on cyclic behaviour of the entire time-series (default = TRUE).
max.days	Numeric, defines the number of days which the mw and dr methods will consider for their moving windows.

ed.window	Numeric, defines the length of the period considered for assessing the environmental conditions and stable ΔT_{max} values.
vpd.input	An is.trex-compliant object (zoo time-series or data.frame) with a timestamp and a value column containing the vapour pressure deficit (*vpd*; in kPa) with the same temporal extent and time steps as the input data.
sr.input	An is.trex-compliant object (zoo time-series or data.frame) with a timestamp and a value column the solar radiation data (*sr*; e.g., global radiation or *PAR*).
sel.max	Optional zoo time-series or data.frame with the specified ΔT_{max}. This option is included to change predawn ΔT_{max} values selected with the ed method.
criteria	Numeric vector, thresholds for the ed method. Thresholds should be provided for all environmental data included in the function (e.g. c(sr = 30,vpd = 0.1); coefficient of variation, *cv* = 0.5)
df	Logical; if TRUE, output is provided in a data.frame format with a timestamp and a value (ΔT or ΔV) column. If FALSE, output is provided as a zoo object (default = FALSE).

Details

There are a variety of methods which can be applied to determine zero-flow conditions. Zero-flow conditions are required to calculate $K = (\Delta T_{max} - \Delta T)/\Delta T$. A detailed description on the methods is provided by Peters *et al.* (2018). In short, the pd method entails the selection of daily maxima occurring prior to sunrise. This method assumes that during each night zero-flow conditions are obtained. The algorithm either requires specific times within which it searches for a maximum, or it analyses the cyclic pattern within the data and defines this time window. The mw method uses these predawn ΔT_{max} values and calculates the maximum over a multi-day moving time-window (e.g., 7 days). The dr methods is applied by calculating the mean over predawn ΔT_{max} with a specified multi-day window, removing all values below the mean, and calculating a second mean over the same multi-day window and uses these values as ΔT_{max}. The ed method selects predawn ΔT_{max} values based upon 2-hour averaged environmental conditions prior to the detected time for the predawn ΔT_{max}. These environmental conditions include low vapour pressure deficit (in kPa) and low solar irradiance (e.g., in W m-2). In addition, the coefficient of variation (cv) of predawn ΔT_{max} are scanned for low values to ensure the selection of stable zero-flow conditions.

Value

A named list of zoo time series or data.frame objects in the appropriate format for further processing. List items include:

max.pd ΔT_{max} time series as determined by the pd method.

max.mw ΔT_{max} time series as determined by the mw method.

max.dr ΔT_{max} time series as determined by the dr method.

max.ed ΔT_{max} time series as determined by the ed method.

daily_max.pd daily predawn ΔT_{max} as determined by pd.

daily_max.mw daily predawn ΔT_{max} as determined by mw.

daily_max.dr daily predawn ΔT_{max} as determined by dr.

daily_max.ed daily predawn ΔT_{max} as determined by ed.

all.pd exact predawn ΔT_{max} values detected with pd.

all.ed exact predawn ΔT_{max} values detected with ed.

input ΔT input data.

ed.criteria data.frame of the applied environmental and variability criteria used within ed.

methods data.frame of applied methods to detect ΔT_{max}.

k.pd K values calculated by using the pd method.

k.mw K values calculated by using the mw method.

k.dr K values calculated by using the dr method.

k.ed K values calculated by using the ed method.

Examples

```
#perform Delta Tmax calculations
raw <- is.trex(example.data(type = "doy"),
     tz = "GMT", time.format = "%H:%M", solar.time = TRUE,
     long.deg = 7.7459, ref.add = FALSE)
input <- dt.steps(input = raw, start = "2014-05-08 00:00",
         end = "2014-07-25 00:50", time.int = 15, max.gap = 60,
         decimals = 6, df = FALSE)
input[which(input<0.2)]<- NA
output.max <- tdm_dt.max(input, methods = c("pd", "mw", "dr"),
                 det.pd = TRUE, interpolate = FALSE,
                 max.days = 10, df = FALSE)

str(output.max)

plot(output.max$input, ylab = expression(Delta*italic("V")))

lines(output.max$max.pd, col = "green")
lines(output.max$max.mw, col = "blue")
lines(output.max$max.dr, col = "orange")

points(output.max$all.pd, col = "green", pch = 16)

legend("bottomright", c("raw", "max.pd", "max.mw", "max.dr"),
        lty = 1, col = c("black", "green", "blue", "orange") )
```

---

tdm_hw.cor *Heartwood correction*

Description

The function corrects for the proportion of the probe that is installed within the non-conductive heartwood according to Clearwater *et al.* (1999). The function requires ΔT_{max}, the probe length and the sapwood thickness. The correction is applied on the ΔT (or ΔV) values and K is recalculated accordingly. When an `is.trex`-compliant object is provided, the K values for each method are determined (see `tdm_dt.max`.

Usage

```
tdm_hw.cor (input, dt.max, probe.length = 20,
            sapwood.thickness = 18, df = FALSE)
```

Arguments

`input`	A `tdm_dt.max` ouput or `is.trex`-compliant object of ΔT (or ΔV) values containing a timestamp and a value column.
`dt.max`	Optional `zoo` object or `data.frame` (columns = "timestamp" or "value") containing the ΔT_{max} when no `is.trex`-compliant object is provided.
`probe.length`	Numeric, the length of the TDM probes in mm.
`sapwood.thickness`	Numeric, the sapwood thickness in mm.
`df`	Logical; If `TRUE`, output is provided in a `data.frame` format with a timestamp and a value column. If `FALSE`, output is provided as a `zoo` vector object (default = `FALSE`).

Details

The function applied the correction provided by Clearwater *et al.* 1999. ΔT (or ΔV) was corrected (denoted as ΔT_{sw}) for the proportion of the probe that was inserted into the conducting sapwood vs the proportion of the probe that was inserted into the nonconductive heartwood (γ in mm mm-1). Together with ΔT_{max}, ΔT was corrected according to the following equation:

$$\Delta T_{sw} = (\Delta T \ - \ (1 \ \breve{} \ \gamma) * \Delta T_{max})/\gamma$$

ΔT_{sw} together with ΔT_{max} is then recalculated to K.

Value

A `zoo` object or `data.frame` in the appropriate format for other functionalities. See `tdm_dt.max` for output specifications. All K values for each method are provided when an output from `tdm_dt.max` was provided. If individual time series are provided for `input` and `tdm_dt.max` an alternative output is provided:

input = ΔT input data.

dt.max $\Delta T_{max} \Delta T max$ input data.

dtsw Corrected ΔT data.

k.value K values calculated according to Clearwater et al. (1999).

settings data.frame of the applied `probe.length` and `sapwood.thickness`

References

Clearwater MJ, Meinzer FC, Andrade JL, Goldstein G, Holbrook NM. 1999. Potential errors in measurement of nonuniform sap flow using heat dissipation probes. Tree Physiology 19:681–687 doi: 10.1093/treephys/19.10.681

Examples

```
#correct for probes being inserted into the heartwood
raw   <-is.trex(example.data(type="doy"),
          tz="GMT",time.format="%H:%M",solar.time=TRUE,
          long.deg=7.7459,ref.add=FALSE)
input <- dt.steps(input=raw,
                  start="2014-05-08 00:00",
                  end="2014-07-25 00:50",
                 time.int=15,max.gap=60,decimals=6,df=F)
input[which(input<0.2)]<-NA
input <-tdm_dt.max(input, methods=c("pd","mw","dr"),
                det.pd=TRUE,interpolate=FALSE,max.days=10,df=FALSE)
output.data<-tdm_hw.cor(input,probe.length=20,
                  sapwood.thickness=18,df=FALSE)
plot(output.data$k.dr,col="orange")
lines(input$k.dr)
```

`tdm_uncertain` *Uncertainty and sensitivity analysis*

Description

Quantifies the induced uncertainty on SFD and K time series due to the variability in input parameters applied during TDM data processing. Moreover, it applies a global sensitivity analysis to quantify the impact of each individual parameter on three relevant outputs derived from SFD and K, namely: i) the mean daily sum of water use, ii) the variability of maximum daily SFD or K values, iii) and the duration of daily sap flow. This function provides both the uncertainty and sensitivity indices, as time-series of SFD and K with the mean, standard deviation (sd) and confidence interval (CI) due to parameter uncertainty. **Users should ensure that no gaps are present within the input data and environmental time series**.

Usage

```
tdm_uncertain(
  input,
  vpd.input,
  sr.input,
  method = "pd",
  n = 2000,
```

```
  zero.end = 8 * 60,
  range.end = 16,
  zero.start = 1 * 60,
  range.start = 16,
  probe.length = 20,
  sw.cor = 32.28,
  sw.sd = 16,
  log.a_mu = 4.085,
  log.a_sd = 0.628,
  b_mu = 1.275,
  b_sd = 0.262,
  max.days_min = 1,
  max.days_max = 7,
  ed.window_min = 8,
  ed.window_max = 16,
  criteria.vpd_min = 0.05,
  criteria.vpd_max = 0.5,
  criteria.sr_mean = 30,
  criteria.sr_range = 30,
  criteria.cv_min = 0.5,
  criteria.cv_max = 1,
  min.sfd = 0.5,
  min.k = 0,
  make.plot = TRUE,
  df = FALSE,
  ncores
)
```

Arguments

input	An is.trex-compliant object (zoo object
vpd.input	An is.trex-compliant object (zoo object, data.frame) containing a timestamp and a vapour pressure deficit (VPD; in kPa) column with the same temporal extent and time steps as the input object. This input is required when using the environmental dependent ("ed") method.
sr.input	An is.trex-compliant object (zoo object, data.frame) a timestamp and a solar radiation data (sr; e.g., global radiation or PAR) column with the same temporal extent and time steps as the input object. This input is required when using the environmental dependent ("ed") method.
method	Character, specifies the ΔT_{max} method on which the sensitivity and uncertainty analysis are to be performed on (see tdm_dt.max). Only one method can be selected, including the pre-dawn ("pd"), moving window ("mw"), double regression ("dr") or the environmental dependent ("ed") method (default = "pd").
n	Numeric, specifies the number of times the bootstrap resampling procedure is repeated (default = 2000). Keep in mind that high values increase processing time.
zero.end	Numeric, defines the end of the predawn period. Values should be in minutes (e.g., predawn conditions until 8:00 = 8*60; default = 8*60).

`range.end` Numeric, defines the number of time steps for `zero.end` (the minimum time step of the input) for which an integer sampling range will be defined (default = 16, assuming a 15-min resolution or a 2 hour range around `zero.end`).

`zero.start` Numeric, defines the start of the predawn period. Values should be in minutes (e.g., predawn conditions from 1:00 = 1*60; default = 1*60).

`range.start` Numeric, defines the number of time steps for `zero.start` (the minimum time step of the input) for which an integer sampling range will be defined (default = 16, assuming a 15-min resolution or a 2 hour range around `zero.start`).

`probe.length` Numeric, the length of the TDM probes in mm (see `tdm_hw.cor`; default = 20 mm).

`sw.cor` Numeric, the sapwood thickness in mm. Default conditions assume the sapwood thickness is equal to a standard probe length (default = 20).

`sw.sd` Numeric, the standard deviation for sampling sapwood thickness sampling from a normal distribution (default = 16 mm; defined with a European database on sapwood thickness measurements).

`log.a_mu` Numeric, value providing the natural logarithm of the calibration parameter a (see `tdm_cal.sfd`; $SFD = aK^b$). This value can be obtained from `tdm_cal.sfd` (see `out.param`). Default conditions are determined by using all calibration data as described in `cal.data` (default = 4.085).

`log.a_sd` Numeric, the standard deviation of the a parameter (see `log.a_mu`) used within the calibration curve for calculating SFD (default = 0.628).

`b_mu` Numeric, the value of the calibration parameter b (see `tdm_cal.sfd`; $SFD = aK^b$). This value can be obtained from `tdm_cal.sfd` (see `out.param`). Default conditions are determined by using all calibration data as described in `cal.data` (default = 1.275).

`b_sd` Numeric, the standard deviation of the b parameter (see `log.a_mu`) used within the calibration curve for calculating SFD (default = 0.262).

`max.days_min` Numeric, the minimum value for an integer sampling range of `max.days` (see `tdm_dt.max` for the `"mw"` and `"dr"` ΔT_{max} method). As the `"mw"` and `"dr"` method apply a rolling maximum or mean, the provided value should be an uneven number (see `tdm_dt.max`; default = 15; required for the `"mw"` and `"dr"` ΔT_{max} method).

`max.days_max` Numeric, the maximum value for an integer sampling range of `max.days` (see `tdm_dt.max` for the `"mw"` and `"dr"` ΔT_{max} method). As the `"mw"` and `"dr"` method apply a rolling maximum or mean, the provided value should be an uneven number (see `tdm_dt.max`; default = 5; required for the `"mw"` and `"dr"` ΔT_{max} method).

`ed.window_min` Numeric, the minimum number of time steps for the `ed.window parameter` (see `tdm_dt.max`; the minimum time step of the input) for which an integer sampling range will be defined (default = 8, assuming a 15-min resolution or a 2 hour range; required for the `"ed"` ΔT_{max} method).

`ed.window_max` Numeric, the maximum number of time steps for the `ed.window` sampling range (default = 16, assuming a 15-min resolution or a 4 hour range; required for the `"ed"` ΔT_{max} method).

criteria.vpd_min
: Numeric, value in kPa defining the minimum for the fixed sampling range to define the vapour pressure deficit (VPD) threshold to establish zero-flow conditions (default = 0.05 kPa; see tdm_dt.max; required for the "ed" ΔT_{max} method).

criteria.vpd_max
: Numeric, value in kPa defining the maximum for the fixed sampling range to define the VPD threshold to establish zero-flow conditions (default = 0.5 kPa; required for the "ed" ΔT_{max} method).

criteria.sr_mean
: Numeric value defining the mean sr.input value around which the fixed sampling range for the solar irradiance threshold should be established for defining zero-flow conditions (see tdm_dt.max; default = 30 W m-2; required for the "ed" ΔT_{max} method).

criteria.sr_range
: Numeric, the range (in %) around criteria.sr_mean for establishing the solar irradiance threshold (see tdm_dt.max; default = 30%; required for the "ed" ΔT_{max} method).

criteria.cv_min
: Numeric, value (in %) defining the minimum value for the fixed sampling range to determine the coefficient of variation (CV) threshold for establishing zero-flow conditions (default = 0.5%; see tdm_dt.max; required for the "ed" ΔT_{max} method).

criteria.cv_max
: Numeric, value (in %) defining the maximum value for the fixed sampling range to determine the coefficient of variation (CV) threshold for establishing zero-flow conditions (default = 1%; see tdm_dt.max; required for the "ed" ΔT_{max} method).

min.sfd
: Numeric, defines at which SFD ($cm^3cm^{-2}h^{-1}$) zero-flow conditions are expected. This parameter is used to define the duration of daily sap flow based on SFD (default = $0.5cm^3cm^{-2}h^{-1}$).

min.k
: Numeric value defining at which K (dimensionless, -) zero-flow are expected. This parameter is used to define the duration of daily sap flow based on K (default = 0).

make.plot
: Logical; If TRUE, a plot is generated presenting the sensitivity and uncertainty analyses output (default = TRUE).

df
: Logical; If TRUE, output is provided in a data.frame format with a timestamp and a value column. If FALSE, output is provided as a zoo vector object (default = FALSE).

ncores
: Numeric, number of cores to use for parallel processing. If missing, defaults to available cores - 1, or 1 if single-core machine.

Details

Uncertainty and sensitivity analysis can be performed on TDM ΔT (or ΔV) measurements. The function applies a Monte Carlo simulation approach (repetition defined by n) to determine the variability in relevant output variables (defined as uncertainty) and quantifies the contribution of each

parameter to this uncertainty (defined as sensitivity). To generate variability in the selected input parameters a Latin Hypercube Sampling is performed with a default or user defined range of parameter values per ΔT_{max} method (see `tdm_dt.max()`). The sampling algorithm generates multiple sampling distributions, including an integer sampling range (for `zero.start`, `zero.end`, `max.days`, and `ed.window`), a continuous sampling range (criteria for `sr`, `vpd` and `cv`), and a normal distribution (for `sw.cor` and calibration parameters `a` and `b`). Within this algorithm no within-day interpolations are made between the ΔT_{max} points (see `tdm_dt.max`, `interpolate = FALSE`). This approach ensures near-random sampling across different types of sampling distributions, while avoiding the need for increasing the number of replicates (which increases computation time). For the application of this approach one needs to; i) select the output of interest, ii) identify the relevant input parameters, and iii) determine the parameter range and distribution. For a given time-series three output variables are considered, calculated as the mean over the entire time-series, to be relevant, namely; i) mean daily sum of water use (or Sum, expressed in $cm^3 cm^{-2} d^{-1}$ for SFD and unitless for K), ii) the variability of maximum SFD or K values (or CV, expressed as the coefficient of variation in % as this alters climate response correlations), and iii) the duration of daily sap flow based on SFD or K (or Duration, expressed in hours per day dependent on a threshold, see `min.sfd` and `min.k`). A minimum threshold to define zero-flow SFD or K is required for the duration calculation as small variations in night-time SFD or K are present. All data-processing steps (starting with `"tdm_"`) are incorporated within the function, excluding `tdm_damp()` due to the need for detailed visual inspection and significantly longer computation time.

For the sensitivity analysis the total overall sensitivity indices are determined according strategy originally proposed by Sobol' (1993), considering the improvements applied within the sensitivity R package. The method proposed by Sobol' (1993) is a variance-based sensitivity analysis, where sensitivity indices (dimensionless from 0 to 1) indicate the partial variance contribution by a given parameter over the total output variance (e.g., Pappas *et al.* 2013). This global sensitivity analysis facilitates the identification of key parameters for data-processing improvement and highlights methodological limitations. Users should keep in mind that parameter ranges represent a very critical component of any sensitivity analysis and should be critically assessed and clearly reported for each case and analytical purpose. Moreover, it is advised to run this function on one growing season of input data to reduce processing time.

Value

A named `list` of `zoo` or `data.frame` objects in the appropriate format for other functionalities. Items include:

output.data data.frame containing uncertainty and sensitivity indices for SFD and K and the included parameters. This includes the mean uncertainty/sensitivity [,"mean"], standard deviation [,"sd"], upper [,"ci.min"] and lower [,"ci.max"] 95% confidence interval.

output.sfd zoo object or data.frame with the SFD time series obtained from the bootstrap resampling. This includes the mean uncertainty/sensitivity [,"mean"], standard deviation [,"sd"], upper [,"CIup"] and lower [,"CIlo"] 95% confidence interval.

output.k zoo object or data.frame with the K time series obtained from the bootstrap resampling. This includes the mean uncertainty/sensitivity [,"mean"], standard deviation [,"sd"], upper [,"ci.max"] and lower [,"ci.min"] 95% confidence interval.

param a data.frame with an overview of selected parameters used within `tdm_uncertain()` function.

References

Sobol' I. 1993. Sensitivity analysis for nonlinear mathematical models. Math. Model Comput. Exp. 1:407-414

Pappas C, Fatichi S, Leuzinger S, Wolf A, Burlando P. 2013. Sensitivity analysis of a process-based ecosystem model: Pinpointing parameterization and structural issues. Journal of Geophysical Research 118:505-528 doi: 10.1002/jgrg.20035

Examples

```
#perform an uncertainty and sensitivity analysis on "dr" data processing
raw   <- example.data(type="doy")
input <- is.trex(raw, tz="GMT", time.format="%H:%M",
           solar.time=TRUE, long.deg=7.7459, ref.add=FALSE, df=FALSE)
input<-dt.steps(input,time.int=15,start="2013-04-01 00:00",
             end="2013-11-01 00:00",max.gap=180,decimals=15)
output<- tdm_uncertain(input, probe.length=20, method="pd",
               n=2000,sw.cor=32.28,sw.sd=16,log.a_mu=3.792436,
               log.a_sd=0.4448937,b_mu=1.177099,b_sd=0.3083603,
               make.plot=TRUE, ncores = 2)
```

TREXr *TREXr: TRree sap flow EXtractor.*

Description

Performs data assimilation, processing and analyses on sap flow data obtained with the thermal dissipation method (TDM). The package includes functions for gap filling time-series data, detecting outliers, calculating data-processing uncertainties and, generating uniform data output and visualisation. The package is designed to deal with large quantities of data and apply commonly used data-processing methods. The functions have been validated on data collected from different tree species across the northern hemisphere (Peters et al. 2018 <doi: 10.1111/nph.15241>).

---

vpd *Vapor pressure deficit measurements (raw)*

Description

Returns an example dataset of vapour pressure deficit (VPD) monitoring from 2012-2015 at 1300 m a.s.l. in the Swiss Alps (Loetschental, Switzerland; Peters *et al.* 2019). Sensors were installed at the site on a central tower (~2.5 m above the ground) within the canopy to measure air temperature and relative humidity (Onset, USA, U23-002Pro) with a 15-min resolution. Vpd (kPa was calculated from the air temperature and relative humidity measurements according to WMO (2008).

Usage

```
vpd
```

Format

Provides an `is.trex`-compliant object with 135840 rows and 1 column.

index Date of the measurements in solar time ("yyyy-mm-dd") (`character`)

value kPa values obtained from the site-specific monitoring (`numeric`)

References

Peters RL, Speich M, Pappas C, Kahmen A, von Arx G, Graf Pannatier E, Steppe K, Treydte K, Stritih A, Fonti P. 2018. Contrasting stomatal sensitivity to temperature and soil drought in mature alpine conifers. Plant, Cell & Environment 42:1674-1689 doi: 10.1111/pce.13500

WMO. 2008.Guide to meteorological instruments and methods of observation, appendix 4B, WMO-No. 8 (CIMO Guide). Geneva, Switzerland: World Meteorological Organization.

Index