Tree Physiological Data Processing in R

WK 24 - A Comprehensive Toolbox for Tree Physiological Data Processing in R

ESA Virtual Meeting
Vital connection in ecology

Organizer
Christoforos Pappas - Centre d'étude de la forêt, Université du Québec à Montréal & Téluq

Co-Organizers
Alexander G. Hurley - GFZ Potsdam

Richard L. Peters - Ghent University & Gembloux Agro Bio-Tech, University of Liège

Roman Zweifel - Swiss Federal Institute for Forest, Snow and Landscape Research WSL

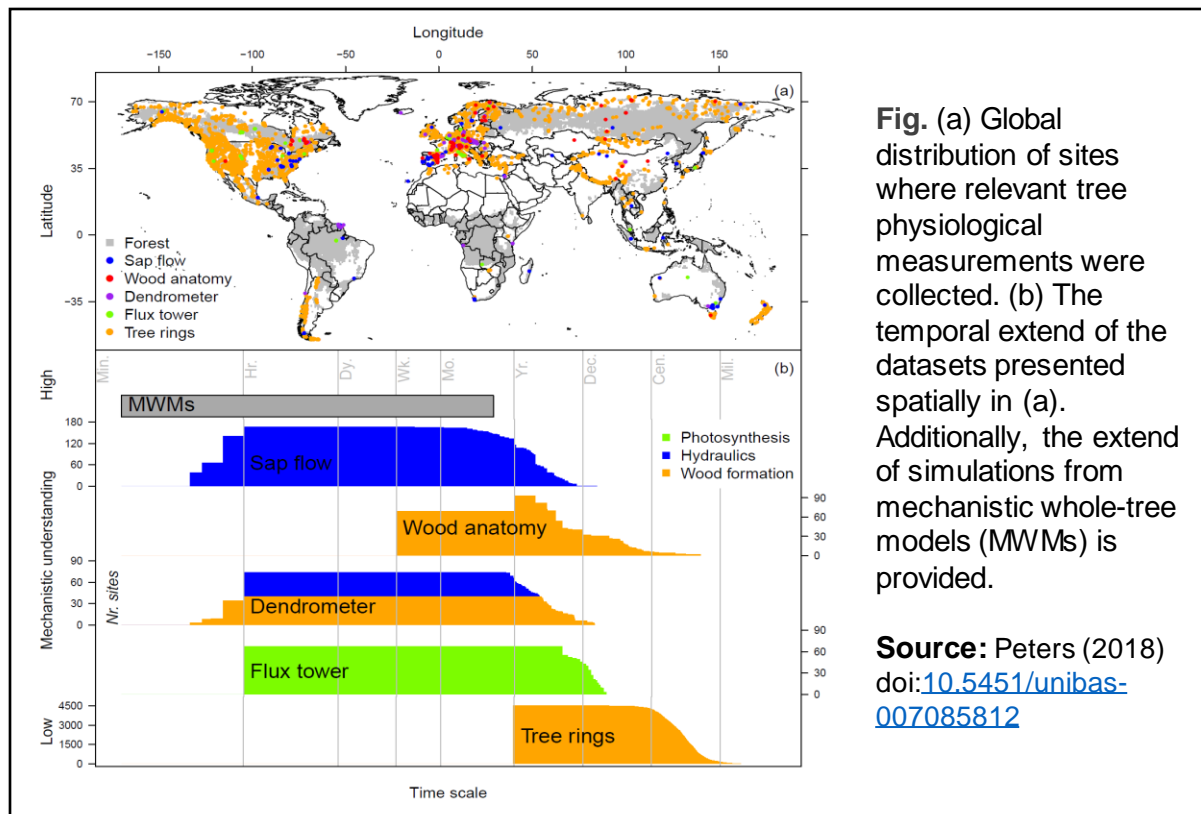## 1. Course - Relevance of processing high-resolution tree physiological measurements


Fig. (a) Global distribution of sites where relevant tree physiological measurements were collected. (b) The temporal extend of the datasets presented spatially in (a). Additionally, the extend of simulations from mechanistic whole-tree models (MWMs) is provided.

Source: Peters (2018) doi:10.5451/unibas-007085812

Time-series measurements

Increasing availability of tree physiological time series on tree growth and water use [Fig.] provide unique opportunities for exploring tree and forest function, health, and resilience to ongoing environmental changes. However, processing such time series data is challenging, due to data quantity and quality, varying time steps, labor-intensive data cleaning, and assumptions for converting raw measurements to physiologically meaningful quantities. To resolve such issues, software tools should be utilized for facilitating data pre- and post-processing in a fast, efficient and reproducible way.

Tree Physiological Data Processing in R

1. Course - Target of the workshop

R | Data cleaning | Sapflow | Dendrometer | treenetproc | TREX
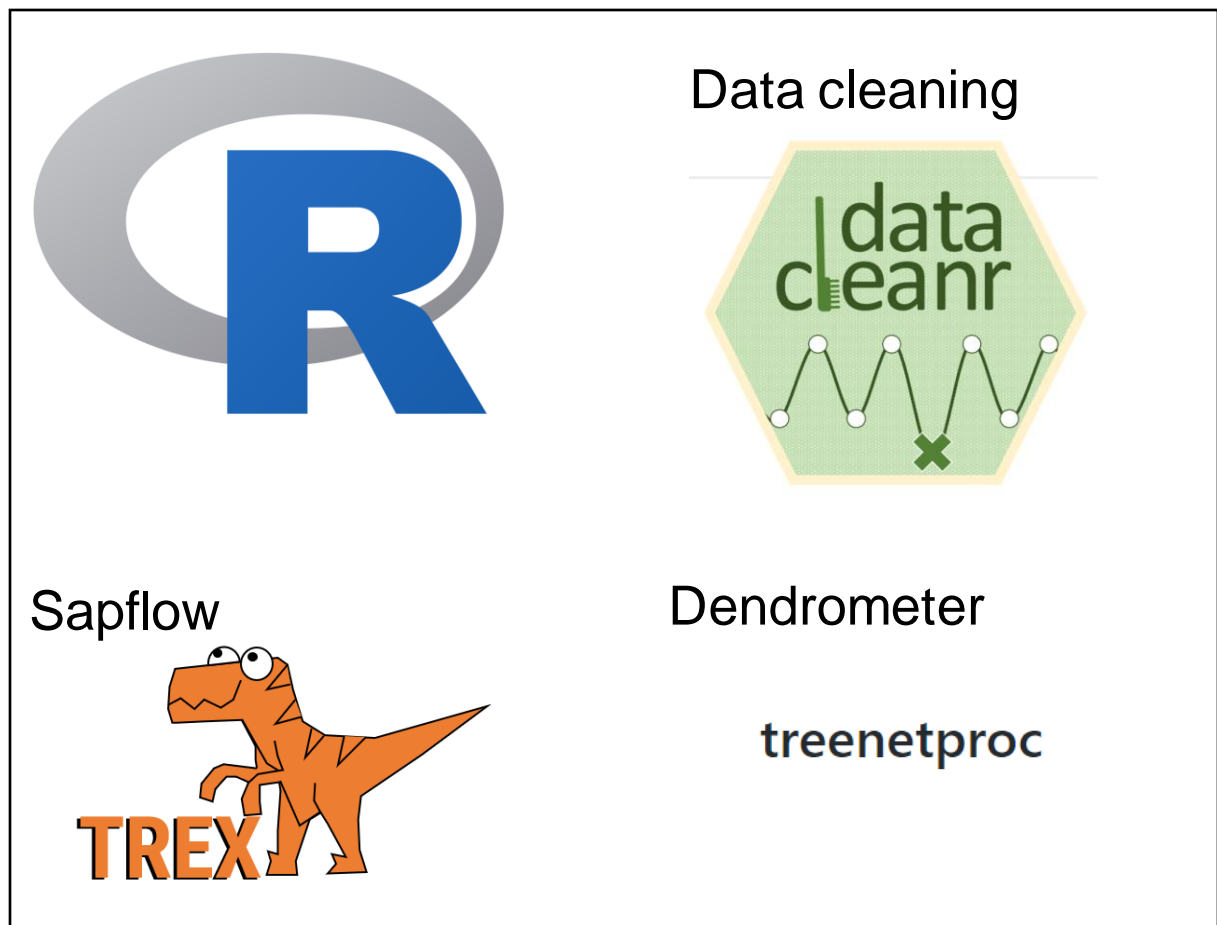
Aim
Within this workshop we aim at providing participants hands-on training on free software tools and their synergies for tree physiological data processing, spanning from interactive visual inspection and cleaning of raw data to advanced data analyses and uncertainty quantification.

Practice
Guided by example datasets (within dedicated tutorials), this interactive workshop presents a benchmark for interactive and reproducible data cleaning, dendrometer and sap flow data processing methods.

Tree Physiological Data Processing in R

1. Course - Logistics for software and code availability
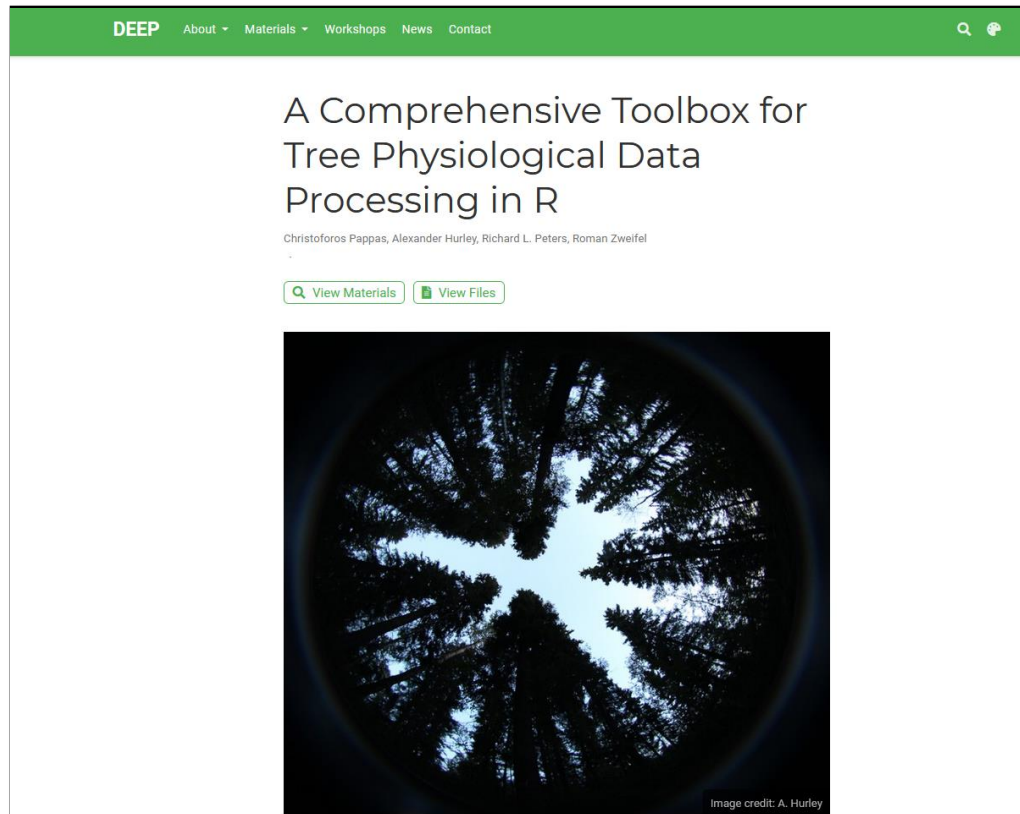
Participants should use their own laptops with pre-installed copies of R/RStudio. Data files and codes for this workshop are available via the below provided link.
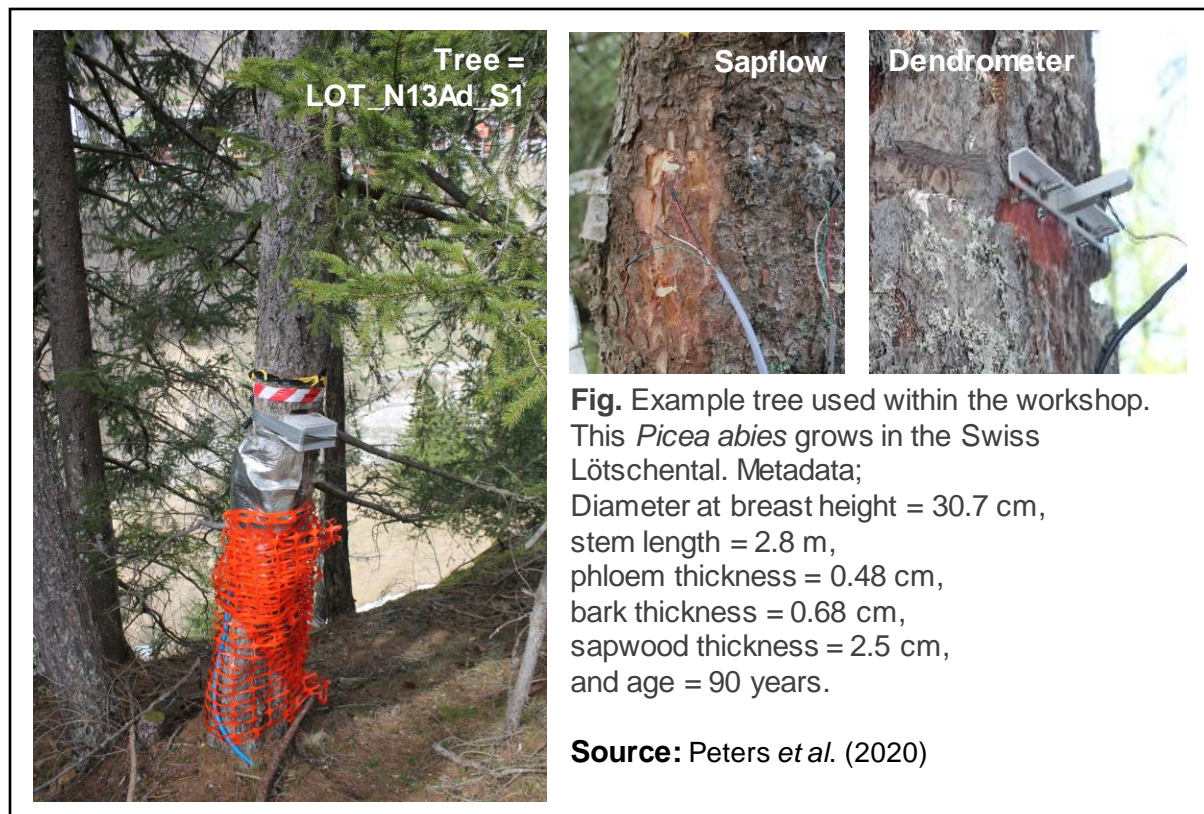
DEEP tools Link [Fig.]
https://deep-tools.netlify.app/talk/esa-2021-cpappas-ahurley-rpeters-rzweifel/

R software
Install R and R studio.
https://www.r-project.org/; https://rstudio.com/

Tree Physiological Data Processing in R

1. Course - Content and structure of the workshop

Fig. Example tree used within the workshop. This *Picea abies* grows in the Swiss Lötschental. Metadata;
Diameter at breast height = 30.7 cm,
stem length = 2.8 m,
phloem thickness = 0.48 cm,
bark thickness = 0.68 cm,
sapwood thickness = 2.5 cm,
and age = 90 years.

Source: Peters *et al.* (2020)

1. Data cleaning. Interactive data cleaning with 'datacleanr' (https://the-hull.github.io/datacleanr/), an R package designed to ensure best data-handling practices of spatiotemporal tree ecophysiological data.

2. Sapflow. Sap flow data processing with the 'TREX' (https://the-hull.github.io/TREX/), including gap-filling thermal dissipation data, converting heat metrics to sap flow, and estimating data-processing uncertainties.

3. Dendrometer. Dendrometer data processing with 'treenetproc' (https://github.com/treenet/treenetproc/), an R package with advanced functionalities on partitioning stem growth and hydraulic signals from dendrometer data and on detecting growing season dynamics.

1. Data cleaning - Utilizing the datacleanr R package
(https://the-hull.github.io/datacleanr/)

Introduction
datacleanr is developed using the shiny package, and relies on informative summaries, visual cues and interactive data selection and annotation. All data-altering operations are documented and converted to valid R code (reproducible recipe), that can be copied, sent to an active RStudio script, or saved to disk.

Availability
The R package datacleanr and all source code is available on GitHub (https://the-hull.github.io/datacleanr) as well as CRAN. In the R software, the package can be installed with the following commands:
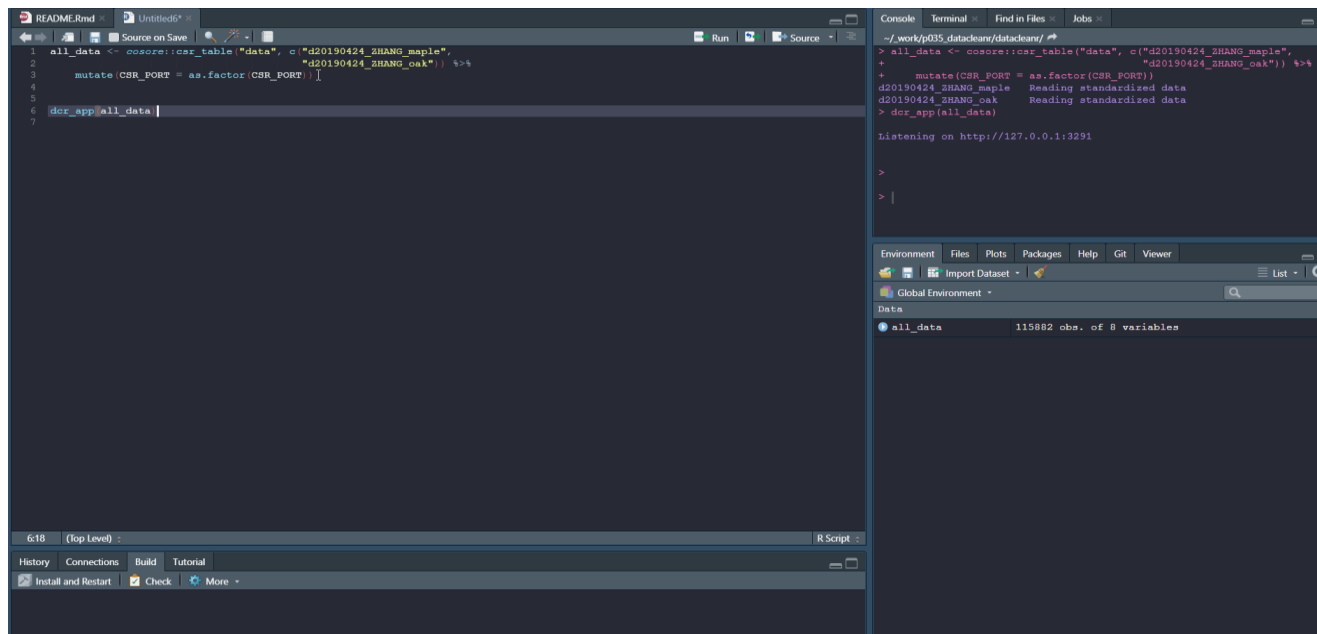
```
# install.packages("devtools")
library(devtools)
devtools::install_github("the-Hull/datacleanr")
# or
install.packages("datacleanr")
```

Link
An R tutorial on the use of datacleanr is accessible via this link:
https://deep-tools.netlify.app/docs-workshops/esa-workshop2021/01_datacleanr/

References

1. Data cleaning - Utilizing the datacleanr R package (https://the-hull.github.io/datacleanr/)

datacleanr *Removing data should finally be reproducible! – A.G. Hurley*

Properties:
- Uses R (links with other packages);
- Freely available (avoid license costs);
- Uses R shiny (interactive approach).

Structure of the tool:
- Set-up & overview;
- Filtering;
- Visual cleaning and annotating;
- Extract.

Tree Physiological Data Processing in R

1. Data cleaning - Utilizing the datacleanr R package
(https://the-hull.github.io/datacleanr/)

Extract *Reproducible recipe to cook up some fresh data*

Fig. Example of dendrometer data recorded for three trees and its resulting datacleanr recipe.

2. Sap flow - Utilizing the TREX R package
(https://the-hull.github.io/TREX)
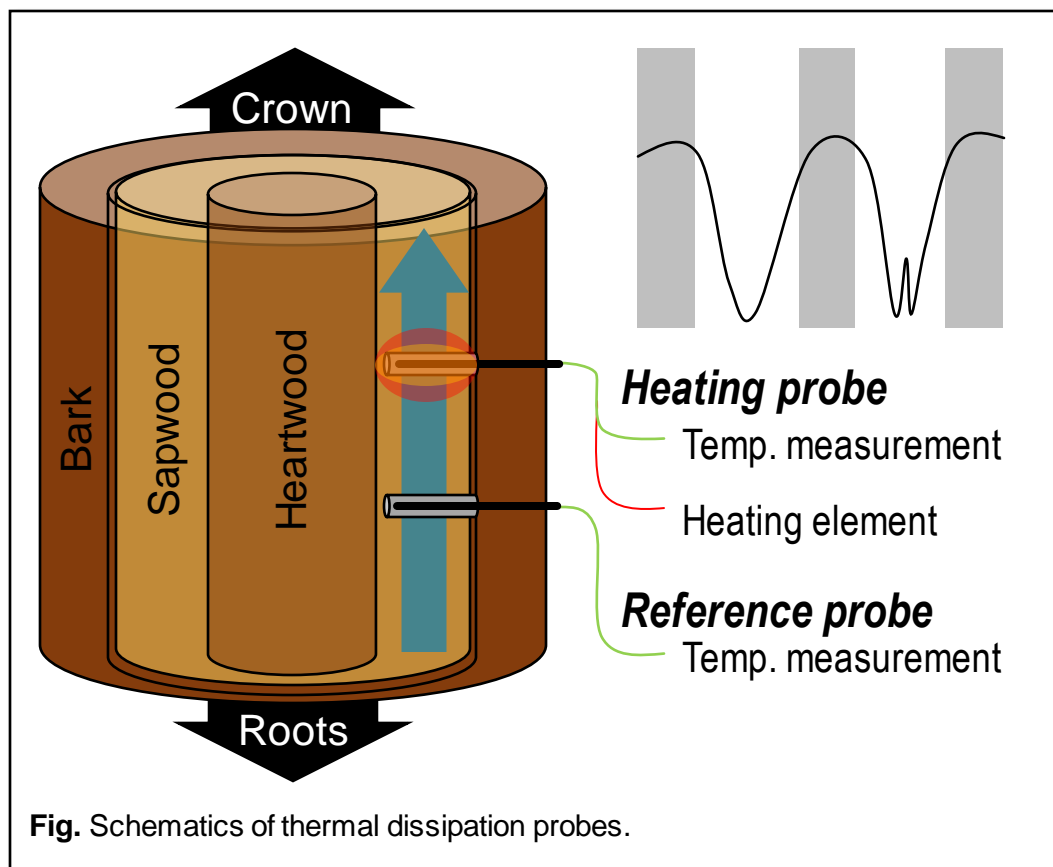
Fig. Schematics of thermal dissipation probes.

Methods in Ecology and Evolution

Assimilate, process and analyse thermal dissipation sap flow data using the TREX R package

Richard L. Peters[1,2] | Christoforos Pappas[3,4,5] | Alexander G. Hurley[6,7] |
Rafael Poyatos[8,9] | Victor Flo[7] | Roman Zweifel[2] | Willem Goossens[1] |
Kathy Steppe[1]

Introduction

Sap flow measurements provide semi-continuous (e.g., hourly) monitoring of stem water use and can provide unique insights into tree water use strategies. Sap flow data can be also upscaled at the stand level and quantify the transpiration, a major component of the terrestrial water balance. One of the most widespread techniques for measuring sap flow in tree stems is the thermal dissipation method.

2. Sap flow - Structure of the TREX R package and accessibility

Source: Peters *et al. 2020*
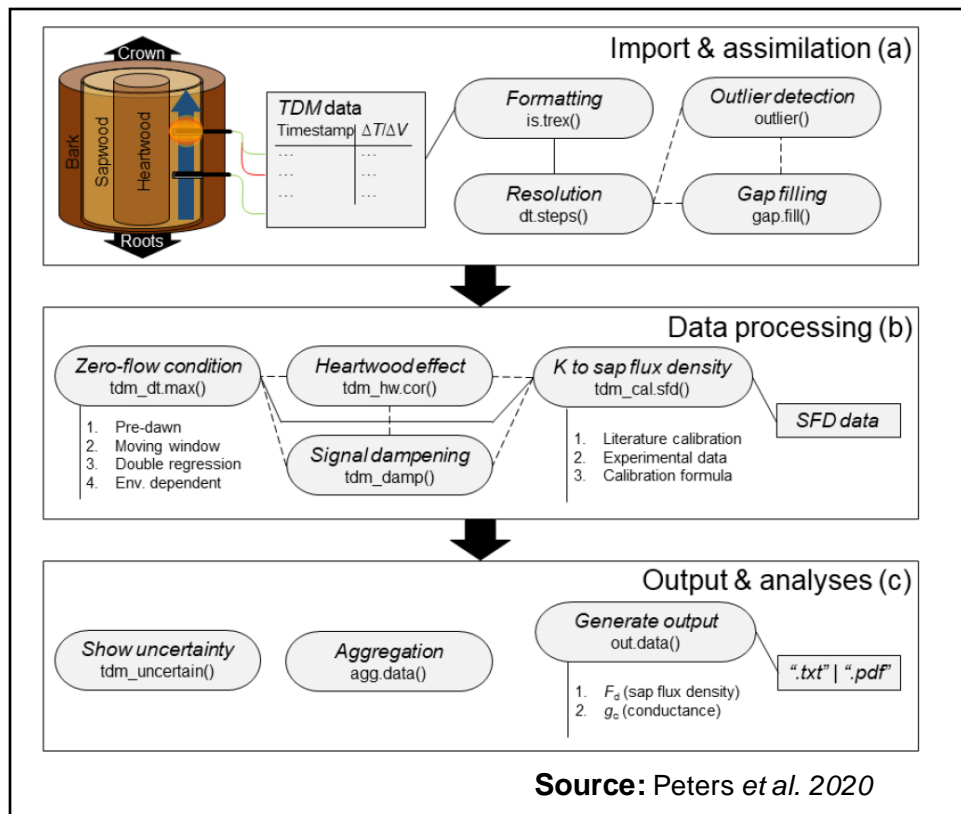
Availability

The R package TREX and all source code is available on GitHub (https://the-hull.github.io/TREX). In the R software (Team, 2019), the package can be installed with the following commands:

```
# install.packages("devtools")
library(devtools)
devtools::install_github("the-Hull/TREX")
```

Link

An R tutorial on the use of TREX is accessible via this link:
https://deep-tools.netlify.app/docs-workshops/esa-workshop2021/03_trex/

Structure

The general workflow of TREX is composed of three main steps including multiple functions [Fig.]. In step 1 (import & assimilation), the raw sap flow data and the associated auxiliary meteorological data are imported in R and the consistency of the time series object is tested and if necessary corrected (i.e., regular time steps of time series objects, outlier detection, gap filling). In step 2 (data processing), zero-flow conditions can be derived with several approaches, and corrections can be applied (i.e., to heartwood correction and dampening). Then, sap flux density can be estimated using user-specific or literature values of the calibration parameters. In step 3 (output & analyses), the uncertainties associated with the sap flow pre-processing assumptions can be quantified with state-of-the-art statistical methods, the temporal resolution of the generated data can be adjusted, and the crown conductance to water can be estimated

References

Peters *et al.* 2020 doi: 10.1111/2041-210X.13524

2. Sap flow - TREX-compatible time-series objects
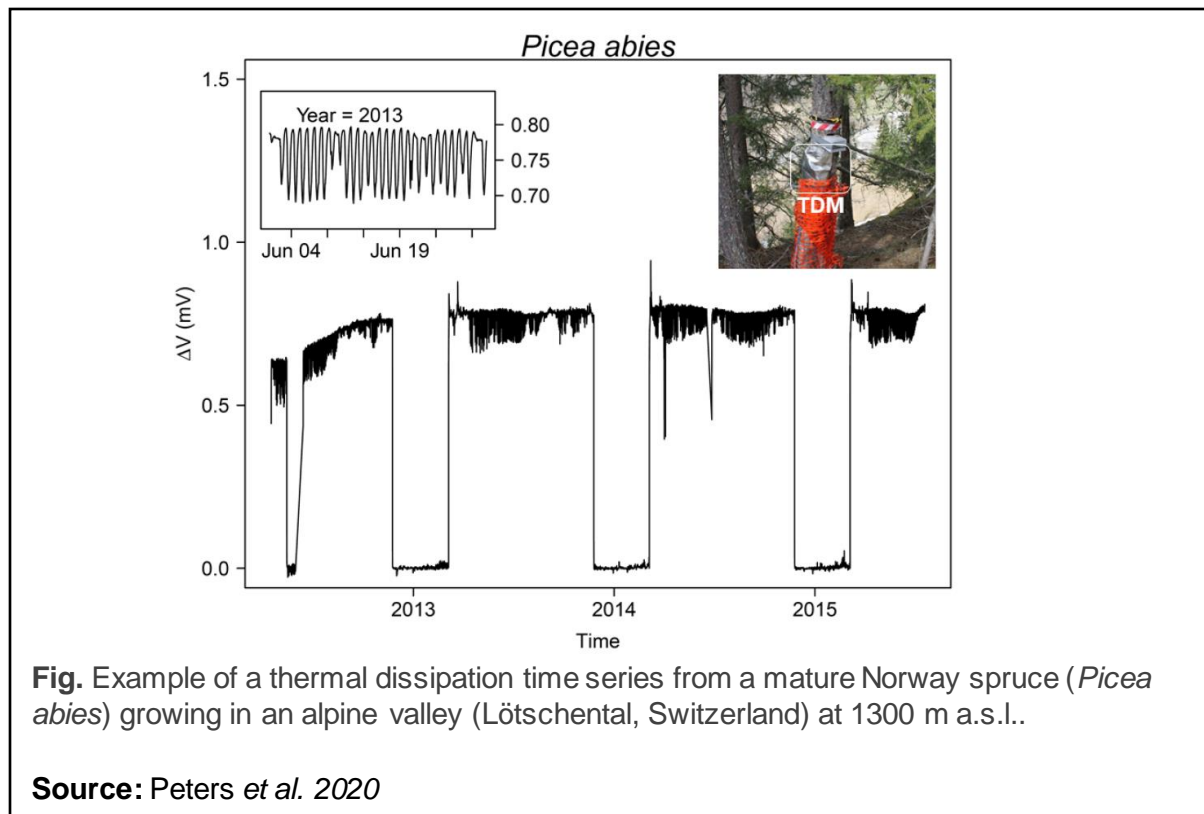
Fig. Example of a thermal dissipation time series from a mature Norway spruce (*Picea abies*) growing in an alpine valley (Lötschental, Switzerland) at 1300 m a.s.l..

Source: Peters *et al. 2020*

Step 1a: Data import & assimilation

The initial module of the TREX package deals with importing and assimilating the data. This includes check for time zone, define the temporal output resolution, data cleaning and gap filling. More specifically:

- the **is.trex** function tests if the structure of the input matches the requirements of TREX functions and specifies the time zone;
- The **dt.steps** function performs minimum time step standardization, gap filling and start/end time selection;
- The **gap.fill** function fills gaps by linear interpolation between observations.

Code

?is.trex
?dt.steps
?gap.fill

2. Sap flow - Data processing of raw sap flow measurements
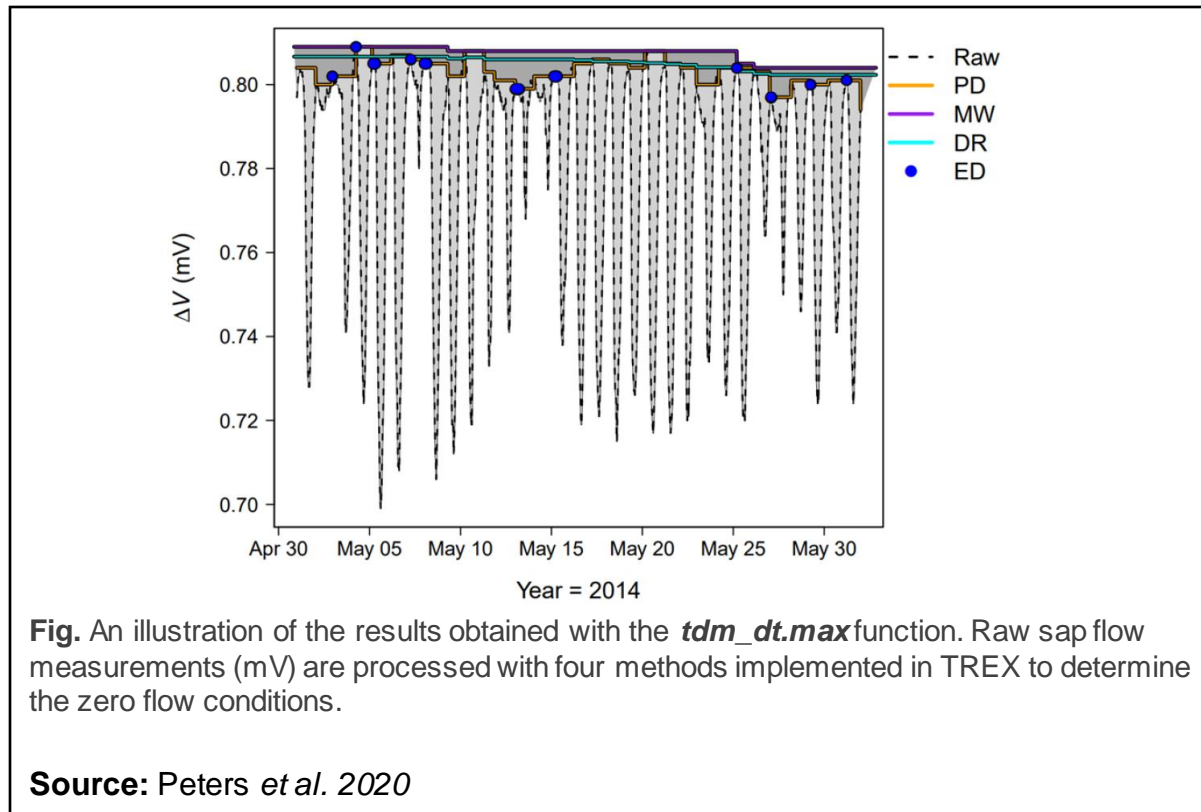
Fig. An illustration of the results obtained with the *tdm_dt.max* function. Raw sap flow measurements (mV) are processed with four methods implemented in TREX to determine the zero flow conditions.

Source: Peters *et al. 2020*

Step 2a: Zero flow conditions

The second TREX module includes the data processing, such as the determination of zero-flow conditions. In TREX four methods are implemented for determining the zero flow conditions, namely, 1) predawn (pd), 2) moving-window (mw), 3) double regression (dr), and 4) environmental-dependent (ed) as applied in Peters *et al.* (2018; doi: doi.org/10.1111/nph.15241). The *tdm_dt.max* function can provide ΔTmax (or ΔVmax) values and subsequent K values for all methods [Fig.].

Code
?tdm_dt.max

2. Sap flow - Data processing of raw sap flow measurements

Fig. An illustration of the results obtained with the ***tdm_damp*** function. A non-liner model was fit to describe the signal dampening effect and then the raw sap flow data are corrected accordingly.

Source: Peters *et al. 2020*
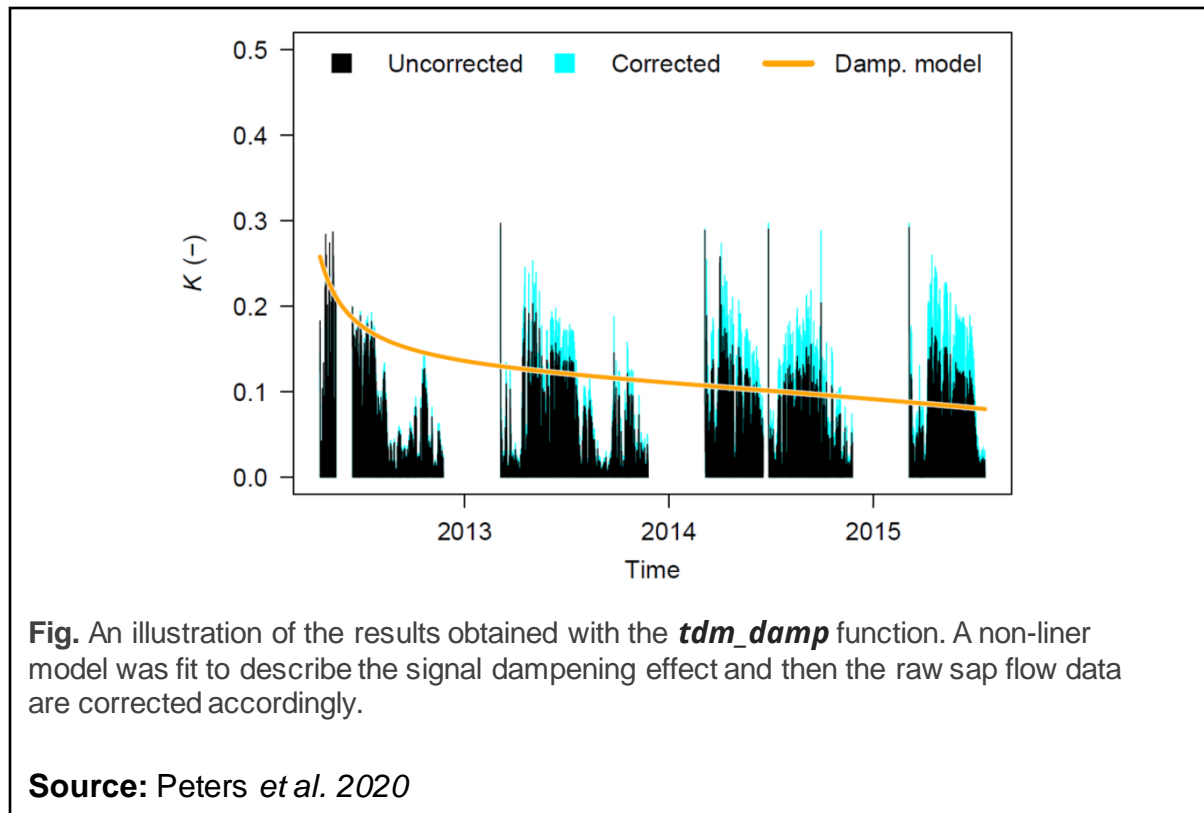
Step 2b: Heartwood correction

Depending on the tree-specific sapwood allometry as well as the needle length of the deployed sap flow sensors, it might be necessary to apply a heartwood correction in the obtained raw sap flow data. The ***tdm_hw.cor*** function corrects for the proportion of the probe that is installed within the non-conductive heartwood according to Clearwater *et al.* (1999) doi: doi.org/10.1093/treephys/19.10.681. The function requires ΔTmax, the probe length and the sapwood thickness.

Code
?tdm_hw.cor

Step 2c: Signal dampening

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 doi: doi.org/10.1111/nph.15241). Applying the signal dampening correction requires visually inspecting the correction curve (see [Fig.]). 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.

Code
?tdm_damp

2. Sap flow - Data processing of raw sap flow measurements
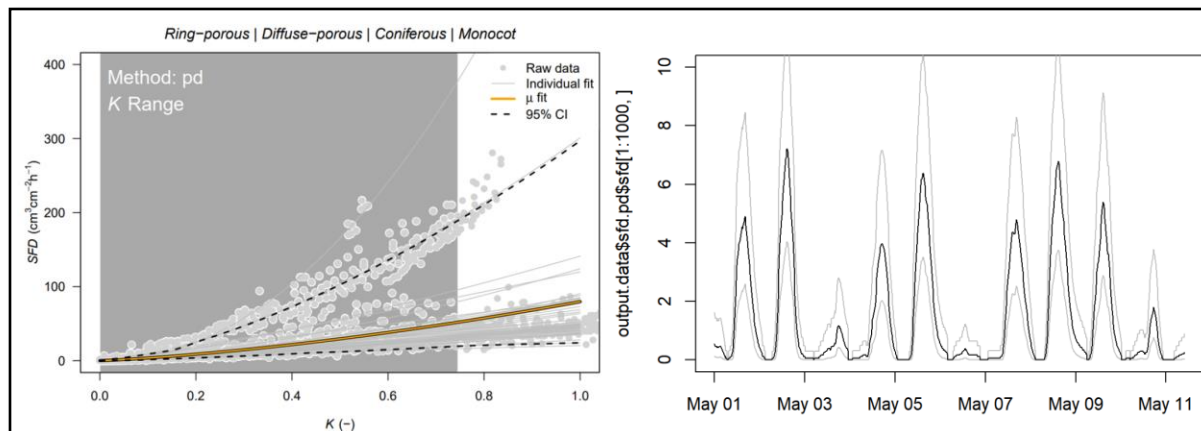
Fig. (left panel) Calibration curves obtained from literature, including all wood types, i.e., ring-porous, diffuse porous, coniferous and monocot. Power functions are fitted through each individual dataset. (right panel) Estimated sap flux density and 5 to 95% confidence intervals.

Source: Peters *et al. 2020*

Step 2d: Sap flux density

After having processed the raw sap flow data (K values) the next step is to calculate the sap flux density (SFD; cm^3 cm^{-2} h^{-1}) with the ***tdm_cal.sfd*** function. This step requires some empirical (species-specific) parameters. As many calibration curves exist (see Peters *et al.* 2018 doi: 10.1111/nph.15241; Flo *et al.* 2019 doi: 10.1016/j.agrformet.2019.03.012), the function provides the option to calculate SFD using calibration experiment data from the meta-analyses by Flo *et al.* (2019); see ***cal.data*** function. Additionally, raw calibration data can be provided or directly the parameters *a* and *b* for a specific calibration function (a K^b). The algorithm determines for each calibration experiment the calibration curve (SFD = a K^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 [Fig.].

Code
?tdm_cal.sfd
?cal.data

2. Sap flow - Output generation and data analyses

Fig. Visual output from the ***tdm_uncertain*** function when considering the double-regression ΔTmax method. Total Sobol' sensitivity indices of the investigated parameters with their mean (coloured dots) and 95% confidence intervals (vertical lines) are provided for K and SFD, respectively.

Source: tdm_uncertain() and Peters *et al. 2020*
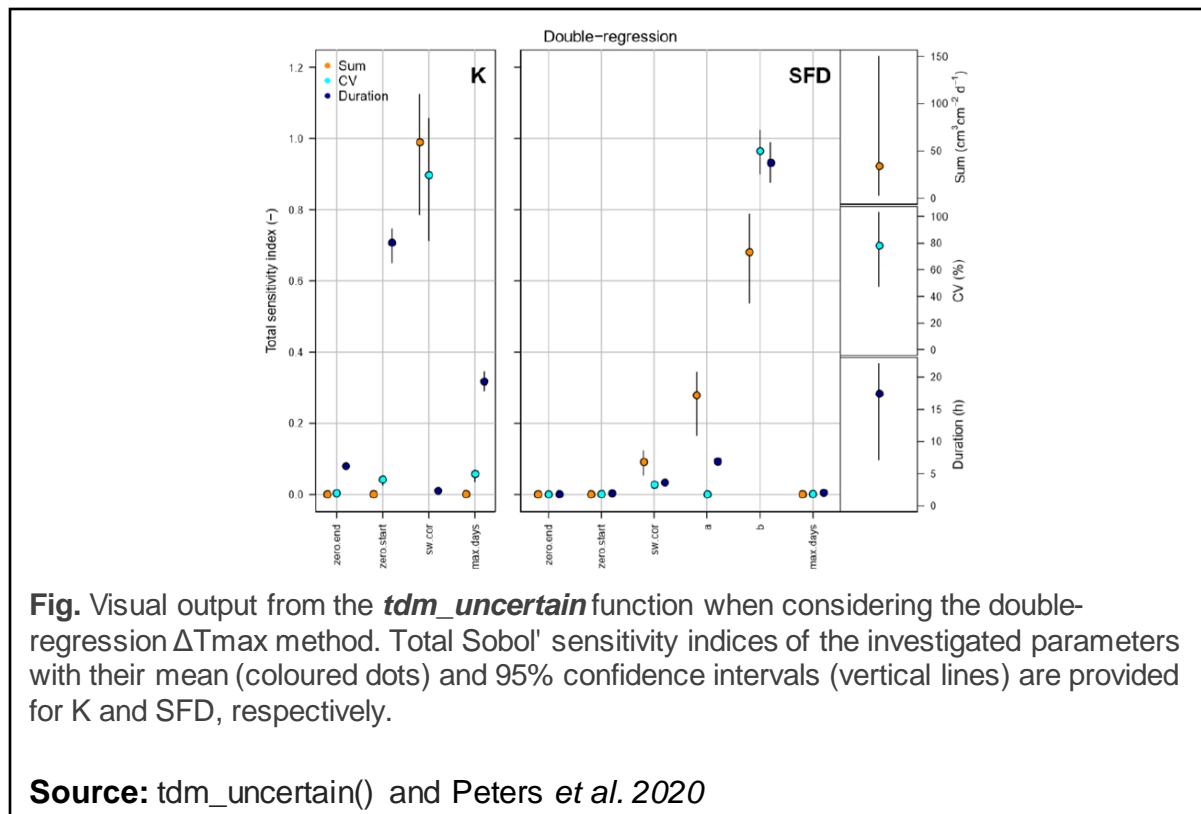
Step 3a: Uncertainty quantification

Several of the previous steps include user-specific assumptions as well as empirical parameters. To assess the robustness of the obtained results (e.g., sap flux density), within the frame of TREX package, we provide functionalities for advanced global sensitivity and uncertainty analyses (Pappas et al. 2013 doi: 10.1002/jgrg.20035). The ***tdm_uncertain*** function 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 [Fig.].

Code
?tdm_uncertain

2. Sap flow - Output generation and data analyses
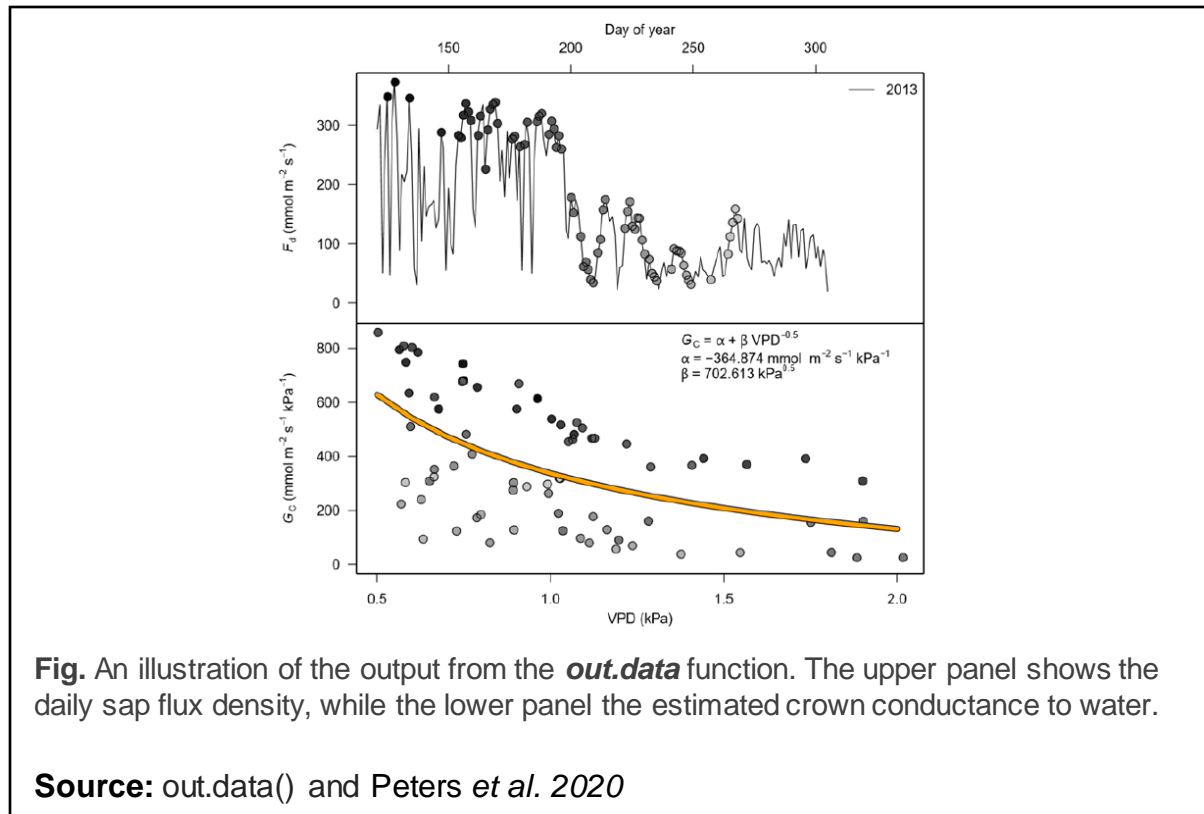
Fig. An illustration of the output from the ***out.data*** function. The upper panel shows the daily sap flux density, while the lower panel the estimated crown conductance to water.

Source: out.data() and Peters *et al. 2020*

Step 3b: Generate output

After having gone though the previous steps, the ***out.data*** function allows the user to generate relevant outputs from the sap flux density (SFD) values. This function provides both SFD expressed in mmol m^{-2} s^{-1} and crown conductance values (Gc; an analogue to stomatal conductance) in an easily exportable format. Additionally, the function can perform environmental filtering on Fd and Gc and model Gc sensitivity to vapor pressure deficit (VPD) [Fig.].

In addition, in the ***agg.data*** function allows the user to aggregate the variables of interest to specific temporal resolution and to select time periods of interest.

Code
?agg.data
?out.data

3. Dendrometer - Utilizing the treenetproc R package (https://github.com/treenet/treenetproc)

Fig. Examples of band and point dendrometers

forests

Technical Note

Processing and Extraction of Seasonal Tree Physiological Parameters from Stem Radius Time Series

Simon Knüsel [1,2], Richard L. Peters [2,3,*], Matthias Haeni [2], Micah Wilhelm [2] and Roman Zweifel [2]

Introduction

Radial stem size changes measured with automated dendrometers [Fig.] at intra-daily resolution allow us to link environmental conditions with tree physiology (i.e., radial stem growth and tree water relations).
However, measured time series need to be cleaned of outliers and data shifts, and must be separated into the reversible (water related) and the irreversible (radial growth) components in order to conduct physiologically meaningful interpretations.

3. Dendrometer - Structure of the treenetproc R package and accessibility

Fig. Workflow of treenetproc. Functions of the package are shown in rounded rectangles, including a description of the functionality (first line) and the name of the function (second line). Outputs of the functions are shown as rectangles (data output) and octagons (graphical output). Solid arrows indicate the regular workflow, dashed arrows indicate the optional, user-induced workflow.

Source: Knüsel *et al.* 2021

Availability

The R package treenetproc and all source code is available on GitHub (https://github.com/treenet/treenetproc). In the R software (Team, 2019), the package can be installed with the following commands:

# install.packages("devtools")
library(devtools)
devtools::install_github("treenet/treenetproc")

Link

An R tutorial on the use of treenetproc is accessible via this link: https://deep-tools.netlify.app/docs-workshops/esa-workshop2021/02_treenetproc/

Structure

The general workflow of treenetproc is composed of three main steps including multiple functions [Fig.]. In step 1 (time alignment), the raw data of dendrometer is aligned to user-defined, regular time steps (L1 data). In step 2 (data cleaning), outliers and shifts in the L1 data are detected and corrected. In step 3, the L2 data is analysed and several derived variables are calculated (Table 1).

References

Knüsel *et al.* (2021) doi: 10.3390/f12060765
Zweifel *et al.* (2016) doi: 10.1111/nph.13995
https://github.com/treenet/treenetproc

3. Dendrometer - Raw dendrometer series and common issues

Fig. Example data of point dendrometer measurements from LOT_N13Ad_S1. Here we plotted all three years of raw dendrometer data (L0; radial variability in µm) collected from a Picea abies tree (S1) growing in the valley bottom (N13Ad) in the Lötschental (LOT; Switzerland). Multiple data issues are present within the data, including; outliers in 2008 (purple circles), sensor failure in 2009 (orange square) and a measurement jump due to reinstalling the sensor in 2010 (blue square).

Source: View(dendro_data_L0)
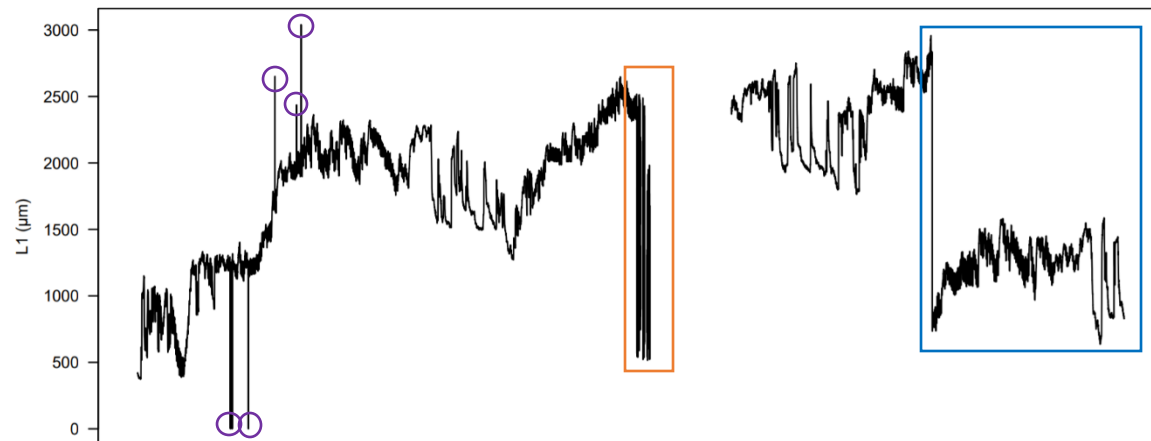
Common data issues

Within this tutorial we will present the treenetproc functionalities on LOT_N13Ad_S1 [Fig.]. The analysis of dendrometer measurements can be challenging, as raw measurements often contain outliers, errors, shifts or jumps in the data due to adjustments of the device in the field, electronic failures or external mechanical disturbances. Therefore, data cleaning is often a manually performed, time-consuming, and usually a poorly reported part of the data treatment.
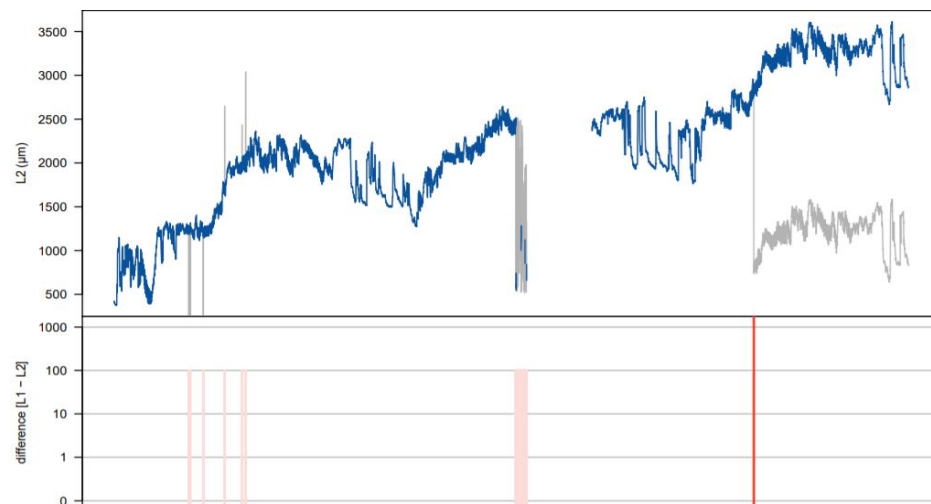
Code
See section: #0. Import data (L0)

3. Dendrometer - Removing outliers and correcting for data jumps

Fig. Example of data cleaning of the point dendrometer measurements from LOT_N13Ad_S1, using the function proc_dendro_L2. The first panel shows cleaned L2 data. The grey lines show the raw measurements, while the blue lines indicate the cleaned time-series. The second panel shows the data jump correction-induced differences between L1 and L2 data (red) on a logarithmic scale.

Source: proc_dendro_L2()

Step 1: Time alignment

Key to the following processing steps is the proper time alignment of dendrometer and temperature data to a user-defined, regular time step with the function proc_L1.

Code

See: ?treenetproc::proc_dendro_L1

Step 2: Data cleaning

Automatic outlier error detection of the dendrometer data is based on first order differences of the L1 data (diff) and the respective frequency analysis of the function proc_dendro_L2. Since the density distribution of diff is typically very narrow (most values of diff are close to zero) the most frequent values within the 30% and 70% percentiles are removed before calculating thresholds for data outliers and shifts. To increase the quality of error detection, time-aligned temperature data is provided to temp_data_L1.Cleaned data series are gap-filled according to the custom-set parameter interpol (in min.) which defines the maximum gap size that is linearly interpolated. The cleaned and gap-filled data is used to compute timeseries of tree water deficit (twd) and annually accumulated growth (gro_yr) which are returned as csv-files [Fig.].

Code

See: ?treenetproc::proc_dendro_L2

3. Dendrometer - Manual cleaning of time-series
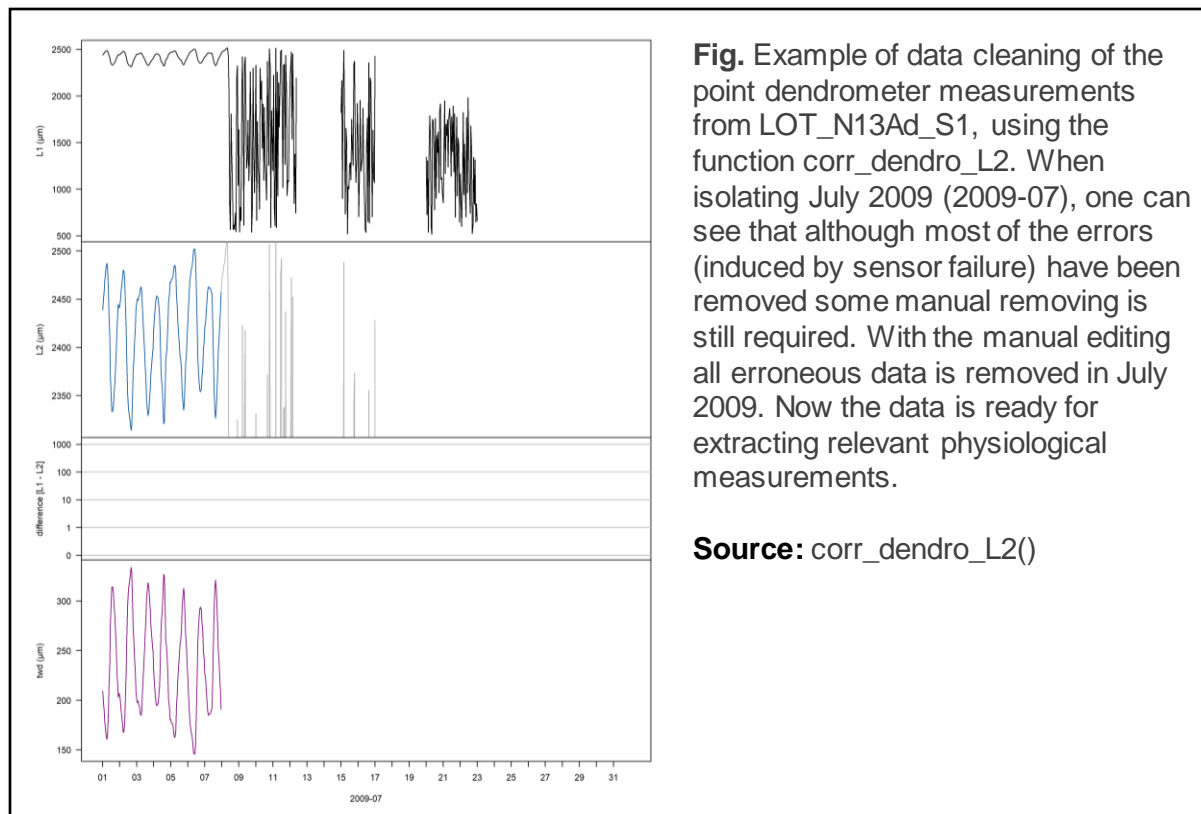when automatic detection fails

Fig. Example of data cleaning of the point dendrometer measurements from LOT_N13Ad_S1, using the function corr_dendro_L2. When isolating July 2009 (2009-07), one can see that although most of the errors (induced by sensor failure) have been removed some manual removing is still required. With the manual editing all erroneous data is removed in July 2009. Now the data is ready for extracting relevant physiological measurements.

Source: corr_dendro_L2()

Manual editing
The processed L2 data can be exported and visually inspected in user-defined plots to identify remaining errors and inconsistencies due to an inaccurate parameter setting. For example, a too flexible value of tol_jump (too high value) will not correct all jumps in the dataset. In contrast, a too rigid one (too low value) may, in the worst case, lead to an attenuation of the shape of the stem size curve over time. For remaining errors treenetproc offers functions to overrule introduced changes or force changes that were not automatically made. Generally, it is advisable to remove periods of obvious erroneous data already after time-alignment with the function corr_dendro_L1. This function can be used to reverse erroneous changes or force changes that were not automatically made. All corrections are also reflected in the returned data.frame and all changes are documented in the column flags

There are three possibilities to manually correct remaining errors:
1) Reverse: specify the ID numbers of the changes that should be reversed. Remaining changes are renumbered starting at 1;
2) Force: force a shift in the data that was not corrected for by specifying a date up to five days prior to where the shift should occur;
3) Delete: delete an entire period of erroneous data by specifying a date range. This can also be done for L1 data with the function corr_dendro_L1().

Code
See: ?treenetproc::corr_dendro_data_L2

3. Dendrometer - Workflow to obtain physiologically relevant parameters

Fig. From raw stem radius measurements on trees to valuable growth and tree water relations data. The automated point dendrometer mounted on a tree stem measures highly resolved stem size changes in a micrometre resolution. Stem radius changes origin in the cambium due to newly built cells (growth), and in the bark due to hydraulically induced pressure changes and the tree water deficit-induced tissue shrinkage and swelling (TWD). Measured raw data needs to be aligned, cleaned for data outliers and shifts, and gap filled before being separated into the growth- and the TWD-induced fractions of the size changes.

Source: Knüsel *et al.* 2021

Physiological parameters

After error detection and processing and the manual removal of remaining errors, the package offers functions to calculate additional physiological parameters that may be of use for later analyses [Fig.].

The cleaned dendrometer time series is partitioned into growth and water-related components according to the zero growth (ZG) concept. The ZG concept assumes that growth starts once the previous stem diameter maximum is exceeded and ends as soon as stem shrinkage occurs. Consequently, the ZG concept only considers growth to occur during periods without stem shrinkage.

Moreover, treenetproc calculates a number of characteristics such as the start and end of stem growth for the years measured or the timing and rate of change in phases of stem shrinkage and expansion offering a wide range of opportunities to better set stem radius data into value.

Source

Zweifel *et al.* (2016) doi: 10.1111/nph.13995

3. Dendrometer - Tree water deficit patterns provide information on drought stress

Fig. Example of tree water deficit (twd) time-series.
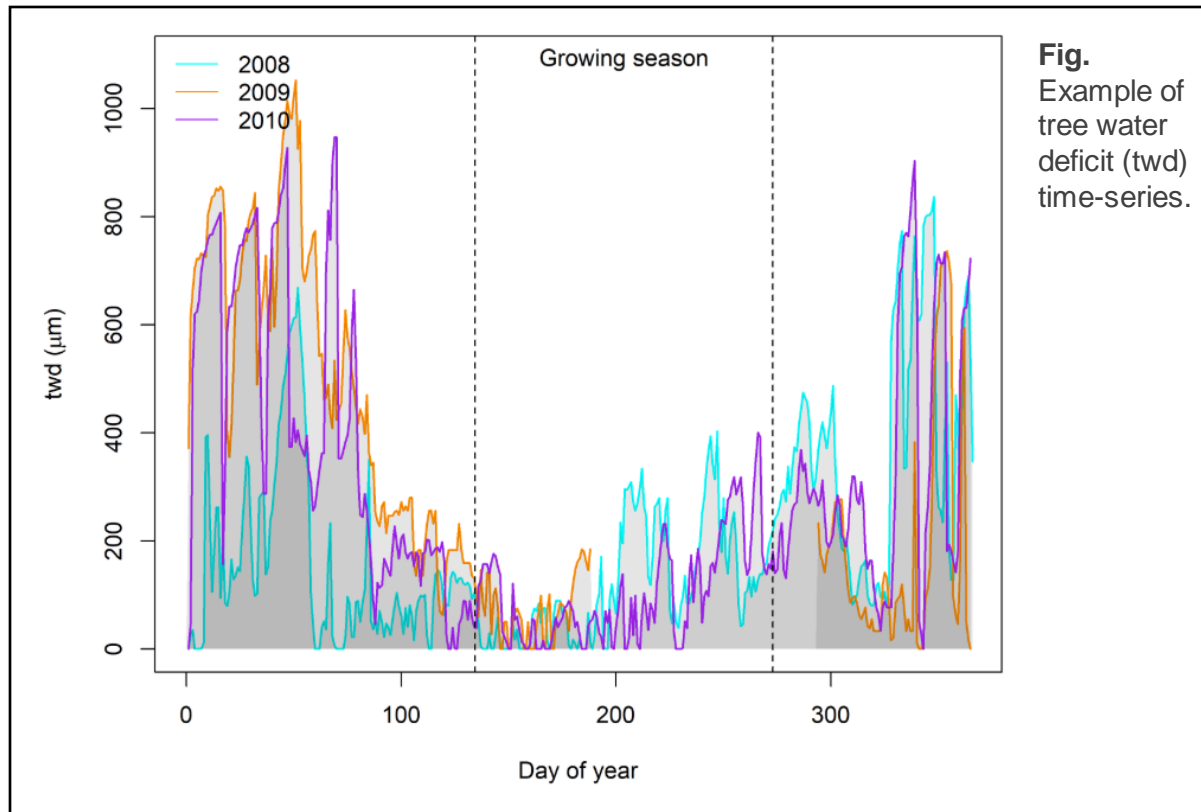
Tree water deficit and drought stress

High-resolution dendrometers capture reversible elastic reductions in stem diameter following capacitive water release, most often referred to as tree water deficit (twd). This physiological parameter can be used as an indicator of drought stress, as larger shrinkage indicates higher depletion of the water storage pools within the stem. Particularly the minimum daily twd highlights the capability of a tree to fully refill their storage pools during drought events and prevent drought induced damage to their hydraulic architecture.

Interpretation

Tree water deficit (twd) dynamics [Fig.] reveal that the tree mainly shrinks during the night in winter. Besides the temperature (frost) induced winter shrinkage, drought impacts are detectable within the growing season. During the growing season 2008 showed more shrinkage compared to 2010, revealing the tree was experience stronger water limitation during growth within this period.

Code

See section: #3. Data aggregation (L3)

3. Dendrometer - Extracting growth for determining the start and end of radial growth

Fig. Example on how the calculation of start (gro_start, red dashed line) and cessation (gro_end, red solid line) of the growing season are performed and how these are influenced by the function grow_seas with a 5% uncertainty tolerance (tol_seas = 0.05).

Source: Knüsel *et al.* (in Rev.)
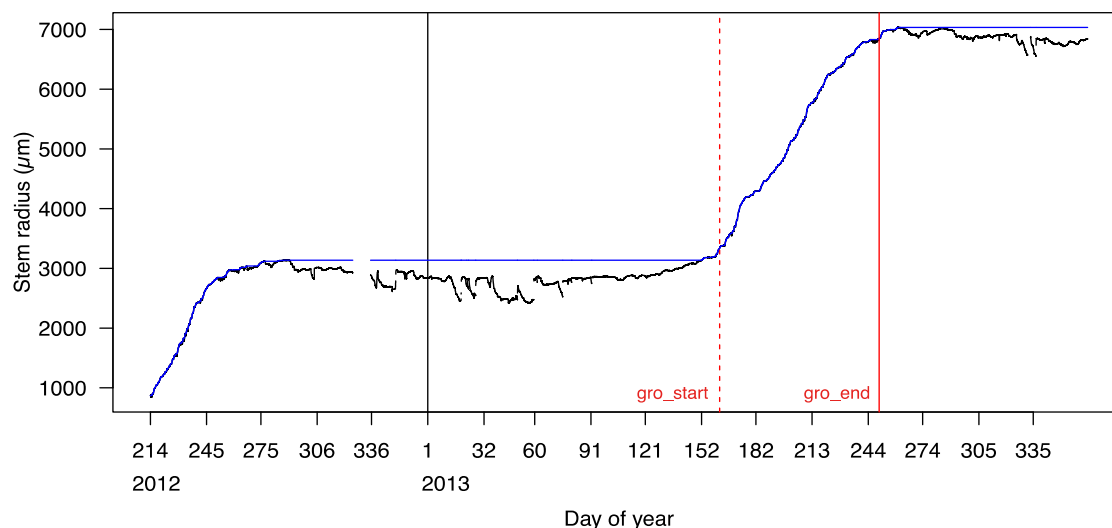
Determining the growing season

The function grow_seas returns the day of year of growth onset and growth cessation [Fig.] for clean L2 dendrometer data. Growth onset (gro_start) is defined as the day of year at which the maximum value of the past year is crossed, based on the ZG concept. Growth cessation (gro_end) is defined as the day of year at which the maximum value is reached.

Interpretation

The function returns a data.frame containing the day of year (doy) of the start and end of the growing season. Values are returned starting from the second year only, since gro_start and gro_end depend on the values from the previous year. The results generate in the R markdown illustrate that the tree growth starts around the end of May, which is realistic for these species growing in the Alps. The asymptotic nature of the growth curves makes the definition of the end of the growing season more sensitive to tol_seas. Make sure to validate these numbers for rationality.

Code

See: ?treenetproc::grow_seas

3. Dendrometer - Diurnal cyclic patterns to extract hydraulic activity

Fig. Example of phase statistics.
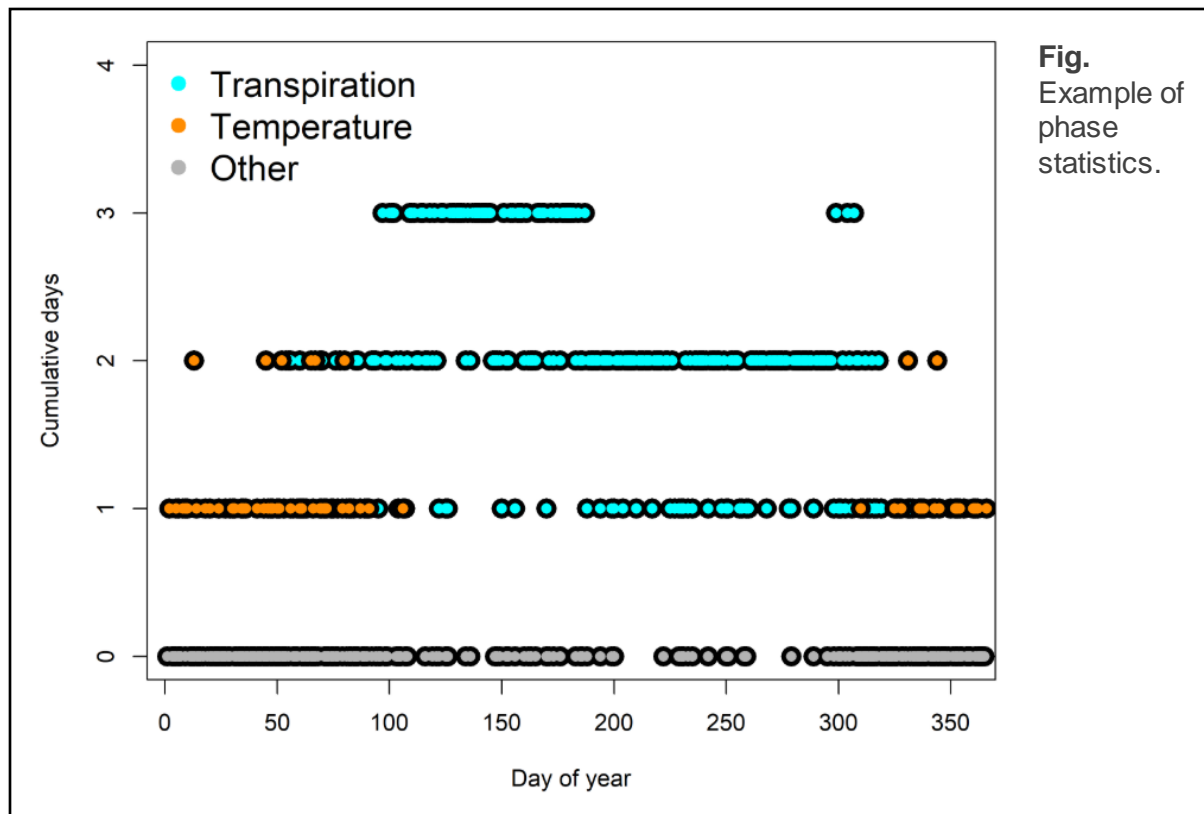
Phase statistics
Several characteristics of the diurnal shrinkage and expansion phases can be calculated with the function phase_stats. The function phase_stats calculates the timing, duration, amplitude and the rate of change of shrinkage and expansion phases. The function returns the timing, duration, amplitude and slope of the shrinkage and expansion phases. This information can be plotted daily for visual inspection.

Interpretation
The information on the diurnal phases could be used to identifies days on which radial change is likely driven by transpiration (phase_class = 1) or temperature (phase_class = -1). This plot [Fig.] shows the cumulative days of the three years (2008, 2009 and 2010) of LOT_N13Ad_S1 where the daily cycle is likely explained by transpiration, temperature or something else (other). These results show that transpiration for this individual tree is starting around day of year 50 and continues until day of year 325.

Code
See: ?treenetproc::phase_stats

Tree Physiological Data Processing in R

WK 24 - A Comprehensive Toolbox for Tree Physiological Data Processing in R

ESA Virtual Meeting
Vital connection in ecology

Organizer
Christoforos Pappas
[christoforos.pappas@teluq.ca]

Co-Organizers
Alexander G. Hurley
[hurley@gfz-potsdam.de]

Richard L. Peters
[richard.peters@ugent.be]

Roman Zweifel
[roman.zweifel@wsl.ch]