Available online at www.sciencedirect.com
ScienceDirect
SoftwareX 5 (2016) 139–143
SoftwareX
www.elsevier.com/locate/softx

Baseliner: An open-source, interactive tool for processing sap flux data from thermal dissipation probes

A. Christopher Oishi[a,*], David A. Hawthorne[a], Ram Oren[b]

[a] USDA Forest Service, Southern Research Station, Coweeta Hydrologic Laboratory, Otto, NC, USA
[b] Duke University, Nicholas School of the Environment, Durham, NC, USA

Received 5 July 2016; accepted 25 July 2016

Abstract

Estimating transpiration from woody plants using thermal dissipation sap flux sensors requires careful data processing. Currently, researchers accomplish this using spreadsheets, or by personally writing scripts for statistical software programs (e.g., R, SAS). We developed the Baseliner software to help establish a standardized protocol for processing sap flux data. Baseliner enables users to QA/QC data and process data using a combination of automated steps, visualization, and manual editing. Data processing requires establishing a zero-flow reference value, or "baseline", which varies among sensors and with time. Since no set of algorithms currently exists to reliably QA/QC and estimate the zero-flow baseline, Baseliner provides a graphical user interface to allow visual inspection and manipulation of data. Data are first automatically processed using a set of user defined parameters. The user can then view the data for additional, manual QA/QC and baseline identification using mouse and keyboard commands. The open-source software allows for user customization of data processing algorithms as improved methods are developed.

Keywords: Sap flux; Thermal dissipation probe; Granier probe; Data processing

Code metadata

Current Code version	4.0
Permanent link to code/repository used of this code version	https://github.com/ElsevierSoftwareX/SOFTX-D-16-00055
Legal Code License	*GNU General Public License v3.0*
Code Versioning system used	*git*
Software Code Language used	*MATLAB*
Compilation requirements, Operating environments & dependencies	
If available Link to developer documentation/manual	https://github.com/ElsevierSoftwareX/SOFTX-D-16-00055/blob/master/README.md
Support email for questions	*acoishi@fs.fed.us*

Software metadata

Current software version	4.0
Permanent link to executables of this version	https://github.com/ElsevierSoftwareX/SOFTX-D-16-00055
Legal Software License	*GNU General Public License v3.0*
Computing platform/Operating System	*Microsoft Windows, MATLAB*
Installation requirements & dependencies	
If available Link to user manual — if formally published include a reference to the publication in the reference list	https://github.com/ElsevierSoftwareX/SOFTX-D-16-00055/blob/master/README.md
Support email for questions	*acoishi@fs.fed.uc*

* Corresponding author.
E-mail address: acoishi@fs.fed.us (A.C. Oishi).

http://dx.doi.org/10.1016/j.softx.2016.07.003

1. Motivation and significance

Quantifying water use by woody plants is a vital component of research in plant physiology, hydrology, forest ecology, and environmental science. Currently, the most effective approach for estimating continuous, in-situ, whole-tree water use is with sap flux sensors. Various types of sap flux sensors have been developed over the past several decades, but all operate on similar principles. Sensors apply heat to a portion of the tree's water-conducting tissue and measure how the flow of xylem water (henceforth sap) affects changes in sensor temperature over time.

One of the most commonly used type of sap flux sensor is the thermal dissipation probe (TDP) [1,2]. These probes are commercially available or can be constructed by researchers with relative ease and low cost. TDPs measure the difference in temperature between a heated and unheated sensor (details in Section 2: Theory of Thermal Dissipation Probe Operation) and output a raw voltage signal that can be recorded by an automated datalogger. Accurately estimating sap flux from the raw TDP signal requires addressing several important challenges, including data quality assurance and quality control (QA/QC), and converting raw data to sap flux. Improper data processing will propagate error in subsequent analyses, resulting in incorrect estimates of tree- and stand-level water use and nocturnal sap flux [3,4]. Data from TDPs may exhibit considerable variability, both for an individual probe over time and among different probes. Currently, there is no universal algorithm available for data QA/QC and processing.

Here, we present software that enables users to QA/QC and convert raw TDP data into sap flux. This software combines both automated and manual approaches to data QA/QC and processing. The open-source design of the software enables improved data processing algorithms to be incorporated as they are developed.

2. Theory of thermal dissipation probe operation

Based on Granier's original design [1], one TDP consists of two, cylindrical metal tubes, each typically 20 mm in length and 2 mm in outer diameter, containing a T-type copper-constantan thermocouple. The tubes are installed radially into the tree's active xylem (sapwood) with vertical separation of approximately 15 cm. The upper probe includes a heating element made of constantan wire, wound around the probe, supplied with 0.200 W. Thermocouples produce a small voltage that varies with temperature and the pair of thermocouples is used to measure the temperature difference ($\mathrm{d}T$; °C) between heated and unheated probes. As the velocity of water movement increases, more heat near the upper probe is dissipated and the $\mathrm{d}T$ declines (Fig. 1a).

$\mathrm{d}T$ is inversely related to sap flux (F; m^3 m^{-2} s^{-1}) as a function of the maximum temperature differential between probes when flux is zero ($\mathrm{d}T_{\max}$; °C). Thus, F can be related to $\mathrm{d}T$ with the following equations:

$$K = (\mathrm{d}T_{\max} - \mathrm{d}T)/\mathrm{d}T = (\mathrm{d}T_{\max}/\mathrm{d}T) - 1, \quad (1)$$

$$F = \alpha \times K^{\beta}, \quad (2)$$

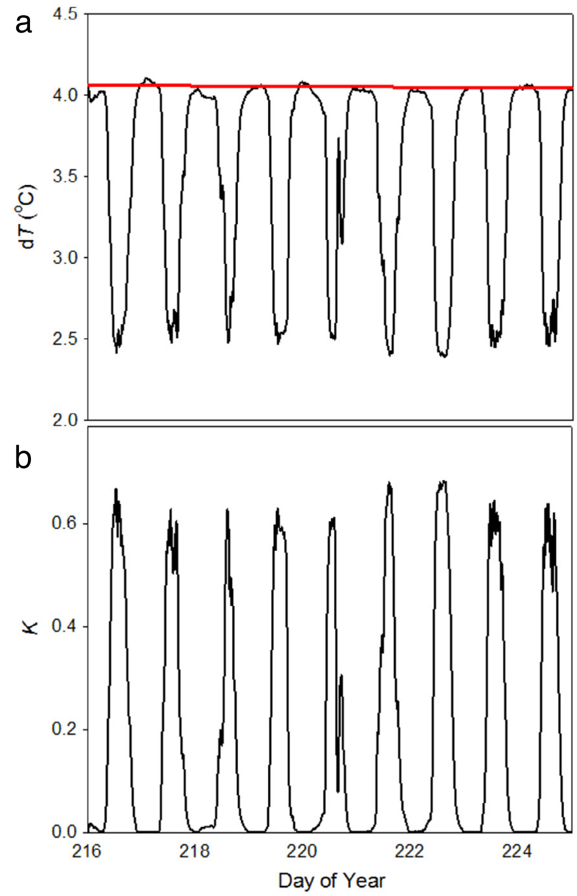

Fig. 1. Example time series of (a) raw temperature differential ($\mathrm{d}T$) from a thermal dissipation probe with zero-flow reference "baseline" ($\mathrm{d}T_{\max}$) in red, and (b) converted flow index (K). This example shows a relatively stable $\mathrm{d}T_{\max}$. Note that K during the second half of day of year 220 was lower than other days due to an afternoon rain event.

where K is a dimensionless "flow index" and α and β are empirical coefficients [1].

The $\mathrm{d}T$ signal has some predictable characteristics. $\mathrm{d}T$ is typically greater than 2 °C and less than 15 °C, with a diel wave-like pattern with a typical amplitude of less than 5 °C (Fig. 1a). However, the amplitude will vary among days with sap flux and the wave-like pattern may be interrupted within a day if sap flux declines dramatically (e.g. as the result of an afternoon rain event; Fig. 1).

TDPs are relatively simple and reliable devices; however, they can output data within the range of expected values, but exhibiting erratic patterns, inconsistent with plant physiological behavior. The $\mathrm{d}T$ signal may be compromised by faulty wiring, power supply interruptions, or other electrical interference (Fig. 2a). These data do not represent the true sap flux and should be filtered out prior to data processing. Additionally, short gaps in reliable data can occur, often due to an interruption in power (Fig. 2b).

Identifying a reliable $\mathrm{d}T_{\max}$ also presents a methodological challenge. $\mathrm{d}T_{\max}$ varies among sensors and changes over time due to tree water status, ambient temperature fluctuations, and variability in the power supplied to the heating element, among other factors [3]. Thus, over time, $\mathrm{d}T_{\max}$ may appear stable (Fig. 1a), drift upward or downward (Fig. 3a), or exhibit

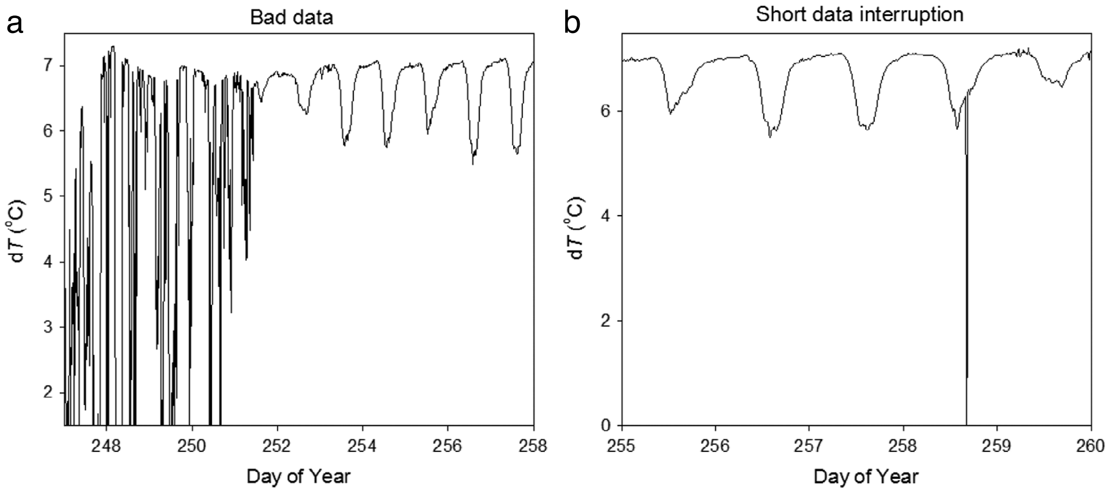

Fig. 2. Example time series of temperature differential (dT) data from a thermal dissipation probe illustrating (a) unreliable data, likely due to a faulty wiring connection and (b) a single data-point interruption in the dT signal.

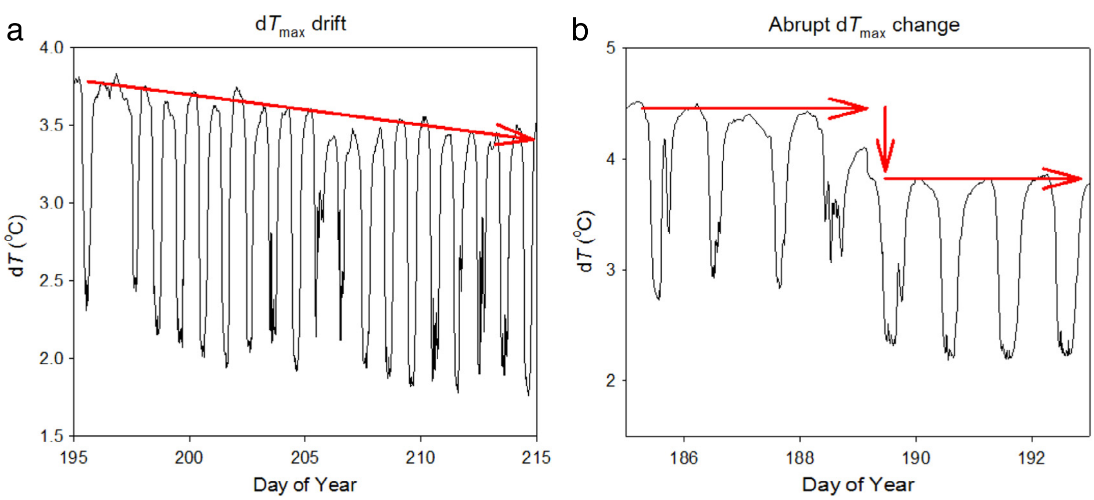

Fig. 3. Example time series of temperature differential (dT; black line) data from a thermal dissipation probe and maximum dT representing zero-flow conditions (dT_{max}) illustrating (a) a drift in dT_{max} and (b) a step change in the signal, likely due to a change in heat supplied to the sensor.

a step-change (Fig. 3b). Additionally, "noise" in sensor data, such as a spike or step-change in nighttime dT, can lead to misinterpretation of the likely dT_{max}. In some systems, dT_{max} may occur every night if air humidity within the tree canopy is very high, or vapor pressure deficit (D; kPa) reaches 0 kPa. However, this assumption is not valid for many systems and numerous studies have demonstrated water movement through the stem during night, either due to nighttime transpiration or recharge of stem water [4–8]. For these reasons, no universal rule or algorithm exists for identifying dT_{max}. The Baseliner software takes a hybrid approach to dT_{max} estimation by first identifying points in time where flow is likely zero, based on dT stability and biophysical conditions, then allowing the user to visually inspect and modify those points. After identifying individual dT_{max} "anchor points", a continuous dT_{max}, or "baseline" is linearly interpolated for the entire time series.

3. Software framework

Baseliner (version 4.0) is open source software written in MATLAB (R2015b, Mathworks Inc., Natick, MA) that is designed to run within the MATLAB environment or compiled as a standalone executable program for Windows. The software incorporates much of the user interface appearance of earlier versions of the Baseliner program, but has been rewritten to incorporate greater functionality. It is released under an open source license and allows for user modifications to the core algorithm for processing TDP data.

Baseliner provides a graphical user interface (GUI) for visualizing data from each TDP and performing two types of operation: QA/QC of raw data and conversion of raw data into flow values. Each type of operation is described in detail below.

3.1. User interface

Baseliner's main window has options for opening and saving files and exporting converted data (Fig. 4). The GUI presents a time-series plot of dT and K for visual inspection of data from each TDP. The mouse and shortcut keys can be used to zoom in and out, pan forward and backward in time. Since sap flux is often highly dependent on the atmospheric demand for water, the GUI displays D (if measured) alongside K data, which can

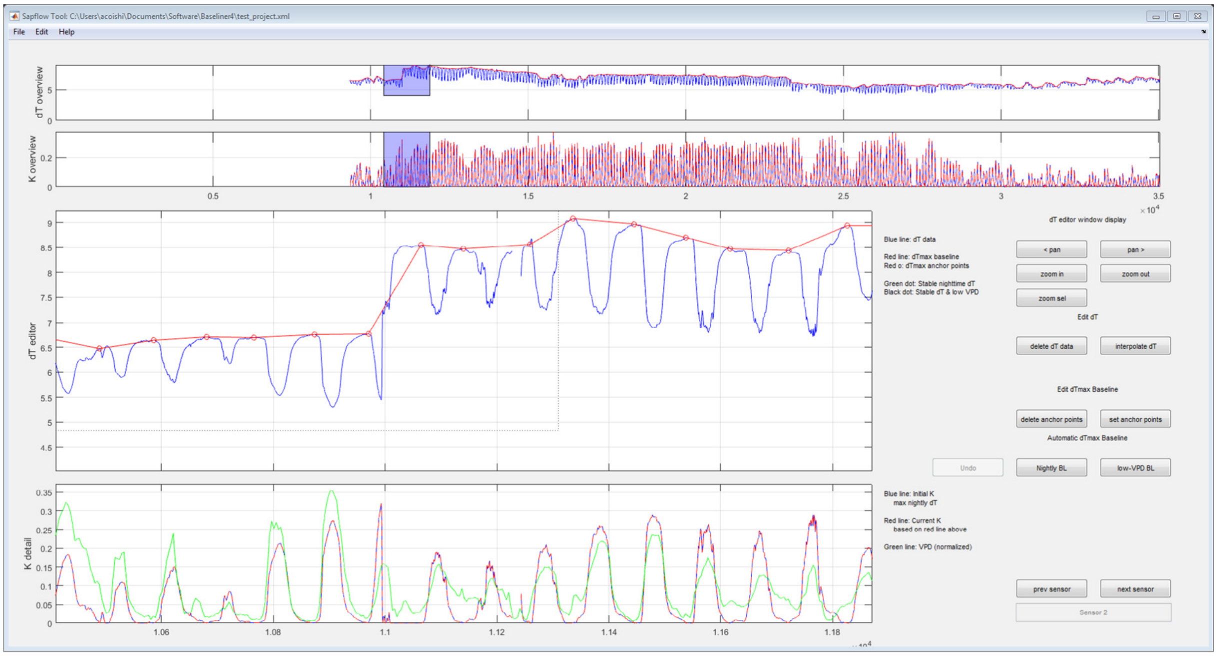

Fig. 4. Example view of Baseliner user interface window, including one year of TDP data.

provide a reference for the timing and magnitude of diurnal trends.

3.2. Input data

Baseliner imports raw data as a Comma Separated Values (.csv) file containing columns with time and date, radiation, D, and TDP output. Since measurements are not consistent among research sites, flexibility has been incorporated for input variables. Measurements can be at any increment of time (typically 15 min to 1 h records), but must be continuous, and day-of-year must increase monotonically. The type of radiation measured is not specified and can include photosynthetically active radiation (PAR), incoming shortwave radiation, or net radiation. This variable is only used to partition nighttime and daytime data, so in the event that no radiation data are available, a placeholder vector of zeros and ones can be used to represent night and day, respectively. If D data are available, they can be used for conditional dT_{max} identification; however, this variable is not required. Units for dT are arbitrary and can be expressed as raw mV values from the thermocouples or converted to °C. Baseliner does not alter the raw data and tracks any changes made during processing.

3.3. Data QA/QC

Preliminary QA/QC of the dT data is done automatically based on several user-defined parameters. Data will be set to not-a-number (NaN) values if they exceed a minimum or maximum threshold, outside the reliable operational range for the sensors. Short segments of data interrupted by NaN values are often indicative of a sensor malfunction and, if they are shorter than one day, the segments may be too short to provide reliable sap flux estimates. Thus, the user can define the minimum length of consecutive valid data to accept (e.g., filter-out data segments if valid consecutive measurements are less than 4 h). Although rapid changes in dT are possible, some changes may be unreasonably large and indicate questionable data. The user can define a maximum absolute change over one unit time to omit. Additionally, manual omission of data can be performed using the mouse and shortcut keys. The user can select a box encompassing a range of data to be deleted. The mouse can also be used to select a range of data to interpolate linearly. The interpolation function is useful for filling small gaps in the data or replacing short spikes in the data.

3.4. Data conversion

Baseliner provides users the opportunity to use both automatically-selected dT_{max} points as well as manually identified values. The default option for automatic dT_{max} point selection assumes that flow ceases every night and identifies the maximum dT value each day, between midnight and 7:00 AM. This approach accounts for the environmental conditions in which temperature is most likely to drop below the dew point as well as the physiological response in which cessation of flow past the sensor may lag behind stomatal closure at the leaf level.

The second option for automatic dT_{max} point selection finds a joint set of conditions when sap flux is likely zero. These conditions are (1) nighttime hours, characterized by near-zero radiation, (2) stable dT, characterized by a low coefficient of variation, and (3) low D, characterized by values below a set threshold for a designated length of time [4]. If these conditions are met on more than one point in time on a given night, the last such point is selected.

Since conditions may exist where an automated approach may not identify a reasonable dT_{max} point, the GUI also allows the user to use the mouse to add dT_{max} points by clicking near the dT line or remove unreasonable dT_{max} points by highlighting these points with a box. The unitless flow parameter, K, is automatically recalculated after any changes to the dT_{max} baseline or the dT data are made. The K time series is displayed in the GUI.

The structure of the open source code allows users to modify the core algorithm for dT_{max} point selection, as well as processes for filtering data.

3.5. Data output

The user has two options for exporting converted K data to a.csv file. The first option is to export K using all modifications to dT and dT_{max} anchor points. For any sensors that have not been modified, the default dT_{max} values will be used to calculate K. The second option uses the user-modified dT data, but re-evaluates the dT_{max} baseline by selecting the peak nightly value. Any modified dT_{max} anchor points are not used for these K estimates, but they are also not reset in the main project file. The second option allows the user to compare results from the interactive approach and the forced nightly dT_{max} approach.

The user also has the option to export an estimate of the error associated with the precision of point selection. The software runs Monte-Carlo simulations of the K estimates based on randomly selected dT_{max} points, varying around the selected dT_{max} anchor points within a normal distribution with a standard deviation of 1 h (independent of time step). The output from this file is the standard deviation of exported K. Columns follow the raw input file.

Baseliner also creates and saves an Extensible Markup Language (.xml) "project file" that tracks any changes made to the dT values and the dT_{max} baseline. This .xml file allows the program to reopen an existing project file and make additional modifications to data.

4. Conclusions

Baseliner is designed to help users avoid some of the most common mistakes in processing data from TDPs and to help establish a standardized approach for data QA/QC and processing. The structure of the software is intended to provide users with maximum flexibility, both in terms of the data processing interface and the ability to modify the core algorithms in the source code. We acknowledge that no approach for processing TDP data will be immune to errors and that some level of subjectivity of data analysis is necessary, either as pre-defined thresholds and algorithms or through user interaction. Therefore, the format of Baseliner strikes a balance between automation and visual inspection. This approach helps users to produce the best estimates of sap flux, accounting for variability in probe heat supply, variability in tree water status, and nocturnal sap flow.

Acknowledgments

We would like to thank Chelcy Miniat and Ken Krauss for constructive feedback on the development of this software, for providing sample test data, and providing comments on the manuscript. We thank Doug Aubrey for providing comments on the manuscript, also Pete Anderson and Shelly Hooke for testing the software.

This software development was funded by the USDA Forest Service, Southern Research Station. Previous versions of Baseliner were developed at Duke University, Nicholas School of the Environment at Duke University with programming support by Huang Ce (v.1) and Yavor Parashkevov (v.2 and v.3) and funding support from the Biological and Environmental Research (BER) Program (Grant No. DE-FG02_00ER63015), U.S. Department of Energy, through the Southeast Regional Center (SERC) of the National Institute for Global Environmental Change (NIGEC; DOE cooperative agreement DE-FC030-90ER61010), and through the Terrestrial Carbon Process (TCP) Program (DE-FG02-95ER62083 and DESC0006967).

References

[1] Granier A. Une nouvelle méthode pour la mesure du flux de sève brute dans le tronc des arbres. Ann Sci For 1985;42:81–8.
[2] Granier A. Evaluation of transpiration in a Douglas-fir stand by means of sap flow measurements. Tree Physiology 1987;3:309–20.
[3] Lu P, Urban L, Zhao P. Granier's thermal dissipation probe (TDP) methodology for measuring sap flow in trees: theory and practice. Acta Bot Sin 2004;46:631–46.
[4] Oishi AC, Oren R, Stoy PC. Estimating components of forest evapotranspiration: A footprint approach for scaling sap flux measurements. Agricult Forest Meteorol 2008;148:1719–32.
[5] Daley MG, Phillips NG. Interspecific variation in nighttime transpiration and stomatal conductance in a mixed New England deciduous forest. Tree Physiol 2006;26:411–9.
[6] Dawson TE, Burgess SSO, Tu KP, Oliveira RS, Santiago LS, Fisher JB, Simonin KA, Ambrose AR. Nighttime transpiration in woody plants from contrasting ecosystems. Tree Physiol 2007;27:561–75.
[7] Oren R, Phillips N, Ewers BE, Pataki DE, Megonigal JP. Sap-flux-scaled transpiration responses to light, vapor pressure deficit, and leaf area reduction in a flooded *Taxodium distichum* forest. Tree Physiol 1999;19:337–47.
[8] Phillips N, Oren R, Zimmermann R. Radial patterns of xylem sap flow in non-, diffuse- and ring-porous tree species. Plant, Cell Environ 1996;19:983–90.