

## ***Interactive comment on “Global transpiration data from sap flow measurements: the SAPFLUXNET database” by Rafael Poyatos et al.***

**Stan Schymanski (Referee)**

stanislaus.schymanski@list.lu

Received and published: 16 November 2020

### GENERAL COMMENTS

The manuscript entitled "Global transpiration data from sap flow measurements: the SAPFLUXNET database" is a great addition to ESSD, and I would like to thank the authors for making such a large and helpful dataset public. It is also really good to see that a lot of the software used for data processing, quality control and subsequent analysis is also publicly available, but this is not described very well in the Data availability section (Section 6). In fact, I was not able to entirely understand the data harmonisation and quality control described in Section 2.4 (see some of my detailed comments below), and without the accompanying scripts, it may be difficult to reproduce the data

C1

in the database from the original data, should questions emerge.

The database itself is very large and it is a pity that the whole database has to be downloaded and de-compressed before inspecting it, but this is a limitation of the zenodo repository. To make access to the data more convenient, the authors could also place a copy on github or gitlab, where users would be able to download individual files or folders. I was also confused about the organisation of the data into 'leaf', 'plant', and 'sapwood'. If a dataset contained information about leaf area and sapwood area, would it then be repeated in all three folders? This would not seem very efficient, as it would unnecessarily inflate the database in my eyes.

I went through the Quick Guide (<http://sapfluxnet.creaf.cat/sapfluxnet/articles/sapfluxnet-quick-guide.html>) and found that working with the database in R is relatively easy even for readers with very limited experience with R. What I found a little confusing is that the names of the columns in 'arg\_maz\_sapf' are not explained anywhere, and the units are also not readily accessible, although they appear correctly in the plots. It would be helpful if the Quick Guide explained how to find out what the column names mean and what the units are. It would also be helpful to include code that would allow easy reproduction of the figures in the paper and, ideally, also figures in the original papers linked to each dataset. This would help the reader to better understand the intended use of the data. Perhaps the latter could be followed for the next version of the database.

The manuscript contains a very inspiring section on potential uses of the database. However, in this section and other places in the manuscript, it seems to equate sapflow dynamics with whole plant transpiration dynamics, which can be quite misleading, given capacity-caused delays between those two variables, which are mentioned in passing but not explicitly discussed in the context of shifts in timing between transpiration and sapflow. See detailed comments below.

I share the editor's concern that data uncertainty is not adequately addressed in the

C2

paper and in the data base. The manuscript discusses various aspects of uncertainty, cites evidence for consistent underestimation of sapflow rates by some methods and suggests that a first-order correction is possible, but it does not provide any numbers or details about the correction. Similarly, it mentions that estimation of sapwood area can have substantial errors without providing numbers and even an attempt to estimate the propagation of such errors to whole plant sap flow rates. I find that this is a missed opportunity, given the exquisite selection of co-authors on this manuscript. It would be great if the authors could give some guidance to the readers on the likely error bounds of the data in the database.

Below, I will answer to the ESSD review criteria one-by-one.

Are the data and methods presented new? I believe that the database and underlying methods are cutting edge, as described in the manuscript.

Is there any potential of the data being useful in the future? Definitely, as described in the manuscript, with the caveat that it should be clarified that the diurnal dynamics of sapflow should not be expected to be the same as that of transpiration, and subject to some quantification of potential error / uncertainty.

Are methods and materials described in sufficient detail? Mostly yes, but I was not able to easily understand some of them, as pointed out above.

Are any references/citations to other data sets or articles missing or inappropriate? Not that I am aware of.

Is the article itself appropriate to support the publication of a data set? Yes, it is very detailed and honest.

Is the data set accessible via the given identifier? Yes.

Is the data set complete? I did not really check, but a colleague of mine has been using the data more intensively and has not found any issues.

### C3

Are error estimates and sources of errors given (and discussed in the article)? Unfortunately, no quantitative error estimates were given, although sources of errors are discussed in the article. Please see above.

Are the accuracy, calibration, processing, etc. state of the art? I could not find any issues, but I am also much less qualified to assess this than the many co-authors on this manuscript.

Are common standards used for comparison? Unfortunately, I am not aware of any common standards for comparison against sapflow data. Lysimeter studies might be helpful, but I did not see any such studies cited in the manuscript.

Is the data set significant – unique, useful, and complete? I am convinced it is, except for error estimates.

Are there any inconsistencies, implausible assertions or noticeable problems that would suggest the data are erroneous (or worse)? Most datasets included in the database have already been used and published elsewhere but I did not assess whether the results of the original publications can be reproduced using the database, nor did I see such an analysis as part of the data quality checks. Given that most of the analysis is done with R, it would have been great if the authors included code to reproduce the figures and tables presented in this manuscript, and, perhaps for future data submissions, to reproduce figures in the original publications related to each dataset.

Is the data set itself of high quality? The quality checks on the data were quite systematic and the metadata included is also very helpful, instigating trust in the quality of the data.

Is the data set usable in its current format and size? Yes.

Are the formal metadata appropriate? Yes.

Is the length of the article appropriate? Yes.

### C4

Is the overall structure of the article well structured and clear? Yes.

Is the language consistent and precise? Yes, with a few exceptions, see detailed comments below.

Are mathematical formulae, symbols, abbreviations, and units correctly defined and used? I did not find any problems.

Are figures and tables correct and of high quality? Since the code used to generate the figures and tables in the manuscript was not provided, I was not able to verify within the time I had for this review.

Is the data set publication, as submitted, of high quality? Yes.

Finally: By reading the article and downloading the data set, would you be able to understand and (re-)use the data set in the future? Very likely.

#### DETAILED COMMENTS

L290: "understanding of"

L295: This sentence does not make much sense to me: "In practice, transpiration is relatively easy to isolate from the bulk evaporative flux, evapotranspiration, only from the leaf to the plant levels." First of all, it is not easy to separate transpiration from ET at the leaf level, as the only other component of ET at leaf level is evaporation of intercepted water, which is not easily separated from transpiration. Secondly, transpiration measurement at the plant level is exactly what sapflow is about, so why are you saying that it is relatively easy?

L314: When stating that sap flow sensors can be deployed in almost any terrestrial ecosystem, it would be important to point out that they are limited to woody plants (unless I am wrong).

L338: Where are the quality assurance and control procedures described?

C5

L364: What do you mean by "methodological uncertainty"? Uncertainty about what method was used for a given dataset due to inadequate documentation, or uncertainty about what is the most suitable method for a given case? Could you clarify in the text?

L372: This sounds as if the stem heat balance method was superior to the others. Could you clarify?

L432-434: What does the Hampel filter do? How does it identify outliers and estimate correct values? Where is the code used to determine out of range values and outliers? Where is this R Shiny application? I could not find it at <https://github.com/sapfluxnet>. What is the role of expert knowledge here? Why were out of range values replaced by NA and outliers by the Hampel filter value?

L440: I think the first "Granda et al." should be removed.

L445: Was sap flow per unit leaf area computed even for deciduous species using constant leaf area all year round? Does it make sense to do this?

L450-465: These details could be moved to the appendix, as I did not find it very informative without seeing the database.

L490: Grammar: A sentence should not start with "Because".

L503: By how much did HD and CHD underestimate sap flow rates?

L505: Grammar: reads as if plants were doing the data processing.

L510: What does it mean for the upscaling? Could it not be done at all for datasets with missing sapwood area or depth?

L529: Is there any reliable way to estimate sapwood area for trees where this information was not provided? I did not understand what role tree height plays for the use and/or interpretation of the data.

L583: What does the "compared to..." mean? PAR and Rnet were provided less often

C6

than shortwave? Was shortwave converted to PAR or vice versa to obtain a homogeneous dataset?

L594: It would be good to mention that sub-daily sap flow variation is not necessarily consistent with sub-daily variation in transpiration.

L599: What do you mean by "scale up to"? Maybe "translate to"?

L504: How could the data be used to estimate hydraulic conductivity?

L605-610: This may be misleading, as diurnal patterns in sap flow do not necessarily translate to diurnal patterns in transpiration, due to capacitance. So the relevant time lags are not only between evaporative demand and sap flow, as stated here, but also between sap flow and transpiration. This needs to be clarified to avoid confusion.

L616: Why would SAPFLUXNET allow quantification of nocturnal sap flow for data sets less suitable for quantification of night-time fluxes? More explanation would be helpful here.

L625: How are treatments documented in the database? Qualitatively or quantitatively?

L658: Has any of this been attempted already, and if so, how big are the uncertainties?

L676: To know if this up-scaling is promising, it would be important to know the uncertainty in sap flow derived individual-based or sapwood area based transpiration.

L704: How could uncertainty propagation ever be done for the commercial methods? This might be a good place to emphasise the need for open source data-processing software.

L710-711: How accurate, how big is the under-estimation?

L711: Grammar: don't start a sentence with "Because"

L715: So what error bands should we expect for the data before and after corrections?

C7

L728: How large can the errors due to sapwood area estimation errors be?

L773: Could you add the relevant packages used for data processing and quality control, some of which are mentioned in <https://github.com/sapfluxnet>, and specify which version of each package was used?

---

Interactive comment on Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2020-227>, 2020.

C8