Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2018-56-AC1, 2018

Biogeosciences Discussions
Open Access
EGU

Interactive comment on "Basic and extensible post-processing of eddy covariance flux data with REddyProc" *by* Thomas Wutzler et al.

Thomas Wutzler et al.

twutz@bgc-jena.mpg.de

Received and published: 4 April 2018

This is an extended version of the immediate short comment by Thomas Wutzler, agreed by all coauthors. Here we repeat each reviewer's comment (RC) before each of our author's comment (AC) replies in blue.

We thank referee 1 for his constructive comments.

RC: The paper adds to a growing segment of extensive method descriptions for reproducible computational research. While the technical focus is laudable, it is also the reason that the manuscript in its current form somewhat misses the scope of BG ("interactions between the biological, chemical, and physical processes"; https://www.biogeosciences.net/index.html). I imagine this might

Printer-friendly version

Discussion paper

be part of the reason that the other referees declined. Alternative Copernicus journals like AMT (“techniques of data processing”; https://www.atmospheric-measurement-techniques.net/index.html) or GMD (“statistical models”, “technical papers”; https://www.geoscientific-model-development.net/) should provide a much better fit.

AC: While the suggested alternative journals maybe fit the content slightly better, they target an audience primarily interested in data processing or model development. Contrary, we want to target an audience of researchers who *use* EC flux data in their studies. We argue that Biogeosciences is the best open access journal to reach this audience. For instance Biogeosciences Journal recently hosted a special issue on eddy covariance data collected in the Australia and New Zealand, which include also more methodological papers "OzFlux: a network for the study of ecosystem carbon and water dynamics across Australia and New Zealand, 2016". Also a search of the keyword "eddy covariance" in the main text on April 4th, 2018 give back 1620 records. We also looked at the number of views in the open review phase and they were more than 350 (April 4th), supporting the idea that Biogeosciences can be the right audience argument for the article.

RC: Here a few points for consideration in such re-submission: - I suggest shortening the Discussion paper manuscript. It should be straightforward to consolidate 22 pages of heavy methodological detail by about 1/3, and focus on novel aspects.

AC: We agree with the reviewer that the manuscript can be shortened and in the revised version we will work in this direction. In particular we will move some paragraphs (sec 2.2.3, Fig3, parts of sec 2.3.3)) in the appendix and part of the appendices to online supplementary materials (Appendix A1 and C). We strive to report the details in some way for reproducibility.

The length of the manuscript results also from the fact, that essentially three different issues are explored, and all of them are compared to existing methods: 1) filtering,

Printer-friendly version

Discussion paper

i.e. u_* threshold estimation 2) gap-filling, and 3) flux-partitioning. When splitting the manuscript into three papers, big parts, e.g. in the introduction, will be redundant.

RC: - l 5: "standard tools available in open source environment for processing high frequency (10 or 20 Hz) data into half-hourly quality checked fluxes". At this time open-source environments for eddy-covariance data processing that actually facilitate open development are only emergent. REddyProc provides a substantial and much appreciated contribution to this movement. I suggest to either substantiate the claim of an abundance of open-source high-frequency data processing environments through providing examples, or to provide a more differentiated overview.

AC: We agree with the reviewer that open source tools for eddy covariance processing are only now emerging. There are now available some open source tools for processing of high frequency data, for instance the Eddy Pro software, which is open source and widely used. Moreover, there are also an increasing amount of packages (see this URL[1]): We will modify the sentence to clarify ths aspect.

RC: - l. 8: While it is true that R is a cross-platform language, this does not mean that research is reproducible by using an R-package across platforms. Known as "dependency hell", installing e.g. REddyProc on a standard Debian Linux distribution requires the co-installation of several operating-system-side libraries (libudunits2-0, libudunits2-dev, udunits-bin, libnetcdf-dev) and even more R-side dependencies (backports, praise, evaluate, highr, mlegp, logitnorm, ncdf4, RNetCDF, minpack.lm, segmented, rprojroot, testthat, knitr). In some operating systems such as Windows, there is hardly any automation available for resolving operating-system-side dependencies, making R-packages with heavy dependencies inaccessible to less experienced users. Most importantly, dependency resolution itself is not reproducible among operating systems, thus rendering reproducible research impossible. A balanced discussion of how REddyProc can be used for reproducible research alongside examples for dependency

[1]http://fluxnet.fluxdata.org/2017/10/10/toolbox-a-rolling-list-of-softwarepackages-for-flux-related-data-processing/

resolution would add much substance and usability to the manuscript.

AC: The library dependency issue are caused by the NetCDF packages. We listed this packages as "suggests" instead of "depends". Hence, REdddyProc can be used and installed without these dependencies really with a single line as exemplified on page 22 line 14 (We tested it also on a standard DEBIAN distribution using docker). Only when trying to read NetCDF files, REddyProc issues an error advising to install these packages before.
For users who want to read NetCDF files but have not yet installed the required sysem libraries, together with the revised manuscript, we will provide a Docker image that already includes all required system libraries and R-packages.
However, in the revised manuscript we wil keep these technical details short for the sake manuscript length and targeted audience.

RC: - l 15: It could be pointed out that REddyProc has already been adopted for computational research by the flux community, such as in Metzger et al. (2017). These authors also point to a community solution for “dependency hell”, an pre-compile REddyProc alongside its dependencies into compute images that contain a turn-key, reproducible and shareable processing environment.

AC: Thanks for this suggestion. We will cite the paper when referring to the provided Docker image, but will keep technical details short because of the targeted audience.

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-56, 2018.

Printer-friendly version

Discussion paper