Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2018-56-RC1, 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.

Anonymous Referee #1

Received and published: 7 March 2018

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 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.

Here a few points for consideration in such re-submission: - I suggest shortening the

Printer-friendly version

Discussion paper

manuscript. It should be straightforward to consolidate 22 pages of heavy methodological detail by ∼1/3, and focus on novel aspects.

- 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.

- 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 resolution would add much substance and usability to the manuscript.

- 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.

References Metzger, S., Durden, D., Sturtevant, C., Luo, H., Pingintha-Durden, N., Sachs, T., Serafimovich, A., Hartmann, J., Li, J., Xu, K., and Desai, A. R.: eddy4R 0.2.0: a DevOps model for community-extensible processing and analysis of eddy-covariance data based on R, Git, Docker, and HDF5, Geosci. Model Dev., 10, 3189-3206, doi:10.5194/gmd-10-3189-2017, 2017.

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

Printer-friendly version

Discussion paper