

## ***Interactive comment on “eddy4R: A community-extensible processing, analysis and modeling framework for eddy-covariance data based on R, Git, Docker and HDF5” by Stefan Metzger et al.***

### **Anonymous Referee #3**

Received and published: 26 March 2017

This study presents a radically new way to process eddy-covariance data. It combines R-coded EC software that are wrapped in a portable Docker image that can be used on various platforms. It is meant to be scalable and to make use of parallel processing of large quantities of data.

Major comments:

I line with the other reviewers I think that the paper currently lacks a clear scientific question. I could imagine that for GMD a clear description of a software environment would suffice, but this paper seems to describe “work in progress”. I am a big fan of

C1

Docker and directly downloaded the Docker image. I was disappointed in the fact that the image did not contain clear examples (e.g. the three examples outlined in the paper). I could see that the eddy4R.base and eddy4R.qaqc packages were part of the Docker image. I think it is a missed opportunity not to provide examples of (simple) data processing and plotting. Now the advantage of Docker images remains untraceable to the readers and remains rather theoretical. For instance, the HDF5 section (2.4) is clear but a rather standard description that is available on internet (meta-data, directory structure, self-documenting). Again this is a missed opportunity to guide users through an example (download raw data, process the data, and HDF5 output and visualization of results). You want to convince the “traditional ASCII” community. Section 2.5 presents the way NEON wants to deploy Docker images. Again, this remains rather high level, while the stated goal is to “empower the Science community at large by putting the key to the scientific algorithms into the hand of scientists”. Again a clear running example in a Docker container would convince these scientists more than a NEON brochure. Section 2.6 would be an ideal starting point for further “Docker-assisted” data analysis, but unfortunately stops at a reference to the eddy4R wiki pages. In section 5 there is a reference to the raw data, but again unfortunately no examples are given in which a Docker image automatically reads, processes, and presents results. In the remainder of the paper, three examples are given, which is basically fine, but without a traceable and “hands-on” exercise does not add much. It is (and should be) part of the standard software testing. In summary, I very much like the concept presented in this paper. However, without more in depth possibilities for potential users of the software, the paper seems more suitable for internal documentation than convincing readers that this is a promising way for the community to process eddy covariance data.

Minor comments

Page 1: line 34: mention where the NEON site is and also where the aircraft data were collected.

C2

Page 1, line 38: “streaming generation of science-grade EC fluxes”: please explain better what this means.

Page 6, line 185: current recent

Page 6, Figure 3, introduced at line 191. This hardly adds anything. A link would do here. Also figure 4 and figure 7 seem illustrations that do not add much.

Page 7, line 231: CI?

---

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-318, 2017.