Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-318-RC2, 2017

Geoscientific Model Development Discussions

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 #2

Received and published: 26 March 2017

Metzger et al. describe a data processing framework which is illustrated on ad-hoc examples from NEON's eddy covariance tower and airborne measurement datasets. Overall this technical concept seems potentially valuable for streamlining automation of specific data processing steps from different measurement stations but it is extremely difficult to recognize the broader scientific values in the current version of the paper as written. I must admit that I was rather disappointed to find the description of the tools to be fragmented and poorly supported by the scientific results and conclusions. The whole analysis is very descriptive and in many cases misleading as to what is possible. There is little effort to synthesize what

can be actually learnt from using the tool other than what its potential applications might be in the future. Most importantly, the paper does not specify scientific goals and does not even address the scope of modeling which is what the main interest of the journal's audience is. The manuscript seems to need much work to make the results and discussion useful for the scientific community but could be worthwhile to reconsider after major clarifications. The other reviewer has already provided a useful detailed guidance how this could be achieved and I agree with her/him. I also have other concerns which hopefully can be addressed in the revision.

General issues:

1) One fundamental issue in this paper intended for GMDD is that the work is not even connected to any model or modeling framework. The journal scope does not overlap with what paper is about or at least the connection is not made clear. Because there is no model, there is no model version – a requirement of the journal. There are only two words "model" in the whole paper, one of which is included in the last sentence of conclusions but probably in a different meaning: "We hope this framework can serve as a *model* for implementing community-sourced, distributed-development scientific code while combatting the deficiencies of current computational frameworks that limit accessibility, reproducibility, and extensibility."

2) It is not apparent how exactly this technical set of workflows adopted by NEON can be useful for a broader scientist/modeler community and what scientific problems it can solve as the idea wraps around different open-source products dedicated essentially to crunching of eddy covariance measurement data. In the abstract, it is promised that the framework is applicable beyond EC but it is completely unclear how. Maybe one way to overcome this issue would be to make a strong connection to a modeling framework where measurement and model outputs are evaluated together or elucidate aspects where this data processing framework would add to novelty and usefulness for the broader GMD community.

3) There are no clear scientific objectives of the paper and the title does not help either “eddy4R: A community-extensible processing, analysis and modeling framework for eddy-covariance data based on R, Git, Docker and HDF5”. The use of “modeling framework” is misleading (see also comment 1) because the paper fails to present any modeling or prediction which could be achieved from this framework.

4) The story basically presents a rather ambitious idea of automating data processing including quality control. The latter is not shown yet that it already works well so the product is not yet ready to be fully useful for the community. Once QC is implemented it could be interesting to see how it is done and how flexible the options are for the user. For instance, on page 14 L32 it is concluded “Once scientific QA/QC and uncertainty budget is implemented, the computational expense will likely increase by a factor of two to three. This suggests that eddy4R performs comparably to other flux processors.” As presented, the value from another EC flux processor tool is unclear in where it would really help but what is interesting is that the development is directed to a modeling audience who might also be able to use this tool if it was better explained. However, without clearly stated goals and sufficient supporting material to assess its quality and usefulness, it is difficult to evaluate the code framework for all its ambitious features. The paper is incoherent in its presentation (e.g. different components, datasets are presented separately without a clear thread creating multiple fragmented methods and results) and in many places the quality is diverging from the standards of a scientific paper.

Specific issues

5) The number of figures seems rather large and not all of them seem necessary. A heavy detail from different settings and configurations (e.g. Sect. 3.1.1, 3.2.1) could be nicely summarized in a table. The examples in Figures 9-13 require specific understanding of eddy covariance and do not help a modeler to adjust the framework for their needs. Even for the eddy covariance community, it might seem surprising that

airborne and tower data can be automatically compared, because there is a comprehensive quality control that needs to be performed on these data and it is not easy to automate such EC comparison at different scales (e.g. Mahrt, 1998), at least without multiple user interactions. For example, in Sect. 3.3 “Validation and verification” it is stated: “eddy4R includes a verification script which automatically processes subsets of the tower and aircraft data introduced in Sect. 3.1 and Sect. 3.2, and verifies the results against a reference, e.g. generated with a different software.” Where is this validation shown? Do you actually mean that you duplicate the processing (e.g. also with Eddy Pro) or just check selected files for consistency? The agreement in Figure 13 definitely seems surprising. It almost looks like the same dataset was plotted against the same dataset? The significant figures inconsistently range from 1 to 5. R2=1 is surprisingly good but not too meaningful (did you mean 1.0000, 0.9999 or 0.99)? I am also confused why the measured variables (e.g. w, q, CO2 mixing ratio) are compared with each other as they should have been the same unless the software interferes with the measurement data.

6) The data quality control does not seem careful. For example, in Figure 9 the periods of latent heat and CO2 flux were not rejected when the friction velocities were at their minima which look definitely below 0.1 m/s on days 114, 115, 116, 119, and other. It is also unclear what the gaps correspond to (rejected data, power interruption). I would be surprised if it was not possible to choose an uninterrupted dataset in the NEON’s large EC measurement network. It also seems weird in the same figure that the general temperature trend is anticorrelated with sensible heat flux (doy 115-118). Does the data output use normalized flux units for CO2?

7) The results and discussion also do not focus on the science but rather on what the software can do before the QC/QA are implemented. The QC/QA are the most important component of any data processing, so I am a little bit shocked that this has not been done before the submission and only raw data are reported. It is all about QC/QA so if it does not work well, the whole infrastructure could be in vain. Was it not

possible to wait until the QC steps are implemented? When will it happen?

8) Sect. 3.2.1 “Algorithm setting and profiling”. Can you define algorithm setting? There is no model algorithm here. By algorithm you probably mean the data processing routine which deals with technical issues of EC data handling such as “despiking of unphysical data”. This and other similar sections can be confusing for the journal readers. It is also unclear what you mean by profiling in this context as it can also have different meanings (I suppose you meant vertical profiling of EC fluxes rather than algorithm profiling). The authors should be careful not to use ambiguous terms and define clearly what they mean by model, algorithm, and other terms where the meaning is not unambiguous in the modeling context.

9) The choice of example figures 10 and 11 is not optimal because they are not well explained or sufficiently informative for the story. The figures are described only superficially what they represent but are not interpreted scientifically. The blue areas represent high deposition of methane? I am not convinced it is fair to show these data without the discussion of uncertainties which are definitely different in airborne and ground fluxes. The results section 3.3.2 are less than a paragraph so it cannot be informative. This should be made more general or explained much better for general audience of GMD.

10) There are other issues which are uncommon to see in a peer-review paper. For example, on Page 7, L225-239 the information is shown as bullet points more like a web-based manual or technical report which almost feels like from a magazine advertising IT Systems. It is interesting, that not even one paper is cited from the GMD community, and majority of the references are authors’ own papers published in specialist eddy covariance journals. I think it would make more sense to send the paper to one of those journals or make a better and balanced connection to the GMD literature realm. The EC data handling does seem promising but the approaches vary in various details among the groups so I found the author’s EC method review particularly unbalanced.

Overall, I like the general concept of the real-time data/model processing framework, but I expected the paper would be much more than just a teaser of an EC data-processing framework in progress. The revised paper should be guided by clearly defined science question(s) through a coherent story thread throughout the paper. If the intention is to publish the science in GMD, I would strongly recommend the authors to refocus the story on a solid connection between measurement data and modeling. One example could be a model-measurement testbed to validate models on observation data which could be very novel and useful for a larger audience including GMD.

References:

Mahrt L. Flux sampling errors for aircraft and towers. Journal of Atmospheric and Oceanic technology. 1998 Apr;15(2):416-2

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