Response to reviewer comments

bg-2016-189: OzFlux Data: Network integration from collection to curation; Isaac et. al.

We would like to thank the 2 reviewers for the time they have taken to read the manuscript and for their comments, suggested changes and proposed additions.

We begin with the comments from Reviewer 1 and for each reviewer we deal with their general comments first and then their specific points. Original wording from the reviewers comments are given in plain text and our responses are given in *italics*.

Reviewer 1

General comments

Paragraph 1

"On the negative side, some of the analyses and results drawn in the paper could be seen as not being fully supported by the evidence presented, in particular the comparison of data products. It seems likely that additional information has to be included to properly support conclusions or at least more detailed clarification on limitations and assumptions should be added."

We believe that the reviewer's concerns expressed here will be addressed by the changes suggested under Paragraph 3 below.

Paragraph 2

"How can the authors guarantee that the differences observed are due to differences in the methods and not in the implementation/coding of these methods?"
"For some of the methods, the authors indicate they had access to the original code (e.g., the SOLO ANN method), but there is no indication of cross-validation of results."

The primary purpose of the current paper is to document the processing techniques underlying the data used for this special issue and the data available to researchers from OzFlux and from the FluxNet 2015 synthesis. While cross-validation of different implementations of the same algorithm is an important task (the numerical modelling community has gone before us here), it needs to be done very carefully (see Fratini and Mauder, 2014) and in our opinion is better suited to a stand alone paper. A full investigation would need to cover not just different implementations of, say, the same u-threshold detection technique, the same respiration model etc but also the gap filling techniques (marginal distribution sampling, ANN etc) used for the meteorological drivers.*
To the specific point, the SOLO ANN, mentioned by the reviewer, the original code is called directly from OzFluxQC, we did not implement this in a different language. By definition, cross-validation of this would show identical results.
The current paper does introduce the issue, arguably to a limited extent due to the priority given to other material, by presenting results from different implementations of the same method in Figure 9 (OzFluxQC, ReddyProc and FluxNet nocturnal methods using the Lloyd-Taylor ER model, OzFluxQC and FluxNet day time methods using the Lasslop method) and it is discussed in the second paragraph of Section 7 Future Directions as an important area for further work.

Paragraph 3

“How "the redundant steps at L3" are skipped when already applied is not very clear. Is this done from metadata generated by other software packages, does OzFluxQC try to determine that from the data, or is this dependent on user input?”

We will amend the last sentence of Section 3.1 to clarify that the skipping of redundant steps requires user input but that OzFluxQC will fail with an error message if the choice made by the user is not consistent with the input data.

“This is particularly relevant in the comparison with EddyPro results. How can the results really be equivalent if steps such as filtering based on stationarity checks (mentioned in the paper) are not applied? At the same time, comparing results from EddyPro using processing options selected to match what is being done by OzFluxQC (when getting datalogger data directly) does not seem like a fair comparison.”
“So my suggestion is to remove this comparison (leaving the detailed version for a follow-up paper), and focus on exploring the NEE/RE/GPP comparisons in more detail as suggested above. This would entail removing the paragraphs that make up section 5.2 and Figure 8, which would not impact the paper significantly, in my opinion.”

The reviewer suggests removing Section 5.2 Comparison with EddyPro and the associated Figure 8 and to “... focus on exploring the NEE/ER/GPP comparisons in more detail as suggested above.” We are happy to remove Section 5.2 and Figure 8 and to leave a more detailed comparison of OzFluxQC and other flux processing tools (such as EddyPro) to a subsequent paper. This is consistent with the primary focus of the paper as set out in our response to Paragraph 2 of the reviewer comments above.
We are also happy to further explore the comparison of NEE, ER and GPP but we are not completely sure what the reviewer means by “... as suggested above” and would appreciate their comments on the following proposal.
The existing Figure 9 shows annual sums of NEE, ER and GPP for 9 OzFlux sites in 2013 calculated using 7 different methods, 3 of which are implemented in OzFluxQC, 2 of which come from the FluxNet 2015 synthesis data set and 1 each from ReddyProc and DINGO. The figure shows that while the 7 different values for NEE, ER and GPP show the same site to site variability, there is considerable variability between methods within individual sites. The variability between methods at individual sites is discussed in the existing manuscript but not pursued.
We propose to expand the comparison of NEE, ER and GPP presented in the existing Figure 9 by focussing on annual sums of NEE, ER and GPP for 3 sites, AU-Cpr, AU-DaS and AU-Whr, from 2 groups of methods; nocturnal, u-filtered (OzFluxQC, FluxNet and ReddyProc) and light response curve (LRC) intercept (OzFluxQC and FluxNet). Our objective is to explore the variability between the nocturnal methods and between the LRC methods as functions of u* threshold (nocturnal only) from the 3 implementations, of gap filling techniques (all methods) and of parameter values from the non-linear curve fit process (all methods). The hypotheses are:*

a) *that variability between methods at the sites tested can be primarily attributed to one of the possible causes listed above and*
b) *that this information can be used to direct efforts aimed at resolving discrepancies between different implementations of the same method.*

Most of the additional analysis required by this proposal can be done with data prepared for the original manuscript. The new material would be presented as Section 5.3 Comparison of Methods (the existing Section 5.3 will be renamed to Section 5.2 following the removal of the existing Section 5.2).

Specific comments

1) It is not clear in the paper how the alternative data from AWS, ACCESS-R and ERA-Interim are reconciled and used in the gap filling process. How one (or more?) of the three sources are selected and used for gap filling of driver variables?

   This is a good point. To address this, we propose to add a short paragraph (3 sentences should suffice) between the existing paragraphs 1 and 2 on page 14 of the current manuscript. The paragraph will emphasise the differences in data available from AWS, ACCESS-R and ERA_Interim (these are introduced in Sections 2.5.1, 2.5.2 and 2.5.3), describe the default hierarchy of choice made by OzFluxQC, how this can be changed by the user, the circumstances which may justify such a change and the evidence provided to support such a change.

2) Also on gap filling, at the end of section 3.5, the issue of gaps being close to the window size can seriously restrict the applicability of gap filling methods. A bit more detail on the "minimum points criteria" mentioned in the paper would be very helpful in clarifying how this problem is addressed.

   We will address this point by re-writing the 2 sentences at the end of Section 3.5 to clarify the automated method used by OzFluxQC to determine the optimal window size when the gap size approaches or exceeds the default window size. The description of the role of the minimum points criteria will also be clarified.

3) At the end of section 3.6, there is a description of final NEE, ER, and GPP time series, from what seems to be a combination of the results or the three used for partitioning. However, the same last few lines seem to indicate that NEE and GPP obtained from the hyperbolic light response curve method are not included in this these final versions of the variable. Is this correct? If not, how is the light response curve results incorporated? If yes, it's not clear why this method would be included in OzFluxQC (and in the paper).

   The reviewer is correct. The 3 methods for estimating ER, and hence gap filling NEE and calculating GPP, are introduced at the start of Section 3.6 but it is not made clear that all 3 ER estimates and the corresponding gap-filled NEE and GPP, are available as separate outputs from OzFluxQC. We will address this point by adding a sentence to the end of Section 3.6 that makes it clear values of ER, NEE and GPP from all 3 methods are output by OzFluxQC and that the user is free to choose which one they believe is the best for their site. We will also introduce subscript notation in Eqn. 1 and 2 and in the use of ER in the third paragraph of Section 3.6 to clearly indicate the 3 sources of ER estimates. Eqn. 4 will be expanded to contain 3 equations for GPP with subscripts indicating the different sources of ER.

4) The sentence below (from page 5, lines 23-24) is a bit awkward and could be rephrased for clarity: "The large range in precipitation amount and seasonality has resulted in a large range of biomes across Australia and OzFlux samples the majority"

   The wording of the sentence will be changed to clarify the sentence meaning.

5) Regarding the use of the CF Conventions, how are variables not covered in the current version of the conventions handled? This is probably worth one or two sentences for clarification.

   The reviewer makes another good point. A sentence describing the treatment of variables *not covered by the CF Conventions will be added to the preamble of Section 3. A further sentence will be added to Section 7 Future Directions that notes our intention to propose to the CF Working Group standard names for those flux data set variables that are not currently covered by the CF Conventions.*

6) For all gap filling comparison plots, it would be relevant to include indications of the amounts of missing data. These can be good indicators of variability in the data (or lack of variability, for some gap filling methods).

   The percentage of missing data for each site will be added as text above the top axis of the top panel in each of Figures 4, 5 and 6.

7) Also on the plots that use color, adding a color legend would make the plots considerably more readable compared to the text descriptions of the color associations in the figure captions.

   Figures 4, 5, 6 and 9 will be changed to include a colour bar as suggested.

8) On page 14, line 27: "drives" should be "drivers"

   The wording will be changed as suggested.

9) Section 4, which describes the data portal and data distribution, somewhat breaks the flow of the paper. Maybe moving the text in this section to section 2 would make the discussion on the data pipeline more fluid.

   This is a good suggestion and we will move Section 4 to become Section 2.6. Our intention with the original structure was reflect the typical work flow of a site PI (one processes data and then makes it available) but as the reviewer points out, this is not necessarily the best structure for a research article.

10) Finally, there is an issue that is not necessarily a discussion for this paper, but very relevant. There is a natural conflict between trying to encourage site teams to "know thy site" while providing the comfort of using standard (and vetted) software toolkit. In many cases, even experienced eddy covariance researchers will be tempted to not worry or question much about the results obtained using highly automated tools. Offering visual presentation of intermediate results, as done by OzFluxQC, is certainly a step in the right direction and is appreciated.

    This an excellent comment by the reviewer and goes straight to the heart of a debate that has raged within the first author for years. On the one hand, we want to minimise the time researchers spend in the drudgery of data processing so they have more time for the interpretation of results. On the other hand, sometimes one simply has to look at data, and at lots of data, in order to obtain a deeper understanding what an ecosystem is doing. The pressure to produce quick results does not help this process of deep involvement in the data. The first author has not found a foolproof way of helping others to walk this fine line and so feels unqualified at this stage to put anything into writing. The most effective way has been practical demonstrations of the need to visually examine and to rigorously question the data and the best time to do this demonstration has been during training exercises in the use of the software.
    We acknowledge the reviewer's insight but feel the topic is worthy of it's own paper and that to insert something here could not do the subject justice.