Dear Editor,

thank you for recommending the revised manuscript for publication. In this document we address the minor revisions suggested by Referee #3 and provide the manuscript with marked changes below.

Wutzler and others demonstrate the REddyProc package. I feel that the manuscript will be an important contribution to the flux community given the number of users who will likely use the package.

Thank you.

Abstract was a bit qualitative, numbers would help. For example, how much lower and higher are the uncertainty estimates noted toward the end of the abstract?

We updated the abstract:
Lower uncertainty estimates of both u ∗ and resulting gap-filled fluxes by 50% with the presented tool was achieved by …
Higher estimates of uncertainty in day-time partitioning (about twice as high) resulted

'allowed a boost in' is too wordy.

We replaced by „greatly advanced“

'and temperature' instead of 'or temperature' on line 7 page 2.

We corrected as suggested.

Figure 1 already includes a thumbprint plot for RE that seems a bit implausible. Is the red/orange horizontal line through the middle of the plot really true?

The "reddish" band does not come from "gap filled" NEE estimates but the band coincides from mainly measured data where there is a very high NEE at nighttime (if you look the nighttime NEE immediately before the big gap looks quite high). Hence its in the original measurement data. The color scale is assigned so that the maximum value of the dataset gets the red color code.
We decided to not discuss this distracting detail in the paper.

The legend of Fig. 2 is a bit elusive but can be figured out with careful study. Clarifying it would help.

We reworded the figure caption as follows:
Concept of the u∗-filter: Night-time NEE at low u ∗ is biased towards lower NEE values compared to cases with higher u∗. Unbiased NEE should scatter around the same plateau because environmental conditions are similar. The u ∗ threshold (dashed line), i.e. the value below which this bias is considered significant, is estimated by a moving point method on u ∗ bins (crosses) across half-hourly records (circles). The example uses a subset data from DE-Tha.

P 5 and elsewhere: please use 'van Gorsel'. Regarding section 2.1, will improved approaches following van Gorsel et al. (2007), sigma_w-based approaches (Jocher et al., see earlier work by Acevedo: https://www.sciencedirect.com/science/article/pii/S0168192308001962), estimates of drainage (see Hayek et al., 10.1016/j.agrformet.2017.12.186), atmospheric stability approaches (Novick et al., 2004) or other non-ustar approaches be implemented? It's been my feeling that a large part of the flux community has moved beyond ustar.

We apologize for not correctly citing the two-word surname and corrected it. Currently there are no plans to include more recent approaches. The issue tracker on the ReddyProc github site is a good place to suggest, argue for, and detail such a suggestion.

2.1.1: I've long-wondered what is the best threshold for 'nighttime' especially given the characteristic decline in turbulence with the collapse of the daytime boundary layer may occur earlier. Is there a justification for Rg < 10 W m-2? (see also the last line of page 8).

We selected this threshold and other similar values because we had not reason to deviate from the standard approaches, here DP06. We did not discuss this issue in the manuscript but give some thoughts here in the replies.

The night-time threshold for u* estimation 10 Wm-2 as used by Papale et al 2006 was selected to our knowledge mainly to 1) guarantee a certain amount of nighttime data for the estimation of E0 and Rb (this we think is the most important reason); and 2) to avoid problems in the selection of nighttime data in sites with drift in the pyrometer (i.e. positive values at nighttime due to wrong calibrations or drift that were pretty common in the past, now we think that the quality control is stricter in this regard). Also for the point 2 the crosscheck with the potential radiation is important.

Further, the Rg measurement can be a little unclean and therefore small values below 10 are included, for example bright moon hours at night, slight offsets of the instrument, or a not fully dark sky due to some back scatter from low hanging clouds after sunset. For the partitioning, the algorithm additionally even filters for nighttime by calculating the potential radiation and only "true" nighttime values will be considered.

Layering might already set in before, i.e. at higher radiation values like 30 W m-2 and u* estimation and filtering could include even more data points. We are not aware of any study on this and suspect a good threshold to be very site specific and maybe even season specific - for example thinking of hours with low sun angles in Northern sites. However, for this, REddyProc allows the PIs and users to experiment with their own Rg thresholds.

Regarding 2.2.2, often alternate micrometeorological information is available. Is it possible in REddyProc to add multiple micrometeorological measurements of the same variable?

We do not fully understand the request. E.g. „Using PAR when Rg is missing“ is currently not implemented. However, the user can flexible define different meteorological variables and thresholds with the Lookup-table approach. However, then it is not "the" MDS algorithm any more. There is also currently no feature to use a different temperature column aside from air temperature yet. Again, the github issue tracker would be a good place to propose and details such a feature.

On page 10, BGC16 was used and defined before. Please define it at first instance.

Section 3 is dedicated to describe the benchmark tools. When using acronym BGC16 before, its difficult to describe it sufficiently. Hence those usages refer to section 3 (e.g. Table 1 and p8l10).

‘Papale C-implementation’ on p. 13, just use DP06. A quick re-read for minor issues like this would improve the paper. Also, what is a ‘big scatter’ on page 13?

We hope that a re-read captured those minor issues, as we corrected on page 13.
We reworded „Big scatter“ by „although individual threshold estimates differed“

Regarding the LRC fits, is a least squares cost function applied or least absolute deviation to avoid the influence of outliers (see Richardson 2005/6 papers). Some discussion of the cost function is necessary in my opinion.

The fitting for least absolute deviation is appropriate if measurement error is Laplace distributed, which seems to be the case both for soil respiration and NEE data. However, when comparing fluxes at similar environmental conditions and hence, of similar magnitude, the distribution is better approximated by a Gaussian distribution, and the apparent Laplace shape is probably generated by a superposition of several Gaussian distributions, where standard deviation scales with flux magnitude (Lasslop 2008). Hence, in accordance with Lasslop 2010 we used the usual normal assumption in fitting the LRC curve.
We extended the section on fitting the LRC curve and repeated the cost equation from Lasslop et al. in supplement 1 (section 2 , step 3).