Response to comments from reviewer RC2

Note that reviewer's comments are in black font, and responses in blue font.

Padrón and others analyze nocturnal evapotranspiration measurements from eddy covariance and estimates from models. The analysis is interesting and certainly novel although a few methodological points need to be reconsidered in my opinion, and the text could be improved in multiple instances.

We appreciate the positive opinion about the relevance of our manuscript.

Sentences like 'Lombardozzi et al. (2017) compiled evidence of this from 204 species' aren't particularly instructive. What did they find? In the paragraph at the bottom of page 1 try to make the scientific findings, not the authors, the subject of the sentences. For a discussion of this see https://schimelwritingscience.wordpress.com and the associated book.

We appreciate the suggestion. We modified the text to improve the focus and readability.

A more powerful way to synthesize the literature, which would make the present manuscript more citable, would be to synthesize existing studies in a table to help further motivate the present analysis and be more comprehensive.

We appreciate the suggestion. We now introduce Table 1 in the revised manuscript to summarize nocturnal water loss estimates from the literature.

The points about dew and hoar frost are great.

Thank you.

P 2 line 22: disentangle aerodynamic vs. surface conductances more clearly. The surface has both stomatal and boundary-layer resistances.

We clarified this. We now provide a more complete description of the resistances included in models to compute latent heat flux.

2.1.1: Why is the 10 W m-2 threshold used to differentiate between day and night? Sensors have uncertainty but the solar zenith angle can be calculated with extreme accuracy for environmental science applications. Are results sensitive to the 10 W m-2 threshold? I see that a zenith angle-based analysis is done in section 2.1.2 (sun up and sun down). Why are different approaches used? What are the 'cases described by Hirschi et al. (2017)'?

Here we use this simple threshold because the focus is on the comparison of the lysimeter and EC data, and we wanted to be consistent with the comparison from Hirschi et al. (2017). The results are hardly sensitive to the 10 W m^{-2} threshold.

We extended the sentence to clarify the meaning of 'cases described by Hirschi et al. (2017)'. It corresponds to cases when the tower is upwind of the sensor and thus disturbing the air flow.

P 3 line 30: using a static value for the latent heat of vaporization is fine, but it's easy to add its temperature sensitivity to add a bit more accuracy in the latent heat to water flux conversion.

Yes, we are aware of this, but for simplicity and to avoid dealing with possible missing temperature data, we assume a value of λ corresponding to a temperature of approximately 12 °C. We trade simplicity for a very small loss in accuracy. In addition, note that we do not incur in a highly biased error, given that temperatures are likely to be sometimes greater and sometimes less than 12 °C.

2.1.2: The Bowen-ratio-based assumption is a bit problematic; there is extensive evidence that undermeasured sensible heat flux from large eddies plays a large role in lack of energy balance closure. That being said, these factors are less important at night where low-level jets and decoupling of the eddy covariance sensors and the canopy often dominate.

We appreciate the insight. We now analyze the uncertainty of NWL estimates with and without the Bowen ratio assumption in Figure 2.

2.1.2: instead of emphasizing caution, perhaps don't use gap-filled fluxes for the analysis. This is a hard thing to do at night when eddy covariance data are often less reliable than many people believe.

Gap-filled fluxes are required in order to obtain the total NWL estimates shown in Fig. 2. An alternative would be to estimate a mean hourly NWL rate from the non-gap-filled observations and obtain total sums by multiplying the mean by the total number of nighttime hours. However, this has its own disadvantages. Nonetheless results are rather similar with both options.

When analyzing the correlation of NWL with environmental conditions in Fig. 4 we do not employ gap-filled data.

Thinking broadly, is 'nocturnal water flux' better than NWL given that water can be both lost and gained (but is admittedly a net loss over the time scales mostly investigated here).

In a first draft we also used nocturnal water flux but decided that NWL is more appropriate to communicate our results.

3.1: why is the second threshold chosen? Is it appropriate for the site or just simply half of the previous threshold?

In this case is just half of the defined threshold value to provide an estimate of the sensitivity. We revised the text to make this clearer.

Fig. 2 and elsewhere: what are representative uncertainties of the site-level NWL measurements/estimates?

Figure 2 now includes four different NWL estimates from FLUXNET sites: without energy balance correction, and with the 25, 50 and 75th percentile of the distribution of energy balance corrected fluxes. The text accompanying the analysis was modified to convey this point more clearly.

This statement should be in the Methods: "These annual mean values are computed from monthly climatologies obtained by omitting months with half or more of missing latent heat flux data."

Note that there is no specific "Methods" section in the structure of our manuscript. We thus think the location of the statement is appropriate.

In general, the assumptions made in the flux processing for NWL for the FLUXNET2015 database needs to be explained in more detail.

We expanded the text. Note also that a full description of the processing is provided at https://fluxnet.fluxdata.org/data/fluxnet2015-dataset/data-processing/, as indicated in the text.

The statement 'Nonetheless, deciduous broadleaf forests (DBF) have an overall lower NWLf, whereas evergreen needleleaf forests (ENF) include most cases with higher NWLf' suggests to me that perhaps difficulties in measuring the surface-atmosphere flux is partly responsible here. ENF needles are more closely coupled to the atmosphere on account of their smaller dimensions and I can't think of a discernable reason why DBF would have particularly low NWL. Although perhaps relative NWL given that they are frequently found in mesic regions.

We appreciate the insight and now include it in the text. Note that cross correlations and confounding factors might also be relevant.

Figure 4 is tricky to look at. I'm curious to know if there is a more logical way of presenting these complex data.

We increased the size of the symbol representing the mean to convey the main message. We expect that the text also helps to understand the Figure.

The analysis of models is interesting, and the degree of discrepancy is surprising.

Thank you.