Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2017-573-AC2, 2017

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Local tower-based merging of two land evaporation products" *by* Carlos Jiménez et al.

Carlos Jiménez et al.

carlos.jimenez@estellus.fr

Received and published: 21 December 2017

The authors investigate the added value of merging two land ET products based on their performance with respect to tower-based ET. This is definitively an interesting topic in particular in the light of the existing large uncertainties in ET estimations. General comments and questions:âĂÍThe study is well written and provides interesting insights in the performance of the two used ET models. These seem to perform very similar and the merge of them does not provide a significant added value. I'm wondering thus if the use of other WACMOS- ET models with more diverse performance at the tower sites could be a better test case for the proposed merging procedure (instead of having two already similarly well- performing models with not much of room for improvement). Can the authors comment, why they did not include a more diverse

palette of models?

R. First, we would like to thank the reviewer for taking his/her time to review the paper.

We choose GLEAM and PTJPL to study a possible product merge as we thought that the more interesting challenge would be to merge the two project ET products showing more skills to capture tower fluxes and large-scale inferred evaporation. Being both already closer to the tower fluxes, we wanted to see if adding the tower information could result in a better ET product, which could have been of utility for the project. The reviewer is possibly right than merging with one of the less performing-models would have resulted in a more diverse performance at the towers and perhaps a more illustrative merging exercise, but as mentioned before, our main objective was to improved the performance of the two more skilled models.

In any case, as mentioned at the end of the paper, we plan to continue this with ET estimates of finer resolution, derived with more independent forcings, and for more extended periods. It is likely that the more independent model forcings will result in a more diverse performance at the towers. This, together with a larger pool of tower stations from the more extended period, could make the tower fluxes more informative, contributing more to the merged product. We will also be looking at adding other products, such as a newer global version of SEBS at 5km and daily resolution currently available from the SEBS developer team, or the new MODIS ET product version 6.

Also, how are data gaps in tower data and non-consistent temporal coverage of the towers treated? This might influence the analysis of the derived merging weights if the station sample changes over time.

R. There are certainly gaps in our processed tower data because, even if a station data record was complete, we remove the rainy intervals. We deal with this by requiring a minimum number of 20 daily observations in the running window selecting the time interval to derive the daily weights. Otherwise the weights will not be computed for that specific day. The latter happens in a very few occasions so it is not critical for the results.

Regarding temporal coverage, for the 84 stations considered, 6, 14, 24, 9, and 31 stations had 2, 3, 4, 5, and 6 years of data, respectively. Clearly, the shorter the time period the less confidence we should have in the weights because the weights may be too specific to the climate conditions of these few years (e.g., especially dry or wet conditions). If these weights were then applied to merge the products at that location but for more standard conditions, the merge may not be optimal.

So, yes, there can be an impact due to data gaps and temporal coverage, but how this is influencing the analysis is difficult to judge. A possibility to minimize the impact of the temporal coverage is to keep only towers having data for a relatively large number of years, but that considerably reduces the already small number of towers. For instance, allowing at least 4 years of data removes half of the towers considered in the study.

Specific comments:âĂÍ page 3, line 12: What about other observational data? E.g. lysimeters or catchment wide estimates could be mentioned as well here.

R. We have used catchment wide estimates to evaluate our ET models during the WACMOS-ET project (Miralles et al., 2016), but we would possibly not call them observations, at least not ground observations. Lysimeters are definitely ground observations. However, as far as we know there is no organized network of lysimeter measurements, which may facilitate their widespread use for global evaluation of ET models. Nevertheless, we can mention them. We will rephrase as: "Ground measurements of land heat fluxes are typically conducted during field experiments (Pauwels et al., 2008) and by more permanent lysimeters (Hirsch et al., 2017) and flux tower networks (Baldocchi et al., 2001)".

âĂÍpage 5, lines 24/25: Do you expect an impact on the merging weighs when sub-daily simulations and tower data would be considered?âĂÍ

R. Yes, but only if the ET models skills were markedly different at different times of

the day. In that case the weights for specific times of the day would differ from the daily ones. However, from our sub-daily simulations presented in Michel et al. 2016 we did not notice very different skills at different times of the day when comparing with the tower data, and not much more skill was gained by producing daily ET based on 3-hourly input as opposed to forcing the models with the original daily input. We could then speculate that the weights will not be dramatically different in our case. In any case, we do not envisage to produce more sub-daily runs within this project given that, on the one hand, we do not have had many requests for sub-daily simulations estimates, and on the other hand, ET model updates are mostly focusing only on the daily scale.

page 7, line 18: Obsolete brackets between the two cited papers.âĂÍ

R. Thanks, we will remove them.

page 8, line 1: What happens if the two data sources for precipitation disagree?âĂÍ

R. Certainly there are moments when the precipitation at the tower disagrees with the gridded precipitation, which is expected due to the different spatial resolutions and the unavoidable errors associated to the gridded product. We only leave days when there was no rain from the gridded product and the tower recordings (for the towers where precipitation was measured). We will rephrase to make it clearer as: "(2) masking measurements for rainy intervals, only leaving observations if both the global precipitation product and the local measurements (if available) do not indicate precipitation (eddy-covariance measurements are generally less reliable during precipitation events)".

page 12, line 8, "estimated over the time series of available errors": What about differences in the length of the EC time series or differences in the occurrence of data gaps between the towers? How is this taken into account in the analysis of the weights?

R. As we explained above, all the towers do not have the same number of years of data, and there are data gaps, especially because we filter for precipitating conditions.

This could have an impact in the weights analysis, especially if for the stations with a very short number of years, the existing years are not representative of the typical climate conditions, but we cannot quantify the impact of this. Being much stricter with the data gaps and number of years was not possible here, as the number if towers and weights would have been considerably reduced, as previously discussed.

We will add the information about the number of years in the text: “This processing selects 84 stations for the 2002-2007 study period. Notice that not all stations completely cover this period, with 6, 14, 24, 9, and 31 stations having 2, 3, 4, 5, and 6 years of data, respectively. Their geographical locations . . .”.

We will also comment on the possible impact of a relatively short number of years to derive the weights when describing Figure 7, as the US-SRM station shows large inter-annual variability. We will add: “Figure 7 showed that at some stations inter-annual variability is large. At these stations, if only a relatively short number of years exist to derive the weights, the weights may not be representative of the overall climate conditions at the tower. For instance, for the US-SRM site, weights derived by only considering the year 2005 or the year 2007 would have possibly been quite different. This suggests that there can be an impact regarding the different number of years available per station, and that more confidence should be placed in the derived weights the longer the time period available at the tower. Limiting the study to stations with a relatively large number of years can be used to minimize the impact of this, but in our case that severely reduces the number of towers. For instance, if we only derive weights if at least 4 years of data are available, half of the towers would have been removed.”

page 13, lines 3-5: Rephrase: "A 61-day running window was found to provide ..."

R. We will rephrase like that.

page 13, lines 3-5: Is there a minimum requirement of data availability for deriving the weight within the window?âĂÍ

R. Yes, 20 daily observations as mentioned before. We will add this to the text as: “A number of 30 days before and after each calendar day was found to provide a good compromise between the smoothness of weights and the number of samples required, so a 61-day running window was found to provide the daily weights.âĂÍ Notice that due to the masking of the tower data at very few occasions the 61 daily estimates are present in the running window, and 20 daily values reasonably spread in the running window are required as a minimum number of ET estimates to derive the weights”.

page 13, lines 9-13: Sounds a bit confusing and not so clear (at this point of the paper at least). Try to re-formulate being a bit more specific (reasons to believe that?).

R. We are just stating here the impossibility to evaluate interception as we mask the data record for precipitation conditions. We will rephrase trying to be clearer as: “As the tower measurements are masked for rainy intervals, we cannot evaluate the interception loss of the modelled ET. Therefore, only the sum of soil evaporation and transpiration is compared with the tower data and weighted. This is clearly a limitation as in most cases the total ET, and not only the soil evaporation and transpiration, are required. This means that an estimate of interception loss should also be provided, either by assuming that GLEAM and PT-JPL interception loss are equally uncertain and adding their average to the weighted soil evaporation and transpiration, or by adding just one of the GLEAM or PT-JPL individual interception losses, if there are reasons to believe that one is overall more certain than the other. âĂÍ page 14, lines 14-20, "optimum product": How often is the optimum product one of the two models (i.e., weight of one)? From Fig. 5 it looks like a weight of one is never reached.âĂÍ

R. On average over all stations, 80% of the times, so quite often. Or in other words, on average over all stations only once every 5 days the tower ET is between the GLEAM and PT-JPL estimates, and the optimum product is the tower ET. But notice that this is not related to the weights of Figure 5. The reviewer is possibly confusing the weights derivation with the construction of the optimum product. See please our response to your last comment, where we clarify the relation between weights and the optimum product.

page 19, Fig. 4: The legend interferes with the figure information, please increase the y-axis range a bit.âĂÍ

R. Sure, we will fix that.

page 19, Fig. 4 caption: Is it the optimum product you are comparing with the tower ET, not the WA-product? From page 14, I get that the optimum product is either the tower ET or one of the two models. Assuming that it's often the tower ET (judging from Fig. 5 where weights never reach one), shouldn't the RMSDs become zero? Perhaps, the relation of the optimum product and the WA-average product is not completely clear (at least to me) and might deserve some clarifications in the text.

R. Yes, it is the optimum product. We will try to clarify the differences between the weights of the merge product and the construction of the optimum product in the following lines.

The weights of the merge product are derived by calculating the variance of the error distribution over a time interval, as described in Section 2.3. For the weights to be close to one the error variance of one of the models has to be much smaller than the other one, which is seldom the case. Maximum values tend to be around 0.9, as it is illustrated in Figure 6. This means that the merge product is rarely just one of the GLEAM or PT-JPL estimates.

Now, the optimum product is not derived by applying those weights, but by comparing tower and model estimates at every day, as also described in Section 2.3. If the tower ET is between the GLEAM and PT-JPL estimates, the tower ET is then the optimum estimate, but if both the GLEAM and PT-JPL estimates are below or above the tower ET, the closest model ET is the optimum estimate. So this is done independent of any weight estimation, and it is not related to the weights presented in Figure 5, which are the weights of the merge product derived by using the error variances.

For the RMSD of the optimum product to be zero at a given tower, that would imply that the tower ET is always between the GLEAM and PT-JPL estimates. This is never the case for the 84 considered stations.

We try to make this clearer this by rephrasing the description of the optimum product as: " When improving a product and comparing with a reference, the common targets are correlation unity and zero RMSD. Here, instead, we define a product that minimizes the RMSD with the tower ET, and call it, in the context of our merging strategy, the optimum product. At the days when the tower ET is between the GLEAM and PT-JPL estimates, the tower ET is the optimum estimate. But there is also the possibility that both GLEAM and PT-JPL estimates are below or above the tower ET. For the 84 towers considered here this is the most common situation, happening on average over all stations for around 80% of the days. In this case the merge product will never be the tower ET because, independent of the weight values assigned by the error variance analysis, the merge product is bound by the original model ET estimates. This is precisely the reason why correlation unity and zero RMSD can never be achieved here. For the optimum product, the closest model ET to the tower ET is the value that minimizes the RMSD with the tower ET, and this is the value selected for the optimum product in this case".

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-573, 2017.