Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2017-147-RC1, 2017

Hydrology and
Earth System
Sciences
Discussions
Open Access

Interactive comment on "Derived Optimal Linear Combination Evapotranspiration (DOLCE): a global gridded synthesis ET estimate" *by* Sanaa Hobeichi et al.

P. Dirmeyer (Referee)

pdirmeye@gmu.edu

Received and published: 3 May 2017

The authors present a method to more optimally combine global ET estimates using FLUXNET sites as anchor points for cal/val (which they name DOLCE), producing a demonstrably better product when validated at FLUXNET sites (even with cross-validation), although there are many instances where DOLCE is outperformed by one or more other products at individual sites. The approach presented here is limited to providing a single weighting function per input data set (see Table 3) and is limited to a 10-year period. The approach is flexible and can incorporate other / more input data sets and validation sites easily.

This is an interesting, useful and worthy paper, demonstrating real improvements but

also casting light on the difficulty of producing vastly improved globally distributed estimates of ET. I very much appreciate the attempts to identify the effects of misrepresentation of vegetation classes. I do have a number of comments:

General comments:

1. There is obviously a question regarding the representativeness of the FLUXNET stations to their regions (grid box average or dominant surface conditions), which the paper addresses fairly well. However, one specific issue jumps out at me: what about rare representatives, particularly in the tropics? FLUXNET is woefully thin on stations at low latitudes. The out-of-sample tests are like an OSSE, but there is little sampling in the tropics to play with here. What was the specific effect of denying low-latitude stations? Can we use this study to make case to prioritize FLUXNET expansion into the tropics, S/SW Asia and the Southern Hemisphere? I would like to see this issue discussed more, where the global maps are presented and also in the Discussion. Motivating targeted FLUXNET network expansion would be a good "broader impact" of this study.

2. What about the uncertainty/error in the FLUXNET observations? They certainly contain random errors and systematic biases. For a measured quantity that has a red-noise spectrum (following a Markov process) the random error can be estimated from the behavior of lagged autocorrelations (cf. http://dx.doi.org/10.1175/JHM-D-15-0196.1). Random error will systematically degrade correlation and affect other statistics (cf. http://dx.doi.org/10.1175/JHM-D-15-0063.1). Systematic error is more difficult to identify, but the energy balance corrections in FLUXNET2015 give some clue (see specific comments below). All these mean that using FLUXNET for cal/val is itself flawed and imperfect. On the other hand, model ET is inherently precise (no random error) and can be used to estimate/differentiate this type of error from others by noting its statistical differences from the instrument records. This issue should be acknowledged and discussed, including how the assumptions in DOLCE regarding errors (that they are uncorrelated) affect results. In other words, more discussion about uncertain-

ties.

3. In the Supplement: There are captions for Fig S1 and Fig S2, but the PDF does not contain any figures!! Please recheck the rendering.

Specific comments:

P3 L6: Change "85 FLUXNET tower data" to "85 FLUXNET towers"

P3 L8: “ground-truthed” is not an appropriate verbification for this context. Change ’...gridded ET data sets are “ground-truthed” using flux tower data from FLUXNET...’ to ’...tower data from FLUXNET provide ground truth for gridded ET data sets...’

P5 L16: Change "where" to "were"

P7 L22: RSD cannot convey whether the variance is too large or too small. So for the changes shown in Fig 3b - we cannot tell whether the improvement is due to an increase or decrease in standard deviation. Furthermore, what about locations where mean ET (denominator) is near zero - does that explain some of the very large values and changes?

Fig 3: So the spread is across 5000, and each of those is an average of 25% out-of-sample stations, right? It is not 5000x172x0.25 points. Please make clear. Also, Fig 4 suggests individual o-o-s stations frequently fare worse. Transferability of calibrations appears to be kind of weak, which is not a surprise. Calibration transferability is a very difficult enterprise. This should also be acknowledged, either here or in the Discussion section (wishing PILPS-San Pedro had been completed as it would have really shone a light on this problem).

P9 and Fig 4: "widespread out-of-sample improvement that this approach offers over existing gridded ET products" seems like a bit of an overstatement - there are frequently locations that fare worse, and sometimes much worse. In Fig 4, there is typically a tremendous range, and usually the central two quartiles encompass the zero line. For RMSE it appears that ∼20-49% of the time the estimates are worse than other

products and methods, 30-55% for RSD, and 20-55% for COR. While there is definitely a net (and welcome) improvement in almost all cases, often the DOLCE estimate is worse. This should be acknowledged.

P9 L26: "Standard Deviation (SD) difference" - this is clearly not the same as RSD defined earlier - please define, which minus which?

Table 2 has little accompanying discussion, and what is there is very shallow adding little to comprehension. Please discuss more or remove the Table if it does not warrant discussion.

Fig 6: Here it could be said you use DOLCE to estimate MPI, and most places have a positive bias. But the energy-balance-corrected fluxnet data, which close the surface energy balance by construct conserving Bowen ratio, consistently increases ET compared to the raw measurements (at 107 of 122 FLUXNET2015 Tier-1 stations during JJA by my quick calculation). There is recent independent indication that tower sites bias low because of errors in the turbulence theory applied to estimate fluxes (see: http://dx.doi.org/10.1002/2017GL073499). Could FLUXNET instrument error (and the simplicity of the energy closure correction) contribute to systematic biases in DOLCE? Please discuss here or in the Discussion section.

Re Fig 7: Much of the largest differences are at low latitudes where there is little FLUXNET data for calibration. Please discuss, as I see this as a major issue (more with FLUXNET station distribution than your methods, but the problem of representative ET estimates in the tropics is an ongoing concern).

P10 L21: "ET doesn’t exhibit any seasonal change over Greenland and the deserts in North Africa..." - not expected in the absolute, because mean values are tiny. It would be more informative to also show relative (percentage) changes, which are more relevant for local water balances.

Discussion §: Please also speculate whether some spatial variability in weighting could improve estimates further, even if only in 2 or 3 categories of weights.

Fig 9: Only stations in the tropics are 2 EBF stations; the savannah stations are in southern Africa. These are not o-o-s results, right? I would have expected these types to stand out more, but perhaps the samples are biased to the extratropical stations in the categories. Thoughts?

Also Fig 9: Crops are tricky. The category is a catchall that is unsatisfactory because there is such variability in phenology, seasonality, stomatal resistances, etc. Are any of the CRO stations rice (which acts very different because of seasonal flooding). Not surprising many of the biggest errors are there.

P13 L1: "...for example anthropogenic water management..." - but it was stated earlier that irrigated sites were excluded. Please expand on this comment.

Conclusion: Please also state plans for updates, future versions, perhaps (hopefully) covering a longer time period!?

References: There seem to be a lot of redundancies in author lists for papers with names showing up 2 or 3 times for the same entry. Please check.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-147, 2017.