Biogeosciences Open Access
Discussions

Interactive comment on "Evaluating terrestrial CO$_2$ flux diagnoses and uncertainties from a simple land surface model and its residuals" *by* T. W. Hilton et al.

T. W. Hilton et al.

thilton@ucmerced.edu

Received and published: 17 October 2013

Author responses to Matthew Smith (Referee #1)
Referee comments in boldface, author responses in normal typeface.

Page 13755, sentence beginning line 10. This is hard to read. Please reword.

Thank you for pointing this out. We have reworded the sentence.

Page 13757, sentence at the end of line 8. Use of "We" seems inappropriate here

Thanks for pointing this out. We have reworded the sentence.

On "Cross validation". Arguably you do not use rigorous cross validation. You have one set of test data. A more rigorous approach would be to do 5-fold cross validation on the 65 sites, assess model performance using those 5 sets of held out sites and leave the 27 FluxNet sites for final evaluation. Having just one cross validation dataset can lead to overfitting through many loops of model improvement, refitting, and recomparing with the no-longer independent test data. That said - this is an overly critical criticism of me and I applaud the authors for assessing model performance against a substantial independent test dataset. However, what is lacking from the methods for me is a description of how and why the particular 27 FluxNet sites were selected for assessment - where they randomly selected? Were they the most recent sites? If they were always next door neighbours to the original FluxNet sites then they're not independent. Basically - it'd be good to convey your objective reasons for selecting these sites for model evaluation. This comes back again in the caveats where you say "The 27 cross-validation sites (Fig. 2, Table 2) generally have shorter observational records than the sites used for VPRM parameterization. Repeating the cross-validation experiment with different, perhaps randomly selected subsets might be a useful exercise." So why where these ones chosen in the first place?

We agree that rigorous testing against independent data is crucial. The 27 tower sites used for cross validation were initially set aside for data-availability reasons. It occurred to us later in the process that these sites were more useful as an independent test dataset than as additional parameterization data.

In the methods section 2.2. you don't really convey any hypotheses for why parameters might be best represented at these various degrees of spatial and temporal aggregation. It would be good if you could state any a-priori hypotheses or evidence for whether you should expect these parameters to aggregate at any particular level - the Introduction is probably the place for this though.

Thanks for this comment. Hilton et al (2013) goes into this in a little more detail, and

we added a summary of those arguments to the methods section:

As argued by Hilton et al (2013), an ideal model parameter estimation scheme should permit parameter values to vary at space and time scales matching variations in NEE. Because NEE varies on numerous space and time scales, the space and time scales in which NEE variations are deemed "of interest" (as opposed to "noise") will vary with the modeling goals and spatial domain to be modeled. The range of temporal parameterization windows considered (annual, seasonal, monthly, and ten-day) allow variation consistent with a number of first order NEE drivers: annual climate variation, seasonal ecological cycles, and synoptic weather. Evaluating PFT-specific parameters as well as parameters estimated across many sites helps to evaluate the performance of PFTs as a land surface classification method for NEE diagnosis.

As mentioned, in the methods when I read "plus 27 "cross-validation" sites" I thought - how were these selected? Cross-validation needs to be done right.

We agree enthusiastically that cross validation is crucial and must be performed rigorously. We believe that the even more rigorous five-fold cross validation on 65 sites described by the reviewer is a good idea. That said, we believe that that the cross validation setup described here and the results presented in figure 3 are a useful step toward identifying the most relevant spatio-temporal scales for regional NEE diagnosis via light-use efficiency models.

REFERENCES

Hilton, T. W., Davis, K. J., Keller, K., and Urban, N. M.: Improving North American terrestrial CO_2 flux diagnosis using spatial structure in land surface model residuals, Biogeosciences, 10, 4607-4625, doi:10.5194/bg-10-4607-2013, 2013.