

## ***Interactive comment on “Evaluating terrestrial CO<sub>2</sub> flux diagnoses and uncertainties from a simple land surface model and its residuals” by T. W. Hilton et al.***

**M. Smith (Referee)**

Matthew.Smith@Microsoft.com

Received and published: 30 September 2013

In this study the authors assess the accuracy of a simple light use efficiency ecosystem model for predicting Net and Gross Ecosystem Exchange of Carbon Dioxide (NEE and GEE, respectively) - when constrained to data from 65 North American FluxNet sites using a least sum of squared errors technique. They find, as other studies have implied before, that the scale of spatial and temporal aggregation of eddy covariance sites when conducting the parameter estimation give very similar overall fits in terms of sum of squared errors, but very different spatial patterns of predictions.

Overall I think the results are sufficiently interesting for publication and the research

C5570

has been conducted to a high standard and my critical comments are minor.

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

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

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 re-comparing 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 were these ones chosen in the first place?

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.

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.

C5571

Page 13796 - I have no idea what a "nugget" is. Please make sure it is explained in the text.

---

Interactive comment on Biogeosciences Discuss., 10, 13753, 2013.

C5572