Biogeosciences Discuss., 10, C6330–C6336, 2013
www.biogeosciences-discuss.net/10/C6330/2013/

Biogeosciences Discussions

BGD

10, C6330–C6336, 2013

Interactive Comment

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: 1 November 2013

Author responses to anonymous referee #2)
Referee comments in boldface, author responses in normal typeface.

We thank referee 2 for the careful reading and helpful suggestions. Responses to specific points are below.

Given that I am the 2nd of 2 referees to upload comments I will limit myself to things not covered by R1 (I note that I am in complete agreement with his comments).

This is an excellent paper overall and a worthy contribution to the corpus of upscaling Earth System Earth System FLUXNET literature. After minor technical fixes this is ready for publication. I have a few higher-order comments:

I would prefer a more comprehensive review of upscaling to date. The authors cite and detail a few studies but there are several that are left off. As examples: the Jung Nature paper that upscales ET (not strictly a 1:1 correspondence with C fluxes as here but very much a game changer for FLUXNET-inspired upscaling wrt visibility). Schwalm et al (2010,2011ab) have done FLUXNET-based global upscaling of changes in NEP, GPP, and TER that are solely attributable to hydrological intensification and drought. Yuan et al 2010 detail the derivation of the EC-LUE model. Yang et al 2007 use FLUXNET sites in conjunction with MODIS and SVM to get at GPP. There are others. These are all worthy contributions that have advanced this field. While a summary as per the original is certainly too verbose perhaps a summary sentence (or two) that showcases the depth and breadth of upscaling approaches would be useful.

Thanks. We have added brief discussions of these studies to the introduction.

I am intrigued why RMSE was chosen as the metric? I am curious if you looked at how your results would change if you scored fit differently? Note that I am not suggesting submitting another set of parallel results. But, how sensitive are the conclusions of your study to the skill metric chosen?

This is a very good question and worthy of extensive further investigation. We think a better approach for quantifying model fit would be a statistically proper likelihood function. We chose SSE because the a statistically proper likelihood function would require integrating likelihood functions for all of the sources of error that contribute to model error: model structural error, parameterization error, eddy covariance observation error, etc. These error sources' distributions may be approximated, yielding a likelihood function function for each. Reducing that integral to a computationally tractable form is

BGD

10, C6330–C6336, 2013

Interactive Comment

difficult and beyond the scope of this study. In the absence of a statistically proper likelihood function we chose to use the mathematically simple SSE. This is equivalent to a maximum likelihood approach if the model errors may be assumed to be independent and identically distributed (i.i.d). Model errors are not i.i.d. (Ricciuto et al, 2008), but we have made this simplification in light of the points mentioned above.

We have added the above text to the paper.

As R1 I am intrigued by the 27 vs. 65 split and was very curious as to selection criteria. But R1 has discussed this already and I have nothing to add. But I would emphasize the issue's importance. Consider that you show different maps based on different parameter sets. How about different maps based on different site splits?

This is also an excellent point. To summarize our response to referee 1, the 27 cross-validation sites were initially left out of model parameterization because of data availability problems. It occurred to us later that these were more useful for cross validation than for further model parameterization. We agree with referee 1 and referee 2 that a rigorous cross validation exercise is an important extension of this work, and we are currently working toward that end.

Did you consider downscaling your driver data, e.g., 16-day MODIS data? You use 3h data to drive VPRM and there are canned routines for downscaling (for "imposing" the diurnal cycle).

We agree with reviewer 2 that this would be a worthy line of further inquiry. We did not explore this path during the analyses presented here.

Wrt Eq. [5], this has minimal skill (r2 = c. 0.3). I'm not sure I have a high degree of comfort with any map generated based on this equation. I would like some words on why, given the clear lack of skill of Eq. [5], it has any value wrt uncertainties as discussed in the original.

BGD

10, C6330–C6336, 2013

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

Thanks for this careful consideration of the error variance model. We have added the following two paragraphs to the discussion of eq 5:

The multiple r^2 value of 0.289 achieved by eq 5 may at a glance appear relatively low. However, our ultimate goal in this exercise is a spatial estimate of VPRM NEE uncertainty. In this context it is more important to successfully diagnose the distribution of error magnitudes than to accurately capture every local rise and fall of the error magnitude as a function of its drivers. This is because spatial aggregation of high-resolution VPRM error diagnoses will smooth out the high-resolution inaccuracies without sacrificing the more important regional accuracy. Hilton et al (2013) provides the spatial error covariances needed to perform this aggregation.

In spite of its r^2 of 0.289, eq 5 performed well across 27 cross validation sites on two performance measures: First, 55 of 56 predicted errors (98%) fall within the 95% prediction confidence interval (fig. 13, top panel). Second – and crucially – the distribution of predicted errors matches the distribution of observed errors (fig. 13, bottom panel) at the cross-validation sites. This suggests that the distribution of diagnosed VPRM NEE error magnitudes is consistent with observations.

In addition, by virtue of diagnosing the difference between modeled NEE and observed NEE, the VPRM errors estimated by eq 5 include all error sources that contribute to VPRM error. That said, there are error sources that are not included in the regression model drivers, such as VPRM structural error or land surface classification error, although these are integrated into the final error estimate because eq. 5 seeks to predict the difference between observed and modeled NEE. The 70% of the observed VPRM error variance that eq 5 does not explain is caused by these types of drivers along with random error. As mentioned in section 2.5, in the absence of a statistically rigorous joint likelihood function we feel that the statistical model of equation 5 is a useful first step toward uncertainty quantification.

Use NEP and NOT NEE. NEE is the integrated vertical exchange of CO2.

FLUXNET does not measure this (so you can't upscale this C term either). FLUXNET gets at CO2 exchange as the disequilibrium between GPP and TER only. The processes that are part of NEE (e.g., aquatic evasion, disturbance emission [fLUC or fire flux as examples] and product decay) are not "seen" by FLUXNET at all. Put another way, you cannot compare your NEE values to something that comes from an inversion framework. So I find this misleading, NEP is what FLUXNET does well, not NEE.

We respectfully disagree with referee 2 on this point. All processes occurring beneath an eddy covariance tower including disturbance emissions and aquatic evasion, when present, are integrated into the net flux measured by the tower. Thus an eddy covariance tower observes the NEE from its footprint, not NEP. It is perhaps true that eddy covariance towers are usually sited to exclude some of these influences (e.g. aquatic evasion). That said, several of the towers used in this study (CA-NS2, CA-NS3, CA-NS4, CA-NS5, CA-NS6, CA-NS7) were specifically aimed at quantifying net ecosystem exchange of CO_2 during burn recovery (Goulden et al 2006), though not fire emissions themselves. The upscaled VPRM net fluxes presented here are as representative as the eddy covariance tower sites used to parameterize them allow. For these reasons we believe that the shortcomings of the upscaled VPRM in diagnosing some of the processes mentioned here are better classified as model structural error than by changing NEE to NEP throughout the text.

We also follow the terminology used by past Fluxnet upscaling studies (e.g. Xiao et al (2008), Jung et al (2011), Sun et al (2011)) and define the terminology. We acknowledge that the terminology is important, but think that we are not out of bounds with past literature and logic in saying that these towers observe NEE.

Figures: What does black represent? This is not detailed/explained.

Thank you for pointing this out. We did not produce calculations for these regions. In figures 7, 8, 9, 10, 11, 12 and 14 the black regions are outside of the study domain. In

BGD
10, C6330–C6336, 2013

figure 15 the black regions were not calculated because the figure highlights features in the Southeastern USA. We have added text to captions of figures 7, 8, 9, 10, 11, 12, 14, and 15 to clarify this.

Pg 13770: wrt "Once again, instead of concluding that respiration is causing the mixed forests of the southeastern USA to release on the order of 150 g Cm$-$2 yr$-$1 to the atmosphere, the explanations discussed in Sect. 3.3 seem more plausible." Could you parenthetically include what Sect 3.3 stated? This would help the reader.

Thanks. We have added a parenthetical summary of section 3.3 to the text.

REFERENCES

Goulden, M. L., Winston, G. C., McMillan, A. M. S., Litvak, M. E., Read, E. L., Rocha, A. V. and Rob Elliot, J. (2006), An eddy covariance mesonet to measure the effect of forest age on land–atmosphere exchange. Global Change Biology, 12: 2146–2162. doi: 10.1111/j.1365-2486.2006.01251.x

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.

Jung, M., et al. (2011), Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, J. Geophys. Res., 116, G00J07, doi:10.1029/2010JG001566.

Ricciuto, D., Butler, M., Davis, K., Cook, B., Bakwin, P., Andrews, A., and Teclaw, R.: A Bayesian synthesis inversion of simple respiration and GEP models with eddy covariance data in a northern Wisconsin forest: determining the causes of interannual variability, Agr. Forest Meteorol., 148, 309–327, 2008.

Sun, G., et al. (2011), Upscaling key ecosystem functions across the conterminous United States by a waterâĂŘcentric ecosystem model, J. Geophys. Res., 116,

BGD

10, C6330–C6336, 2013

Interactive Comment

G00J05, doi:10.1029/2010JG001573.

Jingfeng Xiao, Qianlai Zhuang, Dennis D. Baldocchi, Beverly E. Law, Andrew D. Richardson, Jiquan Chen, Ram Oren, Gregory Starr, Asko Noormets, Siyan Ma, Shashi B. Verma, Sonia Wharton, Steven C. Wofsy, Paul V. Bolstad, Sean P. Burns, David R. Cook, Peter S. Curtis, Bert G. Drake, Matthias Falk, Marc L. Fischer, David R. Foster, Lianhong Gu, Julian L. Hadley, David Y. Hollinger, Gabriel G. Katul, Marcy Litvak, Timothy A. Martin, Roser Matamala, Steve McNulty, Tilden P. Meyers, Russell K. Monson, J. William Munger, Walter C. Oechel, Kyaw Tha Paw U, Hans Peter Schmid, Russell L. Scott, Ge Sun, Andrew E. Suyker, Margaret S. Torn, Estimation of net ecosystem carbon exchange for the conterminous United States by combining MODIS and AmeriFlux data, Agricultural and Forest Meteorology, Volume 148, Issue 11, October 2008, Pages 1827-1847, ISSN 0168-1923, http://dx.doi.org/10.1016/j.agrformet.2008.06.015. (http://www.sciencedirect.com/science/article/pii/S0168192308001925)

BGD

10, C6330–C6336, 2013