Biogeosciences Discuss., 9, C2538–C2544, 2012
www.biogeosciences-discuss.net/9/C2538/2012/

Biogeosciences
Discussions

Interactive comment on "Improving terrestrial CO_2 flux diagnosis using spatial structure in land surface model residuals" *by* T. W. Hilton et al.

Anonymous Referee #2

Received and published: 18 July 2012

General comments

Overall, this paper addresses an important and difficult question of how to improve our ability to diagnose and predict regional CO2 fluxes. However, there is not enough effort made to tie the specific analyses performed and results of the study back to the ultimate goals of the science. First and foremost, how does specifying a good land surface model residual covariance matrix help us to improve models like VPRM, or improve flux estimates from kriging maps or atmospheric inversions? What's the marginal benefit of these covariance matrices in light of the ultimate goals? Is improving diagnosis and prediction of CO2 fluxes at regional scales a tractable problem given land surface heterogeneity?!! (You should at least discuss/ mention this.)

In terms of the specific analyses, it is also not clear from the outset what are the hy-

potheses or overall questions that you're trying to address. Is this study primarily an evaluation of VPRM? Or are these results more widely applicable to other land surface models? What is the goal in terms of trying to figure out how many parameters to estimate? (What are we trying to optimize? Is it the flux maps that you mention in the 2nd to last paragraph in the conclusions?) You should also give some theoretical introduction as to how we expect the four estimated parameters in VPRM to vary in space and time. What do plot-level studies and physiological understanding tell us? Why do you assume that the 3 scaling factors (for water, phenology and temperature) don't vary in space and time? How much model complexity (in terms of # of parameters) is needed in order to improve the ultimate goal of improved CO2 flux model prediction or diagnosis? How do we evaluate the benefits and costs associated with parameter lumping/ separation?

You mention multiple times that we expect, a priori, that VPRM parameters should vary by plant functional types (PFT's). You need to back this up with more with ecological understanding and references, rather than assume that the reader is familiar with this argument.

Some of the writing is overly-abstract and hard to follow. One example starts on p3, line 94 with "Michalak et al. (2004) point out that spatial structure, if existent, contains information that constrains fluxes and suggests weights for fluxes to identify and remove redundant information." I've read this paper multiple times, but I still don't understand this sentence! Also, there are a few abrupt transitions where the flow of argument is lost. For example, the paragraph starting at line 35 is quite abrupt, as compared to the previous paragraph. You mention a number of issues associated with poor diagnostic and predictive ability in land surface models. Then, you launch into the data sources available to constrain the problem. Why do carbon cycle scientists struggle with weak performance in land surface models? Is this a tractable problem? Is there reason to believe that more observations will help to solve this problem? How is your approach helping to bring data to bear in order to improve models?

How did you choose the 9 different groupings for parameter estimation (lines 288-290)? Spell out clearly what these are in the text, and your justification for choosing these groupings. Also, talk more generally about the results beyond just the finding that the PFT's don't seem to matter. Does site-level or seasonal (monthly) variability matter at all? Where do you see significant differences? (Isn't there any way to get uncertainties on your ML parameter estimation?)

What do you make of the fact that the 9 different groupings give remarkably different sill variances (Fig. 4)? How important is a reduction in sill variance in light of your ultimate goal (to create flux maps from VPRM, improve inversions, etc.)? What's the trade-off in terms of reducing the sill variance (explaining more of the variability in the residuals) and estimating more parameters?

Did you try any cross-validation, excluding some of the sites from the parameter estimation? For example, you could test which grouping of parameters best minimize the model-data residuals, with a penalty for the number of estimated parameters. There should be an appropriate statistical test for this.

Finally, is the 400km length scale for VPRM residuals model-dependent? Would we expect other land-surface model residuals to have similar length scales, or are the spatial scale of model errors dependent on the model itself? If so, what have you learned about VPRM model formulation?

Specific comments

Make a distinction between running VPRM with grid-scale remote sensing data (wall-to-wall estimation), vs. point-based with site-specific data collected at eddy covariance sites, as in this study. What's the representation error associated with extrapolating from the site-specific to the grid-scale form of estimation? (Does this matter for your analysis or ultimate goal of producing maps?) As a corollary, you should also mention in the introduction that residual correlation length scales are likely scale-dependent. For example, correlation lengths of CO2 fluxes at a 1km resolution are likely to be

shorter than at a 500km resolution. Given the poor diagnostic ability of flux models like VPRM, there is reason to believe that the correlation lengths of model residuals from VPRM are also likely scale-dependent.

Does it matter that the exponential distribution was only selected by AIC in 74 out of 1000 GRF's in the pseudo-data test, and that the median covariance range of these exponential distributions is 936km (almost double the "true" length of 402km)? That seems like a relatively low ability to detect the correct distribution, and estimate the correct correlation lengths. What do you make of the fact that the exponential covariance function was chosen for 92 of 252 of the observed residuals, a much higher proportion than in the pseudo-data test?

How much confidence do you have in the median length scale of 400km, given the large range in your estimates (from 100 to 900km)? In terms of recommending 400km for an estimation scale in inversions, this may still be too large directly around measurement locations, where there is a risk of misinterpreting small wiggles in the data, and attributing local influences to much larger regions.

In the review of inversions and correlation length scales, please also mention geostatistical inversion techniques that don't use priors (on page 3): Michalak et al., 2004; Gourdji et al, 2008, 2008, 2012 & Mueller et al, 2008. These are terrestrial flux papers, which may be more relevant than an ocean inversion (the Jacobson et al., 2007a,b references). In these approaches, where fluxes are estimated at fine scales without prior flux estimates from a land surface model, the flux covariance structure is critical. Also, mention the use of RML for estimating correlation length scales (of fluxes or flux residuals) from atmospheric CO2 data in an inversion framework (Michalak et al, 2005; Gourdji et al., 2010, 2012) in the last paragraph on p. 3. Finally, these papers are relevant for your discussion of the optimal estimation scale for inversions (lines 477 to 484), as they estimate at fine scales and post-aggregate fluxes to larger, more meaningful scales. Concerns about aggregation error should be acknowledged (Kaminski et al., 2001; Engelen et al., 2002; Schuh et al., 2009) as a concern when choosing the appropriate estimation scale for inversions.

As a minor note: how is Pscale different from EVI? Doesn't EVI get at phenology? This should be folded into a few more sentences on the scaling parameters, and how they are derived, and why we don't expect them to vary by site or season.

The paragraph beginning at line 369, introducing the analysis of inter-annual variability, is very confusing! First, you should introduce the concept of inter-annual flux and residual variability, and why we expect this to vary at larger spatial scales than NEE itself. Then, you can move on to how you calculated anomalies from the long-term mean of both NEE (VPRM & observed), and model residuals in order to investigate this question. The ordering in that paragraph currently makes it hard for the reader to follow. Also, is VPRM structured well enough to get at drivers of inter-annual variability (e.g. El Nino/ La Nina, disturbances, extreme climate events, volcanoes, etc.)? If not, is this analysis meaningful at all?

Finally, the outcome of the investigation into length scales of inter-annual variability is unclear (paragraph beginning with line 377). The main result seems to be: "Of the seven years examined, NEEobs anomalies show correlation at scales of roughly 1000 km only for 2006." I would add another phrase, saying that the AIC chose a nugget distribution for all other years, indicating no spatial structure at all, and that this result with the observations is roughly consistent with that from VPRM estimates. The reader shouldn't have to go to Fig. 6 to draw this conclusion for him or herself. Also, in the following paragraph you discuss how the residuals seem to have more spatial structure than NEE estimates themselves. I would add an aside that this occurs mainly with more lumped parameters, e.g. annual, but not monthly, and all sites as compared to individual sites or PFT's.

In the Caveats section (4.1), you mention that VPRM doesn't capture long-term drivers of NEE well, e.g. disturbance and carbon pools. Then, you conclude the paragraph with "These simplifications caution us against attempting detailed ecological interpretation

of the VPRM NEE results and VPRM residuals." Does this invalidate the whole study then?!! If you can't learn anything about nature from your analysis, what have you learned that helps to advance the science and the ultimate goals of the science?

You mention that disturbance history is likely a driver of carbon cycle dynamics and land surface heterogeneity (lines 448 to 458). Can VPRM indirectly capture this at all? For example, this likely affects EVI, and perhaps the estimated parameters or the scaling factors, which might vary by stand age?

The last sentence in the conclusions should tie this paper and analysis back to the ultimate goal of the science. E.g. "With an accurate land surface model residual co-variance matrix, CO2 flux estimates from inversions can be improved, for the ultimate goal of diagnosing regional CO2 fluxes and improving predictive skill in land surface models…". This is kind of rough, but for someone who only reads the introduction and conclusions, they should come away with why your paper is useful, and how it fits into the larger goals of the science.

Technical corrections

* In the abstract, you say that spatial structure exists in data-model residuals at a length scale of $\sim$1000km. The body of the manuscript says 400km. Correct or clarify.

* Think of a better way to rephrase the last sentence in the abstract. What do you mean by "carbon cycle participation"? Maybe say that VPRM model parameters do not differentiate by plant functional type?

* First sentence in introduction: update emissions from fossil fuel combustion, which are now around 9 PgC/yr in 2010 (http://www.globalcarbonproject.org/carbonbudget/10/hl-compact.htm).

* Lines 141 to 143: I assume that temperature and PAR are collected at the flux tower sites? Don't these count as meteorological driver data? (Clarify the statement "It can thus be run globally, with no need to compile temporally-filled meteorological driver

data.")

* Lines 163-167: Please explain why site phenology, land surface water, and land surface cover type are needed in terms of the VPRM model formulation (eqs. 1 & 2). In general, after introducing the equations for VPRM (after line 137), specify which are the 4 estimated parameters, and how you get the 3 scaling factors.

* Line 381: shouldn't this be 74 of 1000 attempts?

* Add a reference to the statement on lines 445-446, starting with "Second, PFTs are commonly assumed. . .".

Interactive comment on Biogeosciences Discuss., 9, 7073, 2012.