Biogeosciences Discuss., 9, C8900–C8904, 2013
www.biogeosciences-discuss.net/9/C8900/2013/

Biogeosciences
Discussions

Interactive comment on "Constraints from atmospheric CO_2 and satellite-based vegetation activity observations on current land carbon cycle trends" *by* D. Dalmonech and S. Zaehle

Anonymous Referee #1

Received and published: 8 March 2013

Title: 'Constraints from atmospheric CO2 and satellite-based vegetation activity observations on current land carbon cycle trends'

Authors: D. Dalmonech and S. Zaehle

The major scheme of this study is to improve land surface models through comprehensive comparisons with different sets of observations. The paper describes an extension to existing benchmarking systems towards a more objective evaluation scheme. Since the global carbon cycle represents a first order uncertainty in future projections of global climate, studies like this that evaluate the skill of land surface models in reproducing the observations in a rigorous way are clearly very important.

Overall, the study is thoughtfully executed, and in my humble opinion does provide a valuable contribution. The framework could indeed be the first stepping stone towards a uniform model data evaluation scheme enabling the testing of many models, offline and online, in a time efficient manner such that model developers can quickly improve deficiencies. I, thus feel that, after some substantial changes that emphasize more the key contributions and make the paper more transparent (outlined below), it could be suitable for publication in BG.

Major comments:

(1) The title is a bit misleading, and reads more like a 'data-model assimilation scheme'. I suggest changing it such that is more consistent with the key objectives of the paper (e.g. Towards a more objective evaluation. . ..).

(2) In the abstract it states that ' . . .The selection of observational characteristics (traits) specifically considers the robustness of information given the uncertainties in both data and evaluation scheme. . .'. When I first red this I became very excited, because this would indeed be a significant improvement. But after reviewing, I didn't really see how uncertainties were explicitly treated. Just not using the observations when they are not as robust (e.g. optical satellite data over the tropics) does not have much merit in my opinion. One could rank the observations in regards to robustness. For example, satellite vegetation data are in general most robust at seasonal time scales, and least robust at quasidecadal time scales (Tucker et al. 2005; ref. given in ms). There is a related discussion about this point in the Introduction on Page 16090 (lines 1-8) as well, but the stated 'philosophy' based on a null-model and how this deals (or does not deal) with uncertainties in observations and evaluation scheme are not clear to me. Since this is a key point of the paper, the specific approach and thinking behind it should be made as transparent as possible to increase the impact of the study.

(3) Since this study uses only a small subset of available observations (e.g. see Randerson et al. 2009), it appears that the key contribution to the existing state of the art in

model data comparisons is the extension towards a more objective evaluation scheme including quantitative model performance (ranking) measures. I, therefore, suggest emphasizing this portion of the study much more through restructuring the paper. The following comment (main comment (4)) is also related to this.

(4) Figure 2 captures a big portion of the key results. But after reviewing the paper I was unable to understand how the global scores were really derived. I suggest to make this point much more transparent.

(5) Section 3.5.3. I wonder how meaningful any model data comparisons are that involve the atmospheric CO2 growth rate (e.g. Fig. 10). This metric integrates numerous carbon sink processes at various spatial and temporal scales. This also points to a general problem in model-data intercomparisons in regards to using observations that are difficult to interpret. I would refrain from using such 'traits' as it is not clear what understanding could be gained in such comparisons. If the model does do well in reproducing observations that are relatively straightforward to interpret, then they could be useful to understand complex observations.

Minor comments:

(1) Page 16096, lines (13-21). Using observations in the most meaningful way. The CO2 record derived from observing stations is a complicated signal as it contains a long-term trend (from ff burning emissions) and a superimposed seasonal cycle due to plant activity. Simply doing a trend analysis on monthly CO2 data, as stated is not a meaningful metric, and one has to extract the seasonal cycle in a prior step (see Keeling et al. 1996; ref. given in ms). Extraction of the seasonal cycle, however, relies on knowledge about the underlying trend (which is not known), and for that reason the amplitude of the seasonal cycle is considered the most robust signal (and hence used by several authors as a carbon cycle metric) followed by the downward and upward shifts. Unless the authors can convincingly show that their seasonal cycle extraction method does account for the complexity in these data, I would not have any trust in the

observed monthly CO2 trends (MT) and thus also not in the corresponding data model comparison.

(2) Page 16096, line (9). '....in both standard and modeled fluxes...'. What are the standard fluxes?

(3) Page 16097, lines (4-6). 'A direct comparison of ...'. I would highly disagree with this sentence. We have now observations, including satellite-derived products (e.g. MODIS GPP products, or GPP from upscaled FluxNet which could be used to evaluate magnitudes of simulated GPP. In fact, the absolute magnitude is as important as the phase to get the carbon sinks right (which is also discussed later in the paper when modeled and observed atmospheric CO2 are compared). I suggest changing this sentence.

(4) Page 16097, lines (10-26). How did the authors calculate the 't-onset' dates from the fAPAR records? The explanation given here is entirely non-transparent.

(5) Page 16099, lines (17-18). Why didn't the authors also exclude tropical Africa? Are there viewer clouds over this portion of the tropics?

(6) Page 16103, lines (12-13). The text here states that data for Pt. Barrow are shown in the corresponding Fig. 5, but in Fig. 5 caption it seems it is the Alert observing station. For this comparison shown in Fig. 5, see also my minor comment (1).

(7) Page 16104, lines (1-2). Not sure about the logic here. Since the mean seasonal cycle comparison showed good performance of JSBACH, the mismatch in Fig. 5b may not be as much due to an 'asynchrony of photosynthesis and respiration' but more related to divergence in observed and modeled climate sensitivities of photosynthesis and respiration. Here, I may also add that reproducing interannual or longer-term variability is a much stringer test than corresponding comparisons at seasonal time scales.

(8) Page 16106, lines (11-14). In relations between spring phenology and land sur-

face warming, why did the authors use annual temperatures? I would expect spring temperatures, or at least cold season temperatures, would be more suitable here.

Interactive comment on Biogeosciences Discuss., 9, 16087, 2012.