Biogeosciences Discuss.,
doi:10.5194/bg-2016-557-RC1, 2017

Biogeosciences
Discussions
Open Access

Interactive comment on "Reviews and syntheses: Systematic Earth observations for use in terrestrial carbon cycle data assimilation systems" *by* Marko Scholze et al.

N. MacBean (Referee)

nlmacbean@gmail.com

Received and published: 16 February 2017

Review of Scholze et al. "Reviews and syntheses: Systematic Earth observations for use in terrestrial carbon cycle data assimilation systems"

Scholze et al. present a rigorous, in-depth review of the major observation types that are used to constrain the carbon cycle and related variables in terrestrial biosphere models. Such a synthesis is timely given the increasing availability of a variety of observations, as well as the pressing need to reduce uncertainty in model simulations of the current and future carbon budget. The strength of the paper lies in the extensive description of the main types of EO observations that can be used in a carbon cycle data assimilation (DA) system. As such I would strongly recommend this paper to any

colleagues wishing to use these data for model optimization (and benchmarking). I have made some comments below that I think would further improve the structure and clarity of the manuscript, as well as suggesting some extra references – although I appreciate the authors are not aiming to provide an exhaustive review.

General comments

Firstly, I appreciate the distinction the authors try to make between their review and that of Raupach et al. (2005) and Ciais et al. (2014) by placing an emphasis on EO data versus in situ atmospheric CO2 and eddy covariance data; however, these two sources of data are one of the most widely used in carbon cycle DA studies, and therefore I think it is worth having a separate section that briefly summarizes these data and their uncertainties, while keeping the focus on EO. Otherwise the description of updates to the eddy covariance uncertainty estimates in the general section on observational uncertainties (Section 3.1) could feel out of place. In addition, Section 3.2 discusses operational carbon observing systems which currently include many in situ networks.

Secondly, I suggest a slight re-structuring so all the examples of DA studies with these data are incorporated into one specific section, and possibly after the description of the different types of observations. Currently, there are examples in Section 2.3 and the introduction to 3.3. Whilst the examples given in the latter are specifically pertinent to EO data, the use of EO data in an assimilation system has been discussed already in Section 2.3, and therefore the lines are somewhat blurred.

Finally, it would be good to include websites/references for data access in all data tables (as in Table 1), and, given the important emphasis on observation uncertainty, note if uncertainty estimates come with the data.

Specific comments

Section 2.1

Lines 109-111: worth pointing out that a better fit between the posterior maximum likelihood simulation and the observations does not necessarily mean you have the correct parameters and/or model structure (e.g. MacBean et al., 2016)

Section 2.2: The distinction between sequential and variational DA could be slightly confusing for the lay reader. I suggest the following:

Line 133: make it clear that sequential assimilation happens at the point of having an observation – otherwise one may wonder "at which discrete time steps?"

Lines 137-139: I think this could read as if J is only evaluated in the variational approach (though that may be helped by changing the caption of Figure 1 – see below). I suggest that instead of just discussing the inner loop you could make a distinction about when J is evaluated and at what point the minimum is found for both approaches. In addition, it might be helpful to the reader to have a sentence that qualitatively describes what the cost function represents and to explicitly say that the aim is to minimize the cost function around lines ∼132-139.

Figure 1: I like this figure, but I cannot see a "Model-data comparison" box as you describe in the caption. I guess you mean "Evaluation of J"?

Section 2.3: Line 195: Maybe add paper by Bloom and Williams (2015) and latest CLM paper by Post et al. (2017)? Line 200: maybe worth adding "…same cost function value at the minimum"? Line 203: Could add Thum et al. (2017) here

Section 3.1: Worth mentioning that observation errors in a DA system should include the models errors, and what could give rise to errors in the model?

Section 3.2: This section is very focused on Europe. It would be worthwhile detailing efforts that are underway in other regions, e.g. example the NASA Carbon Monitoring System (http://carbon.nasa.gov). This section also feels a little out of place. I might suggest incorporating it into the introduction to Section 3 or having it as a perspectives section at the end of the article.

Section 3.3: Lines 306-308: worth pointing out that using level 2 products may in-

crease the observation uncertainty, particularly given parameters/processes implemented in retrieval algorithm may not be consistent with corresponding equivalent parameters/processes in the underlying model (and that this may be a benefit of using level 1 products – e.g. Quaife et al., 2008). Also perhaps worth explaining that for vegetation activity that VIs are an intermediate step in that they are "lower order" products – i.e. they are raw radiances but also do not require a complex retrieval algorithm; instead they require an atmospheric transport model and limited calculations

Line 316-318: worth including that NDVI has also been used (e.g. MacBean et al., 2015a), and the advantages/disadvantages of using VIs

Line 318: Forkel et al. (2015) is another example of the use of FAPAR with a terrestrial model.

Line 321: and by optimizing parameters related to phenology and photosynthesis (MacBean et al., 2015b)

Line 322: Saying "Also assimilation of XCO2" comes a bit out of the blue here as you have just been talking about vegetation activity. Please could you say what is meant by XCO2, or refer to section 3.3.1.

Lines 325-332: Other examples of the impact of soil moisture (and LAI and FAPAR) data assimilation on LAI and C fluxes include the work at CNRM with the ISBA-A-gs model, e.g. Barbu et al. (2014).

Line 334: several studies have demonstrated the added benefit of aboveground biomass, including articles already cited (Richardson et al., 2010; Williams et al., 2005; Keenan et al., 2012). Might be worth listing a few examples, or, combining this section with aforementioned examples of C cycle related DA studies (section 2.3).

Line 340: LAI has been used in C cycle DA (see Barbu et al., 2014). Further to my comment on VIs above, perhaps it would be worth explaining somewhere in the text the differences between using VIs, FAPAR and LAI, why one would use one vs another?

Line 341: Worth mentioning the dataset of Li et al. (2011) that has been used in several studies investigating trends in biomass. In fact, I expect that VOD data will be increasingly widely used for optimizing biomass in terrestrial biosphere models, and therefore I would suggest adding a discussion of what these data actually represent in Section 3.3.5 (i.e. how reliably can you estimate biomass (leaf or total aboveground?) from what is essentially a measure of water content).

Line 343: I think that LST might be used in a similar manner to soil moisture in DA in the future, and not just as an input/boundary condition. Therefore perhaps it can be included with VOD in this context?

Section 3.3.2 Lines 477: you mean significant difference in the absolute magnitude between the products (as the temporal and spatial patterns are quite consistent, as you state)? This was also a conclusion drawn by D'Odorico et al. (2014) and, to some extent, Tao et al. (2015); therefore, it is worth mentioning that these studies agree on this point.

As mentioned above, here or elsewhere I think it would be beneficial to have a discussion of the use of VIs and LAI. Arguably LAI is the variable that is most closely linked to standard terrestrial model state variables, therefore the reader should understand why one might use any of these three options for optimizing vegetation dynamics/activity, and the advantages/disadvantages of each. For example, if a modeler is mostly concerned with optimizing the overall magnitude of vegetation activity, careful choice of which FAPAR product to use is important (likely the same for LAI). If they are more concerned with temporal dynamics, one could argue that using a normalized lower order product (e.g. NDVI) that does not require such a sophisticated retrieval algorithm might be more appropriate. Perhaps you do not agree! But in any case, a discussion would be useful here.

At the end of this section there is a particular focus on the JRC-TIP FAPAR product as opposed to one of the others, MODIS for example. It would be good to explain

the reason for this choice, or to see more information on some of the other commonly available products.

Lines 484-485: Please could you be clearer how the products in this sentence link to Table 2? JRC MGVI is not described in Table 2 for example.

Section 3.3.4 Lines 489-491: Although I appreciate you do not wish to provide an exhaustive description of retrieval algorithms, I think it would be helpful to qualitatively describe the difference between passive and active retrieval algorithms in one or two sentences here, as well as the fact different algorithms may produce either volumetric water content (absolute values) vs relative soil moisture values.

I would be interested to see a discussion of GRACE land water content in this section.

Section 3.3.5 Line 670: do not need to reiterate what an active sensor is here.

Line 675: I see you do refer to the VOD product of Liu et al. here. Still, I think it would be beneficial to detail that this is based on VOD data and describe briefly how VOD are derived (following on from the mention of VOD in the soil moisture section) and how biomass is estimated from VOD and their expected use/value for optimizing biomass (as discussed above), as well as for better understanding discrepancies in other sources of biomass data that you discuss towards the end of Section 3.3.5.

Lines 676-684: updated reference: Santoro et al. (2015) – update to aforementioned papers providing biomass estimates across a wider range of biomes in the northern hemisphere.

Line 711: Could you provide the biomass limit that the P-band BIOMASS mission will be able to resolve (to compare with the NISAR mission)?

Given you mention the international soil moisture network in Section 3.3.4, it may be worth mentioning the international tree ring data bank in this section (https://data.noaa.gov/dataset/international-tree-ring-data-bank-itrdb), as these data represent a promising new direction for optimizing biomass across a range of biomes.

Minor comments and typos Line 252: maybe “between” better than “among”? Line 309-310: sentence could be simplified Line 111: benchmark Line 135: measurement Line 145: knowledge Line 234 and 241: related Line 255: diagonal Line 311: terrestrial Line 315: biogeochemical Line 420: that than Line 440: reflectance Line 576: complementarily Line 727: satellite Line 1395: Updated Thum et al. (2016) reference – see below.

References:

Barbu, A. L., Calvet, J.-C., Mahfouf, J.-F., and Lafont, S.: Integrating ASCAT surface soil moisture and GEOV1 leaf area index into the SURFEX modelling platform: a land data assimilation application over France, Hydrol. Earth Syst. Sci., 18, 173-192, doi:10.5194/hess-18-173-2014, 2014.

Bloom, A. A. and Williams, M.: Constraining ecosystem carbon dynamics in a data-limited world: integrating ecological "common sense" in a model–data fusion framework, Biogeosciences, 12, 1299-1315, doi:10.5194/bg-12-1299-2015, 2015.

Forkel, M., Carvalhais, N., Schaphoff, S., v. Bloh, W., Migliavacca, M., Thurner, M., and Thonicke, K.: Identifying environmental controls on vegetation greenness phenology through model–data integration, Biogeosciences, 11, 7025-7050, doi:10.5194/bg-11-7025-2014, 2014.

Liu, Y. Y., R. A. M. de Jeu, M. F. McCabe, J. P. Evans, and A. I. J. M. van DijkÂă(2011), Global long-term passive microwave satellite-based retrievals of vegetation optical depth,ÂăGeophys. Res. Lett., 38, L18402, doi:10.1029/2011GL048684.

MacBean, N., Maignan, F., Peylin, P., Bacour, C., Bréon, F.-M., and Ciais, P.: Using satellite data to improve the leaf phenology of a global terrestrial biosphere model, Biogeosciences, 12, 7185-7208, doi:10.5194/bg-12-7185-2015, 2015a.

MacBean, N., F Maignan, P Lewis, L Guanter, P Koehler, C Bacour, P Peylin, J Gomez-Dans, M Disney, F Chevallier, A Model-Data Fusion Approach for Constrain-

ing Modeled GPP at Global Scales Using GOME2 SIF Data, Abstract #B43H-0649 presented at 2015, Fall Meeting, AGU, San Francisco, CA, 14-18 December, 2015b http://adsabs.harvard.edu/abs/2015AGUFM.B43H0649M

Post, H., J. A. Vrugt, A. Fox, H. Vereecken, and H.-J. H. Franssen (2017), Estimation of Community Land Model parameters for an improved assessment of net carbon fluxes at European sites, J. Geophys. Res. Biogeosci., 122, doi:10.1002/2015JG003297.

Quaife, T., Lewis, P., De Kauwe, M., Williams, M., Law, B.E., Disney, M. and Bowyer P. (2008) Assimilating canopy reflectance data into an ecosystem model with an Ensemble Kalman Filter. Remote Sensing of Environment, 112, 1347-1367

Santoro M, Beaudoin A, Beer C, Cartus O, Fransson JE, Hall RJ et al. (2015).ÂăForest growing stock volume of the northern hemisphere: Spatially explicit estimates for 2010 derived from Envisat ASAR. Remote Sensing of Environment, 168, 316-334.

Thum, T., N. MacBean, P. Peylin, C. Bacour, D. Santaren, B. Longdoz D. Loustau and P. Ciais (2017) The potential benefit of using forest biomass data in addition to carbon and water flux measurements to constrain ecosystem model parameters: case studies at two temperate forest sites, Agricultural and Forest Meteorology, 234, 48-65.

Interactive comment on Biogeosciences Discuss., doi:10.5194/bg-2016-557, 2017.