Comment: Assimilating solar-induced chlorophyll fluorescence into the terrestrial biosphere model BETHY-SCOPE: Model description and information content

Authors: Alexander Norton et al.

Summary: This paper uses satellite observations of Solar-Induced Fluorescence (SIF) in an inversion scheme (CCDAS) to reduce uncertainty in *a posteriori* estimates of model parameters and outputs, specifically GPP. Interestingly, no attention is given to actual parameter values or GPP estimates; the focus is entirely on how much reduction in uncertainty can be expected due to the inclusion of SIF.

The paper is reasonably well written, and uses a novel approach to attempt to reduce uncertainty in *a posteriori* estimates of model parameters and output. However, I feel that the paper needs clarification and perhaps some reorganization to help readers to follow the story. Furthermore, I believe that the critical issue of observational uncertainty is given too little attention and must be clarified.

The authors provide reasonably comprehensive citations for CCDAS, but the paper reads is if it were written (as it probably was) by someone who is a Data Assimilation (DA) expert. To this reviewer it seems that some details are either implied or 'skipped over'. It is likely that many readers will be DA experts themselves, but the inclusion of SIF will probably draw in readership that may not possess the DA expertise to easily understand what is going on. I may be a member of that part of the audience, so some clarification is warranted. Specifically, the relationship between covariance matrices (C_x, C_d) and standard deviation (σ) is not entirely clear.

The description of grids used and observation area ("GOSAT grid cell"; section 2.4) needs clarification. Two grid sizes are mentioned in Section 2.4, but we don't learn much more about them until Section 2.5. I would like to see a more deliberate explanation of "here is what we are going to do, and here is how we are going to do it". That might fit better in Section 2.1. Some specific Issues:

- Figure 3, showing observational uncertainty, is not referred to in the section describing observational uncertainty. It needs to be.

Observational uncertainty, Eqn 4: I see two ways that this value can be small: 1) there are many observations, and σ^2 is small. 2) There are very few observations, and Area is small. Parazoo et al. (2013) estimated uncertainty as the standard error. This has the effect of allowing a large error in regions with very few observations, like the tropics. Figure 3 in the manuscript under review shows some of the smallest observational uncertainty in the tropics, and that makes absolutely no sense to me. I've worked with the GOSAT data, and over the deepest tropics there are very few observations, which makes me suspicious that your uncertainty is small because of reason 2). Parazoo et al. did not extend their analysis to the wetter parts of Amazonia because they just didn't have enough data to justify it. Now the authors claim that this region has some of the smallest observational uncertainty on the globe! A detailed justification of how uncertainty can be very small over a region with few or no datapoints is an absolute necessity.

I do not think multiplying by square-root-2 is sufficient to remedy what might be unrealistically low uncertainty values.

When GOSAT 2010 data is aggregated onto the 1.25x1.0 degree MERRA grid, I see that the maximum number of retrievals for a given month, anywhere on the globe, is between 30-35 or so. Looking at South America, I see that very few MERRA gridcells have more than 10 retrievals in a given month during 2010, and many gridcells have 5 or fewer. Aggregating up to 7x10 (or 2x2) you are not going to get very much increase in sample size. I'd like to see the authors address the sparseness of the GOSAT data and explain how this will or will not effect their method.

The number of GOSAT observations is invariant and does not change with grid size. The aggregation of GOSAT observations changes with grid size (Section 2.4). This should be clarified.

An individual GOSAT retrieval has pixel size of around 10 km^2, I believe. OCO-2 will have a pixel size of ~5 km^2, and GOME-2 is a 40-80 pixel, or 3200 km^2. This will have a large impact on your inversion scheme and the calculation of observational uncertainty. Since this paper only uses GOSAT, the other products probably don't need too much (or any?) explanation, but I do have questions about GOSAT and the grids used:

1. There is the possibility for (possibly) many 10km^2 GOSAT retrievals to be included in a 7.5x10 degree gridcell. For that matter there can be many of them in a 2x2 gridcell too. BETHY-SCOPE tiles in 3 PFTs; how are GOSAT retrievals registered to these PFTs? Are GOSAT retrievals marked with a specific land cover type, and accumulated on a per-PFT basis? What about GOSAT retrievals that are not associated with one of the 3 PFTs tiled into the BETHY-SCOPE gridcell? Are they discarded? Why or why not?
2. If all GOSAT retrievals within a gridcell are utilized, is the mean taken and used for DA with all 3 PFTs? In this case aren't you 'smearing out' the information that SIF provides? Guanter et al. (2012) demonstrate that the linear relationship between SIF and individual PFTs is heterogeneous. Do you take this into account? If so, how? If not, why not?
3. In August 2010 the GOSAT scan strategy was changed; the area observed was decreased, but the number of retrievals over a given region was increased. How does this effect the two questions above?

The reduction in uncertainty for global GPP is dramatic (79%). However, this reduction is critically dependent upon C_d (observation uncertainty) according to equation 1. Therefore, I think it is absolutely essential that the questions surrounding the determination of this observation uncertainty are answered in a clear and categorical manner.

I'm not a DA expert, but I do collaborate with quite a few people who are, and I think I understand the basics. The covariance matrices are absolutely fundamental to the outcomes of a DA experiment: If the observational uncertainty is small and the model uncertainty large, the *a posteriori* outcome can be pulled strongly towards the observations. If the opposite is true, then it will be hard to budge the inversion away from the model prior. Is this correct?

In this paper the first case is presented: the observational uncertainty is, to my eye extremely small and therefore results in an amazing reduction in uncertainty in the *a posteriori* result.

The absence of evaluation of actual posterior values of either parameter or flux values may actually hinder the analysis. If the result of the study is an outlandish value for global GPP, then that might indicate a problem. Of course, estimates of global GPP vary by about a factor of two (Huntzinger et al., 2012), so maybe this wouldn't help as much as one might hope. However, posterior parameter and flux values might offer insight, and a comprehensive evaluation of method and results (values of parameters and flux) could provide more support for the authors' conclusions. Was this considered? Why or why not? I'm suspicious that posterior flux and parameter values *were* outlandish, and a choice was made to focus on method even though results may be untrustworthy. I suspect many readers will have this suspicion too.

A detailed description of the construction of the observation uncertainty may detract from the paper's readability, but including it in an appendix would be appropriate. Additionally, I would like to see, perhaps in supplemental material, a step-by-step description of the calculation of the observation uncertainty, perhaps in the 7x10 gridcell that contains Manaus, Brazil.

To see such a large reduction in error sent warning bells ringing with me; I don't think it is an overstatement to say that the entire paper depends on the observation uncertainty. If the authors can demonstrate that the values shown in Figure 3 are justifiable, then the paper has merit. If not, I think the whole endeavor falls apart, as the structural underpinning would have disintegrated. In that case the paper is not worthy of publication.

Specific Comments:

- Figure 2: The information here is too dense (small labels, tiny resolution on the plot) to follow. If the only pertinent information is in the lower-right-hand of the plot, why not omit the rest and enlarge this sector of the graph?

- Figure 2: There is very little description of the graph and what it means. Again, this may be another case where the authors are assuming that their readers look at graphs like this every day and know what it is showing.
- Figure 3: What are the units?
- Figure 4: Absolute uncertainty annual GPP will of course correlate directly with productivity. If you standardize the time series and look at relative uncertainty I imagine that map will look very different. Have you done this? If you have, do Figures 4-6 look similar or different?
- Table 1A: There is no description of what these parameters are and what they do. There are sporadic mentions in the text, but for the most part the reader is left to one's self to figure out what these parameters are for. I would like to see a column added (there appears to be room, as the uncertainty reduction columns could be re-formatted) with a couple of words or a phrase describing each variable. Section 3.2: line 14 on page 11 mentions that τ_W makes up 82% of the global annual uncertainty in posterior global GPP. The reader does not know what τ_W is. At the end of Section 3.1 there are several other parameters listed, and again the reader is not told what they are. It might be helpful to have a short description in parentheses following the listing of each parameter, but I would prefer to see that information in table 1A.
- Boilley and Wald (2015) discuss a high bias in the radiation from reanalyses. I'm not sure this is the same as the uncertainty mentioned in sections 2.4 and 3.4. Can you elaborate?
- Page 17, lines 7-8: "…we can predict and quantify how SIF wil constrain the uncertainty of process parameters and GPP, but we cannot predict how their values will change". Why not? Can't you back the posterior values out of the *a posteriori* covariance matrices and the Jacobian? Isn't the whole point of DA to obtain these posterior values?

References

Boilly and Wald, 2015: Comparison between meteorological re-analyses from ERA-Interim and MERRA and measurements of daily solar irradiation at surface. Renewable Energy, 75, 135-143.

Guanter et al 2012: Retrieval and global assessment of terrestrial chlorophyll fluorescence from GOSAT space measurements. Remote Sensing of Environment, 121, 236-251.

Parazoo, N.C., K. Bowman, C. Frankenberg, J.-E. Lee, J.B. Fisher, J. Worden, D.B.A. Jones, J. Berry, G.J. Collatz, I.T. Baker, M. Jung, J. Liu, G. Osterman, C. O'Dell, A. Butz, S. Guerlet, Y. Yoshida, H. Chen, C. Gerbig, 2013: Interpreting seasonal changes in the carbon balance of southern Amazonia using measurements of XCO2 and chlorophyll fluorescence from GOSAT. Geophys. Res. Let., 40, 2829-2833, doi:10.1002/grl.50452.