1 Author General Response

1.1 Observational Uncertainty Calculations

As pointed out by both reviewers the calculation of the observational uncertainty requires clarification. To address these recurring comments we have done the following:

- Re-written the section in the methods on the calculation of observational uncertainties. We have gone through the calculation and justified it step by step to help readers follow what is being done and why.
- Provided a simplified equation in that section that approximates the (seemingly confusing) area-weighted uncertainty.
- Attached an additional section in supplementary material giving further details of this calculation and the exact formula. As part of this we include an example calculation for a grid cell over the Amazon with accompanying figures showing the original data and the final calculated uncertainty (for the whole 12 months).

We also clarify here. In calculating the observational uncertainties we make the assumption that the observations are independent, i.e. have uncorrelated errors. This is the same assumption made in Parazoo et al. (2013,2014).

This means, effectively, that with the aggregation of GOSAT grid cells into a larger region (i.e. the course model grid cells) there is a larger number of observations therefore the uncertainty goes down by the $1/\sqrt{n}$ law (the same occurs when calculating the standard error). This is a well-known occurrence in dealing with satellite observations and it can be surprising to see the effect of going from single sounding precision (relatively large uncertainty) to aggregated regions (relatively low uncertainty). Another way to describe this is that if you aggregate a region you're taking many independent observations (from each sub-region) and getting out just one independent observation, so to preserve the information content of those sub-regions independent observations the uncertainty goes down; this is called the Jacobian rule of probabilities.

Characterizing correlations in errors is a known problem with satellite measurements. For SIF correlated errors may be due to, for example, error in the retrieval zero-level offset. We are currently looking into the effect of the zero-level offset and will add a additional sensitivity test in the results and discussion accounting for this. If measurements have correlated errors the information content is less than without. To be on the more conservative side we scale our uncertainties by $\sqrt{2}$ which increases the uncertainty.

One reviewer also noted that the observational uncertainties over the tropics (and in particular the Amazon) in Figure 3 appear much smaller than expected. We recognise that this needs explaining. Amendments have been made to the methods section clarifying this, but we also clarify here. Again, the two main points above are relevant. Another element of the small uncertainty over the tropics in Figure 3 is that this is an "annual" uncertainty, so this accounts for the fact that during parts of the year the high-latitudes have no data, while the tropics almost always have data, therefore the tropics have more observations which leads to lower uncertainty.

1.1.1 Inclusion of Structural Uncertainties

This point relates to the calculation of the covariance matrix C_d. Formally, this is the uncertainty covariance matrix representing observational and model uncertainty. We agree that we must specify this in the methods and have thus changed it.

There are two general types of structural uncertainties.

- First, is a structural uncertainty in the model (i.e. model structural error). This may be due to incomplete process formulation in the model equations. One can address this error by looking at statistics in the model-observation mismatch following an assimilation of the data (Kuppel et al., 2013). This is therefore only feasible following an assimilation of the data to estimate posterior SIF, posterior parameters, and posterior GPP. In the present study, we are only interested in error propagation so we do not perform an assimilation of the data.
- Second, is a structural uncertainty in the observations. This may be due to certain unknown errors in space and/or time due to (for example) systematic errors in the instrument or retrieval algorithm. One example of this for SIF is an error in the zero-level offset (Frankenberg et al., 2011;2014).

We address this issue by conducting a sensitivity test. We introduce a structural uncertainty into the error propagation system to assess the effect on the calculated posterior uncertainties. We incorporate this sensitivity test into the results and discussion to approximate the effect this extra uncertainty may produce on uncertainty in GPP.

2 Anonymous Referee # 1

Author comments are shown in blue.

2.1 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 (Cx, Cd) and standard deviation () is not entirely clear. Good point. We want readers from different audiences to be able to follow what was done easily. We have added and clarified text in the methods section to help non-DA readers relate covariance matrices to standard deviation uncertainty simplified other points where possible. We have also modified the last paragraph of the introduction to make it clearer what the specific aims are.

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. Good point. We have amended this as suggested by shifting the grid resolution information to the beginning of section 2. 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. Amended.

2.2 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. Please refer to general response section above.

I do not think multiplying by square-root-2 is sufficient to remedy what might be unrealistically low uncertainty values. Please refer to general response section above.

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. Id like to see the authors address the sparseness of the GOSAT data and explain how this will or will not effect their method. In the amended manuscript we show an example calculation over a 7.5x10 degree grid cell including the GOSAT sub-grid cells to show how this scales across 2010. We see that aggregating from 3x3 to 7.5x10 you get actually see a big increase in sample size. For example in Jan 2010 any GOSAT sub-grid cell may have between 0-20 soundings, but aggregating to the 7.5x10 there is almost 80.

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. In fact, the number of GOSAT observations does vary with grid size. With a larger grid size you capture more GOSAT soundings. You may refer to the general response section for further details. We have clarified this in the methods section.

An individual GOSAT retrieval has pixel size of around 10 km^2, I believe. OCO-2 will have a pixel size of $\sim$ 5 km^2, and GOME-2 is a 40-80 pixel, or 3200 km2. 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 dont 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? This is a good point. The observations are not separated per PFT, doing so would effectively triple the information content as there would be three times more observations, which would in fact improve the results. We compare observations at the grid cell scale. Thus information is transferred/split to PFTs through the

Jacobian sensitivities, which account for PFT fractions. E.g. if a grid cell is 90% C3Gr, then the SIF sensitivity over that grid cell will be dominated by parameters relating to C3Gr, with smaller contributions from the PFTs that make up the remaining 10%. Thus, the information content of the observations is split accordingly. 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? We do not attempt to disaggregate observations in this way. We assume there is roughly even coverage across the PFTs, even though the absolute footprint of a GOSAT sounding is about $10 km^2$, it has a wide swath of around $750\ km^2$ with 5 footprints. Thus we assume decent coverage. This will be more important to consider in a full assimilation of the data i.e. for estimating parameter values and fluxes.

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 arent 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? This is true for a full assimilation and parameter estimation but in this study, we do not consider the mean values of the observations, only their uncertainties as we're only interested in information content. Thus these issues are not present.

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? Yes, good point. The observational uncertainties used in section 2.4 are standard errors (although slightly adjusted to increase the uncertainty as described in section 2.4), thus they account for the number of observations per grid cell.

The reduction in uncertainty for global GPP is dramatic (79%). However, this reduction is critically dependent upon Cd (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. Agreed. We have clarified our calculation of the observational uncertainties in the manuscript. Please refer to general response above.

Im 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? Essentially, yes this is true. However, we note that this question primarily applies to a an assimilation with real data. In this paper we assess the information content of SIF observations, i.e. only uncertainties of model parameters and GPP, not their values. We can do this because this is a linear problem, whereas the full assimilation is a non-linear problem and the subject of subsequent study. The point regarding observational uncertainty vs model uncertainty is pertinent however, and we address this in the general response section.

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 wouldnt 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? Im 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. Assessing information content of the observations is a linear problem which can be performed independently of comparing actual values of model and observed data. This is convenient as an assessment of the information content tells us whether SIF is going to be a useful constraint on GPP before we have to go through the challenging process of fully assimilating the data. We also believe that the information content study here is substantial enough. Adding in a full assimilation to estimate parameter and GPP values is a complicated non-linear problem and adding this into the current manuscript would make for too large a study. An assimilation of the data where one actually estimates global GPP is the subject of subsequent study.

A detailed description of the construction of the observation uncertainty may detract from the papers 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. Agreed. Refer to general response section.

To see such a large reduction in error sent warning bells ringing with me; I dont 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.

2.3 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? Good point. We have edited this figure to make it clearer. We have removed any rows/columns that have no correlations and increased the font size.

- 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. Yep fair enough. We've provided a better description in the text and caption.

- Figure 3: What are the units? Amended.

- 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? We need some clarification here from the reviewer. We can do the following: prior uncertainty divided by prior GPP and posterior uncertainty divided by prior GPP. But we cannot do the following: posterior uncertainty divided posterior GPP. As this is an error propagation study we have not estimated posterior GPP.

- 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 ones 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 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. Good point. We will amend Table A1 and make it clear what the parameters mean if referring to them in text.

- Boilley and Wald (2015) discuss a high bias in the radiation from reanalyses. Im not sure this is the same as the uncertainty mentioned in sections 2.4 and 3.4. Can you elaborate? We were not aware of the Biolley and Wald (2015) study, so we thank the reviewer for the citation and we have included it in the manuscript. A known bias in the radiation such as this should be removed from the reanalyses data before it is distributed. We consider an uncertainty of unknown sign, as shown in Kato et al. (2012), which can be accounted for in the prior uncertainty and constrained through the error propagation system as we demonstrated. We clarify this in the manuscript.

- 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? The process of getting posterior parameter values and obtaining posterior fluxes is a non-linear problem that is therefore arduous and challenging. So, before one goes down this path they can actually assess whether it is worthwhile doing by first assessing the information content, this is linear problem and therefore simpler. However, as SIF has not been used in a full DA system with a process-based model like this before it is valuable to show, in detail, what SIF may constrain, how it does it, and any caveats to this. It seems we have not made it clear enough exactly what this study is and exactly why we are doing it. We have therefore added in some extra points to the introduction and methods section to clarify this.