Response to Anonymous Referee #4

We thank the reviewer for their valuable comments. Throughout the document reviewer comments are displayed in standard font while author responses are presented with blue text.

This work details a set of pseudo-data experiments in which the prior flux and prior flux uncertainty are varied in a systematic manner in order to assess the sensitivity of atmospheric inversions with satellite data to the prior flux constraint. The manner in which this is carried out is a good template for assessing sensitivity to inversion system ingredients.

General comments: 1. The presentation in terms of images makes some of the conclusions and discussion difficult to assess. Specifically I'm referring to Figures 3-5, where the "meat" of the results is contained. I would suggest that the NEE range differences are great material for a table, and that the other two columns are perfect to be condensed into stacked bars or something similar for ease of visibility.

We appreciate the comment from the reviewer regarding Fig. 3-5 in the original manuscript (now Fig. 6-8 of the updated version of the manuscript). While producing the manuscript we spent significant time and effort developing these figures in order to present as much information as possible in a clear manner. Due to the large amount of information provided in these figures, it proved a difficult task. We tested numerous figure types (line graphs, stacked bar graphs, pie charts, etc.) and determined the current layout was the most effective. Specifically, for the stacked bar charts, when multiple prior or posterior NEE estimates are very similar (which happens frequently) for a specific region/season, they will not be visible as there is not enough difference in the values. After numerous versions of Fig. 3-5 we reverted back to the grouped bar chart format as the best way to display the main results of this study.

We point the reviewer to Table 3 of the manuscript where NEE range values illustrated in Fig. 6 are already listed (along with "true", multi-model prior and posterior means, and standard deviations of NEE values shown in Fig. 6). We have added the following text to the Fig. 6 caption to guide the reader to this information: "Detailed statistics of the "truth", multi-model means of prior and posterior NEE estimates, standard deviations, and ranges displayed in this figure are listed in Table 3.". In response to the reviewer, we have added Table S1 and S2 to the supplementary material of the revised manuscript in order to provide the "truth", mean, standard deviation, and range of the NEE values calculated in the sensitivity studies of prior error statistics and ocean glint (OG) simulations, respectively (detailing Fig 7 and 8).

2. My main complaint with this is that there are a few papers out there now that use ensembles of models for inference (Basu et al, 2018; Crowell et al, 2019; older ones like Peylin et al, 2013), and that this would be a much more effective paper if it were to place itself in the context of the "uncertainty budget" for these other papers. For example, Crowell et al (2019) presents results in the form of ensemble means and standard deviations, and Basau et al (2018) presents ensemble members individually, and the results here could be placed beside the Basu et al (2018) results to attempt to explain the scatter in the flux results in Crowell et al (2019). That sort of analysis would elevate the messages in this paper to a greater impact.

We fully agree with the reviewer that the comparison of our results with Crowell et al. (2019) and Basu et al. (2018) is important and will greatly improve the impact of this study. We thought of this prior to submission of our manuscript, however, Crowell et al. (2019) (along with the OCO-2 Multi-model Inter-comparison Project (MIP) Level 4 (L4) CO_2 flux data) and the supplementary tables of Basu et al. (2018) were not yet published. However, now that this information is available, we have performed the suggested comparison and have added the following text to the conclusions section of the revised manuscript:

"As explained earlier in this study, estimates of surface CO_2 fluxes from numerous inversion systems in the OCO-2 MIP ensemble model framework, using identical OCO-2 observations, result in different optimized/posterior regional NEE fluxes (Crowell et al., 2019). This inverse model variance can be due to numerous factors (e.g., model transport, inversion methods, observation errors, etc.) including prior model mean and uncertainty estimates. In order to estimate the amount of variance in the results of posterior NEE values from the OCO-2 MIP which could be due to prior flux estimates, we compare our OSSE derived residual posterior NEE range (using LN+LG) to the range in the posterior Level-4 OCO-2 Flux data (using both LN and LG) (https://www.esrl.noaa.gov/gmd/ccgg/OCO2/index.php) in each TransCom-3 region. This comparison suggests that prior NEE and uncertainty statistics could contribute 10-30% (average ~20%) of annually-averaged NEE variance calculated for each TransCom-3 region in the OCO-2 Level-4 MIP flux data. Comparing this contribution of prior model impact to the OSSE study by Basu et al. (2018), which calculated the impact of atmospheric transport on posterior NEE estimates when assimilating OCO-2 observations, this contribution is ~50% less compared to the impact of atmospheric transport. From our study and Basu et al. (2018) it is estimated that the combination of prior flux/uncertainty assumptions and atmospheric transport could contribute on average ~50% of the annually-averaged posterior NEE variance of the OCO-2 MIP study."

Also, the references to "Crowell et al. in prep." have been updated to "Crowell et al., 2019".

Specific comments: Page 4, Line 28 - how much do the models really vary using this approach? I'd guess not much. Can you provide a figure at a well known flux tower site as a demonstration?

A similar comment was made by the editor on behalf of an additional referee. In order to demonstrate the difference between the "true" NEE and prior model data in regards to diurnal variability, monthly-averaged 3-hourly (MsTMIP) and hourly (for the 4 prior biosphere models) NEE values were plotted for July 2015 at the well-known Park Falls flux tower site (45.95°N, 90.27°W). This figure has been added to the revised manuscript in the supplementary section as Fig. S1. We added additional text to Sect. 2.1 of the revised manuscript which reads: "We allow the "true" and prior models to have different diurnal variability in order to represent a realistic scenario, as prior models will differ some from the actual diurnal variability of NEE in nature. In general, the diurnal variability of NEE is similar between the "true" and individual prior models. An example is shown in Fig. S1 where it can be seen that the diurnal NEE from the "true" and prior models for July 2015 at the Park Falls flux tower site (45.95°N, 90.27°W) have near identical temporal diurnal patterns and only differ in NEE magnitude.".

Page 7, Line 16-18 Chevallier et al also went after this by looking at structural uncertainties in the ORCHIDEE ecosystem model. Chevallier, F., Viovy, N., Reichstein, M., & Ciais, P. (2006). On the assignment of prior errors in Bayesian inversions of CO_2 surface fluxes. Geophysical Research Letters, 33(13), L13802. https://doi.org/10.1029/2006GL026496

We appreciate the reviewer pointing out this study. We have added this reference to the revised manuscript and additional text in Sect. 2.4.5 which reads: "…, using continuous in situ measurements of CO_2 flux compared to model simulations to inform prior errors (Chevallier et al., 2006), …"

Page 8, Lines 5-9 - these errors are even smaller than those predicted by the Level 2 retrieval, and those are known to be underestimated (from various uncertainty quantification talks). I wonder, could this really be a sign that your prior errors need tuning, rather than your observation errors? In the OSSE setting, it's equivalent which way you go, but in real data settings, this choice can matter a lot. There are other metrics to optimize the prior errors vs. the observational error statistics, such as the Desrozier approach (commonly used in numerical weather prediction)

We agree with the reviewer that the observational error values applied to our OSSE simulations for synthetic OCO-2 data are smaller than that in the real data. It should be noted that the simulated synthetic OCO-2 data produced in this study is done so using known fluxes and transport and sampling a model-predicted atmosphere, thus errors should be small. This manner of producing synthetic satellite observations makes it so there is no model-data mismatch or systematic error which is the major fraction of OCO-2 error values.

The reviewer is correct in the fact that the manner in which observational and prior error are adjusted is equivalent for OSSE simulations (such as this work), however, can impact the results of posterior estimates in “real” inverse model simulations. In addition to the fact we use model produced observations in our OSSEs, we also decided not to reduce our prior error statistics in order to be representative of how prior error is typically calculated/treated in “real” data assimilations. Calculating the difference or spread between state-of-the-science model ensemble members as the best estimate of our knowledge of a process is commonly done to define prior error values. Furthermore, if prior error values are reduced and observation error remain large, this results in an inversion system which is limited in the ability to deviate from the prior and posterior NEE spread calculated would be due to both observational error and prior NEE flux/error. Therefore, by reducing observational error (attached to data which have no model-data mismatch or systematic error in our OSSE framework), and defining prior error as is done in “real” inversions, provides a true representation of the spread in posterior estimates primarily due to prior flux/error.

Page 10, Lines 34-35 - this is often called the "uncertainty reduction", assuming the standard deviation is a proxy for the uncertainty.

We have modified the following text in Sect. 3.4 of the revised manuscript to demonstrate this: “The reduction in the SD of NEE (also known as uncertainty reduction) in most regions/seasons, calculated as 100 × (1 – (posterior NEE SD)/(prior NEE SD)) is generally > 70% and up to 98%”.

References

Basu, S., Baker, D. F., Chevallier, F., Patra, P. K., Liu, J., and Miller, J. B.: The impact of transport model differences on CO_2 surface flux estimates from OCO-2 retrievals of column average CO_2, Atmos. Chem. Phys., 18, 7189-7215, https://doi.org/10.5194/acp-18-7189-2018, 2018.

Chevallier, F., Viovy, N., Reichstein, M., and Ciais, P.: On the assignment of prior errors in Bayesian inversions of CO_2 surface fluxes, Geophys. Res. Lett., 33, L13802, https://doi.org/:10.1029/2006GL026496, 2006.

Crowell, S., Baker, D., Schuh, A., Basu, S., Jacobson, A. R., Chevallier, F., Liu, J., Deng, F., Feng, L., McKain, K., Chatterjee, A., Miller, J. B., Stephens, B. B., Eldering, A., Crisp, D., Schimel, D., Nassar, R., O'Dell, C. W., Oda, T., Sweeney, C., Palmer, P. I., and Jones, D. B. A.: The 2015–2016 carbon cycle as seen from OCO-2 and the global in situ network, Atmos. Chem. Phys., 19, 9797–9831, https://doi.org/10.5194/acp-19-9797-2019, 2019.