Response to Reviewer 1

Original text is in black. Our responses are in blue text. References to the revised text take the following format: RP78,L54-55 (Revised Page 78, Lines 54-55). The references are to the manuscript with changes tracked.

This paper compares the atmospheric distributions of CO2 resulting from two sets of optimized fluxes derived from GEOS-Chem using different observing systems based on in situ data and GOSAT data respectively. The results show the differences in the optimized fluxes and how their correction is transported in the atmosphere. An evaluation of the seasonal cycle and inter-hemispheric gradient is also provided. Finally, the zonal variability of the flux correction signal at different vertical levels (boundary layer, free troposphere and stratosphere) is also explored. The differences between the two sets of posterior fluxes and their atmospheric distributions highlight problems associated with spatial and temporal coverage of observing systems and their ability to constrain the surface CO2 fluxes at different temporal and spatial scales. Overall, the results point to the conclusion that the in situ observations do a better job at constraining the fluxes at global and annual time scales, leading to smaller biases in their fit with independent observations. While GOSAT data is able to better capture the seasonal cycle at northern extratropical sites. The paper is well written and well structured. However, I have some concerns on the use of atmospheric differences associated with flux correction patterns to draw conclusions on the potential representation of zonally asymmetric patterns by different observing systems. It is not possible to say that GOSAT is (potentially) better at constraining the zonal patterns without substantiating this with an assessement of the errors in zonal variability based on independent observations (e.g. zonal gradients using TCCON or in situ data). The analysis of the seasonal cycle could also be improved by looking at the seasonal amplitude and phase, instead of just providing seasonal biases which is too qualitative in my opinion. The results and conclusions would also be more robust if more than just one year and a half of data was used.

Response: The question raised by the Reviewer here and in the second and third bullets under General comments is an important one. The question ultimately concerns our approach of drawing conclusions from comparing posterior atmosphere adjustments due to flux increments with posterior atmospheric adjustments due to uncertain meteorology. Because it is such a different approach from any other in the literature, it should have been better explained. Therefore, we are grateful for the opportunity to clarify this. In the revised manuscript, we add a new section which mathematically describes the posterior atmospheric adjustment (referred to as the "flux signal" in the original manuscript) and shows that it comprises components due to flux adjustments, initial state adjustments and meteorological uncertainty. Clearly, if the posterior atmospheric adjustment due to flux increments is not the dominant term (for certain spatial or temporal scales), then caution must be exercised when utilising the retrieved fluxes on those scales. Thus we introduce a new diagnostic for retrieved fluxes and we use this to show that assimilating GOSAT observations gives zonal standard deviations that exceed those associated with meteorological uncertainties. We also agree (and we noted this in several places in the original manuscript) that the conclusion of whether GOSAT is better at constraining final spatial scales must rely on the availability of dense network of independent observations. Such a network does not currently exist. The TCCON and in situ surface networks are not sufficiently dense. But, but we can still point to the *potential* ability of GOSAT to see finer spatial scales, using our diagnostic since spatial scales produced by in situ-constrained fluxes do not exceed those produced by uncertain meteorology. In general, larger values of zonal standard deviations means more spatial structure is seen, but *more* does not mean *better*. However, if the threshold (of posterior atmospheric adjustments due to uncertain meteorology) is not exceeded, the spatial structure seen should not be trusted since a perturbed but equally valid wind field would give different spatial structures. Finally, it is always better to have longer runs but we all balance this desire with what is computationally feasible. Because of the systematic and consistent differences between the posterior atmospheric adjustments obtained with the two different posterior fluxes over the course of our experiments, we believe that our conclusions concerning the global/annual time scale and seasonal cycle are robust. We also note that our diagnostic is model dependent so there is inherent difficulty in generalizing results but our use of two very different models yielded the same patterns for atmospheric flux adjustments. To be sure, the data assimilation problem refers to a "best" estimate projected onto a given model's basis and is thus diagnostics of data assimilation systems are inherently model dependent. We had emphasized this in our original conclusions section along with the need for independent confirmation with other models. We further clarify this point in the revised manuscript.

In summary, we appreciate the thoughtful and thought-provoking comments of the Reviewer and we feel that in the course of addressing the questions raised by the Reviewer, the manuscript has been significantly improved.

GENERAL COMMENTS

- The use of CO_2 flux signal to denote the cumulative impact of the flux corrections/adjustments in the atmosphere is a bit misleading. A flux signal gives the impression that it is associated to a process or phenomenon, while here it just reflects a correction (or analysis increment) which depends on the specific model, prior flux and observation used. I would think that using the term 'posterior atmospheric adjustment' would be a better term to describe the difference between posterior and prior atmospheric distributions of CO_2 or alternatively 'flux correction signal'.

  Response: We struggled with to find a good name for "flux signal" and had considered "analysis increment". However, since it was the flux that was incremented not the concentrations, we thought that term would not work. We also agree that the term "flux signal" is not entirely appropriate as it suggests a physical phenomenon. We like the reviewer's suggestion of "Posterior Atmospheric Adjustment (PAA)" and have replaced the term "flux signal" with PAA in the revised manuscript.

- Spatial variability of flux correction signal in the atmosphere does not necessarily translate in better provision of information by observations nor an improvement in spatial/regional patterns. If the observations are very noisy (e.g. GOSAT has larger errors than the in situ observations assimilated in flux inversion systems) or observations are not homogeneously distributed (e.g. many more data over land than sea as it is the case in northern extratropical regions) then the flux corrections can create artifacts in the zonal variability which increase the zonal variability but are nevertheless not realistic.

  Response: We agree with this point: zonal structures in the flux signal do not necessarily imply that the inversion is capturing the true zonal structures. As the reviewer points out, biases in the observations or artifacts due to the uneven spatiotemporal distribution of the observations could plausibly result in zonal structures in the posterior fluxes. Therefore, since we do not validate the zonal structures introduced by the inversion, we cannot conclude that their presence implies that the GOSAT inversion is better capturing the true zonal structures. Although we do point this out in the previous manuscript (Page 21 lines 12-18, Page 22, line 23-28), there were also some places where we erroneously made this connection and have removed these statements. In addition, we have expanded the text discussing the possibility of spurious results. Here are some details. Newly added text is in red, deleted text is struck-out:

  Page 1, line 9 (Key point 1 was revised to): ~~The potential for GOSAT data to better resolve zonally asymmetric structures in the tropics year round and in the northern extratropics except during boreal winter is demonstrated.~~ Inversions constrained with GOSAT data introduce zonally asymmetric structures in posterior atmospheric adjustments that exceed those due to uncertain meteorology in the lower troposphere year round in the tropics and outside the boreal winter in the northern extratropics. (RP1,L10-13)

  Page 19, line 23: ~~This suggests GOSAT is capable of picking up finer spatial scales due to the high density of observations in this region when the satellite shifts its view to the northern hemisphere (Figure 2).~~ (RP25,L20-21)

  Page 20, line 18-25: "In the lower troposphere, zonal asymmetry in GOSAT flux signals exceeds that arising from wind field uncertainty except in November, December and January (Figure 1~~7~~6a). However, for in situ data, the zonal structure can only be trusted in boreal summer (June, July and August). Thus the satellite data are potentially able to retrieve fluxes on finer spatial scales than are in situ data through most of the year, but it is important to note that more spatial structure does not mean better spatial structure. Validation of spatial structures in posterior distributions needs to be made against a dense network of independent observations in order to determine if the increased spatial variation is correct. Given the difference in observation densities (Figures 1 and 2), this result is not surprising. The lack of ability of in situ data to produce zonal asymmetry in flux signals that are larger than those arising from uncertainty in wind fields outside of boreal summer may indicate why it has been difficult for flux inversions to regionally attribute sources with this observation network (e.g. Gurney et al., 2002, Peters et al., 2010, Bruhwiler et al., 2011, Peylin et al., 2013)." (RP26,L19-22)

Page 22, lines 2-6: "Zonal standard deviations of the ~~flux signal~~ PAAF (which reveal spatial structures in the zonal direction) are much larger when GOSAT-informed posteriors are used (in the northern extratropics outside of boreal winter and in the tropics throughout the year) (Figure 16, 17). ~~This indicates a potential for GOSAT data to retrieve finer scale fluxes since the accuracy of such finer scale features requires a dense network of independent measurements to validate.~~ (RP28,L22-23)

Page 22, line 18-30: "In situ observations were found to generate zonal standard deviations larger than this minimum level only in boreal summer whereas GOSAT data exceeded this threshold through most of the year (Figure 16, 17). This potential for retrieving finer spatial scales with GOSAT sampling relative to the in situ network makes sense given the density of GOSAT observations (Figure 2) and is consistent with the prediction of Takagi et al. (2014) or Deng et al. (2016). Moreover, the ability to retrieve zonal structure is evident throughout the year in the tropics and in all seasons except boreal winter in the northern extratropics is rather encouraging. However, verifying such finer scales will be challenging given the limited spatial coverage of validating measurements from TCCON or aircraft platforms and temporal and spatial scales resolved may depend on the characteristics of the flux inversion system. Indeed, the current dispute over the enhanced European sinks obtained with GOSAT data (Feng et al., 2016; Reuter et al., 2014; Houweling et al., 2015) indicates that the finer spatial scales retrieved are not necessarily correct and are difficult to validate. It is plausible that spurious zonal structures in the PAAF could be introduced by spatially varying biases in the observations or uneven spatial coverage. However, there is also evidence supporting the ability of space-based observations to recover zonal asymmetries in the CO_2 fields. Liu et al. (2017) use observations from GOSAT and OCO-2 to isolate tropical flux anomalies between continents during the 2015-2016 El Niño event, while Chatterjee et al. (2017) found that zonal asymmetries in XCO_2 anomalies could be isolated during the same El Niño event. Furthermore, the fact that the spatial structure seen in ~~flux signals~~ PAAFs obtained with in situ data surpassed the minimum uncertainty level only in boreal summer implies that regional attribution of fluxes may be challenging with the in situ observation network alone when the inversion integrates signals over many seasons." (RP29,L15-19)

Page 23, line-14-15: As noted in response below.

Newly added citations:

Liu et al. (2017): J. Liu et al., Contrasting carbon cycle responses of the tropical continents to the 2015–2016 El Niño. Science 358, eaam5690 (2017).

Chatterjee et al. (2017): A. Chatterjee, Influence of El Niño on atmospheric CO2 over the tropical Pacific Ocean: Findings from NASA's OCO-2 mission. Science 358, eaam5776 (2017)

- The paper would benefit from a better quantification of error reduction at different scales based on TCCON and in situ observations which could be presented in tabular format.
  Response: Given that there are only 14 TCCON stations available worldwide in our period of study, it is not possible to assess spatial scales with this network. The surface network is more plentiful, but with only around 40 stations reporting hourly, this too will not be enough to compute zonal asymmetry since most of the sites are clustered in Europe and North America. This is why we had stressed (abstract, last line; p.21, lines 14-15; p.22, line 5) that a dense network of independent observations is needed to verify an improvement in spatial scales achieved by any system, but that such a network does not yet exist.

- The fact that the minimal level of uncertainty in the zonal variability associated with imperfect knowledge of winds is around 0.5 ppm and the global zonal variability of flux corrections is of similar magnitude does not make the posterior zonal flux correction pattern is unreliable. The objective of the flux inversion systems is to reduce the uncertainty of the posterior fluxes and the flux corrections on their own do not necessarily reflect the uncertainty reduction. The posterior zonal patterns should be assessed with independent observations and their standard error compared to the minimal level of uncertainty associated with transport.

Response: The reviewer suggests that comparing posterior atmospheric adjustments due to fluxes with adjustments due to imperfect knowledge of winds is not the appropriate comparison. We disagree. The posterior atmospheric adjustment due to flux increments can and should be compared to the atmospheric adjustment due to uncertain meteorology and this can lead to new insights on the information provided by different observing systems. This becomes clear when we look at the mathematical definitions. Let us first define an atmospheric transport model (*T*) that integrates an initial state for CO_2 (c_0), a set of surface fluxes ($s_{0,n-1}$) and a set of meteorological fields ($x_{0,n-1}$) to yield a CO_2 distribution at time step *n*:

$$c_n = T\left(x_{0,n-1}, c_0, s_{0,n-1}\right). \quad (1)$$

Subscripts refer to time steps and the model integration starts at time step 0 and yields a final CO_2 state at time step *n*. The posterior atmospheric adjustment (using the Reviewers' suggested name above) is simply the difference between the constituent distribution obtained with posterior fluxes and that obtained with prior fluxes:

$$\Delta c_n = T\left(x_{0,n-1}^a, c_0^a, s_{0,n-1}^a\right) - T\left(x_{0,n-1}^b, c_0^b, s_{0,n-1}^b\right) \quad (2)$$

The superscripts *a* in (2) can be viewed as the "after adjustment" values and the superscript *b* refers to the before adjustment value. Note that (2) is a general form which allows for the adjustment of the initial concentrations (c_0), and imperfect meteorological states ($x_{0,n-1}$) at the same time that the fluxes are adjusted. Because initial concentrations are not adjusted in our GEOS-Chem flux inversions, we ignore the impact of potential variations of c_0 on Δc_n. However, we allow for the possibility that the meteorological states are not perfect. Thus $x_{0,n-1}^a$ is a set of meteorological states which are perturbed by realizations of meteorological analysis error (computed as explained in our supplemental section, using our operational Ensemble Kalman filter). Then (2) can be written as:

$$\Delta c_n = T\left(x_{0,n-1}^a, c_0^a, s_{0,n-1}^a\right) - T\left(x_{0,n-1}^a - \varepsilon_{0,n-1}, c_0^a, s_{0,n-1}^a - \Delta s_{0,n-1}\right) \quad (3)$$

Now, expanding (3) in Taylor series reveals that to first order, the posterior atmospheric adjustment has two components

$$\Delta c_n \cong \Delta c_n^s + \Delta c_n^x \quad (4)$$

where

$$\Delta c_n^s = T\left(x_{0,n-1}^a, c_0, s_{0,n-1}^a\right) - T\left(x_{0,n-1}^a, c_0, s_{0,n-1}^b\right) = PAAF \quad (5)$$

$$\Delta c_n^x = T\left(x_{0,n-1}^a, c_0, s_{0,n-1}^a\right) - T\left(x_{0,n-1}^b, c_0, s_{0,n-1}^a\right) = PAAM \quad (6)$$

PAAF is the component of PAA due to flux adjustments while PAAM is the component of PAA due to uncertain meteorology. PAAF is computed by integrating the transport model with a set of posterior fluxes and again with the prior fluxes but both integrations use the same set of meteorological analyses ($x_{0,n-1}^a$) and initial concentrations. However, this is only one component of the posterior flux adjustment because the meteorological analyses are not perfectly known, and we can simulate that uncertainty by perturbing the meteorological analyses with realizations of meteorological analysis error (see supplemental material). In other words, for a given set of fluxes, the meteorological fields could have been slightly different but equally valid in the context of the meteorological analysis errors. This is what PAAM defines and it is computed by integrating the model twice (with perturbed and unperturbed meteorology) for a given set of posterior fluxes and where we again use the same initial concentrations in both integrations. A novel aspect of our work is the ability to compare the component of posterior atmospheric adjustment due to flux increments (PAAF) with that due to meteorological uncertainty (PAAM). If the PAA component due to flux increments alone (PAAF) does not exceed the component due to meteorological errors (PAAM), then it is not the dominant contribution in (5) and should not be accorded great significance.

In summary, a novel aspect of our work is the ability to compute two components of PAA with our coupled meteorological and tracer transport model. This comparison provides new insights into the atmospheric adjustments arising from flux increments associated with different types of observing systems. This information is complementary to direct comparisons of posterior CO_2 distributions to measurements made with a single set of meteorological analyses.

In the revised manuscript, the PAA is now defined mathematically in a new section (2.3). This greatly clarifies the arguments and results presented in our manuscript.

SPECIFIC COMMENTS

- Page 7, Line 14: Isn't the uncertainty of 22% associated with NEE very low?

Response: The uncertainty of 22% is applied to both GPP and Respiration because these are the quantities that are optimized in the GEOS-Chem inversion. These quantities are significantly larger than NEE. So too are their uncertainties.

- Page 7, Line 19: I would not call GOSAT coverage "dense".

Response: We changed "dense" to "more dense" since it is fair to say GOSAT coverage is more dense than the surface network. (RP9,L12)

- Page 8, Lines 2-3: "Note that ... " sentence is not clear.

Response: We clarified this statement from "Note that posterior fluxes contain total fluxes from unoptimized as well as optimized fluxes." To "Note that posterior fluxes contain the total of all optimized (GPP, Respiration, ocean, biomass burning and anthropogenic) fluxes and the small amount of un-optimized fossil fuel emissions from shipping (~0.19PgC/yr) and aviation (~0.16PgC/yr)." (RP9,L31-32 to RP10,L1-2)

- Page 9: Please provide a quantitative estimate of standard error and bias per month/season for the surface in situ evaluation in order to assess the seasonal cycle quantitatively. When the bias is shown to be smaller, it would help to know by how much

Response: This is a good suggestion. Table 2 was added to the manuscript. It contains the seasonally aggregated statistics (means and standard deviations) for 6 seasons used in the TCCON figures and table, for the four experiments (two models forced by posterior fluxes from GEOS-Chem inversions with GOSAT or in situ data). This new table is indeed helpful to the discussion of Figure 4 (revised Figure 5) on page 9 (RP13-14).

- Page 10, Line 2, Page 11, Line 23: Is the flux correction signal in the atmosphere "propagated" or "transported"?

Response: In the first instance, we changed "…before considering the vertical propagation of the flux signal in section 3.2." to "…before considering the vertical structure of the PAAF in section 3.2." (RP15,L6-7). In the second instance, we changed "…flux signal then propagates vertically…" to "flux perturbation is then vertically transported…" (RP17,L4-5)

- Page 12, Line 16: Why is GOSAT reducing meridional gradient? From Figs 8 and 10 it looks that the meridional gradient from the GOSAT posterior fluxes is worse than that from the in situ data.

Response: From our study, we cannot say why the meridional gradient is worse than that from the in situ data, but we noted here (page 3, line 19 and page 10, line 22) that other studies have seen the same thing. We also noted in Polavarapu et al. (2016, ACP on Page 3, Line 18): "It has been suggested that the GOSAT-based inversions shift some uptake from North Africa to Europe which reduces the north-south gradient in CO_2 and reduces agreement with observations (Houweling et al., 2015)"

- Page 14, Lines 22-25: How do you reconcile this with the larger bias of GOSAT versus TCCON in SH?

Response: It is difficult to reconcile the larger bias of GOSAT-based CO_2 distributions with TCCON in the southern hemisphere sites in all seasons (relative to CO_2 distributions informed by in situ data) (Figure 9) with the lower bias when compared to HIPPO3 data (Figure 10). However, we can speculate that it is due to difference in specific locations sampled by the two measurement networks. In addition, transport error does play a role because HIPPO comparisons using GEM-MACH-GHG are consistent with the TCCON results (Figure S12). Note that comparisons to independent surface (continuous) sites in the southern hemisphere also show a larger bias for GOSAT-based posteriors even with GEOS-Chem (e.g. at South Pole in new Table 2). For example, in Figure R1, we show monthly statistics from Cape Matatula for GEOS-Chem integrations (top panel) and GEM-MACH-GHG integrations (lower panel). This is also consistent with the TCCON results of Figures 8 and S11. Because there are some results that we cannot reconcile, we made conclusions based only on the results that are consistent. What is consistent is that lower CO_2 values are produced with in situ-constrained fluxes in both comparisons and both models and for long (12 or 18 month) time averages, CO_2 fields forced by in situ-based posteriors agree better with independent measurements.

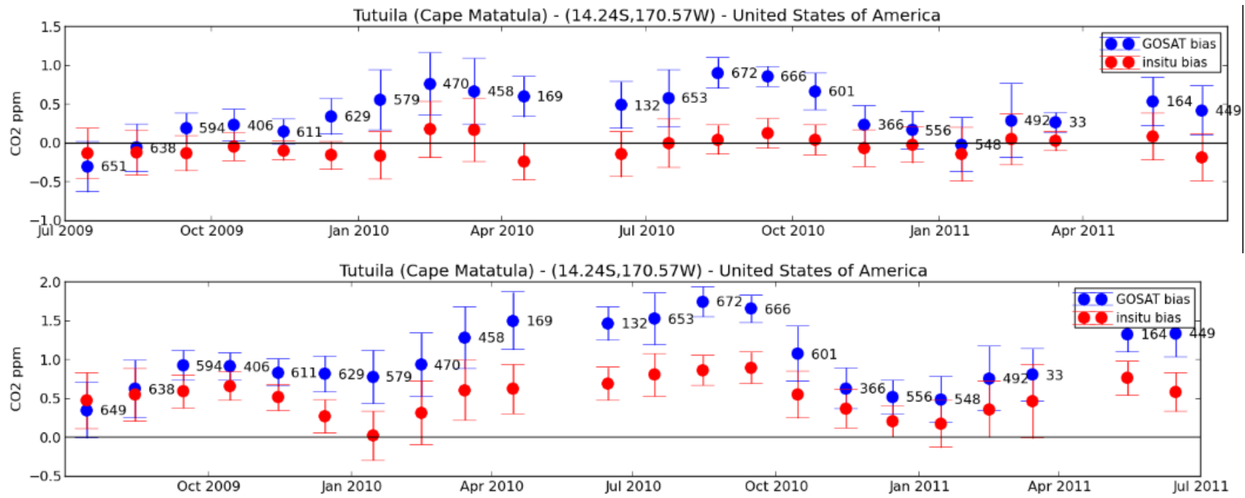

Figure R1: Comparison of model and in situ hourly observations at NOAA's Cape Matatula site with GEOS-Chem (top panel) and GEM-MACH-GHG (bottom panel). All observations (including night time) are used in statistics. The whiskers correspond to one standard deviation and the number to the right of each dot is the number of observations used in each calculation. Both GEM-MACH-GHG simulations produce more CO2 than the corresponding GEOS-Chem simulations due to transport error mismatches. For both models, in situ informed posterior distributions better match measurements.

- Page 16, Line 2: How do you explain that GOSAT produces better fit with observations in middle to upper troposphere in boreal winter?

Response: We can only speculate on this. We saw that in situ posteriors better match TCCON total columns at all times except boreal summer (Figs. 9, S11) and, on long time scales, they better match both TCCON (Figs. 8, S10) and NOAA aircraft profiles (Fig. S13). However, seasonal anomalies are better captured with GOSAT posteriors (Table S1) and there is seasonal variation to the agreement with the aircraft profiles (Figs. 11, S14). So, it is possible that the better fit of GOSAT posteriors to aircraft profiles is specific to the winter season, for North America. We cannot speculate on a reason for this. However, Figure 18 shows that zonal asymmetries (standard deviations) in CO_2 adjustments are well below those due to meteorological uncertainty (Fig. 18b,c) for the northern extratropics in boreal winter. So for the whole zonal band, zonal asymmetries (e.g. results for North America only) are not robust because they are likely sensitive to transport error.

- Page 21, Lines 11-18: The message that GOSAT observations have the potential benefit of improving the zonal structure seems to be contradicted by the results from flux inversions using GOSAT data published in Houwelling et al (2015). Therefore, the conclusion of the potential benefit of GOSAT highlighted in the abstract can be misleading.

Response: We did note that the "potential benefit" aspect is contradicted by the Howelling result a few lines later in the same paragraph. However, we revised the abstract to better correspond to the revised manuscript which focuses on our new diagnostic and its potential utility for understanding flux inversion results.

- Page 23, Line 15: "GOSAT better captures zonally asymmetric structure ..." should be rephrased as this has not been proven in the paper.

Response: The reviewer is correct. We only showed that there was more zonally asymmetric structure, not that it was better. We changed the statement to "GOSAT better captures the seasonal cycle at northern extratropical TCCON sites." (RP30,L6)

- Page 23, Lines 24-27: Note that this type of comparison has already been done by Locatelli et al. (2013, ACP) for CH4.

Response: What Locatelli et al. (2013, ACP) did was to use 10 different transport models with identical initial conditions and prior fluxes to simulate observations which were then used in an inversion system with a single model. Their experiment includes transport errors in the synthetic observations which are then assimilated with the reference inversion system. While this is an interesting and useful method of exploring transport error impact on flux estimation, what we suggested in the text was rather different. Instead of forcing different models to use specific initial conditions or fluxes, we simply suggest taking the results (posterior fluxes) from different inversion models and integrating them all with a reference transport model. This will create a convolution of transport model errors from the inversion model and the reference model (as seen in our work). If transport characteristics of the reference transport model are known, relative transport errors of the flux inversion models can be inferred, to some extent. Obviously, there are limitations with this method since it is *relative* transport errors that are obtained, but since it is very easy to do (does not require a common protocol), we suggest that it is worth considering.

- Page 24, Line 10: .. seasonal correlation of "error" covariances.
Response: Corrected. (RP31,L2)

- Figure 4 and Page 9: It would be good to include GEOS-Chem in Fig. 4.
Response: We agree that because of the content of the discussion in the text, it would be good to have curves corresponding to both models on the same plot. However, there would be too many curves if all the existing ones were retained. Therefore, we have modified Figure 4 (revised Figure 5) to show the GEM-MACH-GHG and GEOS-Chem posterior CO_2 distributions obtained with in situ data compared to observed time series. The analogous figure for posterior CO_2 distributions obtained with GOSAT posterior fluxes is not shown but the newly added Table 2 (as suggested by the Reviewer in specific comment on page 9) supports the discussion concerning both types of posterior fluxes. We also now use the quality flags on the observed data and show only those not flagged as suspicious. The text on page 9 was modified accordingly. (RP13-14)