Full Response to Referee Comments
J. S. Wang et al.

Referees' comments below are in *italics*, and our responses are in regular text style.

Anonymous Referee #1

This paper, entitled "A global synthesis inversion analysis of recent variability in CO2 fluxes using GOSAT and in situ observations" by Wang et al., combines satellite and ground-based observations of CO2 concentrations to infer sources/sinks of carbon over the world.
One of the unique aspects of this work is the adoption of a batch Bayesian synthesis inversion approach that is at higher spatial (perhaps; see comments below) and temporal resolution relative to conventional global inversions that allow full-rank posterior error correlations to be evaluated. However, I have major questions about the results, and I found the conclusion about GOSAT observations to be weak. Therefore, I recommend major revisions prior for consideration for publication in ACP.
We thank the reviewer for the insightful comments. We have made major revisions as detailed below.

MAJOR COMMENTS I. Utility of satellite observations The authors carried out extensive comparisons between inversion results using in-situ data versus those from using the GOSAT satellite data, as well as both data. The results often differed dramatically, and the comparisons against independent aircraft data highlighted potential biases in the GOSAT data. I am left wondering whether it is worth assimilating the GOSAT data at all, given the fact that the inversion-derived spatial and temporal patterns of carbon fluxes could be highly erroneous, with significant biases that are not reflected in the posterior errors (see point II. below). After highlighting several issues with GOSAT in the Discussion and Conclusions section, the final paragraph was anticlimactic and weak. Can the authors provide a broader guidance on how the GOSAT data can be properly used (if at all)? Is there an important take-home message that this study can provide to the reader? A clear, concise message is lacking in this lengthy paper.
We now highlight the following take-home message: Our inversions demonstrate that passive satellites such as GOSAT can in principle reduce flux uncertainties relative to in situ networks because of their greater coverage, and that in situ and satellite observations complement each other. However, accurate estimates are currently hampered, especially on subcontinental/sub-ocean basin scales, by retrieval biases and remaining coverage gaps. With the current bias-corrected GOSAT data, interannual variations seem to be captured over certain regions, though still with some bias. To remedy the situation, bias corrections would need to be improved, especially over regions that are not currently sampled by TCCON and are challenging for forward model simulations, e.g. Africa. In addition, future expansions of in situ networks and coverage by potential active and passive satellite sensors would be helpful in filling gaps in joint in situ-satellite inversions.

II. Uncertainties (prior and posterior errors) Several issues arose as I considered how the prior and posterior errors were dealt with in this paper: 1) More quantitative information regarding the prior "observation" uncertainties at different sites based on the "model-data mismatch" would be helpfulâ˘A˘Te.g., Table or a map of the prior errors at different sites.
We have now created a table (Table S1 in the supplement) containing the model-data mismatch errors for the in situ sites (as well as basic site information such as latitude, longitude, etc.).

2) The handling of transport errors requires more discussion.
The authors adopted a single transport model (instead of an ensemble of transport models as some other studies do), PCTM, and assumed that the "model-data mismatch" accounts for the transport error. How did PCTM perform as part of previous TRANSCOM experiments, and other tracer experiments (e.g., SF6)? Its interhemispheric mixing? How about its vertical mixing and rectifier strength?
We now elaborate on PCTM's transport performance, including in the context of TransCom experiments, in Section 2.3 (Atmospheric transport model and model sampling). The new text is as follows:
Evaluation of PCTM over the years has shown it to a reliable tool for carbon cycle studies. For example, Kawa et al. (2004) showed that the SF_6 distribution from PCTM was consistent with that of observations and of the models in TransCom 2, suggesting that the interhemispheric and vertical transport were reasonable. PCTM performed well in boundary layer turbulent mixing compared to most of the other models in a TransCom investigation of the CO_2 diurnal cycle (Law et al., 2008). The TransCom-CH_4 intercomparison (Patra et al., 2011) showed that a more recent version of PCTM performed very well relative to observations in its interhemispheric gradients of SF_6, CH_3CCl_3, and CH_4 and interhemispheric exchange time, and follow-on studies (Saito et al., 2013; Belikov et al., 2013) demonstrated through evaluation against observed CH_4 and ^{222}Rn that the convective vertical mixing in PCTM was satisfactory overall.

Where does PCTM fit within the well-known "diver-down" figure in Schimel et al. [2015] (www.pnas.org/cgi/doi/10.1073/pnas.1407302112; its Fig. 3)?

We have now calculated the north-south land flux partitioning for our various inversion cases and display them in a diver-down plot (shown below, and now included in the Supplement for our paper). We find that our results for the in situ inversions and especially the GOSAT inversions lie outside of the 1-sigma range consistent with the GCP global land flux estimate, with rather large positive values (carbon source) for the south+tropics. An exception is the tight-priors case for the in situ inversion, which lies on a 1-sigma boundary. One reason for the weak overall land sink in our results is the large ocean sink compared to GCP for all cases, which is a consequence of our loose prior ocean fluxes and necessitates a small land sink for global mass balance. In addition, the GOSAT inversion total budget over relatively short 12-month periods may deviate substantially from a budget based on in situ measurements as in GCP (for the reasons proposed at the beginning of Section 3.2 of our manuscript). We also find that the north-south partitioning is somewhat sensitive to the time period considered, which could be due to the short 12-month periods here. For example, the Sep 2009-Aug 2010 period gives a smaller source in the south than the Jun 2009-May 2010 period for the GOSAT inversion, by ~0.7 Pg C, with a correspondingly smaller sink in the north.

We now refer to the diver-down plot (Fig. S2) in Section 3.2 of the manuscript in discussing the shift in the sink from the south + tropics to the north in the GOSAT inversion relative to the in situ inversion, as well as in several places in Section 3.3 in discussing the land-ocean flux partitioning and sensitivity to prior uncertainties.

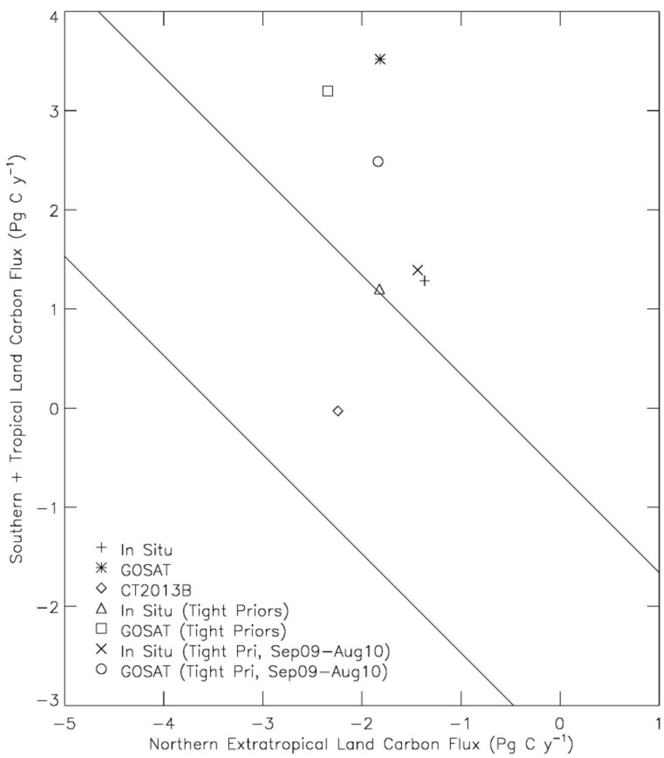

Figure S2. Posterior north-south land flux partitioning after Schimel et al. (2015). The diagonal lines are based on the global land carbon exchange (= land-use change emissions – land sink) estimated by GCP (2015) for the years relevant to the present analysis, i.e. 2009 and 2010, ± 1σ. Fluxes are for June 2009-May 2010 except where specified in the legend (for September 2009-August 2010). CT2013B refers to the CarbonTracker data assimilation system.

3) What is the logic behind neglecting a priori spatial and temporal error covariances? The stated reasoning is "The neglect of a priori spatial error correlations is justified by the size of our flux optimization regions, with dimensions on the order of one thousand to several thousand km, likely greater than the error correlation lengths for our 2_ x 2.5_ grid-level fluxes. For example, Chevallier et al. (2012) estimated a correlation e-folding length of ~500 km for a grid size close to ours of 300 km x 300 km based on comparison of a terrestrial ecosystem model with global flux tower data." If the inversion for flux adjustments is carried out at scales of ~1000 x 1000 km, following the TRANSCOM regions, this is imposing an artificial error covariance lengthscale of ~1000 km (significantly larger than the ~300 km lengthscale), thereby resulting in aggregation errors. Also, given this coarse lengthscale, can the authors make the claim that the adopted inversion is really high spatial resolution? If 2_ x 2.5_ grid is considered high resolution, this is not much different from CarbonTracker's spatial gridding.

We acknowledge that the size of our flux optimization regions, though smaller than that of previous batch inversions, does still impose a correlation that can result in aggregation error. Note that the proper correlation length scale is actually ~500 km, not ~300 km, and the number refers to e-folding, so the correlation is not negligible beyond 500 km. Regarding our claim about the "relatively high spatiotemporal resolution" of our inversion in line 110, we were specifically comparing the *flux optimization* resolution to that of previous *batch* inversions, e.g. TransCom 3 and Butler et al. (2010). We now make that more clear in the paper. We similarly clarify the statement about the "high spatiotemporal resolution of our inversion relative to other Bayesian synthesis inversions" in lines 150-151. Actually, the spatiotemporal resolution of our flux optimization (108 regions, 8-day) is comparable to or even higher than some Kalman filter/smoother systems in use, e.g. CarbonTracker (156 regions, weekly), Feng et al. (2009; 144 regions, 8-day), and Takagi et al. (2011; 64 regions, monthly). As for the transport grid resolution, we did mention in lines 212-214 that "the relatively high-resolution transport model used here (Sect. 2.3) captures much of the variability in the observations beyond background levels." We have now changed "relatively" to "reasonably" since how the transport resolution compares to other models is irrelevant in this statement.

4) Are the prior uncertainties of proper magnitude? While the authors enlarged the observation uncertainties by a factor of 2 to lower the normalized Chi-squared value closer to 1, Table 2 shows that the in situ only inversion results in a Chi-squared of 4.0, which suggests that the prior uncertainties may still be underestimated. Should the prior uncertainties be inflated even more?
We wanted to strike a balance between achieving a chi-squared value reasonably close to 1 and not inflating the uncertainties derived from our in situ data uncertainty formulation too much. In addition, in a test in which we further inflated the uncertainties to 3 times the original values, the posterior fluxes were not very different from those for the reported inversion at the scale of TransCom regions, so that our results appear to be robust. Finally, the use of the chi-squared criterion to scale the prior and/or observation uncertainties is inherently a rough approach, as pointed out by Rayner et al. (1999, *Tellus*). We now add a brief statement in the manuscript mentioning the overall robustness of the results with respect to observation uncertainties.

5) Can we place much faith in the posterior errors for calculated sources/sinks that are derived through the inversion? There are several cases in the paper where different inversions yield different carbon fluxes whose values do not overlap even if allowing for the posterior error bars. Since there is only one true carbon cycle, the posterior errors should allow for the "true fluxes" to be encompassed within the error bars even under different inversion setups. How much trust (if any), therefore, should we place on the posterior errors? I realize that other global inversion studies often face the same situation of underestimating posterior errors, but this would be an opportunity for the authors to comment on the posterior errors.

The Bayesian, least-squares inversion framework adopted here, as in other CO_2 studies, assumes Gaussian error distributions with no bias (observation, transport, prior, etc.), so the posterior uncertainties do not account for possible biases. That's why we discussed in much depth the possible biases in the results, using independent observations for evaluation. We now explicitly state this limitation of the posterior uncertainties after Eq. 3 for the posterior error covariances in Section 2.4. Given that limitation, one wouldn't necessarily expect posterior error bars to overlap for different inversions, e.g. in situ vs. GOSAT, which may have different regional biases.

Note that we mentioned overlaps (or lacks thereof) mostly at the 1-sigma uncertainty level, as that is what the error bars in our figures represent. But two rather different flux estimates could still be considered to encompass the "truth" if they overlap at a 2-sigma level, for example. Also, in a comparison of monthly fluxes from our in situ inversion and CT2013B, we highlighted the similarity, not difference, of the two sets of fluxes: "with overlapping 2σ ranges at all times except in the Northern Oceans region" (lines 491-492). In another part of the paper, we mentioned large differences that even exceed 3-sigma (for Europe), but that refers to comparison of two different years--2010 and 2009 summer posterior fluxes.

III. European sink The large European sink of -1.5 PgC/yr is difficult to reconcile with ground-based information. While the large sink could be caused by biases in GOSAT *data, as pointed out by the authors, I am puzzled by the fact that the summertime drawdown is so large in the prior fluxes as well, reaching fluxes of -5 PgC/yr and comparable with the summertime drawn in boreal and temperate North America. Is there really enough biomass/vegetation to sustain a summertime drawdown of -5 PgC/yr in Europe? Also, I suggest the authors consider and cite a recent paper focusing on the European sink [Reuter et al., 2017, BAMS; http://dx.doi.org/10.1175/BAMS-D-15-00310.1].*

Our GOSAT inversion results are not unique, as the average GOSAT-only posterior annual flux for Europe over the different models in the intercomparison of Houweling et al. (2015) was around -1.5 Pg C/y as well. As for our prior fluxes, comparing them with the mean seasonal cycle over the early 2000s for a set of in situ inversions reported by Peylin et al. (2013) shows that our prior for Europe falls squarely within the range of inversions, for example peaking at ~-0.5 Pg C/month in the summer (equivalent to ~-6 Pg C/y). (There is agreement for North America and North Asia as well.) Also, the TransCom 3 Europe region includes western Russia, so it is actually of comparable size to Boreal or Temperate N. America and contains much forested land.

We now cite Reuter et al. (2017) in our discussion of the shift in the global sink in the GOSAT inversion relative to the in situ inversion, noting that they highlight a similar discrepancy between satellite-based and ground-based estimates of European CO2 uptake and cite retrieval and sampling biases as possible sources of error in the former (while also noting sampling issues with in situ networks for the region).

OTHER COMMENTS Length of text: This is a long paper, with the main text reaching over 40 pages (double-spaced). The Results section as a whole felt unnecessarily long. Can the authors seek to condense?

We now omit some details/condense the text especially in the Results section, for example in the evaluation of the model against Amazonica profiles (omitting details on numbers of cases in which one inversion is better than another in lines 404-412), and in the evaluation against HIPPO (lines 566-597). We have also moved Figures 13 and 14 (comparisons with HIPPO and surface observations for tight priors case) to the Supplement, saving some space in the main document. Although the remaining text is still lengthy, we have put much effort into making the wording as concise as possible and the overall paper is comparable in length to other global CO2 inversion papers that assimilate satellite and in situ data and use multiple evaluation data sets.

Spatiotemporal resolution: From the start of the paper the fact that the inversion is at "high resolution" is mentioned numerous times. This is supposedly a strength of the inversion technique adopted in this study. However, it is not until deep in the Methodology section that the reader finds out exactly what resolution is adopted in both space and time. Can the authors specify exactly what the resolution is early on (even in the Abstract)? Also, can the spatial gridding really be considered to be high resolution (see comment above in comparison to CarbonTracker)?

As described in our response to the earlier comment, we have revised statements in the Introduction and in the beginning of Section 2 to clarify that "high resolution" refers to the flux optimization and is relative to previous batch inversions. Also in the beginning of Section 2, we provide some background on previous batch inversions. We thus believe there is a sufficient level of detail early on to communicate the key point. Also as described earlier, we have now modified the statement about the relatively high grid resolution.

Lines 462-464: It appears that the in-situ only inversion also produces winter-time uptake in Boreal Asia (Fig. 6). This is not just a feature in GOSAT-based inversion.

Why?
It's true that the in situ inversion exhibits a flux with a 1-sigma range entirely below zero in Boreal Asia in December. Given that the in situ inversion is generally noisier than the GOSAT inversion, as we pointed out in line 457, we considered the GOSAT feature more notable and focused on that.

Figs 8 and 12: The different colored bars are hard to distinguish. I suggest separating out the Global flux numbers as a separate panel and zooming in the figure to a smaller range for the regional results.
OK, we have divided each figure into 2 panels as suggested.

Anonymous Referee #2

General comments
This paper describes an impact of satellite observation data on carbon cycle inversion by using multiple settings (observation data sets and prior flux uncertainties) of high-resolution batch Bayesian inversion. The new results are well considered (ex. inversion bias can vary with data coverage). I consider this article should be acceptable after some minor revisions for publications for ACP. One important issue is that the number of observation sites (87 sites (only 10 continuous sites for in situ inversion) is considered insufficient to constrain 108 regional CO2 flux. The inversion results (chisquared value, dipole behavior, mismatch against independent observation, etc.) show this issue. One option to avoid this issue is to add observation sites (JR station data and amazonica aircraft data) for in situ inversion. The other issue is inadequate description of satellite retrieval bias. I consider the difference between in situ inversion and GOSAT inversion comes from not only satellite sampling bias but also satellite retrieval bias. The authors should discuss retrieval bias of satellite from validation of multiple inversion results and show some choices. Because modification of satellite biases (sampling and retrieval) is significantly important to the future use of satellite observation data in carbon cycle analysis.

We thank the reviewer for the insightful comments and the positive overall evaluation. We agree that using a larger number of sites would allow for better constraints on fluxes and reduce some of the posterior flux error correlations seen in the in situ inversion. Note that the number of sites, 87, is similar to that of CarbonTracker 2013B, as we mentioned in the paper, and actually larger than most of the models in the Houweling et al. (2015) inversion intercomparison, many of which had even higher flux optimization resolution than our inversion. The reviewer's suggestion of using JR-STATION and Amazonica data in the inversion is a good one (though the latter observations were made during only a part of our analysis period and at a lower frequency--about once every 2 weeks--than the 8-day resolution of our inversion). In fact, we pointed out in our Discussion and Conclusions that those data sets, and others, could also be incorporated in the inversion. We had instead reserved them for use as independent evaluation data. For example, in the discussion of the 2010-2009 boreal summer flux differences, we chose to focus on the constraints provided by GOSAT (the in situ inversion is not shown in Figs. 16 and 17 and Table 3), and used JR-STATION data to evaluate the GOSAT inversion over Siberia. Note that some regions of the world, such as tropical Africa, are not well sampled by any long-term in situ networks. A final comment is that we delineated the flux optimization regions based not only on the distribution of in situ sites but also on the higher density of the GOSAT observations, with the expectation that the latter would provide greater reductions in flux uncertainties. We have now added the phrase, "which allows us to take advantage of the relatively high density of the GOSAT observations", in Section 2.4 after describing our large number of regions compared to previous batch inversions.

As for retrieval bias, we now cite that in more places than in our original manuscript (details are given below). Note that effects of retrieval bias and sampling bias are often intertwined in inversion results. Since there is a lack of evaluation data in certain important regions, e.g. Africa, we highlight this as a need for improved carbon cycle studies in the future in the Discussion and Conclusions.

Specific comments

Line 38: The authors should also mention the influence of satellite observation errors on this sentence.

In this part of the abstract, we are discussing how the coverage of the in situ and GOSAT data sets affects their ability to constrain fluxes, as reflected in posterior flux uncertainties and correlations. Satellite retrieval errors do not affect the inversion-generated error covariance estimates, since the Bayesian, least-squares method used assumes Gaussian error distributions with no bias. Thus, the retrieval errors are not relevant in the statement about the spatial scales at which the data sets can resolve fluxes. However, we do discuss the effects of satellite retrieval biases soon after that in lines 42-44.

Line 54: At 2010, a relatively high temperatures around eastern North America event (all year round) occurred. This event might have affected greater uptake over the region.

High temperatures could indeed have contributed to greater uptake, especially at higher latitudes, where insufficient warmth could be more of a limiting factor for NEP than insufficient moisture during late spring-early summer. We looked at plots of the 2010-2009 flux difference in May-June for the inversions, and did see some areas of increased uptake in 2010 at high latitudes (including in the east), although the geographic pattern was not entirely consistent across the prior, in situ posterior, and GOSAT posterior. We now mention in Section 3.4 the possible contribution of higher temperatures to increased uptake especially at higher latitudes.

Line 68: The authors should refer to the high precision feature of observation data in this sentence.

OK, we have modified the sentence after that one to include that feature: "However, the accuracy of top-down methods is limited by incomplete data coverage (especially for highly precise but sparse in situ observation networks),…." (We modified that sentence rather than the one in line 68, given that high precision is not a general feature of top-down estimation, applying only to in situ mole fraction observations and not to current satellite retrievals.)

Line 178: In the experimental settings, natural (biosphere and ocean) net sink is too small comparing with current knowledge. This means that it is necessary to redistribute a large amount of CO2 flux by inversion. Errors tend to occur due to transport model, observation data and inversion settings. The author should mention about it.

It's true that the prior fluxes should ideally be consistent with the best available knowledge on a global, annual basis, e.g. a sizable net land sink rather than close to neutral. Note however that the uncertainties we assigned to the prior fluxes are large relative to the possible prior underestimates of the net sinks (especially for land). It can be seen in Fig. 8 for example that the 1-sigma range for the prior global land flux easily encompasses the CarbonTracker posterior estimate of -2.3 Pg C y^{-1} as well as our own posterior estimates using the in situ and GOSAT data sets. In fact, the GOSAT and in situ + GOSAT inversions infer a larger *source* than the prior, so that CO2 is actually redistributed in the opposite direction. And for the global oceans, the 1-sigma range for our prior encompasses the CT posterior. Interestingly, the prior is not strong enough of a constraint to prevent any of our three inversions' ocean fluxes from exceeding the lower bound of the prior range. We now point out in the manuscript that the prior uncertainties are large enough to accommodate possible biases in the prior fluxes.

Line 204: 87 sites seem to be insufficient comparing with 108 regions (numbers of continuous sites are only 10 and geographically unevenly distributed). The authors should use more observation sites to constrain these regional CO2 flux.
Please see our response to the general comments above.

Line 226: The minimum value (0.01ppm) of the standard deviation of the observations within a particular looks too small. The authors should clarify the reason.
We have now added the phrase “to avoid uncertainty values of 0” for explanation.

Line 307: The authors should show a thickness of the lowest model layer.
We now specify that the thickness of the lowest layer is ~100 m on average.

Line 308: The authors should show numbers of dimensions of matrix (especially GOSAT and in situ + GOSAT inversion).
OK, we now show the dimensions for the GOSAT and in situ + GOSAT inversions (at the end of Section 2.4).

Line 392: The authors should also mention the bias of satellite observation.
We were actually referring to bias when we stated that “the in situ and GOSAT data sets are not necessarily consistent with each other”, and earlier in the same sentence, “Comparison of posterior mole fractions with the data set not used (Fig. 3b, c), on the other hand, gives mean differences not as close to 0 as in the comparison with the assimilated data”. Elsewhere in the paper, we make it clear that the bias occurs in the satellite observations.

Line 398: The authors should discuss the reason why chi-square of GOSAT inversions is almost same.
The chi-squared for the GOSAT tight-priors case is slightly larger than that for the baseline inversion, 0.823 vs. 0.778 (the tight-priors case for the in situ inversion also has a larger chi-squared than the baseline), but the numbers both round to 0.8 for one decimal place, which is what we show in the Table. We now add a note in the discussion of the results for the tight-priors cases (in Section 3.3) about the difference in chi-squared being concealed by the rounding.

Line 422-425: The authors need multifaceted discussions. It is difficult for this inversion setting (PCTM horizontal resolution (about 200km) and 8 day mean flux) to reproduce CO2 concentrations near surface at regions where vegetation activities are active like tropical rainforests.
Our prior biospheric fluxes actually have 3-hourly temporal resolution and our transport model has hourly resolution, so the model can potentially reproduce some of the short-term variability in the aircraft observations. (We now add text to the transport model description section to highlight the hourly resolution.) But it is true that our inversion is designed to improve the fit to observations at broad spatial and temporal scales and not at small scales. One factor that likely contributes to the posterior model-observation differences having greater variance than those for the prior at low altitudes is that the GOSAT data are sparse over the Amazon, so that there is little data averaging over the 8-day intervals and flux regions. Thus, random errors can have a substantial impact. We failed to mention this in the text, so we now add it. Also, our original statement about the GOSAT random errors not being correlated with those of the aircraft

observations is obvious and actually irrelevant, since the aircraft observations, based on flask samples, have negligibly small measurement errors. We now omit that statement.

Line 434-437: The authors should show the usefulness of increasing the number of observation data.
The focus in this paper is on a comparison of inversions using satellite and in situ data. As such, we believe it is sufficient to show the usefulness of the larger number of observations provided by GOSAT without presenting multiple in situ inversions using different numbers of observations. Please also refer to our response to the General Comments above for a discussion of the number of sites used in the in situ inversion.

Line 448-453: The authors should show degrees of freedom for signal and noise for in situ + GOSAT inversion as previous paragraph.
OK, we have now added those numbers and some text commenting on them.

Line 537: The authors should unify expression (prior/in situ/GOSAT) for N. Pacific and N. Atlantic.
We intentionally left out the priors for those regions because the GOSAT inversion differs from only the in situ inversion and not the prior over northern oceans, as stated in the sentence before that line. We wanted to show only the relevant numbers here.

Line 572: The lack of ocean observations at southern high latitudes brings analysis results closer to a priori information. The authors should consider satellite retrieval bias.
Our original thinking was that the underestimate of the GOSAT inversion relative to HIPPO in the southern extratropics reflected a negative correlation of the posterior fluxes with excessively positive fluxes in the tropics, which is caused by insufficient observational constraints. But now we agree that retrieval bias likely plays a role in the region, since the magnitude of the discrepancies relative to HIPPO, as much as several ppm, is probably too large to be caused solely by sampling bias. Note that sampling bias still plays a role (via error correlations), since the posterior mole fractions are worse than the prior ones even where there are no GOSAT observations. We now add retrieval bias as a factor in the text here.

Line 577-579: 67ppm difference seems too large. The author should identify and remove the cause observation data from inversion.
We agree the difference is unreasonable but do not find specific reason to omit the comparison for that particular latitude bin. We suggested that both the in situ and GOSAT inversions may be *under-constrained* in that region and season (high-latitude North Pacific and Alaska in the fall), so removing data in the inversions would be unlikely to help. Also, we pointed out that the prior is substantially higher than HIPPO here as well, and suggested that transport error might be a common cause for all the discrepancies.

Line 586-588: In general, current transport models could not well reproduce tropopause. The authors should use only tropospheric HIPPO data in figure 10 (c,f) for discussion.
Although many transport models have problems in simulating the tropopause, we expect PCTM to perform well, given that it has high vertical resolution (with more than 20 levels in the stratosphere) and the tropopause in the underlying GEOS-5/MERRA meteorological data assimilation system is considered to be accurate. Evidence is provided by the TransCom-CH_4 evaluation of models of Patra et al. (2011) and the ozone tropopause transport analysis of Wargan et al. (2015). (Wargan, K., S. Pawson, M. A. Olsen,J. C. Witte, A. R. Douglass, J. R.

Ziemke,S. E. Strahan, and J. E. Nielsen (2015), The global structure of upper troposphere-lower stratosphere ozone in GEOS-5: A multiyear assimilation of EOS Aura data, J. Geophys. Res. Atmos., 120, 2013–2036, doi:10.1002/2014JD022493.) Thus, we think it is reasonable to include HIPPO data in the upper troposphere-lower stratosphere region as part of the evaluation of the inversion results. We have added a note in the text justifying the inclusion of the UTLS data.

Line 642-643: It seems that the figure and explanation sentences do not match, so more detailed explanation is needed.
We think the source of confusion for the reviewer is perhaps our lack of a definition for "South America", so we now specify in the text which regions in Fig. 11 correspond to "South America and Africa": "Trop Am", "Temp S Am", "N Africa", and "S Africa".

Line 644: The larger anti-correlations is visible among land area (Bor. N. America and Temp. N. America, Trop. America and Temp. S. America, N. Africa and S. Africa, Temp. Asia and Trop. Asia). The authors should discuss such anti-correlations.
That is a keen observation (the GOSAT inversion exhibits negative correlations of larger magnitude than the in situ inversion for those regions). Note that the point we were trying to make was that "There are a larger *number* of sizable correlations between land regions in the in situ inversion than in the GOSAT inversion." But to provide a more complete discussion, we now add a comment about the larger magnitude of the GOSAT correlations over land, speculating that although GOSAT observations are of higher density over many regions, the column averages tend to reflect mixtures of air from a broader source region than for in situ observations, and may thus result in larger error correlations for immediately adjacent regions.

Line 756: A decreased sink in parts of North America (Eurasia) almost matches high temperature anomaly area at 2010 summer. The authors should mention this point in this paragraph. Amazonica aircraft data also could constrain tropical America CO2 flux. https://www.ncdc.noaa.gov/sotc/global/201007

We had discussed the heat waves and droughts of summer 2010 in North America (and Eurasia) in the previous paragraph (citing Guerlet et al., 2013). So it would be repetitive to mention the temperature anomaly in this paragraph.

As for Tropical America, please see our response to the general comments above, including the point about the Amazonica data being available for only a part of our analysis period (namely 2010, so they provide no information on 2010-2009 differences).

Line 767: Measurements from the JR-STATION are significantly important to constrain Eurasia CO2 flux. The authors should include these data to inversion.
Please see our response to the general comments above.

Line 860: The author should consider possibility of using CONTRAIL dataset.
We did suggest that various aircraft data sets could be used as input, although we did not list specific ones besides Amazonica. We now specifically mention CONTRAIL, as well as NOAA's regular aircraft profiles over mostly North America.

Line 863: Satellite retrieval bias also reflected in this paragraph.
This is the same issue as for Line 38; please see our response above.

Line 889-891: GOSAT TIR retrieval also could provide high latitude winter observation. The authors should mention it.

OK, we have added the following sentence after those lines: “Ongoing development of thermal IR (TIR) CO_2 retrievals for GOSAT and the future GOSAT-2 with sensitivity to several layers from the lower troposphere to the lower stratosphere shows promise for producing sufficiently accurate data that could also help to fill NIR retrieval coverage gaps (Saitoh et al., 2017a; b).”

Figure 1: The authors should show validation sites (JR stations, HIPPO and amazonica aircraft).
We have added the JR-STATION and Amazonica sites to the figure (with a different symbol to represent validation sites). We decided not to show the multiple, irregular HIPPO flight paths for each of two missions on this figure, since they would make the figure rather crowded. We believe that our description of the HIPPO sampling region in the text should be sufficient, and that readers can refer to the HIPPO mission description paper we cited for details and figures.

Figure 5: The authors should show Tropical America to discuss validation against amazon aircraft data.
We appreciate that suggestion. We had carefully selected only three TransCom regions to display in that figure based on space considerations, so we’d prefer not to add another panel. Furthermore, for the purpose of evaluating the model against the aircraft data, it is more illuminating to provide information on both the model and the observations in a plot, as we’ve done in Fig. 4 and S1, than to show only the model fluxes as in Fig. 5.

Figure 6: The number of observation sites should be shown in the bottom of the figures to know how much the region was constrained.
Regional fluxes are constrained by sites both within and outside of the region. A more appropriate measure of constraint are the uncertainty reductions, which are shown in Table 1.

Figure 10 and 13: The authors should remove or mention outlier (stratospheric observation data?) in the figure. The authors should show from (a) to (f) in the figure.
We assume the reviewer is referring to the outliers in panels (c) and (f), which show the upper altitude bin. We don’t see a need to remove outliers, given that we expect our transport model to simulate the tropopause region reasonably well (see response for lines 586-588 above). Also, those particular outliers do not affect our main conclusions about the relative performance of the GOSAT and in situ inversions in that region.
We now add the labels (a) to (f).

Figure 11: The GOSAT inversion seems to enhance dipole phenomena comparing with the in situ inversion. The authors should explain it.
Please see our response for line 644 above.

Anonymous Referee #3

The authors analyze the first year and a quarter of GOSAT column CO2 data (June 2009 - Sept 2010) with a Bayesian synthesis inversion approach, comparing it against a similar inversion of surface in situ CO2 measurements as well as to independent data from the JR-STATION network over Siberia, from the HIPPO transects, and from partial-column CO2 profiles over the Amazon. In the Bayesian synthesis approach, fluxes are estimated across 8-day spans for 108 pre-defined flux regions (obtained by sub-dividing the 22 TransCom3 regions) across late March 2009 through the end of September 2010; the flux patterns assumed inside of each region/span are taken to

be the absolute value of the prior fluxes; transport is given by the PCTM off-line model run at 2.0x2.5 deg resolution (lat/lon) with 56 vertical layers. A key advantage of the Bayesian inversion is that a full-rank covariance matrix is obtained for this discretization, providing accurate estimation errors and correlations for analysis purposes. A disadvantage is that flux patterns inside each 8-day span and region cannot be optimized, leading to possible representation errors.
The paper, though long, provides a clear and careful analysis of this initial period of the GOSAT data, attempting to tease apart the influence of errors in the ACOS v3.4 retrievals used from the true flux signals of interest. I believe it is a useful addition to the existing GOSAT literature and should be published here after a few points of clarification (listed below) are addressed.
The main weakness of the work here, in my view, is that the measurement span addressed is quite short and the influence of errors in the initial conditions are likely to be significant further into the flux analysis span than the 40 days at the beginning of the span that have been discarded here. In particular, the June-August 2009 period used to analyze the impact of the 2010 climate drivers for the northern land regions may be feeling the effects of these spin-up errors, since the inversion span begins on March 22, 2009. It is true that the authors attempt to correct for errors in the initial condition by solving for two scalars (a multiple of the initial pattern and an offset) and this may work well, but I would have liked to have seen some sort of sensitivity study addressing the impact of the initial conditions. The comparison to the JR-STATION data suggests that this impact could be substantial. Also, the 2 PgC/year difference in the global flux total estimated by the GOSAT-only inversion in comparison with the in situ inversion clearly points to the impact of the short inversion span: it might have been better to add an additional constraint on the total (land+ocean) flux solved for in the inversion to prevent this difference, since the data themselves do not contain enough trend information to constrain the total.
Overall, though, a lot of good analysis is presented here. I have made some suggestions below for clarifying certain points in the text.

We thank the reviewer for the insightful comments and positive overall assessment. Regarding spin-up, we did put much effort into ensuring a minimal impact of errors in the initial conditions on posterior fluxes after about April 2009, including examining the results of various sensitivity tests. We do mention our conclusion from these tests in Section 2.4, although we had decided not to include much detail on them. But to address the reviewer's concerns, we now add the following details: "For example, for an in situ inversion in which we did not allow adjustments in the i.c. and offset parameters, 8-day average flux results are very similar to those of the baseline inversion, especially after the first two months, with a mean correlation coefficient of 0.95 from June 2009 onward across all TC3 regions and a mean difference of 0.03 Pg C/yr." Discrepancies between the GOSAT posterior concentrations and the JR-STATION data reflect errors in the GOSAT data and not in the initial conditions, as explained below in the response for lines 767-799.

Regarding a constraint on the total flux, the only information on this available is from atmospheric mole fraction data. We already use most of the long-term surface monitoring sites available in the in situ inversion. And it wouldn't be appropriate to apply a long-term (e.g. several years) average atmospheric growth rate as an inversion constraint, since we are solving for fluxes specific to our analysis period. Also, note that GOSAT observations actually have the advantage of seeing the entire atmospheric column, whereas the in situ observations normally used to estimate the atmospheric growth rate see only the surface or lower troposphere in parts of the world, and thus may not give an accurate estimate for the atmosphere as a whole over a period shorter than the global atmospheric mixing time. On the other hand, GOSAT observations are hampered by the separate problem of retrieval biases. The in situ + GOSAT inversion presumably has the best constrained total flux (though still possibly influenced by GOSAT biases), so our inclusion of those results in the paper gives some indication of the robustness of regional flux results.

Detailed comments:
line 36: Add "Northern Hemisphere" before "high-latitude ocean"? Or do you mean that this applies in the south, as well?
Yes, we do mean both the north and the south, as GOSAT ocean glint observations do not extend into high latitudes in either hemisphere. We have added "northern and southern" before "high-latitude ocean" for clarification.

72: Add "land" before "vegetation"?
OK. (We assume this refers to line *62*.)

96: There is no "Chevallier et al. (2014)" in the reference list. Should the data on this reference be 2013?
We incorrectly wrote "2013" in the reference list. We've changed it to "2014".

112: After "exact solution" add "of the linear equations relating the targeted flux variables to the measurements"? Because the fluxes have been discretized at a fairly coarse spatial and temporal resolution, the approach here does not give an exact solution for the fluxes at fine scales, but it does do so at the coarser resolution targeted here, given the assumed shape of the flux patterns corresponding to each basis function used in the inversion.
Solving for flux scale factors at a coarse resolution actually isn't specific to our batch inversion technique--some of the Kalman filter/smoother methods in use have similar or even coarser spatial resolution (64-156 regions) and temporal resolution (monthly to weekly). Since we are introducing the advantages and disadvantages of the batch technique here, we don't think there's a need to mention the flux resolution.

120-122: A downside of this short span is that much of it may be corrupted by spin-up errors, which may last many months after the start of the inversion span.
Please see our response to the overall/general comments above.

184-185: For clarity, replace "...from fossil and biospheric gases" with "from the oxidation of non-CO2 gases from fossil fuel and biospheric burning"?
We have added "fuel" after "fossil" as suggested. For conciseness, we leave out "oxidation" since it is used in the previous sentence and can be inferred from "chemical production" in the current sentence. We also omit "non-CO2" since it can be inferred from context. "Biospheric burning" is inaccurate as we mention both biogenic and biomass burning gases in the previous sentence. (Examples of biogenic gases include wetland methane and isoprene from vegetation.)

198-200: For the NOAA in situ data, you should give the specific ObsPack file name (which includes the version number) if it came from an ObsPack file, or something equivalent if from some other source.
We did not use an ObsPack file, but rather, the original data files for each individual site. We give the citations recommended by the data providers, namely "Dlugokencky et al., 2013; Andrews et al., 2009", and in the full references we include the version dates.

226: "and apply a minimum value of 0.01 ppm": it is not clear what this phrase indicates.

If there is already an error of 0.3 ppm for the first of up to two possible samples, why is there a need for an additiona 0.01 ppm?
The minimum of 0.01 ppm is needed to avoid standard deviations of 0 for multiple samples (up to "two pairs", or four samples). We now add the explanation: "to avoid uncertainty values of 0".

232: Please indicate before this which measurements come at this 30-second frequency – the Japanese continuous sites?
We indeed failed to indicate that. We now add the phrase "30-second-average continuous" early on when we discuss NOAA data flags in the previous paragraph. (We had already specified "hourly" in the context of the Japanese (JMA) data flags.)

251: add a comma before "other"?
OK, done.

252: Add "all" before "GOSAT"?
OK, done.

309-310: Please give the exact equations that implement what you have described here in words. This is needed, because there are different ways to implement what you describe, and these differences can matter to the inversion.
OK, we now provide the equation corresponding to the words, namely Eq. 15 in the paper by Connor et al. (2008), which we cited in that sentence.

374: Add "assumed" before "well-mixed"? For the time after 13 months, was the pattern obtained at the end of the 13 months used in the Jacobian, or was the completely-mixed value used?
This is a statement rather than an assumption, as the detail we provided after that, "(within a range of 0.01 ppm)", is a quantity based on actual transport model runs. After 13 months, we repeated the pattern at the end of the 13 months until the end of the analysis period.

377-380: It would be useful to describe this SVD procedure in more detail, since discarding singular vectors can completely remove corrections to certain regions at certain times. It might be useful to plot the projection of the singular vectors retained in the fit onto the regions so that the reader can see where the corrections to the prior fluxes are possible and where they are not. What fraction of the original singular value spectrum is truncated and what is retained? Also, usually if one can take the SVD of a matrix, not much more work is required to obtain the full solution: it is not clear how using an SVD approach helps you deal with the large matrix. Please explain this more. What aspect of your SVD approach allows you to handle the otherwise too-large Jacobian successfully?
The SVD technique avoids the inefficient and numerically unstable inversion of a large Jacobian matrix involved in the direct computation of the Bayesian solution. This is described in more detail in the standard inversion literature, including by Rayner et al. (1999), whom we cite. We actually kept all of the singular vectors in the results presented in the paper; as noted by Rayner et al., the SVD technique avoids amplifying the contributions of small singular values, and thus eliminates the need to set a truncation threshold. We now provide more explanation in the text: "A singular value decomposition (SVD) approach is used *instead of direct computation of Eq. 2 and Eq. 3* to obtain a stable inversion solution *without any need for truncation of singular values below a certain threshold* (Rayner et al., 1999)…."

389-390: "gives mean differences not as close to 0 as in the comparison with the assimilated data": They are actually closer to zero for the in situ inversions, but, yes, quite a bit farther from zero for the GOSAT inversions.
We were actually referring to a comparison of Fig. 3b with 3d (not with 3a) and Fig. 3c with 3a (not with 3d). We acknowledge that the original text was ambiguous; we now add the clarifying phrase "(Fig. 3d and 3a, respectively)" after "not as close to 0 as in the comparison with the assimilated data".

392: "...and have independent random errors": how does the fact that the fit to data not used in the inversions is worse allow you to say that the errors are independent?
Our explanation was indeed flawed. We were comparing the posterior differences with those for the prior model when we stated: "standard deviations that are larger than for the prior" in lines 390-391. So a relevant fact is not that the in situ and GOSAT observations are two different data sets with independent errors, but that the prior model generally exhibits less variability than the posterior model, which has assimilated mole fraction observations, due to random instrument/retrieval errors in the observations, the small sample footprint of the observations compared to the model grid and thus larger variability, and noise created by the inversion process (via error correlations). Thus, posterior model-observation differences would be expected to have a larger standard deviation (when the observations are different from the ones assimilated in the model). We now replace the phrase "have independent random errors" with "combine to produce larger standard deviations than with the less variable prior model, which has not assimilated any data".

430: "fractional": it is not clear here what you mean by this – clarify?
We mean fractional (or percentage) as opposed to absolute (e.g. in Pg C y^{-1}). Smaller regions may have large fractional uncertainties despite having relatively small absolute uncertainties. We have added "(percentage)" after "fractional" for clarity.

432-433: "accounting for error correlations": since you are just aggregating means instead of uncertainties, it is not clear why you need to worry about error correlations – why do you mention it?
We are aggregating "results" here, which include both means and uncertainties.

434-439: Since you have the exact covariances for each region, you could aggregate these (accounting for correlations) and get a posteriori uncertainties for these larger regions. Then you could compare the observed variability to these to see whether random estimation uncertainties do indeed account for this variability or not. Doing this would be better than just speculating, as you do now.
We did aggregate the exact covariances up to larger regions (perhaps see the response before this one for lines 432-433). Also, the estimation uncertainty for a particular 8-day mean is not really equivalent to an estimate of the variability across different 8-day periods. Finally, we do go beyond speculation, as we discuss the calculated degrees of freedom for signal and noise in that paragraph.

443-447: It is important that you mention here that the observed variability could also be due to systematic errors in either the measurements (especially for the GOSAT case) or in the transport model (especially for the in situ case). By computing the expected random error from your a posteriori covariance matrix, you could potentially rule out random error as the cause, allowing you to attribute the new variability to either a real flux signal or to systematic errors. This is a key reason why you should use your covariance matrix calculation in this analysis.

We did compute posterior uncertainties for aggregated regions, please see the two responses before this one. We also attributed the variability to either signal or noise, as explained in the response before this one. Systematic errors are not as relevant here, since we are discussing the amount of fluctuation from 8-day period to 8-day period rather than biases sustained over longer periods.

446: The dipole behavior mentioned here, in particular, would be reflected in the covariance matrix, if that is in fact the cause of much of the variability.
As described in the three responses before this one, we did utilize the covariance matrices. And as stated in line 447, we discuss dipole behavior associated with negative error correlations later in the paper. Also, note that the numbers of degrees of freedom for signal and noise reported in this paragraph are calculated using the posterior covariance matrix, in a manner described by Rodgers (2000), whom we cite.

474: It might be helpful to mention here in the text that you are comparing your in situ results (not GOSAT results) to CarbonTracker, which also uses only in situ measurements.
We do actually mention that we are comparing our in situ-only inversion with CarbonTracker in this paragraph, in lines 489-490.

489: The sentence starting with "Results" could perhaps be deleted to save space, as it repeats the first sentence of the paragraph.
We appreciate this suggestion. However, we intentionally separated the information about Fig. 7 into the two sentences, since we did not want to include too much detail about the figure in the very first sentence of the paragraph. The two sentences contain distinct information, as the second one specifies that we are comparing only our in situ inversion with CarbonTracker (we had just described the characteristics of the CarbonTracker system in the last few sentences) and that we are showing large regions (aggregated from TransCom regions).

499-502: This is another place where the text could be compacted somewhat – it seems repetitive.
We have now tightened the text here to make it sound less repetitive.

Fig.8 Caption and elsewhere: To avoid having to use the "NEP (x-1)" phraseology everywhere, why not just say you are solving for NEE (which is approximately equal to -NEP)?
NEE includes other fluxes such as emissions from fires and other disturbances; since we prescribed fire emissions and only optimized NEP in the inversions, we needed to be precise in terminology when showing only the NEP component in figures, e.g. Fig. 5 but not Fig. 8.

515-516: "Such a large difference ... is plausible": what evidence can you give to back up your assertion? You have pointed to some plausible causes for the difference, but even so, the difference seems larger than expected. Why did Houweling et al get a difference that was an order of magnitude lower for inversions across a similar span?
For the Houweling et al. results, we gave the total fluxes averaged across 8 models. Within those averages is a substantial amount of inter-model variability. For example, Basu et al. (2013), whose model was one of those included in the Houweling et al. study, reported a 12-month global flux that is ~1.3 Pg C/y less negative for their GOSAT inversion than for their in situ inversion (Sep 2009-Aug 2010, RemoTeC GOSAT retrieval). And Chevallier et al. (2014), who reported results from two of the models included by Houweling et al., found *more negative* global fluxes, by up to 1 Pg C/y, in their GOSAT inversions (ACOS and UoL retrievals) than in their in situ inversions, though for a different 12-month period, 2010. Overall, the results show

that the global growth rate can differ substantially over a short period based on either GOSAT or surface observations and among different models. We did show in our paper how the total flux from the inversion can be very sensitive to the time period considered for a 12-month time frame. Also, note that GOSAT column measurements may actually sample the atmosphere better for constraining the short-term growth rate than the surface network, but on the other hand, the result may be affected by retrieval biases. To address the reviewer's concern about the smaller difference in the Houweling et al. results, we now replace "though the difference may not be statistically significant" in lines 523-524 with "with a substantial amount of inter-model variability within those averages".

line 538 and Figure 8: You have used the same term, "Southern Ocean", to refer to the true Southern Ocean (as defined, for example, by TransCom3 as everything south of about 45 deg S) as well as the extra-tropical southern oceans (everything south of 23 deg S). I suggest changing what you call the latter area for clarity.
OK, we now use the phrase "extratropical southern oceans" in the text to refer to the latter. (Likewise with "extratropical northern oceans.") In the figures, for compactness we use the phrase "Southern Oceans," with the "s" at the end to distinguish it from the TransCom "Southern Ocean." (For consistency we add "s" at the end of the other ocean aggregations as well.) Note that this makes Fig. 8 consistent with Fig. 7, in which we had already used plural forms.

568, 571, and Fig 10: The text refers to sub-panels of Figure 10 (a-e), but these labels are absent from the actual figure – please add these labels on the figure.
We now add the labels (a) to (f). (Also for Fig. 13.)

568-570: "Evaluation of the inversions against latitudinal profiles constructed from HIPPO aircraft measurements, which provide additional sampling over the Pacific, indicates an overestimate by the GOSAT inversion relative to HIPPO in parts of the tropics at lower altitudes": my reading of the figure shows only one, maybe two, points from the GOSAT case that are outside of 1 standard deviation of the observations – this certainly does not seem to be a strong feature of the plots, according to my reading of them. It is not until about 40 deg N that the GOSAT results move positive in panel a).
We agree that there is not a general overestimate by the GOSAT inversion in the tropics relative to HIPPO, as there are similar numbers of points higher and lower than HIPPO. We now modify the sentence to read "does not indicate any systematic overestimate by the GOSAT inversion relative to the observations in the tropics (Fig. 10a-f), unlike what was seen in the comparison with the more globally distributed surface observations."

632: Replace "elaborate on the subject of" with "discuss"? Less wordy...
We think "elaborate" is a more precise term here, but we do shorten the phrase to "elaborate on".

695-696: "accounting for the riverine flux, the 1_ range for the in situ inversion overlaps with that of GCP": I believe you are incorrectly applying the riverine flux correction here. The GCP number of -2.5 should be decreased to -2.0 PgC/yr when turning it from an anthropogenic uptake into a total net (anthropogenic+natural) uptake, since the natural cycle (driven by the riverine fluxes into the ocean at the river mouths) has a net 0.5 PgC/yr outgassing – that outgassing counteracts a corresponding amount of anthropogenic uptake, reducing the total uptake to -2.0 PgC/yr. I.e., -2.5 + (+0.5) = -2.0. Given that, both your GOSAT-only and in situ-only ocean uptakes are still outside the 1 sigma ranges for the GCP number.

We actually did apply the riverine flux correction correctly. Two paragraphs before this one, we explained that "The difference between our inversion estimates and the GCP estimate is actually even larger than suggested by those numbers, given that a background river to ocean flux of ~0.5 Pg C y^{-1} should be subtracted from our ocean flux to make it comparable to the GCP ocean sink, which refers to net uptake of *anthropogenic* CO_2". Rather than converting the GCP number to a net uptake, we just left it as an anthropogenic uptake, and implied that 0.5 Pg C y^{-1} should be subtracted from our ocean fluxes for comparison with GCP. (We did not feel it was necessary to explicitly give the adjusted numbers.) We think the source of misunderstanding is that we were comparing the 1-sigma ranges for our estimates with the 1-sigma range of GCP, rather than our central estimates with the 1-sigma range. To clarify, we have modified the text to: "accounting for the riverine flux, the 1σ range for the in situ inversion overlaps with the 1σ range of GCP, while the 1σ range for the GOSAT inversion is still just outside of that of GCP."

Figure 14: I would suggest using some color other than cyan to depict the tight-prior GOSAT results here. As things stand now, it is much too easy to confuse that case with the in situ-only results on Figure 9. One has to read the caption and legend carefully to see that you have changed what is shown in cyan at the moment.
OK, we have changed the color to green.

724: "substantially larger global total budget": it would be clearer to say that the total flux is more positive in the GOSAT case, since "larger" depends on whether the fossil fuel has been added onto the total or not.
OK, we have changed the wording to "substantially more positive global total budget".

734-735: "and an increased source in the tropics of _2 Pg C y-1 in the GOSAT inversion relative to the in situ inversion.": I think that it is important to note that this change from the in situ-only results in the tropics is accompanied by a change in the global total of the same magnitude and sign; in other words, the change is directly related to the fact that the global total is not well-constrained in this short-span inversion. This might be expected to change in an inversion over a longer span, for which the global total is better constrained.
The difference between the in situ and GOSAT budgets in the tropics could indeed be related to insufficient constraints on a 12-month time scale for one or both of them (as we hypothesized for the global budgets), so that the difference could be much smaller when averaged over a span of several years or longer. However, the in situ + GOSAT inversion gives an indication of the impact of an under-constrained global total flux on regional flux estimates, given that the combined data sets provide better constraints than either of the data sets alone. The in situ + GOSAT inversion produces a global flux close to mid-way between the in situ-only and GOSAT-only inversions, while it produces a Tropic Land + Ocean flux much closer to that of the GOSAT inversion than to the in situ inversion. This suggests some degree of independence of the GOSAT-inferred regional result from the global result. We now mention this in the manuscript.

767-799: This whole discussion of the Eurasian source in 2010 and the examination of JR-STATION sites suggests to me that the growing season results in 2009 could well be affected by spin-up issues in the inversion. That could explain why the GOSAT inversion results agree with the data at VGN, AZV, and KRS in 2010, but are too negative in 2009. If that is the explanation, the agreement with the Guerlet (2013) result would be more due to that modeling issue, rather than any real climate-related driver.
Please see our response to the general comments above regarding spin-up issues. In addition, it is important to be aware that inaccurate initial conditions do not affect the fit of the posterior concentrations to observations, but rather, the correctness of the posterior fluxes. The inversion

optimizes the fit to observations through flux adjustments regardless of the i.c., but there would be errors in the final fluxes if the i.c. were incorrect (after adjustment of the i.c. and offset parameters in the case of our inversion setup). Thus, any discrepancies between our GOSAT posterior concentrations and the JR-STATION data would reflect errors in the GOSAT data and not errors in the i.c.

899-901: "Thus, it may not be accurate to assume that year-to-year posterior flux differences are insensitive to satellite retrieval biases, as was done in the other study."
This would be a good place to note that spin-up errors in this study (as well as the Houweling study) could also be adversely affecting the 2009 flux results, as well as the 2010-2009 shift.
Please see our response to the previous comment (on lines 767-799).