We thank the referee for carefully reviewing our manuscript. We have accepted several of his suggestions, which we believe have significantly improved the manuscript. The referee had two major general comments, (i) the results presented in §4.2 and §4.4 were mostly qualitative, and (ii) it was difficult to draw conclusions about total column vs planetary boundary layer (PBL) measurements from the experiments because of differing spatiotemporal coverage.

Regarding the qualitative nature of results in §4, it is difficult to present general quantitative results in a work like this, because the impact of transport model uncertainty varies by region, time and data stream. Previous work that tackled the question of assessing transport model uncertainty, such as Locatelli et al (2013), also struggled with drawing quantitative yet general conclusions. The referee mentions comparing our transport-derived uncertainties with posterior uncertainties of the flux estimates. This is also not robust, since the posterior uncertainty from TM5 4DVAR is an overestimate (Meirink et al, 2008), a problem common to most iterative flux inversion techniques. Instead, this work considers the cross-model spread of OCO-2 inverse models (Crowell et al, 2018) a measure of the uncertainty of our knowledge of inversion-derived surface fluxes, and tries to estimate whether those uncertainties are consistent with what we would expect just from transport model uncertainty.

However, we agree that our existing results could be presented in a more quantitative form. We have added several tables in the text to facilitate this. Table 2 shows the number of observations assimilated from each data stream, Table 3 gives the uncertainty (spread across five transport models) in the global budget and its partitioning (land/ocean or latitude band), and Table C1 gives the uncertainty in the annual flux from all the geographical regions considered, along with the prior and true fluxes. Comparing the uncertainty with the true flux gives an idea of the significance of the uncertainty.

Regarding the differing coverage of OCO-2 and in situ measurements, this is a very good point. To make a clearer distinction between the impact of total column measurement (vs PBL) and the impact of a spatially distributed sampling pattern, we created two hypothetical in situ networks. IS-LNLG (IS-OG) had PBL samples 30m above ground level at the times and locations of all OCO-2 land (ocean) soundings used in the LNLG (OG) inversions. A comparison between LNLG and IS-LNLG (OG and IS-OG) inversions, therefore, reveal the impact of having total column vs PBL measurements over land (ocean), and not differences in spatiotemporal coverage. We have replaced our figures and tables to include these new (hypothetical) data streams, and have reworked our results and conclusions to incorporate the new results. The short summary is that over most land regions, total column samples do lower the transport-driven uncertainty in flux estimates compared to PBL samples. This holds less strictly over ocean regions, likely due to lower convective fluxes (and hence lower model to model differences). The land/ocean partitioning within a zonal band is more uncertain with land PBL samples, but the aggregate over the zonal band is not. Flying a remote sensing instrument with higher PBL sensitivity has been a goal of space-based greenhouse gas missions (Wang et al, 2014). Our results suggest that if such an instrument were to fly, the uncertainty in transport modeling would become a severe bottleneck, and considerable improvement in transport modeling would be needed before we could use such an instrument to improve on the precision of estimated surface fluxes.

<u>Specific comments</u>

"What motivates the chosen time span?"

The time span was motivated by two factors. (a) When this study was initiated (late 2016), OCO-2 data were available up to July 2016. Allowing for some spin-up and spin down, this left 2015 as the only full calendar year we could address. (b) The initial goal of the OCO2 model intercomparison project (Crowell et al, 2018) was to perform inversions to estimate and compare flux estimates for 2015. Since this work was supposed to help them test the robustness of their conclusions, it made sense to perform our work over the same time period. Having said that, we certainly want to extend this work to at least three years in the near future, to study questions such as trend and interannual variability that cannot be addressed with one year's fluxes.

"Is daytime sampling used for marine background sites also?"

Yes, the majority of sites in the MBL network were sampled in the local mid-afternoon. Mountaintop sites such as Mauna Loa were sampled in the early morning in both the IS and MBL networks to reduce the possibility of updrafts. In fact, the samples in the MBL network are a subset of those in the IS network. We simply chose those samples in the IS network that belonged to sites used by Baker et al (2006).

"Why were posterior fluxes from CarbonTracker chosen as prior? They are not independent from the data that are used to derived the truth ... to me it seems more logical to take the CarbonTracker prior. How consistent is the choice of prior covariances in this case?"

The CT posterior was chosen as the prior for two reasons. (a) The CT prior does not have a net ecosystem sink, and using such as obviously biased prior guarantees a biased posterior in an inversion. This is why some long term inversions evaluate their fluxes with respect to fluxes that already guarantee the correct atmospheric $CO_2$ trend (Chevallier et al, 2010). (b) Several inversions in the OCO2 model intercomparison project (Crowell et al, 2018) is also using a climatological CT posterior as the prior flux, and our goal was to make our experiments maximally relevant to that effort. We also note here that the 2000-2015 average posterior, which we used as the prior, will have little information specifically from 2015 observations. In any case, our conclusions primarily concern transport-driven uncertainty, which is not expected to be sensitive to the choice of prior (a fact the referee notes later).

The question of the prior covariance is an interesting one. In our system we specify the prior error as a fraction of the CASA heterotrophic respiration and not the NEE. As such, it is not strongly coupled to our choice of the NEE; the fractional change in the NEE from prior to posterior is a very small change in comparison to the heterotrophic respiration. Moreover, since our prior uncertainty uses the same CASA vegetation map as CT, it is guaranteed to be large (small) where CT thinks there is a lot of (no) ecosystem activity.

"What is Nret typically? Does epsilon^2/Nret yield a realistic systematic error?"

$N_{ret}$ can be anywhere between 1 and 24. See Figure 1 for histograms of $N_{ret}$ over land and ocean. To answer the second part of the question, we plotted the standard deviation of $XCO_2$ coming from the retrievals in 1s bins in Figure 1. The purpose of the epsilon$^2$/$N_{ret}$ is to prevent the value of this spread from getting too low, as might be the case when there are only a couple of shots in the bin and they happen to have close to the same $XCO_2$ value (or the extreme case, when there is only a single shot, in which case the standard deviation is zero).

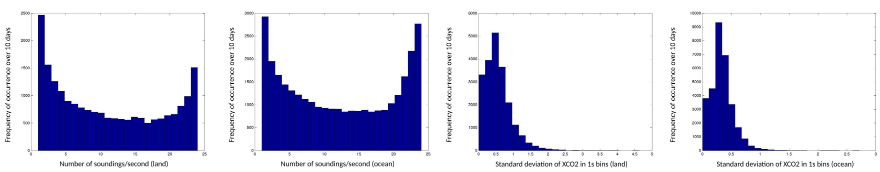


**Figure 1:** Frequency distribution of $N_{ret}$ over land (far left) and ocean (near left) over ten days chosen randomly from the OCO2 record. Over land (LN+LG), $N_{ret}$ = 1 19% of the time, $N_{ret}$ = 2 7.5% of the time, etc. The standard deviation of retrieved $XCO_2$ in 1s bins over land and ocean are shown in near right and far right respectively.

"Even if all the models had the same random noise added to the data, this would not have changed the inter model spread in the fluxes. However, if the importance of transport model uncertainty is assessed in relation to overall posterior flux uncertainty then measurement uncertainties do matter (whether or not you would add random noise to the data in this case depends on the method for calculating posterior flux uncertainties)."

This is true. However, as explained earlier, we are not comparing the transport model uncertainty to the analytical posterior flux uncertainty from any single inversion, because our system cannot provide a robust estimate of that. Rather, the goal is to use these transport uncertainties as guides when comparing flux estimates (not their uncertainties) from different assimilation systems. And usually an assimilation system does not add random noise to the measurements.

"So, in this case the difference between 2 models doesn't even depend on the choice of prior flux. This means that my earlier remark about the use of CarbonTracker posterior fluxes actually doesn't matter. It would still be useful to point out that the prior fluxes that are described in detail aren't really relevant to the problem.. well, they are to the extent that the a priori fluxes are used to define a priori flux uncertainties. Some further sentences clarifying this would be useful."

This is a good point, and we have added a sentence clarifying this.

"what do you mean by 'lateral' grid cell? Each individual grid cell?"

Yes. We had used the term 'lateral grid' to distinguish it from the 'vertical grid', but realize that that is clear enough from the context. We have deleted the word 'lateral' from that sentence.

"'due to chance' you mean 'due to differences in transport'?"

What we meant was that the proximity of the flux from the TM5 data stream to the true flux reflected the posterior uncertainty of our inversion system. If a different model's result happened to be closer to the true flux for some region, it should not be taken to mean that that model somehow provided "better" than perfect transport. However, we agree that meaning was not clear. We have changed that sentence to read "should not be interpreted as significant".

"but since the sampling is also very different between surface and satellite there is no way to isolate the impact of PBL versus total column."

See our description above of the hypothetical networks IS-LNLG and IS-OG. We have added several inversions with these networks in the revised version.

"I wonder if this difference between this study and Baker et al (2006) could be influenced by the choice of an El Nino year for the current inter-comparison, which may not be well representative of a typical year (so my earlier remark about justifying the chosen times window)."

That is possible, but it's not clear to us why tropical and temperate transport uncertainties would change differently in an El Nino year. This is definitely something to look at when we extend out study to multiple years.

"but could also be due to a more even sampling coverage."

This has now been addressed with our new model runs.

Technical corrections

All three technical corrections have been implemented.

References

D. F. Baker, R. M. Law, K. R. Gurney, P. Rayner, P. Peylin, A. S. Denning, P. Bousquet, L. Bruhwiler, Y.-H. Chen, P. Ciais, I. Y. Fung, M. Heimann, J. John, T. Maki, S. Maksyutov, K. Masarie, M. Prather, B. Pak, S. Taguchi, and Z. Zhu, "TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO2 fluxes, 1988-2003," Glob. Biogeochem. Cycles, vol. 20, no. 1, p. GB1002, Jan. 2006.

F. Chevallier, P. Ciais, T. J. Conway, T. Aalto, B. E. Anderson, P. Bousquet, E. G. Brunke, L. Ciattaglia, Y. Esaki, M. Fröhlich, A. Gomez, A. J. Gomez-Pelaez, L. Haszpra, P. B. Krummel, R. L. Langenfelds, M. Leuenberger, T. Machida, F. Maignan, H. Matsueda, J. A. Morguí, H. Mukai, T. Nakazawa, P. Peylin, M. Ramonet, L. Rivier, Y. Sawa, M. Schmidt, L. P. Steele, S. A. Vay, A. T. Vermeulen, S. Wofsy, and D. Worthy, "$CO_2$ surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements," J. Geophys. Res., vol. 115, no. D21, pp. D21307–D21307, Nov. 2010.

S. Crowell, et al, "The 2015-2016 Carbon Cycle As Seen from OCO-2 and the Global In Situ Network", in prep (2018).

R. Locatelli, P. Bousquet, F. Chevallier, A. Fortems-Cheney, S. Szopa, M. Saunois, A. Agusti-Panareda, D. Bergmann, H. Bian, P. Cameron-Smith, M. P. Chipperfield, E. Gloor, S. Houweling, S. R. Kawa, M. Krol, P. K. Patra, R. G. Prinn, M. Rigby, R. Saito, and C. Wilson, "Impact of transport model errors on the global and regional methane emissions estimated by inverse modelling," Atmos. Chem. Phys., vol. 13, no. 19, pp. 9917–9937, Oct. 2013.

J. F. Meirink, P. Bergamaschi, and M. C. Krol, "Four-dimensional variational data assimilation for inverse modelling of atmospheric methane emissions: method and comparison with synthesis inversion," Atmos. Chem. Phys., vol. 8, pp. 6341–6353, Jun. 2008.

J. S. Wang, S. R. Kawa, J. Eluszkiewicz, D. F. Baker, M. Mountain, J. Henderson, T. Nehrkorn, and T. S. Zaccheo, "A regional $CO_2$ observing system simulation experiment for the ASCENDS satellite mission," Atmos. Chem. Phys., vol. 14, no. 23, pp. 12897–12914, 2014.