Reply to Reviewer

Thank you for reviewing our paper.

General comments:

1. Throughout the paper, the expression "drawdown" is used when what the authors really mean is "negative anomalies".

"Drawdown" refers to the variation amplitude in the seasonal cycle that mostly resulted from strong plants activity. We will clarify this point in the revised paper.

2. Additionally, the reader might be lead to the somewhat implicit conclusion that XCO2 can't serve as a constraint on surface fluxes, but this might be an artifact of the multimonth averaging that is used in the analysis in later sections. At least some discussion of the ability of models such as CT to capture transient features needs to be put forth, because these transient spatial gradients could indeed be attributed to fluxes given sufficient accuracy and adequate transport modeling, as has been shown in numerous OSSEs (Liu et al [2014], Miller et al [2007], Rayner and O'Brien [2001], to name just a few). Connecting the analysis in this paper to those earlier studies is critical to readers trying to assess the conclusions of inversion work with OCO-2 and GOSAT.

We agree with the reviewer that transient spatial gradients could indeed be attributed to fluxes. The ability to capture transient feature is important criteria for modeling performance. Transient features, for example, the seasonal cycle, are generally easier to model since the signals are much larger compared with annual average. However, very small biases in seasonal cycle can still cause drastically biased annual fluxes and unrealistic compensating fluxes. The ability to capture transient features and perfect transport in modelling are not enough. We think it is critical to reduce biases in measurements as inputs for inversions. Previous study by Masarie et al. (2011) had evaluated the impact of CO2 measurement bias on CarbonTracker flux estimates and found that 1 ppm bias at one site, the Park Fall ,Wisconsin (LEF site in our study), can cause 68 Tg C/yr bias in flux estimate for Temperate North America (~ 10% of the estimated North American annual terrestrial uptake). Flux estimate errors are also found in Europe and boreal Eurasia to compensate for the errors in North America.

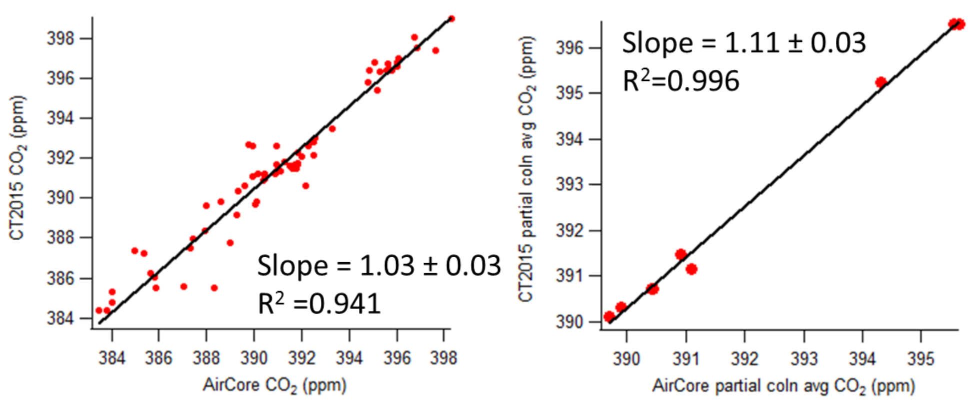
Whether or not the column CO2 can serve as a good constraint on surface fluxes really depends on the biases in column CO2 retrievals. High accuracy is needed from column CO2 product to be useful to constrain surface fluxes because column CO2 is a total column average, thus it is not as sensitive to surface fluxes as surface measurements. For example, a simple mass balance argument shows that all U.S. CO2 emissions from fossil fuel burning (~1.4 Pg yr-1) create a total column enhancement of only **0.6 ppm** on average in air parcels over the East Coast compared to the West Coast and Gulf Coast if we assume a residence time of the emissions of 5 days to pass the contiguous U.S. (~8×10^{12} m2). From the noise to signal ratio perspective, it is important to have high accuracy in column CO2 products. Considering current state of biases in remote sensing products, we think extensive works are needed to reduce biases. This can be done given enough well-calibrated surface measurements and vertical profiles.

Our study provides an approach to check whether column CO2 retrievals are satisfying the high accuracy requirement on some levels. Since spatial gradient is the core of deriving reliable regional fluxes, we look into the spatial gradient in long-term multimonth averages, which should be easier for models and column CO2products to achieve compared with some short-term metrics (e.g. diurnal cycles).

Specific comments:

1. Figure 2: Are there statistics of goodness of fit? These two examples may not be representative.

We have 9 independent AirCore profiles during study period, which are all sampled nearby CAR and SGP sites. That's why we show one sample from each site. We will clarify this point in the revised paper. The flowing figures show AirCore vs CT2015 modelled CO_2 above 330hPa. The left panel shows direct point to point comparisons from all 9 profiles at the vertical levels of CT2015; the right panel shows partial column (0- 330hpa) averages. Due to the limited amount of AirCore profiles, we prefer not to put too much discussion of estimate uncertainty in our paper. In addition, the upper 1/3 of the atmosphere is not important when we are looking at long-term averaged spatial gradients in total column because there is little spatial variability in the upper atmosphere. We can see the fit is generally good even when we are comparing at profile by profile basis without temporal averaging.



2. Line 207: Can you explain the word "random" here? If all of the profiles are given the same weight in the sampling distribution, then this isn't really a measure of uncertainty, but rather a weighted standard deviation that would be extremely sensitive to the 100 particles you selected. If there were a "prior" uncertainty placed on each profile, how was that done?

'random' means all of the profiles are given the same weight in the Monte Carlo resampling, and the resulting 100 column averages values have a relatively large range since this value are sensitive to the combination of the profiles. That is what we want to use to represent the uncertainty, to account for the atmospheric variability without assuming a giving distribution of the vertical profiles.

A ‘prior’ uncertainty is likely to represent mostly the measurement uncertainty, which is too small without fully consider the atmospheric variability.

3. Line 216: There is also some increase due to the shallow PBL alone. It would be good to know what fraction is from the enhancement is due just to boundary layer dynamics. Similarly for the summer.

We agree with the reviewer. However, putting a number on the fractions of enhancement from either changes of PBL or flux requires model with good PBL dynamics and reasonable flux estimates. Good PBL simulations are still very challenging for models. We think it is beyond the scope of this study.

4. Line 228: The PBL height is different at different locations, and through different seasons. How much is the chosen division of the atmosphere mis-attributing boundary layer into the free troposphere, and vice versa? This might be a small detail, but it does impact the conclusions later about the seasonal strength of sources and sinks by region.

For the < 2 km measurements, we believe both PBL and fluxes are important driving factors for the signals we observe. By choose a giving height (2 km), we have removed some of the PBL effect by compensating the PBL with some free-troposphere air. If we use actual temporally changing PBL height as threshold, we will see stronger influences of PBL in wintertime as the CO_2 levels in shallow PBL are even higher compared to our approach. At this moment, we cannot estimate the total influence of PBL and completely remove it without using a model, which may also have big uncertainty in PBL estimates. Our study does not focus on the discussion of the actual number of the fluxes.

5. Line 257: "The SE region also demonstrates a less pronounced seasonal cycle with weaker summer drawdown compared with other northern regions, which may due to the sea-breeze influence in summer within PBL.” Is this a statement about the actual impact of the sea breeze on the fluxes, or is it an assertion that we can’t interpret the column due to the meteorology?

We state the possible impact of sea breeze on the data and the gradient without implying the interpretation of column data. Given a model with good performance on sea-land breeze and sufficient accuracy in column CO2, we should be able to interpret the column data that influenced by sea-land breeze. The message here is that we should take into account of the sea-land breeze effect when interpreting the data.

6. Figure 4: It would be useful to have a fifth panel that shows full column XCO2 with the CT extension here, to re-inforce the assertion on line 249-250 about the information lost by considering the total column.

It is presented in Figure 7c. We will point the reader to Figure 7c in this part of the description.

7. Figure 9a: The multi-month average XCO2 gradients can easily miss transient features that could, in theory, be well captured by a regional transport model having spatial resolution that is sufficient to capture synoptic features such as fronts. These features could be attributed to fluxes under this assumption, provided biases are small. That doesn't discount this analysis, but it does imply the need to make assertions about the constraint of XCO2 on surface fluxes.

The purpose of Figure 9a is to show that CT2015 can compare well with XCO2 from aircrafts data at this temporally averaged scale, and we should expect a smooth spatial gradient pattern at this temporal scale. This is another baseline to evaluate the performances of models and column CO2 retrievals, in additional to transient features. Please also see our responses in General Comment.

8. Fig 9b: This vertically averaged wind vector plot doesn't really match the spatial gradient well, in some cases actually being perpendicular to the field. It would make more sense to use the potential temperature at 700mb, as did Keppel-Aleks et al in the reference you cite. Alternately, the 500mb geopotential height is a commonly used field for synoptic scale transport in NWP.

We agree with the review that potential temperature and geopotential height patterns can better match with CO2 spatial gradient; however, these terms are less straight forward when we interpretation the transport and point to the source regions. We think the wind pattern is sufficient to show the upwind locations we discuss in this study.

9. Figure 10: Can you show the same plots, but for the partial columns that are depicted in Figure 9c and d? That would really drive home the point about transport versus local fluxes.

We will provide these figures in the revised paper:

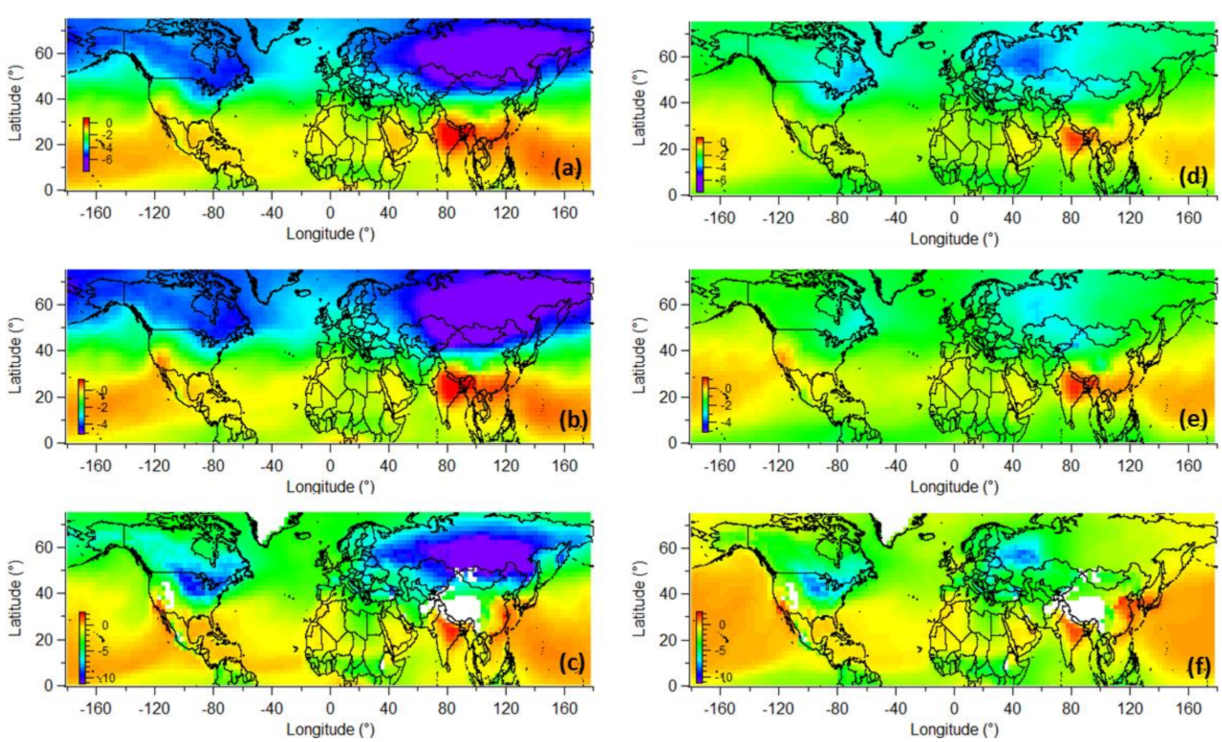


Figure. Total and partial column ΔXCO_2 from Carbon Tracker control (left panels: (a), (b), and (c)) and masked (right panels, (d), (e), and (f). Eurasian boreal flux is masked) runs for 2012 June-August (3° × 2° spatial resolution). MLO trend from each individual scenario is removed before the ΔXCO_2 calculation. (a) and (d) show total column averages. (b) and (e) show partial column averages for free troposphere (800 hPa to 330 hPa). (c) and (f) show partial column averages for lower troposphere (below 800 hPa). (a) and (d) use the same color scale as in Fig. 9a., which reflects maximum 6 ppm gradient. Color scale in (b) and (e) also shows maximum 6 ppm gradients for comparison with (a) and (d); however, the actual values are different. Color scales in (c) and (f) are larger to reflect large spatial gradients in lower troposphere.

10. Discussion of Reuter et al [2014]: Is it possible that the differences are due to the manner in which the anomalies are computed? Or are you asserting that the gradients are due to satellite measurement bias? Stating that they "should not be used" needs a bit more justification here.

We don't think the manner in which the anomalies are computed is the reason for the unrealistic gradient. According to their description, the long-term trend has been removed when computed the anomalies, just as in our study. They use 8-year summertime data, which should provide enough data for a reasonable averaged pattern. Thus the quality of the retrievals is likely the reason for an unrealistic pattern. Since the spatial gradient is the core of inversion study to estimate fluxes, it is critical for the input data to have good spatial gradients. We will add in more discussion in the revised paper.