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.

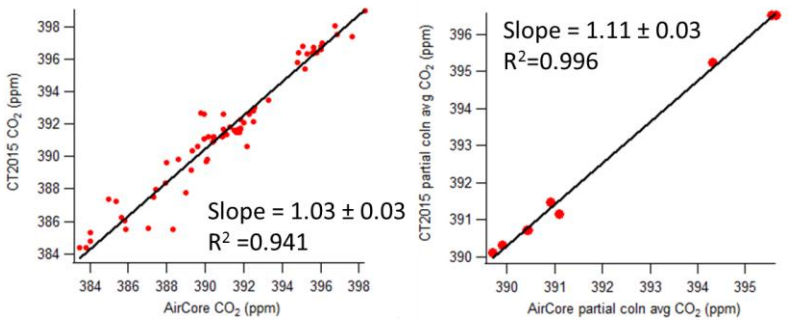
Our study provides an approach to check whether column CO2 retrievals are satisfying the high accuracy requirement on some levels. Since spatial gradient is the core of deriving reliable regional fluxes, we look into the spatial gradient in long-term multimonth averages, which should be easier for models 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:

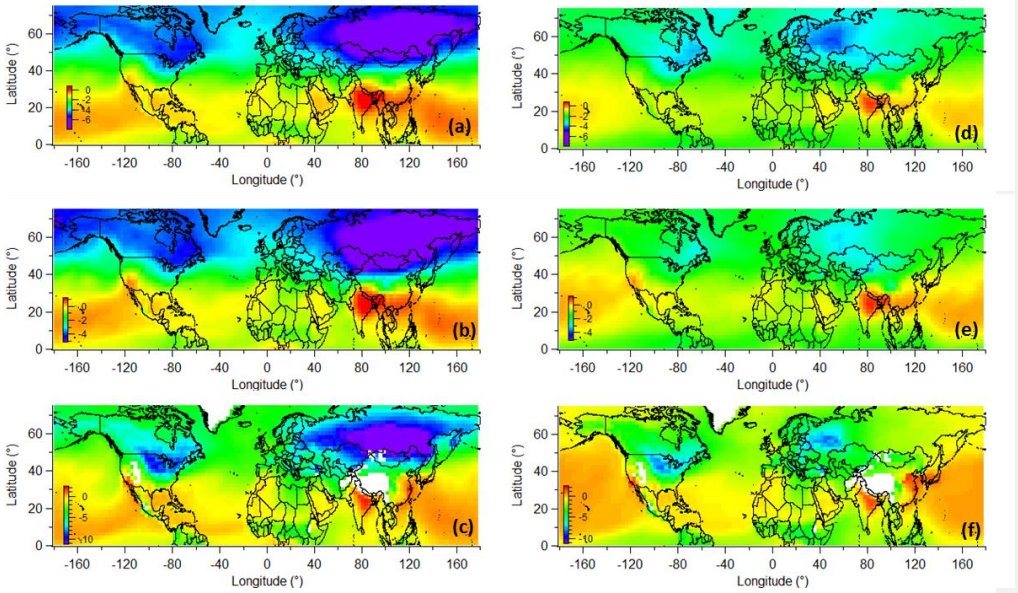


Figure. Total and partial column ΔXCO2 from Carbon Tracker control (left panels: (a), (b), and (c)) and masked (right panels, (d), (e), and (f). Eurasian boreal flux is masked) runs for 2012 June-August (3° × 2° spatial resolution). MLO trend from each individual scenario is removed before the ΔXCO2 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.

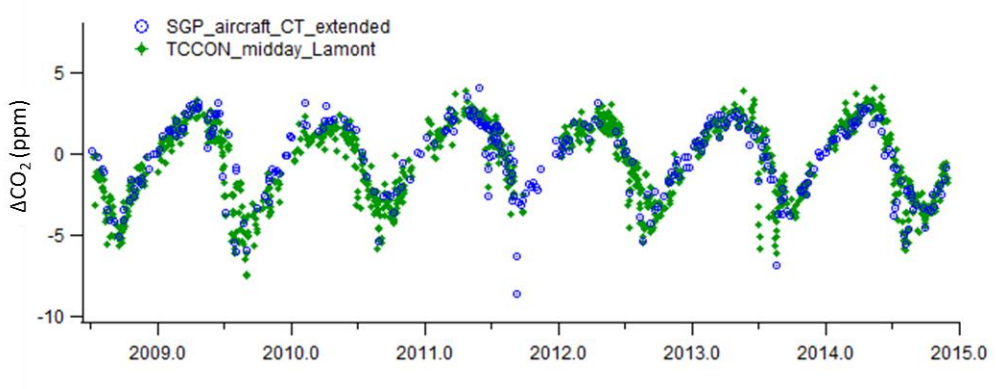
Reply to Reviewer:

Thank you for reviewing our paper.

General comments:

My major comment is that this paper is begging for a direct comparison between the aircraft-derived XCO2 quantity and those measured via remote sensing at coincident TCCON stations. The obvious two sites would be at LEF (Park Falls) and SGP (Lamont) for which TCCON measurements exist since 2004 and 2008, respectively. There is also a TCCON station at ETL, but it has been measuring for less than one year and would therefore be less useful for this study. It would significantly strengthen this paper if the authors could show that their method of integrating the aircraft and extending the profile into the stratosphere with CT compares well with, or improves upon, the total column measurements from TCCON. The authors could average the TCCON XCO2 within an hour or so of the aircraft profile times for an apples-to-apples comparison. They could then use the remaining TCCON data (or at least the near-noon data) to investigate any biases or missing information caused by the relatively infrequent aircraft measurements. As part of this comparison, a more rigorous error analysis of the aircraft-derived XCO2 would be necessary.

We agree with the reviewer that it will be interesting to compare TCCON with aircraft + CT based total column CO2. The following figure can give us some ideas about this comparison. TCCON values in this figure are daily averages computed from mid-day data (10 to 15 LST). Outliers in TCCON data are filtered out, by using 3.S.D threshold of residuals (with iteration) from a quadratic fit on data for each day.



TCCON Data DOI 10.14291/tccon.ggg2014.lamont01.R0/1149159

TCCON reference Wunch, D., G. C. Toon, J.-F. L. Blavier, R. A. Washenfelder, J. Notholt, B. J. Connor, D. W. T. Griffith, V. Sherlock, and P. O. Wennberg (2011), The total carbon column

observing network, Philosophical Transactions of the Royal Society - Series A: Mathematical, Physical and Engineering Sciences, 369(1943), 2087-2112, doi:10.1098/rsta.2010.0240. Available from: http://dx.doi.org/10.1098/rsta.2010.0240

However, it is not an apple-to-apple comparison between our aircraft-based column CO2 and TCCON. The sampling time and location are different. Aircraft sampling typically takes ~ 30 min to get the whole profile through downward spiral, mostly within 0.1^{o} of the site location. For TCCON that sampling the whole column instantly at an angle, the actual sampling location, especially at high altitudes, may be far away from our aircraft inlets even though the distance on the ground is much smaller. Mismatches caused by both sampling time and location should be examined carefully if we want answers for accuracy, which should be the topic of another in-depth study. In addition, the aircraft and tower measurements are calibrated on WMO scale; however, direct calibrations of remote sensing instruments are not available. It makes more sense to compare TCCON against calibrated measurements for accuracy instead of the other way around. Our study focuses on the long-term mean spatial patterns of column CO2 over North America and the influence of large-scale transport; we don't think analysis on two individual TCCON sites can provide extra information regarding to our goal. Previous study by Keppel-Aleks et al. (2012) already provides analysis on TCCON data and meridional spatial patterns.

Specific comments:

1. L19: What does a "stronger summer drawdown" mean? Larger amplitude? Lower minimum?

"Drawdown" refers to the variation amplitudes in the seasonal CO2 signals that mostly resulted from strong plants activity. We clarify this point in the revised paper.

2. L20: This sentence is contradicted by the conclusions:

"The spatial gradients of total column XCO2 across North America mainly reflect largescale circulation patterns rather than regional surface sources and sinks."

Conclusions:

"By comparing the spatial gradients of XCO2 with wind vectors across North America, we find that total column XCO2 patterns are equally affected by large-scale circulation patterns as by regional surface sources and sinks."

Your paper seems to corroborate the sentence in your conclusions and not your abstract.

Please modify the abstract accordingly.

We do not mean that local fluxes have no influence on long-term averaged column CO2 spatial pattern. But the major contributor is transport, instead of surface flux. We have results from CT experiments showing that 50% of North-South gradient are due to Eurasia flux, which is stated in both abstract and conclusion parts. Since Eurasia flux is not the only transport signals we observed (upwind Canadian region is likely contribute to transport signal too), we actually expected larger than 50% transport influence in total. We will clarify the statement in both abstract and conclusion.

3. L55: in-situ observations are sparse in global and regional coverage, and, with the exception of AirCore measurements are limited in vertical extent - most cannot measure more than 80% of the atmospheric mass.

We agree with the reviewer that aircraft measurements have vertical limitations. However, we should also acknowledge that most of the flux signals reside in the lower part of the atmosphere. There is little spatial variation of atmospheric CO2 in the stratosphere over mid-latitude region (Andrews et al., 2001). Thus this part of the atmosphere has little contribution to the spatial patterns of total column CO2 that are used to retrieve surface sources and sinks.

(Citation: Andrews, A., Boering, K., Daube, B., Wofsy, S., Loewenstein, M.,Jost, H., Podolske, J., Webster, C., Herman, R., Scott, D., et al.: Mean ages of stratospheric air derived from in situ observations of CO2, CH4, and N2O, J. Geophys. Res. Atmos., 106, 32295–32314, 2001)

4. L73: What are the uncertainties of total column XCO2 calculated from in situ measurements? According to Wunch et al. 2010, given the lack of measurements above the aircraft ceiling, the total column aircraft uncertainty is 0.4 ppm, which is similar to the TCCON measurement uncertainty. I would expect the errors for the profiles discussed in this paper to have larger uncertainties, since the altitude coverage for the NOAA flights is significantly smaller than the profiles used in the Wunch et al., 2010 paper.

We have provided the uncertainty in partial column CO2 derived from tower and aircraft measurements in SI, figure S4, which ranges from 0.12 to 0.96 ppm with average value of 0.32 ppm over all eight regions for long-term monthly averaged values (uncertainty in summer months are higher). Uncertainty in vertical profile mainly reflects atmospheric variability instead of just instrument errors. In our study, several regions have more than one aircraft sites. When we are averaging over those sites over a total of 11-year vertical profiles, the uncertainty goes down since we have reduced atmospheric variability significantly. The uncertainty in Wunch et al.,2010 is based on individual profile without long time averaging, thus it is not the same as in our study .

We cannot properly assign an exact number as the uncertainty for the top 1/3 of the atmosphere without routine in-situ measurements. However, spatial gradient and atmospheric variability is very small in this part of the atmosphere, and thus it is not important in our study.

5. L190: As you mention, nine AirCore profiles is inadequate to evaluate CT2015 stratospheric CO2. Perhaps you could use the other, (much) older balloon-borne or ER-2-borne stratospheric CO2 measurements (e.g.,BOS (https://espoarchive.nasa.gov/archive/browse/bos/Balloon), STRAT (https://espo.nasa.gov/strat/content/STRAT_Science_Overview), ASHOE, etc.), or the more recent HIPPO and ATom aircraft profiles that often reach above the tropopause altitude, especially in wintertime.

Stratospheric CO2 has very small spatial gradient (Andrews et al., 2001) that contribute to total spatial gradient of total column CO2. Thus it is not important for our study. We only need reasonable stratospheric CO2 results to compensate the missing column information of stratospheric CO2. We could've simply expand the partial column results to the total column since we know we have ~ 70% of the column with most of the spatial gradient signals; however, we opt to use the CT modelled stratosphere CO2 to account for the effect of slightly delayed seasonal cycles in stratosphere.

6. L201: It's not clear to me why you compute the high CT bias by using the partial column comparisons. Why don't you integrate the entire CT and AirCore profiles and compare those values?

The top 1/3 partial column is the part from CT that we use to calculate total column CO2, in addition to aircraft and tower measurements. Integrating entire vertical profiles of AirCore will introduce extra uncertainty of AirCore measurements near surface, which is mostly related to the seal time of the AirCore that is not relevant to the stratospheric part of the measurements. We have modified the sealing process for recent AirCore samples.

7. L206: This is unclear: if the stratosphere doesn't matter for your analyses, why include it at all?

Please also see answers for question 3 and 5. We want the results for total column in addition to the partial column that ceiled at 8 km, to provide baseline spatial patterns for comparison with total column retrievals from remote sensing instruments. For the stratosphere, we show in our study that Carbon Tracker's performance is good when compared with individual AirCore profile. For long-term averages, Carbon Tracker results are good enough for our study.

8. Fig 3 caption: What are the black dots?

Black dots in Fig.3 shows the aircraft data above 2km. It is showed in the figure legend.

9. L250: I don't understand why this sentence does not end after "It demonstrates that there is a lot of important information in the vertical profile."

We state "It demonstrates that there is lot of important information in the vertical profile that is diminished in observations of the total column". We think it is important conclusion in our study: most spatial gradient signals reside in lower troposphere (that's why it is important for measurements to be sensitive to this part of atmosphere). This conclusion cannot be demonstrated by column measurements.

10. Fig. 7: It seems strange that you would not include de-trended averaged TCCON XCO2 for comparison in regions that have long-term TCCON measurements (i.e., MW and SM). Additionally, Figure S1 would be an ideal place to show the SGP TCCON total column measurements.

Please see response to general comment.

11. L338: What does "deepest drawdown" mean? The lowest minimum value? The largest amplitude?

We mean the largest amplitude in seasonal cycles among the 8 regions. We will clarify this in the revised paper.

12. L337-339: "It is interesting that the deepest drawdown is seen in region NM, not in region MW that encompasses the very intensive agricultural activities in the U.S. midwest, which suggest the possibility of strong upwind influence in the NM region." I agree this is interesting. Could the authors say something more about possible causes of this effect?

This is evidence that column CO2 is not dominated by surface flux. Even though the boreal uptake is the largest in MW in summer due to intense plant activities, the column CO2 does not show the lowest value in MW because transported signals have major contribution on column CO2.

13. L343: Again, what does “strength of summer drawdown” mean here?

“strength of summer drawdown” means the variation amplitude in seasonal cycles. We will clarify this in the revised paper.

14. L373: This sentence is misleading, regarding Keppel-Aleks et al., 2012: “Thus they also propose that the variations in column CO2 are mainly driven by large-scale flux and transport.” In Keppel-Aleks et al., 2012, they also state (in the abstract): “Rather than obscure the signature of surface fluxes on atmospheric CO2, these synopticscale variations provide useful information that can be used to reveal the meridional flux distribution.”

Keppel-Aleks et al.(2012) finds similar results as our study, and states in their abstract that "New observations of the vertically integrated CO2 mixing ratio <CO2> from ground-based remote sensing show that variations in <CO2> are primarily determined by large-scale flux patterns." They specifically state in sector 3.1 that “In summary, the comparison between drawdown in <CO2> and eddy covariance flux confirms that while regional information is contained in column observations, **these regional flux signals are obscured by larger-scale variations in <CO2>** even on the short timescale.” We both agree with this. Since column CO2 primary reflects large-scale flux, it is logical that local signals are significantly “diluted” in the total column.

The later statement is actually referring to their method to use both potential temperature and column CO2 to estimate flux gradients (the distribution), but not the flux itself. This method is useful to evaluate biosphere models.

15. L373: That large-scale circulation drives almost half (~40%) of the N-S gradient in XCO2 was also shown in Wunch et al. (2013) through the interannual variability of the seasonal cycle amplitudes.

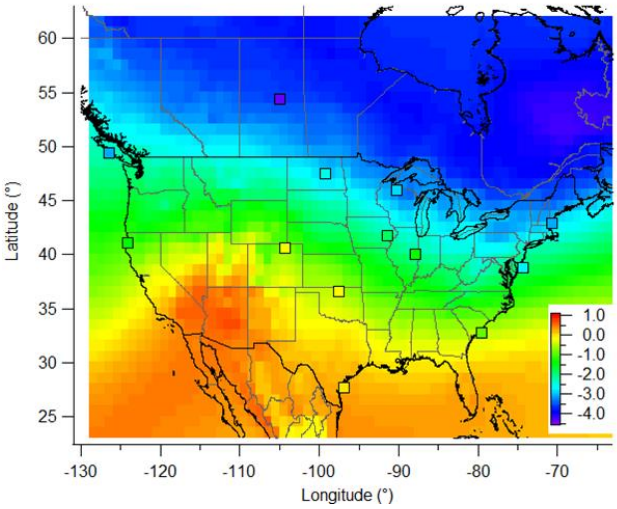
We will cite Wunch et al. (2013) in revised paper. However, we should also acknowledge that interannual variability is not the same statistic metric as the long-term averaged spatial patterns we discuss in our study. We should expect that smaller scale features (local/regional fluxes) are further dampened in long-term time scale.

16. Fig 9: Why do the aircraft measurements appear to disagree significantly with CT in panel (c) in the SGP/Colorado region, and the site just south of lake Michigan?

CT in panel (c) shows the partial column CO2 between 800 hpa to 330 hpa. For SGP and CAR (-0.30 and -0.23 ppm, respectively, based on aircraft measurements), CT results show -0.74 and -0.57 ppm (these numbers are provided in the SI in the revised paper). The mismatches are within ~0.4 ppm which is similar to our uncertainty levels. There are a few possible reasons for these mismatches. The aircraft is designed to sample $0.1^o \times 0.1^o$ area, while CT is $1^o \times 1^o$ spatial resolution. Thus CT may show smoother results. In this case, CT compared better with aircraft measurements in background sites than those influenced by local anthropogenic emissions significantly. SGP is highly polluted by Oil and Gas operation. CAR is also influenced by ONG operations; however, the mountain terrain is challenge for the model too.

For the site HIL (south of Lake Michigan), we have recently found unrealistic high CO2 levels from some profiles that might associated with cabin air contamination. As a precaution, we have removed this part of data (including both 2012 and 2013 summer) in our study after comparing mid-troposphere measurements with the other 4 aircraft sites (LEF, WBI, CAR, and DND) around. We have updated relevant results (including figures) in the revised paper. This update has minimal influence on the regional averages, since we have abundant data in MW region. This update yields slightly lower column averaged CO2 in HIL, which better match the CT 2015 (see updated figure below). Another update has been made on aircraft based column CO2

to fully account for the influence of the temporally uneven sampling frequency. In addition, there are other reasons that model cannot produce perfect results, for example, the transport error may also play a part in the mismatches.



17. L443: I would call 4 ppm a large difference. Please remove “only” from this sentence and quantify the “large spatial gradients” observed below 2 km.

We will remove ‘only’. 4 ppm is smaller compared to the 6 ppm in the SCIAMACHY retrievals in Europe. Spatial gradients below 2 km are discussed in details in section 3.1, and presented in Figure 4, with a maximum difference of ~15.5 ppm between MW and SM in summer on long-term averaged basis.

Technical comments:

We have made changes according to the reviewer suggests.

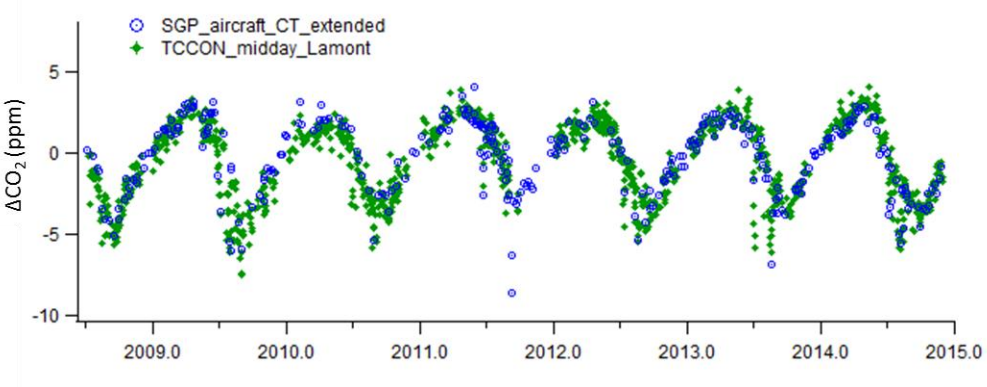
Reply to Reviewer:

Thank you for reviewing our paper.

General comments:

In general, this paper is well-written and is a useful contribution to the literature. I have only a couple serious issues with the paper. First, since this paper is about columns, it would be useful to show direct comparisons of their aircraft-derived XCO_2 to TCCON, to get a sense of differences in remote sensing with respect to a similarly accurate measurement. Second, sometimes they seem to argue that nearly all spatial variability in XCO2 comes from large-scale + remote fluxes and transport, rather than regional-scale fluxes. Other times (such as in the main body), they argue that a significant contribution comes from large-scale + remote fluxes and transport, but a significant contribution also comes from more regional-scale fluxes (within North America in this study). The latter is more in agreement with their data and specifically their removal experiment of section 3.4, so we suggest they rewrite the paper to stress that both mechanisms affect the horizontal structure of XCO2.

We agree with the reviewer that it will be interesting to compare TCCON with aircraft + CT based total column CO2. The following figure can give us some ideas about this comparison. TCCON values in this figure are daily averages computed from mid-day data (10 to 15 LST). Outliers in TCCON data are filtered out, by using 3.S.D threshold of residuals (with iteration) from a quadratic fit on data for each day.



TCCON Data DOI 10.14291/tccon.ggg2014.lamont01.R0/1149159

TCCON reference Wunch, D., G. C. Toon, J.-F. L. Blavier, R. A. Washenfelder, J. Notholt, B. J. Connor, D. W. T. Griffith, V. Sherlock, and P. O. Wennberg (2011), The total carbon column

observing network, Philosophical Transactions of the Royal Society - Series A: Mathematical, Physical and Engineering Sciences, 369(1943), 2087-2112, doi:10.1098/rsta.2010.0240. Available from: http://dx.doi.org/10.1098/rsta.2010.0240

However, it is not an apple-to-apple comparison between our aircraft-based column CO2 and TCCON. The sampling time and location are different. Aircraft sampling typically takes ~ 30 min to get the whole profile through downward spiral, mostly within 0.1^{o} of the site location. For TCCON that sampling the whole column instantly at an angle, the sampling time and the actual sampling location (especially at high altitudes), may be far away from our aircraft measurements. Mismatches caused by both should be examined carefully if we want answers for accuracy, which should be the topic of another in-depth study. In addition, we need to acknowledge that the aircraft and tower measurements are calibrated on WMO scale; however, direct calibrations of remote sensing instruments are not available. It makes more sense to compare TCCON against calibrated measurements for accuracy instead of the other way around.

Regarding to reviewer comments on large-scale + remote fluxes and transport:
We do not mean that regional/local fluxes have no influence on long-term averaged column CO2 spatial pattern. But the major contributor is transport, instead of surface flux. In addition to our observation that long-term column CO2 spatial gradients pattern is smooth, and mainly reflects large-scale circulation pattern, our results from CT experiments showing that 50% of North-South gradient are due to Eurasia flux, which is stated in both abstract and conclusion parts. Since Eurasia boreal region is obviously not the only upwind location with significant CO2 sources, (upwind Canadian region is likely contribute to transport signal too), we actually expected **larger than 50%** transport influence in total. We clarify this point in the revised paper.

Though the authors discuss at length their use of CT2015 to extend the aircraft profiles from 330 hPa to the top-of-atmosphere, they don't really discuss the potential errors from the fact that the aircraft measurements don't go all the way to the surface. While some sites do sample very close to the surface (e.g. 0.2 km at SGP), other sites don't even sample as low as 1 km AGL (e.g. CAR: 2.2 km agl, HIL: 1.1 km agl). Because the concentration can change quickly near the surface, the authors need to acknowledge this source of error, and ideally estimate its potential magnitude. They could do this by taking full CT profiles for their sites and comparing the column extended their way (lowest value held constant) vs. the CT value. They can also estimate it based on tower observations for a few sites.

The reviewer has mistaken the unit we use for elevation and altitude. We use 'above sea level (asl)' instead of 'above ground level (agl)' in our study, and the information is presented in S Table 1. CAR is located on a high elevation terrain, as sated in Line 143. There is no < 1km (asl) measurements because the averaged surface elevation is already 1.5 km asl. The lowest layer of aircraft measurements at CAR is ~ 2.1 km asl, which is ~ 0.6 km agl. HIL does sample below 1 km asl; the lowest level s is ~ 600 m asl (surface elevation is ~200 m asl). Most regions we discuss in this study also have tower measurements below 1km asl, although aircrafts can sample even lower than the highest inlets of some towers, as we can see from Fig. 5 and Fig. 6. We have a few layers of measurements below 800 hPa for all regions (except MC with 1.5km surface elevation), as we can see from in Fig.5 and Fig.6.

Regarding to the reviewer's comment that "the concentration can change quickly near the surface", is the reviewer referring to nighttime measurements? The aircraft profiles are taken at late morning and early afternoon (data outside 10:00-17:00 local time are excluded in this study), during which time the PBL are generally well-mixed. The tower data we use in our study are also daytime data only, which are stated in our section 2.1. We have observed consistent daytime CO2 from different heights in tower measurements, even during summer. The following figures are from our tower data: https://www.esrl.noaa.gov/gmd/ccgg/about/co2_measurements.html

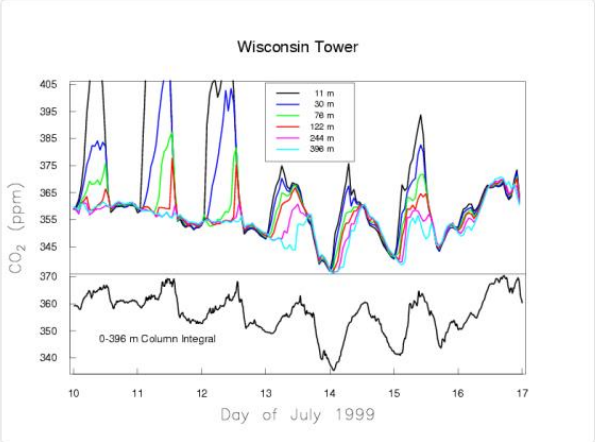


Figure 5a. Measurements of CO_2 by a continuous in-situ analyzer at a tall tower during the summer in northern Wisconsin. The top plot show the CO_2 values at six different heights on the tower.

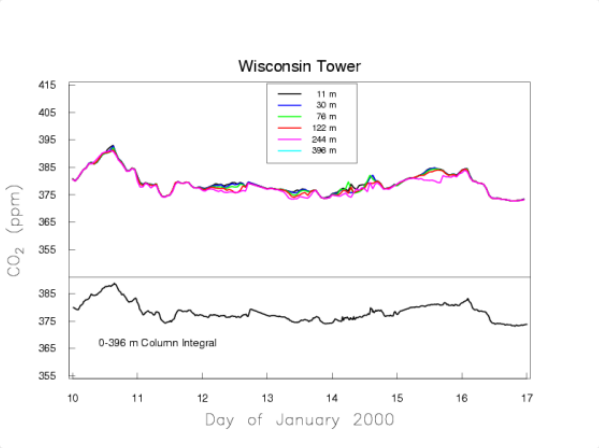


Figure 5b. Measurements of CO_2 by a continuous in-situ analyzer at a tall tower during the winter in northern Wisconsin.

Finally, their section on Reuter et al. (2014, hereafter R14) doesn't fully fit in with the rest of the paper. Specifically, they show that CarbonTracker (CT) agrees well with aircraft over North America. This is not surprising, given the excellent coverage of surface sites in North America, which CT assimilates. However, this does NOT guarantee that CT will be right everywhere, which they seem to imply throughout the paper. They need to stress that this work validates CT in North America specifically. Anywhere else, and they are merely speculating. Also, they focus on the "hot spots" in the R14 map, which are likely due to inhomogenous SCIAMACHY spatiotemporal sampling as much as anything, and may disappear in a regional inversion. Regarding the European sink, the stations that CT assimilates are all in Western Europe, and much of the controversy is really how much sink is portioned between the Europe and Eurasian Boreal Transcom regions, as discussed in both R14 and Reuter et al. (2016). Their central argument that the spatial patterns of XCO_2 from R14 are unphysical is qualitative at best, and, from this author's perspective, partially based on the unfortunate color scheme (rainbow) chosen by R14 which can accentuate very small spatial differences. They should tone down the language to something more like to say that the spatial patterns seem unphysical, but more work would be required to really rebut their physicality. They also need to cite Reuter et al. (2016), which gives a thorough overview of the "European sink controversy".

The reviewer agrees that the column CO2 spatial pattern from SCIAMACHY retrievals is "unphysical". From our comparison with CarbonTracker, which is proven to be trustworthy for the long-term spatial pattern over the U.S., we think the column CO2 spatial pattern from SCIAMACHY retrievals is unrealistic and disagree with our understanding of transport. It is important to point out these in our study.

We think it is reasonable to believe that CT produces trustworthy results even outside North America, for the following reasons: (1) CT-Europe that simulate more surface measurements shows similar results as CT. (2) The transport in CT is reasonably well, supported by the good simulation of SF6, a tracer for large-scale transport (see https://www.esrl.noaa.gov/gmd/ccgg/carbontracker/CT2016_doc.php#tth_sEc6, section 6.2). (3) The reviewer seems to suggest that we cannot judge the model permanence in Europe because we don't have the truth to compare with, unlike the well-calibrated aircraft profiles we have in North America. But that's the case for all CO2 models over Europe. At least for CT, it

compared well with N.A. aircraft measurements, even though aircraft measurements are not assimilated in CT. This gives us more confidence for the CT model.

Regrading to the spatial patterns of SCIAMACHY in Reuter et al.(2014), it is unlikely the inhomogeneous spatiotemporal sampling is the major reason for spotted feature, since 8 years of data are used to generate this spatial pattern. We do not propose that the conclusion of R14 about the large European carbon sink is wrong. However, it is important to point out the serious issue that this study use unrealistic spatial pattern. Based on our knowledge of transport, a smooth spatial pattern is expected, especially after averaging 8 years of data. We do not agree with the reviewer that the "unfortunate color scheme (rainbow) chosen by R14" is the reason for the unrealistic spatial pattern. We use the same rainbow color scheme and the same 6 ppm maximum color scale as R14 in our Fig. 11, in order to compare with Reuter et al. We do not see the spotted feature on similar spatial resolution; instead we see gradual spatial patterns that match our understanding on transport (Keppel-Aleks et al., 2012 find similar results about the significant influence of large-scale transport).

We will cite Reuter et al. 2016 for European sink. However, the main purpose of our study is not to discuss the European sink. We intend to show a smooth spatial pattern of column CO2 (based on well-calibrated measurements) that match our understanding of transport, which should be a baseline for current and future remote sensing retrievals.

Specific and technical comments:

L54: Suggest you add language such as "or else regional-scale biases can result" and cite Chevallier et al. (2014).

We cite the Chevallier et al. (2014) in the Line 88.

L65: While satellite retrievals of XCO_2 can certainly have regional biases, the Inoue et al. (2016) paper is hardly thorough. It discusses one particular retrieval (the NIES retrieval) of GOSAT data, and in fact the authors appear to cherry-pick the biases of 2.09 and 3.37 ppm they list, which come from stations with the fewest colocations (3 and 1, respectively). Most of the stations in that paper have biases less than 1ppm, consistent with their comparison to TCCON, and that of Kulawik et al (2016). The authors should give a fairer representation of Inoue et al, including their TCCON comparisons, and mention the mean biases relative to TCCON given in Kulawik et al. (2016).

The reviewer suggests that comparison between GOSAT and TCCON is fairer. The Inoue et al., 2016 compared GOSAT retrievals with aircraft based column CO2 after correcting GOSAT retrievals against TCCON. This is independent evaluation; however, comparison with TCCON after correction (using TCCON data as reference) is not. We don't think comparing with TCCON is a better approach than comparing with aircraft-based column CO2. While TCCON shares some similarities with satellite retrievals column CO2 considering both use remote sensing technique, TCCON itself has known bias. TCCON cannot be directly calibrated, unlike aircraft profiles. Well-calibrated measurements should be a prefer 'truth' to compare with.
We will provide the information that 20 out of 27 stations in Inoue et al. (2016) study shows difference smaller than 1 ppm compared with aircraft-based column CO2.

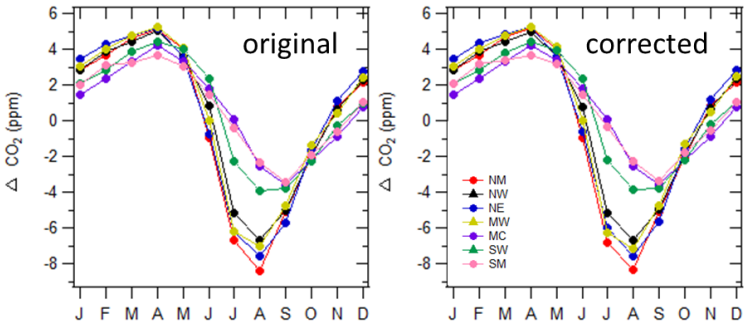
L88: To the 2001 and 2004 papers they cite, they should add Chevallier et al. (2014), which more explicitly demonstrate this.

We cite the Chevallier et al. (2014) in revised paper.

L119-125: Regarding the recently-discovered high humidity bias. To me, this seems like a pretty large error, especially if the mean / typical correction is 0.5 ppm! Since this is the first paper to discuss it, a plot seems warranted – authors should plot the estimated error based on humidity, and justify this 1.7% humidity cut-off value they give. Also, the authors aren't entirely clear on the units of "humidity" here. Presumably they mean volumetric mixing ratio = dry air mole fraction; please be more clear, or use a more typical unit (such as mass mixing ratio or specific humidity).

While a 0.5 ppm correction seems to be large, only ~ 4% of total aircraft measurements or ~ 12% of those below 2 km are impacted. The 1.7 % threshold is preliminary, based on data from

2 tower sites; however, our more recent aircraft measurements with larger amount of data find that this threshold is sufficient for data correction. We are still working to better understand and quantify this bias, and testing dryer design to avoid influence of high humidity. To evaluate the impact of this bias on our study, we have compared the partial column CO2 results (below 330 hPa) before and after the corrections. Please see figures below:



From this comparison, it is safe to say that this correction does not pose an obvious difference on our results. However, for future studies that are not averaging over long-term record like this study, we need to be prudent on this bias.

The unit for humidity here is ppm in mole fraction, which is the mole of water vapor versus whole air. Dry air should not be assumed when we talk about ratio of mole fractions. We clarify the unit in revised paper.

L158: I think the authors mean Figure 3, not 2.

We change it to “Fig. 3”.

L277 – Here is where the authors should discuss potential errors by extending the lowest measurement to the surface (see discussion above in *General Comments*).

Please see response to general comment.

L308 – Please remind the reader here what "partial column" means. I think surface to 330 hPa, but it's not entirely clear.

We clarify "partial column" in revised paper.

L318 – Suggest you change the sentence to stress that CT captures these features of XCO2 *over North America* reasonably well. There is absolutely no guarantee that this will be true elsewhere, as discussed in *General Comments*. You can speculate that it will work in other locations, but you don't show it, so as written this statement is not justified.

We modify the sentence. Please also see response to general comment.

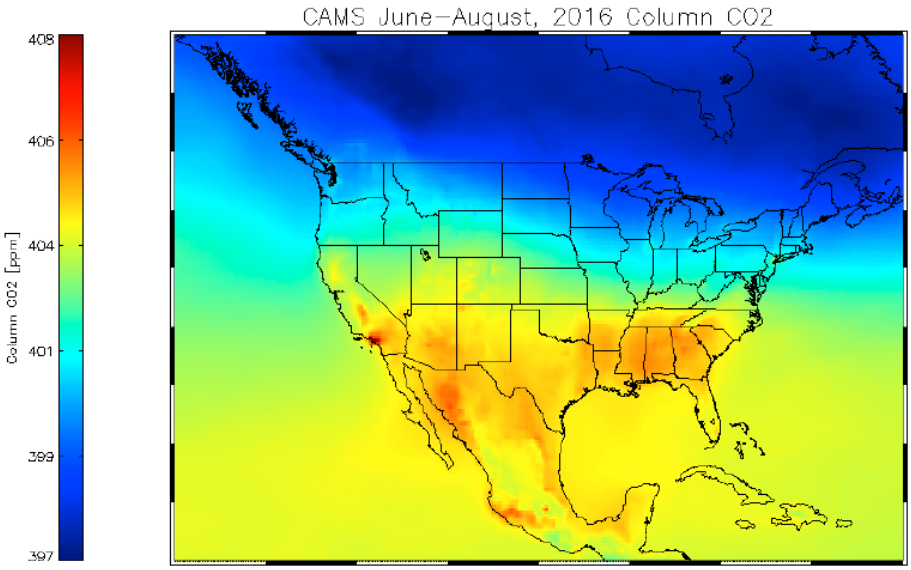
L329 – Please define "drawdown". Does it mean peak to trough of CO2 in the seasonal cycle, or the minimum CO2 relative to MLO? The former is more typically accepted as the definition of drawdown. Then make Fig 8 consistent with your chosen definition.

"Drawdown" refers to the variation amplitudes in the seasonal CO2 signals that mostly resulted from strong plants activity. We clarify this point several places in the revised paper.

L346 – Again emphasize that the agreement is only shown for North America.

Please see response to general comment.

L348-350 Regarding CT not showing hotspots. This is not necessarily surprising and could be due to CT's limited resolution. I've attached an unofficial plot of the CAMS nrt XCO2 over North America, for the June-August 2016 average (see e.g. Massart et al. 2016 for details), which has a resolution of roughly 0.25°, and hotspots over L.A. and Mexico city are clearly visible, consistent with the findings of Kort et al. (2012) and others.



It is reasonable that finer grid can show more pronounce hot spots. However, from the CAMS results for 2016 summer, we cannot see random spotted feature like the 8-year mean SCIAMACHY retrievals shows (which are in a much coarser grid).

It is interesting to see whether the hot-spot in LA and Mexico city also exist in the a priori. How does the long-term averaged spatial pattern look like from CAMS? Is the hot-spot feature still pronounced? The long-term averages are those that both Reuter et al. 2014 and our study focus on.

L374-6 : This implies that local/regional fluxes account for ~50% of the horizontal gradient in this case. Authors should mention this fact, rather than argue that virtually all of the spatial gradients come from large scale transport, which is what they state in the abstract.

Please see response to general comment.

L392: Reuter et al (2014) doesn't only use SCIAMACHY; they use GOSAT as well, and in fact, they use a number of different GOSAT retrievals, not just one, and get consistent results from all of them. Statements in this paper should reflect this.

The unrealistic spatial patterns in Reuter et al. (2014) Fig. 2 are SCIAMACHY retrievals. The main purpose of our study is not to discuss the European sink. We intend to show a smooth spatial pattern of column CO2 that match our understanding of transport, which should be a baseline for remote sensing retrievals. Since our study is not an inversion study, we cannot prove that the inversion results in Reuter et al. (2014) are wrong. We do not intent to focus on the resulting fluxes. Instead, we focus on pointing out that the spatial pattern is unrealistic. The spatial pattern from GOSAT is actually drastically different from SCIAMACHY. Spatial gradient is the backbone of inversion modeling that resolve surface sources and sinks. Their consistent results from different inversions that use drastically different spatial patterns are not assuring, instead, they raise the question that why different inversions based on drastically different spatial gradients (or the unrealistic gradients in SCIAMACHY) can produce consistent results. It is not clear that whether the data are serving as good enough constraints in inversions or the a priori and transport are dominating the results.

L409: The Feng et al. (2016) study is only applicable to global inversions. R14 uses a regional inversion with only observations in Europe, to make themselves insensitive to an overall regional bias, and only use the gradients in the satellite data to infer the European sink. This is somewhat at odds with the Feng et al. study, which shows that much of the apparent European sink comes from satellite observations outside of Europe.

We do not agree with the reviewer that Feng et al. study only applicable to global inversion. From the prospect of inversion setup, there is no guarantee that regional inversion is better than global inversion since reliable boundary condition is still needed for regional inversion. Global constraints are needed to be satisfied. From the perspective of regional satellite data, using regional data makes the inversion less susceptible to extreme bias, such as land-sea bias; however, there are still potential sources of bias within the European region which may be

caused by spatial differences in clouds and aerosols, surface albedo, etc. Small bias is enough to produce drastically fluxes results. 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. The importance of small bias is also shown in RT14 Figure 2c that an overall 0.5 ppm spatial gradient over Europe is enough to produce the extra European carbon sink.

Gradients of Column CO_2 across North America from the NOAA Global Greenhouse Gas Reference Network

Xin Lan[1, 2], Pieter Tans[1], Colm Sweeney[1, 2], Arlyn Andrews[1], Andrew Jacobson[1, 2], Molly Crotwell[1, 2], Edward Dlugokencky[1], Jonathan Kofler[1, 2], Patricia Lang[1], Kirk Thoning[1], Sonja Wolter[1, 2]

[1]1National Oceanic and Atmospheric Administration, Earth System Research Laboratory, Boulder, 80303, Colorado, USA

[2] University of Colorado, Cooperative Institute for Research in Environmental Sciences, Boulder, 80309, Colorado, USA

Correspondence to: Xin.Lan (xin.lan@noaa.gov)

Abstract. This study analyzes seasonal and spatial patterns of column carbon dioxide (CO_2) over North America calculated from aircraft and tall tower measurements from the NOAA Global Greenhouse Gas Reference Network from 2004 to 2014. Consistent with expectations, gradients between the eight regions studied are larger below 2 km than above 5 km. The 11-year mean CO_2 dry mole fraction (XCO_2) in the column below ~330 hPa (~ 8 km above sea level) from NOAA's CO_2 data assimilation model, CarbonTracker (CT2015), demonstrates good agreement with those calculated from calibrated measurements on aircraft and towers. Total column XCO_2 was attained by combining modeled CO_2 above 330 hPa from CT2015 with the measurements. We find large spatial gradients of total column XCO_2 during June to August, and the north and northeast regions have ~3 ppm stronger summer drawdown (variation amplitude in seasonal cycle) than the south and southwest regions. The long-term averaged spatial gradients of total column XCO_2 across North America show a smooth pattern that mainly reflect large-scale circulation patterns ~~rather than regional surface sources and sinks~~. We have conducted a CarbonTracker experiment to investigate the impact of Eurasian long-range transport. The result suggests that the large summer time Eurasian boreal flux contributes about half of the north-south column XCO_2 gradient across North America. Our results confirm that continental-scale total column XCO_2 gradients simulated by CarbonTracker are realistic and can be used to evaluate the credibility of spatial patterns from satellite retrievals, such as the long term average spatial patterns from satellite retrievals reported for Europe which show larger spatial difference (~ 6 ppm) and scattered hot spots.

1 Introduction

Atmospheric measurements of carbon dioxide (CO_2) from ground and airborne platforms have greatly increased our knowledge of the global carbon cycle. Observations of CO_2, including the NOAA Global Greenhouse Gas Reference Network (GGGRN), initially emphasized ground-based measurements. These observations, started by C.D. Keeling, have monitored the CO_2 trend on both regional and global scales for over 50 years (e.g., Keeling and Rakestraw, 1960; Tans et al., 1989). In addition, the frequency and spatial distribution of airborne measurements have increased rapidly in the last two decades, providing important information about horizontal and vertical

variability of atmospheric CO_2 (e.g., Gerbig et al., 2003; Choi et al., 2008; Biraud et al., 2013). Routine aircraft measurements from the NOAA/ESRL GGGRN monitor the large-scale distributions of a suite of trace gases, including CO_2, under the influence of continental processes (Sweeney et al., 2015). A very successful approach has been to employ commercial aircraft as a platform for CO_2 measurements, such as Japan's CONTRAIL (Comprehensive Observation Network for TRace gases by AIrLiner) project which has provided valuable information for CO_2 in the high troposphere and lower stratosphere (Machida et al., 2002; Machida et al., 2008). Vertical profiles of atmospheric CO_2 reflect the combined influences of surface fluxes and atmospheric mixing. Vertical profiles are particularly useful for evaluating vertical mixing in atmospheric transport models that are used for inverse modeling (e.g. Stephens et al., 2007) to derive estimates of regional- to continental-scale CO_2 sources and sinks (e.g., Tans et al., 1990; Gurney et al., 2002; Gurney et al., 2004; Ciais et al., 2010;).

While CO_2 sources and sinks are better constrained at the global scale by global mass balance, it remains challenging to accurately resolve CO_2 sources and sinks at regional-to continental-scale, the apportionment of which depends on relatively minor variations of the observed spatial and temporal patterns of CO_2. When averaging over a few months and longer the largest portion of the variations over continents results from hemispheric-scale terrestrial uptake (photosynthesis)/emissions (respiration) and fossil fuel emissions, while regional net fluxes can make a relatively small contribution to the signal. For example, a simple mass balance argument shows that all U.S. CO_2 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} m^2).

With careful calibration, air handling, and analysis, the uncertainties of in-situ measurements are less than 0.1 ppm. However, in-situ observation networks are sparse in global and regional coverage. Remote sensing data radically increase the number of observations and capture under-sampled regions. It is likely to have a valuable impact on our understanding of the carbon cycle. However, both the precision and the potential of even very small systematic biases in remote sensing measurements need to be carefully evaluated. Vertical profiles from in-situ CO_2 measurements have been used to evaluate ground-based total column XCO_2 (X stands for dry mole fraction) determinations, such as those from the Total Carbon Column Observing Network (TCCON) (Washenfelder et al., 2006; Wunch et al., 2010; Messerschmidt et al., 2011; Tanaka et al., 2012). The uncertainty of TCCON total column CO_2 is reported to be 0.4 ppm (1σ) after comparison to aircraft measurements (Wunch et al., 2010). Vertical profiles are also used to evaluate satellite retrievals of total column XCO_2, such as those from the Tropospheric Emission Spectrometer (TES)(Kulawik et al., 2013) and the Greenhouse Gases Observing SATellite (GOSAT) (Inoue et al., 2013, 2016; Saitoh et al., 2016). Satellite retrieval products have known and unknown biases (due to errors in spectroscopy, viewing geometry, spatial differences in clouds and aerosols, surface albedo, etc.) that can result in false horizontal gradients in total column XCO_2 for inverse estimates of sources (Miller et al., 2007; Crisp et al., 2012; Feng et al., 2016). After correction for known biases, the mean GOSAT total column CO_2 (NIES retrievals) biases range between -2.09 to 3.37 ppm (mean = 0.11 ppm, S.D.= 1.11 ppm; 20 out of 27 stations show biases lower than 1 ppm) across different aircraft sites over land, compared with aircraft-based total column XCO_2 (Inoue et al.,2016). By comparing with TCCON, the Orbiting Carbon Observatory-2 (OCO-2) retrieval of total column XCO_2

was estimated to have a mean difference less than 0.5 ppm with RMS differences typically below 1.5 ppm after bias correction (Wunch et al., 2016). The overall uncertainty of satellite retrievals is relatively large compared with the total column XCO_2 calculated from in-situ measurements. Total column XCO_2 calculated from vertical profiles from the Japanese CONTRAIL project (Machida et al., 2008) and from the NOAA Carbon Cycle and Greenhouse Gas aircraft program (Sweeney et al., 2015) complemented with simulated profiles from a chemistry–transport model above the maximum altitude of the data have uncertainty less than 1 ppm (Miyamoto et al., 2013). The relatively small uncertainty of the in situ-based total column XCO2 suggests that they can be used to evaluate satellite retrievals of column averaged CO_2. Since aircraft profiles co-located with satellite retrievals are rare, it is useful to consider the statistics of total column XCO_2 fields derived from repeated aircraft profiles over particular locations.

The effect of satellite column averaging kernels and a priori profiles when comparing aircraft-based column XCO_2 with GOSAT retrievals has been assessed by Inoue et al. (2013). For the case considered, application of the averaging kernel and a priori profile to simulate total column XCO_2 was generally within ± 0.1 ppm of the density weighted total column, suggesting that the averaging kernels can only account for small part of the overall uncertainty of the GOSAT total column XCO_2 (Inoue et al., 2013).

Transparent and objective estimates of CO_2 sources and sinks derived from atmospheric measurements are paramount for validating emissions reduction efforts and other mitigation policies, and for lowering the uncertainties of carbon cycle-climate feedbacks. The latter are major ambiguities in predicting future climate, such as potential uncontrolled CH_4 and CO_2 emissions from warming permafrost in Arctic regions. Satellite retrievals of total column XCO_2 can significantly improve estimates of sources and sinks only if they are sufficiently precise and accurate (~~Rayner and O'Brien, 2001;~~ Houweling et al., 2004; Chevallier et al., 2014), meaning that even very small systematic errors (biases) must be eliminated. Here, we analyze the spatial and temporal variability of column CO_2 over North America using well-calibrated CO_2 measurements from aircraft and tall tower, and we use model results from NOAA's CarbonTracker, version CT2015 (Peters et al. 2007, with updates documented at http://carbontracker.noaa.gov) to investigate the primary drivers of variability in total column XCO_2. The aircraft data enable direct analysis of column CO_2 characteristics, which is the fundamental step for accurate apportionment of sources and sinks. This study focuses on the long-term averaged column CO_2 gradient and the contributions of different vertical layers to the total column variability. It can serve as a reference for evaluating current and future column CO_2 retrievals from both ground and satellite platforms.

2 Methods

2.1 Aircraft and tall tower sampling

Aircraft sampling in the NOAA GGGRN intends to provide vertical profiles of long-lived trace gases to capture their seasonal and interannual variability. The aircraft sampling system consists of 12 borosilicate glass flasks in each programmable flask package (PFP), a stainless-steel gas manifold system, and a data logging and control. These flasks (0.7 L each) are pressurized to obtain 2.2 L of sample air from each target altitude. Air samples are then shipped back to NOAA/ESRL for carefully calibrated and quality-controlled measurements. Carbon dioxide is

measured using a nondispersive infrared analyzer. Long-term measurements at ~15 sites are carried out using light aircraft that can reach 8.5 km. Air samples are collected mostly during late morning to early afternoon, when the air mass within the planetary boundary layer (PBL) is generally well mixed, and CO_2 enhancement near the ground from plant respiration during the night has been mixed throughout the boundary layer. Normally, the aircraft follows a pre-decided route such that most samples are collected within 0.1° of the site location. The sampling frequency varies from site to site, currently from twice a month to once every 1.5 months. For more sampling details, quality control discussions, and an evaluation of the sampling frequency, please refer to Sweeney et al. (2015). More information on the aircraft sites can be found at http://www.esrl.noaa.gov/gmd/ccgg/aircraft/. We estimate the uncertainty of individual measurements of CO_2 in flask air (68% confidence level) at 0.08 ppm. However, we have seen evidence of positive biases for samples collected using older flasks that may contain contaminants. Andrews et al. (2014) reported biases that increased from <0.1 ppm in 2008 to an average offset in 2013 of 0.36 ppm. The aircraft sampling protocol was modified starting in August 2014 to mitigate this bias. For samples collected prior the protocol change, laboratory tests showed that new/clean flasks have zero bias, but some older/dirty flasks could have biases of > 1 ppm. This bias is not consistent among individual flasks and increasing over time (Andrews et al., 2014), the potential bias is hard to quantify for measurements before August 2014. Thus, the high bias is not corrected in our study. More recently, low bias has been found in PFP measurements when the ambient humidity is high, based on comparisons of PFP measurements with data from in-situ analyzers at tall towers. We are working to understand and quantify this bias, and for this study we have derived a preliminary correction factor, which shows a linear trend with −1.4 ppm CO_2 offset per 1% above 1.7% of ambient water content (in mole fractions versus whole air). Only ~ 4% of total aircraft measurements or ~ 12% of those below 2 km are impacted by humidity higher than 1.7%, for which we have applied corrections before data analysis. The mean correction applied is 0.53 ± 0.4 (1 σ) ppm for the impacted data.

The NOAA tall tower network measures CO_2 and other trace gases within the continental boundary layer. Continuous in-situ measurements are conducted using nondispersive infrared (NDIR) absorption sensors and cavity ring-down analyzers. The long-term stability of these systems is typically better than 0.1 ppm for CO_2 (Andrews et al., 2014). Most tall tower sites have more than one air intake height. In this study, continuous in-situ measurements from the highest intake are used to minimize potential influences from local sources. More information concerning the tower sites can be found at http://www.esrl.noaa.gov/gmd/ccgg/insitu/. For the column XCO_2 calculation, tower data only from 10:00-17:00 local standard time (LST) on flight days are averaged to one data point per day, as a complement to vertical profiles within the PBL.

2.2 Site description

We analyze data from 19 aircraft sites and 6 tall tower sites during 2004 to 2014 (see Table S1 for a summary of site conditions). After considering the geographic distribution of these sites in North America, we group them into eight regions for spatial comparisons (Fig. 1). The northern west (NW) and southern west (SW) regions represent the inflow area in the west coast of US, directly downwind of the Pacific Ocean at both higher elevations. The northern mid-continent (NM) region represents the boreal forest and agriculture region in north-central North America. The

mid-continent (MC) region represents a dry landscape due to its high elevation (above 1.5 km on average) and semi-arid climate. The mid-west (MW) region is strongly influenced by agriculture and temperate forest. The southern mid-continent (SM) represents the south-central humid temperate region, with inflow from the Gulf of Mexico during summer. The northeast (NE) region represents the temperate forest in north-east coast of U.S., which is mostly downwind of regions to the west above the PBL, and downwind of its south-west regions within the PBL. The southeast (SE) region represents the warm temperate region in the south-east coast of U.S.

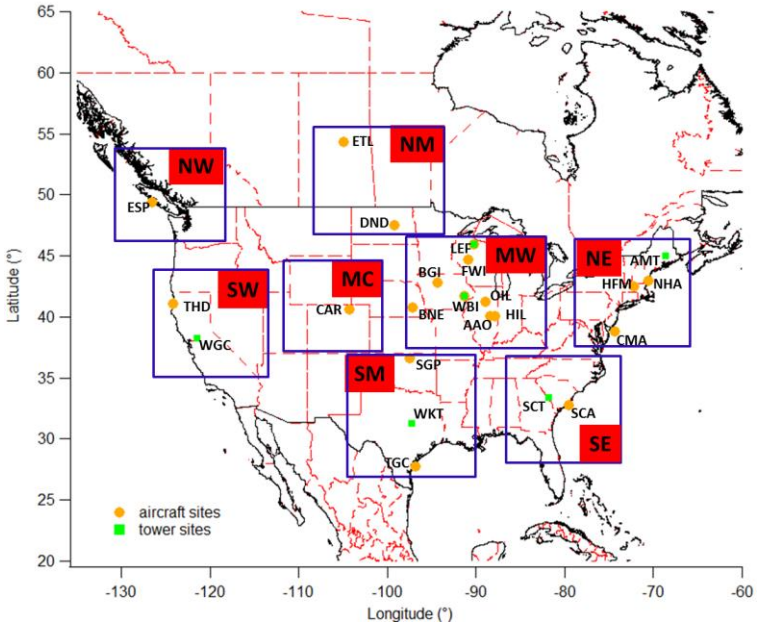


Fig. 1. Aircraft, tall tower, and high elevation/tower sites in the NOAA GGGRN. The eight boxes define regions that are further discussed for spatial pattern comparison.

2.3 Smoothing of the reference data and column XCO_2 calculation

We use Mauna Loa Observatory (MLO) as a reference site. The l~~L~~ong-term trend of CO_2 measurements from this site is removed before combining multiple years of data to calculate long-term averages. MLO is located at 19.536°N, 155.576°W, and 3397 m above sea level. Carbon dioxide measurements from this site are widely used to represent background CO_2 in the Northern Hemisphere. For our study, a function consisting of a quadratic polynomial and four harmonics is fitted to the MLO data, adopted from the method described by Thoning et al. (1989). Residuals of the data from this function are smoothed by a low-pass filter with full-width at half-maximum in the time domain of 1.1 years. The smoothed residuals are then added back to the polynomial part of the function to produce the long-term deseasonalized trend. This trend (see Fig. ~~2~~3) is subtracted from all aircraft and tall tower measurements, as well as from CarbonTracker model results (CarbonTracker - MLO deseasonalized trend,

CarbonTracker results presented in this study are the differences relative to observed MLO deseasonalized trend). We use 'Δ' to represent detrended data in the following text and figures. The choice of reference site is not important for this study, since we focus on examining the relative seasonal patterns of the detrended spatial and vertical distributions of CO_2 instead of the total changes in CO_2 abundance attributed to global surface fluxes.

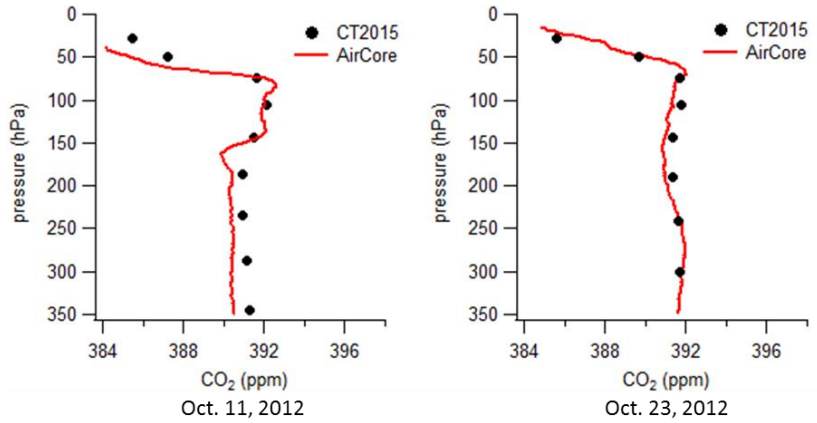


Fig. 2. Carbon Tracker (CT2015) simulations compared with AirCore in-situ measurements in upper atmosphere. AirCore profiles in the left and right panels are sampled near CAR and SGP, respectively.

We calculate partial column average CO_2 dry mole fraction using tall tower and aircraft data, and the total column by adding simulations of high altitude CO_2 (above 330 hPa, ~ 8 km above sea level) from CarbonTracker. Since geometric height from the onboard Global Positioning System (GPS) (after 2006) or inferred from the aircraft altimeter or pressure altitude is archived with each aircraft measurement, we first convert geometric height (in meter) to pressure (in hPa) for the pressure-weighted column XCO_2 calculation. This conversion uses geopotential data from NOAA/NCEP North American Regional Reanalysis (NARR) (Mesinger et. al, 2004), available at https://www.esrl.noaa.gov/psd/data/gridded/data.narr.html, in which the geopotential is a function of latitude, longitude, pressure altitude and time. We interpolate the geopotential field vertically to retrieve pressure, and then calculate dry pressure by incorporating specific humidity data from NARR. Eventually we use a trapezoidal method to integrate over detrended vertical profiles for dry-pressure-weighted column average. For the long-term averaged column ΔXCO_2 calculation, a long-term mean vertical profile is first constructed for each month by combining 11-year detrended data together and then average data in each 40 hPa vertical bin. To look at the long-term averaged total column ΔXCO_2 from individual aircraft sites, we combine aircraft data with upper-layer CT2015 simulations.

The NOAA CarbonTracker model assimilates CO_2 measurements from surface sampling networks and tall towers to generate global 3D fields of atmospheric CO_2 mole fraction. The Carbon Tracker model has evolved significantly since Peters et al. (2007). A detailed description of this model is provided in documents available at http://carbontracker.noaa.gov. Our study utilizes CarbonTracker results from the 2015 release (CT2015), publicly

accessible at ftp://aftp.cmdl.noaa.gov/products/carbontracker/co2/CT2015/molefractions/. This version provides CO_2 mole fraction over North America with 1° × 1° spatial and 3 hour temporal resolutions, which are analyzed in Sect. 3.2 and 3.3. Total column CO_2 calculated from CT2015 global data with 3° × 2° spatial resolution is also presented in the supporting information (SI). We have evaluated the performance of CarbonTracker in upper atmosphere (330 to 0 hPa) by comparing its simulations with in-situ measurements from 9 AirCore profiles (Karion et al., 2010) sampled in 2012-2014. AirCore is a ~150 m stainless steel tube that utilizes changes in ambient pressure for passive sampling of the vertical profile. It is released using balloons and it collects a continuous sample as it descends. It is then measured by an analyzer after it is recovered. More information about AirCore system can also be found at https://www.esrl.noaa.gov/gmd/ccgg/aircore/. All 9 AirCore profiles are taken near SGP and CAR sites. Figure 2 shows examples of AirCore profiles compared with CT2015 in the upper atmosphere, which demonstrates good agreement. We also compare partial column (330 to 0 hPa) averages from the 9 AirCore profiles and CT2015. Results from CT2015 agree generally well with AirCore, with difference ranging from 0.03 to 1.22 ppm (mean value equals 0.66 ppm), which suggests that CT2015 may have a high bias that can contribute to 0.66×1/3=0.22 ppm overestimate on average to the total column average. However, AirCore is in the process of rigorous evaluation, the differences between AirCore and CT2015 are not well characterized yet since we only have a limited amount of AirCore data. It is unclear whether the potential bias of CT2015 in this partial column is dependent on time or sampling location. Adding a constant bias correction to all regions will not change the spatial gradients that we focus on in this study. Thus no correction is applied when using CT2015 simulations to represent the upper 1/3 of the total column.

For uncertainty estimates, we use a 'bootstrap' method that uses random resampling and repetition of individual vertical profiles (low bias due to high humidity was corrected), with 100 Monte Carlo runs for each column average calculation. Uncertainty is then defined as one standard deviation of the 100 Monte Carlo results.

3 Results and Discussions

3.1 Seasonal patterns and spatial gradients

Typically one aircraft profile contains measurements at 12 different altitudes. Column ΔXCO_2 can be computed for each profile using the method described in Sect. 2.3 (Fig. S1). Figure 3 shows aircraft (at all altitudes) and tower data (daily averages for 10:00-17:00 LST data) from all sites used in this study. Aircraft data above 2 km exhibit much smaller seasonal variations than the full dataset, because the variations are mainly driven by CO_2 sources and sinks near Earth's surface. CO_2 concentration is enhanced in the shallow wintertime PBL primarily due to reduced plant photosynthesis and ecosystem respiration combined with slightly increased fossil fuel emissions. During summer the PBL is deeper, and depletions within the PBL are due to strong terrestrial uptake that dominates over emissions especially during June through August. During summer of 2010 to 2012, CO_2 from aircraft measurements appears higher than other years in Fig.3; however, similar characteristics are not present in tower data. This difference is due to a decrease in sampling frequency at several aircraft sites that resulted in an aliased picture of the

full summer ~~drawdown~~signals. Since we focus on climatological mean of 11 years of data in our study, this influence is eliminated by combining 11 years of data together into one "average year".

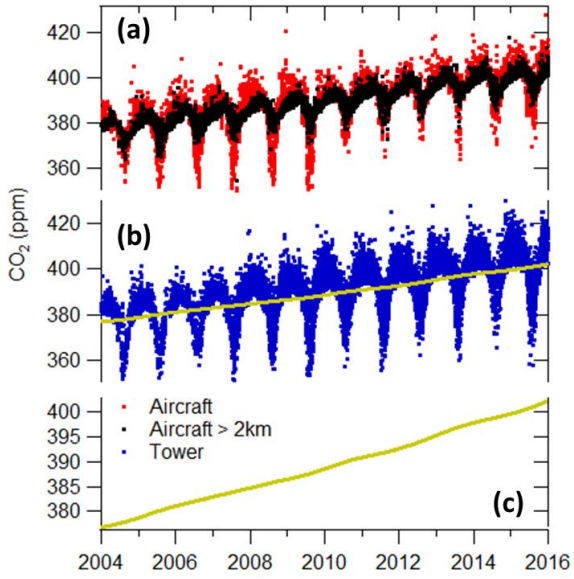


Fig. 3. CO_2 observations from aircraft (a) and towers (b). The yellow line in (b) illustrates the deseasonalized trend at Mauna Loa (MLO), same as in (c), in which y-axis expanded.

To investigate the contributions of different altitudes to spatial gradients between regions, we divided all measurement data into three layers according to their sampling altitudes: below 2 km, 2 - 5km, and 5 - 8.5 km masl (Fig. 4). Smooth seasonal curves are attained from fitting data with four harmonics using the method described by Thoning et al. (1989). The peak-to-valley amplitudes of the seasonal cycles below 2 km are the largest among the three layers for most regions, with a minimum of 10.3 ppm in SM and a maximum of 25.0 ppm in MW. The seasonal variation amplitudes decrease to 7.7-11.5 ppm in the 2 - 5 km layer, and further decrease to 7.2-10.0 ppm in the 5 - 8.5 km layer. We also observe that the seasonal cycle drawdown occurs later in the layers above 2 km (see Fig. S2, which provides similar information as Fig. 4, but seasonal curves from different vertical layers are grouped by regions to facilitate comparisons of the phases of seasonal cycles). The seasonal CO_2 drawdown below 2 km is mainly influenced by terrestrial photosynthesis and gradients are influenced by local to regional fluxes, with an earlier onset of drawdown in southern regions than in northern regions. The seasonal cycle aloft is damped and lagged compared to the PBL, with influences from throughout the Northern Hemisphere and with spatial gradients likely driven by large-scale transport. The NW, SW, SM, and SE inflow regions have significant delays of more than one month in the 2 - 5 km layer compared with the surface layer, which is likely due to the delayed phase of the

seasonal cycle in well-mixed air coming from the oceans. Vertical homogeneity of air over ocean was observed during the HIAPER Pole-to-Pole Observations (HIPPO) aircraft campaign (Wofsy et al., 2011; Frankenberg et al., 2016). As air masses are transported further inland, we observe reduced discrepancies of the timing of CO_2 drawdown between surface and upper layer air (2-5 km), which may be associated with the increased influence of the land surface in the mid-troposphere due to strong convection over land. CO_2 drawdown in the 5 - 8.5 km layers also occurs later than in the 2 - 5 km layers in most regions; however, differences between these two layers are small. The declining amplitude and delayed phase of the seasonal cycle with altitude have been noted often (e.g., Tanaka et al., 1983; Ramonet et al., 2002; Gerbig et al., 2003, Sweeney et al. 2015). It demonstrates that there is lot of important information in the vertical profile that is diminished in observations of the total column.

We find that the largest horizontal spatial gradients between regions occur below 2 km during summer time (Fig. 4), with a maximum difference of ~15.5 ppm between MW and SM. SM and SW exhibit less pronounced seasonal cycles, which is likely associated with air masses from the Gulf of Mexico and the Pacific Ocean, respectively, whereas MW exhibits a deep summer drawdown (amplitude in seasonal cycles) partially as a result of strong regional forest and crop uptake. Crevoisier et al. (2010) estimated the surface flux over North America using vertical CO_2 measurements and average wind vectors, and reported that annually averaged land carbon flux at the western (including SW region) and southern regions (including SM region) were neutral. The SE region also demonstrates a less pronounced seasonal cycle with ~~weaker summer drawdown~~ higher summertime levels compared with other northern regions, which may be due to the sea-breeze influence in summer within PBL. In wintertime, CO_2 levels in NE and MW are higher than in other regions, which result from regional fossil fuel and terrestrial biogenic emissions combined with transport from the west and south.

Higher altitude data (above 2 km) exhibit only small spatial gradients. In the 2 - 5 km layer, the largest gradient is 4 ppm in summer (Fig. 4b). It further decreases to less than 3 ppm in the 5 - 8.5 km layer (Fig. 4c). Figure 4d shows modeled CO_2 mole fractions from CT2015 for the upper troposphere and above (330 hPa to 0 hPa), which are used to fill in above the aircraft profiles for calculation of total column ΔXCO_2. Spatial gradients in this layer are less than 0.5 ppm, suggesting that the top third of the total column has little contribution to the spatial gradients of the total column.

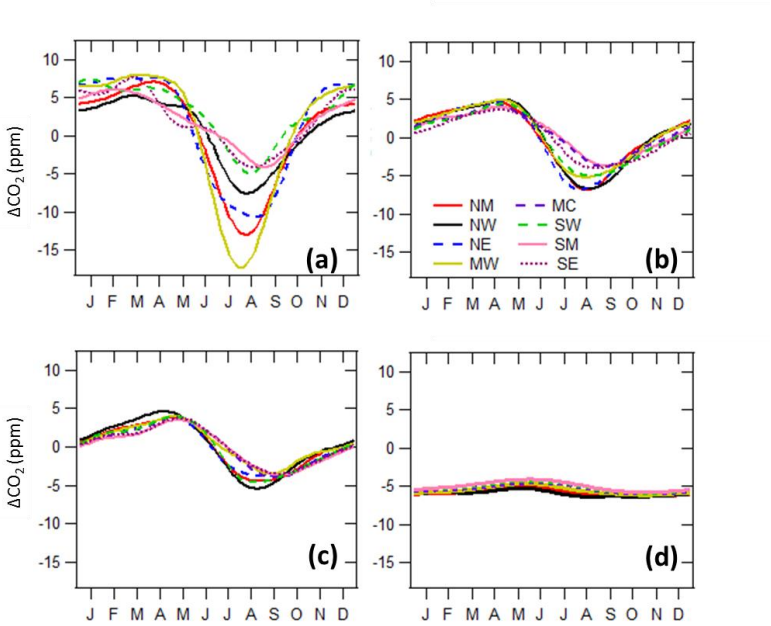


Fig. 4. Multi-year (2004-2014) average smooth seasonal curves of CO_2 relative to the long-term de-seasonalized trend at Mauna Loa for different vertical layers: (a). Aircraft and tower data under 2 km, MC is not presented because only limited data were available due to high surface elevations (>1.5 km on average) in this region; (b). Aircraft data from 2 - 5 km; (c). Aircraft data from 5 - 8.5 km; (d). CT2015 model results for layers above 330 hPa (~8.5 km) to 0 hPa (~80 km).

3.2 Long-term mean vertical profiles

To investigate the mean spatial gradients, we first calculate the long-term mean monthly vertical profiles as described in Sect. 2.3. In addition, each tower serves as one additional layer in the mean profile. The long-term mean tower data generally fit well in the vertical profiles from measurements of aircraft samples (Fig. 5 and Fig. 6), suggesting that the biases described in Sect. 2.1 above do not significantly affect the long-term mean. To attain profiles of the entire atmospheric column, upper layers (330 to 0 hPa) are filled in by CT2015, and the lowest data point of the measured profile is extended to ground level, defined by the mean surface elevation in that region.

Figure 5 presents two examples of long term mean profiles with data variability, which is the one standard deviation for each 40 hPa bin of aircraft data or for all flight-day tower data. Variability as large as 20 ppm is seen within the PBL in the MW region in summer, which is due to strong and heterogeneous surface vegetation uptake and ecosystem respiration combined with day-to-day changes in wind direction. All long-term mean monthly vertical profiles are presented in Fig. 6, which shows the mean temporal and vertical variability of CO_2 in each season, and further demonstrates the vertical propagation of seasonal CO_2 due to changes of surface flux. In

wintertime, monotonic decrease of CO_2 with altitude can be observed from all regions, in which high PBL CO_2 is mainly driven by surface emissions and reduced vertical mixing (Denning et al., 1998; Stephens et al., 2007). Surface CO_2 decreases dramatically in the growing season in those regions influenced by high plant activity, such as NM and MW regions. For the summer vertical profiles in NE and SE region (east coast of the U.S.), the CO_2 mixing ratio is elevated in the layer under 900 hPa followed by significant decreases in upper layers until 750 hPa, and then increases with altitude until tropopause (Fig. 6). This is likely a feature of sea breeze influence. Lower-troposphere air from the sea, lacking terrestrial uptake of CO_2, typically has higher CO_2 in summer compared with inland air. Polluted air previously advected offshore can be brought back along with sea breeze. Without significant vertical mixing over the marine surface, high levels of pollutants remain in those air masses. The convergence of sea breeze with prevailing wind moving offshore may create a period with a stalled frontal structure that can aggregate air pollutants (Banta et al., 2005). The convective internal boundary layer structure of the sea breeze system can significantly reduce mixing height (Miller et al., 2003), and also induces higher CO_2 levels. When the sea breeze is not dominant, air advected from southwest and west (the land) can also bring in polluted air with high CO_2 since this region is downwind of continental U.S. emissions (Miller et al., 2012).

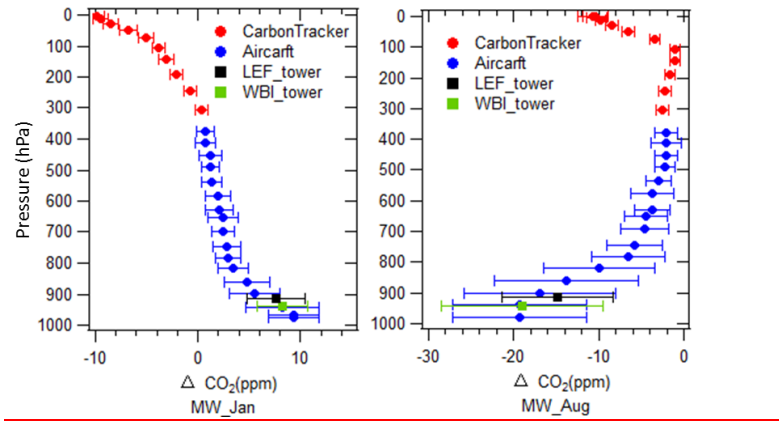
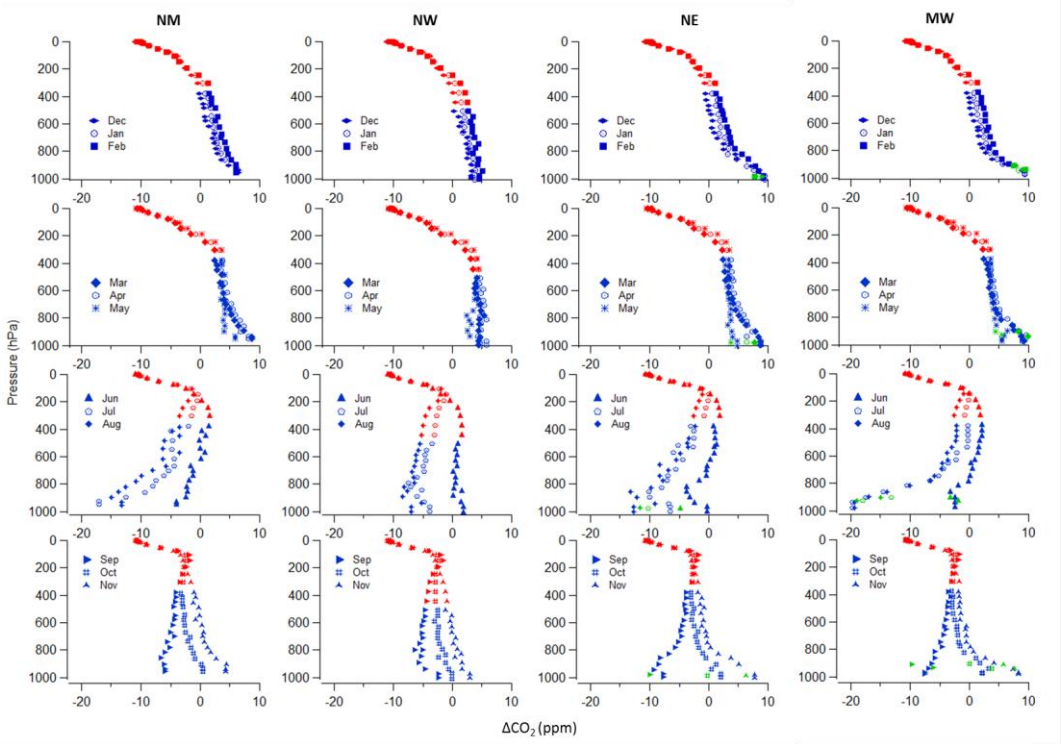


Fig. 5. Long-term mean (2004-2014) average vertical profiles in January (left panel) and August (right panel) in region MW. Error bar shows one standard deviation.



306

307 **Fig. 6a**. Long-term mean (2004-2014) monthly vertical profiles in NM, NW, NE, MW (by column, from left to
308 right). Blue points were calculated from observations, red points were calculated from CT2015, and green points
309 were calculated from tower data.

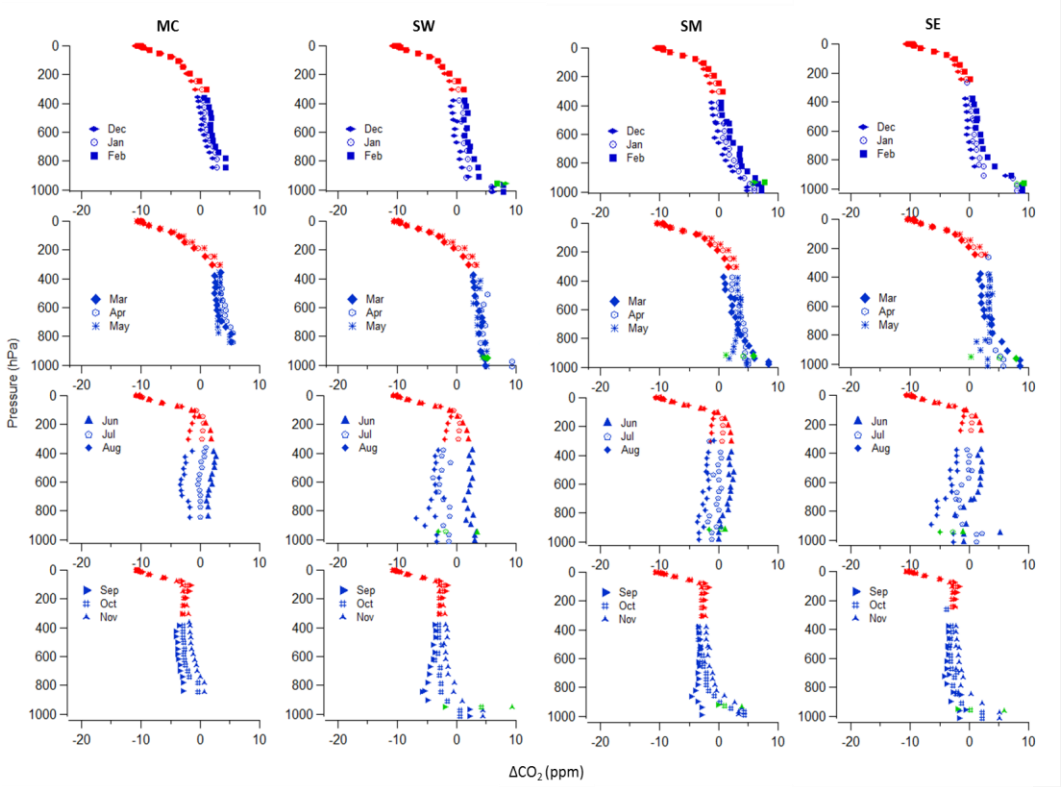


Fig. 6b. Long-term mean (2004-2014) monthly vertical profiles in MC, SW, SM, SE (by column, from left to right). Blue points were calculated from observations, red points were calculated from CT2015, and green points were calculated from tower data.

3.3 Partial column ΔXCO_2 and total column ΔXCO_2

Seasonal variations of monthly averaged partial column ΔXCO_2 (below 330 hPa) demonstrate maximum values in April and minimum values in August or September (Fig. 7a). The largest amplitude appears in NM, with peak-to-valley difference up to 13.5 ppm. SW, SM, SE, and MC have similar amplitudes of 7-8 ppm, smaller than other regions. To evaluate the performance of CT2015 on column ΔXCO_2, CT2015 results are sampled to match the latitude, longitude, altitude and time of actual measurements. Note that aircraft profiles are not assimilated in CT2015, so aircraft data are independent of the CT2015 data assimilation. Figure 7b shows monthly partial columns of ΔXCO_2 calculated from CT2015, which demonstrate good agreement with results from measurements. Only small seasonal biases exist in CT2015, with high bias occurring mostly in spring and early summer and low bias in September and October (Fig. S3). The overall differences of monthly partial column ΔXCO_2 (CT2015 - measurements) mainly fall in the range of -0.64 ppm (5^{th} percentile) to 0.84 ppm (95^{th} percentile) with a mean difference of 0.13 ppm. These differences are of similar magnitude to the uncertainties of partial column ΔXCO_2 calculated from the measurements (Fig. S4). It is clear that CT2015 captures the long-term mean variations of both phase and amplitude of partial column XCO_2 reasonably well when compared with well-calibrated measurements across North America.

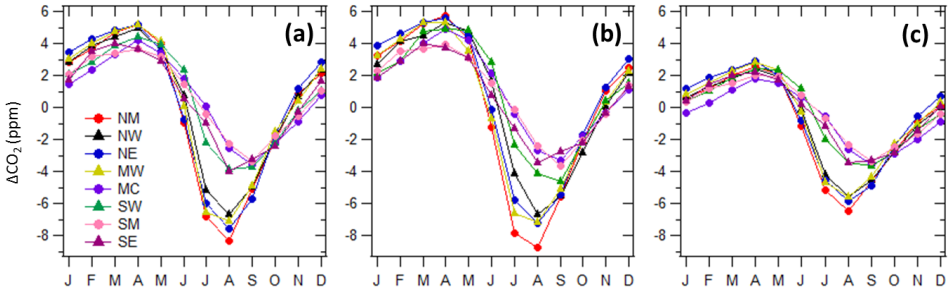


Fig. 7. (a). Partial column ΔXCO_2 calculated from aircraft and tower data; (b). Partial column ΔXCO_2 calculated from CT2015; (c). Total column ΔXCO_2 calculated from aircraft and tower data and including the top layer data from CT2015.

Total column ΔXCO_2 is presented in Fig. 7c. In regions NW, NM, NE, and MW, seasonal variations of total column ΔXCO_2 are very similar in both phase and amplitude (8-9 ppm peak to valley). For SW, SM, SE, and MC, amplitudes are ~5.5 ppm. The smallest spatial gradients occur during May and October, which result in maximum differences among all regions of only 0.9 and 0.7 ppm, respectively. The largest spatial gradients occur during June, July and August, which result in maximum differences of 2.4, 4.5, and 4.1 ppm, respectively. It is interesting that the deepest drawdown (largest amplitude in seasonal cycles) is seen in region NM, not in region MW that encompasses

the very intensive agricultural activities in the U.S. mid-west, which suggest the possibility of strong upwind influence in the NM region. Transported signals have significant influences on total column CO_2. ~~—~~ The summer ~~drawdown of~~ total column ΔXCO_2, represented by the June to August average from CT2015, has a magnitude that is similar to observations with differences no more than 1 ppm (Fig. 8). Based on the seasonal patterns of total column ΔXCO_2 (Fig. 7c) and ~~strength of summer drawdown~~the summer column ΔXCO_2 (Fig. 8), we can separate the eight regions into two groups. The group with NW, NM, NE, and MW, has ~3 ppm stronger drawdown (larger amplitude) than the group with SW, SM, SE, and MC. For winter total column ΔXCO_2 (December to February average), the maximum spatial difference is only 1.6 ppm, with the highest total column ΔXCO_2 of 1.2 ppm in NE and the lowest value of -0.3 ppm in MC.

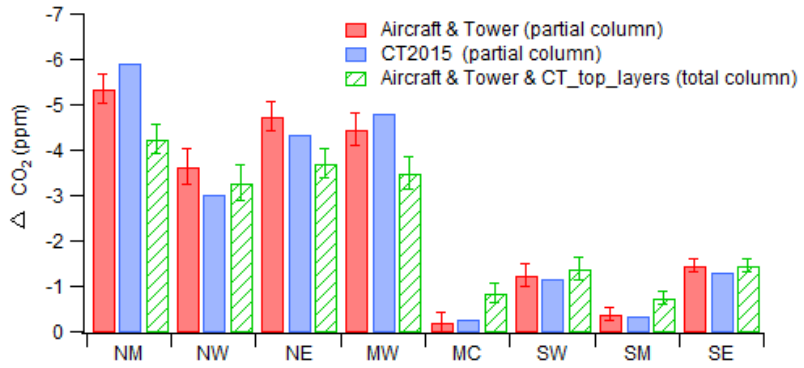


Fig. 8. Long-term mean (2004-2014) June to August partial and total column ΔXCO_2. Error bars represent one standard deviation from the bootstrap uncertainty calculation (see Sect. 2.3).

3.4 Influence of large scale circulation

Figure 9 shows long-term mean summer column ΔXCO_2 calculated from CT2015, together with full column ΔXCO_2 from individual aircraft sites (note that some aircraft sites have less than 11 years of data the CT2015 shows in Fig. 9, only aircraft sites with more than 6 years of data are presented, the actual values are provided in Table S2). The fact that total column ΔXCO_2 from CT2015 agrees well with aircraft sites also supports the performance of CT2015 on a long-term average basis. The observations show a similar summer spatial pattern, with lower column ΔXCO_2 in the north and northeast regions and higher column ΔXCO_2 in the south and southwest regions (Fig. 9a). Scattered hot spots of high column ΔXCO_2 associated with surface emissions from megacities, or cold spots associated with strong local uptake, are not or just barely visible in the long-term average column ΔXCO_2 map at $1^{o}x1^{o}$ resolution. Instead, the wave-like pattern of column ΔXCO_2 over North America reflects large scale circulation. To support our hypothesis on the influence of large scale circulation, we analyze the long term mean wind pattern over North America. We can see that air masses from northwest of the continent bring in low average

Formatted: Subscript

column ΔXCO_2, while air masses from the south (mainly the subtropical Pacific Ocean and the Gulf of Mexico) bring in high column ΔXCO_2 (Fig. 9b). The zonal gradients over the continent, especially north of 40° N, also reflect long-term average wind patterns; southwest wind corresponds to higher column ΔXCO_2 over the western part of the continent until the wind direction shifts to west-northwest over the eastern part of the continent. This wind pattern matches well with the geographic division of the over/under -3 ppm areas colored in green/blue in the column ΔXCO_2 map (Fig. 9b). Figure 9c and 9d shows partial column averages for free troposphere (800-330 hPa) and lower troposphere (below 800 hPa), respectively. The free troposphere spatial gradient also demonstrates a wave-like pattern. A previous study on the total column CO_2 from ground based Total Carbon Column Observation Network (TCCON) found strong correlation between the mid-latitude column CO_2 and synoptic-scale variation of potential temperature (θ, at 700 hPa), a dynamic tracer for adiabatic air transport (Keppel-Aleks et al., 2012). Thus they also propose that the variations in column CO_2 are mainly driven by large-scale flux and transport. Analysis on the interannual variability of the seasonal cycle amplitudes of column CO_2 in North Hemisphere has also found significant contribution of large-scale circulations to the north-south gradient (Wunch et al., 2013).

The strong drawdown over northeast North America in summer is a consequence of long-range transport of low CO_2 from northeast Eurasia, in addition to regional terrestrial uptake. Sweeney et al. (2015) notes well-mixed vertical profiles (up to 8 km) of CO_2, CO, CH_4, N_2O, and SF_6 from THD, ESP and PFA (Poker Flat, Alaska; 65.07°, -147.29°) sites and suggests that air coming across the Pacific was strongly influenced by Asian surface fluxes before being vertically homogenized as it passed over the Pacific Ocean. This well-mixed air forms an important boundary condition in the column CO_2 of air coming into the North American continent. This was best illustrated at sites like PFA where the summertime minimum in CO_2 significantly preceded maximum ecosystem uptake of CO_2, implying significant influence of transported air from lower latitude regions from Asia. We further conduct an experiment using Carbon Tracker to investigate the importance of this effect. A control run and a "masked run" are conducted for 2010-2012, in which the Eurasian boreal flux is turned on/off. The MLO CO_2 trend from each model scenario is used as reference background and thus removed before total column ΔXCO_2 calculation. Figure 10 shows the results for 2012 summer, which is an average summer when compared with the 2004-2014 mean pattern (Fig. 9 and Fig. 11). The maximum north-south difference reduces to ~2.5 ppm after we turn off the Eurasian boreal flux, compared with ~5 ppm from the control run. In both control and masked scenarios, the free troposphere partial ΔXCO_2 demonstrates similar spatial patterns as for total column ΔXCO_2 (Fig. S5). This result combined with results from Sweeney et al. (2015) demonstrates that the transport of low CO_2 resulting from large summertime Eurasian boreal uptake has a large contribution on the overall summer total column CO_2 drawdown in North America.

Formatted: Subscript

Formatted: Not Superscript/ Subscript

Formatted: Not Superscript/ Subscript

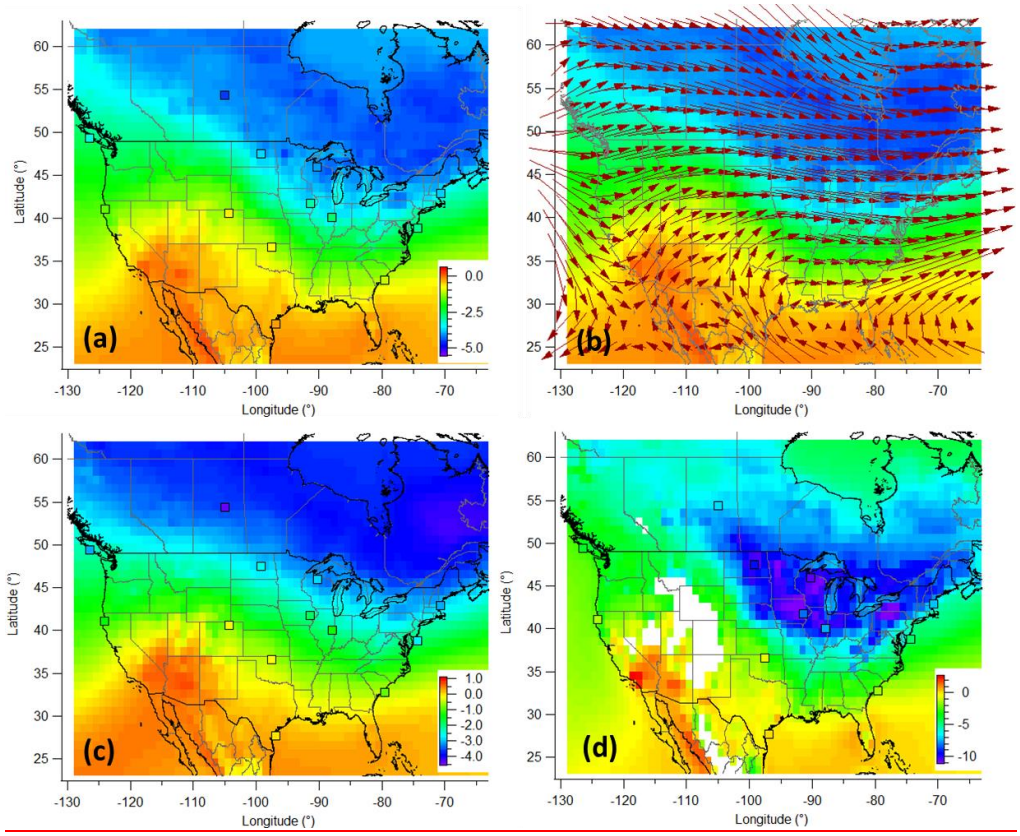


Fig. 9. Long-term mean (2004-2014) June-August total column ΔXCO_2 from CT2015 in 1° × 1° spatial resolution with total column ΔXCO_2 for 13 individual aircraft sites in squares (a), and CT2015 column ΔXCO_2 overlaid with pressure-weighted (1000 hPa to 500 hPa) mean wind vectors for the same period (b). (c) and (d) are similar as (a), except for free troposphere (800 to 330 hPa) and lower troposphere (below 800 hPa), respectively. Note the different color scales.

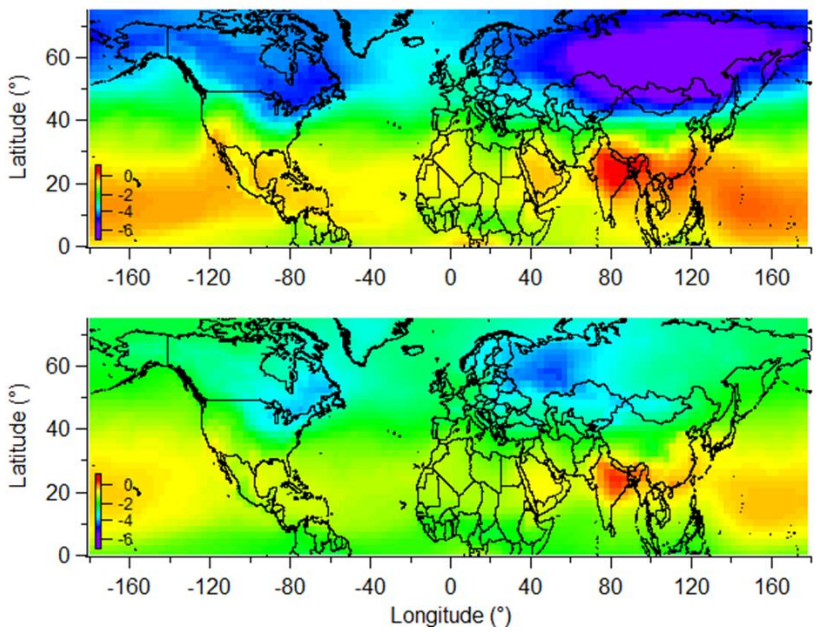


Fig. 10. Total column ΔXCO_2 from Carbon Tracker control (top panel) and masked (bottom panel, Eurasian boreal flux is masked) runs for 2012 June-August (3° × 2° spatial resolution). MLO trend from each individual scenario is removed before the ΔXCO_2 calculation. Same color scale is used as in Fig. 9a. Partial column ΔXCO_2 patterns for free troposphere (800 to 330 hPa) and lower troposphere (below 800 hPa) are provided in SI.

Formatted: Not Superscript/ Subscript

3.5 A comparison with apparent gradients over Europe

Figure 11 shows the climatological June - August mean modeled global column ΔXCO_2 map in 3° × 2° spatial resolution, which presents smooth wave-like patterns. Reuter et al. (2014) use SCIAMACHY and GOSAT satellite retrievals of column CO_2 and inverse modelling to attain surface CO_2 flux over European region, and suggest a large uptake of CO_2 in this region. Column ΔXCO_2 from CT2015 (Fig. 11) exhibits a drastically different summer spatial pattern over Europe compared with the eight year mean (2003-2010) June through August satellite retrievals presented by Reuter et al. (2014, their Fig. 2a). The spatial gradient from CT2015 results in a maximum 3-4 ppm difference and a gradual pattern, instead of as much as 6 ppm from satellite retrievals. There is no sign of XCO_2 hot spots from surface emissions or removals in the CT2015 spatial pattern over Europe (Fig. 11), in contrast to several hot spots that are apparent from the 8-year averaged SCIAMACHY satellite retrievals over Ireland, U.K., northeast of France, Belgium, Netherland, north of Germany, and south of Sweden, and low spots over the Ukraine and Kazakhstan (Reuter et al., 2014). Although the NOAA/ESRL CT2015 (https://www.esrl.noaa.gov/gmd/ccgg/carbontracker/CT2015/) assimilates fewer observations over Europe than Carbon Tracker Europe (http://www.carbontracker.eu/), both models produced similar fluxes over the European

region (see both websites for detailed fluxes). The 3° × 2° grid from CT2015 is not likely responsible for a much smoother pattern for Carbon Tracker, compared with the 2° × 2° grid from satellite retrievals (Reuter et al., 2014) . The North America region on the $3^0 \times 2^0$ grid in Fig. 11 shows similar pattern as the 1° × 1° grid in Fig. 9, with similar spatial difference of ~ 5 ppm. A smoother spatial distribution should be expected in Europe for the long-term mean column XCO_2 (Fig. 11) due to the influences of dominating west and southwest winds in summer. Since the satellite retrievals in Reuter et al. (2014) appear to show unrealistic column XCO_2 spatial gradients over Europe, they should not be used to derive estimates of a European carbon sink. A recent study (Feng et al., 2016) using inverse modeling suggests that satellite retrievals outside the immediate European region and a small bias of 0.5 ppm were sufficient to produce the apparent large carbon sink in the study of Reuter et al. (2014). This is expected from elementary mass balance considerations as in Sec.1. Spatial gradients are the fundamental signals to infer regional fluxes. Since spatial gradients from CT2015 are realistic, boreal fluxes inferred by CT2015, which shows 0.03 ± 2.33 Pg C yr^{-1} for Europe, should be more trustworthy than fluxes estimated based on unrealistic spatial pattern. However, the European carbon sink is still inclusive; the discrepancies among different methods and results are further discussed by Reuter et al. (2017). Increasing the amount of highly precise observations such as the well-calibrated surface measurements and vertical profiles can greatly help to estimate the carbon sink.

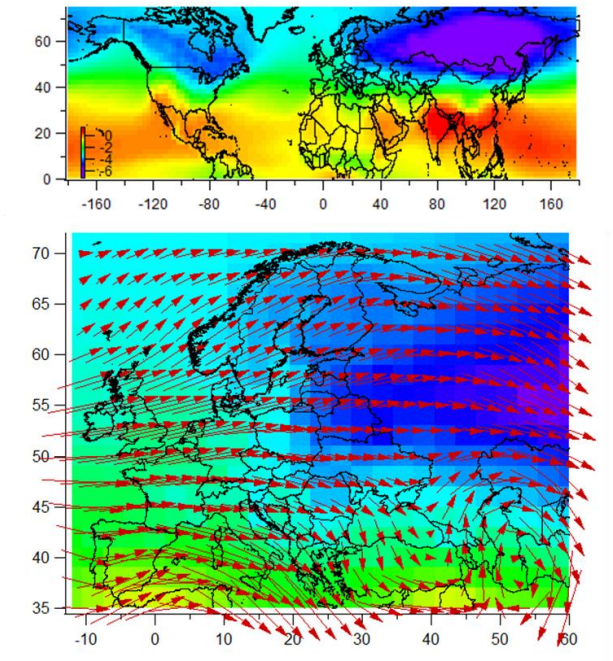


Fig. 11. Long-term mean (2004-2014) June - August total column ΔXCO_2 from CT2015 (top panel) in 3° × 2° spatial resolution, and zoom-in for Europe overlaid with pressure-weighted (1000 hPa to 500 hPa) mean wind

vectors for the same period (bottom panel). The color scale is the same as in Fig. 9a, which is scaled to reflect 6 ppm difference of XCO_2 to compare with satellite retrievals from Reuter et al. (their Fig. 2a, 2014).

4 Conclusion

Aircraft and tall tower measurements from the NOAA GGGRN provide detailed information describing the long-term average temporal and spatial variations of CO_2 in the PBL and the free troposphere. These data provide valuable constraints for evaluating model simulations and satellite retrievals. Seasonal cycle peak-to-peak amplitudes of CO_2 are largest below 2 km, where those maximum values are about twice those in the vertical layers above, indicating that most of the information on surface sources and sinks resides in the continental PBL. Large spatial gradients of CO_2 over North America are observed below 2 km during summer (with maximum difference of ~15.5 ppm between MW and SM), while higher altitude data (above 2 km) have much smaller contributions to spatial gradients, with a maximum difference of ~~only~~ 4 ppm. The spatial differences of CO_2 in the upper troposphere and above (330 hPa to 0 hPa) are less than 0.5 ppm, according to CT2015. Comparison with Aircore measurements shows that the upper troposphere and lower stratospheric simulations from CT2015 are reasonably trustworthy.

Our long-term mean vertical profiles show that tower data agree well with aircraft data at similar vertical levels. Partial column ΔXCO_2 was calculated from the long-term mean vertical profiles. By comparing the partial column ΔXCO_2 from measurements with those from CT2015, we verify that CT2015 captures the long-term mean patterns of both phase and amplitude of partial ΔXCO_2.

Large spatial gradients of ΔXCO_2 only appeared in summer, during which time the north and northeast regions had ~3 ppm stronger drawdowns than the south and southwest regions. Scattered hot spots of high column ΔXCO_2 associated with surface emissions from megacities, or cold spots associated with strong local uptake, are not or just barely visible in the long-term average column ΔXCO_2. Instead, the wave-like pattern of column ΔXCO_2 over North America matches well with large scale circulation. ~~By comparing the spatial gradients of ΔXCO_2 with wind vectors across North America, we find that total column ΔXCO_2 patterns are equally affected by large scale circulation patterns as by regional surface sources and sinks.~~ A CarbonTracker experiment to investigate the impact of Eurasian long-range transport suggests that the large summer time Eurasian boreal flux alone contributes about half of the north-south column ΔXCO_2 gradient across North America. Considering the transported signals from other upwind regions, including northern Canada, we expect that the transported signals overall has major contribution to the column ΔXCO_2 spatial gradient.

Formatted: Not Superscript/ Subscript

Author contributions

Xin Lan was responsible for study design, data analysis, and manuscript writing. Pieter Tans was responsible for study design, data analysis, and manuscript improvement. Colm Sweeney and Arlyn Andrews provided measurement data and improved manuscript. Andrew Jacobson provided modelled data and improved manuscript. Edward Dlugokencky analyzed measurements and ensured data quality, and improved manuscript. Jonathan Kofler

conducted tower measurements and improved manuscript. Molly Crotwell, Patricia Lang, and Sonja Wolter analyzed measurements and ensured data quality. Kirk Thoning provided data smoothing method.

Acknowledgements

We especially thank John Mund for extracting NARR meteorological variables for our measurements. This research was supported by a fellowship from the National Research Council Research Associateship Programs.

References

Andrews, A. E., Kofler, J. D., Trudeau, M. E., Williams, J. C., Neff, D. H., Masarie, K. A., Chao, D. Y., Kitzis, D. R., Novelli, P. C., Zhao, C. L., Dlugokencky, E. J., Lang, P. M., Crotwell, M. J., Fischer, M. L., Parker, M. J., Lee, J. T., Baumann, D. D., Desai, A. R., Stanier, C. O., De Wekker, S. F. J., Wolfe, D. E., Munger, J. W. and Tans, P. P.: CO2, CO, and CH4 measurements from tall towers in the NOAA Earth System Research Laboratory's Global Greenhouse Gas Reference Network: instrumentation, uncertainty analysis, and recommendations for future high-accuracy greenhouse gas monitoring efforts, Atmos. Meas. Tech., 7, 647-687, 2014.

Banta, R. M., Senff, C. J., Nielsen-Gammon, J., Darby, L. S., Ryerson, T. B., Alvarez, R. J., Sandberg, S. R., Williams, E. J. and Trainer, M.: A bad air day in Houston', B. Am. Meteorol. Soc., 86, 657, DOI: http://dx.doi.org/10.1175/BAMS-86-5-657, 2005.

Biraud, S. C., Torn, M. S., Smith, J. R., Sweeney, C., Riley, W. J. and Tans, P. P.:A multi-year record of airborne CO2 observations in the US Southern Great Plains, Atmos. Meas. Tech., 6, 751-763, 2013.

Buchwitz, M., Reuter, M., Bovensmann, H., Pillai, D., Heymann, J., Schneising, O., Rozanov, V., Krings, T., Burrows, J. P., Boesch, H., Gerbig, C., Meijer, Y. and Loscher, A. : Carbon Monitoring Satellite (CarbonSat): assessment of atmospheric CO2 and CH4 retrieval errors by error parameterization, Atmos. Meas. Tech., 6, 3477-3500, 2013.

Chevallier, F., Breon, F. M. and Rayner, P. J.: Contribution of the Orbiting Carbon Observatory to the estimation of CO2 sources and sinks: Theoretical study in a variational data assimilation framework, J. Geophys. Res. Atmos., 112, D09307, doi:10.1029/2006JD007375, 2007.

Choi, Y. H., Vay, S. A., Vadrevu, K. P., Soja, A. J., Woo, J. H., Nolf, S. R., Sachse, G. W., Diskin, G. S., Blake, D. R., Blake, N. J., Singh, H. B., Avery, M. A., Fried, A., Pfister, L. and Fuelberg, H. E.: Characteristics of the atmospheric CO2 signal as observed over the conterminous United States during INTEX-NA, J. Geophys. Res. Atmos., 113, D07301, doi:10.1029/2007JD008899, 2008.

Ciais, P., Rayner, P., Chevallier, F., Bousquet, P., Logan, M., Peylin, P. and Ramonet, M.: Atmospheric inversions for estimating CO2 fluxes: methods and perspectives, Climatic Change, 103, 69-92, 2010.

Conway, T. J., Tans, P. P., Waterman, L. S. and Thoning, K. W.: Evidence for interannual variability of the carbon-cycle from the national-oceanic-and-atmospheric-administration climate-monitoring-and-diagnostics-laboratory global-air-sampling-network, J. Geophys. Res. Atmos., 99, 22831-22855, 1994.

513 Crevoisier, C., Sweeney, C., Gloor, M., Sarmiento, J. L. and Tans, P. P.: Regional US carbon sinks from three-
514 dimensional atmospheric CO2 sampling, Proc. Natl. Acad. Sci. U. S. A., 107, 18348-18353, 2010.
515 Crisp, D., Fisher, B. M., O'Dell, C., Frankenberg, C., Basilio, R., Bosch, H., Brown, L. R., Castano, R., Connor, B.,
516 Deutscher, ¨ N. M., Eldering, A., Griffith, D., Gunson, M., Kuze, A., Mandrake, L., McDuffie, J.,
517 Messerschmidt, J., Miller, C. E., Morino, I., Natraj, V., Notholt, J., O'Brien, D. M., Oyafuso, F., Polonsky, I.,
518 Robinson, J., Salawitch, R., Sherlock, V., Smyth, M., Suto, H., Taylor, T. E., Thompson, D. R., Wennberg, P.
519 O., Wunch, D., and Yung, Y. L.: The ACOS CO2 retrieval algorithm – Part II: Global XCO2 data
520 characterization, Atmos. Meas. Tech., 5, 687–707, doi:10.5194/amt-5-687-2012, 2012.
521 Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A.,
522 Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C.,
523 Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Holm, E. V., Isaksen, L.,
524 Kallberg, P., Kohler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K.,
525 Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J. N. and Vitart, F.: The ERA-Interim reanalysis:
526 configuration and performance of the data assimilation system, Q.J.R. Meteorol. Soc., 137, 553-597, 2011.
527 Denning, A. S., Takahashi, T., and Friedlingstein, P.: Can a strong atmospheric CO2 rectifier effect be reconciled
528 with a "reasonable" carbon budget?, Tellus B, 51, 249–253, 1999.
529 Feng, L., Palmer, P. I., Parker, R. J., Deutscher, N. M., Feist, D. G., Kivi, R., Morino, I. and Sussmann, R.:
530 Estimates of European uptake of CO2 inferred from GOSAT X-CO2 retrievals: sensitivity to measurement bias
531 inside and outside Europe, Atmos. Chem. Phys., 16, 1289-1302, 2016.
532 Frankenberg, C., Kulawik, S. S., Wofsy, S., Chevallier, F., Daube, B., Kort, E. A., O'Dell, C., Olsen, E. T., and
533 Osterman, G.: Using airborne HIAPER Pole-to-Pole Observations (HIPPO) to evaluate model and remote
534 sensing estimates of atmospheric carbon dioxide, Atmos. Chem. Phys., , 16, 7867–7878, 2016. doi:10.5194/acp-
535 16-7867-2016, 2016.
536 Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S. and Grainger, C.
537 A.: Toward constraining regional-scale fluxes of CO2 with atmospheric observations over a continent: 1.
538 Observed spatial variability from airborne platforms J. Geophys. Res. Atmos., 108, doi:10.1029/2002JD003018,
539 2003.
540 Chevallier, F., Palmer, P. I., Feng, L., Boesch, H., O'Dell, C. W., and Bousquet, P.: Towards robust and consistent
541 regional CO_2 flux estimates from in situ and space-borne measurements of atmospheric CO_2, Geophys. Res.
542 Lett., 41, 1065–1070, doi:10.1002/2013GL058772, 2014.
543 Gourdji, S. M., Mueller, K. L., Yadav, V., Huntzinger, D. N., Andrews, A. E., Trudeau, M., Petron, G., Nehrkorn,
544 T., Eluszkiewicz, J., Henderson, J., Wen, D., Lin, J., Fischer, M., Sweeney, C. and Michalak, A. M.: North
545 American CO2 exchange: inter-comparison of modeled estimates with results from a fine-scale atmospheric
546 inversion, Biogeosci., 9, 457-475, 2012
547 Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y. H., Ciais,
548 P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Maki, T., Maksyutov, S., Masarie, K.,
549 Peylin, P., Prather, M., Pak, B. C., Randerson, J., Sarmiento, J., Taguchi, S., Takahashi, T. and Yuen, C. W.:

Formatted: Subscript

Formatted: Subscript

550 Towards robust regional estimates of CO2 sources and sinks using atmospheric transport models, Nature,
551 415(6872), 626-630, 2002.
552 Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Pak, B. C., Baker, D., Bousquet, P., Bruhwiler, L., Chen,
553 Y. H., Ciais, P., Fung, I. Y., Heimann, M., John, J., Maki, T., Maksyutov, S., Peylin, P., Prather, M. and
554 Taguchi, S.: Transcom 3 inversion intercomparison: Model mean results for the estimation of seasonal carbon
555 sources and sinks, Global Biogeochem. Cycles, 18, GB1010, doi:10.1029/2003GB002111, 2004.
556 Houweling, S., Breon, F. M., Aben, I., Rodenbeck, C., Gloor, M., Heimann, M. and Ciais, P.: Inverse modeling of
557 CO2 sources and sinks using satellite data: a synthetic inter-comparison of measurement techniques and their
558 performance as a function of space and time, Atmos. Chem. Phys., 4, 523-538, 2004.
559 Inoue, M., Morino, I., Uchino, O., Miyamoto, Y., Yoshida, Y., Yokota, T., Machida, T., Sawa, Y., Matsueda, H.,
560 Sweeney, C., Tans, P. P., Andrews, A. E., Biraud, S. C., Tanaka, T., Kawakami, S. and Patra, P. K.: Validation
561 of XCO2 derived from SWIR spectra of GOSAT TANSO-FTS with aircraft measurement data, Atmos. Chem.
562 Phys., 13, 9771-9788, 2013.
563 Inoue, M., Morino, I., Uchino, O., Nakatsuru, T., Yoshida, Y., Yokota, T., Wunch, D., Wennberg, P. O., Roehl, C.
564 M., Griffith, D. W. T., Velazco, V. A., Deutscher, N. M., Warneke, T., Notholt, J., Robinson, J., Sherlock, V.,
565 Hase, F., Blumenstock, T., Rettinger, M., Sussmann, R., Kyrö, E., Kivi, R., Shiomi, K., Kawakami, S., De
566 Mazière, M., Arnold, S. G., Feist, D. G., Barrow, E. A., Barney, J., Dubey, M., Schneider, M., Iraci, L.,
567 Podolske, J. R., Hillyard, P., Machida, T., Sawa, Y., Tsuboi, K., Matsueda, H., Sweeney, C., Tans, P. P.,
568 Andrews, A. E., Biraud, S. C., Fukuyama, Y., Pittman, J. V., Kort, E. A., and Tanaka, T.: Bias corrections of
569 GOSAT SWIR XCO2 and XCH4 with TCCON data and their evaluation using aircraft measurement data,
570 Atmos. Meas. Tech. 9, 3491–3512, 2016, doi:10.5194/amt-9-3491-2016.
571 Karion, A., Sweeney, C., Tans, P., and Newberger, T.: AirCore: An Innovative Atmospheric Sampling System, J.
572 Atmos. Ocean. Tech., 27, 1839–1853, doi:10.1175/2010JTECHA1448.1, 2010.
573 Karion, A., C. Sweeney, S. Wolter, T. Newberger, H. Chen, A. Andrews, J. Kofler, D. Neff, and P. Tans (2013),
574 Long-term greenhouse gas measurements from aircraft, Atmos. Meas. Tech., 6(3), 511–526, doi:10.5194/amt-6-
575 511-2013.
576 Keeling, C. D. and Rakestraw, N. W.: The concentration of carbon dioxide in the atmosphere, J. Geophys. Res., 65,
577 2502-2502, 1960.
578 Keppel-Aleks, G., Wennberg, P. O., Washenfelder, R. A., Wunch, D., Schneider, T., Toon, G. C., Andres, R. J.,
579 Blavier, J.-F., Connor, B., Davis, K. J., Desai, A. R., Messerschmidt, J., Notholt, J., Roehl, C. M., Sherlock, V.,
580 Stephens, B. B., Vay, S. A., and Wofsy, S. C.: The imprint of surface fluxes and transport on variations in total
581 column carbon dioxide, Biogeosci., 9, 875– 891, doi:10.5194/bg-9-875-2012, 2012.
582 Kulawik, S. S., Worden, J. R., Wofsy, S. C., Biraud, S. C., Nassar, R., Jones, D. B. A., Olsen, E. T., Jimenez, R.,
583 Park, S., Santoni, G. W., Daube, B. C., Pittman, J. V., Stephens, B. B., Kort, E. A., Osterman, G. B. and Team,
584 T. E. S.: Comparison of improved Aura Tropospheric Emission Spectrometer CO2 with HIPPO and SGP
585 aircraft profile measurements, Atmos. Chem. Phys., 13, 3205-3225, 2013.

586 Lauvaux, T., Schuh, A. E., Uliasz, M., Richardson, S., Miles, N., Andrews, A. E., Sweeney, C., Diaz, L. I., Martins,
587 D., Shepson, P. B. and Davis, K. J.: Constraining the CO2 budget of the corn belt: exploring uncertainties from
588 the assumptions in a mesoscale inverse system, Atmos. Chem. Phys., 12, 337-354, 2012.
589 Machida, T., Kita, K., Kondo, Y., Blake, D., Kawakami, S., Inoue, G. and Ogawa, T. : Vertical and meridional
590 distributions of the atmospheric CO2 mixing ratio between northern midlatitudes and southern subtropics, J.
591 Geophys. Res. Atmos., 108(D3), 8401, doi:10.1029/2001JD000910, 2002.
592 Machida, T., Matsueda, H., Sawa, Y., Nakagawa, Y., Hirotani, K., Kondo, N., Goto, K., Nakazawa, T., Ishikawa, K.
593 and Ogawa, T.: Worldwide Measurements of Atmospheric CO2 and Other Trace Gas Species Using
594 Commercial Airlines, J. Atmos. Ocean. Tech., 25(10), 1744-1754, 2008.
595 Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., Jovic, D., Woollen, J., Rogers,
596 E., Berbery, E. H., Ek, M. B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., and
597 Shi, W.: North American regional reanalysis, B. Am. Meteorol. Soc., 87, 343–360, doi:10.1175/BAMS-87-3-
598 343, 2006.
599 Messerschmidt, J., Geibel, M. C., Blumenstock, T., Chen, H., Deutscher, N. M., Engel, A., Feist, D. G., Gerbig, C.,
600 Gisi, M., Hase, F., Katrynski, K., Kolle, O., Lavric, J. V., Notholt, J., Palm, M., Ramonet, M., Rettinger, M.,
601 Schmidt, M., Sussmann, R., Toon, G. C., Truong, F., Warneke, T., Wennberg, P. O., Wunch, D. and Xueref-
602 Remy, I.: Calibration of TCCON column-averaged CO2: the first aircraft campaign over European TCCON
603 sites, Atmos. Chem. Phys., 11(21), 10765-10777, 2011.
604 Miller, C. E., Crisp, D., DeCola, P. L., Olsen, S. C., Randerson, J. T., Michalak, A. M., Alkhaled, A., Rayner, P.,
605 Jacob, D. J., Suntharalingam, P., Jones, D. B. A., Denning, A. S., Nicholls, M. E., Doney, S. C., Pawson, S.,
606 Boesch, H., Connor, B. J., Fung, I. Y., O'Brien, D., Salawitch, R. J., Sander, S. P., Sen, B., Tans, P., Toon, G.
607 C., Wennberg, P. O., Wofsy, S. C., Yung, Y. L. and Law, R. M.: Precision requirements for space-based X-CO2
608 data, J. Geophys. Res. Atmos., 112, D10314, doi:10.1029/2006JD007659, 2007.
609 Miller, J. B., Lehman, S. J., Montzka, S. A., Sweeney, C., Miller,B. R., Karion, A., Wolak, C., Dlugokencky, J.,
610 Southon, J., Turnbull, J. C., and Tans, P. P.: Linking emissions of fossil fuel CO2 and other anthropogenic trace
611 gases using atmospheric 14CO2, J. Geophys. Res., 117, D08302, doi:10.1029/2011JD017048, 2012.
612 Miller, S. T. K., Keim, B. D., Talbot, R. W. and Mao, H.: Sea breeze: Structure, forecasting, and impacts, Rev.
613 Geophys., 41, 1011, doi:10.1029/2003RG000124, 2003.
614 Miyamoto, Y., Inoue, M., Morino, I., Uchino, O., Yokota, T., Machida, T., Sawa, Y., Matsueda, H., Sweeney, C.,
615 Tans, P. P., Andrews, A. E. and Patra, P. K.: Atmospheric column-averaged mole fractions of carbon dioxide at
616 53 aircraft measurement sites, Atmos. Chem. Phys., 13(10), 5265-5275, 2013.
617 Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L.
618 M. P., Petron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol,
619 M. C. and Tans, P. P.: An atmospheric perspective on North American carbon dioxide exchange:
620 CarbonTracker, Proc. Natl. Acad. Sci. U. S. A., 104(48), 18925-18930, 2007.
621 Ramonet, M., Ciais, P., Nepomniachii, I., Sidorov, K., Neubert, R. E. M., Langendorfer, U., Picard, D., Kazan, V.,
622 Biraud, S., Gusti, M., Kolle, O., Schulze, E. D. and Lloyd, J.: Three years of aircraft-based trace gas

measurements over the Fyodorovskoye southern taiga forest, 300 km north-west of Moscow, Tellus B-Chem. Phys. Meteor., 54(5), 713-734, 2002.

Reuter, M., Bovensmann, H., Buchwitz, M., Burrows, J. P., Connor, B. J., Deutscher, N. M., Griffith, D. W. T., Heymann, J., Keppel-Aleks, G., Messerschmidt, J., Notholt, J., Petri, C., Robinson, J., Schneising, O., Sherlock, V., Velazco, V., Warneke, T., Wennberg, P. O. and Wunch, D.: Retrieval of atmospheric CO2 with enhanced accuracy and precision from SCIAMACHY: Validation with FTS measurements and comparison with model results, J. Geophys. Res. Atmos., 116, 2011.

Reuter, M., Buchwitz, M., Hilker, M., Heymann, J., Schneising, O., Pillai, D., Bovensmann, H., Burrows, J. P., Bosch, H., Parker, R., Butz, A., Hasekamp, O., O'Dell, C. W., Yoshida, Y., Gerbig, C., Nehrkorn, T., Deutscher, N. M., Warneke, T., Notholt, J., Hase, F., Kivi, R., Sussmann, R., Machida, T., Matsueda, H. and Sawa, Y.: Satellite-inferred European carbon sink larger than expected, Atmos. Chem. Phys., 14(24), 13739-13753, 2014.

Reuter, M., Buchwitz, M., Hilker, M., Heymann, J., Bovensmann, H., Burrows, J. P., Houweling, S., Liu, Y. Y., Nassar, F., Chevallier, F., Cias, P., Marshall, J., and Reichstein, M.: How much CO2 is taken up by the European terrestrial biosphere? Bull. Am. Meteor. Soc., 98, 665–671, doi:10.1175/BAMS-D-15-00310.1, 2017.

Saitoh, N., Kimoto, S., Sugimura, R., Imasu, R., Kawakami, S., Shiomi, K., Kuze, A., Machida, T., Sawa, Y. and Matsueda, H.: Algorithm update of the GOSAT/TANSO-FTS thermal infrared CO2 product (version 1) and validation of the UTLS CO2 data using CONTRAIL measurements, Atmos. Meas. Tech., 9(5), 2119-2134, 2016.

Stephens, B. B., Gurney, K. R., Tans, P. P., Sweeney, C., Peters, W., Bruhwiler, L., Ciais, P., Ramonet, M., Bousquet, P., Nakazawa, T., Aoki, S., Machida, T., Inoue, G., Vinnichenko, N., Lloyd, J., Jordan, A., Heimann, M., Shibistova, O., Langenfelds, R. L., Steele, L. P., Francey, R. J. and Denning, A. S.: Weak northern and strong tropical land carbon uptake from vertical profiles of atmospheric CO(2), Science, 316(5832), 1732-1735, 2007.

Sweeney, C., Karion, A., Wolter, S., Newberger, T., Guenther, D., Higgs, J. A., Andrews, A. E., Lang, P. M., Neff, D., Dlugokencky, E., Miller, J. B., Montzka, S. A., Miller, B. R., Masarie, K. A., Biraud, S. C., Novelli, P. C., Crotwell, M., Crotwell, A. M., Thoning, K. and Tans, P. P.: Seasonal climatology of CO2 across North America from aircraft measurements in the NOAA/ESRL Global Greenhouse Gas Reference Network, J. Geophys. Res. Atmos., 120(10), 5155-5190, 2015.

Tanaka, M., Nakazawa, T. and Aoki, S.: Concentration of atmospheric carbon-dioxide over japan, J. Geophys. Res. Ocean., 88(C2), 1339-1344, DOI: 10.1029/JC088iC02p01339, 1983.

Tanaka, T., Miyamoto, Y., Morino, I., Machida, T., Nagahama, T., Sawa, Y., Matsueda, H., Wunch, D., Kawakami, S. and Uchino, O.: Aircraft measurements of carbon dioxide and methane for the calibration of ground-based high-resolution Fourier Transform Spectrometers and a comparison to GOSAT data measured over Tsukuba and Moshiri, Atmos. Meas. Tech., 5(8), 2003-2012, 2012.

Formatted: Font: Not Bold

Tans, P. P., Conway, T. J. and Nakazawa, T.: Latitudinal distribution of the sources and sinks of atmospheric carbon-dioxide derived from surface observations and an atmospheric transport model, J. Geophys. Res. Atmos., 94(D4), 5151-5172, 1989.

Tans, P. P., Fung, I. Y. and Takahashi, T.: Observational constraints on the global atmospheric co2 budget, Science, 247(4949), 1431-1438, 1990.

Thoning, K. W., Tans, P. P. and Komhyr, W. D.: Atmospheric carbon-dioxide at Mauna Loa observatory. 2. Analysis of the NOAA GMCC data, 1974-1985, J. Geophys. Res. Atmos., 94(D6), 8549-8565, 1989.

Washenfelder, R. A., Toon, G. C., Blavier, J. F., Yang, Z., Allen, N. T., Wennberg, P. O., Vay, S. A., Matross, D. M. and Daube, B. C.: Carbon dioxide column abundances at the Wisconsin Tall Tower site, J. Geophys. Res. Atmos., 111, D22305, doi:10.1029/2006JD00715,4 2006.

Wofsy, S. C.: HIAPER Pole-to-Pole Observations (HIPPO): finegrained, global-scale measurements of climatically important atmospheric gases and aerosols, Philos. T. R. Soc. A, 369, 2073–2086, doi:10.1098/rsta.2010.0313, 2011.

Wunch, D., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Stephens, B. B., Fischer, M. L., Uchino, O., Abshire, J. B., Bernath, P., Biraud, S. C., Blavier, J. F. L., Boone, C., Bowman, K. P., Browell, E. V., Campos, T., Connor, B. J., Daube, B. C., Deutscher, N. M., Diao, M., Elkins, J. W., Gerbig, C., Gottlieb, E., Griffith, D. W. T., Hurst, D. F., Jimenez, R., Keppel-Aleks, G., Kort, E. A., Macatangay, R., Machida, T., Matsueda, H., Moore, F., Morino, I., Park, S., Robinson, J., Roehl, C. M., Sawa, Y., Sherlock, V., Sweeney, C., Tanaka, T. and Zondlo, M. A.: Calibration of the Total Carbon Column Observing Network using aircraft profile data', Atmos. Meas. Tech., 3(5), 1351-1362, 2010.

Wunch, D., Wennberg, P. O., Messerschmidt, J., Parazoo, N. C., Toon, G. C., Deutscher, N. M., Keppel-Aleks, G., Roehl, C. M., Randerson, J. T., Warneke, T., and Notholt, J.: The covariation of Northern Hemisphere summertime CO2 with surface temperature in boreal regions, Atmos. Chem. Phys., 13, 9447–9459, doi:10.5194/acp-13-9447-2013, 2013.