Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2018-84-AC3, 2018

Atmospheric Measurement Techniques Discussions

Interactive comment on "Comparison of CO_2 from NOAA Carbon Tracker reanalysis model and satellites over Africa" *by* Anteneh Getachew Mengistu and Gizaw Mengistu Tsidu

Anteneh Getachew Mengistu and Gizaw Mengistu Tsidu

anteneh.getachew7@aau.edu.et

Received and published: 23 August 2018

Authors' response to interactive comment on "Comparison of CO2 from NOAA Carbon Tracker reanalysis model and satellites over Africa" by Dr. Baker

We thank Dr. Baker very much for meticulous and insightful comments. His comments are educational and helpful to improve the manuscript to its current level. We would also like to express our gratitude to Dr. Dietrich G.Feist, the editor, for giving us the chance to get feedback from the two knowledgeable reviewers. In the following, we will

Printer-friendly version

Discussion paper

respond to all the comments and questions of Dr. Baker in details indicating page and line numbers of the manuscript where changes are effected as much as it is possible.

Baker's comments: In this manuscript, CO2 mixing ratio fields over the continent of Africa taken from the CarbonTracker-2016 (CT2016) CO2 flux inversion system are compared to CO2 measurements obtained from the GOSAT and OCO-2 satellites. Though it is not said explicitly, this CO2 comparison is presumably done using column averages (commonly referred to as "XCO2 "), either using a straight pressure weighting or by sampling the CT model using the same vertical averaging kernel and prior CO2 vertical profile that the satellite retrieval uses on the true atmosphere.

Response In the older manuscript, we have not used averaging kernel and a priopri CO2 vertical columns to smoothout the model XCO2 based on our simple inspection of the vertical grids used by the model and the satellite retrievals and the judgment that followed to ignore the difference in vertical resolution. However, during this revision we applied the averaging kernel and a priori profile based on the comments from the anonymous reviewer. This led to substantial improvements in the degree of agreement between the two data sets. We thank both reviewers for bringing it to our attention. As a result, we have now updated the manuscript with the new results. As indicated in our response to the anonymous reviewer's comments, we have also included the procedure used for smoothing as proposed by Rodgers and Connor (2003) and Connor et al. (2008) in the revised manuscript in Section 2.4.

D. Baker's comments: The CT2016 system takes a set of somewhat-realistic surface CO2 fluxes (fossil fuel burning, land use change, wildfires, photosynthesis and respiration from the land biosphere, air-sea fluxes) as a first guess, runs them forward through the TM5 off-line atmospheric tracer transport model to obtain CO2 mixing ratios in the interior of the atmosphere, samples these at the time and place of surface in situ data, and uses the measurement differences to improve the initial flux estimates using a fixed-lag ensemble Kalman smoother as the inversion method. CT2016 thus can be thought of as a model of CO2 in the atmosphere that has been forced to agree with

surface CO2 measurements, or alternatively as a glorified interpolation between the in situ measurements. The difference between CT2016 and the satellite data is mainly viewed here as a deficiency in the CT model as compared to the more accurate satellite data. Explanations for the differences are suggested in terms of likely flux errors in the CT2016 model. Comparisons of this sort are of value, but in my view are better viewed from a different perspective. First, if one is interested in understanding the processes driving CO2 fluxes in Africa (fires, photosynthesis, respiration from plants or soils, etc), one might do better to look at the results of an inversion of the satellite CO2 data compared to the fluxes obtained from a system like CT2016 that inverts in situ CO2 data. The reason for this is that the column averages compared here have much information in the upper part of the column that blows in from outside the bounds of Africa and reflect the effect of fluxes from around the globe; thus, column CO2 differences compared over Africa may not reflect the impact of local fluxes, but also of far-field fluxes that have blown in over Africa. The inversion systems, which model atmospheric transport, are supposed to sort this out and apportion the fluxes to the right locations – they are the tool that I would use if understanding fluxes is the goal.

Response We agree that the tone of the discussion tends to portray accurate satellite observations as compared to CT model. We have made several changes to highlight the work is about comparison of two data sets with their own weaknesses and strengths. Clearly, we are not doing model validation. Our intention is to assess the extent of agreement between satellite retrievals and CT model so that in subsequent work which is currently in progress, we can use CT XCO2 with confidence to address some scientific problems including variability at different temporal scales and their possible drivers. This is due to the fact that CT provides synoptic data and additional parameters such as flux. Understanding fluxes is not our goal in this particular work. Some of our suggestions to explain the differences between XCO2 from CT and Satellite are difficult to substantiate in current context. For example, as suggested by Dr. Baker, some of our explanation can be best supported if inversion of the satellite CO2 data is compared to the fluxes obtained from a system like CT2016. However, that is not our

intention. In situation like this, we changed our wording such that our suggestion indicate only a few of many factors as a possible source of difference. But, there are also instance where CT clearly shows weakness to capture XCO2 distribution over Africa as compared to satellite. For example, because of the climate of Africa, it is generally expected to observe high XCO2 along equator and low XCO2 as we move away from equator towards arid regions over Sahel and Kalahari (Williams et al., 2007). This is generally captured by satellite in contrast to CT implying some level of weakness in CT.

D. Baker's comments: The direct comparisons of column CO2 do have value, but one should not assume that the model is wrong and the satellite measurements right, as has been done here for the most part. The retrieval of CO2 from satellite data is beset with a variety of challenges (scattering due to clouds and aerosols, instrument issues, unknown spectroscopy, surface characterization issues, etc.) that often result in systematic errors that, at the moment, are usually removed in a separate step after the retrieval. A key component of this bias correction process is the comparison to independent data, such as column-averaged CO2 measurements from the TCCON network, for example (using comparisons similar to those in this manuscript, but with the model being replaced by the TCCON data). However, comparison to models that have assimilated in situ data (like CT2016) have also been used in this bias correction step. It has been found that the systematic errors in the satellite retrievals are generally larger than those from an ensemble of models, so that the models may be used to help find errors in the satellite retrievals (instead of vice versa, as is done in this manuscript). Hopefully that situation will change as satellite retrieval schemes improve, but at the moment it is not a good assumption that most of the difference between model and satellite measurements can be attributed to model error. Really, there are errors in both model and measurement and one should quantify both with appropriate uncertainties. The authors do provide satellite CO2 retrieval uncertainties here, but these do not capture the impact of systematic errors in the retrievals and are overly optimistic – one cannot assume that because they are small, the majority of the model-measurement difference must be due to model error.

Response We appreciate the comments from Dr. Baker. As indicated in the previous response, we have now soften the tone of several statements in the manuscript but we have still kept some of the possible factors for the difference between the two XCO2 data sets. For example, as correctly indicated here in the review comments by Dr. Baker himself, satellite XCO2 accuracy is affected by a number of factors including clouds and aerosols, pointing errors, spectroscopy, forward modeling error etc. These have been indicated wherever appropriate in both the old and revised manuscript. On the other hand, CT, as a model, has its own limitation including accuracy of reanalysis data used and TM5 tracer. For example, reanalysis data from ECMWF is not well constrained by in-situ observations over Africa due to gaps in existing in-situ data and insufficient number of observations. There are several studies on the performance of reanalysis data over Africa that supports the above understanding (e.g., Nagarajan and Aiyyer, 2004; Mengistu Tsidu, 2012). The assessment of the impact of these inaccuracies in input data is work on its own but certainly it will severely affect the accuracy of CT XCO2 since it is difficult to merely assume CT to be just a simple glorified interpolation of surface fluxes in particular over Africa with just handful of in-situ observations. Therefore, these are points worth discussing when the two data sets are compared. Discussions that reflects our response here have been added in different parts of the manuscript.

D. Baker's comments: In my view, the model-measurement comparisons done here are most valuable for understanding errors in the satellite retrievals. Africa is a continent with wide extremes in surface type (desert vs. rainforest vs savannah) and aerosol loading. These conditions have a strong impact on the satellite retrievals, so a study of this sort over Africa can tell much about how these systematic errors vary geographically. I would encourage the authors to consider whether their comparisons might be better viewed through that lens.

Response We are afraid that it might not be a good idea since we have indicated in the preceding responses, that there are a number factors that affect both satellite and

CT model. Moreover, the agreement between the two in terms of absolute magnitude (low bias and RMSD) and phase (strong correlation) suggests worth using them for further scientific problems on equal footing or one of them depending on suitability to the problem at hand. For example, where the spatial and temporal resolution as well as synoptic nature of the data are important consideration to address a scientific problem, one might prefer CT over Satellite XCO2. We can not dispute the importance of what Dr. Baker suggested but our mere interest in this work is to assess how CT performs as compared to satellite and from there to use CT to solve other scientific problems since it has better spatial and temporal resolution and agrees reasonably well with the only available data over the region i.e., satellite XCO2.

D. Baker's comments: In general, the English punctuation and usage in this manuscript need extensive editing, which I have not attempted to do here. The authors need to clarify certain aspects of their method, most notably what CO2 quantity they are plotting on their figures (presumably column-averaged CO2) and how they do their vertical averaging. Measurement errors of 2 to 5 ppm are not small or reasonable – they are show-stoppers that prevent the satellite data from being useful for inferring surface CO2 fluxes (since these measurement errors feed through to similar errors in the flux results). Some of the statistical metrics presented here are unconventional and are difficult to understand, and do not add much to the presentation – I suggest below that those sections of the document be removed.

Response In the revised manuscript we have made substantial editing to improve the language. The quantity used in all our plots is XCO2. We understand that there was lack of clarity which is now improved. The statistical metrics used here are common in other discipline when the interest is to compare how the distribution of physical quantity as determined from two observations or observation and model agrees at the extreme ends of the distribution. In several cases including current study, the discrepancy between medians of the distribution determined from two observations usually agree very well as mainly reflected by correlation, bias and RMSE. However, a notable difference

between two measurements of the same quantity exists at the two tails of the distribution. This can be qualitatively assessed using scatter plots. The categorical metrics based on qunatile thresholds can disclose the agreement between model and satellite observations in more quantitative manner. The information from these statistics can be used either by modelers or those working on satellite retrievals. For example, in satellite retrievals, smoothing constraint/regularization (e.g. Tikhonov first or second order-so called shape constraint) heavily penalizes the portion of the profile at either low extreme or high extremes. Moreover as the regularization strength depends on the a priori profile constructed usually from climatology (with strong smoothing of the extremes), the part of the distribution in the tails may not represent true observations. That could create huge discrepancy between the model and satellite observations in the tail region of the XCO2 distribution. Therefore, information from these part based on categorical statistics could provide useful insights to experts in satellite remote sensing for further improvements of retrieval strategy. This is just a single example but one can think of similar benefits for modelers to improve accuracy of XCO2. Therefore, we retain section on categorical statistics but enhanced the methodology and discussion sections to improve clarity. Content that reflects parts of this response is included in the manuscript on page 13, lines 2-12 of revised manuscript.

D. Baker's comments: page 1, Line 1: "poor global coverage and resolution of satellite observations": Usually the satellites are thought to have good coverage over the globe, at least when compared to the in-situ data.

Response Indeed, satellites have good global coverage as compared to in suit observations. However, the global coverage from the model simulation is far better than satellite observations. We put this statement to highlight that the global coverage from models are better than that of satellite observations. Since it might create similar wrong impression by the general reader, we have rephrased it such that it now reads as "poor spatial and temporal resolution of satellite observations" in the revised manuscript on page 1 line 1.

D. Baker's comments: Abstract: Here, the fidelity of the CarbonTracker model is being tested by comparing its a posteriori 3-D CO2 fields to CO2 measured from satellites and by the TCCON network of ground-based sun-viewing spectrometers. That would make sense if the accuracy of those measurements was thought to be better than that of the model. But what if the reverse was true? Then it would make more sense to check the accuracy of the measurements by comparing them to the (more accurate) model. In the early days of satellite CO2 data (and we are still in the early days, really) it was thought that latter situation was actually the case, and models were used to correct the satellite data. Given that, some discussion of the assumptions underlying the comparison of CarbonTracker to the satellite and TCCON measurements would be useful in this paper before moving on to the comparison.

Response We appreciate your suggestions and guidance how we should perform the comparison in the best way. In fact, in our introduction, we put some discussions that the Carbon Tracker is more accurate than satellite observations. For example, see this statement "Kulawik et al. (2016) found root mean square deviation of 1.7, and 0.9 ppm in GOSAT and CT2013b XCO2 relative to TCCON respectively" on page 3 line 18 of old and lines 22-23 of revised manuscript) over TCOON sites. However, this could not guarantee that Carbon Tracker performs best at continental and large scales in particular over Africa with extremely limited number of in-situ flux observations in contrast to region covered by TCOON sites. Dr. Baker's argument may work for region with relatively dense network of flux measurements. However, we are dealing with Africa land mass with hardly a handful of flux measuring sites. So, we are in a difficult position to accept the suggestion in the case of this study which focuses on Africa landmass. It is indeed difficult to convincingly take the suggestion by Dr. Baker which amounts to assume flux measurement in the surrounding regions (e.g. Europe and others) can be extrapolated to produce more accurate XCO2 than satellites. Therefore, we still assume that the satellite is better than CT on certain aspect such as producing expected decrease in XCO2 away from equator (see our previous response). A long as the current data gaps in flux measurements over Africa is concerned, the satellite

Printer-friendly version

Discussion paper

might have some advantages over CT in reproducing true XCO2. Therefore, we prefer to treat the analysis in this manuscript focusing only on how the satellite and CT XCO2 compare and suggesting all possible scenarios for the difference. Our work is a follow up of many other studies (e.g., Chevallier et al.,2010; Feng et al., 2011; Yingying Jing et al.,2018)

D. Baker's comments: p1 L15: "relative accuracies of 1.22 and 1.95 ppm were found between the model and the two data sets"; these were judged to be "reasonably good". The in situ CO2 data have relative accuracies an order of magnitude better than this (0.1 to 0.2 ppm) – why then should measurements an order of magnitude less accurate be considered "good"? Some guidelines concerning what is considered a good error versus what is considered a bad error ought to be given to support this assertion. The density and coverage of a data type might also factor into this assessment, as well as the importance of random versus systematic errors.

Response This comment is based on our older figures. There is significant improvement after CT is smoothed using retrieval averaging kernel and apriori profiles. For example in the revised manuscript, the relative accuracies is now 1.01 instead of 1.22 in the older version of the manuscript for the comparison between CT2016 and GOSAT. Similarly, the relative accuracies changes from 1.95 (older manuscript version) to 1.18 ppm (revised manuscript) for comparison between CT16NRT17 and OCO-2). These are reasonably good in view of results from previous studies. For example, Deng et al. (2016) founds a regional accuracy of 0.62 ppm between ACOS and TCCON and 0.93 ppm between NIES and TCCON for comparison over 11 TCCON sites. These figures are comparable and reasonable given that our comparison is over 426 pixels and region with diverse climate in contrast to limited number of sites (e.g., 11 TCCON sites). Moreover, TCCON is more accurate as compared to satellites whereas in this study we are dealing with less accurate satellite and model, both of which contribute to the relative accuracy. Apart from these factors, other authors have indicated vertical transport is more variable among transport models (Gurney et al., 2002) and probably

more error-prone. Hungershoefer et al. (2010) have shown that an error of up to 3 ppm may be observed over site with complex circulation and fluxes based on numerical experiment (simulation). The above information is highlighted in the manuscript to strengthen our conclusion and help the readers understand how good is good in the context of this study.

D. Baker's comments: p1 L17-18: "...probability of detection ranges from 0.6 to 1 and critical success index ranges from 0.4 to 1..." These statistics may not be familiar to the general reader. Maybe say "(see main text for definitions)" when using these in the abstract?

Response We accepted the comment and made changes in the abstract accordingly. Moreover, we have made a lot of changes that includes basic definitions, equations and the physical acceptable limits of the individual statistical parameter in Section 2.4. These changes are crucial to understand and interpret the results.

D. Baker's comments: p1 L20-21: "GOSAT and OCO-2 XCO2 are lower than that of CT2016 by upto 4 ppm over North Africa (10 N −35 N) whereas it exceeds CT2016 X CO2 by 3 ppm over Equatorial Africa (10 S−10 N)." It seems like the sign of this is wrong. Also, these satellite data are raw or bias-corrected?

Response Dr. Baker is right that these figures are too high. We have now realized that the discrepancy arises mainly from ignoring the difference in the vertical resolution of the CT and satellite retrievals. Following one of previous comments, we have already recalculated XCO2 from CT using averaging kernel weighting function and a priopri profiles of the retrievals. The comparison with smoothed XCO2 has resulted in much lower difference. Moreover, GOSAT XCO2 is lower than that of CT. Therefore, the statement is amended such that it now reads as "Spatially, OCO-2 XCO2 are lower than that of CT16NRT17 by 3 ppm over some regions in North Africa (e.g., Egypt, Libya Mali) whereas it exceeds CT16NRT17 by 2 ppm over Equatorial Africa (100S − 100N)." This change is made on page 1 lines 23-25 of the revised manuscript. We

would like to confirm that data from both satellite are biased corrected. This information is also included in the revised manuscript on page 3 line 11 and 20.

D. Baker's comments: p1 L25: "In these cases, the model overestimates the local emissions and underestimate CO2 loss." Here, it is assumed that the satellite data are correct and the model is wrong. But what if the reverse was actually true? In reality, both model and satellite are wrong, to differing degrees that are quantified using the metric of uncertainty. You should give reasons why you think the uncertainty on the satellite measurements is lower than the uncertainty on the modeled CO2, to defend why you think the model is wrong and the satellites right. The satellites have retrieval biases and random errors that could easily be larger than the modeling errors in CO2 (errors due to transport, errors in the in situ data, inversion assumptions, etc).

Response The comment is well taken and the statement is removed from the revised manuscript.

D. Baker's comments: p2 L25-28: In introducing the TCCON data, you should note that the TCCON network measures a column-integrated CO2 mixing ratio; yes, it is "ground-based", since the sensor sits on the ground, but the measurement is a column average (unlike the ground-based in situ data) and that is why it is useful for validating the satellite data, which are also column averages.

Response We accept your comment and rephrased the statement on page 2 line 31 such that it now reads as "Total Carbon Column Observing Network (TCCON) is a notable one since it provides accurate and high–frequency measurements of column-integrated CO2 mixing ratio"

D. Baker's comments: p3 L3: "Other studies have revealed that significant improvement in estimation of weekly and monthly CO2 fluxes can be achieved subject to CO2 retrieval error of less than 4 ppm" This may be true if the errors are purely random – with enough low precision data of this sort, the errors may average down to something lower and more useful. However, if this is accompanied by systematic errors of similar

magnitude, these will not average out. Therefore some discussion of random versus systematic error must be added when giving out these error quantities.

Response This is a mere reference to other works.

D. Baker's comments: p3 L21: "V02.XX" – you need to complete the version numbers for two products on this line

Response We have indicated the specific version of GEOS-Chem used from the cited paper on page 3 line 31 of the revised manuscript. However, the NEIS L2 V02.xx remained the same. According to Yoshida et al. (2013), "xx" in NIES L2 V02.xx and NIES L2 V01.xx initially planned for the specific versions of NIES L1 used in the retrievals to generate L2 data. However, the L1 data version remains the same in all NIES L2 and therefore it remains xx in all versions to the best of our knowledge.

D. Baker's comments: p4 L3-4: "These findings suggest that it is important to assess the accuracy and uncertainty of XCO2 from models with respect to observations" Again, the authors point to the differences between satellite data and models, then leap to the assumption that the models are responsible for most of those differences when the measurements might actually be more responsible. Some discussion of which is more error-prone (model or measurement) is needed.

Response We have rephrased it such that it now reads as "These findings suggest that it is important to assess how satellite and model XCO2 compare with each other over other regions". This change is made on page 4 lines 4-5 of the revised manuscript.

D. Baker's comments: p4 L27: "CT2016 and CT2017 respectively": you should not refer to the 2016 near real-time CT as "CT2017", as that term is reserved for the subsequent release of the full data span (the release that would have been put out in 2017 if the releases were put out when indicated; the release that finishes at the end of 2016). The near-real-time releases are different from the standard releases in that they bring out a subset of the full data set, without the usual quality assurances, for early

use by modelers. Even if that is only a label for use in this manuscript, it will confuse people.

Response Thank for your suggestion. CT2017 was replaced by CT16NRT17 in the revised manuscript.

D. Baker's comments: p5 L1: Which version of ACOS GOSAT retrievals did you use in this study?

Response Thank you for pointing out the missing information which was also indicated by the anonymous reviewer. The version used in this study is ACOS B3.5 Lite. This information is incorporated in Section 2.2 of the revised manuscript.

D. Baker's comments: p5 L2-3: "GOSAT is the world's first spacecraft to measure the concentrations of carbon dioxide and methane, the two major greenhouse gases, from space." Not true: many satellites/instruments measured both in the thermal IR. Also SCIAMACHY, which came out much earlier, measured both in the near-IR. What you could say is that GOSAT was the first spacecraft dedicated solely to measuring CO2 and CH4.

Response The comment is well taken and included on page 5 line 1.

D. Baker's comments: p5 L15: When discussing the OCO-2 data, you must say what version of retrieval you are using: version 7? Also, you should indicate whether you are looking at raw XCO2 values or bias-corrected values. If bias corrected, you should indicate whether you used the additional "s31" bias correction, designed to correct albedo-related errors not caught in the initial bias correction. This extra bias correction could be particularly important over Africa, with its large differences between desert and tropical rainforest. Also, you should discuss whether you sample your model with the OCO-2 averaging kernel and prior CO2 profile when comparing to the OCO-2 retrievals.

Response This was also a comment from the anonymous reviewer. So, it has already

been corrected. The version of OCO-2 (i.e., OCO-2 V7 lite level 2) used is biased-corrected. However, we have not used albedo bias correction. We have smoothed the CT XCO2 using averaging kernel and a priori profiles with respect both satellite data sets following your previous comments and that of anonymous reviewer's comments (see also one of the previous responses).

D. Baker's comments: p5 L21: (Methods): the satellites estimate both a profile of CO2 on 20 levels, as well as a vertical weighted average, "XCO2", computed from these. Which do you compare to? If the latter, you should discuss the vertical weighting kernel for XCO2 and how you sample your model to account for this. This methods section seems to be the correct place to discuss this.

Response We used the column average XCO2 from the satellites and compared it to the XCO2 computed from the model CO2 profile. This CO2 profile is extracted and interpolated to resolution and vertical level of the satellites. Then following the procedures described in Coner et al., (2008) we transform the model CO2 profile to model averaged values (XCO2). Brief discussion and mathematical expression of the technique are included in section 2.4 of the revised manuscript.

D. Baker's comments: p6 L19: "Fig. 1 shows the five-years average of CT2016 (Fig. 1a) and GOSAT (Fig. 1b) XCO2 distribution." It would have been more useful to do the 5-year average over a true 5-year span, rather than the "April 2009 to June 2014" span that you used. As it stands, you have an extra 3 month span (April to June) that is unbalanced by the remainder of an annual cycle. As a result, this three month span throws off what otherwise would be the average of five full annual cycles. If flux is positive north of the equator during April to June (which it probably is, this being right at the end of the burn season there) then you tend to see positive values there in Figure 1a and 1b. This is due to the seasonality of the flux rather than the annual mean value. It might be easier to discuss annual mean fluxes in the text if these plots reflected averages over pure annual cycles.

Response We agree that the extra 3 months may destroy the symmetry of the annual cycles over Africa. In the revised manuscript we change the study period to span only 5 years (from May 2009 to April 2014).

D. Baker's comments: p7 L3: "respiration from forest in the region is overestimated" Why do you think that this is not due to photosynthesis being underestimated?

Response It is possible that both or one of the two can cause the observed discrepancy. It might be equally possible cloud contamination may affect the satellite XCO2 such that the satellite estimates are unusually high (O'Dell et al., 2012). The equatorial region, where the difference is observed, is known to high altitude thin cirrus clouds. The revised manuscript was edited to reflect these possibilities on page 8 lines 18-23.

D. Baker's comments: p7 last line: "standard deviation of 1.31 ppm indicating better consistency and less potential outliers." Better than what?

Response In the revised manuscript this standard deviation was only 0.98 and the statement was reworded as "standard deviation of 0.98 ppm indicating good regional consistency and low potential outliers" on page 8 lines 30-31.

D. Baker's comments: p8 L3: change "require" to "permit" Also, if the difference can be up to 1.5 degrees on either side, then we are talking about a box of 3 degrees on a side in both latitude and longitude.

Response Accepted. Change is made on page 8 line 33.

D. Baker's comments: On p5 L29 you should then change the box from being 1.5 degrees long and wide to 3 degrees long and wide.

Response Accepted. Change is made on page 8 line 33.

D. Baker's comments: Fig 3 caption: please provide units (ppm)

Response The unit is now indicated as the title of the color bars in the revised manuscript.

D. Baker's comments: p8 L19: A correlation should be unitless – remove "ppm"

Response Removed

D. Baker's comments: Figure 3: It is not clear to me why Figures 3a and 1c are different – they are both showing the difference (or bias) between CT2016 and GOSAT on an annual mean basis. So why aren't they identical?

Response We agree that the mean difference and the bias should be the same. But Fig.1c is based on the full coincident data sets while in Fig.3a, grid points with less than 10 data points are removed to establish reliable statistics. Although such restriction is not a requirement for simple difference which is temporally averaged in Fig.1c, we have now excluded those points in Fig.1c to avoid similar misunderstanding in the revised manuscript.

D. Baker's comments: p8 L20: Good – here you are examining whether some of the difference might be due to the satellite data rather than the model.

Response Appreciated the complement and changes in other parts of the manuscript have similar tone in the revised manuscript.

D. Baker's comments: Section 3.2: The statistics presented here are abstruse and require the reader to go back to the other references not only for what the statistics mean, but even what the acronyms stand for. It is not clear what the message is that these statistics convey. The only point I gleaned from the discussion was that the annual mean bias between model and data is higher where there are fewer data to average over – that is a point that could be made with Figure 1 alone. I would suggest deleting this section altogether.

Response As the statistical tools used in this section are uncommon in our science community, it is not at all surprising to see more inquiry on the tools. We have clearly indicated why it is important by citing specific application on page 6 of this author response and in the manuscript in Section 2.4 as well as in the discussion section from

page 13 lines 2-12. Moreover, amendment in different parts of the manuscript will now provide more information that clarify the importance of the statistical tools (see also the author response to anonymous reviewer).

D. Baker's comments: p13 L9, p14 L11-14: Here you are suggesting that local fluxes are responsible for the changes in CO2 seen over Africa, but it is quite possible that a large part of these changes are due to fluxes elsewhere, over different continents, blowing downwind over Africa. Without inversion results linking concentrations to the fluxes that caused them, your argument will remain weak using this approach.

Response We have taken note of your suggestion and included it as one of possible scenarios in the revised manuscript on page 18 lines 1-3.

D. Baker's comments: p25 L16: "Satellite observations are able to capture the CO2 concentration truly and objectively." I don't understand how you can say this based on the results that you have shown. You have shown large differences between CarbonTracker and the satellite data, on the order of 5 ppm on a seasonal and regional basis. This could be due to satellite retrieval problems in addition to the model problems that have been emphasized here. You haven't shown any data or comparisons that suggest the satellite data are better than the model, really. Maybe you could compare GOSAT and OCO-2 across the short time period that they overlap (late 2014, 2015) and if they agree closely then suggest they are both getting the right answer. From the plots you have showed in this manuscript, however, it looks like there are significant differences between OCO-2 and GOSAT... suggesting that the satellite errors are significant.

Response We have accepted that we are a bit biased in favor of satellites. We removed the statement. However, since we do not have overlap between the two satellites (GOSAT and OCO-2) in the version used in this study, we did not compare them.

References 1. Chevallier, F.; Feng, L.; Bösch, H.; Palmer, P.I.; Rayner, P.J. On the impact of transport model errors for the estimation of CO 2 surface fluxes from GOSAT observations. Geophys. Res. Lett. 2010, 37. 2. Connor, B. J., Boesch, H., Toon,

G., Sen, B., Miller, C., and Crisp, D.: Orbiting Carbon Observatory: Inverse method and prospective error analysis, Journal of Geophysical Research: Atmospheres, 113, 2008. 3. Deng, A., Yu, T., Cheng, T., Gu, X., Zheng, F., and Guo, H.: Intercomparison of Carbon Dioxide Products Retrieved from GOSAT Short-Wavelength Infrared Spectra for Three Years (2010–2012), Atmosphere, 7, 109, 2016. 4. Feng, L.; Palmer, P.I.; Yang, Y.; Yantosca, R.M.; Kawa, S.R.; Paris, J.D.; Matsueda, H.; Machida, T. Evaluating a 3-D transport model of atmospheric CO 2 using ground-based, aircraft, and space-borne data. Atmos. Chem. Phys. 2011, 11, 2789–2803. 5. Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fan, S., et al.: Towards robust regional estimates of CO2 sources and sinks using atmospheric transport models, Nature, 415, 626–630, 2002.

6. Hungershoefer, K., Breon, F.-M., Peylin, P., Chevallier, F., Rayner, P., Klonecki, A., Houweling, S., and Marshall, J.: Evaluation of various observing systems for the global monitoring of CO 2 surface fluxes, Atmospheric chemistry and physics, 10, 10 503–10 520, 2010. 7. Jing, Y., Wang, T., Zhang, P., Chen, L., Xu, N., and Ma, Y.: Global Atmospheric CO2 Concentrations Simulated by GEOS-Chem: Comparison with GOSAT, Carbon Tracker and Ground-Based Measurements, Atmosphere, 9, 175, 2018. 8. O'Dell, C., Connor, B., Bösch, H., O'Brien, D., Frankenberg, C., Castano, R., Christi, M., Eldering, A., Fisher, B., Gunson, M., et al.: The ACOS CO 2 retrieval algorithm-Part 1: Description and validation against synthetic observations, 2012.

9. Mengistu Tsidu, G.: High-resolution monthly rainfall database for Ethiopia: Homogenization, reconstruction, and gridding, Journal of Climate,25 , 8422–8443, 2012. 10. Nagarajan, B. and Aiyyer, A. R.: Performance of the ECMWF operational analyses over the tropical Indian Ocean, Monthly weather review, 132, 2275–2282, 2004. 11. Rodgers, C. D. and Connor, B. J.: Intercomparison of remote sounding instruments, Journal of Geophysical Research: Atmospheres, 108, 2003.

12. Yoshida, Y., Kikuchi, N., Morino, I., Uchino, O., Oshchepkov, S., Bril, A., Saeki, T., Schutgens, N., Toon, G., Wunch, D., et al.: Improvement of the retrieval algorithm for

GOSAT SWIR XCO2 and XCH4 and their validation using TCCON data, 2013.

13. Williams, C. A., Hanan, N. P., Neff, J. C., Scholes, R. J., Berry, J. A., Denning, A. S., and Baker, D. F.: Africa and the global carbon cycle, Carbon balance and management, 2, 3, 2007.

Please also note the supplement to this comment:
https://www.atmos-meas-tech-discuss.net/amt-2018-84/amt-2018-84-AC3-supplement.pdf

---

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2018-84, 2018.