Reply to Reviewer #1

This is potentially an interesting paper that provides an update about the status of OCO-2 data. The authors use an inversion ensemble to compare against a range of independent data. Using an ensemble of models is a strength of this study but since the models do not use common prior information this reviewer was unable to understand how OCO-2 improved the performance of individual models. The manuscript would benefit greatly from a better exposition of the performance of individual models from the standpoint of error reductions. This would make it easier for other readers to fully appreciate the results of this study. Major and minor comments are listed below.

We are grateful for the comments of Reviewer #1 and for taking time to review our manuscript. We answered below the comments with information on page and line numbers that have been changed in the manuscript when necessary.

1- MAJOR. The MIP design claims to mimic past model intercomparison projects. Aside from using a common prior for the fossil fuel emissions, individual modeling groups were free to choose all other model features, e.g. biospheric priors, uncertainties for data, errors associated with model transport, etc (Table 1). These differences will impact posterior flux estimates. The authors argue that variations among inversion systems are considered beneficial for the purpose of characterizing flux uncertainty, but this reviewer would argue that some portion of this variation in the ensemble is unnecessary and a reflection of the design of the MIP. Obviously, it is too late to redo this ensemble experiment but given the expertise and complexity of the inversion systems it would be incredibly useful to report individual model prior and posterior uncertainties. Did some models use stricter constraints? Do some models follow their prior more than others? Did some models overfit data? How well did the individual model fit the net fluxes before the fossil fuel component was removed? Were the net CO2 fluxes consistent with NOAA atmospheric CO2 growth rate estimates? It would be useful to understand the basics of individual model performance before embarking on more elaborate data analysis. Otherwise, it is difficult to understand what knowledge has been gained from this experiment.

Previous model intercomparison experiments – for CO_2 and other long-lived species – have had limited control on individual model setups. The Global Carbon Atlas (http://www.globalcarbonatlas.org/en/content/atmospheric-inversions) infers the state of the carbon cycle over the past decades from a diversity of CO_2 inverse models with no common elements, post-hoc corrected for the assumed fossil fuel emissions (Peylin et al., 2013). The Global Carbon Project (GCP) routinely uses model ensembles, both top-down and bottom-up, to construct CO_2 and CH_4 budgets (Friedlingstein et al., 2020; Saunois et al., 2020). The model ensembles used by the GCP for this purpose also have very limited or no commonality imposed on them. This is by design, since inter-model differences are likely to be a better estimator of the true uncertainty in our knowledge compared to the uncertainty estimate from any single model. It is true that in such model ensembles, not all models "perform" well, i.e., the ensemble is not well calibrated. Despite this limitation, as top-down models have continued to develop, they have started converging on certain key aspects of the global carbon cycle (Gaubert et al., 2019).

Compared to the top-down model ensembles mentioned above, the OCO2 MIP ensemble is more tightly controlled. They all use the same fossil fuel emissions, the same set of OCO2 retrievals with identical errors for assimilation, and the same set of in situ CO_2 observations with recommended errors used by most models. This level of control is, to the best of our knowledge, unprecedented among published top-down CO_2 model intercomparisons. The only aspects of the MIP not controlled are (i) the specification of the prior flux and its covariance, (ii) the assimilation scheme, and (iii) the atmospheric transport model. We argue that (ii) and (iii) provide a necessary diversity in model setup to gauge the true uncertainty in our results and emphasize that they are strengths rather than weaknesses of the MIP protocol. Indeed, inter-model differences due to (ii) and (iii) are often larger than due to the assimilated data (Chevallier et al., 2014), and it is therefore necessary to have some diversity in those aspects to represent the true posterior uncertainty. As for (i), specification of the prior flux and its uncertainty are tied very closely to the model setup and is therefore hard to standardize. The prior flux uncertainty, in particular, is incredibly difficult to standardize in reduced rank techniques such as 4DVAR and EnKF even among frameworks that use the same transport model (Babenhauserheide et al., 2015), and to the best of our knowledge no one has done that successfully. It is in theory possible to standardize the prior biospheric and oceanic fluxes, which is a direction future GCP CO_2 and CH_4 inversion efforts are taking. However, we would argue that the spread among bottom-up flux models (e.g., the TRENDY and OCMIP ensembles) is greater than the prior uncertainty one would construct based on any one model. Therefore, having a diversity of prior fluxes is also a strength of our protocol.

We consequently decided to add a sentence in our manuscript page 4 line 80: "MIP models are more strictly controlled or have more common elements than previous CO_2 inverse model intercomparisons such as Peylin et al (2013)".

The reviewer raises a very good point that given the diversity of model setup, not all models are equally good, and it is important to evaluate the performance of individual models before drawing conclusions about the carbon cycle. As part of the MIP, we did collect extensive validation statistics against non-assimilated in situ and column data. These metrics are available at https://gml.noaa.gov/ccgg/OCO2_v9mip/, both as global aggregate statistics and at a site level. However, we stopped short of using those statistics to weight models for two reasons. First, there is to date no consensus on how to weight the MIP models. We have tried several approaches from Bayesian model averaging (Massoud et al., 2020) to analysis of variance (Cressie et al., 2021), but have yet to find a satisfactory scheme. This remains an area of active research. Second, we suspect and hope that the MIP fluxes will be used by a variety of studies focusing on different regions on the globe. While it may be possible to derive global statistics for the models' performance, those statistics may not correlate with how well the models do over individual regions. We have therefore left it to the discretion of future users to choose MIP models wisely for their specific studies, aided by the metrics published at https://gml.noaa.gov/ccgg/OCO2_v9mip/.

Lastly, we would like to address a few specific questions raised by the referee which we think have already been answered in the manuscript:

- "Did some models use stricter constraints?" On the data, all models used the same constraints. On the prior fluxes, the models had different constraints.
- "Did some models overfit data?" Fit statistics are available at https://gml.noaa.gov/ccgg/OCO2_v9mip/co2tser.php. To the best of our knowledge, none of the models overfit the data.
- "How well did the individual model fit the net fluxes before the fossil fuel component was removed?" We do not understand this question, since none of the models try to *fit* the net fluxes.
- "Were the net CO_2 fluxes consistent with NOAA atmospheric CO_2 growth rate estimates?" Over the four years considered in the study, all the models fit the NOAA atmospheric CO_2 growth rate.

2- MAJOR. Differences in median prior fluxes between Crowell et al and this study are not explained. It is unfortunate that this group did not use consistent fluxes for their ensemble study or use the same fluxes used by Crowell et al. Figure 5 shows a wide range of values being used. Again, as described above, it is impossible to see which models track their priors more than others. This is particularly relevant for the analysis of the IS data in the tropics where coverage is sparse.

In our paper, we are not evaluating the different models, but the retrieval version of OCO-2 used in a model ensemble. As previously mentioned, information of each prior used and how they compare to the LNLG and IS posterior of each model are available in the OCO-2 MIP website. While we understand the desire for consistent priors across v7 and v9, it is worth noting that the MIP activity is largely supported by leveraging individual and separately funded research projects and thus is subject to necessarily changing models describing both the biosphere and atmospheric transport.

If we look at the figure on the website and our following figures for each individual model between their priors, IS and LNLG posterior over the tropics, we can see that neither IS nor LNLGv9 for each models follows their priors over the long period of 2015-2018. This can be observed for both the seasonality and the magnitude of the fluxes.

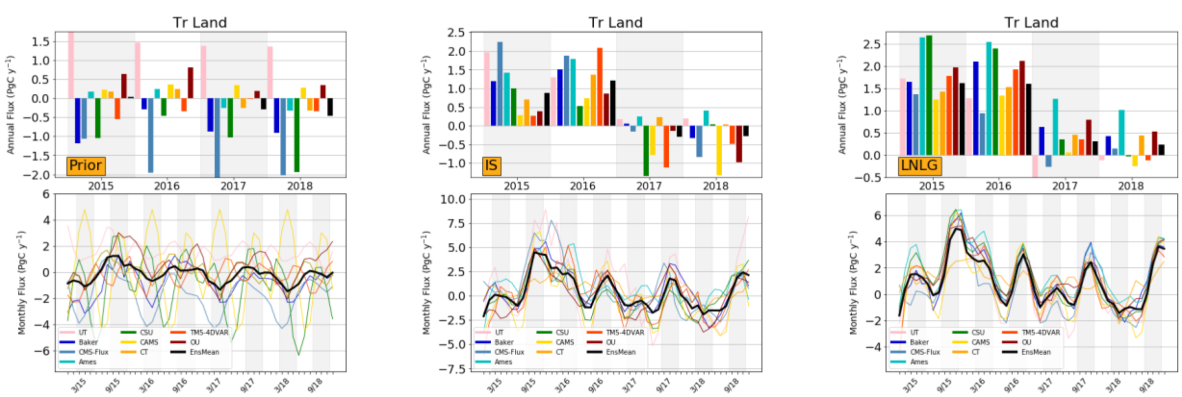

Fig 1. Monthly mean fluxes in PgC/month (bottom side) and annual mean flux in PgC/yr (top side) of Tropics Land from 2015 through 2018 among the different models for Prior (left side), IS (middle) and LNLGv9 (right side).

We mentioned this in our manuscript page 18 line 345: "We can also observe that the IS ensemble spread does not deviate from the prior spread. Neither IS nor LNLGv9, for each model, follows their priors in the tropics during the period of study (not shown here)."

3- MAJOR. Section 3.3 is largely superficial from the perspective of understanding reported flux estimates. Some versions of the fluxes agree with others versions... This is only interesting, if you tell the readers why you think this is important or relevant. This reviewer is surprised by the result over North America. Given this is a continental with a wealth of independent data this result is worrisome. Are there any model outliers that would explain some of the variations that are being reported? This needs more thought. Similarly, over Europe they are implying that their data have a large carbon sink but do not explicitly say it. Follow-up studies suggested this might not be correct (not cited) so Peiro et al could be a useful addition to the broader debate. The authors go on to suggest the enhanced summertime uptake might be due to a dipole between Europe and northern Africa and cite Houweling et al (2015). That's lazy. What did Peiro et al find? They have substantial computational machinery at their disposal. Any consensus among their models about this dipole? The authors have a great opportunity with their ensemble to do some insightful analysis.

For North America, there is a larger spread among the models with IS than with LNLGv9. Some models observed a weaker sink with IS than with LNLGv9 (information from the ensemble spread and median values in Fig. 7 of the manuscript). Schuh et al., 2019 and Schuh et al., 2021 demonstrated large differences in the flux inversions primarily over the Northern Midlatitudes mainly due to transport. This has not been investigated in our study, but could be the first reason for differences observed for Northern America (Fig. 7.a) and North Asia (Fig. 7.e).

It was mentioned in the manuscript (lines 339) that larger sink over Europe is observed with v9 than with v7. We can observe with the monthly fluxes that this sink is strong during the summer months. We were not interested in evaluating or comparing the different models but were interested in looking only to the ensemble mean.

But if we examine the individual models; we can also see the dipole between Europe and Northern Africa among each model (see Fig. 2) with sinks over Europe balanced with sources over Northern Africa.

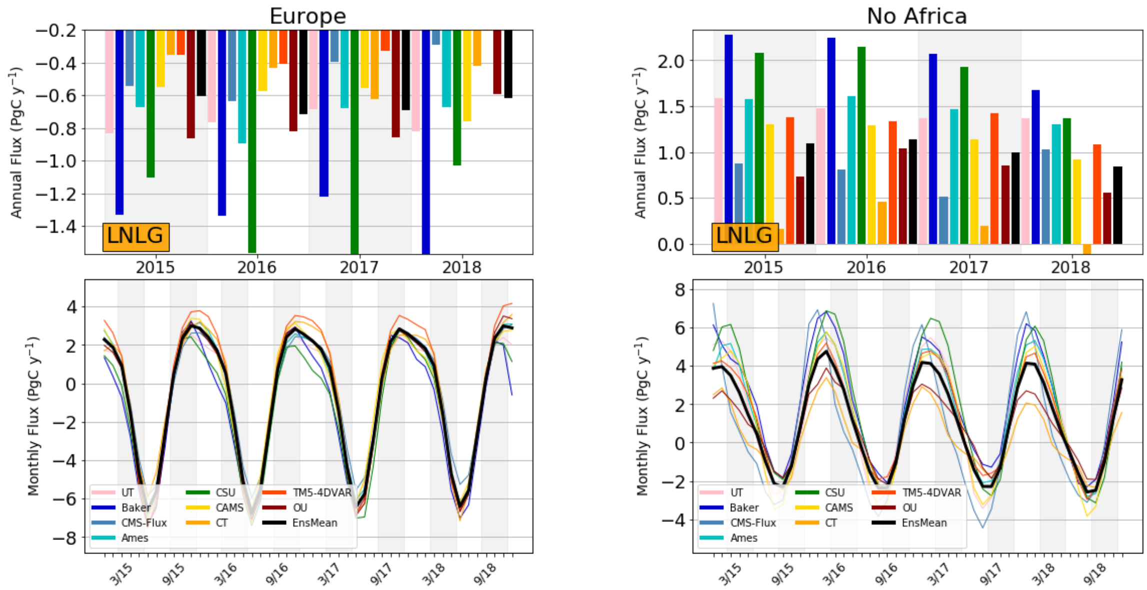

Fig 2. Monthly mean fluxes in PgC/month (bottom side) and annual mean flux in PgC/yr (top side) from 2015 through 2018 among the different models for LNLGv9. Fluxes over Europe are on the left side of the figure and Northern Africa is on the right side of the figure.

4- MAJOR. Section 3.3 continues with a discussion about northern Asia and weaker/stronger sinks that could be due to different amounts of data being assimilated. Surely, the authors could find this out with their analysis. This reviewer would like to see more definitive statements based on their analysis...We find XXX based on our ensemble analysis. These statements might not be generally true but it would add something to the literature.

We are not sure what the reviewer would like us to add. The reviewer should be able to find the estimated median fluxes from the ensemble analysis in our main text (such as line 346). We have already mentioned this information for Northern Asia for instance:

"For Northern Asia (including Eurasia Temperate and Eurasia Boreal), while IS gives large sinks (with an ensemble spread between -2.5 PgC/yr and -0.5 PgC/yr for the whole period), v9 and v7 both show weaker sinks (with a ensemble spread between -1.25 PgC/yr and -0.25 PgC/yr for 2015 and 2016) and a decrease with v9 for 2017 (-0.5 +- 0.5 PgC/yr) and 2018 (-0.25 +- 0.5 PgC/yr).".

5- MAJOR. Line 361: ...we see a trend towards a weaker sink from 2015 to 2018 for northern Asia….Why? The authors should provide some explanation. If they cannot find one then they should admit it.

The trend over Northern Asia has not been investigated and no available reference can provide an explanation for this trend observed in the sink. We have therefore added this sentence page 20 line 395: "Further investigation is needed to explain the decrease in the carbon sink in this region with both IS and v9".

6- MAJOR. Line 368: ...IS seems to follow the pattern of the prior… All priors, some of them? How about the error reduction? Did the authors learn anything from the IS data?

The annual median fluxes for IS appeared to follow the pattern of the prior in 2017 and 2018 over Northern and Southern Tropics. In particular, the median values with the IS ensemble mean give similar values than for the prior, however the ensemble spread is slightly different. When looking at the model individually, the IS does not completely follow the prior of each model (See figure below).

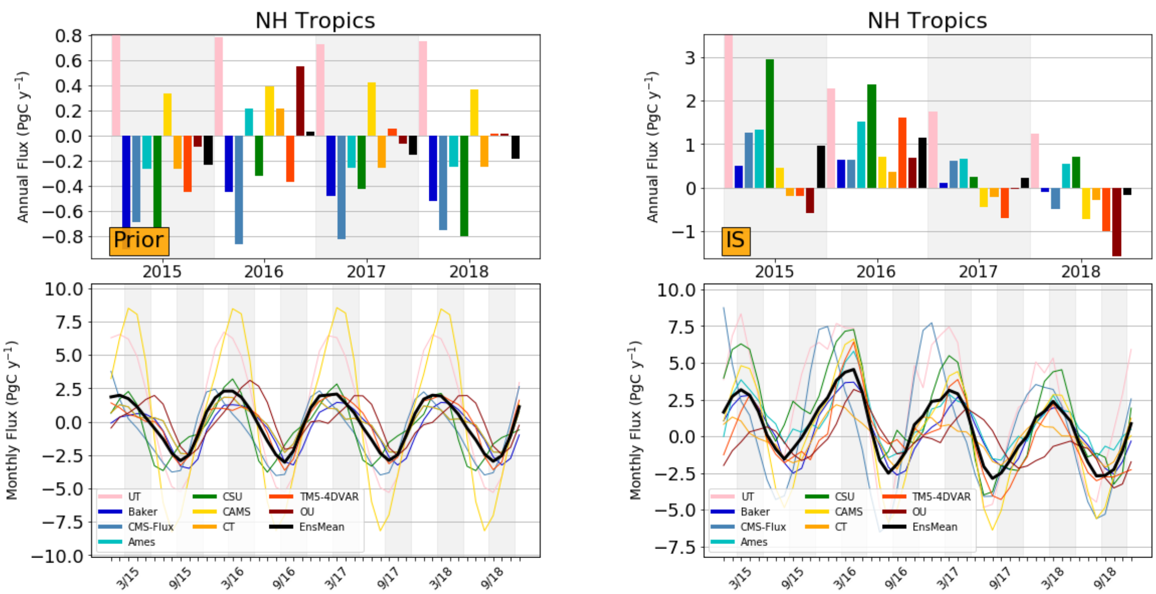

Fig 3. Monthly mean fluxes in PgC/month (bottom side) and annual mean flux in PgC/yr (top side) of Northern Tropics from 2015 through 2018 among the different models for Prior (left side) and IS (right side).

However, the goal of the paper was not to give an analysis for each individual model but to give an analysis of the overall average and to compare the v7 MIP and v9 MIP versions.

Additionally, the website provided by the OCO-2 MIP v9 projects shows interactive figures of each model or experience by regions or globally.

If readers would like further information on individual models, the figures are available on this website:

https://gml.noaa.gov/ccgg/OCO2_v9mip/index.php

We have added the link of this website page 35 line 600: "Further information including figures on individual models can be accessed at the following link https://gml.noaa.gov/ccgg/OCO2_v9mip/index.php"

7- MAJOR. Line 375 Fluxes during the recovery period differ between data sources. Why is this important? Any dipoles in neighbouring regions? What about reductions in uncertainties? What have the authors learnt?

Fluxes during the recovery period (2017 and 2018) over the tropics (Northern and Southern Tropics) differ among the data (v9 and IS). This difference seems to come from Northern and Southern Tropical Africa which balance each other, but also from Tropical South America. As we have already mentioned when focusing on these regions: "But this could be also because most of the *in situ* data are located inside the Amazon and not in the Cerrado savanna of Brazil, resulting in IS inversions being dominated by tropical forest seasonality. Alternatively, the OCO-2 inversions could be dominated by savanna seasonality (Baker et al., 2021a in prep)."

Reduction in uncertainties cannot be observed for this period between v7 and v9 as we do not have data with v7 in 2017 and 2018. However, as previously mentioned in the manuscript and for the El Nino period (2015-2016), the reduction in uncertainties are particularly present over smaller regions both in the Tropics or Extra-Tropics. On average, LNLGv9 ensemble spread is smaller than LNv7 and LGv7. Comparing IS and v9, the ensemble spread is, on average, smaller with v9 than with IS, except for Tropical South America where the ensemble spread seems to be significant in this region for both v9 and IS.

8- MAJOR. The discussion is interesting, although this reviewer notes that the authors have made a claim that a previous study (Palmer et al, 2019) using v7 data would probably have been similar using v9 data. And then the authors proceed to compare their results for v9 and those from Palmer et al 2019. This reviewer is unconvinced this is a valid scientific approach. Surely, a cleaner comparison would be to compare their own ensemble values between the two data versions? The discussion about Gloor et al, 2018 and Liu et al, 2017 is a bit odd. What is the authors' point? This reviewer was also concerned that over the tropics the authors suggest that a good test for fluxes inferred from OCO-2 was fluxes inferred by sparse IS data. That seems like a weak argument. Admittedly, this is a difficult situation (evaluating satellite data using models with poorly characterized errors and sparse in situ data) but in this reviewer's opinion relying on tropical fluxes inferred from IS data is not a great strategy.

The goal of the paper was, as mentioned line 63, to assess if there are any differences between MIPv7 and MIPv9, what would be the implications in the carbon cycle community of using v9 regarding precious studies that used v7. We have been able to observe over the tropics, changes between MIPV7 and MIPv9. Particularly, as observed in figure 9, the differences between the two datasets are not observed for the Northern Tropics but for the Southern Tropics where sources of carbon are observed with v9 while sink of carbon are observed with v7. In our discussion, we are comparing the new results obtained with our v9 ensemble with results provided in the study of Palmer et al., 2019 where they used OCO-2 v7 data in their inversions. Annual flux in PgC/yr are provided in the study of Palmer et al., 2019 allowing us to compare their results with our v9 over the same tropical regions. Our discussion does not mention that this previous study using v7 would have been similar using v9 data. On the contrary, the results of the two studies show that the conclusion of Palmer et al., 2019, in Tropical South Africa, would have been different if they had used v9. Indeed, as we mentioned in our discussion:

"Palmer et al. (2019) assimilated OCO-2 v7 land data and GOSAT v7 data separately during the El Niño period and analyzed the posterior emissions over the pan-Tropical regions. With their inversions, they found carbon sources of 1.56 PgC/yr in 2015 and 1.89 PgC/yr in 2016 over the Northern hemisphere of the tropics. The carbon sources they found with OCO-2 v7 are similar to what we obtain with our OCO-2 v9 inversions. Palmer et al. (2019) found with v7 that the largest seasonal cycle of carbon fluxes in the tropics was over Northern Tropical Africa. This analysis would have probably been similar with v9. However, while they had -0.21 PgC/yr in 2015 and -0.12 PgC/yr in 2016 over the Southern Tropical latitudes, we observe around 0.70 PgC/yr in 2015 and 0.4 PgC/yr in 2016 with v9 (Fig. 9). Analyzing our inversions at regional scales, we saw that this opposite sign in emissions was coming from Tropical South Africa. In our Fig. 11, we have been able to observe a source of around 0.25 PgC/yr in 2015-2016, while v7 emissions were around -0.25 PgC/yr for the same period, which is also what Palmer et al. (2019) found approximately.

They concluded that the largest carbon uptake was over the Northern Congo basin, situated in the Southern Tropical Africa MIP region. This result and the difference between v7 and v9 suggest that the conclusion over the Southern Tropical Africa MIP region with v7 would have been different with v9."

As we mentioned in our paper and summarized in the following table, we can see that the mean values are close between the v7 results of Palmer et al., 2019 and our v7 results:

	PgC/yr	North Trop.		North Trop. Africa		South Trop.		South Trop. Africa	
		2015	2016	2015	2016	2015	2016	2015	2016
Mean	Palmer et al., v7	1.56	1.89	1.78	1.96	-0.21	-0.12	-0.13	-0.18
	Crowell et al., 2019 Peiro et al., v7	0.89	1.21	1.18	1.33	-0.06	-0.2	-0.21	-0.18
	Peiro et al., v9	1.42	1.70	1.24	1.42	0.68	0.35	0.21	0.25
Median	Palmer et al., v7	1.55	1.64	1.89	2.02	-0.26	-0.05	-0.06	-0.20
	Crowell et al., 2019 Peiro et al., v7	0.87	1.15	1.26	1.35	-0.14	-0.19	-0.13	-0.165
	Peiro et al., v9	1.40	1.42	1.26	1.36	0.70	0.38	0.18	0.27

Table 1: Mean values (top) and median values (bottom) between the ensemble values using OCO-2 v7 from Palmer et al., 2019 and our v7 and v9 results. Yellow colors represent positive fluxes and blue colors represent negative fluxes. Fluxes are in PgC/yr.

Our IS data also suggests a source of carbon over Southern Tropical Africa with median fluxes like our v9 results. The study of Gloor et al., 2018 took a similar approach to that of Liu et al., 2017, but instead of using OCO-2 data, they assimilated NOAA surface station network. Our results with v9 and IS are close to those found in Gloor et al., 2018 suggesting a release of carbon over Southern Tropical Africa and particularly near the Congo basin during the 2015 El Nino event. We had to mention this study by Gloor et al., 2018 as they were the first ones to observe sources of carbon in both Northern Tropical and Southern Tropical Africa. And as we also mentioned in our discussion, CO anomalies were found at the same period over this region. Recently, several studies are focusing on Africa to investigate the sources of carbon observed in Tropical Africa.

9- MINOR/MAJOR. Line 384: ...we find better agreement between LNv7 and LNLGv9… So what? Without any context (which this reviewer is sure the authors can provide) this statement about two independent data products is redundant.

Throughout the paper, and when we started looking at latitudinal bands and the MIP regions, we could observe that LNLGv9 is particularly closer to LNv7 than LGv7. As we mentioned in the first notification of this sighting, page 18, line 324: "As we can observe for all other latitudes bands and we will observe for smaller regions, LNLGV9 tends to be closer to LNv7 than to LGv7. This points to previously known issues with the v7 LG data that were resolved with a unified bias correction in OCO-2 v9." This being already explained line 324 and being redundant line 384, we removed the sentence "Here again, we find better agreement between LNv7 and LNLGv9".

10- MINOR/MAJOR. This reviewer was not totally convinced by the authors' use of normalized bias - the MDM has a dynamic range of two orders of magnitude? It would be useful to also report the ensemble and individual model bias as a function of latitude. Similar argument goes for the standard deviation. Appendix?

MDM has a dynamic range of two orders of magnitude. So, the more meaningful interpretation of the residuals comes from the normalized residuals. For instance, a 0.1 ppm residual at the South Pole is more meaningful than a 10 ppm residual at the LEF tower during the summer months.

We have already included the ensemble and individual model bias as a function of latitude for the withheld evaluation since the first submission of our paper. Indeed, figure 12 shows the normalized bias and standard deviation for the ensemble mean of IS and LNLGv9 by latitude. Additionally, figure 14 shows the normalized bias and standard deviation by latitudes for each individual model. It is not necessary to add them as an annex since they have always been available in the main text.

11- MINOR. TCCON and in situ data are treated as the gold standard. My understanding is TCCON data, despite herculean efforts, still contain biases and those should be acknowledged. Admittedly, those biases are likely smaller than those for OCO-2 but they could be important.

The TCCON data used for validations are average over 30 minutes bins of local solar time and for each site. This gives different bins for different sites, even at the same latitudes or longitudes. TCCON observations are not assimilated but are only used for evaluation purposes, so the uncertainty assigned to the binned observations, which is of 0.5 ppm, does not matter much.

The figure below shows the uncertainty mean (in ppm) over 2015-2018 for each TCCON measurement. We can see that there is large variability in the uncertainties between the different TCCON sites with values between 0.3 ppm and 1.1 ppm. However, as mentioned previously, TCCON uncertainties from the data used in the evaluation are around 0.5 ppm.

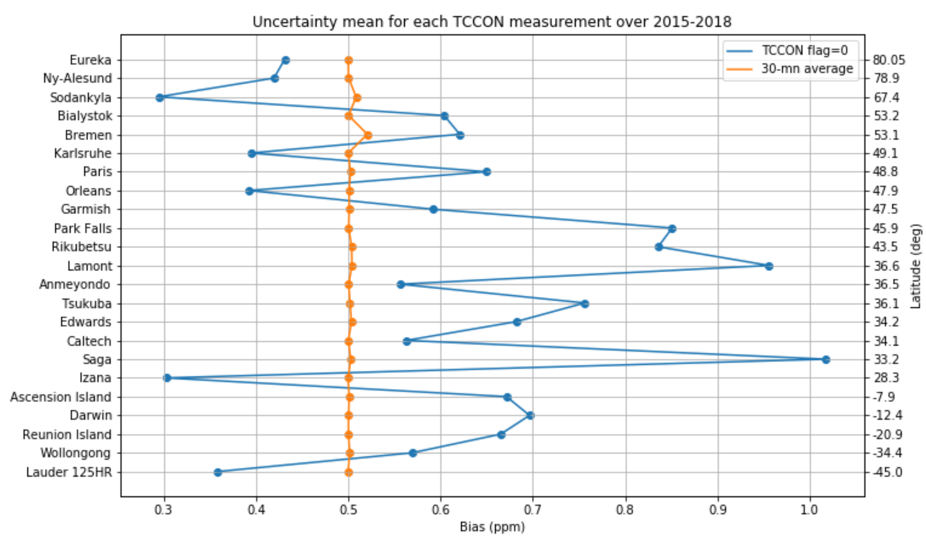

Fig 4. Uncertainty mean for each TCCON measurement over 2015-2018. The uncertainty from the original TCCON data is represented in blue while the uncertainties assigned to the 30 minute bins are orange. TCCON site names are on the left y-axis and latitude in degree corresponding to each site are on the right y-axis.

12- MINOR. This reviewer was disappointed by the incremental update in the length of analysis compared to Crowell et al 2019. OCO-2 was launched in 2014 and it is still to my knowledge still producing data so why has this study ignored 2019 and 2020? That's 50% more data than they have analyzed! Even if the author took into consideration that they needed 6 months of data at the end of their flux reporting period, they could at least report 2019.

The reviewer and readers need to understand that it takes a year for the modelers to update the MIP protocol and to run the simulations with the new retrievals. Then it takes some months to run the validations of these simulations and some more months to write the paper. This explains the laps between the OCO-2 retrieval versions used and the time the MIP paper is submitted.

Additionally, the incremental update in the length of analysis from our paper is the same length as for Crowell et al., 2019. Crowell et al., 2019 was submitted in 2019 and includes only the years 2015 and 2016. There were hence two years of data, 2017 and 2018, that were not included in the paper. This time range is the same as for our paper. We submitted our paper in 2021 and included four years of data (2015, 2016, 2017 and 2018) and did not include two years of data which are 2019 and 2020.

13- MINOR. On a related note, all the transport models are reasonably well described. However, the description of the inversions is lacking for some models. Uniformity in individual descriptions is needed. For example, what assumptions were made about uncertainties and spatial/temporal correlations - some model descriptions are more comprehensive than others.

We thank the reviewer for this remark where some information was indeed missing for some inversions. We added particularly prior uncertainties for the simulation CT, CSU, and CMS-flux.

The correction can be found on page 38 line 659 for CMS-flux:

"The prior ocean error is 100%. Fire is not optimized separately; they are part of the NBE."

The correction can be found on page 38 line 675 for CSU:

"Prior ocean comes from a climatology based on Landschutzer v18. Prior standard deviations (independent, no prescribed correlations) for ocean exchange, respiration and GPP were 10% of the net exchange (or respiration or GPP). Fires are from GFED4.1s."

The correction can be found on page 39 line 686 for CT:

"Prior standard deviation is equivalent to 50%. Prior covariance is applied such as the correlations between the same ecosystem types in different Transcom regions decrease exponentially with distance scale L = 2000 km.

More information can be found in:

https://gml.noaa.gov/ccgg/carbontracker/CT2019B_doc.php#tth_sEc8.2"

References

Babenhauserheide, A., Basu, S., Houweling, S., Peters, W., and Butz, A.: Comparing the CarbonTracker and TM5-4DVar data assimilation systems for CO_2 surface flux inversions, Atmospheric Chem. Phys., 15, 9747–9763, https://doi.org/10.5194/acp-15-9747-2015, 2015.

Chevallier, F., Palmer, P. I., Feng, L., Boesch, H., O'Dell, C. W., and Bousquet, P.: Toward robust and consistent regional CO_2 flux estimates from in situ and spaceborne measurements of atmospheric CO_2, Geophys. Res. Lett., 41, 1065–1070, https://doi.org/10.1002/2013GL058772, 2014.

Cressie, N., Bertolacci, M., and Zammit-Mangion, A.: From many to one: Consensus inference in a MIP, https://arxiv.org/abs/2107.04208, 2021.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S., Aragão, L. E. O. C., Arneth, A., Arora, V., Bates, N. R., Becker, M., Benoit-Cattin, A., Bittig, H. C., Bopp, L., Bultan, S., Chandra, N., Chevallier, F., Chini, L. P., Evans, W., Florentie, L., Forster, P. M., Gasser, T., Gehlen, M., Gilfillan, D., Gkritzalis, T., Gregor, L., Gruber, N., Harris, I., Hartung, K., Haverd, V., Houghton, R. A., Ilyina, T., Jain, A. K., Joetzjer, E., Kadono, K., Kato, E., Kitidis, V., Korsbakken, J. I., Landschützer, P., Lefèvre, N., Lenton, A., Lienert, S., Liu, Z., Lombardozzi, D., Marland, G., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Niwa, Y., O'Brien, K., Ono, T., Palmer, P. I., Pierrot, D., Poulter, B., Resplandy, L., Robertson, E., Rödenbeck, C., Schwinger, J., Séférian, R., Skjelvan, I., Smith, A. J. P., Sutton, A. J., Tanhua, T., Tans, P. P., Tian, H., Tilbrook, B., van der Werf, G., Vuichard, N., Walker, A. P., Wanninkhof, R., Watson, A. J., Willis, D., Wiltshire, A. J., Yuan, W., Yue, X., and Zaehle, S.: Global Carbon

Budget 2020, Earth Syst. Sci. Data Discuss., 2020, 1–3, https://doi.org/10.5194/essd-2020-286, 2020.

Gaubert, B., Stephens, B. B., Basu, S., Chevallier, F., Deng, F., Kort, E. A., Patra, P. K., Peters, W., Rödenbeck, C., Saeki, T., Schimel, D., Van der Laan-Luijkx, I., Wofsy, S., and Yin, Y.: Global atmospheric CO_2 inverse models converging on neutral tropical land exchange, but disagreeing on fossil fuel and atmospheric growth rate, Biogeosciences, 16, 117–134, https://doi.org/10.5194/bg-16-117-2019, 2019.

Massoud, E. C., Lee, H., Gibson, P. B., Loikith, P., and Waliser, D. E.: Bayesian Model Averaging of Climate Model Projections Constrained by Precipitation Observations over the Contiguous United States, J. Hydrometeorol., 21, 2401–2418, https://doi.org/10.1175/JHM-D-19-0258.1, 2020.

Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., Niwa, Y., Patra, P. K., Peters, W., Rayner, P. J., Rödenbeck, C., van der Laan-Luijkx, I. T., and Zhang, X.: Global atmospheric carbon budget: results from an ensemble of atmospheric CO_2 inversions, Biogeosciences, 10, 6699–6720, https://doi.org/10.5194/bg-10-6699-2013, 2013.

Saunois, M., Stavert, A. R., Poulter, B., Bousquet, P., Canadell, J. G., Jackson, R. B., Raymond, P. A., Dlugokencky, E. J., Houweling, S., Patra, P. K., Ciais, P., Arora, V. K., Bastviken, D., Bergamaschi, P., Blake, D. R., Brailsford, G., Bruhwiler, L., Carlson, K. M., Carrol, M., Castaldi, S., Chandra, N., Crevoisier, C., Crill, P. M., Covey, K., Curry, C. L., Etiope, G., Frankenberg, C., Gedney, N., Hegglin, M. I., Höglund-Isaksson, L., Hugelius, G., Ishizawa, M., Ito, A., Janssens-Maenhout, G., Jensen, K. M., Joos, F., Kleinen, T., Krummel, P. B., Langenfelds, R. L., Laruelle, G. G., Liu, L., Machida, T., Maksyutov, S., McDonald, K. C., McNorton, J., Miller, P. A., Melton, J. R., Morino, I., Müller, J., Murguia-Flores, F., Naik, V., Niwa, Y., Noce, S., O'Doherty, S., Parker, R. J.,

Peng, C., Peng, S., Peters, G. P., Prigent, C., Prinn, R., Ramonet, M., Regnier, P., Riley, W. J., Rosentreter, J. A., Segers, A., Simpson, I. J., Shi, H., Smith, S. J., Steele, L. P., Thornton, B. F., Tian, H., Tohjima, Y., Tubiello, F. N., Tsuruta, A., Viovy, N., Voulgarakis, A., Weber, T. S., van Weele, M., van der Werf, G. R., Weiss, R. F., Worthy, D., Wunch, D., Yin, Y., Yoshida, Y., Zhang, W., Zhang, Z., Zhao, Y., Zheng, B., Zhu, Q., Zhu, Q., and Zhuang, Q.: The Global Methane Budget

2000--2017, Earth Syst. Sci. Data, 12, 1561–1623, https://doi.org/10.5194/essd-12-1561-2020, 2020.

Schuh, A. E., Jacobson, A. R., Basu, S., Weir, B., Baker, D., Bowman, K., Chevallier, F., Crowell, S., Davis, K. J., Deng, F., Denning, S., Feng, L., Jones, D., Liu, J., and Palmer, P. I.: Quantifying the Impact of Atmospheric Transport Uncertainty on CO2 Surface Flux Estimates, Global Biogeochemical Cycles, 33, 484–500, https://doi.org/10.1029/2018GB006086, 2019

Schuh, A. E., Byrne, B., Jacobson, A. R., Crowell, S. M. R., Deng, F., Baker, D. F.,Weir, B.(2021), Chinese land carbon sink obscured by transport model uncertainty, Arising Matters, (Accepted Reply to Nature Communications, https://doi.org/10.1038/s41586-020-2849-9, not in press yet)