Atmospheric Chemistry and Physics Discussions

Atmos. Chem. Phys. Discuss., referee comment RC2
https://doi.org/10.5194/acp-2021-373-RC2, 2021

Comment on acp-2021-373

Anonymous Referee #1

Referee comment on "Four years of global carbon cycle observed from the Orbiting Carbon Observatory 2 (OCO-2) version 9 and in situ data and comparison to OCO-2 version 7" by Hélène Peiro et al., Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2021-373-RC2, 2021

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.

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.

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.

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 amongst 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 CO_2 fluxes consistent with NOAA atmospheric CO_2 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.

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.

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.

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 Piero 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 Piero 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.

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.

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.

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?

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?

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.

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?

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 characterised 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.