Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2020-170-RC2, 2020

Interactive comment on "Spatially resolved evaluation of Earth system models with satellite column averaged CO_2" *by* Bettina K. Gier et al.

Anonymous Referee #2

Received and published: 16 July 2020

This study used satellite-observed column-average CO_2 (XCO_2) to evaluate how well the current generation of earth system models reproduces atmospheric CO_2 variability. The authors compared spatially resolved model simulations of XCO_2 with observed XCO_2 in terms of biases, growth rates, seasonal cycle amplitudes (SCA), and trends in SCA. They found that most models overestimate XCO_2 and the growth rates of XCO_2, but underestimate the seasonal amplitudes of XCO_2. The study is novel and interesting, and could be considered for publication after concerns are addressed.

General comments

1. The apparent "trend reversal" of SCA in satellite XCO_2 caused by the sampling coverage bias raises an important question whether other characteristics of the observed XCO_2 were subject to the same bias. It seems that the "sampled" vs. "unsampled" comparison was done only for the trend in SCA. Would this comparison give different results for the growth rate of XCO_2?

2. Although the effect of different spatial coverage between the two instruments has been addressed in Figs. 9 and 10, there has been no mention of whether the measurement scales of XCO_2 were calibrated between the two instruments. This could be another source of bias. During the overlapping period, did SCIAMACHY and GOSAT measurements agree well with each other?

3. In addition to the spatial distribution of SCA shown in Fig. 6, it would also be interesting to compare the simulated latitudinal gradients (i.e., the zonal mean) of SCA from different models on the same plot.

4. Given that there was no correlation between the growth rate and the growing-season temperature anomaly in the *observations* (Fig. 5), I'm not convinced that a robust emergent constraint relationship can be established for the period of 2003–2014. Note that this is different from Fig. 2 in Cox et al. (2013) *Nature* in which a clear correlation is seen in the observations. In the absence of a relationship of the same kind in the observations, one could not distinguish a real emergent constraint from an artifact of model assumptions. It may as well exist for a longer period with more data, but we couldn't tell. I suggest moving the paragraph of P8L248–P9L263 as well as Fig. 5 to the supplement.

5. In all figures that show a regression line, it would be better to show both the Pearson correlation and the p-value.

6. The study would benefit from a more structured discussion of lessons learned from model evaluation, for example, what the likely causes of model biases in XCO_2 and the SCA of XCO_2 would be. There are some good points made in section 5.3 and the summary, which could be better organized.

Specific comments

- Abstract: The abstract is quite long and technical. Describe the major findings concisely and leave nonessential details to the main text.
- P1L20: The multi-model means give a feeling that CMIP6 models had a great improvement relative to CMIP5 models. But a closer look at Fig. 3 would tell that this was mainly because negative biases canceled positive biases. I suggest adding the multi-model standard deviations of the bias (or other statistics that characterize the spread) in parentheses.
- P1L25–32: The "trend reversal" in SCA caused by sampling bias is not clear on a first reading. It would be helpful to write this in a way that is less entangled.
- P2L46–60: If the purpose of this paragraph is to introduce SCA, then the first few sentences seem quite redundant. Better get to the point straight away.
- P2L64: Missing Ciais et al., 2013 in the References.
- P3L65: "downlooking" → "downward-looking"
- P3L89–90: "such as a general overestimation of photosynthesis [relative to data-driven models]." We don't know the magnitude of the global photosynthesis with certainty. The number could range from 112 to 169 PgC yr^{-1} (Ryu et al., 2019, *RSE*). The MTE GPP data product (Jung et al., 2011, *JGR*) that was used to evaluate the CMIP models would sit near the lower end of this range.

- P5L131: Missing Dlugokencky et al., 2018 in the References.
- P7L203: "parameters" → "variables"
- P7L205: Surface observations sample the air within or closer to the boundary layer, and therefore may have a larger seasonal swing.
- P8L241: What is the correlation between the observed GR and the multi-model mean GR?
- P9L256: The increase of GR sensitivity to temperature after including 2015 and 2016 data could have been due to El Niño.
- Table 1: Use whole numbers in the "Altitude" column.
- Figure 2: I suggest lightening the gray background to prevent it from interfering with the reading of the curves.
- Figure 3: It seems that after detrending, the seasonal variability in the multi-model mean XCO_2 would match quite well with that in the observed XCO_2, and better than the seasonal variability in any individual models. I wonder what the correlation coefficients would be.
- Figure 4: I think ranking the models by their average growth rates, from low to high, would make the figure clearer.
- Figures 6a and 6b: Why not use the same scale? One could reserve the purple region of the colormap for the high SCA values from MPI-ESM-LR. This would not affect other models that have values represented in colors from blue to red.

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2020-170, 2020.