Review of "The Information Content of Dense Carbon Dioxide Measurements from Space: A High-Resolution Inversion Approach with Synthetic Data from the OCO-3 Instrument" by Roten et al.

This work applied observing system simulation experiments (OSSE) to evaluate the ability of OCO-3's SAM (synthetic) measurements to constrain fossil-fuel CO2 emissions using Bayesian-inversion modelling. The Vulcan 3.0 emission inventory and the X-STILT model were used to create synthetic CO2 enhancements. In the inverse model, Vulcan 3.0 emission inventory is considered as a true emission, whereas ODIAC emission inventory is considered as a prior emission.

This manuscript raises interesting and fundamental questions about the how the emission assessments are influenced by factors such as grid size of prior emission and transport error, as well as how will constraining in the locations and uncertainties of large point sources affect the inversion scheme and how aggregated SAMs affect bias correction. The answers to these questions could improve our understanding of inversion results as well as how to optimally use satellite measurements for emission assessments.

However, the manuscript is not easy to follow for people who are not inversion experts, and therefore requires some efforts for rewriting and restructuring. The results section can be more focused and shortened, so that the readers can easily grasp the outcome of the sensitivity studies. A table can be utilized to summarize the test conditions and achieved results (similar to Table 2). A better presentation will be needed for the readers to follow the concepts and appreciate the benefits of the results. The description of the Bayesian inversion scheme is a bit confusing in the end of section 2.4, i.e. how (6) is obtained. Also It is confusing that in the test 4 different criteria are used compared to other test cases.

Also I find it a confusing message that the posterior emission estimates are further from the truth compared to the prior emission, which are shown in Fig. 9, Fig. 13, and Fig.16. It introduces doubt to the general applicability/benefit of Bayesian inversion. In my opinion, the setup of the inversion framework regarding Q and R should be improved.

In addition, the authors uses terms such as "effectiveness of optimization", "corrective power" which sometimes just refer to the difference between posterior and prior estimates or the improvement of the fit to the data. The optimization of the emission estimates should be given by comparing with the true fluxes. Consequently, there are also misleading conclusions that may confuse the readers.

Apart from the general comments listed above, below please find some specific comments:

Line 114: Is the total emission of ODIAC and Vulcan different from one another? It's worth knowing because ODIAC typically underestimates road emissions because it uses night-time light data as a proxy.

Line 164: The first term on the right hand side…

Line 170: "R reflects uncertainties in DXCO2 observations from various components" is a bit misleading because it sounds like it only consists of observation errors. R consists also of transport errors, etc. as listed in Table 3.

Line 180: and instrument error (ε). Could you please elaborate how did you incorporate other error sources such as the transport error?

Figure 3: the color bar should be denoted as DXCO2 [ppm]

Figure 5: Histogram plot of difference between prior (customized and non-customized) and true emission will be more intuitive.

Line 328: "It demonstrates that the effectiveness of the optimization is directly proportional to the observed enhancement"
This, to our understanding, implies that good emission optimization necessitates a significant increase in CO2. In most low wind speed cases, CO2 will be significantly increased, but transport error could be high. Therefore, it is preferable to see an error bar in plot 6 (c). In addition, I do not think the effectiveness of optimization can be represented by Posterior flux – Prior Flux, it should be compared with the "true flux".

Figure 6 (b-d) and section 3.1.1 (319-334): Similar to the point before, to understand the inversion's ability to optimize estimates, the posterior flux should be compared with true flux (posterior flux - true flux) i.e., corrective power. If so, the difference between posterior flux and prior flux (posterior flux – prior flux), i.e., amount of correction, don't indicate the inversion's ability. If my understanding is correct, then it is also applicable to figure 7, 8, 12 and 15, and its discussion part.

Figure 9: The y axis should be "differences" in total emissions. It also applies to Figure 13 and 16. I would also write "Overestimate" instead of "Over Estimate"

Line 519: The background approach mainly accounts for major uncertainty. In this study, the authors used synthetic CO2 enhancement. So, they cannot assess the inversion ability for varying background uncertainty. However, in the paper, they only considered the background error from one paper (Kiel et al, 2021). The authors could consider and discuss the background approaches from other cases. For example:
Wu, D., Liu, J., Wennberg, P. O., Palmer, P. I., Nelson, R. R., Kiel, M., and Eldering, A.: Towards sector-based attribution using intra-city variations in satellite-based emission ratios between CO2 and CO, Atmos. Chem. Phys. Discuss. [preprint], https://doi.org/10.5194/acp-2021-1029, in review, 2022