Review of Santaren et al (DOI:10.5194/amt-2020-138)

The authors have done a series of inversions in an OSSE framework to explore the feasibility of a wide swath CO_2-sensing satellite mission such as CO2M to quantify the fossil fuel (FF) CO_2 emissions from individual cities, clusters of cities, and regions. They have relied on uncertainty reduction as the primary metric for evaluating their inversions. The work they have done is mathematically correct, and the results follow expectations from prior experience with inversions, i.e., they "make sense". However, there are three critical shortcomings, owing to which I cannot recommend publication unless they make the suggested changes. Since making these changes is likely to require a fair amount of additional work, and may change their conclusions, I am classifying them as "major" changes.

First, in any CO_2-based effort to quantify FF CO_2 emissions, the biggest confounding factor is the non-FF variation of CO_2. This can be due to biosphere fluxes over continents, or, in a regional study, due to inflow/outflow at the boundary. This has been a well-known problem both in the context of estimating FF CO_2 from in situ and satellite CO_2 measurements (e.g., https://doi.org/10.1002/2014GL059684 and https://doi.org/10.1029/2019JD030528 respectively). Biases in the assumed NEE – which is very likely in any biosphere model – **will** lead to biases in derived FF CO_2, which makes any uncertainty reduction irrelevant. And yet, in this set of OSSEs the authors skirt this very important issue, assuming the prior NEE to be unbiased. While this does not make the results wrong, it makes them less than useful for evaluating the potential of a CO2M-like mission, which will surely have to contend with unknown biospheric CO_2 fluxes.

Second, in the context of satellite CO_2 instruments, an additional complication is the data gap or poor data due to cloud cover, aerosol loading, and other factors. For example, for the currently flying OCO2 spectrometer, only a few percent of its footprints result in good quality XCO_2 retrievals. For the domain the authors have chosen (NW Europe with lots of urban centers), both cloud cover and aerosol loading are important limiting factors. Yet, the authors explicitly ignore this complication, "The cloud cover and the corresponding gaps in the spaceborne passive XCO2 sampling are ignored" (L100). I understand that simulating realistic clouds and aerosols is difficult, but for an OSSE to be realistic, some attempt must be made. For example, the authors could have used the statistics of past cloud and aerosol data to introduce sampling gaps. Or, they could have taken the fraction of good quality retrievals among all footprints from an existing NIR XCO_2 instrument like OCO2 and then downsampled their footprints. The results presented here without considering realistic sampling gaps are mathematically correct, but not very useful for evaluating the capability of any real CO_2 mission.

Third, the authors use the posterior covariance matrix and uncertainty reduction as performance metrics for their inversions throughout the paper. Given the importance of uncertainty calculation to the work (as opposed to reduction of biases in their priors), I would like to see a more realistic specification of flux and data uncertainty. Currently they assume (i) uncorrelated retrieval errors (L193), which is unrealistic for the small footprint and dense sampling that they've given their satellite instrument, and (ii) no spatial correlation in their prior flux covariance **B** (L334), also unrealistic given the high spatial resolution of their fluxes. I do not understand why they need to make either simplification, since in a batch inversion (as opposed to an iterative approximation like EnKF or 4DVAR) one can actually specify off-diagonals in both **B** and **R**. The simplifications (i) and (ii) make their posterior covariance, and the conclusions based thereon, not very relevant for real-world inversions of CO2M-like satellite data.

A general comment I would like to make about the three issues I raise above is that in inversions of real satellite data, modelers often have to make simplifying choices to make the problem tractable. E.g., ignoring off-diagonal elements in **R** is pretty common, although more recent inversions try to at least account for correlations in **R** by error inflation, aggregating, or data thinning (e.g., https://doi.org/10.1029/2007GL030463, https://doi.org/10.5194/acp-13-8695-2013 and https://doi.org/10.5194/acp-19-9797-2019). Similarly, inversions with an iterative approximation like EnKF or 4DVAR often have an inexact posterior covariance due to computational limitations, while many inversions ignore biases in satellite retrievals because it's still an open problem. However, making too many simplifications in a single OSSE, as the authors have done here (unbiased priors and retrievals, no sampling gaps, uncorrelated retrieval errors, no spatial error correlation in fluxes) makes the results inapplicable to any real-world satellite instrument. I will also point out that several co-authors of this manuscript have previously authored papers stressing the importance of these complicating factors (e.g, https://doi.org/10.1029/2007GL030463, https://doi.org/10.5194/amt-11-681-2018, and https://doi.org/10.1088/1748-9326/ab7835) and published inversions with far more realistic assumptions, which makes the current set of simplifications all the more surprising.

Beside these major issues, here are a few minor points that need correcting or clarifying:

1. L156: Delete "and vertical". The vertical resolution comes later.
2. L161: A high spatial resolution (~2x2 km^2) implies higher temporal resolution as well. If the driving winds are 3-hourly, what provides high frequency variation in the CHIMERE winds?
3. L175: The gradients in column CO_2 due to the top 30% of the atmosphere would be small, agreed, but how large are they? Signals in column CO_2 are deceptively small, so terms that seem to be negligible are not always negligible.
4. L180: Switch 92.8° and 705 km.
5. L227: Since the quantity directly estimated is the FF CO_2 emission between 5 and 11 local time, to estimate the total emissions one would need an accurate diurnal cycle. What is the uncertainty in the diurnal cycle of FF CO_2 emissions?
6. Section 2.2.2 and elsewhere: The word "controlled" keeps confusing me. Do you mean "estimated", as in part of the control vector? Or do you mean "controlled" as in kept in control, static, not changed? I suspect you mean the former, but "controlled" in English can also signify "not allowed to change". I'd suggest using the word "estimated" or "optimized" if you mean the former.
7. L280: Is this an over-determined problem? Then that's not very common in flux inversions, and is likely due to the unrealistic correlations in **R** and **B** (one of my major concerns).
8. L290: Is random noise added to **y**, consistent with **R**?
9. L304: Typo, change XO_2 → XCO_2
10. L348: With the assumptions detailed here at the grid scale, what is the uncertainty on the (say) annual total or seasonal total NEE and FF? Aggregate numbers are easier to make sense of than grid-scale specifications.
11. L399: "Figures 2i" likely means all the subplots of Figure 2. In that case, just say "Figures 2", no need to add the "i".
12. L437: Speed is one aspect of the wind, direction is the other. Since wind direction determines how well plumes present themselves to a satellite that is going one way, uncertainty in wind direction must be considered as well as speed uncertainty. Was that done here?
13. L477: Remove "uncertainty", **B** is just the prior covariance matrix.
14. L519: Again, I'd like to see the uncertainties on aggregated fluxes, such as annual totals.
15. L560: In Figure 7, why do larger emissions have smaller uncertainty reductions?
16. L582: "*… and thus by the variability of these fluxes, during the month of May*". This only matters because the uncertainty on the NEE is larger in May, right? Because this metric/score does not care about the actual prior NEE.
17. L794: "*Efforts have been made to limit the amplitude of such errors in the concept of the new CO2M mission. Our new inversion framework allows accounting for a realistic simulation of the observation sampling and errors.*" I disagree with this statement in the context of this paper, especially the part about a realistic simulation, because of the three major points raised above.