

Updated review of Lan et al, 2017.

The paper is generally improved over the previous version, and will be ready for publication after the following minor revisions are made, and one major revision is considered.

This review is ordered from most significant to least.

- The abstract makes this statement: “Our results confirm that continental-scale total column XCO₂ gradients simulated by CarbonTracker are realistic and can be used to evaluate the credibility of spatial patterns from satellite retrievals.

First, the authors only demonstrate that it is reliable in North America. No statements within the paper address why it should be reliable everywhere, considering that far more data are assimilated into CT from North America than anywhere else. The authors made such statements in the response to reviewers – these needed to be added to the paper to corroborate this claim.

Second, the authors make no mention of potential sampling bias in the observations. In the case of their focus on the SCIAMACHY observations over Europe, section 3.5 still makes no mention of the possible sampling biases in the observations. The authors are sampling all data points over Europe all the time from their model. SCIAMACHY only makes measurements when it flies overhead, and when it is sufficiently clear (and this is true of all the CO₂-measuring sensors). When you average these irregular observations together, *you do not get a mean spatial pattern*. You get a mis-mash that includes whatever samples you happened to take. The authors need to mention this in the revised manuscript as a possible explanation of the appearance of an unphysical spatial pattern in the Reuter et al (2014) figure. How the data are used/assimilated is the critical factor. If the data are used ignoring this fact, it is of course a problem. But nearly all inversion systems sample the data at the times and places of the observations, so this effect is at least partially taken into account. Therefore, the claim in the abstract that CT can be used to evaluate mean spatial patterns from satellites is dubious, since *they are simply not the same thing*. The only way to get around this issue is to sample the model like the satellite, which the authors currently do not do nor discuss. For the paper to be acceptable, this statement in the abstract, and all related statements throughout the paper, needs to be either eliminated or qualified with this caveat.

- Regarding the comparison to TCCON, this comparison should at least be in the supplementary materials, along with error statistics. People have been comparing TCCON to aircraft for a long time and while it is not perfectly apples-to-apples, it can give a good idea of consistency. The authors saying “we don't need to do this” is unacceptable to this reviewer, considering they've already done it, they just need to include it.

- Regarding the error analysis and the missing of the surface layers, it was a misunderstanding on my part, so I withdraw that criticism.
- Line 434: Again the authors need to stress that CT is realistic over North America. Please rewrite to be: “Since spatial gradients from *CT2015* have been shown to be realistic on continental scales over North America, boreal fluxes inferred by CT2015 ... may be more trustworthy than....”
- Line 436: “However, the European carbon sink is still *elusive*;”.