Review of amt-2023-194

The authors present performance estimates for their retrieval algorithm, FOCAL, applied to the upcoming CO2M mission. The performance estimates are based upon more detailed radiative transfer simulations using SCIATRAN, and are focused on the core retrieval algorithm biases rather than instrumental, spectroscopic, or meteorological. The authors estimate that FOCAL in it's current form meets (just!) the extremely stringent XCO2 and XCH4 requirements that have been placed on CO2M over land.

Overall the paper is well written, provides useful information, and is well suited for AMT. However I have concerns about how representative the bias correction method is when being trained on the exact truth. More justification and information on the bias correction should be included, and conclusions updated accordingly. Specific comments follow below.

General Comments

I recognize the need for bias correction, however I think what is done here may go beyond what is typical and requires at minimum more justification. The authors train their bias correction scheme against actual true values, which will not be known for CO2M. No results of the bias correction are shown, except it is clear that it likely has an impact on their final uncertainty quantification for large-scale fluxes based upon Figure 11 showing the best performance for the one month of data (April) that the bias correction is trained on. I would like to see more detail on what effect the bias correction is having, as well as what parameters it has deemed important. After all since this is entirely simulated data the results of the bias correction should be fully explainable. I realize the authors state that their results "do not consider any systematic errors in meteorology" but what this means in terms of the bias correction is missing.

In some places I find the conclusions of the authors a little optimistic. For example the last sentence of the manuscript is "However, the current results give good confidence that the FOCAL-CO2M retrieval is able to fulfil the product quality requirements of the CO2M mission." But my interpretation of the results is if you neglect things not taken into account in the simulation (polarization, full sphericity, 2d/3d effects, among others), assume you have perfect spectroscopy and input ancillary data, assume you have a perfect instrument, and perform bias correction using perfect truth values, then you barely meet the requirements. This is not a knock on the retrieval method, the CO2M requirements are very stringent, but it is hard for me to have a takeaway other than that it is unlikely FOCAL will meet the requirements when applied to real CO2M data in its current form.

Specific Comments

p.3 l.57:
I realize that a full description of the three algorithms is beyond the scope of this paper, and that they can be found in the corresponding citations, but a brief statement here about their differences would be helpful for the reader.

p.3 l.78: "The SCIATRAN calculations are more complex than the FOCAL forward model. For example, they consider surface BRDF (bidirectional reflectance distribution function) effects, different aerosol types and distributions as well as clouds."
This needs some further elaboration. While SCIATRAN itself can be much more complicated than the FOCAL forward model, it may not be configured that way. I see later on that polarization is neglected in the SCIATRAN calculations for example. I assume other features of SCIATRAN such as line of sight sphericity are not included even though they are supported as well.

p.5 l.143: "we filter out all cloudy data"
Surely not every scene with any clouds are filtered out or else there would be no purpose in including clouds

in the SCIATRAN runs?

p.6 l.171: "based on a set of training and test data"
Most algorithms now recommend splitting data into three groups: Train, test, and validation. The reasoning being that you can overfit to your "test" data by tweaking your model parameters, or even what model you are using.

p.6 l.183: "bias correction does not consider any additional errors resulting from systematic differences between the estimated meteorological conditions and the actual atmosphere"
Could this be included in some way? I see from previous publications that for real data truth is taken from a generated database. Is there no error estimate for this database that could be artificially included here?

p.7 l.204: "...cloud-free data ..."
As before, do you mean cloud VOD is identically 0? Or are some thin clouds included?

Section 4.1: For the plume specifically we are probably getting down to the scale where 2d/3d effects are important. I don't expect the authors to quantify these effects since I know SCIATRAN is not capable of doing so and it is a large amount of work, but I was surprised to see no mention of the effect

Section 4.3: Does the bias correction use the FOCAL retrieved optical depth in a significant way? I assume the idea would be to use the MAP aerosol parameters in the bias correction instead. Could you not test this directly by using the true aerosol OD + estimated errors of the MAP retrieval in the bias correction?

Technical Corrections

p.2 l.45: "welling" $\rightarrow$ "upwelling"

p.2 l.53: I would recommend using "better" here instead of "higher", since "higher spatial resolution" can be ambiguous.

p.9 l.255: "is not much sensitive"
Reword

p.9 l.257: "an" $\rightarrow$ "any"