We thank the anonymous reviewer for taking the time to review the revised manuscript a second time and providing helpful feedback. Your advice has been very helpful and allows to better link our truth proxy to 3D cloud effects.

The supplement contains the reviewer comments in black and our responses are shown in red *Italic* with citations from the manuscript in " ".

Reviewer 1:

Thank you for addressing the concerns raised in the review. The quality of the manuscript has been increased greatly.

My main remaining concern is the "truth metric", which uses the deviation from the local median in small area (maximum 100 km). This approach could underestimate the cloud-related bias, when the majority of OCO-2 pixels is in the vicinity of clouds (<10 km). To avoid such bias, it would be necessary to remove all pixels near clouds before computing the median. Since cloud distance is already available from the MODIS product, I think it should be easy to check if filtering for cloud distance affects the results and conclusions of this study.

That is an excellent idea and links the truth proxy more directly to 3D cloud effects. Thanks for the suggestion. We recalculated the small areas truth metric as suggested, recomputed the various steps with our algorithm, and regenerated the figures and tables. While the overall conclusions are similar the previous approach slightly underestimated existing 3D cloud biases.

Discussion of the new truth proxy in 3.1 L96: 'As a pre-processing step we match the 3D cloud variables, OCO-2 soundings, and TCCON by time and location. Afterwards, we remove soundings where no 3D cloud variables are available. To develop the bias correction model, we use the small areas analysis, which is based on the assumption that CO2 is a well-mixed gas and assumed to be constant over spatial scales of less than ~100 km (though, there can be exceptions for strong CO2 emitters such as mega cities). To exploit this constraint on XCO2 we split OCO-2 soundings from the same orbit into small areas with a maximum size of 100 km. Each small area is generated by collecting soundings (sorted by observation time) until the distance between the first and last sounding exceeds the 100 km threshold. Afterwards, the collection process of the next small area is started. For each small area we identify soundings that are assumed to be free of 3D cloud biases (nearest cloud is at least 10 km away). From those soundings we define the median retrieved XCO2 as the true XCO2 of a given small area and any differences to this median are treated as biases. Small areas that contain less than 10 soundings free from 3D cloud biases are removed from the dataset. Since this process biases the remaining small areas towards longer cloud distances, we resample the remaining soundings so that the distribution of nearest cloud distances is similar to the original data set with about 40% of the sounding having a nearest cloud distance of less than 4 km.'

Minor comments:

L45: The sentence is a bit unclear. Do you mean that 25% of all OCO-2 spectra are removed by the pre-processors that mainly look at clouds?

Removed that sentence to not confuse the reader.

L129f: Are the data "with QF=0 and QF=1" filtered here using the values from Table 1?

That is correct. Changed the sentence to clarify this. It now reads in section 3.1 L110: 'Next, we remove outliers with large XCO2 errors from the data set by applying a series of thresholds to the variables from the state vector. The variables and their thresholds are given in Table 1. Note that these filters remove only a small fraction of soundings (4%).'

L479: "operational bias-corrected OCO-2 retrieved XCO2": this is quite difficult to read.

Agreed. Simplified the sentence to: '… allows to mitigate 3D cloud biases in OCO-2 retrieved XCO2.'