

In the following, the review is shown in plain text, our answers are in blue, and text that has been added to the manuscript is in green.

Reviewer 2 : Sihe Chen

Overview

This paper corrects a problem in a previous work, David et al., (2021). In that work, the NN is overfitted to be able to predict the latitude of the sounding. In this work, the problem from the previous work is fixed, and the weak CO₂ band is excluded from the analysis. Good results are obtained in comparison to different XCO₂ sources.

Comments:

The article provides a very good general network trained for prediction of XCO₂. I recommend that this article be added to the literature with several of the following minor issues addressed.

We thank the reviewer for his work on our paper and for the positive comments

The authors have clearly shown a nice hypothesis of how D21 could have made a precise prediction on its location. From my perspective of view, this hypothesis is not hard to test. For example, the authors could try to train a NN with wCO₂ as the input only and try to retrieve the time information, or they can show a figure just like Figure 1 for an NN with wCO₂ removed.

We have made a number of attempts to try to understand the information content that is used by the Neural Network. Our analysis shows that, as explained in the manuscript, the NN behavior is very different when the wCO₂ band is used, or not, as input. We do not understand the reviewer's suggestion concerning Figure 1. The reviewer may have meant Figure A1, in which case this is what we have done : Figure A1 shows that no plume feature is retrieved by the NN when the wCO₂ band is used. Conversely, this feature is retrieved when it is not used (Figure 3).

Regarding the description of NN: can you provide the loss function that is used? Also, for the NN structure, I suggest that you show what specific hidden layer numbers are chosen in a table and how they are chosen. Doing something similar for the number of hidden layers could be good too, which should be better than simply stating the decision to be related to experiences. An example is Chen et al., 2022, fig. 8: <https://www.sciencedirect.com/science/article/pii/S0022407321005409>

We use the Mean Absolute Error as the loss function. We have added the information suggested by the reviewer in the revised version of the manuscript

Line 267: At first it was stated that sCO₂ band was considered but later an O₂ band albedo increase is specified. Also, it would be good if both bands' albedo could be compared to the standard deviation, so that we can see which factor plays a more important role.

Indeed, there was a typo. Only the sCO₂ band was supposed to be mentioned. We made the correction (thanks!) and added a sentence: The estimate precision is also a function of the O₂ band albedo, but this effect is not as strong and the O₂ band albedo shows less variability than that of the sCO₂ band.