Reviewer-1

Dear Authors, thank you very much for your revision of the manuscript. I appreciate that you evaluated your model on plumes in Section 5.2 and looked at some ablations studies and feature importance to better understand your model.

However, I still have major concerns with the proposed study:

[1] The results of your analysis of retrieving CO_2 enhancements, or plumes, are concerning and seem to indicate problems with the retrieval approach. The plume in figure 13 b) indicates that the ML retrieval did not retrieve the full XCO_2 enhancement as compared to OCO-2. If one were to calculate emissions from both retrievals, they would end up with very different answers. For Figure 13 c) it is challenging to evaluate how both algorithms compare due to the many overlaying observations (A running mean for each retrieval would have been helpful to better differentiate between both retrievals.). However, it seems that the ML retrieval is systematically higher compared to OCO-2 except for where OCO-2 measured the XCO_2 plume. Figure 13 a) shows the clearest plume structure but suffers from systematic low bias compared to OCO-2 in the proximity of the plume. Depending on how the XCO_2 background would be estimated, emission estimates would be vastly underestimated for all three cases. (The plots shown in Figure 12 don't add much value for comparison since their color bars change for all subplots.) In contrast to my conclusion given your plots, the manuscript summarizes the plume evaluations as "a powerful confirmation to our model's capability to retrieve genuine atmospheric XCO_2 from OCO-2 spectral data." I can't share this perspective.

Thank you sincerely for your insightful feedback. We have carefully analyzed your comments and refined our methodology to incorporate your valuable suggestions. Your critical input throughout this process has been truly invaluable. With these modifications, we are confident that our manuscript has significantly strengthened.

In this revision, we've made significant improvements to enhance the accuracy and reliability of our model. Additionally, we've conducted further analyses to better quantify and compare the enhancements in XCO_2 predictions and CO_2 plume detections. Perhaps the most crucial modification is the inclusion of a pre-processing layer in the MLP model. This layer serves as a way to address issues related to the measurement quality, specifically those represented by the bad sample list from the OCO-2 instrument's grating.

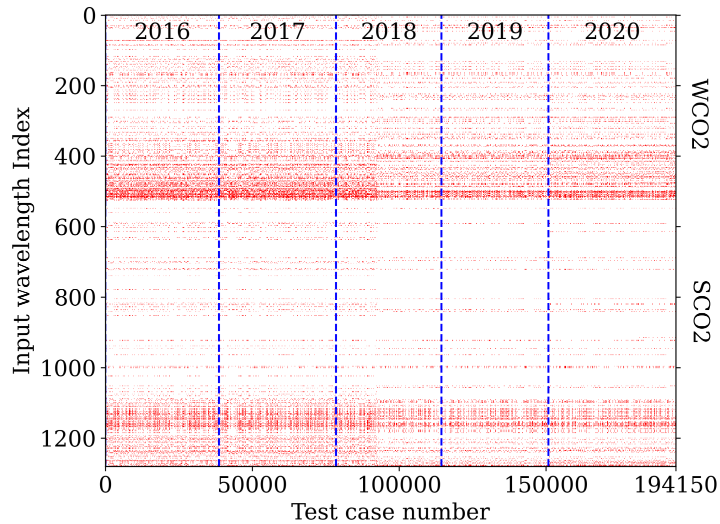

Figure 1: Visualization of the OCO-2 satellite data quality across wavelength grid indices. The color map illustrates the bad sample list extracted from OCO-2 Level 1B files for all test cases. On the *x*-axis, case numbers range from 0 to 194,150, while the *y*-axis represents various wavelength grid indices, ranging from 0 to 1,280. Red coloration denotes values greater than zero, indicating problematic data.

The grating of the OCO-2 satellite undergoes subtle changes due to the natural degradation of the instrument and more pronounced changes following updates to the L1B pixel mapping algorithm. As depicted in Figure 1, the period from 2016 to mid-2018 represents one phase, and the period after the second half of 2018 marks another, due to an algorithm update in the OCO-2 bad sample map [1]. Consequently, we've introduced a new pre-processing layer, called the "bad sample filter," which is applied to the spectral input before it enters the initial layer of the MLP. As depicted in Fig. 2 in the latest manuscript, the model filters out potentially low-quality wavelengths based on the largest union of the bad sample lists from reference data in 2016 and the initial list following the 2018 algorithm update, ensuring that only inputs marked as good radiance remain. To address bad samples resulting from natural degradation, we've implemented a dropout layer between the initial and the first intermediate MLP layer, thus enhancing the model's generalizability with the remaining spectral inputs.

After these modifications to the model, the noticeable biases previously observed in plume detection of figures 13 a) and 13 c) have disappeared (as depicted in Fig. 2). This improvement emphasizes the crucial role played by the bad sample filter, which effectively excludes pixels that would otherwise be replaced by interpolation due to poor quality. This exclusion ensures a more accurate provision of data regarding the relationship between radiance and atmospheric parameters. Additionally, important statistical information regarding plume comparisons has been incorporated in the latest manuscript, in line with your valuable suggestions and comments.

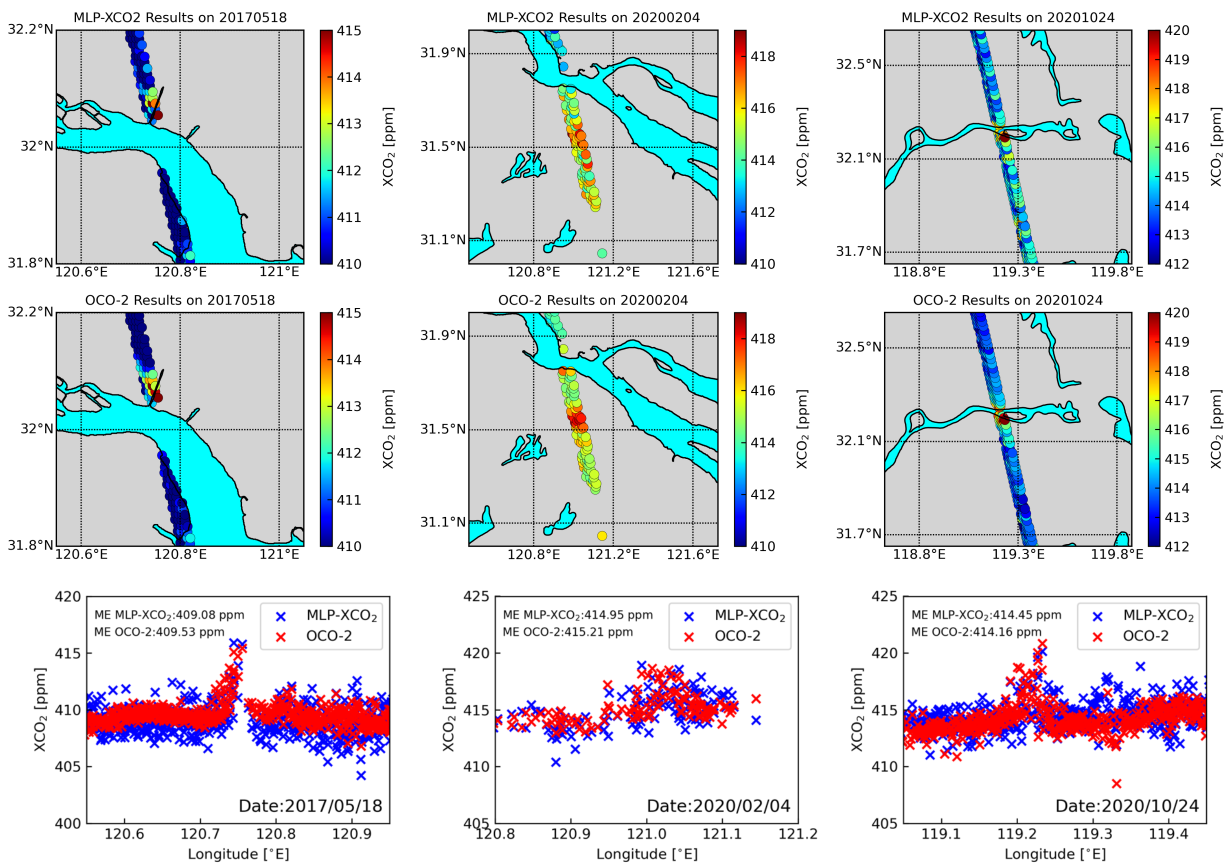

Figure 2: Comparisons of XCO_2 within CO_2 plume regions.

[2] You state in L 181 that your updated model now includes "year" as an additional feature due to changes in the OCO-2 instrument over longer time frames. What makes

you believe that the "year" feature addresses these changes? Would you expect steps in the predictions when switching from one year to another? Are you aware of the operational retrieval having to include such time dependent components? How does the model learn the long-term instrument trends when your observations come only from 2016? Since CO_2 is following a pattern in time, explicitly including temporal information might lead to a model that does not retrieve XCO_2 but simply finds statistical relationships between latitude, longitude, and time. That was one of my major points in the first review "Most importantly I am concerned that the model might get the right result for the wrong reasons and uses some of the parameters like solar zenith and azimuth angle to estimate the location of a given OCO-2 observation, rather than using the information contained in the measured spectrum." Now you added the "year" feature which moves the model further away from retrieving XCO_2 and more towards interpolating CO_2 over space and time.

Our updated model architecture now includes a bad sample filter and a dropout layer which are specifically designed to mitigate the effects of long-term instrument degradation, as outlined in response to the previous question and Section 2.2 of the updated manuscript. As for the "year" feature, it's been demonstrated that in traditional inversion algorithms, a lack of good prior information leads to bad CO_2 inversion results. For instance, as outlined in Figs. 3-4 of Ref. [2], if the prior profile for CO_2 is insufficient (i.e., too low) , the inversion process can lead to minimal or even opposite updates in the stratospheric CO_2 profile, while the tropospheric CO_2 profile near the surface tends to be overestimated to match the satellite spectrum. This discrepancy isn't due to flaws in the inversion algorithm but rather to the limitations of the satellite's signal-to-noise ratio and the radiative properties of the SWIR band, as indicated by the averaging kernel used in the OCO-2 retrieval algorithm [3]. It's crucial to recognize the necessity of certain good prior information.

Regarding your concern about the model potentially focusing solely on interpolation rather than learning actual CO_2 increases within spectra, and thus compromising its ability to detect plumes, we understand your concern. In our last response to reviewers, we demonstrated that relying solely on non-spectral data inputs results in very poor predictions. Therefore, the model is not performing spatial and temporal interpolation; the spectral information indeed plays a role in the retrieval process. Here, we believe that incorporating the "year" as a very conservative and simple way of providing CO_2 prior information, only offers contextual information for XCO_2 retrievals. Seasonal variations in XCO_2 in East Asia, for instance, can exceed about 10 ppm, but all within the same "year" input; which also indicates that the model does not perform the spatial and temporal interpolation.

The explanation above outlines why we've chosen to include the "year" as one of our inputs. While our understanding might not be perfect, adding this basic "year" input can noticeably enhance the retrieval outcomes. In our revised manuscript, we offer explanations based on our current understanding while recognizing the potential limitations of these explanations.

[3] In this revision you introduce an additional model called MLP-P. You state that the operational retrieval extracts its surface pressure information from O2-A band. However, your approach uses all features except the O2-A band. Where does the surface pressure information come from? Why was the additional model necessary? (The surface pressure in the L2MET files does not come from OCO-2, but from a reanalysis product.)

Incorporating surface pressure is critical for accurately interpreting the atmospheric distribution of XCO_2 in dynamically changing environments. Without these parameters, the model might incorrectly attribute changes due to factors like optical path length or CO_2 concentrations. This misattribution could lead to significant errors in XCO_2 estimation.

Furthermore, in regions with significant emissions of high-temperature water vapor and CO_2, which lead to enhanced plumes, the number of dry air molecules near the surface and the atmospheric pressure change significantly. Without including parameters such as surface pressure, the model would not be able to accurately capture the true XCO_2 levels in these dynamically changing environments. The subplots in the third row of Fig. 2 and the second row of Fig. 3 demonstrate that including surface pressure

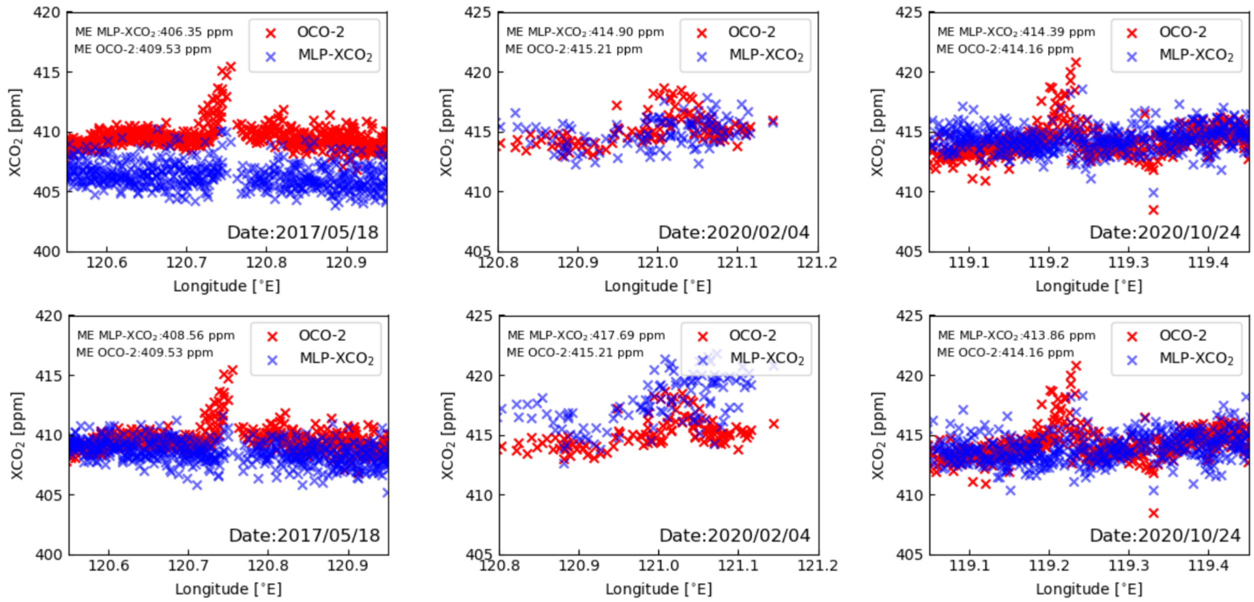

Figure 3: Comparisons of XCO_2 within CO_2 plume regions are presented. In the first row, model outputs are shown after removing both the bad sample filter and surface pressure. In the second row, model outputs are displayed after removing only surface pressure. All other model setting parameters remain consistent with those in Fig. 2.

substantially improves the effectiveness of near-surface plume detection.

Thank you for your insightful comments regarding the integration of non-spectral parameters of surface pressure into our MLP-XCO_2 model. We recognize the significance of surface pressure in accurately retrieving XCO_2, particularly for enhanced XCO_2 plumes. However, introducing an additional MLP-P model may confuse readers. Therefore, in our latest manuscript, we have decided to use surface pressure data retrieved from the OCO-2 L2std database for our mode instead of relying on an extral MLP-P model, and we have made that clear in the updated manuscript.

References

[1] Y. Marchetti, R. Rosenberg, D. Crisp, Classification of anomalous pixels in the focal plane arrays of orbiting carbon observatory-2 and-3 via machine learning, Remote Sensing 11 (24) (2019) 2901.

[2] C. Iwasaki, R. Imasu, A. Bril, S. Oshchepkov, Y. Yoshida, T. Yokota, V. Zakharov,

K. Gribanov, N. Rokotyan, Optimization of the photon path length probability density function-simultaneous (PPDF-S) method and evaluation of CO_2 retrieval performance under dense aerosol conditions, Sensors 19 (5) (2019) 1262.

[3] A. Braverman, N. Cressie, E. Kang, M. Katzfuss, P. Ma, A. Michalak, H. Nguyen, T. Stough, V. Yadav, Fusion of AIRS and OCO-2 carbon dioxide data for mapping lower-atmospheric CO_2.