We thank the anonymous reviewer for taking the time to review this manuscript and providing helpful feedback. Your advice has been very helpful and lead to significant changes to the manuscript that improved the overall quality.

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

Reviewer 1:

The authors develop a bias correction algorithm for the OCO-2 XCO_2 product using a (non-linear) random forest model with the aim to correct 3D cloud effects. The model is trained with bias derived from a "small area analysis" assuming no variation of XCO_2 at local scale (< 100 km). Variables from the OCO-2 CO2 dataset and dedicated cloud features are considered as model features. A feature selection is conducted to identify 3 and 4 features for glint and nadir, respectively, that are used for training the final model. The dedicated cloud features are excluded at this step. The analysis of the model shows a reduction of biases in the validation dataset and compared to TCCON measurements.

The manuscript is well written and the topic is within the scope of AMT. A correction of 3D cloud effects in XCO_2 retrievals would be an important scientific progress, because it would increase the number of observations with good quality. Thus, it would make CO_2 satellites better suited for studying anthropogenic and natural sources. However, I am skeptical that the presented machine-learning model will correct only non-physical variability due to cloud effects. Based on the current description of the method, I suspect that the model also wrongly corrects true variability in XCO_2, for example, from anthropogenic and natural sources (megacities, power plants, wildfires, etc.). My major concerns are the following:

Data filtering: The model is currently trained with QF=0 and QF=1 retrievals. The latter also includes retrievals where the quality is poor, for reasons other than (3D) cloud effects (e.g., high SZA, AOD, snow cover), which affects training, analysis and validation, because correlation coefficient and standard deviation are sensitive to outliers. I therefore think it is necessary to still apply some filtering to the QF=1 retrievals to remove poor-quality retrievals due to non-cloud effects.

We fully agree and added a step in our data preparation pipeline where we remove outliers from the dataset as described below:

L108 "Next, we remove outliers with large X_{CO2} errors 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%) and are not comparable to the quality flags used by OCO-2.

Table 1: Variables and their thresholds used to remove outliers

Variable	*Description*	*Land*	*Sea*
co2_ratio	*Ratio of retrieved X_{CO2} in WCO2 and SCO2 bands*	*x < 1 or x > 1.04*	*x < 1 or x > 1.03*
co2_grad_del	*Change between the retrieved CO_2 profile and the a priori profile*	*x < -100 or x > 100*	*x < -50 or x > 100*

deltaT	*Retrieved offset to a priori temperature profile*		*x < 0*
dpfrac	*Retrieved XCO2 multiplied by difference in retrieved and a priori surface pressure (Kiel et al., 2019)*	*x > 7*	
rms_rel_sco2	*Root Mean Squared error of the L2 fit residuals for the SCO2 band, relative to the continuum signal*		*x > 0.5*
snr_sco2	*Signal-to-noise ratio in SCO2 band*		*x < 200*

"

Truth metric: The generation of the truth metric is described insufficiently. Apparently, a k-mean algorithm is used to divide OCO-2 orbits into small areas where XCO2 does not vary strongly. The true bias is then defined as the deviation of the XCO2 retrieval from the median in this area. I would like to see some examples for some orbits to better judge how well this method works. In particular, it is unclear what happens in proximity of CO2 sources (megacities and power plants) where XCO2 deviates from the mean due to local emissions. I assume that a filtering algorithm needs to be applied to remove these areas to avoid false (positive) biases in the training dataset, but such filtering is not mentioned in the manuscript.

Model validation: The validation needs to be able show that true XCO_2 enhancements are not wrongly corrected by the model. Since TCCON is not well suited for this task as the instrument are generally not located downstream of a source, I suggest conducting some case studies near known CO_2 sources to show the effect of the bias correction in OCO-2 data. There is a large amount of literature on the use of OCO-2 to estimate power plant emissions with suitable cases (e.g., Nassar et al. 2017, Hakkarainen et al 2021). The validation should also be done for the B10-cloud model.

We simplified the generation of the small areas to enhance transparency of how they are generated:

L100 "To exploit this constraint on X_{CO2} 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 (ordered by observations 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."

Power plant sources are indeed interpreted as positive biases in XCO2. However, they are rare and we assume that they don't affect the model significantly. We added this disclaimer to the description of the small areas and added an analysis of three overpasses over power plants to the manuscript to back up our claim:

L106 "Additionally, this processing will interpret real X_{CO2} enhancements, for example from power plants, as positive biases. However, we postulate that these cases are rare and that a model that is robust to outliers can still learn a useful bias correction from these data."

*L439 **"5.3 Effect of bias correction on true CO_2 Enhancements***

As discussed in section 3.1 we use the small areas analysis as a truth proxy to develop our model. This assumes that CO_2 is well mixed and constant over short spatial scales (<100 km). However, this assumption is violated for strong CO_2 emitters such as power plants. Even though these strong emitters *are rare in the data and likely don't influence the bias correction model, there is a risk that the model would "correct", i.e. remove real local CO_2 enhancements. To confirm that real CO_2 enhancements are still present after the proposed bias correction, we compare OCO-2 retrieved and corrected X_{CO2} from three OCO-2 overpasses over large coal power plants (see Figure 11), that have been used in a previous study (Nassar et al., 2017). The CO_2 enhancements of the retrieved and corrected X_{CO2} for the three overpasses (the singular spikes in X_{CO2} in the middle of the graphs) agree closely and demonstrates that the bias correction does not erroneously remove true CO_2 enhancements from the OCO-2 data record.*

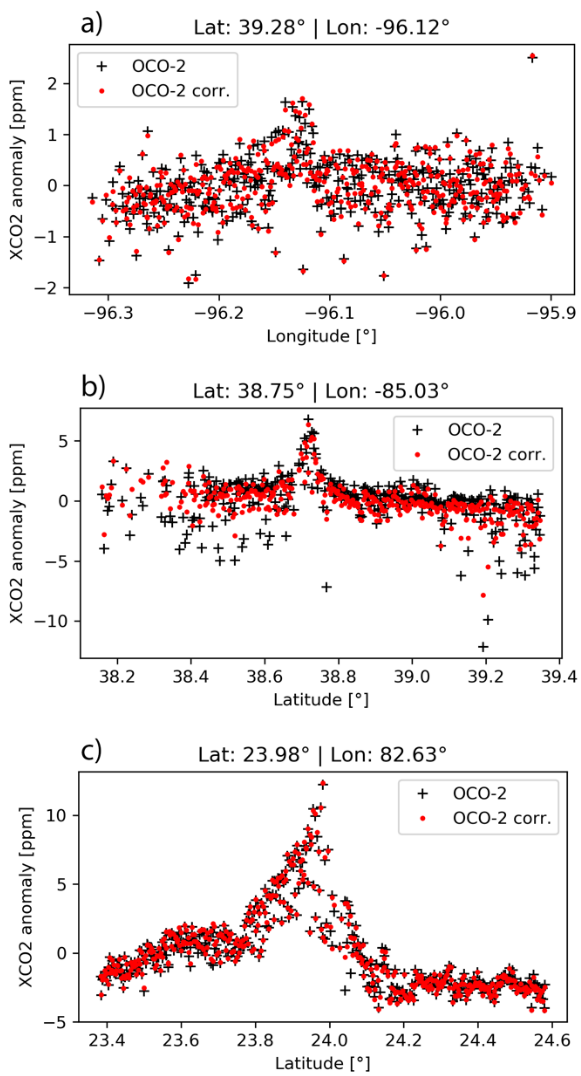

Figure 11: XCO2 anomalies for OCO-2 and bias corrected OCO-2 retrievals in the proximity of coal power plants. Power plant a) Westar at Lat: 39.28° Lon: -96.12° on 12/04/2015, b) Ghent at Lat: 38.75° Lon: -85.03° on 08/13/2015, c) Sasan at Lat: 23.98° Lon: -82.63° on 10/23/2014. Anomaly is calculated by subtracting the average. "

Feature selection: The feature selection method ignores that some variables could correlate with the "truth metric", which is computed from the same dataset and might have some issues (see previous point). These variables cannot be used in the model. In particular, the presented model uses "xco2_strong_idp" as feature, which is XCO_2 retrieved from the IMAP-DOAS in pre-processing "normalized by subtracting the mean of each small area" (L155). This is extremely similar to the XCO_2 bias used for training, i.e. the difference between the XCO_2 retrieved from ACOS and the mean of each small area (L77). As a result, I strongly suspect that the bias correction correct not only cloud biases, but also any deviation from the local mean including enhancements due local sources (e.g. megacities and power plants). Applying this bias correction model to the OCO-2 CO_2product, would make it impossible to estimate accurate CO_2 emissions from OCO-2 observations. To avoid this issue, the features used in the model need to be selected based on their correlation with cloud properties. The B10-cloud model shown in Section 4.3 is likely a good choice. It could be given more emphasis in the manuscript.

We agree that XCO2_strong_idp should be correlated with true changes in X_{CO2} and removed the variables as a feature for our bias correction. Thanks for catching that. However, the other selected variables from the state vector should not be correlated with true variations in XCO2 over spatial scales of less than 100 km. Therefore, we did not remove those features. A bias correction based only on cloud properties has been the topic of a previous study by Massie et al. (2021), thus, we don't go into too much detail in this manuscript.

Specific comments

L17: Not clear if you find the bias in the XCO_2 product with or without (cloud) bias correction.

Clarified sentence. It now reads:

L18 "Overall, we find that the published OCO-2 data record underestimates X_{CO2} ... "

L33: The effect of 3D cloud effects on TROPOMI NO_2 was recently studied: Emde et al. 2022, Yu et al. 2021, Kylling et al. 2022.

L38 Added citation for the three papers

L62ff: It would be nice to provide some more details on the 3D cloud effect features, so the reader does not need to check the cited citations.

Added additional information:

L82 "Finally, we make use of four variables indicative of 3D cloud effects (Massie et al . 2021): H3D, HC, CSNoiseRatio, and Cloud Distance. H3D (Liang et al., 2009; Massie et al., 2017) describes the normalized standard deviation of the MODIS radiance field, and is calculated based on the Cronk (2022) off-line MODIS radiance data files. The radiance standard deviation is calculated in a circle of radius 10 km surrounding each OCO-2 data point. HC is calculated from differences in O2 A-band continuum radiances of an observation point and adjacent points in three rows (frames) of footprints. A frame has eight adjacent OCO-2 footprints, with each footprint on the order of 2 km in size. CSNoiseRatio is the ratio of the O2 A-band continuum radiance spatial standard deviation and noise level, calculated within a footprint (which has 20 "colorslice, CS" sub-pixel elements). These three variables are indicative of 3D cloud effects since radiance gradients are present when clouds are next to observation footprints (radiance enhancements become larger as cloud distance decreases). Cloud Distance (Massie et al., 2021) is the distance of the nearest cloud to each observation point, as determined from the Cronk (2022) off-line radiance data files, *which contain 500 m MODIS radiances, geolocation and cloud mask data. Calculated 3D cloud features can be found for OCO-2 from September 2014 to July 2019 at https://doi.org/10.5281/zenodo.4008764."*

L67: Since overpass time would matter for "CloudDistance", please specify if you use MODIS Aqua and/or Terra.

The MODIS Aqua data used to determine CloudDistance is acquired six minutes after the OCO-2 measurements.

L95: Please provide reasoning why and how 3D cloud effects cause negative biases.

Line 386 "In particular the 3D cloud effect enhances, or brightens, the radiances as compared to no clouds being present. To compensate for this brightening the forward model decreases the retrieved surface pressure (reduction in dp_abp), increases the optical depth of cloud water (aod_water) and increases the surface albedo in the WCO_2 band. These relationships are shown empirically in Figure 7 As shown in Fig. 2 of Massie et al. (2021), the spectral signature of the 3D cloud effect (the optical depth structure of the radiative perturbation of the 3D effect) differs from the spectral signatures of perturbations in surface pressure, surface reflectivity, aerosol, and X_{CO2}. Fig. 2 illustrates that a decrease in surface pressure and X_{CO2}, and an increase in surface reflectance will increase the observed radiance. In order to provide for extra radiance enhancement in the cloud brightened observed radiance, a variety of state variable adjustments (and their unique spectral contributions) are utilized by the retrieval to bring forward model radiances in agreement with the observed radiances. The relationship of 3D cloud biases to surface pressure differences and surface albedo are likely due to a combination of physically-based 3D cloud radiative effects and operational retrieval algorithmic considerations. "

L101: Since OCO-2 might drift in time, please check if splitting by time affects your conclusions.

A possible drift of OCO-2 or nature (climate change, XCO2 increasing steadily, ...) is exactly why a train test split by time is so important. Such a drift would negatively affect the accuracy and validity of our model. Thus, by using soundings from one time period to train the model and then a later time period to test the model we get an honest assessment how the model generalizes to new data (including any instrument drift).

L143: Since the correlation coefficient is not sensitive to a bias in your model, it would be useful to use also other parameters (e.g. RMSE).

Interesting idea. We repeated the recursive feature elimination with RMSE and got the same ordering of the most important features. This is not surprising since the random forest algorithm internally uses the mean square error to construct the individual trees.

L180ff: The results here depend strongly on the definition of the truth metric, which might not contain only "non-physical variability" (see general point).

That is a valid caveat of our analysis. We already have a disclaimer in In Section 3.1 where we note:

L98 "To develop the bias correction model, we use the 'small areas analysis', which is based on the assumption that CO_2 is a well-mixed gas and assumed to be constant over spatial scales of less than ~100 km ***(there can be exceptions for strong CO_2 emitters such as mega cities)."***
Added disclaimer to 4.1 L245: "Figure 4 compares remaining X_{CO2} biases in OCO-2 ***(as determined by the small areas analysis)*** *with biases after our correction is applied (OCO-2 corr.)".*

L299ff: The analysis here assumes that the truth metric is caused primarily caused by 3D cloud effect, which is likely a wrong assumption (see general points). I think that it is necessary that you describe quite clearly here why a feature would be effect by 3D cloud effects, e.g., why ACOS retrieves a wrong surface pressure (dp) in the proximity of clouds.

It is correct that separating biases caused by 3D cloud effects and other retrieval biases (e.g. due to aerosols) is challenging. Our plots in Figure 5 indicate that at a minimum we remove systematic negative biases in XCO2 in the proximity of clouds which indicates that we correct for 3D cloud effects. More theoretical work has been done by Massie et al. (2021) to show that these biases stem indeed from 3D cloud effects and not, for example, cloud shadows or other effects. Based on this work we added some information of how the individual variables relate to 3D cloud biases.

Line 383 "To understand why some variables of the OCO-2 retrieved state vector are correlated with 3D cloud biases it is important to remember that the operational retrieval, based on optimal estimation, tries to match the observed radiances with a forward radiative transfer model. However, while the observed radiances can be perturbed by 3D cloud effects, the forward model tries to match those radiances with an independent pixel approximation that does not physically include 3D cloud effects. In particular the 3D cloud effect enhances, or brightens, the radiances as compared to no clouds being present. To compensate for this brightening the forward model decreases the retrieved surface pressure (reduction in dp_abp), increases the optical depth of cloud water (aod_water) and increases the surface albedo in the WCO$_2$ *band. These relationships are shown empirically in Figure 7. As shown in Fig. 2 of Massie et al. (2021), the spectral signature of the 3D cloud effect (the optical depth structure of the radiative perturbation of the 3D effect) differs from the spectral signatures of perturbations in surface pressure, surface reflectivity, aerosol, and* X_{CO2}*. Fig. 2 illustrates that a decrease in surface pressure and* X_{CO2}*, and an increase in surface reflectance will increase the observed radiance. In order to provide for extra radiance enhancement in the cloud brightened observed radiance, a variety of state variable adjustments (and their unique spectral contributions) are utilized by the retrieval to bring forward model radiances in agreement with the observed radiances. The relationship of 3D cloud biases to surface pressure differences and surface albedo are likely due to a combination of physically-based 3D cloud radiative effects and operational retrieval algorithmic considerations. "*

Technical corrections

L11 (and others): CO2 -> CO_2

Corrected

L26: fraction -> fractions

Corrected

L33 (and others): The citations style does not follow AMT requirements (here: "Massie et al. 2017")

Updated citation style throughout the manuscript

L118: form -> from

Corrected

References

- Emde et al. (2022) https://doi.org/10.5194/amt-15-1587-2022
- Hakkarainen et al. (2021) https://doi.org/10.1016/j.aeaoa.2021.100110
- Kylling et al. (2022) https://doi.org/10.5194/amt-15-3481-2022
- Nassar et al. (2017) https://doi.org/10.1002/2017GL074702
- Yu et al. (2021) https://doi.org/10.5194/amt-2021-338