First of all, we thank reviewer 2 for his/her effort in carefully reviewing our manuscript and his/her constructive comments.

Point-by-point answers to the comments of reviewer 2

General comments

Reviewer 2: *For example, to retrieve the XCO2 and XCH4, the instrumental model is very important. In this manuscript, only the random noise is assessed. The authors should concern the other parameters at least the uncertainly of instrumental line shape function and its wavelength depended response. In addition, the authors were used the actual space-based observation data such as OCO-2 during the FOCAL development. To evaluate the new NRG-CO2M algorithm with actual space-based observation data with realistic uncertainty is also important and informative. However, the authors are only focused the simulation-based dataset. I understand the CO2M will not be launched until 2026. The authors should be considered the evaluation plan with the updated instrumental model data and the realistic characterization error, and these impact on the NRG-CO2M processing. Furthermore, the application for the actual space-based observation dataset, currently available dataset, is also informative and productive for the evaluation purpose. The authors should be considered the evaluation plan for the NRG-CO2M with currently available observation dataset. I recommend the authors will add the sentences and clarify for some of unclear sentences. For these reasons, I recommend this paper for publication with minor changes to the technical content.*
Authors: The reviewer raises two general points related to the instrumental model used and the fact that our study is based only on simulations and not on actual measurements from existing satellite instruments. Since both points are also raised in the section “specific comments”, we'll address them in that section.

Specific comments

Reviewer 2: *Page 1, line 13: Spell out first for “NRG-CO2M”. -> Neural networks for Remote sensing of Greenhouse gases from CO2M (NRG-CO2M)*
Authors: Done.

Reviewer 2: *Page 1, line 19: The definition of “spatio-temporal systematic errors” is unclear. The authors should add the definition or more clear explanation for the condition.*
Authors: We now define the term "spatio-temporal systematic errors" earlier in the abstract: "According to the CO2M mission requirements, the spatial and temporal variability of the systematic errors (or spatio-temporal systematic

errors) of XCO2 and XCH4 ...".

Reviewer 2: *Page 2, line 39: add the "," between "5ppb" and "respectively".*
Authors: Done.

Reviewer 2: *Page 2, line 41: Spell out first for "CO2I".*
Authors: Done.

Reviewer 2: *Page 2, line 42: Spell out first for "MAP".*
Authors: Done.

Reviewer 2: *Page 2, line 42: Spell out first for "BRDF".*
Authors: Done.

Reviewer 2: *Page 2, line 43: Spell out first for "CLIM".*
Authors: Done.

Reviewer 2: *Page 2, line 44: XCO2 or XCH4 -> XCO2 and/or XCH4*
Authors: Done.

Reviewer 2: *Page 2, line 47: 2017b,a -> 2017 a, b*
Authors: Done.

Reviewer 2: *Page 2, line 49: Spell out "EUMETSAT".*
Authors: According to the AMT author guidelines, abbreviations that are better known than their full form need not be defined, which is the case here.

Reviewer 2: *Page 2, line 57: The meaning of "3D effects" is unclear. The authors should add the explanation.*
Authors: The radiative transfer (RT) models used in atmospheric greenhouse gas retrievals are so-called 1D RT models, because they consider changes of the atmospheric properties only in one dimension. I.e., all properties change only with height. As a consequence, photon transport between neighboring columns with different properties is not possible so that atmospheric columns can be considered independent. However, in reality such photon transport happens which results in inaccuracies of 1D RT models. Especially near cloud edges these inaccuracies can become important. RT-models that are able to account for atmospheres with varying properties in three dimensions are called 3D RT models. As they are usually computationally more expensive and as 3D properties of the atmosphere are often not known, they are not used in operational satellite greenhouse gas retrievals. Whenever limiting changes in the atmospheric properties to 1D results in inaccuracies, we speak of 3D effects.

Since we use the term "3D effect" only as a keyword for further reading and as one of several examples, and since the term is common in the context of RT modeling, we would like to avoid a more detailed description in the paper.

Otherwise, we could also remove the term.

Reviewer 2: *Page 3, line 61: Spell out first for "OCO-2".*
Authors: Done.

Reviewer 2: *Page 3, line 62: Spell out first for "GOSAT".*
Authors: Done.

Reviewer 2: *Page 3, line 77: the meaning of "meteorology and angles" are unclear. The authors should add the explanation.*
Authors: We rephrased to " meteorological profiles, observation angles". In the context of page 3, line 77, these are only unspecific examples. A detailed description of the input features to the MLPs trained by us is given later in the paper.

Reviewer 2: *Page 3, line 83: Krasnopolsky and Schiller (2003). -> (Krasnopolsky and Schiller, 2003).*
Authors: Done.

Reviewer 2: *Page 4, line 116: Spell out first for "OSSE".*
Authors: Starting on P4 L115, the manuscript reads: "... is based on simulated measurements from an extensive observing system simulation experiment (OSSE), which is a refinement of ..."

Reviewer 2: *Page 4, line 116: In the previous works, the authors were developed FOCAL full physics algorithm. During the development phase of FOCAL, the authors are actually used the space-based observation data such as OCO-2 and GOSAT. To evaluate the new NRG-CO2M algorithm with actual space-based observation data is quite realistic and import. However, the authors are only focused the simulation-based dataset. So, the authors should be considered the evaluation plan with actual space-based observation dataset or current limitations.*
Authors: We agree with the reviewer that the application to real data from OCO-2, GOSAT, or GOSAT-2 could provide interesting additional results. However, due to the complexity of the work required, this would change the FOCUS of the paper significantly which is to use simulations to develop a technique to modify measured spectra in such a way that they can be used as representative training data for hitherto unprecedented atmospheric conditions and that MLP-based methods are able to fulfill the CO2M mission requirements under suitable conditions. In this respect, the results shown should rather be interpreted as a proof-of-concept (see also reviewer 1). Accordingly, we discuss in the revised manuscript: "In the analysis of real data, several effects, the detailed investigation of which is beyond the scope of this paper, may lead to somewhat degraded retrieval quality. These include unknown systematic errors in the training truth, a priori, and met profiles, non-ideal sampling of the training data set, and potential instrument or RT features that are not well

approximated by our spectrum modification method. Therefore, the actual retrieval quality achievable can only be determined after NRG-CO2M has been trained on and applied to real data." We would also like to point out that due to the differences between the OCO-2 and CO2M instruments, perfect transferability of the results would not be guaranteed.

Reviewer 2: *Page 5, line 154: What is the instrumental line shape model? It also has several uncertainties. It is not clear how to take account spectrally depended uncertainties. The authors should add the explanation.*
Authors: As our OSSE setup including the instrument model is widely adapted from Noël et al. (2024), we only briefly describe the setup. However, we now added to Sect. 2.1: "The simulated main instrument CO2I consists of four imaging spectrometers for the wavelength ranges 405 nm–490 nm (VIS, NO_2), 747 nm–773 nm (NIR, O_2), 1590 nm–1675 nm (SWIR-1, CO_2 and CH_4) and 1990 nm–2095 nm (SWIR-2, CO_2) having spectral resolutions of 0.6 nm, 0.12 nm, 0.3 nm and 0.35 nm, respectively. In line with currently available information about CO2I, the instrument line shape functions are assumed to be Gaussian with full width at half maximum, corresponding to the respective spectral resolution."

Reviewer 2: *Page 7, line 193: How to consider the bias in a priori? Especially in the future prediction, not only a standard deviation but also the bias has to be considered. The authors should add the explanation.*
Authors: Biases in the a priori or the training truth have the potential to introduce biases in the prediction. In case of the a priori this is usually less problematic because the influence of the a priori is reduced when applying the averaging kernels when using the prediction for emission estimation. However, systematic errors in the training truth bear the risk that incorrect relationships are learned, which is particularly possible if biases in the training truth correlate with input features. (e.g. systematically too high CO2 concentrations at high latitudes, or over bright surfaces). Unfortunately, reliable information on such biases and their covariance statistics do not exist which is why we have not considered them and assumed Gaussian noise for convenience. At least our results become better comparable to those of Noël et al. (2024) who also used an unbiased a prior and an unbiased training truth for their machine learning based post processing bias correction. In order to make the reader aware of this point, we discuss in the introduction: "Obviously, such errors would have the potential to reduce the accuracy of the prediction, but a realistic estimate of the to be expected error patterns of the training truth is difficult and beyond the scope of this study."

Reviewer 2: *Page 18, Figure 5: How is the slope? It seems that the linearity can be directly estimated from this analysis. However, it is not mentioned in the text.*
Authors: We updated Fig. 5 of the manuscript (Fig. 1 in this document) which now also includes the results of a linear regression.

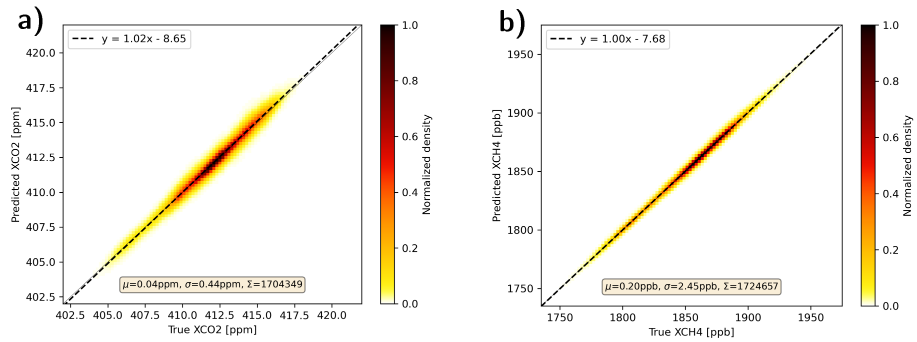

Figure 1: Comparison of postprocessed predicted XCO2 (a) and XCH4 (b) with corresponding true values for noise-free 2020 subset input data. Δ represents the average prediction error (prediction minus true), σ the standard deviation of the prediction error, and Σ the total number of soundings. The figure also contains the results of a linear regression.

References

Noël, S., Buchwitz, M., Hilker, M., Reuter, M., Weimer, M., Bovensmann, H., Burrows, J. P., Bösch, H., and Lang, R.: Greenhouse gas retrievals for the CO2M mission using the FOCAL method: first performance estimates, Atmospheric Measurement Techniques, 17, 2317–2334, https://doi.org/10.5194/amt-17-2317-2024, 2024.