Dear reviewer 1,
We are very grateful for your comments and suggestions, which have helped to improve our manuscript significantly. We have revised the manuscript accordingly. The following is a point to point response to your comments and suggestions. Corresponding changes in the manuscript are also made available below at the appropriate places, if applicable.

Sincerely,
Maximilian Rißmann and Jia Chen on behalf of all co-authors.

---

**There are a number of areas within the manuscript that refer to XCO2 correctly in ppm, but then also refer to this quantity as both a concentration and/or a mixing ratio in various places within the text. The unit of ppm is a mole fraction and does not include a volume quantity, so please be cautious and consistent with the unit of XCO2 (ppm) throughout the manuscript.**
*Thanks for this comment. We indeed have not been too consistent in the use of the term column averaged dry air mole fractions (DMF). We updated the manuscript accordingly.*

**In general, the manuscript is very well-written and concise, but there are a number of grammatical errors and typos within the manuscript that could be removed with more thorough editing. I have noted some corrections below in the technical comments.**
*We incorporated all the technical comments mentioned below. Thanks so much for taking the effort of compiling the list of corrections!*

**Figure 2: The units of XCO2 are missing on subplots. It also looks like the target observations overlap quite a bit – how do you deal with this in the comparison methods and does averaging overlapping soundings affect your additional quality flag results in Figure 3? Similarly, Figure 15 is also missing XCO2 units and overlapping observations make it difficult to compare to the model. Is it possible to average these observations in some way to better visualize this comparison?**
*Thank you very much for this suggestion. We updated all the OCO-2 target figures in our manuscript by averaging OCO-2 soundings into 0.02° x 0.02° bins.* $XCO_2$ *units are added to the plots. We do not introduce any weighting terms to deal with overlapping soundings in spaceborne* $XCO_2$ *retrievals. Each good-quality OCO-2 sounding contributes the same to the mean. Hence, the averaging of overlapping soundings also does not affect the additional quality flag results in Figure 3.*

**For the model comparison, it wasn't clear within the text whether or not the WRF XCO2 calculation takes into account the OCO-2 averaging kernel. Does it? Without doing so, the comparison between OCO-2 and the model is not a true 1:1 comparison.**
*Thanks for the suggestion. We performed an averaging kernel correction. Please see our answer in the comment referring lines 144-155.*

**L24: please replace "concentrations" with "mole fractions"**
*Done, thanks!*

**L31: Reference is not valid and is also not included in the References section**
*We updated the reference in the manuscript to "Montzka, S.: The NOAA Annual Greenhous Gas Index (AGGI), NOAA Global Monitoring Website, https://gml.noaa.gov/aggi/aggi.html, Last accessed on 07.12.2021, 2021"*

**L39: It would be worthwhile to state that the TCCON monitors the long-term atmospheric growth of total-column CO2, CO and CH4…**
*Thanks for this correction. We changed the according sentence in the manuscript as followed:*

| | |
|---|---|
| L39 | It monitors the long-term atmospheric growth of XCO2, XCO and $XCH_4$ along with other atmospheric trace gases. Regular calibrations against aircraft measurements make the TCCON |

| | |
|---|---|
| | stations currently the primary validation source for most space-based $XCO_2$ data products (GOSAT, GOSAT-2, OCO-3, TROPOMI). |

**L54: I think that the 14-day repeat cycle is incorrect, isn't it 16 days?**
*Correct, thanks. We updated the manuscript.*

| | |
|---|---|
| L54 | NASA's Orbiting Carbon Observatory instruments (OCO-2 and OCO-3) capture XCO2 in four different measurements modes: nadir, glint, target and snapshot area mode (SAM). OCO-2 captures XCO2 on a 16-day ground-track repeat cycle. |

**L66: Some care should be taken here to define the scale at which OCO-2 may be able to resolve XCO2 fluxes, which is dependent upon the spatial scale of the target itself.**
*We assume cities larger than Munich, which currently has a population of 1.48 million, to generally have larger $CO_2$ emissions and consequently larger spatial gradients. Thanks to your comment we now emphasize in the manuscript that the spatial scale is limited by the 15x20 km target field size of OCO-2. Due to the small size of Munichs inner city (319.71 square kilometers), OCO-2's target area covers a large part of the city. Complementarily, the bigger OCO-3 SAM field-of-view is better suited for emission studies in larger agglomerated areas. Our paper for the first time covers spatial gradients within one OCO-2 target observation, which is made possible by the proximity of the ground-based MUCCnet spectrometers. The relatively low gradients measured during the overpasses of our study enabled us to test the lower detection limits for intra-urban XCO2 fluxes of OCO-2.*

| | |
|---|---|
| L66 | This way, we test the capability of OCO-2 to resolve small-scale urban $XCO_2$ fluxes in Munich and other cities from space, which is needed to study sector dependent emissions in the future. Due to OCO-2's relatively small target size of around 300 square kilometers, the instrument is best suited for spaceborne emission studies in smaller cities while OCO-3's SAM measurements cover a wider field-of-view, which enables the assessment of large metropolises around the globe. |

**L106: Because it is not possible to truly calibrate XCO2, it is somewhat misleading to call this quantity XCO2 "calibrated". Rather, this is a bias-corrected and scaled XCO2 retrieval that is presented.**
*Appreciated. We updated our manuscript accordingly.*

| | |
|---|---|
| L108 | All results of this paper are based on the scaled and bias-corrected retrievals $XCO2_{corrected.}$ |

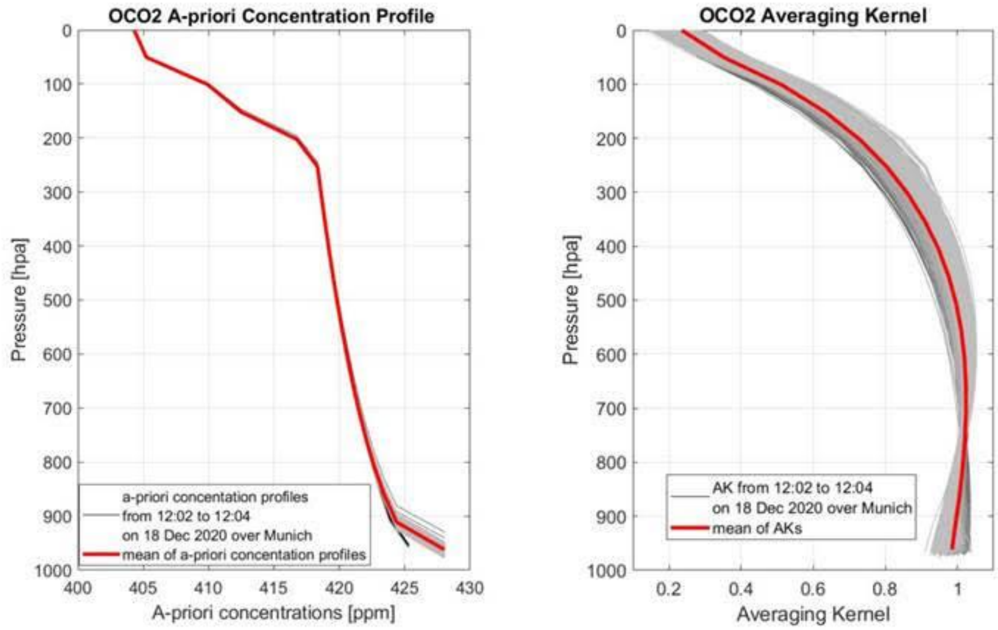
**L125: Please state to which CO2 scale these retrievals are tied for the reader.**
*The OCO-2 v10 data scaling is part of the OCO-2 bias-correction procedure that is applied to the XCO2 soundings in the OCO-2 lite files. The global scaling factor applied, ties the OCO-2 data to TCCON and consequently to the WMO X2007 trace gas scale.*

**Figure 3 caption: Should "spectrometer locations […]" be moved to Figure 2?**
*Done.*

**L138: Given the reference to Figure 2 and this statement, perhaps the order of Figure 2 and Figure 3 should be switched?**
*Thanks a lot for this suggestion. We are of the same opinion and reversed the order of Figure 2 and Figure 3.*

**L144-155: Does the model XCO2 calculation take into account the OCO-2 averaging kernel?**

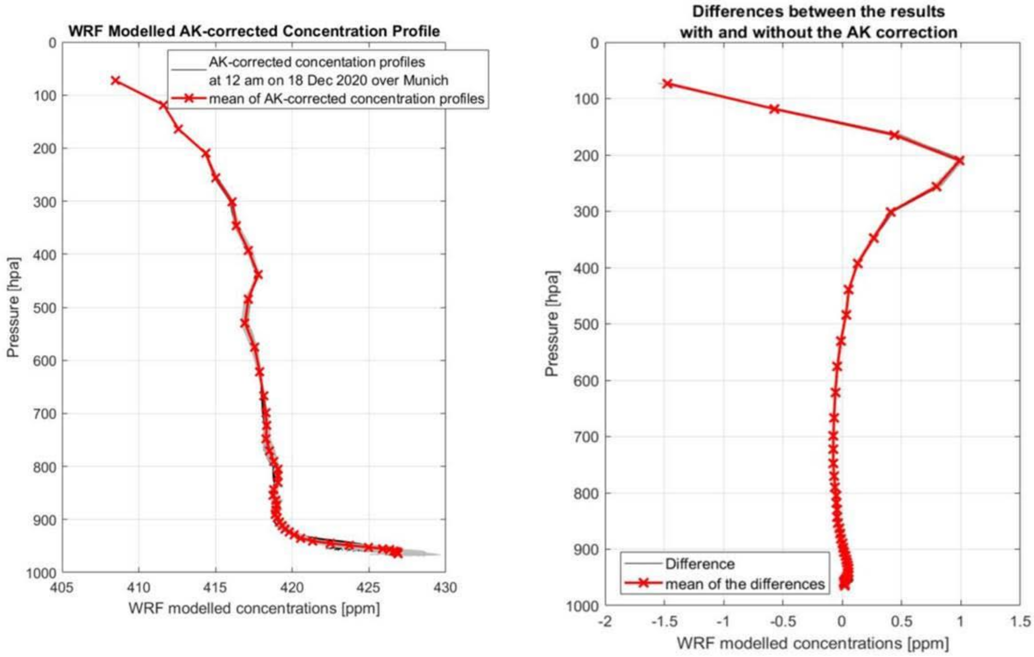
*We did not account for the difference in the averaging kernels initially. Thanks to your suggestion we have added the OCO-2 averaging kernel in the calculation of the model XCO2. We revised the simulation results, by using the OCO-2 column averaging kernel (AK, right) and the a-priori CO2 profile (left) provided by ACOS L2 Lite output.*




*We performed the averaging kernel correction following the method proposed by O'Dell et al. 2012 ([https://amt.copernicus.org/articles/5/99/2012/amt-5-99-2012.pdf](https://amt.copernicus.org/articles/5/99/2012/amt-5-99-2012.pdf)):*

| | |
|---|---|
| Eq. 3 | $C_{ak} = \boldsymbol{AK} * C_{raw} + (\boldsymbol{I} - \boldsymbol{AK}) * C_{apriori}$ |

*The plot below shows the AK-corrected WRF-profile and the difference between the uncorrected and corrected WRF-profile.*



**L172: Can you explain more how the collocation radius is chosen? Is this 6 km chosen to equally segment the target area around the EM27 sensors?**

*Correct, the 6 km radius was chosen, to segment the target area around the EM27 sensors as evenly as possible. It provides the largest number of retrievals within each of the three comparison domains within the target field of view while minimizing the overlap of our comparison domains as stated in lines 179-182.*

| | |
|---|---|
| L179 | For a collocation radius of $r_{col}$ = 6 km around the spectrometer locations we achieve the highest number of collocated soundings for each site while having almost no overlap of collocated soundings between the sites (most soundings are collocated to only one MUCCnet site). This way, we segment the target observation data into three comparison domains - centre, west and north. |

*A larger comparison set of soundings also reduces the effect of random errors in our computed mean XCO2. We assume this relatively large comparison domain to best represent the actual XCO2 around our ground-based measurement sites. We also added this explanation to the manuscript to further describe our thought process when choosing the comparison domains.*

| | |
|---|---|
| L182 | A large comparison set of soundings also reduces the effect of random errors in our computed mean XCO2. We assume this relatively large comparison domain to best represent the actual XCO2 around our ground-based measurement sites. |

**L192: Please include punctuation around equations here and elsewhere.**

*Implemented.*

**L226: Caution should be taken in stating that OCO-2 "measures" XCO2 mole fractions. Rather, OCO-2 measures a radiance that is converted to a mole fraction of XCO2 via a retrieval algorithm. In this sense, care should be taken to refer to XCO2 from both ground-based and spaceborne instrumentation as "retrievals" here, in Figure 8, and elsewhere throughout the manuscript.**

*Again, we want to excuse this oversight. We have modified the statement in Figure 8 and now use retrieve\retrievals throughout the manuscript.*

**L247: Other potential causes for differences in biases are explained in L254-257, so perhaps it's worth moving this discussion to this paragraph.**

*We moved the sentence to the paragraph starting in L254. Thanks for this suggestion. The paragraph now reads as follows:*

| | |
|---|---|
| L255 | The daily offsets in each domain are depicted in Fig. 10. We assume, measurement uncertainties and the relatively small sample size of eleven overpass days to cause the discrepancies in the computed mean biases of the three collocation domains. OCO-2 retrieves higher CO2 mole fractions than MUCCnet in all three domains, during each overpass except for August 12,2020. For most overpasses the by-site offsets are consistent in each of the three collocation areas. The largest discrepancies in daily offsets in the three domains could be observed on overpass days with a smaller than average number of good quality soundings (e.g. November 9, 2020 and July 27, 2020) |

**Figure 11: These error bars look very small given the error bars in OCO-2 XCO2 in Figure 9. Can you explain a bit more how you've calculated this error? In addition, given that the gradient in XCO2 is very small, it might be worth indicating minor x-axis grid lines here for the reader.**

*Thanks for this question. We calculated the standard error of the mean and introduced a weighting term (number of samples in the domain) to the error propagation of the standard because we wanted to set the focus on the robustness of the mean. Here we need to address, that we took a really simple approach to the error in the mean $X_{CO2}$ in the comparison domain. As another reviewer pointed out, we made the simplifying assumption, that the errors of XCO2 soundings to be randomly distributed within our comparison domain, which is not the case. We* *neglected systematic errors in the OCO-2 retrievals. To maintain consistency in our results (consistency with Figures 7 and 8) we removed the weighing term when computing the error for the XCO2 gradient assessment, to compute the combined standard deviation:*

| | |
|---|---|
| Eq. 6 | $sd_{domain1\text{-}domain2} = sqrt(\sigma_{domain1}^2 + \sigma_{domain2}^2)$. |

*We revised our representation of errors in the gradients and updated the plot accordingly.*
*This representation better shows the variance of single soundings within the comparison domains especially for the spaceborne retrievals. Even though the standard deviation is relatively large on some days, the overall RMSD of captured spaceborne gradients when compared to MUCCnet is 0.31 ppm, which indicates the instruments capability of resolving spaceborne gradients on a sub-city scale. We updated figure 11 as shown below.*

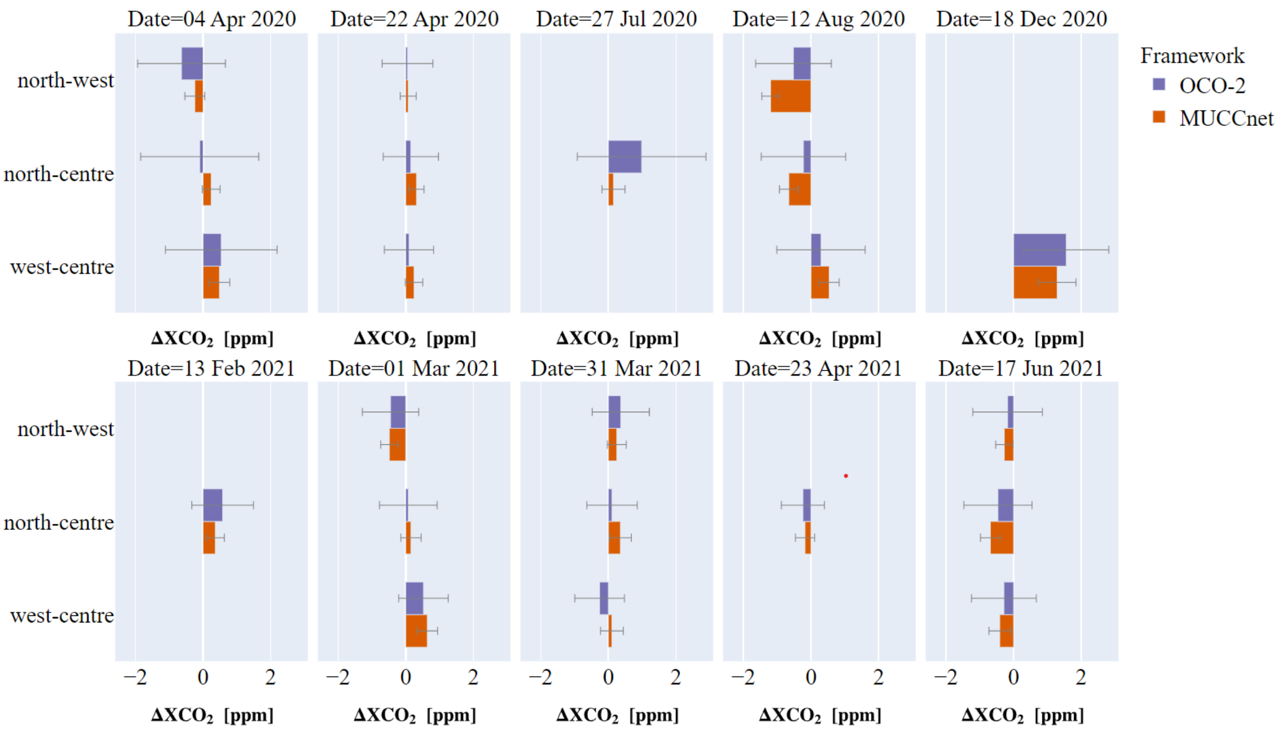



**Figure 11.** XCO2 gradients in Munich on overpass days. Blue bars represent the gradients present in the OCO-2 target observations. Orange bars denote XCO2 gradients captured by MUCCnet. On most days, OCO-2 sees elevated XCO2 in the same region as the ground-based MUCCnet instruments. Error bars are computed using the combined standard deviations of the XCO2 samples in the two domains which are used to compute gradients. (see Eq. 6).

**L299: Given that the gradients presented in Figure 11 do not have overlapping error bars and therefore, do not represent, qualitatively, similar mean XCO2 differences, would it be more accurate to state that OCO-2 is capable of detecting intra-target XCO2 gradients of a similar sign as MUCCnet XCO2?**
*You are correct. The non-overlapping error bars were caused by our previous, non-ideal representation error in the computed gradients. Normalizing the error in each domain with the number of soundings in that domain was not a realistic representation of the actual variance of the computed mean gradients.*
*We also noticed OCO-2's strength of replicating the actual sign (or direction) of the ground-based gradients, which we stated in two paragraphs of the manuscript.*

| | |
|---|---|
| L277 | Considering the rather small XCO2 gradients in Munich, OCO-2 detects the elevated XCO2 in the same domain as MUCCnet for 20 of the 22 computed gradients and therefore qualitatively determines the area of enhanced XCO2 correctly in 91% of cases. |
| Figure 11 | On most days, OCO-2 sees elevated XCO2 in the same region as the ground-based MUCCnet instruments. |

*Nonetheless, the overall RMSD of captured spaceborne gradients when compared to MUCCnet is around 0.31 ppm, which indicates the instrument's capability of capturing similar mean XCO2 gradients on a sub-city scale. This is also stated in the manuscript in line 291:*

| | |
|---|---|
| L291 | For the entire set of gradients OCO-2 achieved an RMSD of 0.31 ppm and a linear correlation with a strong correlation of $R^2$= 0.68 between OCO-2 and the MUCCnet measurements |

**L327: Again, here I am wondering how the XCO2 from the WRF model is computed because L144-155 does not describe whether or not this XCO2 quantity derived takes into account the OCO-2 averaging kernel.**
*We take the averaging kernels into account now. See our answers to your comments above.*

**L331: 'mixing ratios' are described in this same sentence in addition to mole fractions. Please use mole fractions and maintain consistency throughout the manuscript.**
*Done.*

**L344: It would be useful either here or in a previous section to describe what "good measurement conditions" entails.**
*We updated the sentence. Generally, our filter criteria are determined by the number of good-quality soundings OCO-2 was able to retrieve in each comparison domain. The number of good quality soundings is mostly determined by the cloud coverage during the overpass time. In this study we only consider overpass days on which OCO-2 retrieved a minimum of 500 good quality soundings. We cover the selection criteria for the gradient comparison in more details in Section 3.3 – Gradient Comparison.*

| | |
|---|---|
| L344 | *These results suggest that for high gradients and cloud free measurement conditions, OCO-2 target observations can be utilized for an accurate assessment of urban XCO2 and its spatial distribution.* |

.

**L345: Similar to L66, "middle" to "larger" sized cities might be irrelevant without a spatial scale. It would be helpful to state the spatial scale that you would expect OCO-2 to be able to resolve urban XCO2 gradients, given the swath area of the target.**
*This statement was also made under the assumption, that in "larger" more populated areas there are higher XCO2 emissions, that cause greater spatial XCO2 gradients during a measurement. It's right, that the study domain will always be constrained by OCO-2's limited target area. As mentioned above, larger areas then could be covered using OCO-3 SAM observations. We removed the vague statement from the manuscript.*

**Technical comments:**
**L10: Please change "constraint" to "constrained"**

| | |
|---|---|
| L10 | Due to this more constrained collocation, we observe improved agreement between space-borne and ground-based XCO2 in all three comparison domains. |

**L46-47: Consider rephrasing for readability**

| | |
|---|---|
| L46 | In recent years, EM27/SUN instruments have been used in measurement campaigns that aim to quantify urban anthropogenic emissions by combining differential column measurements (DCM) and atmospheric transport models. Multiple field campaigns have been carried out in Berlin, Munich, Indianapolis, San Francisco, California, Poland, St. Petersburg and Hamburg. These studies show the potential of top-down emission estimates as they can help uncover unknown emission sources and constrain bottom-up emission inventories. |

**L93: … determined “by” performing**

| | |
|---|---|
| L93 | The sensors are calibrated by subtracting constant offsets which are determined in side-by-side measurements. |

**L153: Please change “weighing” to “weighting”**

| | |
|---|---|
| L153 | XCO2 in the study area is derived from the modelled concentration profiles with an appropriate pressure weighting, following the method described in Zhao 2019. |

**L173: Please change “none” to “no”**

| | |
|---|---|
| L173 | For a collocation radius of $r_{col}$=6 km around the spectrometer locations we achieve the highest number of collocated soundings for each site while having almost no overlap of collocated soundings between the sites (most soundings are collocated to only one MUCCnet site). |

**L223: “Observing systems observe” is somewhat redundant.**

| | |
|---|---|
| L223 | Both observing systems capture a similar seasonal behaviour of urban XCO2 in Munich in the time period analyzed here |

**L240: Please replace “due to the ” with “by”**

| | |
|---|---|
| L240 | This improvement is caused by more specific collocation, that reduces the effect of averaging over potential spatial XCO2 gradients in the OCO-2 target observation. |

**L290: Please replace “shows” with “shown”**

| | |
|---|---|
| L290 | These improved results are shown … |