Dear Chris,

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.

The first major comment involves error bars. The error bars plotted on OCO2 biases (relative to MUCCnet), such as shown in Figures 7 & 8, appear to be standard deviation only. *IF* one were to assume that these errors were randomly distributed over the small area (<10 km) over which they are evaluated, we could estimate the standard error of the mean as sigma/sqrt(N), where N is the number of observations. However, it is well known that OCO-2 errors are NOT randomly distributed, over small or large areas (see e.g. Kulawik et al., 2019). It is likely that the local scale mean of OCO-2 includes systematic errors that cannot easily be evaluated. This point should clearly be made. Therefore, it is difficult to evaluate errors on mean OCO-2 values over the whole domain or the 3 sub-domains. Further, it is equally difficult to state them for the gradients. On this note, the error treatment on the gradients is both incorrect (as it assumes 1/sqrt(N) Gaussian averaging) and inconsistent with the stated error bars on XCO2 overpass means (which simply uses the standard deviation of the XCO2 values in a given domain). The current error bars listed on the gradients are unrealistically small (of order or less than 0.1 ppm). Therefore, please expand your discussion of errors to include these points, and be sure to treat errors consistent on the mean quantities and the gradient quantities.

Thank you very much for pointing this out. We indeed approached the errors of the computed XCO2 mean in a very simplistic way. We made the simplifying assumption, that the errors of XCO2 gradients are randomly distributed within our comparison domains, which is not the case. We neglected systematic errors in the OCO-2 retrievals. To maintain consistency in our results we removed the weighing term when computing the error of the X_{CO2} *gradients, to compute the combined standard deviation. We revised our representation of errors in the gradients and updated the plot accordingly to be consistent with the standard deviation of the* X_{CO2} *samples we show in Sections 4.1 & 4.2 (Figure 7 & 8).:*

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

We updated Figure 11 accordingly:

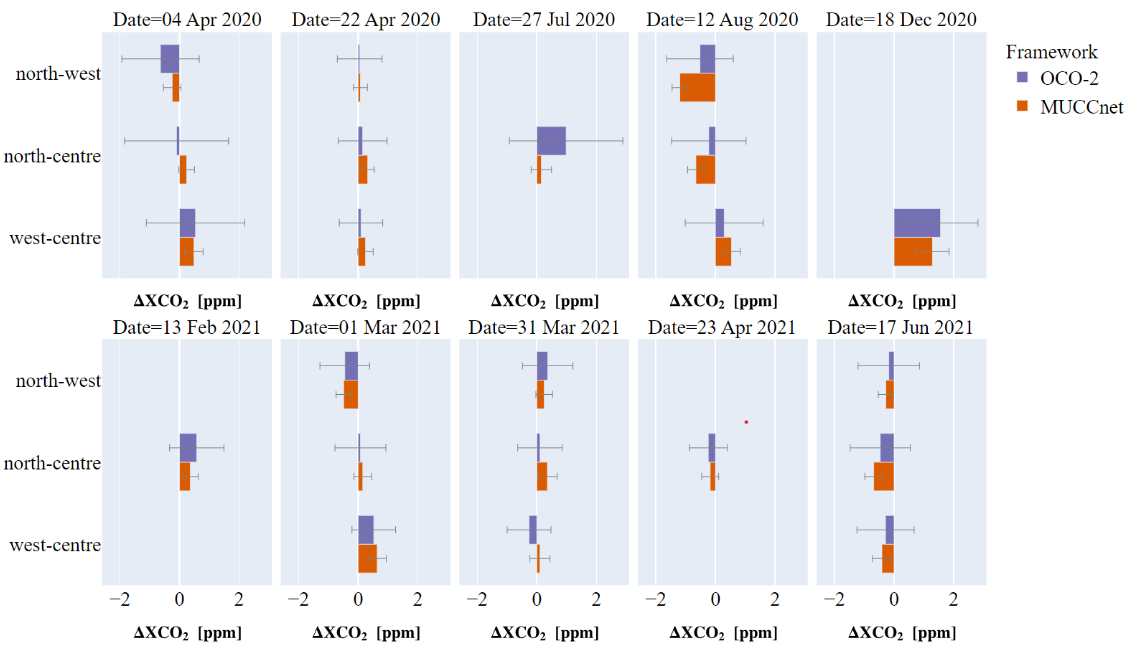


Figure 11. XCO2 gradients in Munich on overpass days. Blue bars represent the gradients present in the OCO-2 target observations. Orange bars denote XCO2 gradients captured by MUCCnet. On most days, OCO-2 sees elevated XCO2 in the same region as the ground-based MUCCnet instruments. Error bars are computed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

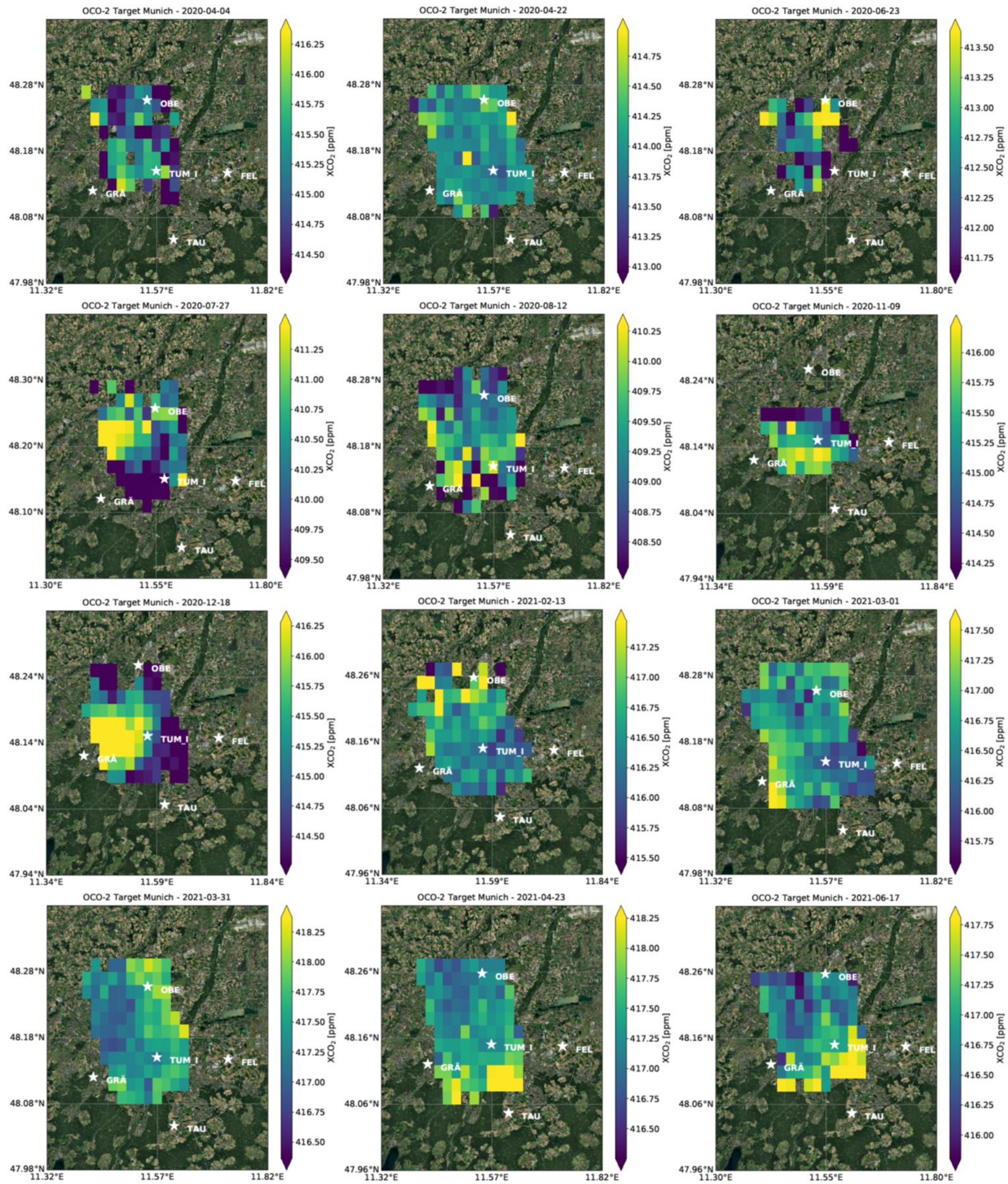
We further added a short paragraph to emphasize, that we are not representing a fully correct description of the error in the mean XCO2 gradients.

L215	Rather than the error of the mean we represent the combined spread of XCO2 in the two domains.

In case there is a better way of representing the error in the XCO2 gradients, we would greatly appreciate further suggestions and are happy to incorporate them into our manuscript.

Second comment (albeit minor) is how the data are presented on a map. Figure 2 shows this for the 12 dates analyzed. I suggest averaging the data onto a 0.5x0.5 km2 grid, because as you show it now, it is large circles that overlap each other. Your plotting method emphasizes the noise in the data, rather than the spatial gradients. It further ignores that rather large horizontal extent of the OCO-2 FOVs, which are parallelograms and in some of these cases are rather wide! You can see an example of this in the figure below, which shows your plot of the Dec 18, 2020 case on the left, the corresponding NASA Worldview plot in the middle (obtainable at https://go.nasa.gov/38MQLRr), and my own home-made generated plot on the right. You can see the two on the right de-emphasize the noise and emphasize the gradients, and further do a better job illustrating the fact that OCO-2 target observations provide relatively complete spatial coverage due to the large degree of overlap of the various soundings in the image. The central image also illustrates the value of including the actual Aqua-MODIS data, at the expense of the spatial resolution of the surface imagery (unfortunately). But often there is cloud contamination in these Munich targets, so it is something to consider when plotting. You may wish to include something like the central image in your Figure 15, to really showcase the cloud context. And in general, please consider using one of these more realistic plotting methods which includes the actual parallelogram-shaped along with some kind of averaging.

We appreciate this suggestion. We updated all the OCO-2 target figures in our manuscript by averaging OCO-2 soundings into 0.02°x 0.02° km bins to comply with the standard guidelines given by JPL. Furthermore, we added the XCO_2 units to the map-plots.



Specific comments:

Section 3.2: The value of 6km radius around each EM27 seems rather arbitrary. Yes, it maximizes the circles while minimizing their overlap, but does it make sense as to the column of air the various MUCCnet sites are really sampling, as compared to OCO-2? For instance, using +/- 30 minutes overpass time for collocation, we could estimate that corresponds to a spatial extent of roughly +/- 15 km for a wind speed of 8 m/s, meaning

that the EM27 signals will be much more "washed out" in the case of higher wind speeds, due to averaging all the EM27 values in that 1 hour time window. You should at least speak to this source of collocation error, and you may wish to mention that more sophisticated schemes (such as using a tighter time window) may reduce the collocation error.

Thank you very much for highlighting this issue. We chose the 6 km radius to equally segment the target area as evenly as possible around MUCCnet's ground-based measurement sites. According to the ERA-5 v10 wind data we used, windspeeds in Munich are rather low. The average wind speed over all was 2.33 +/- 1.54 m/s which corresponds to a spatial extend of around 4.5 km within our colocation time. The highest average windspeeds of 5.08 +/- 1.60 m/s are provided on February 13, 2021. Due to the low wind speeds, we assume the EM27/SUNcolumn signals to still coincide with the spatial extend of our comparison domains. In addition, we usually see rather low temporal variances within the ground-based XCO2 retrievals when compared to the larger variance in the collocated spaceborne XCO2. Consequently, we trade a larger number of spaceborne soundings to reduce the impact of random errors in single soundings against higher specificity in the EM27 signals.

We added a short paragraph in to clarify our thought process:

L194	The relatively long collocation time frame is chosen due to the low average wind speeds of 2.33 +/- 1.54 m/s during the overpasses featured in this study. This may, especially for higher wind speeds, introduce collocation error which can be reduced by adjusting the collocation time frame according to the wind speed.

Near line 150: Please state the spatial resolution of the anthropogenic emissions in TNO-GHG? It appears to be roughly 1x1 km^2, but if it is larger, than could also explain the Dec 18, 2020 discrepancy.

Near surface emission fluxes are taken from the first version of the GHG and co-emitted species emission database provided by the Netherlands Organization for Applied Scientific Research for 2015 (TNO-GHGco_v1.1). The innermost domain covers Munich and its surrounding, initialized by a higher-resolution version of TNO_GHGco_v1.1 at a resolution of approx. 1 km × 1 km. Details of the modelling setup can be found in Zhao et al., (2022). (https://acp.copernicus.org/preprints/acp-2022-281/acp-2022-281.pdf).

Near line 335: Please state the WRF-GHG wind speeds (roughly) in this case. If they are significantly larger than the 0.5-2 m/s wind speeds ERA5 shows, that could explain the ~factor of 10 discrepancy.

WRF modelled mean wind speeds at the 10 meters above the ground within ± 2h of the overpass time are around 1.44 m/s with its standard deviation of 0.48 m/s. Unfortunately, we do not have a better knowledge about the actual windspeed and direction during the overpass, since we did not run any measurement. Nonetheless, we assume the relatively large discrepancies between the modelled XCO2 and the OCO-2 XCO2 to be caused by one or more of the following reasons:

First of all, December 18th, 2020 was an outlier in terms of the captured XCO2 gradients. The gradients are considerably larger, that during all other overpasses. This is true for gradients captured by MUCCnet as well as OCO-2. Since we compare to XCO2 data, that was generated by a yearly averaged emission inventory, we do not expect the model to reliably replicate the actual gradients on that day. In addition, the modeled plume of the WRF-GHG simulation is shifted northwards when compared to the OCO-2 target observation. This causes the highest XCO2 of the simulation to not be fully sampled by the western collocation domain resulting in lower enhancements between the two domains. In the center of the plume, WRF-GHG models XCO2 enhancements of around 0.7 ppm, which is closer to the gradients captured by MUCCnet and OCO-2. We also state this similarly in the manuscript:

L350	We assume this underestimation of XCO2 gradients to be caused by both, uncertainties in the annual emission inventory as well as transport uncertainties. A mismatch in model wind speed and direction causes the area of maximum XCO2 enhancements to be shifted to the north in the modeled XCO2 (see Fig. 15). Furthermore, the XCO2 in the target observation is notably higher than on other days, indicating unusually high emissions in Munich on December 18, 2020, which can't be replicated by a yearly averaged bottom-up emission inventory, while the spatial distribution is reproduced rather accurately.

Near line 365: You ignore the factor of 10 discrepancy in the spatial XCO2 gradients between WRF-GHG and the observations. You should mention this as well, as the obs provide a potential way to improve whatever is going on in the model. I'm guessing perhaps there is something fundamentally wrong about how the model was set up and run, because I think WRF, at sufficiently high spatial resolution, should be able to duplicate the rough magnitude of observed spatial gradients

As stated above we think this factor of 10 is rather caused by the way we compute the gradients in this case. We compare the mean XCO2 of the center domain to the mean XCO2 in the western domain. Due to the shifted plume in the WRF-GHG model, we do not sample the actual highest XCO2 in the model data, while we do that for the OCO-2 measurements.