We would like to thank Ray Nassar for the careful reading and valuable comments that we try to address point-by-point in the following. Reviewer comments are in italics, answers in normal letters and suggested new text in blue color.

"Evaluation of simulated CO_2 power plant plumes from six high resolution atmospheric transport models" by Brunner et al. assesses model simulation capabilities to understand their applicability for use in power plant emission estimation from aircraft and satellite observations. The study uses real aircraft in situ and remote sensing observations from the CoMet aircraft campaign of 2018 over the Belchatow and Jänschwalde power plants and simulated CO2M observations (with 2 noise levels) with six models including Eulerian, Lagrangian and Large Eddy Simulation (LES) models. This is a useful and informative study that enhances our understanding of model capabilities and limitations, but the authors correctly warn that a complete comparative evaluation of the models cannot be carried out based on just a few overflights. The quality of the study is high and the interpretation of the results is generally sound, as the authors are careful not to over-interpret the relatively small number of examples used in the study.

I would have liked the study to go the additional step of actually reporting the estimated emissions of the two power plants based on the real aircraft observations and not only the simulated CO2M data (Table 5). The spread in derived emission estimates that would be obtained with the different models would be informative, although not essential for this study.

Indeed, this additional step could have been taken but we wanted to focus on the model simulations and leave such an analysis to the measurement groups. Nevertheless, we took a step towards a more quantitative analysis. As described in more detail in our response to the other reviewer, we estimated the actual CO_2 emissions during the observation periods from actual energy production rates following your publication (Nassar et al., 2022). Based on this analysis, we scaled all simulated CO_2 fields by a factor 1.23 for Bełchatów and 1.28 for Jänschwalde (actual energy production was higher compared to the annual mean) and regenerated all figures with CO_2 data. The scaled model results agree better with the observations. Furthermore, in Sect. 3.4 (Statistical properties of the plume), we added an analysis of plume integrals to the analysis of plume widths and amplitudes and compared the model results with the observations. This comparison shows that the scaled model results are fully consistent with the observations whereas using annual mean CO_2 emissions would not have been consistent. The two new figures are shown below

a) b)

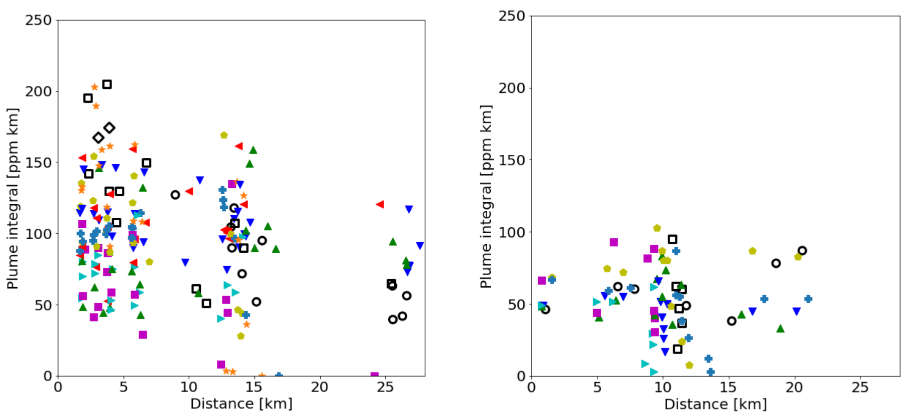

Figure 1: Plume integrals [ppm km] for (a) Bełchatów and (b) Jänschwalde. See Figure 12 and 13 in manuscript for a legend of symbols.

The following paragraphs will be added in Sect. 3.4:

Discussion of results for Bełchatów:

The plume integrals (i.e, the areas under the Gaussian curves) presented in Fig. 12c correspond to the integrated amount of CO_2 along each transect in units of ppm km. Since CO_2 is transported as a passive gas, they are expected to stay constant with distance unless (i) the wind speed or wind direction changes with distance (or with time since the transects were flown at different times), (ii) the plume extent is not fully covered by all transects, or (iii) the plume is not yet homogeneously mixed over the full depth of the ABL, such that a mole fraction measured by an in-situ instrument at a given altitude is not representative for the ABL column mean. The figure suggests that the plume integrals are indeed not constant but decrease with distance, more clearly in the measurements than the simulations. The reason for this could be any combination of the above possibilities. The integrals also enable a quantitative comparison between observations and models. The mean (and standard error of the mean) averaged over all models (excluding ICON-LEM due to its too high wind speeds and excluding points with unrealistically low values below 10 ppm km) and over all distances is 105.6±2.8 ppm km (n=126). The corresponding mean over all observations is 111.9±11.1 ppm km (n=26). The two values agree within their combined uncertainties suggesting that the simulations are consistent with the observations.

Discussion of results for Jänschwalde:

Different from Bełchatów, the plume integrals remain approximately constant with distance. The mean averaged over all models except ICON-LEM is 55.5±2.1 ppm km (n=92) and the corresponding mean over all observations is 57.0±5.6 ppm km (n=13). Again, the two values agree within their combined uncertainties. Using annual mean instead of actual CO_2 emission rates in the simulations would have resulted in too low plume integrals inconsistent with the observations for both Bełchatów and Jänschwalde. This finding agrees with a recent study by Nassar et al. (2022), who demonstrated that it is necessary to account for actual power generation to explain day-to-day variations in CO_2 emissions from Bełchatów estimated from individual OCO-2 and OCO-3 satellite overpasses.

In the conclusions section, the following sentences will be added to the first paragraph:

The CO_2 emissions assumed in the simulations correspond to values officially reported for the year 2018 but scaled by a factor 1.23 for Bełchatów and 1.28 for Jänschwalde to account for the fact that hourly energy production rates were higher during the observations than annual mean production rates. The amount of CO_2 integrated along individual plume transects was highly consistent between simulations and observations when the emissions were scaled in this way.

With or without the above suggestion, I only recommend some minor specific changes before I would deem the study acceptable for publication in Atmospheric Chemistry and Physics.

Specific Points

Line 1: "dominated" is too strong of a term, since there are other major sources like urban CO_2 emissions from which a large fraction is from transportation or residential heating, rather than facilities. Based on the introduction, the contribution from facilities is about 58%.

"Dominated" is indeed to strong. The sentence will be changed to

Power plants and large industrial facilities contribute more than half of global anthropogenic CO_2 emissions.

Line 10: The description should clarify "NWP models extended for atmospheric tracer transport" or something like this rather than just calling them NWP models.

Thank you, this will be changed as suggested.

Line 36: The actual prevalence of stack monitors is somewhat uncertain. Recommend changing "often measured" to "sometimes measured".

In Europe and probably other developed countries such measurements are demanded by regulation, but it is unclear what fraction of industrial and power plants worldwide have such a system. We agree that "sometimes" may be a better description.

Line 46: Recommend including OCO-3 (Nassar et al., 2022 https://www.frontiersin.org/articles/10.3389/frsen.2022.1028240/full), which has enhanced capabilities relative to OCO-2 for locations of interest within the latitude range covered (up to ~52°N), and furthermore is highly relevant to the Belchatów examples.

Thank you for pointing at this publication, which is clearly relevant in the context of our study. We added a reference on line 46 and also added references in Sections 2.2 (Modelling protocol), 3.4 (Evaluation of plume statistics) and 3.5 (Emission quantification with a CO2M like satellite). Note that in response to a question of the second reviewer and considering your recent publication (Nassar et al. 2022), we estimated the actual emissions from Bełchatów and Jänschwalde by comparing hourly energy production data during the observations with annual mean energy production. All simulated CO_2 concentrations were scaled accordingly (with a factor of 1.23 for Bełchatów and 1.28 for Jänschwalde) and all corresponding figures were updated.

Lines 52-56: Should really cite Zheng et al. (2019) as an example for 3D atmospheric transport modelling for CO_2 point source emission estimates https://iopscience.iop.org/article/10.1088/1748-9326/ab25ae, which demonstrated the exact problem with small errors in the wind direction.

We added a reference to Zheng et al. (2019) on line 56.

Line 115: The authors made an appropriate decision to optimize the location used in approximating the power plant as a point source based on actual stack locations.

This was indeed an important consideration, which led to clearly improved results.

Line 277: The authors need to double-check the stated sunrise time of 3:33 CET on June 6, which seems too early, as they are likely reporting the onset of twilight rather than the actual sunrise. I am unsure however which one (onset of twilight or actual sunrise) is more relevant for the ABL height. At minimum, they need to be more careful with wording.

Thanks a lot for spotting this error! According to timeanddate.com, the sunrise (not twilight) was almost one hour later at 4:28 CET. We will change the text accordingly.

Figures 2-4: The comparisons are very interesting and informative regarding the model spread.

Thank you.

Figure 4: Variation in vertical dimension between the models strongly suggests that this will be less of an issue with satellite column data and this is confirmed by comparisons is 3.3 (Fig 9 – 10).

Yes, absolutely. This important point is therefore mentioned also in the conclusions.

Figures 2-11: If the authors can make the model name label on the Figures more prominent in comparison to other text, it would significantly improve clarity for the reader.

We increased the label sizes and placed the model names as the first item in all figure titles.

Figure 14: Units should be specified for colour scale of the figure.

The units are mol cm^{-2}. The corresponding information will be added to the figure caption.

Finally, the conclusion ends somewhat abruptly. I think the manuscript would benefit with an additional paragraph dealing with the bigger picture, where the authors put this study in the context of the expected capabilities and limitations of power plant CO_2 emissions monitoring, verification and support (MVS) with CO2M.

We agree that it would be good to add a few lines placing our study in context. We added the following sentences at the end of the conclusions section:

A potentially important application of high-resolution model simulations as performed in this study is the estimation of point source emissions from satellite observations through inverse modelling. However, accurately simulating the location and structure of the corresponding plumes will remain a challenge especially in the presence of turbulence. Simple Bayesian inversions where simulations and observations are compared locally on a pixel-by-pixel therefore seem little suited but more advanced methods, e.g. using non-local metrics as proposed by Vanderbecken et al. (2022), will be necessary. Whether such methods can outperform simpler methods such as Gaussian plume matching and mass balance approaches that do not require any expensive model simulations, will have to be seen.

High-resolution simulations are invaluable, however, for testing the capabilities of future satellites or other measurement platforms in Observing System Simulation Experiments as shown e.g. by Kuhlmann et al. (2019,2021). Our study shows that simulations performed at a resolution of 1 km or better are able to provide a highly realistic representation of real plumes.

Furthermore, we added the following lines to the paragraph discussing the results from the analysis of synthetic CO2M data.

Our estimate of a 20% uncertainty is higher than the average value of about 12% recently estimated by Nassar et al. (2022) for single Snapshot Area Mapping (SAM) images from the OCO-3 satellite over Bełchatów. They acknowledge that their value could be an underestimate of the total uncertainty, but on the other hand, it was consistent with absolute differences between estimated and expected (from actual power generation) emissions. A 20% uncertainty may thus be a conservative estimate.