

Response to Reviewers

G. Kuhlmann et al.

Anonymous Referee #1

This is truly an excellent paper validating a NO_x emission derivation approach on various global power plants. I was a bit surprised that the paper focused exclusively on power plants, but I understand the reason why. I hope this research team is able to do a follow-up paper focused on urban areas. All of my comments are minor. My most major of the minor comments is that the air mass factor correction section is a bit vague and could use a bit more description and ideally a schematic. All minor comments are below:

- Thank you very much for your positive assessment of our work. We focused on power plants because they provide reliable CEMS data, which makes them ideal for method development and validation. We address all specific comments below.

Line 123. I am curious why you are using a box that extends 30 km downwind as opposed to multiple 10 km boxes following the plume downwind. In this way you could derive the lifetime explicitly, is that correct? I'm not requesting you to re-do your analysis per se, but I am curious the pro's and con's of what I just proposed versus what you are doing. It could be helpful to acknowledge this in one or two additional sentences here.

- Subdividing the 30-km box into three 10-km boxes would only provide three values for estimating the lifetime. From our experience, this would likely be insufficient to accurately estimate the lifetime at overpass due to the complexity of the plume in the nearfield up to about 50 km downwind (e.g., NO to NO₂ conversion, vertical mixing and wind speed variability). A longer plume would be necessary for reliable estimates, but this involves substantially more manual filtering, as shown by Meier et al. 2024. We added the following sentences to the discussion section:
"Nevertheless, further work is needed to assess the applicability in regions with different chemical regimes using dedicated simulations. These simulations can be validated using satellite NO₂ observations, for example, by estimating NO_x decay times from TROPOMI, which is possible for long plumes but requires careful data selection."

Line 134. This is perhaps my most major comment. Can you be a bit more descriptive of how this is done? Are you assuming all NO₂ is only at the stack height, or instead some type of vertically distributed Gaussian enhancement? A schematic of this directly in the paper or the supplement could be very helpful.

We have revised the method section to improve the explanation of the method and added a figure showing the workflow for one example. Revised text: *"To address this limitation, we apply a correction to the AMFs using the averaging kernels (AKs) provided in the standard product and a modified vertical NO₂ profile (Eskes and Boersma, 2003). First, we fit the Gaussian curve with the*

linear background (Eq. 3) to the uncorrected NO₂ image (Fig. 1b). Next, we enhance the standard TM5 NO₂ profile by adding the fitted NO₂ enhancements to the TM5 profile of each pixel. The enhancements are vertically distributed according to the GNFR-A emission profile (Fig. 1c). Finally, we recalculate the AMFs by applying the AKs to the modified NO₂ profile. Figure 1d shows that this approach increases the VCDs in the plume center (by 30–40% in this example), while background VCDs are not modified. The corrected AMFs are then used to update the NO₂ column densities. Subsequently, Equation 3 is re-fitted to the corrected NO₂ columns to obtain the AMF-corrected line density q , which is used in the CSF method to compute the final NO_x emission estimates (Fig. 1f)."

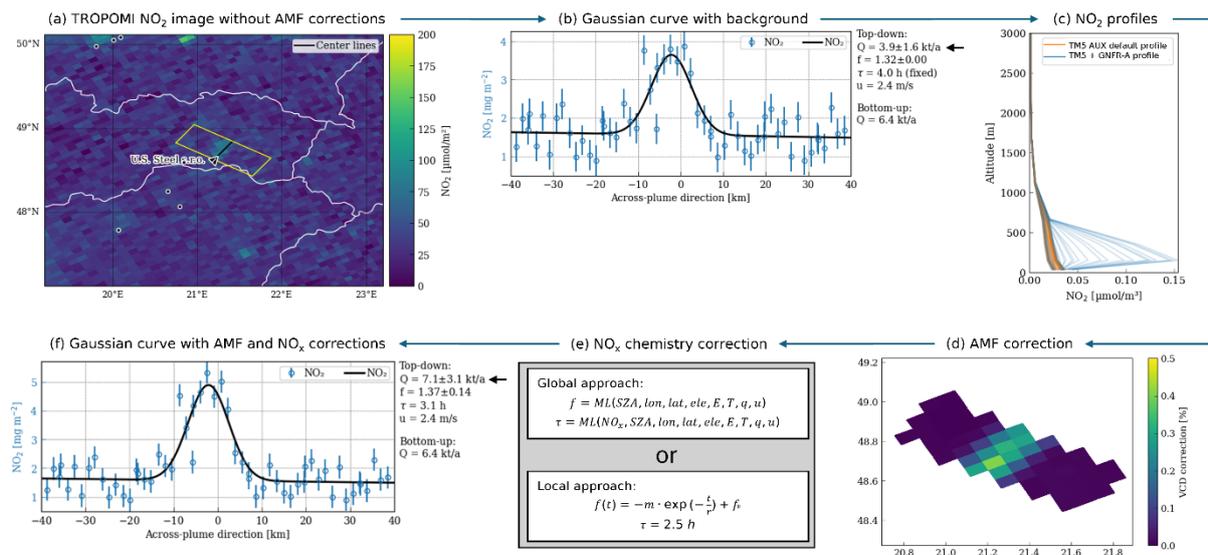


Figure 1: Top-down emission quantification method: (a) TROPOMI NO₂ image with 30 km long and 80 km wide plume area in yellow. (b) Gaussian curve with background (Eq. 3) fitted to the NO₂ columns in the plume area. (c) NO₂ profiles from the auxiliary default profiles from TM5 and profiles with added enhancement from Gaussian curve fit following the GNFR-A emission profile. (d) AMF correction for the pixels in the plume area using the updated profiles. (e) NO_x chemistry correction for f and τ using global or local approach. (f) Gaussian curve with background after AMF correction and final emission estimate Q with NO_x chemistry correction.

Line 175. This is very interesting, and personally, I agree with this approach of using something high-resolution is better than low-resolution. However, the background oxidative environment likely varies between power plant even if emission rates between power plants are similar. For example, a power plant in the presence of a VOC-laden forest may have different oxidative characteristics than a power plant in the desert. I am not sure how you can control for this, but it would be great for you to think about this, and perhaps incorporate this if you do not already, even if it is just a caveat in the text.

- Thank you for this comment. We agree that differences in the local chemical environment can influence plume chemistry. We address this limitation in the discussion, noting that the accuracy of the NO_x corrections depends on the representativeness of the underlying two-day MicroHH simulations for the Janschwalde. Since our study focuses on Europe and the USA, our approach appears to generalize reasonably well. However, dedicated simulations will be necessary to assess

different oxidative environments. We plan to address this in future work with additional MicroHH simulations. We have revised the discussions to emphasize this better.

Line 222. Can you clarify what you mean by 10% systematic error? To me, systematic error implies that the direction of the bias is known (either positive or negative), but that the magnitude of that bias is unknown. It'd be best to modify this phrasing, because I think I am misinterpreting.

- We have clarified the section referring to correlated and uncorrelated errors in order to avoid the ambiguous term 'systematic error'. To estimate the uncertainty, we use Monte Carlo simulations, which assume that errors are uncorrelated, i.e., uncertainties reduce with the number of estimates. For monthly and annual estimates, we account for a temporal sampling error of 30% for each estimate, which was estimated from the bottom-up time series and is consistent with previous studies (Hill and Nassar, 2019). The temporal sampling error is uncorrelated and decreased as the number of estimates increases. Finally, we account for a correlated error of 10% in estimates that do not decrease with the number of estimates; this error is added to the uncertainty of the monthly and annual values.

Lines 226 - 246 (and throughout). For clarity for the reader, please clarify in the text which countries, and US state if applicable, all power plants are located.

- We have added states to Table 1 and 2 and revised the text to indicate country/state throughout the text especially when mentioning the plant the first time.

Figure 1. Ideally please enlarge the text size. My eyes can read it but some numbers are a bit small.

- We have increased the font size in the image (now Figure 2).

Figure 2. Why are you showing CAMS emissions on this plot? Personally I think it'd be clearer to exclude it from this image because it does not seem to be a validation dataset. It's certainly worth mentioning that CAMS disagrees with CORSO, but intercomparing your top-down estimates with CAMS seems unnecessary. I can be convinced either way about this though, if you have a good reason for including CAMS emissions.

- We have added CAMS-GLOB-ANT, because our study is part of the CORSO project, which focuses on the development of the Copernicus CO₂ Monitoring and Verification (CO₂MVS) capacity, where the CAMS inventory is the current standard product. We also use the CAMS inventory to provide spatial context, because not all sources are included in the CORSO database.

Figure 4. For Row 2, it seems the blue line and error bar is simply Equation 4. Is that a correction interpretation? If so, it may be convenient to list this in the white space on the figure. For Row 3, same comment between this and Equation 2. And Row 4 blue line is Equation 2 and 4, right?

- Yes, row 2 and 3 are computed using Equations 4 and 2 for the local approach. In the caption, we only added a reference to Eq. 2, because the figure also shows the global approach, which does not use Eq. 4.

Line 317. power plant —> power plants

- corrected

Line 380. Can you more explicitly define “short-term”? In my opinion, monthly seems very possible. I think your results show that daily is possible only some of the time, and maybe what you are referring to here. And certainly hourly is not possible with once-per-day TROPOMI (which you should also mention).

- In this context, we refer to estimates at overpass or measurement, i.e. with highest available coverage. We have revised the paragraph as follows: *"Despite these challenges, our results demonstrate that satellite observations can effectively resolve seasonal variability in NO_x emissions at the facility level. However, resolving short-term fluctuations from individual satellite images, which are typically only available once or a few times per day, remains more difficult due to data gaps and the high level of uncertainty in individual estimates. This necessitates careful filtering and verification of each estimate."*

Discussion. It'd be helpful to describe how you would alter this method if trying to derive NO_x emissions from a pseudo-point source urban plume.

- We have revised the last paragraph to include how our framework can be applied to other sources especially to urban plumes: *"While this study focused on power plants in Europe and North America -- regions with relatively well-documented emissions from stack measurements -- the same top-down framework can be applied to power plants in regions where bottom-up estimates are more uncertain, as well as to other hot spots such as urban areas and industrial facilities. This broader application is already supported by the ddeq Python library, which allows the user to adapt and modify approach. For urban applications, different vertical NO₂ profiles need to be used for AMF corrections and the effective wind speed, and the NO_x chemistry correction would need to be adapted for urban plumes. Overall, this study demonstrates that satellite observations offer a unique opportunity to constrain emissions and highlights their crucial role in improving global emission inventories for cities, power plants, and a wide range of industrial sources."*

Line 7. Do you mean “when aggregating point sources”?

- Revised as suggested.

Anonymous Referee #2

Review of „Temporal variability of NO_x emissions from power plants: a comparison of satellite- and inventory-based estimates “ by Gerrit Kuhlmann et al.

The study by Kuhlmann et al. evaluates the ability of TROPOMI NO₂ observations to monitor short-term seasonal variability in power plant emissions in the US and Europe, comparing them with bottom-

up emission inventories estimated from electricity generation in Europe and CEMS in the US. With a local and a global approach, two methods for the NO₂ to NO_x conversion are analyzed.

The paper provides more insights into the ability of satellite NO₂ observations to investigate seasonal variations in power plant emissions, as well as their challenges and limitations. I recommend publication after addressing the minor comments below.

➤ Thank you for your positive assessment of our work. We have addressed all comments below.

General comments:

The introduction would benefit from a short overview of some main findings published by other studies closely related to the analyses presented here, see specific comment to P2 L38

The results section would benefit from including comparisons to results from previous studies, see specific comment below.

➤ Thank you for this comment. We have revised the introduction and discussions section to address this comment (see details below).

The study uses several results from other studies, e.g., the MicroHH simulations, power plant emission profiles, and a machine learning model for the global correction factor approach. These studies are briefly mentioned, but I think the reader of this study would have a better experience if some more details are provided for the results taken from these studies. See specific comments below.

➤ We have revised these sections to provide more information on the MicroHH simulations, emission profiles, and the machine learning model (see details below).

Specific comments:

P1 L3: Specify that you are investigating seasonal and day-to-day temporal variability

➤ Revised sentence: "*We assess the capability of TROPOMI NO₂ measurements to quantify the seasonal and day-to-day temporal variability of NO_x emissions from eighteen power plants in Europe and the United States.*"

P1 L5: MicroHH is not well known, please add a short explanation to it.

➤ We revised the phrase to "...a "local" method based on plume-resolving MicroHH simulations...", but cannot add more information here due to 250 word limit of ACP abstract.

P2 L38: spatially **and temporally** resolved emission

➤ Changed as suggested.

P2 L38: I think this section would benefit from a short overview of some main findings published by other studies, e.g., Lorente et al. (2019) and Lange et al. (2022). Especially since Lange et al. (2022) also compared seasonal TROPOMI emission estimates with EPA CEMS data. Maybe even earlier studies by

Frost et al. 2006 and Kim et al. 2006, which compare OMI and SCIAMACHY data with EPA CEMS data, might be relevant.

- *We have revised the introduction and added paragraph addressing previous studies with a focus on current instruments that focus on power plants and comparisons with CEMS data: "Given that CO2MVS envisions assimilating individual emission estimates at the time of the satellite overpass, it is crucial to assess whether temporal variability in emissions can be captured using satellite instruments. This capability would enhance the system's responsiveness to short-term changes in industrial activity, policy interventions, and episodic events such as maintenance shutdowns or fuel switching. In this context, satellite observations of NO2 are of great interest because they are available daily from polar-orbiting satellites (e.g., TROPOMI) and hourly during the daytime from geostationary satellites (e.g., GEMS (Kim et al. 2020) and TEMPO (Zoogmann et al. 2017)). Top-down NOx emissions derived from these observations not only provide information on air quality but can also be combined with CO2:NOx emission ratios to estimate CO2 emissions (e.g., Kuhlmann et al. 2021). NOx emissions have been estimated, for example, from TROPOMI and TEMPO by Goldberg et al. (2019), Beirle et al. (2021,2023), Lange et al. (2022) and Sun et al. (2025). Although NOx emission estimates derived from NO2 observations are largely consistent across different quantification methods, comparisons with bottom-up CEMS measurements for power plants reveal that top-down estimates systematically and substantially underestimate these reported emissions (Lange et al. 2022, Sun et al. 2025). Recent studies based on plume-resolving simulations with chemistry (Krol et al. 2024) suggest that the underestimation is caused by NOx chemistry, where scaling factors used for NO2-to-NOx conversion and NOx lifetime are too small and require larger values for strong point sources (Hakkarainen et al. 2024, Meier et al. 2024)."*

P5 L105: To improve the readability and understanding of this study without further knowledge of Brunner et al., please provide a few more details about the power plant emission profiles, what they consider, and how it depends on the power plant, location, and meteorology.

- *We have added additional explanations: "The effective wind speed was then calculated as weighted mean using the GNFR-A standard emission profile for power plants (Bieser et al. 2011, Brunner et al. 2019). The profile provides a mean distribution for emissions from power plants. It assumes that emissions are distributed between 170 and 990 m with about 50% of emissions between 310 and 470 m. The profile provides a suitable estimate of the NOx distribution near the source."*

P5 L115: I think it is more like an effective lifetime and not the chemical lifetime.

- *We have removed the term 'chemical' to avoid any ambiguity.*

P5 L123: How do you decide the source location, is it based on the inventories?

- *The source locations are provided by the CORSO database, which has manually checked the locations of the largest sources to match stack locations. For the EU power plants, both annual emissions and locations are derived from the integrated Industrial Reporting Database (EEA, 2024,*

doi: 10.2909/9300ec51-d805-4507-9a52-22dabdd9424d), while for the US power plants, we rely on the Emissions and Generation Resource Integrated Database (eGRIDv2021; US EPA, 2023).

P6 L131-137: Please provide more details. Is the correction based on the NO₂ enhancements done on a daily basis? How transferable is the standard emission profile? How strongly does it deviate from the TM5 profiles?

- We have revised the method section to improve the explanation of the AMF correction method and added a figure showing the workflow for one example (see reply to Reviewer #1). The new Figure 1c shows examples of modified NO₂ profiles.

P6 L153: Why are no ozone data used as input for the machine learning model? Where is the NO_x concentration input coming from?

- The model used in this study was specially developed as surrogate NO_x model for ECMWF's integrated forecasting system (IFS) for greenhouse gas emissions, which does provide ozone fields. Therefore, only input variables by IFS are required in this model.

P7 L173/177: You mention that the MicroHH simulations showed variations in lifetime from 1 to 5 hours. What kind of variation, latitudinal, seasonal? How large is the influence of using a fixed lifetime of 2.5 hours?

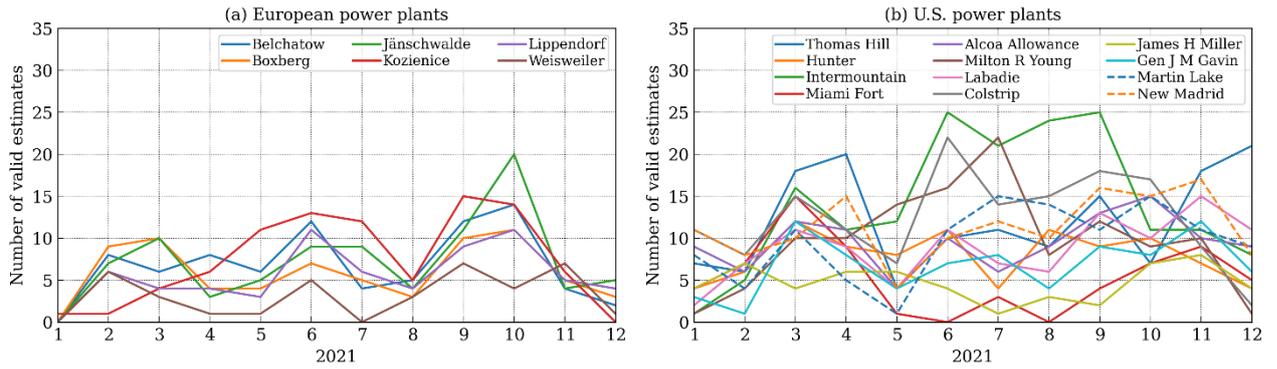
- The variation reflects variability of NO_x lifetime estimated from two days of MicroHH simulations for Jänschwalde and Belchatow. This has been clarified in the manuscript.

P7 L175: Why have you decided for the parameters determined for the Jänschwalde power plant, how much do the parameters vary for the other 3 power plants?

- MicroHH simulations are available for three power plants: Belchatow in Poland, Jänschwalde in Germany and Matimba in South Africa. The fourth simulation is for Lipetsk, which is a steel plant in Russia. We use Jänschwalde, because its emission strength (Q=18 kt/a) lies within the range of power plants analyzed in our study. If we would use the parameters for Belchatow power plant (Q=30 kt/a), NO_x:NO₂ ratios and emission estimates would be about 10% higher, which is within our uncertainty budget. We have revised this section accordingly.

P8 L200: How many individual emission estimates have to be available per month for calculating the monthly mean and having a representative estimate?

- The number varied strongly between months and power plant between 0 to 25 (see Figure S4 in supplement copied below). A small number of estimates will result in a temporal sampling error, which we address by assuming a temporal variability of emissions of 30% in the time series. This assumption is consistent with the temporal variability of emissions in the CEMS measurements.

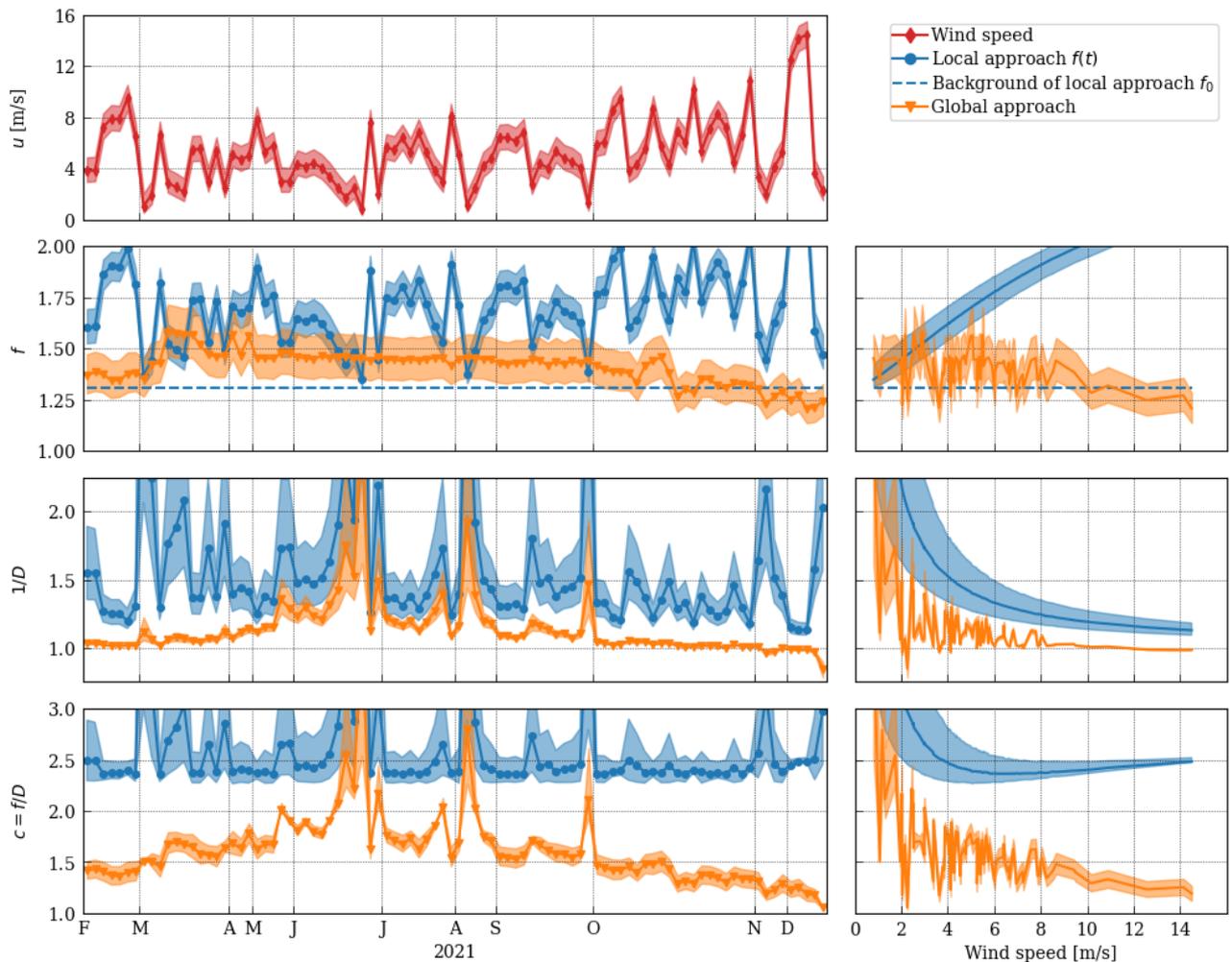


P9 L 246: For me, it wasn't clear at first where one can see the Huntington plume. Maybe add a sentence that it is located north (point) of the Hunter power plant (triangle) within the source area.

➤ Revised as suggested.

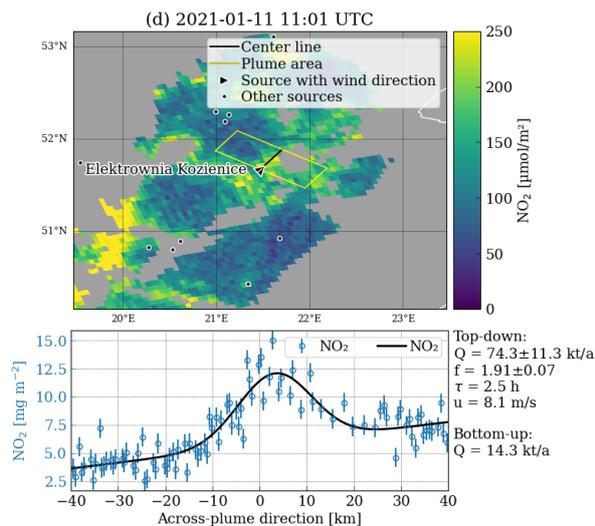
P13 Fig. 4: The limits of the y-axis for the correction factor are ill-chosen to show the seasonal variation. Please adjust the y-axis of the correction factor subplot.

➤ We will provide a second version of the figure in the supplement with smaller y-limits:



P14 L329: Can you see/understand the issue when looking at the outlier results? Is there a way to adjust the filter? This would be interesting when thinking about a more automatic analysis.

- The corresponding scene is shown in the figure below. In this case, background NO_2 concentrations are unusually high. The fitted Gaussian curve likely captures contributions from upstream sources but remains just narrow enough (standard width = 7.8 km) to pass the automatic filter. A potential improvement would be to reject cases with elevated upstream mean concentrations, which could serve as an effective additional filter for more automated analyses.



P14 L332: Do you have an idea why the US power plants show more variability compared to the European ones? Is it an issue with the bottom-up database and the assumption of an emission-power relationship? But I think also the top-down approach shows less variability for the European.

- This difference likely reflects genuine operational differences between many U.S. and European power plants. European coal plants include many large baseload units running at stable capacity factors, especially in Germany and Poland. In contrast, U.S. coal plants apparently have increasingly shifted to load-following due to high renewable penetration in some regions and competition from natural gas.

Results sections: In general, more comparisons to other studies, e.g.:

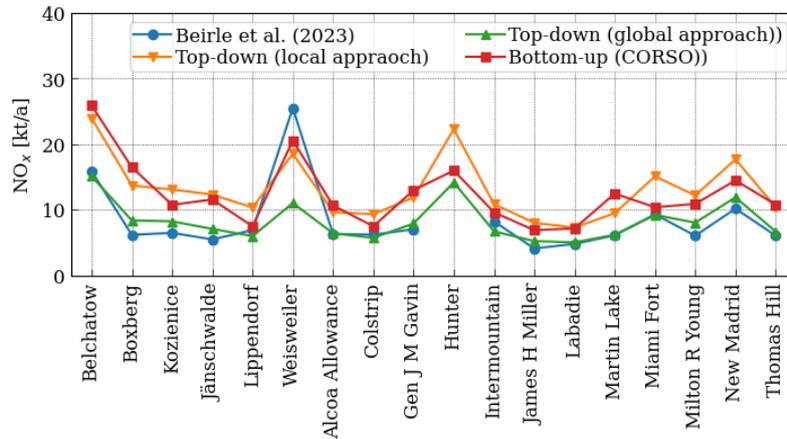
How does the correction factor compare to other studies like Beirle et al. (2021)

Have these power plants studied already in other studies, and how do they compare Beirle et al (2023)?

- We have added additional comparisons with previous studies in the discussions section. The correction factors used are larger than early studies (e.g., Beirle et al., 2023) but consistent with more recent estimates based on the MicroHH simulations (Krol et al. 2024). We added the following text:

"The local correction approach generally yields higher emissions due to larger correction factors, resulting in better agreement with the bottom-up estimates. The NO₂-to-NO_x conversion factor of 1.70 ± 0.08 and the lifetime correction of 1.58 ± 0.38 at about 15 km downstream of the source are

larger than the values of 1.38 ± 0.10 and 1.34 ± 0.25 used by Beirle et al. (2023). Consequentially, our estimates with the local approach are about 70 % larger than Beirle et al. (2023) (cf. Figure S5 in the supplement)."



How do your TROPOMI-based vs CEMS correlations compare with Lange et al (2022)?

- Lange et al. (2022) compare a different time period with only seven estimates, which makes a direct comparison impractical. Correlation coefficients are strongly affected by the strength of the seasonal cycle. We achieve similar correlations as Lange et al. (2022) for power plants with strong seasonal cycle in the CEMS measurements (e.g., Thomas Hill and Miami Fort).

Technical corrections:

P2 L41: of **the** satellite overpass

- Done.

P3 L66: We therefore estimate **hourly** Nox emissions

- Done.

P5 L124: Do you mean +/- 80km or +/- 40km

- Changed to +/- 40. The width in across-wind direction can also be seen in the examples (Figure 1 and 2).

P12 L280: For example, **for** the Hunter power plant, which is ...

- Changed to "One example is the Hunter power plant, which is..."

P17 L379: our results instead of our result

- Done.