## Reviewer 1:

In this article Goldberg et al. present an analysis of how the tropospheric vertical column density (VCD) and surface concentrations of nitrogen dioxide (NO2) depend on cloud coverage. The analysis is carried out over the contiguous U.S. based on measurements from the satellite instruments TROPOMI and TEMPO, in situ measurements at the surface, and simulation data from the chemistry and transport (CT) model WRF-Chem. The influence of cloud cover on these different type of measurements and simulations is often given little attention, at least in the satellite community, where rejection of cloud-contaminated data is mostly the norm. The article deals with an important topic, fits well into the scope of ACP, and presents interesting results. I recommend publication after the following points have been addressed.

Thank you for your comments and suggestions. They have significantly improved the manuscript. Our responses with the revisions are below in red. Text revisions are italicized.

### Major points

I only have two major points of criticism regarding the content of the paper.

**1. NO*y* biases of in situ measurements with molybdenum cartridges** 91 % of the instruments in the EPA AQS dataset use molybdenum cartridges, and previous studies have found that their cross-sensitivity to NO*y* can be very large. Authors like Poraicu et al. (2022) and Kuhn et al. (2024) have reported overestimations in the approximate range of +20 % to +100 % based on model simulations. Examples of empirical studies addressing said issue include Lamsal et al. (2008) and Villena et al. (2012). The authors should consider

- Referencing some of this literature to give the reader an impression of how "far off" these measurements potentially are.

Thank you for this good suggestion. In the Methods Section, we now add additional text and reference the Lamsal et al. (2008), Poraicu et al. (2022), and Kuhn et al. (2024) studies:

*"Lamsal et al. 2008 suggested a correction factor, Equation 1, for converting midday chemiluminescence $NO_2^*$ measurements to $NO_2$ using modelled information of PAN and $HNO_3$.*

$$[NO_2]^* = F_{int} \times [NO_2] \text{ where } F_{int} = \frac{[NO_2]+0.95[PAN]+0.35[HNO_3]}{[NO_2]} \quad (1)$$

*Typically correction factors are in the range of ~1.0 for fresh urban plumes and can be as large as ~3.0 for rural areas during summer, with averages typically in the 1 – 1.5 range for moderate and very polluted regimes, and are important to use for model versus monitor intercomparisons (Lamsal et al., 2008; Poraicu et al., 2023; Kuhn et al., 2024).”*

In this section we also modified the phrase “a small amount” to “some” to better imply that the interference can be substantial in some unique circumstances of summertime photochemically-aged rural air.

- Discussing the influence of these cross-sensitivities on the presented results (e.g. in sect. 3.2). The relevant NO*y* species (PAN, HNO3, etc.) are photo-oxidants, i.e. their concentration (and thus, the “falsely measured“ NO2) decreases under cloud cover. In other words, cloudy scenes are not only expected to have higher NO2 concentrations, but also less measurement bias due to NO*y*. The authors must quantify this effect when comparing in situ measurements with/ without clouds. This could, for example, be attempted through the WRF-Chem simulation data, which lets the authors estimate the contribution of “false“ NO2 using the correction term given by Lamsal et al. (2008).

Thank for you this good suggestion. We have now added a new Figure (Figure 6c) and substantial new analyses in response to this request. We have now calculated $NO_2$* from WRF-Chem to intercompare with the EPA monitors. The normalized mean change using $NO_2$* from WRF-Chem is less than the NMC using only $NO_2$ (+42.1% vs. 58.7%) and in closer agreement with the $NO_2$ from the EPA monitoring network. Below is the newly added text and Figure:

*“91% of monitors in the EPA monitoring network measure using the chemiluminescence method, $NO_2^*$, which quantifies $NO_2$ in addition to some fraction of $HNO_3$. The latter is problematic because the $NO_2$ + OH → $HNO_3$ reaction is often the terminal sink for $NO_2$ during daytime and if $HNO_3$ is additionally being measured then this would appear to buffer photolytically driven changes. We further conducted a sensitivity test in WRF-Chem and found that the NMC is only +42.1% down from 58.7% when a chemiluminescence correction factor from Equation 1 is used (Figure 6c), indicating that some of the perceived differences between WRF-Chem and EPA monitors could be due to monitor interferences from PAN and $HNO_3$.”*

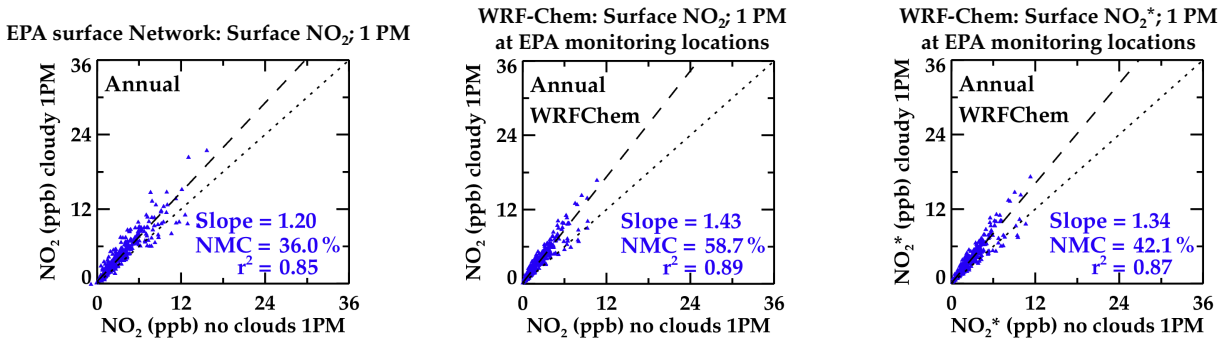



***Figure 6.*** *Scatterplots intercomparing annualized surface $NO_2$ at 13:30 local time during cloudy days vs. no cloud days. (Left) EPA AQS data which is a repeat of Figure 3c. (Center) WRF-Chem collocated with the AQS monitoring sites and using the WRF-Chem cloud filter in lieu of the TROPOMI cloud filter. (Right) WRF-Chem collocated with the AQS monitoring sites, comparing $NO_2$* instead of $NO_2$.*

- In this context, the entry "V2.4 no chemiluminescence" in Table 1 should also be discussed more directly.

Since we now added a new analysis that is more comparable between monitors and WRF-Chem, we feel that the new analysis is better to reference because the sample size of "no chemiluminescence" monitors is small, and these monitors are often only sited in urban areas.

**2. Air mass factors in the satellite retrievals** A good explanation of the air mass factor (AMF) can be found in the TROPOMI PUM (see Eskes et al., 2022). Section 2.2 mentions the AMF, but does not go into the details, which are essential for the retrieved NO2 VCD under cloudy conditions. In particular:

- Line 138: approximately 15 % of these -34.8 % low bias are related to the NO2 profile shapes used to compute the AMF, see e.g. Tack et al. (2021), Judd et al. (2020), Griffin et al. (2019). TROPOMI uses NO2 profiles from the TM5 model which has a horizontal resolution of 1° x 1° (i.e. much lower than the actual measurement resolution).

We have now added a sentence to clarify that some of this bias is due to the operational AMF:

*"Some of this low bias is due to the operational AMF which uses a 1° × 1° model to assume vertical shape profiles; when vertical shape profiles from a regional model are instead*

*used, the bias decreases to between -1% and -23% (Nawaz et al., 2024, Judd et al., 2020, Tack et al., 2021).”*

- Section 3.3: The AMF essentially “fills up“ missing sensitivity with information from the TM5 model. In other words, if the reported NO2 VCD changes in the presence of clouds (as shown in Fig. 5), this is does not necessarily reflect a change of actual NO2 columns as a physical consequence of the clouds - It might just as well be caused by differences between the TM5 model (which then impacts the retrieved NO2 VCD more) and the real world. This aspect should be explained more clearly, and the conclusions in sect. 4 should be adjusted.

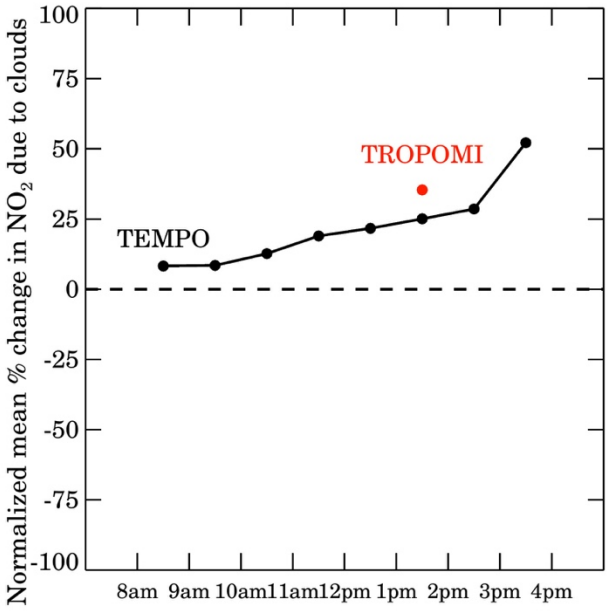
We apologize for not providing enough detail in this section. We are in full agreement. We have now clarified that it is the “model” assumptions in the air mass factor driving this. We have clarified the explanation in this section:

*“In Figure 5, we demonstrate that the vertical column $NO_2$ spatial patterns in the presence of clouds are much different in magnitude than the slant column $NO_2$ whereas the vertical column $NO_2$ spatial patterns in the absence of clouds are similar to the slant column $NO_2$. This is primarily driven by the assumed vertical shape profiles in the model. During measurements when the crf >0.5 as compared to measurements when crf <0.5, the model is “filling in” the missing $NO_2$ and causing small air mass factors as shown.”*

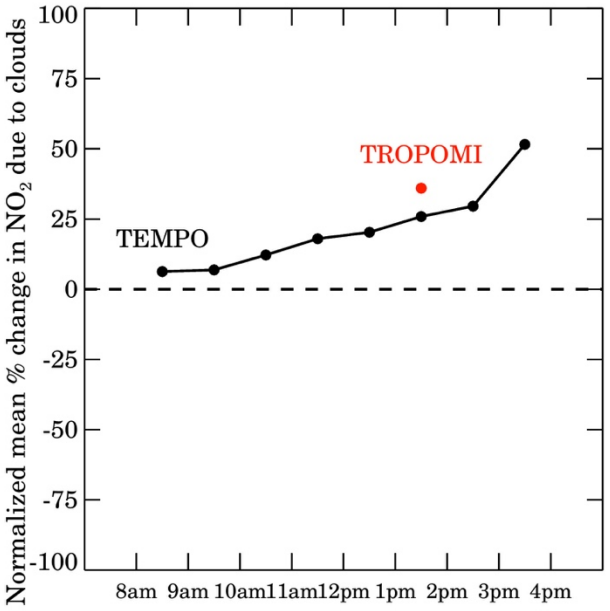
- Section 3.5: This comparison is only justified if both TEMPO and TROPOMI use the same AMFs recipe (i.e. the same NO2 a priori profiles, etc.). Is this the case?

TEMPO and TROPOMI are using different AMFs for this analysis. Below (at the *), we discuss how TEMPO and TROPOMI VCDs change with a regional model to calculate the AMF, but this type of analysis is beyond the scope of this manuscript. We find that using a 4 x 4 $km^2$ regional model to recalculate the AMF has some effects on both TEMPO and TROPOMI in the midday, but it is not dramatic.

However, we acknowledge that in the initial manuscript we could have better matched the temporal timeframe and cloud filters better between TROPOMI and TEMPO. In a major update, we now use the same timeframe, the same cloud filter for TROPOMI, and an updated EPA AQS dataset. Despite this being a major change, this ended up having only a minimal effect on the Figure. The TROPOMI value changed 0.6% due to changing the cloud filter + timeframe + EPA AQS dataset. The TEMPO values changed 0.5 - 2% due to changing the EPA AQS dataset. The “old” and new” figures are below.




New Figure 8.




Old Figure 8.

*In ongoing and unpublished work, for a small Texas, USA domain in September 2023, we find that using a 4 x 4 $km^2$ regional model to recalculate the AMF has some effects on both TEMPO and TROPOMI in the midday, but it is not dramatic. TEMPO and TROPOMI do indeed have better agreement when the same *a priori* vertical profiles are used. However and unfortunately, it is inappropriate to include these results in this paper because this timeframe (1 month) is shorter and different than the timeframe presented in the manuscript (1 year). Recalculating the AMF for TROPOMI and especially TEMPO for 1 year would be a large undertaking and well beyond the scope of this manuscript.

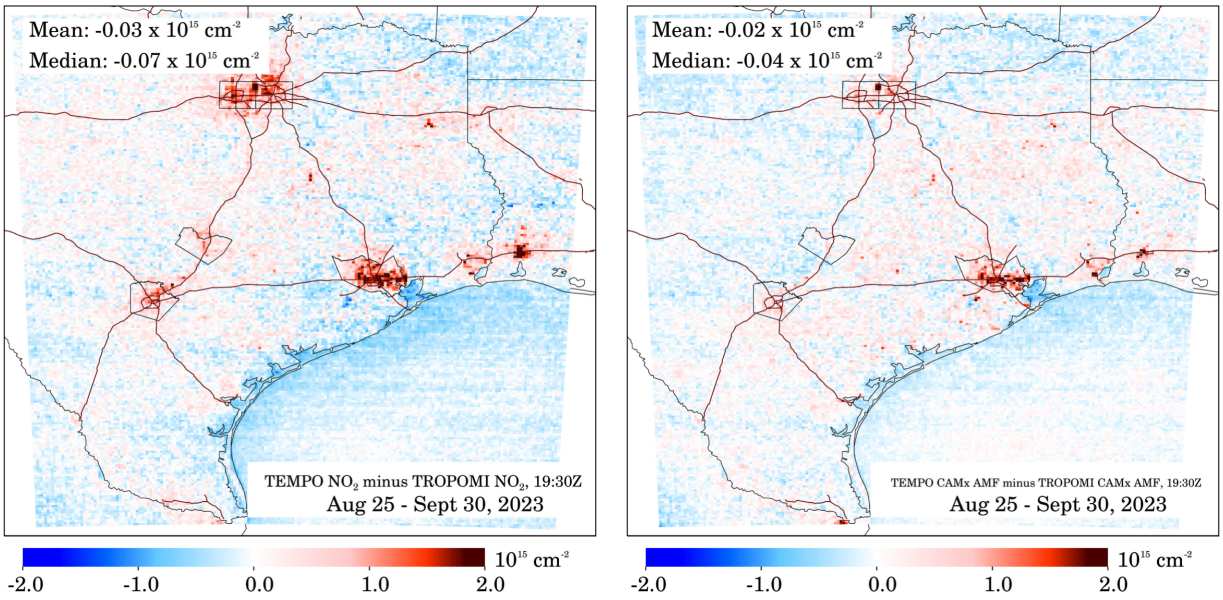



**Figure for private use only.** Difference between TEMPO $NO_2$ and TROPOMI $NO_2$ at the TROPOMI local overpass time (19:30Z). (Left) Using the operational AMFs for both retrievals. (Right) Using a regional model (CAMx) derived AMF for both retrievals.

**Minor points**

Line. 17: Specify that you use WRF-Chem and ERA5 model data.

Added as requested.

Line. 26-27: This sounds as if the authors provided some form of bias-correction method or formula, which is not the case.

Removed “and introduces some corrections to account for these biases”

Line. 49-50: There are numerous more articles on the topic, see e.g. the many references given in Sun et al. (2024). Although there is no necessity to list dozens of reference, the authors might consider referencing: • Cao et al. (2023) and Ghahremanloo et al. (2021), who focus specifically on the U.S. • Kuhn et al. (2024), who address the prediction of surface NO2 with explicit consideration of the molybdenum chemiluminescence biases (see major point above)

Thank you for these additional references. Added Cao (2023), Ghahremanloo et al. (2021) and Ghahremanloo et al. (2023). Kuhn et al. (2024) is be cited when referencing chemiluminescence biases.

Line. 53: better: “up to 5.5 km x 3.5 km“

Modified as requested.

Line. 77: better: “irradiation“ instead of “strength of sunlight“

Modified as requested.

Line. 72-87: The photolysis of NO2 into NO + O should be mentioned somewhere in this paragraph, as this is the main (non-terminal) NO2 sink associated with the high photolytic reactivity of NO2.

Modified to:

*“strong irradiation creates the OH radical which can react with $NO_2$ to create $HNO_3$ – a major terminal sink of $NO_2$ – and also accelerates the photolysis of $NO_2$ into NO and $O(^3P)$ leading to an accumulation of $O_3$ in the presence of VOCs; without VOCs, $NO_2$ cycles more rapidly to NO. Warm temperatures increase biogenic VOC emissions and VOC can react with $NO_2$ directly to create organic nitrates (e.g., peroxyacetyl nitrates and alkyl nitrates) (Zare et al., 2018) which act as a temporary sink of $NO_2$.”*

Line. 92: specify “models“

Modified as requested.

Line. 93: this study does not focus only on the surface “bias“, but also column densities

Added the word “column”.

Line. 95-98: I think the two points given here do not summarize the findings of this article well, because they do not (explicitly) mention the findings associated with TROPOMI and WRF-Chem.

TROPOMI, TEMPO, and WRF-Chem are now explicitly mentioned in this section.

Line. 124: column densities represent the vertically integrated concentration (not molecules per unit area).

Modified as suggested.

Line. 336-341: another option for the verification of reasonable jNO2 values would be to compare simulated NO2/NO ratios to the corresponding in situ observations.

Thank you for this good suggestion, but we have decided not to pursue this for two complementary reasons. First, the chemiluminescence explanation brought up by you seems to be a better explainer of the causes of the bias differences between the monitors and WRF-Chem. Second, it appears that WRF-Chem is simulating reasonable jNO2 through our initial analysis, so we do not feel it is necessary to add an additional verification metric.

**References:**


Lamsal et al. (2008). Ground-level nitrogen dioxide concentrations inferred from the satellite-borne Ozone Monitoring Instrument. Journal of Geophysical Research: Atmospheres, 113(D16).

Poraicu et al. (2023). Cross-evaluating WRF-Chem v4.1.2, TROPOMI, APEX, and in situ NO2 measurements over Antwerp, Belgium. Geoscientific Model Development, 16(2):479–508.

Kuhn et al. (2024). NitroNet – a machine learning model for the prediction of tropospheric NO2 profiles from TROPOMI observations. Atmospheric Measurement Techniques, 17(21):6485–6516.

Villena et al. (2012). Interferences of commercial NO2 instruments in the urban atmosphere and in a smog chamber. Atmospheric Measurement Techniques, 5(1):149–159.

Tack et al. (2021). Assessment of the TROPOMI tropospheric NO2 product based on airborne APEX observations. Atmospheric Measurement Techniques, 14(1):615–646.

Judd et al. (2020). Evaluating Sentinel-5P TROPOMI tropospheric NO2 column densities with airborne and Pandora spectrometers near New York City and Long Island Sound. Atmospheric Measurement Techniques, 13(11):6113–6140.

Griffin et al. (2019). High Resolution Mapping of Nitrogen Dioxide With TROPOMI: First Results and Validation Over the Canadian Oil Sands. Geophysical Research Letters, 46(2):1049–1060.

Eskes et al. (2022). Sentinel-5 pre-cursor/TROPOMI Level 2 Product User Manual Nitrogendioxide. Royal Netherlands Meteorological Institute. https://sentinel.esa.int/documents/ 247904/2474726/Sentinel-5P-Level-2-Product-User-Manual-Nitrogen-Dioxide.pdf Cao (2023). National ground-level NO2 predictions via satellite imagery driven convo-lutional neural networks. Frontiers in Environmental Science, 11.

Ghahremanloo et al .(2021). Deep Learning Estimation of Daily Ground-Level NO2 Concentrations From Remote Sensing Data. Journal of Geophysical Research: Atmospheres, 126(21).