Reviewer 2:

The manuscript by Goldberg et al. title "**NO2 concentration differences under clear versus cloudy skies and implications for applications of satellite measurements"** analyses the differences seen in cloudy versus clear-sky days (using surface in situ, model results and satellite observations) and discusses how these differences may affect satellite observations of nitrogen dioxide columns that are only used under clear sky conditions. The analysis shows that the differences found during clear and cloudy sky conditions need to be taken into consideration when interpreting satellite remote sensing data, as surface data and model results suggest higher surface and column NO2 under cloudy conditions.

I have suggestions to improve this manuscript, which are detailed below. Overall, I think this is a very good fit for publication in Atmospheric Chemistry and Physics after some revisions. It spans multiple disciplines, including satellite remote sensing, air quality, and modelling, making it highly relevant to ACP's readership. I recommend accepting the manuscript after these revisions are made.

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.

General suggestions

In Section 3.5: For TEMPO a much more rigorous cloud fraction is used which will skew the results. For TROPOMI a cloud fraction of 0.5 is used why use 0.15 for TEMPO? A significant issue is that TEMPO's cloud fraction is biased high, there are very few cases where the cloud fraction is that low, a lot of good quality points over clear-sky conditions are filtered when using a cloud fraction cut-off of 0.15. When the coincident cloud fractions from TEMPO and TROPOMI are compared, TEMPO clear-sky is anything below 0.2 (and by clear sky meaning cldF=0 in TROPOMI). I don't think it's fair to compare TROPOMI and TEMPO NO2 that way (as in Fig. 8). Further filters for TEMPO should also include an SZA cut-off (SZA >70 should be filtered) and only snow-free should be used, to have a similar and comparable quality flag as TROPOMI.

Thank you for this good suggestion. We agree that in the initial manuscript we could have better matched the cloud filters and temporal timeframe between TROPOMI and TEMPO. We have decided to match the TROPOMI cloud filter to TEMPO as opposed to the other way around since data processing for TROPOMI is easier (25x smaller file sizes per day).

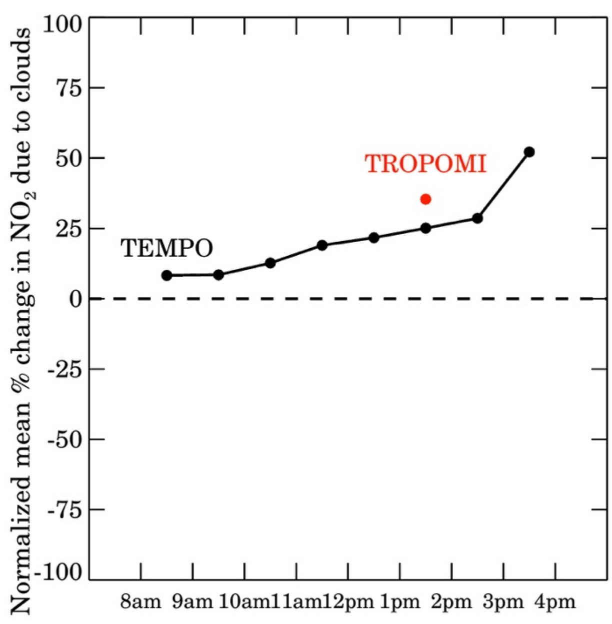
Regarding temporal timeframe, for Figure 8 only, we are now using TROPOMI data during the same August 2023 – June 2024 timeframe as TEMPO.

Regarding the TROPOMI filter for Figure 8 only, we have strengthened the cloud filter to <0.15 during the August 2023 – June 2024 timeframe to match TEMPO. While we agree that there is a high bias in the TEMPO cloud filter, the Users' Guide says it is "biased high by a few percent", which does not give us enough information to warrant using a different cloud filter for TROPOMI than TEMPO.

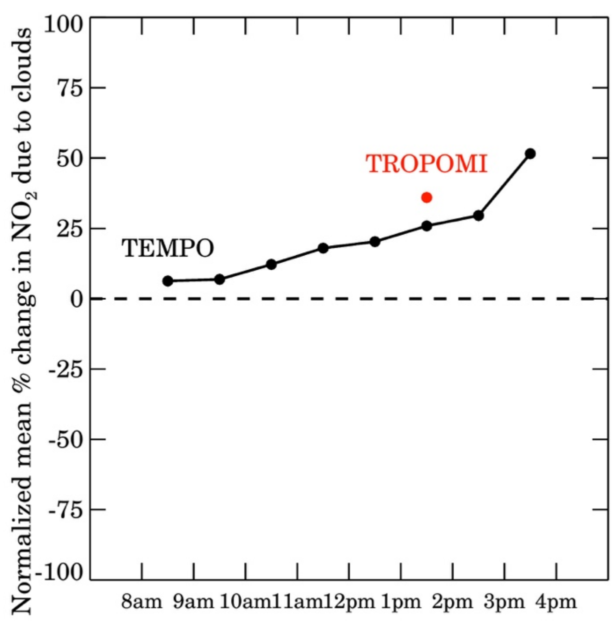
Regarding the TEMPO filter, we continue with the cloud fraction of <0.15 and qa value = 0 as recommended by the Users' Guide. A qa_value=0 already screens out large solar zenith angles.

Last but not least, since the original manuscript submission, EPA AQS has additional ground-level NO2 data in the Q2 (April – June) 2024 timeframe which had been previously missing. These data are now incorporated into the analysis.

With all these changes combined: using the same timeframe, the same cloud filter for TROPOMI, and a complete EPA AQS dataset, 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.

We agree that the difference between TEMPO and TROPOMI are due to how each instrument defines a cloud. While this difference may shrink with future updates to both cloud schemes, it is likely some amount of difference will remain in perpetuity due to inherent differences in instrument characteristics. We have now added the following to the text:

"Differences between TEMPO and TROPOMI are expected because the cloud algorithms and instrument characteristics are different, even though the timeframe and cloud filter threshold used for this analysis are the same."

It would be helpful to include a map of how many days are filtered for TEMPO (per hour?) similar as in Figure 1, and include TROPOMI for the same time period. This could be included in the main text or the supplement. This should also make it clear what is an appropriate filter is for TEMPO when comparing it to TROPOMI.

Thank you for this good suggestion. We have provided Figure S1 to partially answer this question and have now expanded it to additional hours to better match TEMPO (previously 8 AM, 1 PM, and 5 PM; now all individual hours between 8 AM – 4 PM). This figure uses WRF clouds to demonstrate what fraction of days TEMPO "should" be measuring if TEMPO had minimal cloud biases, a regular measurement pattern, and took measurements during 2019.

The reason we cannot create a similar figure for TEMPO is three-fold: 1) TEMPO has an irregular measurement pattern: sometimes missing full hours or days, sometimes acquiring more than one measurement per hour, sometimes missing data because spectra is saturated or the solar zenith angle is too large, etc. Accounting for all of this requires a detailed look at the data that is time-consuming. With TROPOMI's predictable daily pattern this is much easier to account for. 2) The TEMPO dataset is ~25x larger than TROPOMI; the processing to create the TROPOMI image took 1 day of computation time, for TEMPO it would take almost 1 month. 3) The cloud patterns during the TEMPO timeframe (Aug 2023 – beyond) are different than 2019, so we do not have anything to intercompare with (TROPOMI, WRF, and ERA5 are all processed for 2019 in the paper) without significant extra work.

Minor suggestions

Line 53: 5.5x7 km2 before August 2019

Modified.

Line 74: is this in North America (and/or US) or globally?

Clarified that this is referring to statistics in the U.S. Figure 2 shows this.

Line 94: 2019 is a critical year to use for this kind of analysis as the TROPOMI resolution switched from 5.5x7 km2 to 5.5x3.5km2. Could this impact any of the results? It would be good to include a couple of sentences addressing this or at least highlighting this, e.g. the clear versus cloudy pixels could change as with a smaller pixel size more pixels are potentially better quality, also the NO2 columns likely increase in urban areas with the smaller pixel size as they are observing a smaller area.

Thank you for this good suggestion. We found the Krijger et al., 2007 study to be particularly helpful here. When changing pixel size from 25 km^2 to 19 km^2 it is likely increased the fraction of pixel availability by 1 – 2% (Krijger et al., 2007 Figure 1). We have now added text to the Discussion:

"In some ways, the chosen year 2019 was an ideal year to conduct the analysis because it preceded the 2020 global pandemic and its nonlinear and lingering effects on air pollution. But in other ways, this year was less ideal because TROPOMI pixel sizes changed in August 2019 from 7 × 3.5 km^2 (~25 km^2) to 5.5 × 3.5 km^2 (~19 km^2). The fraction of clear-sky pixels likely increased by 1 – 2% after August 2019 as smaller pixel sizes can better "see around" clouds (Krijger et al., 2007). This probably did not meaningfully affect our analysis but is nonetheless a caveat of using 2019 data."

Line 128/129: Include more discussion on the impact of the cloud fraction on the AMF (or the VCDs) it contributes significantly to the final AMF. A cloudy and a clear sky AMF is calculated then the cloud fraction is used to weight these AMFs to get the final AMF. The cloudy AMF is typically smaller leading to higher values of the VCD (VCD = SCD/AMF). See e.g. McLinden et al., 2016; Liu et al., 2021; Nowlan et al., 2025 for more details.

We apologize for not providing enough detail in this section. We are in full agreement with you. 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."

Line 157/158: from Mexico City to the Canadian Oil Sands

Added. Also added the latitude range which is ~17N to ~58N. Added:

“extending from Mexico City (~17°N) to the Canadian Oil Sands (~58°N)”

Line 192: there are other reasons other than clouds for low quality (like snow as mentioned), why not just use the cloud fraction to define clear-sky/cloudy-days? The higher values further north or over the mountains as seen using the TROPOMI qa filter could be due to snow rather than clouds.

We ultimately decided to use a qa_value > 0.75 in lieu only a cloud fraction filter because of its agreement with ERA5 and WRF clouds, with the latter being our best estimate of “true” clouds. Figure 1 (left panel below) shows that using a qa_value > 0.75 threshold matches the observed cloud fractions from ERA5 and modeled cloud fractions from WRF well.

If we use only a cloud fraction filter, then we see worse agreement with ERA5 and WRF (see figure below). It likely that the qa_value calculation is further capturing “true” clouds that the cloud fraction filter alone is not capturing.

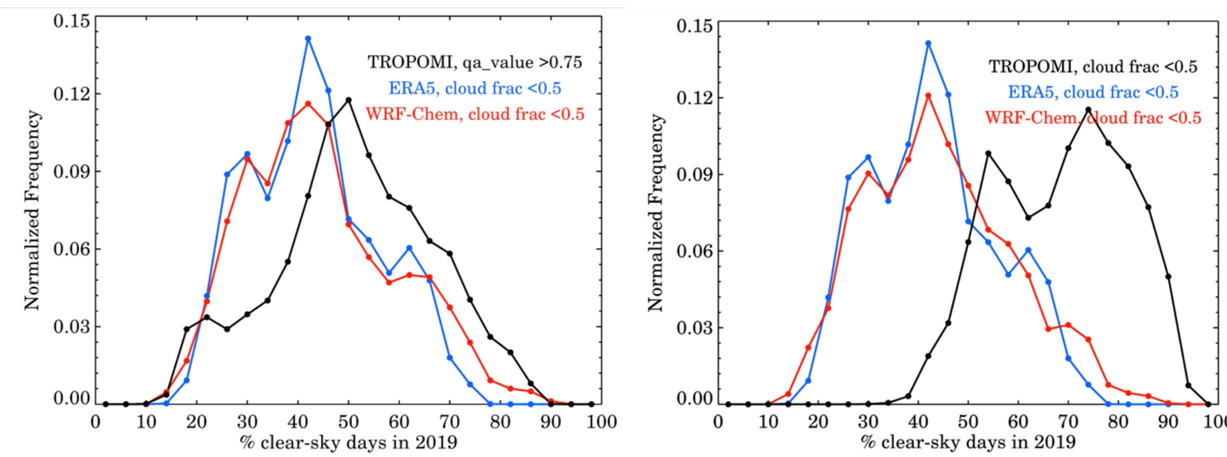


Figure for private use only. Percentage of clear-sky days over the contiguous U.S. during 2019 from the TROPOMI NO_2 V2.4 product using a qa_value > 0.75, ERA5 re-analysis, and WRF-Chem. (Left) Normalized frequency diagram of the binned percentage of clear sky days for the three products. (Right) Same, but now showing TROPOMI cloud fraction < 0.5.

Snow covered pixels are not a cause for this discrepancy between using a qa_value threshold and cloud fraction filter. First, according to the Users' Guide cloud-free snow-covered scenes typically have a qa_value > 0.75. Second, when we analyzed the data ourselves, we find a very small amount of data over the continental US – mostly over very mountainous regions and Canada – are affected by low quality retrievals due to snow alone (see below).

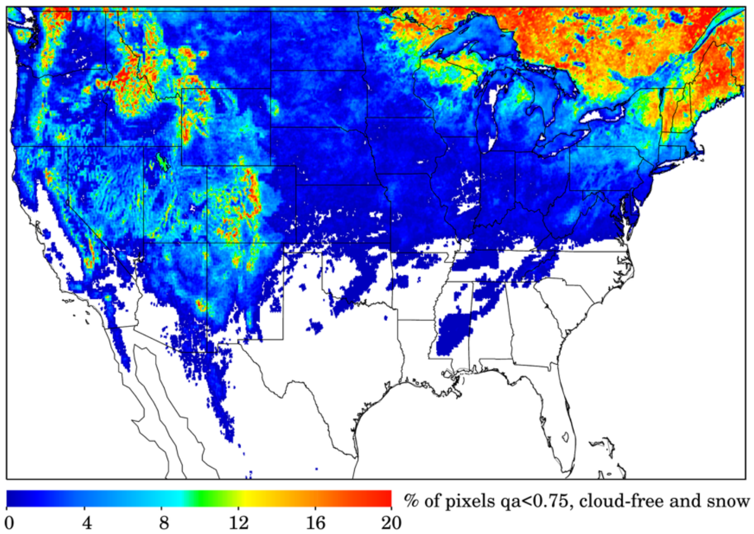


Figure for private use only. Percentage of days with cloud-free snow-covered data filtered out using a qa_value > 0.75

Line 244: The reference in the text to Table 1 is very brief, more details should be included. Here information on chemiluminescence instruments is included, which is quite an important consideration and should be mentioned and discussed more in the text. The no chemilumeneces instruments show a similar slope as the baseline which is encouraging and shows that the increase is not due to increased NOz. It is very encouraging to see that.

Thank you for 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 change using only NO_2 (+42.1% vs. 58.7%) and in closer agreement with the NO_2 from the EPA monitoring network. The NO_z did have an effect, but was not fully responsible for the difference. Below is the newly added text and Figure:

"91% of monitors in the EPA monitoring network measure using the chemiluminescent 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% 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."

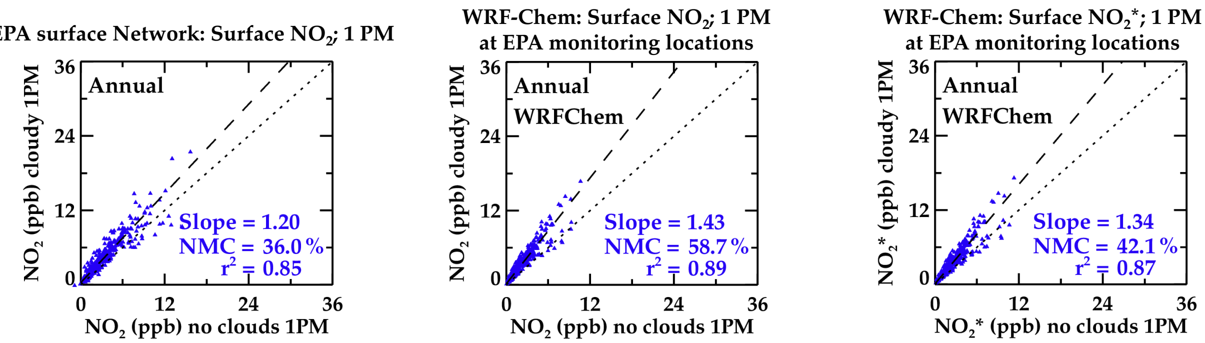


Figure 6. *Scatterplots intercomparing annualized surface NO_2 at 13:30 local time during cloudy days vs. no cloud days. (Left) EPA AQS data which is a repeat of Figure 3c. (Center) WRF-Chem collocated with the AQS monitoring sites, and using the WRF-Chem cloud filter in lieu of the TROPOMI cloud filter.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

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.

What is baseline v2.4 and v2.3.1 there is no explanation of what it means.

These are in reference to the TROPOMI algorithm versions. These acronyms have now been clarified in the Methods Section text and are now consistent throughout the text:

"For our analysis we use the TROPOMI NO_2 version 2.4 (V2.4) re-processed algorithm during Jan 1, 2019 – Dec 31, 2019. We also conducted a sensitivity study using the version 2.3.1 (V2.3.1) algorithm."

How about adding a figure like Figure 5 but for TEMPO and different hours of the day (maybe morning afternoon and evening), either in the main manuscript or in the appendix

Thank you for this good suggestion, but since TEMPO data is 25x larger than TROPOMI this would be computationally expensive. While we have processed a full year of TEMPO data using a cloud fraction of <0.15 and qa=0, we did not retain data when cloud fractions > 0.15. Reprocessing this dataset would take multiple weeks of computational and personal effort.

Figure S1: why switch to 0.3 as a cut-off when most of the paper used 0.5, Figure 2 (in the main manuscript) also uses 0.5, could all 24 h be included or at least show a few more

Thank you for catching this typo in the figure caption. The figure itself shows the correct filter which was indeed 0.5.

We have now also included all individual hours between 8 AM – 4 PM) in the figure (previously only 9:30 AM, 1:30 PM, 4:30 PM).

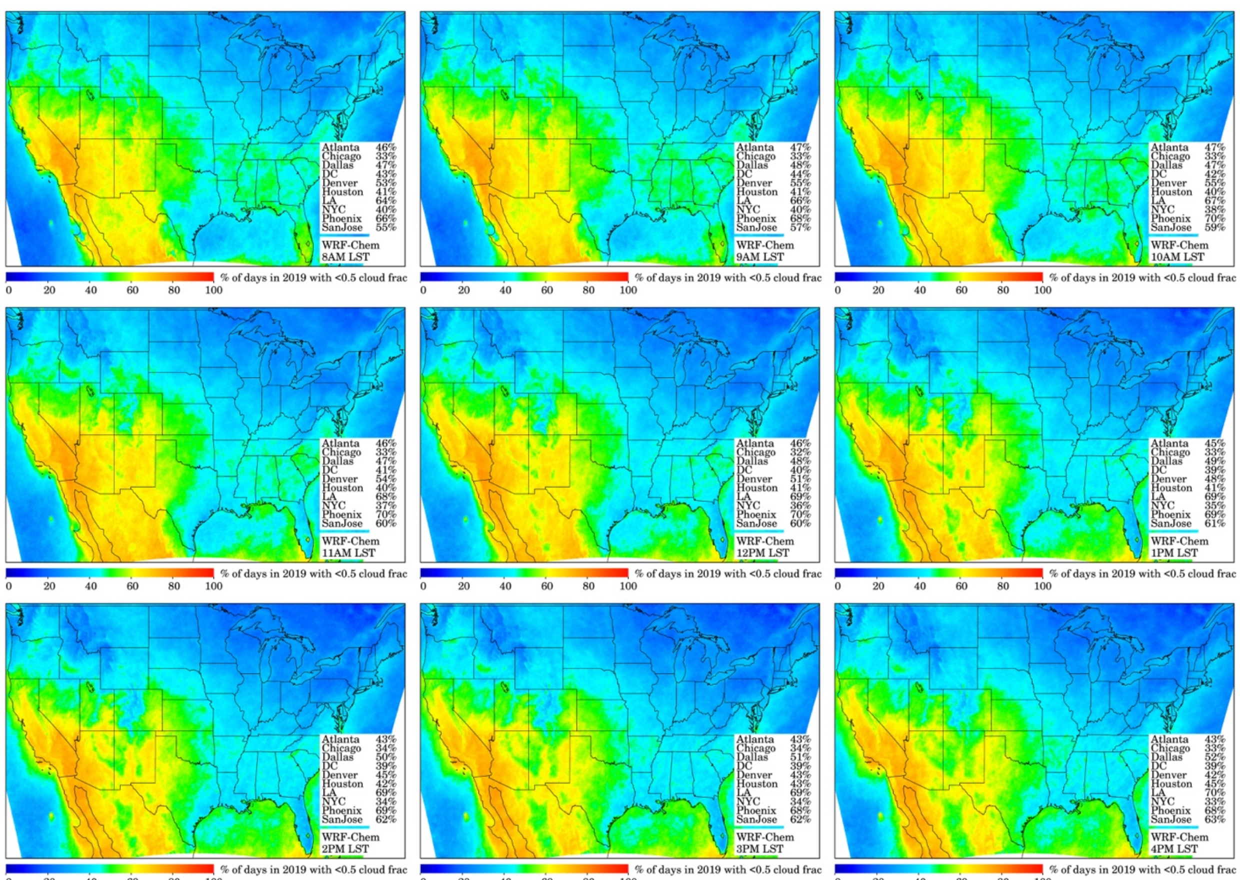


Figure S1. Percentage of days over the contiguous U.S. during 2019 with cloud fractions less than 0.5 as simulated by WRF-Chem at individual hours of 8 AM through 4 PM local solar time.

References

McLinden, C. A., Fioletov, V., Boersma, K. F., Kharol, S. K., Krotkov, N., Lamsal, L., Makar, P. A., Martin, R. V., Veefkind, J. P., and Yang, K.: Improved satellite retrievals of NO_2 and SO_2 over the Canadian oil sands and comparisons with surface measurements, Atmos. Chem. Phys., 14, 3637–3656, https://doi.org/10.5194/acp-14-3637-2014, 2014.

Liu, S., Valks, P., Pinardi, G., Xu, J., Chan, K. L., Argyrouli, A., Lutz, R., Beirle, S., Khorsandi, E., Baier, F., Huijnen, V., Bais, A., Donner, S., Dörner, S., Gratsea, M., Hendrick, F., Karagkiozidis, D., Lange, K., Piters, A. J. M., Remmers, J., Richter, A., Van Roozendael, M., Wagner, T., Wenig, M., and Loyola, D. G.: An improved TROPOMI tropospheric NO_2 research product over Europe, Atmos. Meas. Tech., 14, 7297–7327, https://doi.org/10.5194/amt-14-7297-2021, 2021.

Nowlan, C. R., Gonzales Abad, G., Liu, X., Wang, H., and Chance, K: TEMPO Nitrogen Dioxide Retrieval Algorithm Theoretical Basis Document, https://asdc.larc.nasa.gov/documents/tempo/ATBD_TEMPO_NO2.pdf