This manuscript presents a detailed and comprehensive analysis on the use of the HCHO/NO2 as measured by satellites to characterise the photochemical regimes for ozone production. The manuscript focusses on four different aspects usefulness of HCHO/NO2 as a proxy, the impact of the vertical distribution, spatial heterogeneity, and retrieval uncertainties itself. The analysis draws from a range of model and measured data and makes uses of different statistical approaches. The manuscript provides a wealth of information, but it will be most valuable for the specialist community. I recommend publication in Atmos. Chem. Phys. (although it would also fit well into AMT) after consideration of my comments below.

For the different aspects, different methods and different statistical metrics are used. I would like to get some justification why a specific metric is used and more detail on applied the methods:

Answer
We thank the reviewer for taking their time to provide constructive comments. Our response follows:

- Altitude dependency (section 3.5)
  - Can you please provide some more details on the equation used to compute the first moment of the area (equation 9). The moment of an area is the integral of distance over area. Also, dz is missing.

Answer
Thanks for noticing this. We added "dz". **We also elaborated about the notion of the formula.** **If we rotate the vertical distribution in $HCHO/NO_2$ in Fig. 5 by 90-degree in counter-clockwise direction, we will have the new x-axis as height (z), and the y-axis as the ratio. The centroid along the y direction (the ratio) can be obtained through 1/*A*∫*y (x2−x1) dy* or alternatively *1/2A*∫(*y*2)^2-(y1)^2 dx. Reversing the dimension (x →z and y → $HCHO/NO_2$ →f(z)), we get Eq.9.**
Modifications
We added to the text that: "One can effortlessly fit this function to different bounds of the vertical distribution of FNR such as the 25th and 75th percentiles, and subsequently estimate the first moment of the resultant polygon along *z* divided by the total area bounded to the polygon via:" $$G(z_1, z_2) = \frac{1}{2A}\int_{z1}^{z2} f^2(z)_{75th} - f^2(z)_{25th}\, dz$$

  - Note that a satellite observesa column which is either given by the integral of the concentration over altitude or mixing ratio over pressure, while here mixing ratios seem to be integrated over height which is not correct.

Answer
In satellite observations, we prefer to use pressure (or air density along with mixing ratios) because, in the spectral fitting, we usually estimate the total number of molecules. As a result, it is more convenient to present the concentrations as partial columns or to present the vertical mixing ratios as function of air pressure. In the integral, we calculated the mass centroid of the ratio of HCHO/NO_2 bounded to different percentiles. The ratio is unitless, independent of the unit choice (partial columns or mixing ratios). Therefore, readjusting the equations using the air pressure instead of height should not change any result. However, we wish to keep the formula as a function of height because the *adjustment factor* is suitable when applied for a *given PBL height* which is conventionally given as *km*.

- Why is the standard-deviation of the ratio of the first moment of the interquartile range a good metric for the uncertainty

Answer
There are three uncertainties associated with the adjustment factor: i) the fact that we use observations from only a few campaigns limits our ability to say that the adjustment factor can be generalized to everywhere and every time. We discuss this caveat in detail in the original version of the manuscript. Also, ii) the boundary choice for Eq.9 is subjective. We had tested it for various numbers and as we mentioned in the paper, the adjustment factor became unrobust for large percentiles (> 80^{th} percentile). Finally, iii) there is an error in our assumption about using second order rational functions to describe the vertical distributions in the ratio. This is where the 26% error comes from. We re-estimated the adjustment factors for different coefficients at 1 sigma level (68% confidence level) in Eq.8 (the second-order rational functions) to be able to create the dashed red line in Figure 6. We have elaborated this in the new draft.
Modifications
We added that: "…where *zt* can be interchanged to match the PBLH. This definition is more beneficial than using the entire tropospheric column to the surface conversion (e.g., Jin et al., 2017) because ozone can be formed in various vertical layers. To determine the adjustment factor error, we reestimate Eq.9 with ±1σ level in the coefficients obtained from Eq.8. The resultant error is shown in the dashed red line in Figure 6. This error results from uncertainties associated with assuming that the second-order rational function can explain the vertical distribution of FNRs."

- What is the impact of altitude sensitivity of the satellite column measurement as described by the averaging kernel on the estimate uncertainty.

Answer
This is a great comment. We agree that the magnitude of sensitivity of the radiance to optical thickness within the wavelengths used for HCHO (~350 nm) and NO_2 (~450 nm) is not the same. HCHO tends to have a lower sensitivity to the tropospheric region making VCDs more dependable on the prior model information (AMF). But one of the biggest motivations of using the ratio as described in Martin et al., 2004, and our very recent study in AMT (Johnson et al., 2022), is the fact that the shape of scattering weights is not too drastically different for these two channels. As a result, the first-order discrepancy in scattering weight calculation get normalized after we divide HCHO VCDs by NO2 VCDs. This is why the mean bias in the ratio gets closer to

zero in (Johnson et al., 2022: https://amt.copernicus.org/preprints/amt-2022-237/), despite the fact that individual products can possess a large mean bias.

Regarding the adjustment factor, the shape of scattering weights only matter (rather than the absolute values) which is not drastically different for those two bands (for a generic land pixel) within the first 5 km where the largest variability in the ratio lies in. See Figure 4 in https://amt.copernicus.org/articles/10/759/2017/amt-10-759-2017.pdf. Or the blue line in Figure 2 in https://amt.copernicus.org/articles/11/5941/2018/ within 5 km (~ 600 mbar). So we do not think it will introduce a larger inhomogeneity in the columnar ratio. Our assumption may not hold for a particular scene with variable extinction efficiency induced by complex aerosol optical properties between 350 and 450 nm, or for a specific viewing geometry (particularly, when the geometric AMF is large around early morning or late afternoon). So we added a caveat saying that we had assumed that the shape (the curvature) of the scattering weights of HCHO and NO_2 between surface and 5 km (around 600 mbar) is rather similar.

Modifications

We added the caveat:

"A lingering concern over the application of satellite-based FNR tropospheric columns is that the vertical distribution of HCHO and NO_2 are integrated in columns thus this vertical information is permanently lost. As such, here we provide insights on the vertical distribution of FNR within the tropospheric column. This task requires information about the differences between i) the vertical shape of HCHO and that of NO_2 and ii) the vertical shape in the sensitivity of the retrievals to the different altitude layers (described as scattering weights). Ideally, if both compounds and the scattering weights show an identically relative shape, the FNR columns will be valid for every air parcel along the vertical path (i.e., a straight line). Previous studies such as Jin et al. (2017) and Schroeder et al. (2017) observed a large degree of vertical inhomogeneity in both HCHO and NO_2 concentrations suggesting that this ideal condition cannot be met. A real-time true state of their vertical distribution is not always present, but a natural way of accounting for their distribution is to use retrospective measurements to constitute some degree of generalizations. As for the differences in the vertical shapes (i.e., the curvature) of the sensitivity of the retrievals between HCHO and NO_2 channels (i.e., ~ 340 nm and ~440 nm), under normal atmospheric and viewing geometry conditions, several studies such as Nowlan et al. 2018 and Lorente et al. 2017 showed small differences in the vertical shapes of the scattering weights within first few kilometers altitude above the surface where the significant fluctuations in FNRs usually take place. Therefore, we do not consider the varying vertical shapes in the scattering weights in our analysis. This assumption might not hold for excessive aerosol loading with variable extinction efficiency between ~340 nm and ~440 nm wavelengths or extreme solar zenith angles."

- Spatial heterogeneity (Section 3.6)
    - Please justify the use of the metrics given in equation 14 to quantify the representation error.
    - Important to point out that this is not an absolute but a relative metric (with 3x3 km2) as reference

Answer

The spatial information or variance can be described by the spatial autocorrelation or semivariogram described in Eq. 12 (Matheron, 1963). Our previous study showed how this operator can describe the level of spatial heterogeneity or variance in idealized cases in Figure 1 in https://amt.copernicus.org/articles/15/41/2022/. The semivariogram can be influenced by noise. As a result, we need to fit a function to the semivariogram such as the stable Gaussian distribution used in Souri et al., 2022. The modeled semivariogram then can be used to compare

one dataset to another one allowing for understanding the extent of the spatial variance at a specific length scale each field provides. If two fields show an identical spatial variance (say the first field has a plume, and the second field has the identical plume but rotated 90 degree clockwise), both semivariograms will be identical and the ratio of γ to γ_{ref} will be 1 meaning our target can 100% represent the spatial variance presented in the reference. The ratio of γ to γ_{ref} cannot go above 1 as long as we base the reference on a finer dataset (the baseline). So Eq 14 (1 – gamma/gamma_ref) is proposed to calculate the opposite effect meaning how much of information the target field has lost compared to the reference.

We also added the caveat saying the metric is metric.

Modifications

We added:

"To remove potential outliers (such as noise), it is wise to model the semivariogram using an empirical regression model. To model the semivariogram, we follow the stable Gaussian function used in Souri et al. (2022).."

"where $\boldsymbol{\gamma(h)}$ and $\boldsymbol{\gamma_{ref}(h)}$ are the modeled semivariogram of the target and the reference fields (3×3 km^2). This equation articulates the amount of information lost in the target field for the reference. Accordingly, the proposed formulation of the spatial representation error is relative."

- Satellite errors (section 3.7):
  - 15 assumes uncorrelated random errors between the HCHO and N02 retrieval. This is the case of measurement noise-driven errors but the scatter (standard deviation) in both will also be the result of variable geophysical parameters (e.g. aerosols) which will have some level of correlation.

Answer

We agree. We added a caveat.

Modifications

We added: "where $\boldsymbol{\sigma_{HCHO}}$ and $\boldsymbol{\sigma_{NO_2}}$ are total uncertainties of HCHO and NO_2 observations. It is important to recognize that the errors in HCHO and NO_2 are not strictly uncorrelated due to assumptions made in their air mass factor calculations. The consequence of disregarding the correlated errors is an underestimation in the final error. "

  - What is the role the different averaging kernels between the satellite and ground-based DOAS instruments

Answer

That's an excellent comment which has been mentioned in Verhoelst et al., 2021.

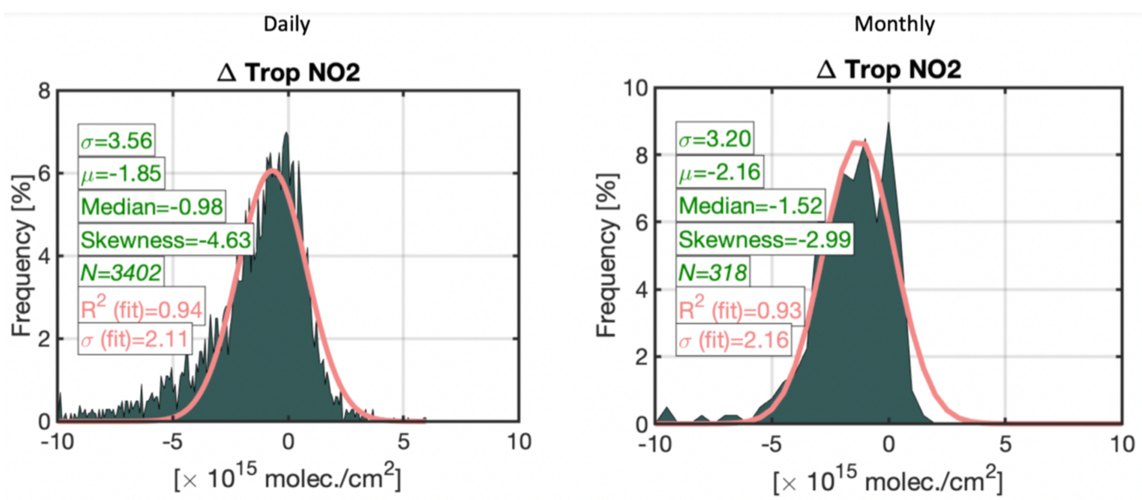
Modifications

We added: "Verhoelst et al. (2021) rigorously studied the potential root cause of some discrepancies between MAX-DOAS and TROPOMI. An important source of error stems from the fundamental differences in the vertical sensitivities of MAX-DOAS (more sensitive to the lower tropospheric region) and TROPOMI (more sensitive to the upper tropospheric area)."

- Total error (Section 3.8)
    - The different error terms are combined into a total error. However, only assumed random components of uncertainties are included (and not systematic ones) so it should be called the total random error. For me, eq. 16 is too some extend trying to combine apples and oranges as the underlying metric in the 3 components is very different and have different meanings.

Answer

Concerning the random error versus systematic errors, we disagree that the sigma values obtained from the histograms are purely random. As a matter of fact, the biggest portion of these errors originate from unresolved systematic errors (or relative errors) in the retrievals. For instance, a relative bias error in the surface albedo manifests in varying biases (systematic) in the retrieval resulting in a large dispersion in the histograms. To better demonstrate this, we recreated the TROPOMI-MAXDOAS histogram based on the monthly-basis observations (Figure S9), and we observed the standard deviation barely changing. If the dispersion were purely random, we would see them going down by 1/sqrt(number_samples). This is why in the beginning of this section, we stated the errors were total uncertainties. Because we do not repeat the same experiment, and the underlying root cause of the errors is unknown, it is impossible to single out the systematic errors from the random errors. The median in the histogram only explains the median of the systematic biases which can vary from pixel to pixel.



In terms of combining different errors, we agree that each error is different in nature, but they all correspond to one quantity, the ratio. Each error explains the extent of information in the satellite-based columns that can become unavailable. Therefore, the sum of them is a good metric to know the combined error. In the satellite retrieval, it is common to have an additive squares errors of different errors such as the temporal representativity, the spatial representativity, the smoothing error, the aggregation error, and the radiative transfer model (RTM) error parameter (Rodgers, 2000). Each of these individual error components come with different meaning but they are translated to one unified quantity. To be able to compare the magnitude of these errors to each other, it is critical to know the combined error which we assumed that their squares are additive. This helped us recognizing the fact that the retrieval error is the largest obstacle in applying the ratio in a robust manner. A very recent group from the same project observed large errors in the satellite retrievals (https://amt.copernicus.org/preprints/amt-2022-237/).

Modifications

To elaborate on the random vs systematic errors we added:

"This fitted normal distribution (R^2=0.94) is used to approximate $\boldsymbol{\sigma_{NO_2}}$ for different confidence intervals and to play down blunders. To understand how much of these disagreements are caused by systematic errors as opposed to random errors, we redo the histogram using monthly-based observations (Figure S14). A slight change in the dispersions between the daily and the monthly-basis analysis indicates the significance of unresolved systematic (or relative) biases. This tendency suggests, when conducting the analysis on a monthly basis, the relative bias cannot be mitigated by averaging. Verhoelst et al. (2021) rigorously studied the potential root cause of some discrepancies between MAX-DOAS and TROPOMI. A important source of error stems from the fundamental differences in the vertical sensitivities of MAX-DOAS (more sensitive to the lower tropospheric region) and TROPOMI (more sensitive to the upper tropospheric area). This systematic error can only be mitigated using reliably high-resolution vertical shape factors instead of spatiotemporal averaging of the satellite data."

Regarding the total error,

"The ultimate task is to compile the aforementioned errors to gauge how each individual source of error contributes to the overall error. Although each source of error is different in nature, combined they explain the uncertainties of one quantity (FNR) and can be roughly considered independent; therefore, the combined error is given by:"

We also changed the total error to the combined error to emphasize that this is simply a linear combination of error:

"To build intuition in the significance of the errors above, we finally calculated the combined error in the ratio by linearly combining the root sum of the squares of the TROPOMI retrieval errors, the…"

We also mentioned our new study too:

"This experiment suggests a standard deviation of 9.4 $\times 10^{15}$ molec./cm^2 with which we again observe the retrieval error to be the largest contributor (>80%) of the combined error (Figure S10). A recent study (Johnson et al., 2022) also suggests that retrieval errors can result in considerable disagreement between FNRs between various sensors and retrieval frameworks."

Minor points:

- Please make sure that all acronyms and abbreviations are spelled out when used for the first time (e.g. NOx, P(O3), DISCOVER-AQ, PAN, VOC, SENEX, SZA, …)

Answer
Sure. We reread the draft and made sure they are spelled out.

- 4, l149: …FNR from **a** chemistry perspective…

Answer
Done.

- 5, l.188: heterogenous chemistry is not considered -> can you add a statement on the importance of that assumption on the study.

Answer
We already mention this is not a major concern in the original version of the manuscript: "Brune et al. (2021) provided compelling evidence showing that the consideration of the HO_2 uptake would make the results significantly inconsistent with the observations suggesting that the HO_2 uptake may have been inconsequential during the campaign. "

- 5, l.206: hv -> h. and define h and (nu)

Answer
Defining them will make the sentence difficult to read so we decided to remove +hv. The photolysis rates of X is meaningful enough.

- 6, eq. 1-3: define k and M, state what the sum is summing up

Answer
Defined.

- 6, l. 239: unconstrained observations -> independent observations

Answer
Changed.

- 6 l. 255: contrary to **an** overestimation in clean ones

Answer
Changed.

- 7, l.262: of NO in **the** chemical mechanism

Answer
Changed.

- 7, l.262: some of **the** oxygenated VOCs

Answer
Changed.

- 7 l264: with larger PAN because -> with larger PAN **mixing ratios** because

Answer
Changed.

- 7, l.277: to reproduce HO2 with -> to reproduce HO2 **mixing ratios** with

Answer
Changed.

- 7, l. 286: 0.62 10^6 cm-3 -> 0.62 x 10^6 cm-3

Answer
Changed.

- 7 l. 288: at least virtually representative -> what do you mean by 'virtually'?

Answer
We meant roughly. Changed to roughly**.**

- 7, l. 291: an analytical solution suggesting… -> solution to what?

Answer
We removed it.

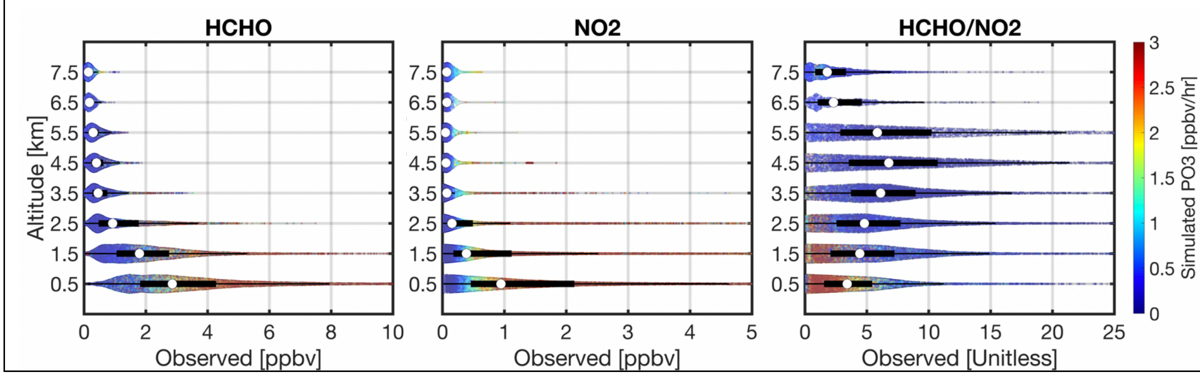
- 8, l. 328: PO3 -> this has been written as P(O3) before.

Answer
We standardized them as PO_3.

- 10, l.399-402: I don't clearly see this larger decrease in NO2 than of HCHO. The media value of the ratio in Fig.5 is more or less 5 with some variability.

Answer
Up to 5 km, the median moves to higher values. This tendency has been well documented in Jin et al., 2017 and Schroeder et al., 2017.

If we combine the data into 1 km layer thickness, the trend will stand out clearly but we want to also keep the variability:



- 31: figure :3 the 3 green lines are very hard to distinguish.

Answer
Thanks we have changed the color.

- 37, Figure 9: I assume the y-axis is not given in %

Answer
Yes, we forgot to multiply to 100. Corrected.