The study by Souri et al. uses a combination of a box model, CTM output, satellite data and aircraft data to try and estimate ozone production rates in the lower boundary layer. The aim of the paper definitely sits within the remit of ACP and TOAR-II, but I believe the manuscript needs major corrections (though mainly textual) are required before it can be accepted for publication.

We thank this reviewer for their constructive comments. Our response is as follows:

Major comments:

1. So, when I read the title and abstract of the paper, it read as if the satellite data was the main dataset/resource used to general the PO3 maps. However, from a detailed read of the manuscript, lots of other data sources are required to achieve the outcome. For instance, the authors use model output from a CTM to derive boundary layer satellite NO2 and HCHO products. This is fine but you have moved a long way from "using satellite data". However, the bias correction of the TROPOMI HCHO and NO2 using surface column measurements is a good practical step to achieve more robust results. There is then the box-model, which is partial tuned to aircraft observations for some tracers, to evaluate key variables which will go into the final scheme (shown nicely in Figure 2) to derive PO3. Overall, I am happy with the methods used to derive PO3 (especially as the satellite data gives you the high spatial resolution) but I think the actual overarching method of the paper needed rewriting (i.e. instead of putting the emphasis on "using satellite data", I think you should make it clearer that you use "a synergy of data products" to derive high spatial maps of PO3.

Response

We agree with the reviewer that the product is not something that solely relies on satellite radiance. As a matter of fact, even satellite retrievals are derived from a combination of models, auxiliary data (albedo, snow/ice information ,…), and the satellite radiance info. To incorporate this valid point, we have done several adjustments to the text:

Modifications

We have renamed the title:

"Feasibility of robust estimates of ozone production rates using a synergy of satellite observations, ground-based remote sensing, and models"

Because we use various models (M2GMI, NCAR TUV, and F0AM) we decided to use a generic name (models) in the title.

In the introduction we added:

"Inspired by those works, we developed a novel product using TROPOMI observations in conjunction with ground-based remote sensing and atmospheric models to estimate PO_3 *and associated errors within the planetary boundary layer (PBL) across the globe."*

In the summary section:

"Providing data-driven and integrated maps of ozone production rates (PO_3*) using a synergy of satellite retrievals, ground-based remote sensing, and atmospheric models enabled us to generate the first satellite-informed product of this kind, offering extensive spatial coverage with significant applications in atmospheric chemistry."*

In a figure caption, we removed a sentence and replaced it with:

"These estimates are based on the proposed algorithm integrating TROPOMI, ground-based remote sensing, and atmospheric models, to estimate PO_3 based upon a statistical approach."

2. In sections 2.2 and 3.1, there are multiple equations but some of the variables are not actually defined, which made it difficult to fully understand the methods without being an expert. So, these sections need to be improved to clearly define and explain what all the variables are in the equations. For instance, on line 161, what are a, b and ε? On line 162, what does i represent? There are also several examples where variables have not been added to the equations (e.g. line 175…I assume these are superscript 2s?). Overall, the method's presentation needs to be improved and discussed more to make it clear to non-experts what you are using the methods for. There are examples below in the Minor Comments supporting this.

Response

Thanks for this comment, we now have improved these sections to better define the equations/methods.

Modifications

We added more description of the variables in Section 2.2.1:

"*To achieve an optimal linear fit ($y = ax + b + \varepsilon$) between the paired observations, where a and b are slope and offset to be determined, we follow a Monte-Carlo Chi-squares minimization such that $\chi^2 = \sum \frac{[y - f(x_i, a, b)]^2}{\sigma_y^2 + a^2\sigma_x^2}$ is minimized. In this equation, σ_y^2 and σ_x^2 are the variances of y (TROPOMI) and x (the benchmark, here MAX-DOAS or FTIR), respectively; i is the subscript refers to i-th observation point, and f is the proposed linear fit subject to optimization.*"

In Line 175, we can either describe the variable or the squares of the variable; it is not necessary to describe the math operation in equations; but to increase the clarity we added:

"*Since there are errors associated with this adjustment resulting from instrument and representation errors, we augment errors of the slope and offset to the total error and label them constant errors (e_{const}) via:*

$$e_{const}^2 = e_{offset}^2 + e_{slope}^2 \times VCD_{bias-corrected}{}^2 \quad (2)$$

where e_{offset}^2 and e_{slope}^2 are squares of errors of offset and slope calculated from the linear regression (Eq. 1). Ultimately, the sum of all three errors constitutes the total errors given:

$$e^2 = e_{const}^2 + \frac{1}{m^2}\sum_{i=1}^{m} e_{random,i}^2 \quad (3)$$

where m is the number of samples for a given grid and timeframe and e_{random}^2 is squares of random errors."

In section 3.1, we added:

"To achieve this, we can use LASSO (least absolute shrinkage and selection operator) (Tibshirani, 1996) consider a regression,

$$Y = X\beta + \alpha + \varepsilon \quad (5)$$

with response $Y = (y_1, \ldots, y_n)^T$, $n \times p$ explanatory variables X, coefficients $\beta = (\beta_1, \ldots, \beta_p)^T$, an intercept α, and noise variables $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$. n is the number of data points, and p is the number of explanatory variables. We can label the regression model sparse when many of β values are zero, and we can label it high dimensional when $p \gg n$. LASSO attempts to select variables such that the following cost function is minimized:

$$(\hat{\alpha}, \hat{\beta}) = argmin\left\{\|Y - X\beta - \alpha\|_2 + \lambda \sum_{i=1}^{p} |\beta_i|\right\} \quad (6)$$

where $\hat{\alpha}$ and $\hat{\beta}$ are optimized intercept and coefficients, λ is a non-negative regularization factor subject to tuning, i is the subscript of the i-th explanatory variable, and $\|.\|_2$ is the L2-norm operator."

Minor Comments:

Line 100: Define NASA and NOAA in the first instances.

Response
We defined them.
Modifications
To study PO_3, we use various aircraft observations from several National Aeronautics and Space Administration (NASA) and National Oceanic and Atmospheric Administration (NOAA) atmospheric composition campaigns.

Line 153: "a fixed additive component that is magnitude-independent", can you provide more detail on what this is and why you use it.

Response
This is the offset derived from the comparison of TROPOMI and the ground remote sensing data. A uniform error that exists everywhere in the scene and it does not vary with the VCD magnitudes.
Modifications
We added: To propagate TROPOMI retrieval errors to the PO_3 product and to remove potential biases, we assume three origins for errors: i) random errors resulting from instrument noise, ii) a fixed additive component that is magnitude-independent (i.e., a uniform offset persisting over all pixels), and iii) unresolved systematic biases that are multiplicative and irreducible by oversampling.

Line 163: Can you give an example of what you mean by "benchmark".

Response
We added FTIR or MAX-DOAS in the parenthesis.
Modifications
… and *x* (the benchmark, here FTIR or MAX-DOAS), respectively.

Line 154: Can you use the satellite column precision as a representation of random errors?

Response
Random errors are mostly dictated by the random errors in the slant column fit. In fact, AMF do not have significant random errors as most of its inputs rely on models (RTM and CTM) or averaged values (except for the O2-O2 algorithm). Here, by the random errors, we strictly refer to errors coming from the pixel SNR (depending on the scene radiance and the instrument noise) and how strong the absorption lines for NO2 and HCHO molecules are depending on their vertical distributions/magnitudes. Both of these components can be well approximated by the error of the fit in SCD projected onto VCD using AMF. The information about this error varying by pixel to pixel is articulated by the precision error variable coming with the L2 product.

Line 156: Can you be clearer on the text "Moreover, to mitigate this error, its squares are average over a month".

Response
Thanks we have clarified it. The random noise gets beaten down by 1/sqrt(n), where n is the number of available pixels in a month for a given area.
Modifications
Moreover, we average the squares of random errors over a month to reduce random noise by the squared number of pixels available at the same location.

Line 161/2: Please define what a, b, ε and i are.

Response
We now defined them (mentioned above).

Line 175: Missing variables in boxes.

Response
Corrected.

Line 184: What does BRDF stand for?

Response
The bidirectional reflectance distribution function.
Modifications
"The product has outperformed traditional LER products such as OMI when both were compared to MODIS surface the bidirectional reflectance distribution function (BRDF) results (Tilstra et al., 2024)."

Section 2.4: Please provide more information on how you use the MERRA2-GMI data to generate the satellite PBL product?

Response
Thanks, we added the equation.
Modifications
To carry out the conversion, we apply the following conversion factor (γ) to the TROPOMI VCDs:

$$\gamma = \frac{\bar{q}_{PBLH}}{\frac{NA}{g \times Mair} \sum q dp} \quad (4)$$

where $\bar{q}_{PBLH}$ is the average of the target trace gas mixing ratios in the PBLH, g is the acceleration of the gravity (assumed 9.81 m/s^2), NA is the Avogadro constant, $Mair$ is the air molecular weight (assumed 28.96 g/mol), q is the target trace gas mixing ratio at a given altitude, and dp is the thickness of each model vertical grid box in hPa. The denominator in Eq. 4 represents the modeled VCD. We integrate modeled partial VCDs up to top of the atmosphere for HCHO, and up to the tropopause pressure layer for NO_2.

Line 217: What do n and p represent?

Response
We now have defined them.
Modifications
n is the number of data points, and *p* is the number of explanatory variables.

Equation 5 RHS: Should "2" be superscript instead of subscript?

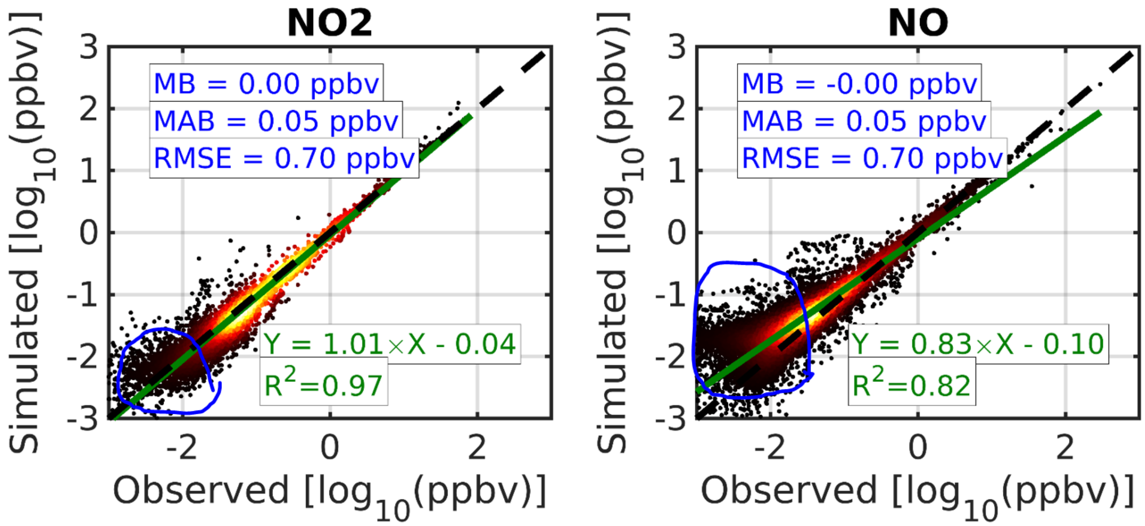
Response
No, ||.||2 is the notation for the L2-norm operator. We now have clarified it.
Modifications
... and $\|.\|_2$ *is the L2-norm operator.*

Line 227: Please make it clearer what you mean by "folds".

Response
This is a generic term used in cross-validation algorithms. We clarified it.
Modifications
To optimize this value, we discretize λ in 100 values between 10^{-4} *up to* 10^{1}*, divide the training dataset into 10 folds (i.e., spliting the dataset into equal size segments),*

Lines 353/354: "Consequently, it is likely that the measurements error resulted in more spread in the comparison". Can you provide some references to support this statement.

Response
The random noise associated with NCAR's NO_2 and NO measurements are reported to be around 0.05 and 0.01 ppbv for NO_2 and NO, respectively. In the log-space, they will be around -1.3 and -2.0. So, the reason that we see a fatter distribution in the comparison over pristine areas is that the uncertainty of the measurements go beyond 100% (blue circles):



We think it is important to acknowledge that NO and NO2 measurements in remote regions (or high altitudes) can be highly uncertain (Shah et al., 2023).

Modifications

We added:

"Consequently, it is likely that the measurements error resulted in more spread in comparison. In particular, Shah et al. (2023) found that these measurements could be contaminated by various reactive nitrogen species in remote regions precluding a robust validation of atmospheric models."

Line 364: Rephase "not to unrealistic" to "reasonable".

Response

Corrected.

Figure 3: NO2 and NO have the same MB, MAB and RMSE. Is this a duplicate of statistics or coincidence? Also, some of the stats legends overlap (e.g. OH), so the presentation needs to be improved here.

Response

We doubled checked the code. The stats are calculated by a function inside the loop iterating over each specie. They are the same within 2 decimal point precision. They are not identical. We have recreated the figure to remove the overlaps.

Line 555: One could argue why don't use just use a CTM or regional model to simulate/output PO3 and supporting variables (e.g. NO2 and HCHO). Would you not benefit from using the satellite and aircraft observations to evaluate the model, identify limitations (e.g. emissions, chemical mechanism etc.), undertake sensitivity experiments to resolve the limitations and then provide more robust estimates of PO3 from the model? That way, you are getting estimates of PO3 but also improving the processed based model providing a better understanding of the processes governing PO3?

Response
This is a valid point. It's more physics-based to incorporate satellite observations into a model to optimize the model's prognostic inputs. This allows us to see the chain of adjustments on various physiochemical processes within a process-based framework. Souri et al. 2020 (https://acp.copernicus.org/articles/20/9837/2020/) was a pioneering work that adjusted NOx and VOCs emissions simultaneously using multi-sensors to better represent PO3 across East Asia. This was motivated by the importance of the chemical feedback between NOx-VOC and HCHO-NOx. However, as we mentioned in the introduction (the paragraph starting with "While the characterization of ozone regimes …"), it is prohibitively expensive to perform joint inversions globally and for a long-term record. Data-driven approaches, like the one described in our work, provide a shortcut. It will neither replace a constrained chemical transport model capable of providing all physiochemical processes and reaction rates, nor will it be a product to understand ozone chemistry. It is simply an estimate of PO3 maps along with the contribution ones that can provide more detailed information compared to binary maps obtained from FNR. As we mentioned in the summary, our parametrization can be enhanced through using more sophisticated algorithms that are better capable of capturing the non-linear chemistry associated with ozone. An upcoming part of our project will demonstrate the use of deep-neural networks that have been shown to predict PO3 and the derivatives without the need for FNR. We believe these new statistical approaches can provide rapid results for regulators to implement emission mitigation in a timely manner, and to prioritize sub-orbital missions over places where in-situ measurements are absent.

Line 560: There are a few instances where you term "significant". However, do you actually use a statistical test to support these statements?

Response
No, we unfortunately didn't use a statistical test. So we removed this term whenever we say something might be significantly different than the other term.

Figure 11: I might have missed this but do you define "CONUS"?

Response
Now defined in the beginning of the paragraph.

Figure 17: "The data is based on 2019 TROPOMI observations". This does not make sense. You list several variables on Lines 652 only which two are actually from TROPOMI. Please update this.

Response
We have modified this sentence along with any other sentences that may wrongly imply that the estimates were purely derived from satellite radiance.
Modifications
"These estimates are based on the proposed algorithm integrating TROPOMI, ground-based remote sensing, and atmospheric models, to estimate PO_3 based upon a statistical approach."

Line 678: I disagree with this statement "satellite-derived product". As to my major comment #1, you use satellite data, aircraft data, MAX-DOAS data, CTM data, box model data and statistical methods to derive PO3 (as depicted in your Figure 2). Therefore, I believe this needs to be reworded and refocussed (e.g. a data-model fusion approach to derive PO3 etc.).

Response
We modified this part. A large fraction of these estimates come from the satellite information so we believe the "satellite-informed" attribute of our product should be highlighted.
Modifications
"Providing data-driven and integrated maps of ozone production rates (PO_3) using a synergy of satellite retrievals, ground-based remote sensing, and atmospheric models enabled us to generate the first satellite-informed product of this kind, offering extensive spatial coverage with important applications in atmospheric chemistry."