The manuscript presents a new global dataset of net ozone production rates (PO3) derived from a neural network framework constrained by satellite observations. The authors develop a deep neural network (DNN) model trained with simulations from the F0AM box model and aircraft measurements with perturbations. The DNN employs as input a set of geophysical parameters derived from multiple sources, including satellite retrievals of HCHO and NO2 (from OMI for 2005–2019 and TROPOMI for 2018–2023), as well as parameters from the MINDS model. The MINDS framework provides the conversion factors from total column to planetary boundary layer (PBL) mixing ratios for HCHO and NO2, simulated O3, and water vapor (H2O).

The authors validate their DNN-derived PO3 product (termed PO3DNN) against the empirical formulation described in Souri et al. (2025). The manuscript further examines several applications: (i) the intercomparison of OMI- and TROPOMI-based products for 2019, (ii) regional PO3 seasonality (2005–2007) across selected sites worldwide, (iii) a heatwave case study over the northeastern United States (August 2007), (iv) seasonal and spatial patterns over the Middle East (2019), and (v) long-term PO3 trends from 2005–2019 in Los Angeles and Tehran.

Comprehensive uncertainty estimates are presented, including both systematic and random errors, with the dominant source attributed to the column-to-PBL conversion factors from MINDS. The total relative errors are reported to range from about 25% in polluted regions to more than 200% in remote areas.

The manuscript is scientifically interesting and presents a valuable global dataset of ozone production rates derived from satellite observations. However, several key methodological and interpretational issues need to be clarified and better quantified before the study can be fully evaluated. Therefore, I consider the following as major comments.

We thank the reviewer for his/her constructive comments and detailed summary, our response follows.

Major Comments

1. *Time period, harmonization, and trend (2005–2023)*

   The title suggests that the study spans the full period of 2005–2023, implying a continuous long-term trend analysis that combines OMI and TROPOMI data. However, the actual trend analysis uses only OMI data (2005–2019), while TROPOMI is primarily used for 2019 onward and inter-satellite comparison. The manuscript should clarify how the two products were harmonized for consistency and provide quantitative evidence of their agreement or bias (for example, regional mean differences, temporal overlap). It would also strengthen the study to explicitly show the magnitude and spatial or temporal characteristics of any applied corrections and to quantify how harmonization affects the derived PO3 trends (for example, OMI-only vs. TROPOMI-only vs. combined). Finally,

please discuss whether a unified 2005–2023 trend is feasible and what systematic offsets might influence its interpretation.

Response

We acknowledge that the title may create confusion about whether both products (TROPOMI and OMI) have been fully harmonized for long-term trend applications.

Harmonization is inherently subjective and depends on user-defined tolerance levels. While our products demonstrate consistency within 10% during a joint year (2019), whether this level of agreement is sufficient for robust trend calculations depends on both the user's tolerance requirements and the magnitude of the trends being analyzed.

Both products use identical models for conversion factors, photolysis rates, water vapor calculations, and a forward DNN estimator. The primary differences come from variations in NO2 and HCHO VCDs.

Complete harmonization requires consistent retrieval algorithms across both TROPOMI and OMI, including:

- **Identical RTMs with consistent assumptions for O2-O2 algorithms, surface properties, ocean properties, and so forth.**
- **Uniform slant column fitting approaches (using identical spectral windows, cross-sections, and fitting methods; whether DOAS or BOAS)**
- **Consistent a priori assumptions**

This needs a substantial undertaking. Well-established, long-term projects such as NASA's MEaSUREs and QA4ECV have been developed specifically to create consistent algorithms for robust long-term analysis. To our knowledge, such comprehensive harmonization has not yet been achieved for TROPOMI, and our current budget constraints prevent us from harmonizing the satellite VCDs to this extent.

We have proactively implemented robust bias correction against ground-based remote sensing observations, which is a critical component of data harmonization. This approach prevents both satellites from diverging significantly from established benchmarks as a first-order approximation, resulting in our products' 10% agreement (Figure*). Whether this makes up sufficient "data harmonization" ultimately depends on users' requirements for trend derivation robustness.

A harmonization effort should be part of a "post-processing" step. Even if we had implemented additional harmonization approaches for deriving 2005-2023 trends (an objective of our final ACMAP-Aura project year), users would still need to apply their preferred harmonization method, as no single standardized approach exists.

An important factor in trend analysis is that TROPOMI's long-term stability for HCHO and NO2 measurements has not been as thoroughly validated as OMI's record. Recent studies have documented potential drifts in TROPOMI data products that warrant further investigation (https://amt.copernicus.org/articles/17/3969/2024/).

In summary, we are unable to fully harmonize TROPOMI and OMI NO2 and HCHO retrieval algorithms to create a consistent PO3 product. While the application of the bias correction using MAX-DOAS/FTIR has made both products agree well within 10%, we think the users may need to apply a harmonization algorithm as a post-processing step. We need to provide this caveat and limitation in the summary section.

Modifications

We added this paragraph to the summary:

While the OMI- and TROPOMI-based PO_3 products maintain algorithmic consistency in several key components, including photolysis rates and water vapor, the underlying satellite retrievals of HCHO and NO_2 VCDs remain unharmonized between the two instruments. To address the resulting inter-instrument biases, we implemented bias correction using ground-based remote sensing retrievals as reference standards. This approach achieved OMI and TROPOMI PO_3 agreement within 10% on average. However, this level of consistency may be insufficient for robust joint trend analysis of the combined OMI-TROPOMI PO_3 record over areas with non-linear or small trends, potentially requiring the implementation of trend harmonization algorithms (e.g., Hilboll et al., 2013) to warrant statistical reliability in long-term analyses.

2. *Bias correction using MAX-DOAS*
   The bias correction procedure for OMI and TROPOMI retrievals using MAX-DOAS observations needs more detail. Were corrections applied globally or regionally, and are they time dependent? How large were the typical corrections? The error treatment also appears simplified for bias correction and may underestimate correlated uncertainties.

Response

The reviewer is right about the fact we did not account for correlated uncertainties among HCHO and NO2 VCDs in the error budget or the linear fit between satellites vs. benchmarks (although the errors in *x* and *y* are considered in Souri et al. 2025 based on weighted chi2 minimization).

TROPOMI HCHO bias correction comes from Souri et al. (2025) who expanded the analysis of Vigrouroux et al. (2021) using FTIR measurements. A similar work was recently done with OMI HCHO with the same dataset (Ayazpour et al., 2025) whose correction factors were used in our work. OMI NO2 correction factors were derived from a established work done by Pinardi et al. (2021), and TROPOMI NO2 follows the work we did in Souri et al. (2025) comparing MAX-DOAS and the satellite observations based on an extension to Verhoelst et al. (2020).

We added a new section right after introducing TROPOMI and OMI retrievals to elaborate on the number of stations, duration, and the magnitude of these corrections.

The ground remote sensing data used are global.

The reviewer raised concerns about the generalizability of our benchmarks, mentioning their sparse distribution and potential for varying satellite discrepancies across seasons and locations. However, numerous validation studies (Verhoelst et al., 2020; Vigouroux et al., 2021; Ayazpour et al., 2025; and Table 1 in https://acp.copernicus.org/articles/21/18227/2021/) have demonstrated that biases in TROPOMI and OMI NO2 and HCHO columns consist of two components: an additive term (offset that exists uniformly regardless of season or location) and a multiplicative term (magnitude-dependent slope).

The rationale for parameterizing retrieval biases as a function of magnitude is to enhance correction factor generalizability across seasons/locations. By understanding how bias changes with magnitude, we can predict seasonal bias variations since column densities vary seasonally.

A remaining question is whether these slopes and offsets remain consistent across different locations and seasons? This is precisely why we included bias correction error terms in our analysis. If the relationship between benchmarks and satellite retrievals varied dramatically by location or season, linear fits would become highly uncertain, resulting in large coefficient errors.
Figure 8 and the validation studies referenced above (https://acp.copernicus.org/articles/25/2061/2025/acp-25-2061-2025.pdf) suggest the opposite, indicating that these parameters are mostly reliable. Similarly, if we had insufficient samples, this would have resulted in larger uncertainties associated with the slopes and offsets.

Lastly, would adding a new instrument over a different area change the correction factors? This question would fall into the area of "unknown unknown". We won't know until we measure. The current correction factors applied are based on the most recent and credible validation efforts (the known known).

Modifications

We clarified that the correlated errors aren't considered in the error characterization:

It is important to acknowledge that the defined total error budget here is only a good guess and optimistic. Some underlying sources of error, which are difficult to quantify, are not included. For example, errors related to the training dataset derived from the F0AM model are challenging to assess because of the lack of PO_3 measurements. We assume other inputs to the PO_3 parametrization, such as the monthly climatology TROPOMI surface albedo to be error-free. Additionally, all datasets used to estimate PO_3 contain spatial representation errors (Souri et al. 2023), which are difficult to measure without knowing their true state of global spatial variability. Moreover, we do not consider correlated errors among HCHO and NO_2 retrievals.

We added a new section right after defining the TROPOMI and OMI datasets:

1.1.1. Bias correction using ground-based remote sensing data

In order to remove large biases in both TROPOMI and OMI products, we bias correct their columns using the offset (additive term) and slope (multiplicative term) determined from a linear fit to paired MAX-DOAS/FTIR and these datasets, as described by Souri et al. (2025). The rationale for defining retrieval biases as a function of magnitude is to enhance correction factor generalizability across seasons and locations. We take advantage of three studies characterizing the bias correction factors, listed in Table 1. The application of these correction factors yields consistency across OMI and TROPOMI NO_2 and HCHO columns within 10% (Section 4.4.4.)

Table 1. The slopes and offsets derived from various validations studies, used to bias correct the satellite retrievals used in the parameterization of PO_3.

Product	**Slope**	**Offset**	**Benchmark**	**Time period of validation**	**Reference**
TROPOMI NO_2	0.59	0.90×10^{15} molecs/cm^2	Global MAX-DOAS observations	2018-2023	Souri et al., (2025)
TROPOMI HCHO	0.66	0.32×10^{15} molecs/cm^2	Global FTIR observations	2018-2023	Souri et al., (2025)
OMI NO_2	0.83	0.26×10^{15} molecs/cm^2	Global MAX-DOAS observations	Varies for each station spanning from 2010-2018	Pinardi et al., (2020)
OMI HCHO	0.79	0.82×10^{15} molecs/cm^2	Global FTIR observations	Varies for each station spanning from 2004-2020	Ayazpour et al., (2025)

3. *Comparison with FNR*

The authors argue that the PO3 framework provides more detailed and continuous information on ozone production and sensitivity than the traditional formaldehyde-nitrogen ratio (FNR). I agree with this conceptual advantage. However, the manuscript does not clearly demonstrate how much additional information PO3 offers beyond FNR in a quantitative or diagnostic sense. It would strengthen the paper if the authors could illustrate specific cases or regions where PO3 reveals gradients or features that are not captured by FNR-based classifications.

The manuscript also describes FNRs as "binary," but it would be more accurate to say

that the interpretation of FNRs is often binary, based on thresholding between VOC-limited and NOx-limited regimes, while the FNR values themselves are continuous. Clarifying this distinction would help avoid oversimplifying the FNR framework.

Finally, the description of Figure 14 qualitatively compares the spatial patterns of PO3 sensitivities to HCHO and NO2. This section could be improved by quantifying those relationships, for example by providing correlation or regression metrics between PO3 sensitivities and FNRs, or by showing how the two indicators diverge under different chemical conditions (for example, high-HCHO/low-NO2 versus low-HCHO/high-NO2). Such quantitative comparisons would make the claimed improvement of the PO3 framework more convincing and scientifically interpretable.

Response

When someone gives us an FNR value (assuming no measurement errors and that HCHO and NO2 perfectly represent VOCR and reactive nitrogen), how do we actually assign a sensitivity value to it in units like ppbv/hr or 1/hr (in case of dPO3/dNO2 or dHCHO)? The main reason we use FNR is to help regulators. What regulators really need from us are the first and second-order derivatives of ozone production rates relative to NOX and VOC emissions, so they can estimate how PO3 will change under different emission scenarios. Where does FNR fit into this? It only tells us the sign of the sensitivities, and by itself it's not enough to translate varying FNR values into actual derivatives or sensitivities.

Going back to the original question: how do we provide these sensitivities given just an FNR value? That requires us to know a lot more about the underlying atmospheric conditions like light levels, humidity, how well HCHO and NO2 represent VOCR and reactive nitrogen, and in some cases heterogeneous chemistry, chlorine chemistry, HONO chemistry, and so on. FNR simply can't capture all these dimensions.

So, can our new product fully meet what regulators need? Not fully; but it is a step forward. It provides first-order derivatives based on how HCHO, NO2, water vapor, and photolysis rates change. We're missing second-order derivatives, and there's a more fundamental hurdle: HCHO and NO2 do not fully represent local emissions because they go through other physical and chemical processes like transport.

That said, we don't think it's very useful to compare an incomplete two-dimensional representation of multidimensional nonlinear ozone chemistry with a product that has more dimensions.

To elaborate more; FNR has three blind spots:

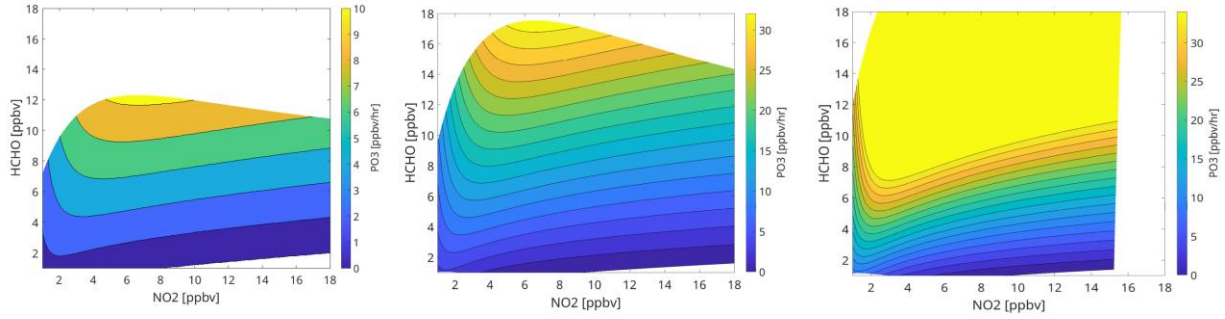
1. **Lack of sensitivity magnitudes: FNR only classifies regimes without quantifying the actual magnitude of ozone sensitivities. For example, if $\partial P_{O_3}/\partial NO_2$ is +10 s^{-1} or +3 s^{-1}, both would be labeled "NOx-sensitive," even though their regulatory implications might be different. What truly matters about emission control is the magnitude of these responses. For this reason, CTM-based calculations (either**

through direct decoupled methods, perturbation or adjoint approaches) are typically used. These, however, require extensive efforts to constrain model inputs with satellite data (see Souri et al., 2020: https://acp.copernicus.org/articles/20/9837/2020/).

Our work provides quantitative first-order sensitivity maps, equivalent to directional derivatives (Appendix A), which is a major innovation of the new algorithm.

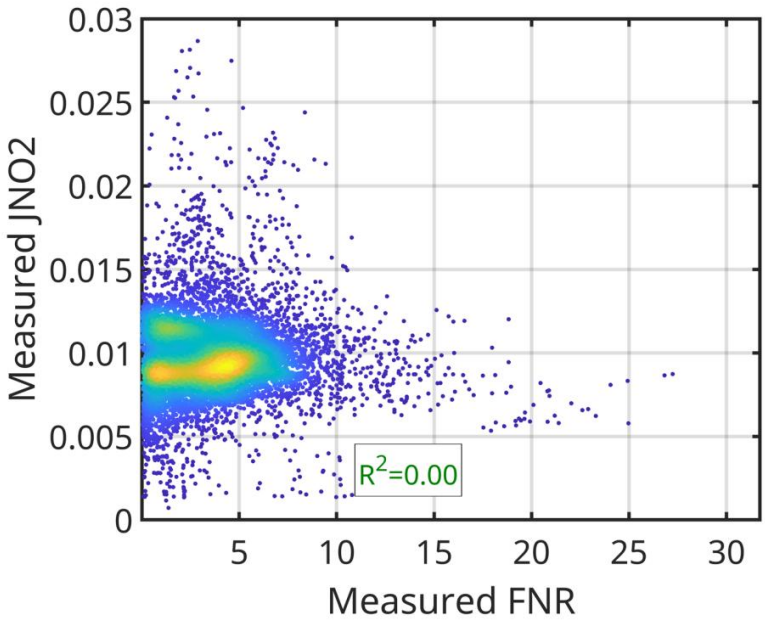
2. **Varying FNR values cannot be directly linked to varying magnitude of sensitivities without accounting for photolysis rates, water vapor, and so forth. These geophysical variables are not articulated by FNR.**

3. **Lack of adequate dimensions: FNR slices the inherently multidimensional, nonlinear system into just two dimensions. To demonstrate this shortcoming, we perturbed photolysis rates over polluted regions during the KORUS-AQ campaign using observationally-constrained F0AM model. Multiplying photolysis rates by factors of 0.5 (dim, left), 1.0 (default, middle), and 2.0 (bright, right) produced three sets of PO3 isopleths.**



The results clearly show that increasing light intensity raises both net PO3 and its sensitivities to NOx and VOC (the contours are more compact in the bright case; each contour corresponds to 3 ppbv/hr).
This means that the same FNR can correspond to entirely different magnitude of sensitivities depending on available light. <u>There is where FNR falls apart.</u>

Although one might expect FNR to indirectly reflect variations in photolysis rates, our analysis of 47,000 data points obtained from KORUS-AQ measurements showed no relationship between measured $j\mathrm{NO_2}$ and FNR:



A similar limitation arises from FNR's inability to account for water vapor effects on PO3. Capturing these complex nonlinear interactions between PO3, light, humidity, and precursor concentrations requires more advanced methods over a simple ratio, lacking any information about light intensity and humidity. In a data-driven framework, this is best achieved using nonlinear parameterizations such as DNNs.

This new product therefore represents a paradigm shift away from oversimplified FNR approaches. It not only provides spatiotemporal sensitivity magnitudes, but also accounts for multidimensional dependencies. We highlight this feature in Section 4.3.

Modifications

To better inform how the new sensitivity maps can eliminate the need for FNR, we added:

"Beyond Binary Maps from HCHO/NO_2: A Deep Neural Network Approach to Global Daily Mapping of Net Ozone Production Rates and Sensitivities Constrained by Satellite Observations (2005–2023)"

While we had provided context about the advances made compared to FNR, we added a paragraph in the introduction describing why we should quantify the multidimensional magnitude of PO3 sensitivity, currently lacking in FNR-based approaches. We added in the introduction:

The overarching goal of producing ozone chemistry sensitivity maps is to inform regulatory agencies about the impact of emission reductions on locally produced ozone. Unlike conventional FNR-based binary maps, these maps must quantify the magnitude of sensitivity rather than merely indicating its direction. This quantitative approach is essential because both

the sign and magnitude of sensitivities are crucial for understanding the impact of emission changes. While detailed sensitivity maps can be derived from chemical transport models by perturbing underlying emissions, the lack of observational constraints on these models can introduce significant biases. Souri et al. (2025) attempted to address this limitation by providing magnitude-dependent sensitivity maps of PO_3 to NO_2 and HCHO using piecewise linear regression. However, their approach yielded derivatives of PO_3 with respect to NO_2 and HCHO that remained invariant with changes in light and humidity conditions. This limitation is problematic because reduced light conditions are known to substantially dampen the sensitivity of PO_3 to NO_x and VOCs, even under identical emission rates. The current work is therefore motivated by the need to capture the complex, multidimensional dependencies of PO_3 on ozone precursors, light intensity, and humidity using a more flexible data-driven approach through a machine learning algorithm. While these maps will not replace process-based chemical transport model experiments, they can efficiently provide first-order assessments to: (i) strategize top-down modeling experiments, (ii) gauge the added value of satellites on predictions of PO_3, and (iii) guide the design of sub-orbital missions in regions with poorly documented elevated PO_3.

In the supplementary, we added a new section describing the fundamental issues with FNR; we did not include it in the main draft because it is more of a reminder for people who may misuse FNR rather than bringing new insights into ozone chemistry.

1. FNR is oblivious to the impact of photolysis rates and water vapor content on PO_3

The primary objective of using the formaldehyde-to-nitrogen dioxide ratio (FNR) is to reduce high-dimensional, non-linear ozone production rates into a two-dimensional framework based on volatile organic compound reactivity (VOCR) and reactive nitrogen. However, beyond the fact that HCHO and NO_2 does not fully represent VOCR and reactive nitrogen, it is crucial to recognize that ozone production rate sensitivities and magnitudes depend on other geophysical variables independent of FNR. Among these variables, photolysis rates and water vapor are major drivers of atmospheric oxidation capacity, modulating numerous reactions related to ozone production (Kleinman et al., 2001).

To demonstrate photolysis rate effects on both PO_3 magnitudes and sensitivities, we conducted F0AM box model simulations constrained by geophysical variables during June 6-9 of the KORUS-AQ campaign (Souri et al., 2025). We perturbed NOx, VOCs, and photolysis rates to generate three sets of isopleths (Figure S1). The results clearly show larger ozone production rates under more intense light conditions. More importantly, the contours corresponding to identical PO_3 intervals (3 ppbv/hr) become more compact under brighter conditions, indicating that PO_3 becomes more sensitive to both NO_X and VOCs with increased light intensity. This pattern suggests that identical FNR values under different photolysis rates can have fundamentally different implications for ozone production rate sensitivities.

To confirm that FNR contains no photolysis rate information, we analyze paired FNR and jNO_2 photolysis rate measurements from over 47,000 data points during the KORUS-AQ campaign, revealing no correlation between these variables (Figure S2). This demonstrates the need for additional dimensions in ozone sensitivity analysis, necessitating more sophisticated algorithms (like our approach) over traditional threshold-based methods.

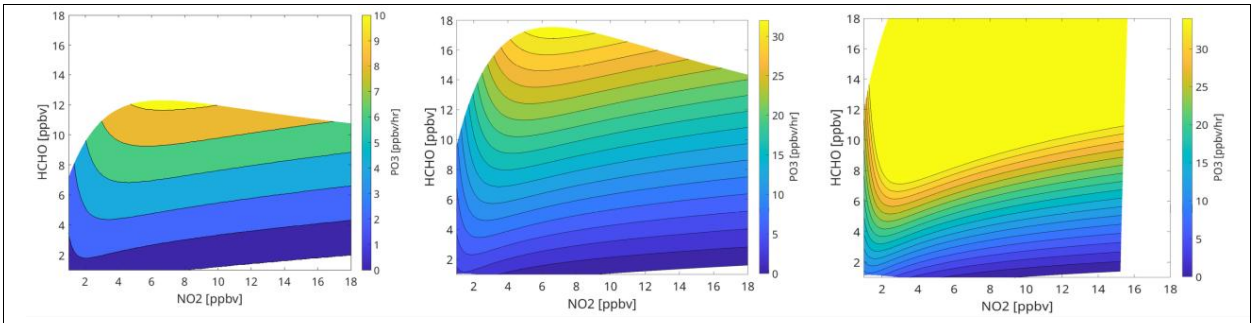


Figure S1. The PO_3 isopleths generated using F0AM box models derived from observations taken during the KORUS-AQ campaign under three different photolysis rates scenarios: (left) multiplied by 0.5, (middle) default, (right) multiplied by 2.0. Each contour represents 3 ppbv/hr.

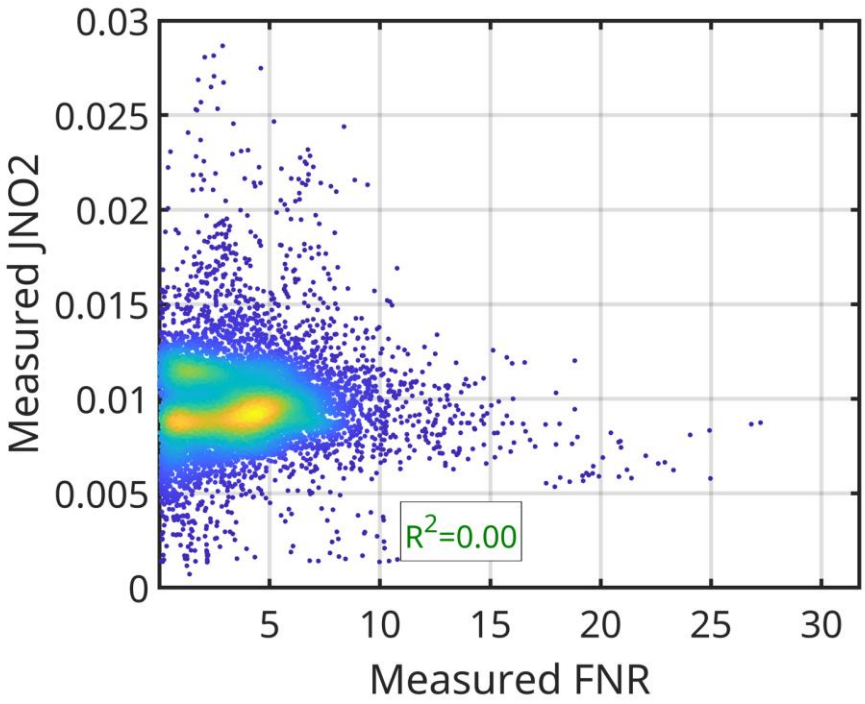


Figure S2. The comparison of measured FNR and measured jNO_2 frequencies taken from aircraft observations during the KORUS-AQ campaigns. All measured points are used to make this plot.

Figure S3 illustrates the representation of ozone sensitivities by mapping five variables derived from TROPOMI and our PO_3DNN parameterization across two seasons over Los Angeles. FNR values are low during colder months due to abundant NO_2 relative to HCHO, qualitatively suggesting the LA region should be predominantly VOC-sensitive. However, the derivatives and sensitivities of PO_3 to both HCHO and NO_2 remain muted due to limited photochemical activity, making PO_3 unresponsive to NO_X and VOC concentrations. Conversely, summer conditions yield larger derivatives, showing much stronger PO_3 responses to both species. This example can be extended to different times of day, such as FNR values

from geostationary satellites or morning versus afternoon measurements from low Earth orbit satellites.

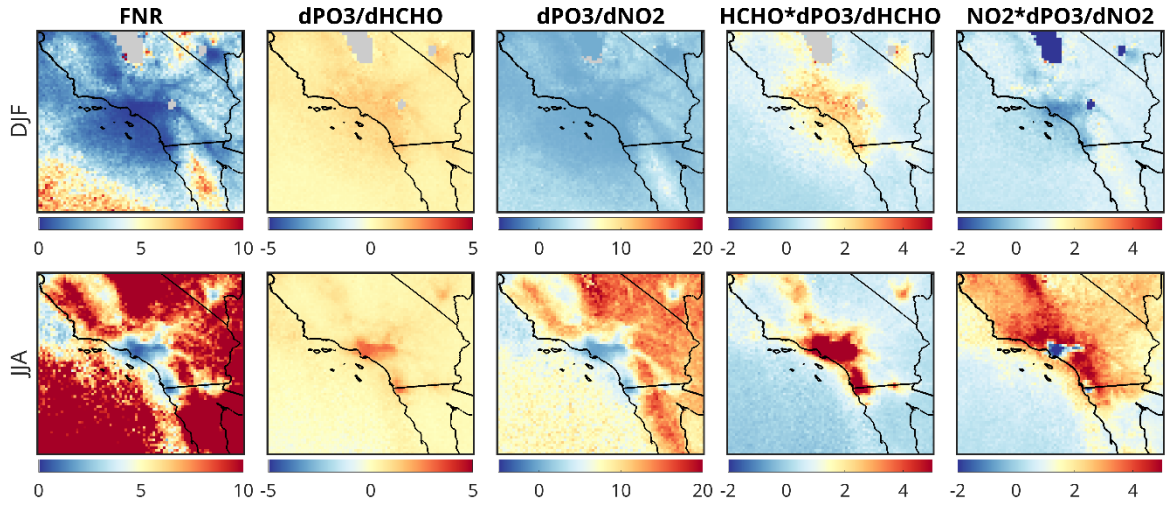


Figure S3. Five variables derived from our PO_3DNN product based on TROPOMI dataset. The first row focuses on December-January-February (DJF), while the second row shows those variables for June-July-August 2023. The calculation of the sensitivities and derivatives are based on perturbation of the DNN algorithm described in the main paper.

The absence of PO_3-relevant geophysical information in FNR also applies to water vapor. F0AM box simulations over polluted regions show that increasing humidity enhances PO_3 through the generation of two OH molecules via H_2O+O^1D reactions (Figure S4). However, FNR contains no water vapor information, as humidity is driven by hydrological and meteorological factors decoupled from the processes determining FNR (Figure S5). This further necessitates adding water vapor as an additional dimension in ozone sensitivity

analysis.

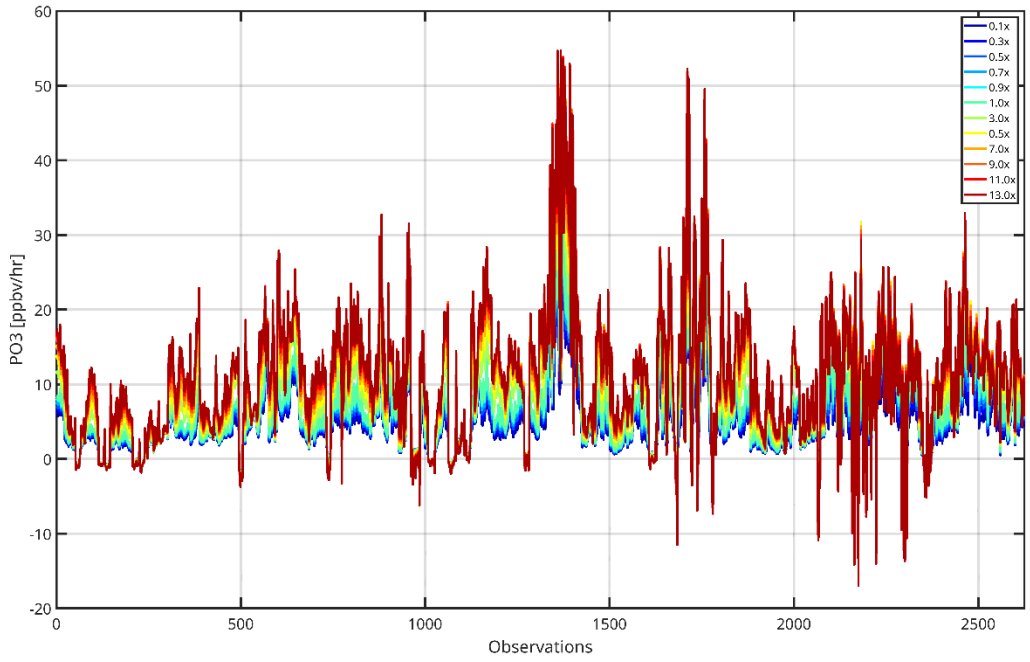


Figure S4. The effect of $H_2O(v)$ on PO_3 during KORUS-AQ campaigns. Only highly polluted regions ($HCHO \times NO_2 > 10$) are selected for this experiment.

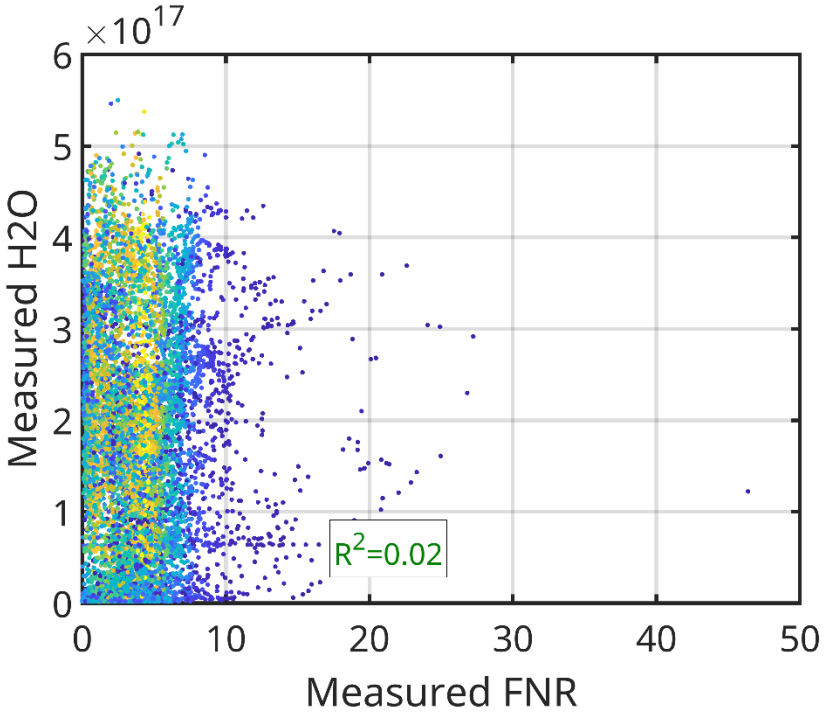


Figure S5. The comparison of measured FNR and measured water vapor density taken from aircraft observations during the KORUS-AQ campaigns. All measured points are used to make this plot.

4. *Conversion factor and averaging kernel*

It is unclear whether satellite averaging kernels were applied when deriving the column-to-PBL conversion factors using MINDS. If they were applied, please specify how; if not, discuss the potential influence on near-surface concentrations and the resulting PO3 estimates.

Response

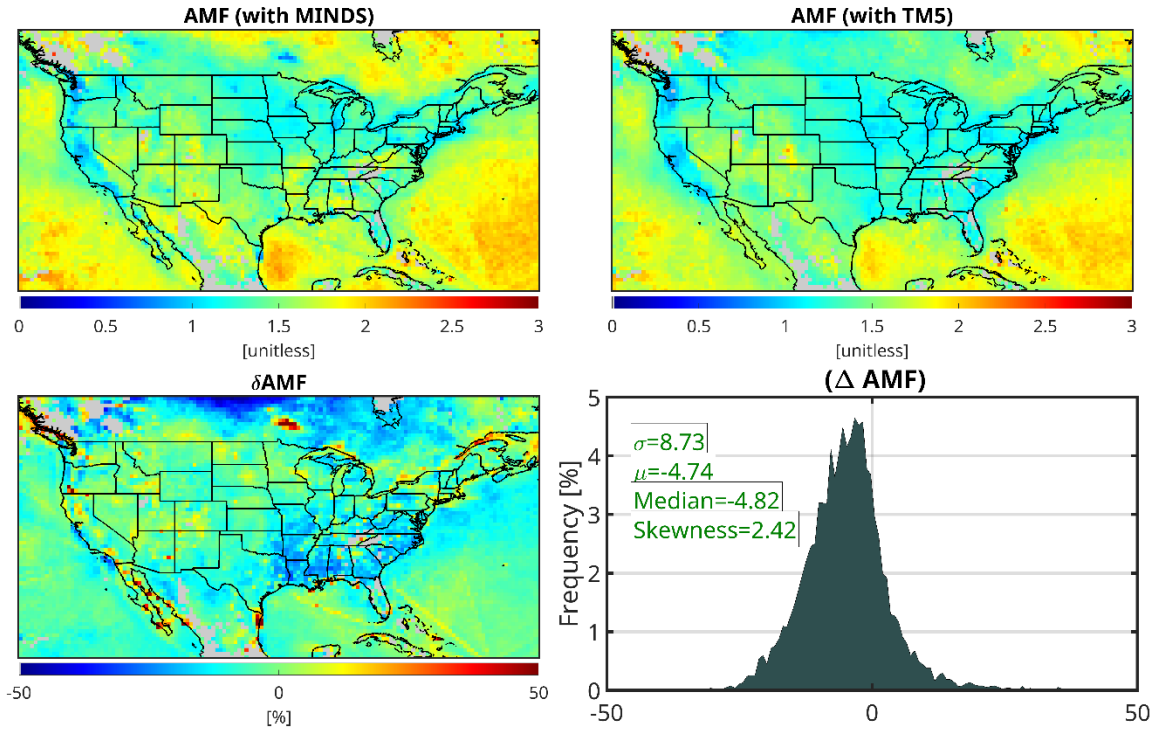
There are two main approaches to remove or mitigate the influence of the a priori assumptions used in OMI and TROPOMI AMFs in order to obtain a consistent, MINDS-driven conversion factor that reflects the satellite vertical sensitivity.

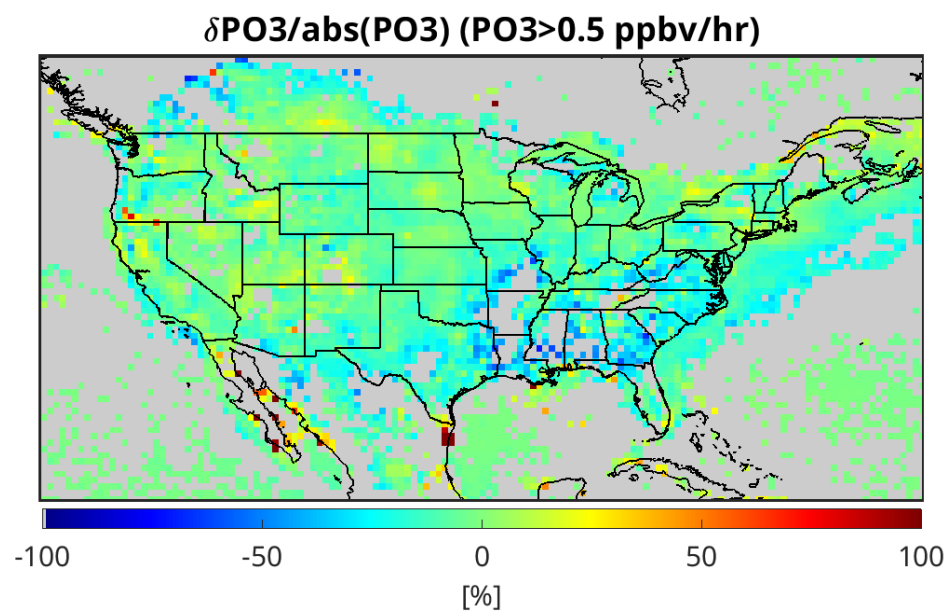
Approach 1: Convolving MINDS Conversion Factors with Satellite Averaging Kernels
In this approach, the conversion factor is defined as
$f_{AK} = q_{PBL}/\sum x$,
where q_{PBL} is MINDS PBL mixing ratio and $x = x_a + A(x_{MINDS} - x_a)$. Here, x_a and x_{MINDS} represent the a priori and MINDS partial columns, respectively, and A is the averaging kernels.
While this method is scientifically sound, it introduces significant complexity: the resulting conversion factor becomes dependent on satellite viewing geometry, scene-specific averaging kernels, and the a priori vertical profiles. This dependency makes validation of the conversion factors against in situ observations extremely difficult.
As noted in the manuscript, the dominant source of systematic error in our product comes from the conversion factors themselves. If these factors are entangled with averaging kernel and a priori uncertainties, they lose generalization and consistency across retrievals and a priori frameworks. By maintaining a sensitivity- and a priori-agnostic formulation (as validated in Appendix B), we ensure that conversion factors can be robustly validated using aircraft observations and applied consistently across models. In other words, the question of "which model does better convert columns to the near surface concentrations?" can be more easily answered without delving into the nuances of satellite sensitivities.

Approach 2: Recalculating AMFs Using MINDS Vertical Shape Factors
This alternative approach recalculates the AMFs using MINDS vertical profiles (section 2.1 in https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2017JD028009), allowing the conversion factor to remain independent of the satellite retrieval. This is a preferred algorithm over approach #1. However, it introduces a circular problem: recalculating AMFs would necessitate revalidating and bias-correcting TROPOMI and OMI NO2 and HCHO columns against ground-based datasets. Repeating the extensive work of Verhoelst et al. (2020), Vigouroux et al. (2021), Pinardi et al. (2021), and Ayazpour et al. (2025) would be a major undertaking.
For these reasons, we chose not to refine the TROPOMI and OMI VCDs using MINDS shape factors at the cost of introducing some biases in our product.

To show the impact of neglecting this step on PO3, we recalculate AMF with MINDS shape factors over the CONUS, an area with varying emissions and meteorological conditions, using our OI-SAT-GMI package (https://github.com/ahsouri/OI-SAT-GMI/tree/main), and quantify the impact on PO3.

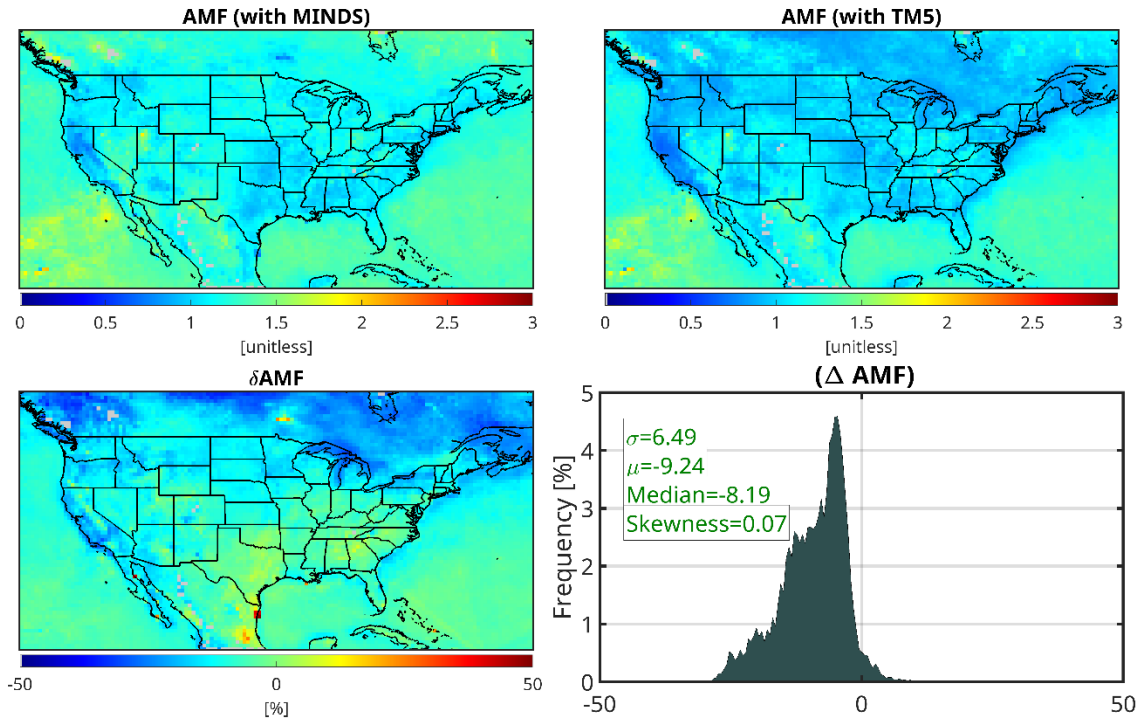
Regarding NO2, we see AMF (thus VCDs and PBL mixing ratios) to vary within 5±9% on average with minimal changes over polluted regions while seeing bigger values over higher latitudes or dark albedo where the retrieval becomes more dependent on the a priori. These changes induce minimal changes on PO3 (<20%) over PO3>0.5 ppbv area, especially hotpots of PO3. The changes can reach to 40-50% over remote high latitude regions, but PO3 errors are already extremely large (>200%) because of the errors in MINDS conversion factors:

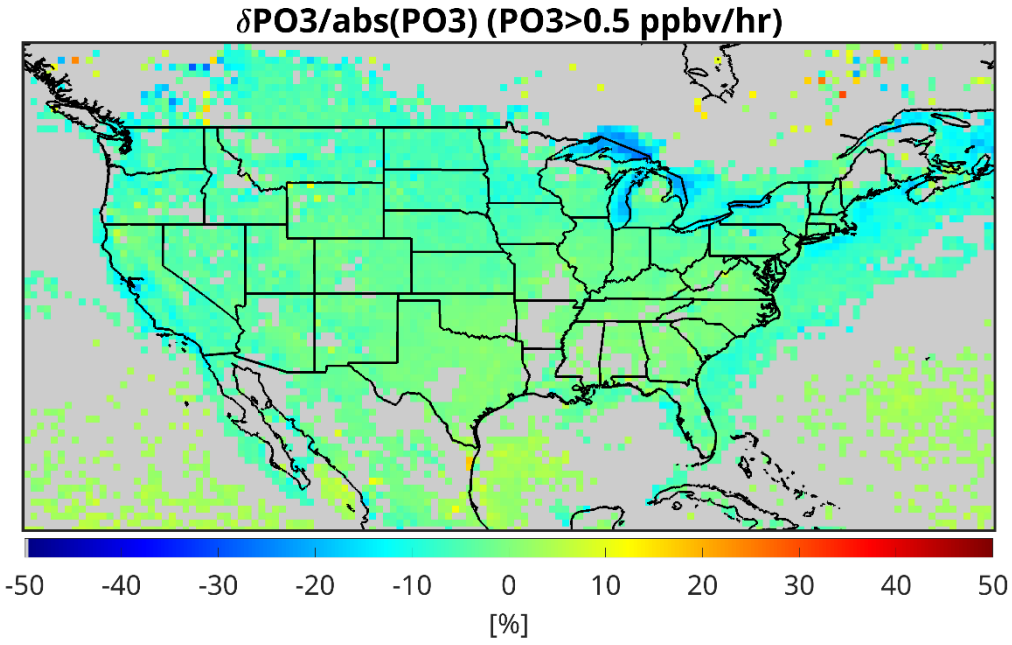




Concerning HCHO, AMFs changed around the same magnitude as those of NO2 (10±7%) resulting in PO3 changing to <15% over PO3>0.5 ppbv/hr.

Therefore, skipping AMFs recalculation should result in ~25% errors in PO3 estimates. However, the consideration of AMFs without redoing the bias-correction would have resulted in the same level of errors, suggesting the most robust way is to adjust both (bias correction and AMFs) at the same time which is not feasible given our budget constraint.





Modifications

We added:

We also quantify the impact of inconsistent shape factors used in the retrievals and the MINDS profile on PO_3 estimates and find them introducing systematic errors of 5-25% over PO_3>0.5 ppbv/hr (Figures S14-S17). Refining TROPOMI and OMI products with MINDS shape factors would require reproducing several large-scale validation efforts (e.g., Verhoelst et al.,

2020; Vigouroux et al., 2021; Pinardi et al., 2021; Ayazpour et al., 2025), which is beyond the practical scope and resources of this study.

In the summary section:

The total errors budget emphasizes on the role of model used for converting satellite-based VCDs to near-surface concentrations and its importance for precisely determining ozone precursors levels near to the surface. Furthermore, in future efforts, we also need to refine satellite retrievals using spatially higher-resolution AMFs derived from MINDS while simultaneously performing retrieval validation against ground-based remote sensing observations.

iii) the inclusion of more sophisticated chemical mechanisms for the generation of the training dataset; and iv) enhanced representation of vertical profiles of NO_2 and HCHO using observationally-constrained chemical transport models with more rigorous column to near-surface conversion factors (Cooper et al. 2020).

We added the above figures to the supplementary material.

Minor Comments

It would be helpful to clarify whether H2O values are directly inherited from MERRA-2 or modified within the MINDS model.

Response

MERRA2 is used to constrain U,V, QV, and T using the replay mode at 3-hourly basis in MINDS. So, meteorology is resolved in MINDS through GEOS. MERRA2 only adds a constraint.

Modifications

We added:

Meteorology is resolved using GEOS with several prognostic inputs, including water vapor, being constrained by MERRA-2 reanalysis using "replay" mode at 3-hourly basis (Orbe et al., 2017).

The description of "Southeast Asia" may be misleading; the text refers to August–September biomass burning, which applies mainly to maritime Southeast Asia, while continental Southeast Asia (Thailand, Myanmar, Laos, Cambodia) experiences its peak burning during February–April. Please clarify the regional definition.

Response

Thanks for pointing out this geographic mistake.

Modifications

We renamed the region to "maritime Southeast Asia" throughout the manuscript.

The expression "SZA acquired from the satellite L2 products" could be misleading, since SZA is not directly observed but computed from geometry information. Suggest rephrasing to "SZA derived from the geometry information in the L2 products."

Response

SZA is actually already computed and provided with L2 products. It's true that we can calculate that given time, location, and altitude, but the operation team has done it already.

Modifications

We modified it to:

Both SZA and surface altitude are provided as auxiliary fields in the satellite L2 products.

Check typographical errors (for example, "trend trends" to "trends"; "Tehan" to "Tehran").

Response

Corrected

The phrase "textbook example of non-linear chemistry" could be softened to "a clear demonstration of non-linear ozone chemistry."

Response

Corrected