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

Amir H. Souri[1,2]*, Gonzalo González Abad[3], Glenn M. Wolfe[1], Tijl Verhoelst[4], Corinne Vigouroux[4], Gaia Pinardi[4], Steven Compernolle[4], Bavo Langerock[4], Bryan N. Duncan[1], Matthew S. Johnson[5]

[1]Atmospheric Chemistry and Dynamics Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA

[2]GESTAR II, Morgan State University, Baltimore, MD, USA

[3]Atomic and Molecular Physics (AMP) Division, Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA, USA

[4]Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Ringlaan 3, 1180 Uccle, Belgium

[5]Earth Science Division, NASA Ames Research Center, Moffett Field, CA, USA

* Corresponding author: a.souri@nasa.gov

Abstract.

Ozone pollution is secondarily produced through a complex, non-linear chemical process. Our understanding of the spatiotemporal variations in photochemically produced ozone (i.e., PO_3) is limited to sparse aircraft campaigns and chemical transport models, which often carry significant biases. Hence, we present a novel satellite-derived PO_3 product informed by bias-corrected TROPOMI HCHO, NO_2, surface albedo data, and various models. These data are integrated into a parameterization that relies on HCHO, NO_2, HCHO/NO_2, jNO_2, and jO^1D. Despite its simplicity, it can reproduce ~90% of the variance in observationally constrained PO_3 with minimal biases in moderately to highly polluted regions. We map PO_3 across various regions in July 2019 at a 0.1°×0.1° spatial resolution, revealing accelerated values (>8 ppbv/hr) in numerous cities throughout Asia and the Middle East, resulting from the elevated ozone precursors and enhanced photochemistry. In Europe and the United States, such high levels are only detected over Benelux, Los Angeles, and New York City. PO_3 maxima are seen in various seasons, attributed to changes in photolysis rates, non-linear ozone chemistry, and fluctuations in HCHO and NO_2. Satellite errors result in moderate errors (10-20%) of PO_3 estimates over cities on a monthly average, while these errors exceed 50% in clean areas and under low light conditions. Using the current algorithm, we have demonstrated that satellite data can provide valuable information for robust PO_3 estimation. This capability expands future research through the application of data to address significant scientific questions about the locally-produced PO_3 hotspots, seasonality, and long-term trends.

1. Introduction

Tropospheric ozone (O_3) is a secondary pollutant formed through complex photochemical reactions involving various precursors, including nitrogen oxides (NO_x = NO + NO_2), volatile organic compounds

(VOCs), aerosols, and halogens (Kleinman et al., 2002, Simpson et al., 2015; Li et al., 2019). Ozone not only poses significant risks to human health (Fleming et al., 2018) and agricultural productivity (Mills et al., 2018) but also influences the radiation budget, thereby affecting the climate (Gaudel et al., 2018). To mitigate the problem of elevated locally-produced ozone, it is crucial to understand the spatiotemporal variability in ozone production rates (PO_3), defined as the number of ozone molecules generated through secondary chemical pathways in the atmosphere. Comprehensive studies of ozone chemistry, informed by observations, are typically confined to observationally-rich air quality campaigns (e.g., Cazorla et al., 2012; Ren et al., 2013; Mazzuca et al.; 2016; Souri et al., 2020a; Schroeder et al., 2020; Brune et al., 2022; Wolfe et al., 2022; Souri et al., 2023), which are sparse in time and space.

Significant advancements have been achieved in using various measurable ozone indicators to simplify the non-linear relationship between PO_3 and NO_x and VOCs into linear forms (Sillman and He, 2002). These forms include NO_x-sensitive (where PO_3 is sensitive to NO_x), VOC-sensitive (where PO_3 is sensitive to VOCs), and the transitional regimes (where PO_3 is sensitive to both NO_x and VOCs). Among the numerous proposed indicators, the ratio of formaldehyde (HCHO) to nitrogen dioxide (NO_2) (known as FNR) has gained popularity (Tonnesen and Dennis, 2000a,b), despite its less effective performance compared to the H_2O_2/HNO_3 ratio in fully explaining the HO_x-RO_x cycle (Silman and He, 2002; Souri et al., 2023). The preference for FNR stems from the fact that both quantities can be informed by UV-Vis radiance data, such as those provided by the Ozone Monitoring Instrument (OMI) and the TROPOspheric Monitoring Instrument (TROPOMI) (Martin et al., 2005; Duncan et al., 2010; Choi et al., 2012; Choi and Souri, 2015a, b; Jin and Holloway, 2015; Jin et al., 2017; Schroeder et al., 2017; Souri et al., 2017; Jeon et al., 2018; Tao et al., 2022). Several limitations associated with the application of satellite-based FNR have been identified such as i) the inherent limitation of understanding the radical termination in the RO_x-HO_x cycle (Souri et al., 2020a; Souri et al., 2023), ii) the challenges associated with converting the column vertical density to the near-surface concentrations (Jin et al., 2017; Schroeder et al., 2017; Souri et al., 2023), iii) spatial representativity associated with large satellite pixels (Souri et al., 2020a, 2023; Johnson et al., 2023), and iv) the retrieval errors (Souri et al., 2023; Johnson et al., 2023). Souri et al. (2023) concluded that the retrieval errors make up the largest portion of total errors associated with FNR. These errors are becoming smaller with better sensor designs, retrieval algorithms, and calibration over time.

While the characterization of ozone regimes offers valuable insights for regulators to prioritize effective emission control strategies, it does not provide information about the magnitude of PO_3 or the absolute quantities of PO_3 derivatives relative to its precursors. Consequently, chemical transport models under various emission scenarios are typically employed (e.g., Pan et al., 2019). These models allow for the execution of process-based scenarios to elucidate the response of PO_3 to different emissions and can simulate four-dimensional PO_3 data. However, the results of these simulations are based on various assumptions and inputs, which carry significant uncertainties. Therefore, it is essential to optimize some of the models' prognostic inputs using observations through inverse modeling/data assimilation. The primary advantage of inverse modeling/data assimilation using satellite observations is its ability to account for satellite errors and eliminate the influence of the a priori profile, thereby carrying only radiance information into the emission estimation. Numerous studies have utilized satellite observations to constrain NO_x and VOC emissions for various applications (e.g., Stavrakou et al., 2016; Souri et al., 2016; Miyazaki et al., 2017; Souri et al., 2017; Souri et al., 2020b; Souri et al., 2021; Choi et al., 2022; DiMaria et al., 2023). Souri et al. (2020b) made an early attempt to simultaneously optimize both NO_x and VOC emissions over East Asia for a more accurate representation of PO_3. Their joint-inversion was able to account for the intertwined relationship between HCHO-NO_x and NO_2-VOC. However, the execution of chemical transport models optimized by multiple satellite observations remains prohibitively expensive, particularly for high-resolution domains demanded by regulatory agencies.

Data-driven methods for estimating PO_3 can become as a more cost-effective alternative to physics-based methods. While using constrained chemical transport models provides a relatively robust framework grounded in some explicit governing equations, they require extensive computation resources and expertise.

Conversely, data-driven algorithms make use of large datasets to identify patterns and make predictions with much reduced computational expenses. However, it is important to recognize that data-driven algorithms lack the ability to provide solid physical interpretability and generalizability. Despite this fundamental limitation, they are sensible tools for applications where rapid analysis over a wide spatial coverage is prioritized. Data-driven parameterizations for several components of atmospheric chemistry such as OH (Anderson et al., 2022) and dry deposition (Silva et al., 2019) have been crafted for this reason. However, to our best knowledge, Chatfield et al. (2010) and Souri et al. (2023) are the only studies that attempted to empirically parameterize PO_3 using the information of HCHO and NO_2 mixing ratios.

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. This enabled us to map PO_3 across various regions at fine scales (i.e., 0.1°×0.1°) for the first time.

2. Data

2.1. Aircraft

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. We have selected three sets of aircraft campaigns for the purpose of PO_3 estimation, targeting: i) urban/suburban air quality, including Deriving Information on Surface Conditions from Column and Vertically Resolved Observations Relevant to Air Quality (DISCOVER-AQ) Baltimore-Washington (2011), DISCOVER-AQ Houston-Texas (2013), DISCOVER-AQ Colorado (2014), and the Korea United States Air Quality Study (KORUS-AQ) (2016) (Crawford et al., 2021); ii) remote areas including Atmospheric Tomography Mission (ATOM) (Thompson et al., 2022) and Intercontinental Chemical Transport Experiment (INTEX) phase B (Singh et al., 2009); iii) a mixture of isoprene-rich environment and large emitters, including SENEX (Southeast Nexus) (Warneke et al., 2016). Figure 1 shows the location of these campaigns. Inspired by the study of Miller and Brune (2022), we list their "when, where, why" characteristics in Table S1.

For aircraft campaigns targeting polluted areas, including DISCOVERs, KORUS-AQ, SENEX, and SEAC4RS, we use 10-sec merged data, whereas, for other measurements taken in relatively remote areas, such as INTEX-B and ATOMs, we used 30-sec merged data. A more detailed description of the measurements is provided in Section 3.2. We exclude times with no measurements of NO, NO_2, or HCHO. The concentrations of OH and HO_2 were only measured during INTEX-B, ATOMs, and KORUS-AQ. Likewise, we void any data points lacking either HO_2 or OH measurements. There are frequent gaps in some measurements, especially for VOCs, because of instrument issues or measurement techniques. Following Souri et al. (2020a), Miller and Brune (2020), Souri et al. (2023), and Bottorff et al. (2023), we fill the gaps in measurements using a linear interpolation method with no extrapolation allowed beyond 15 minutes. We drop any remaining gaps from the analysis. To better capture the rapid fluctuation of VOCs, we pick the PTR-TOF-MS instrument with high temporal resolution over the whole air sampler (WAS) when both instruments have measured the same quantity. Regarding the INTEX-B campaign, we drop isoprene observation due to infrequent samples downgrading the performance of our box model.

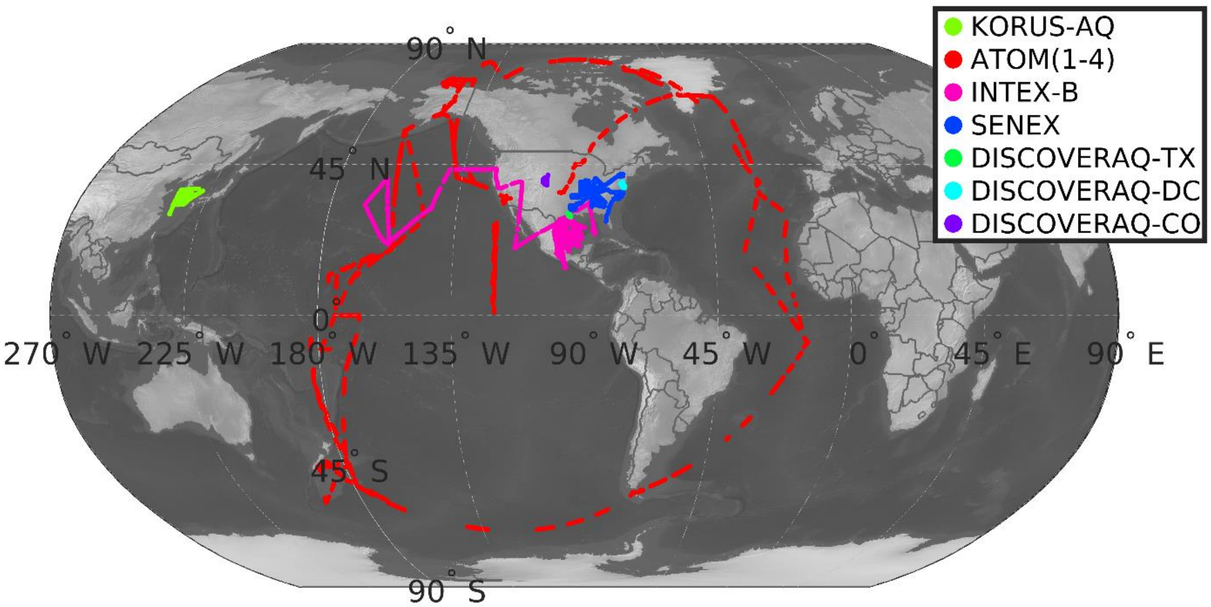

Figure 1. The location of seven different atmospheric composition aircraft campaigns used in this study.

2.2. TROPOMI NO_2 and HCHO

We use the recently reprocessed daily level-2 (L2) TROPOMI tropospheric NO_2 and total HCHO columns (v2.4) derived from UV-visible radiances onboard the European Space Agency's (ESA's) Sentinel-5 Precursor (S5P) spacecraft (~328-496 nm) (Veefkind et al., 2012, De Smedt et al. 2021; van Geffen et al., 2022). This sensor has been operational since May 2018, providing global coverage of NO_2 and HCHO at ~1:30 local standard time at the Equator. Since NO_2 and HCHO are optically thin absorbers in the UV-Visible, meaning their concentrations do not substantially affect the sensitivity of the radiance to the optical thickness of the absorber, the retrieval follows the conventional two-step algorithm involving spectral fitting for Slant Column Density (SCD) retrieval and Air Mass Factor (AMF) calculations for SCD to Vertical Column Density (VCD) conversion. The product has a spatial resolution of 7.2 km (5.6 km as of August 2019) by 3.6 km at nadir. To remove unfit measurements, we use the provided quality flag (*q_value*) and choose only those above 0.75 for NO_2 and 0.5 for HCHO. As the L2 product does not come in a regular grid, we use a mass-conserved regridding technique based on barycentric linear interpolation to map out the data onto a $0.1^{o}\times0.1^{o}$ regular grid.

van Geffen et al. (2022) demonstrated that the reprocessed TROPOMI tropospheric NO_2 columns exhibit a good level of correspondence with those obtained from ground-based MAX-DOAS sky spectrometers, with a correlation of 0.88 and a median bias of -23%, improving on the older product versions which were biased low by about 30% with respect to ground-based measurements at polluted sites (Verhoelst et al., 2021). More information about new modifications and their impacts on the retrieval can be found in van Geffen et al. (2022).

The studies of Vigouroux et al. (2020) and De Smedt et al. (2021) validated the reprocessed monthly-mean TROPOMI HCHO columns against FTIR and MAX-DOAS observations and found a good correlation above 0.8 with a negative bias of 20-30% for polluted sites. The bias tends to be slightly positive or neutral over clean sites.

2.2.1. Error characterization of TROPOMI NO_2 and HCHO using ground-based retrievals

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. The first component is derived from the column precision variable provided along with the L2 product. In the spatial domain, we interpolate the squares of this error the same of way we map the irregular L2 pixels into the $0.1^o \times 0.1^o$ regular grid. Moreover, we average the random errors over a month to reduce random noise by the squared number of pixels available at the same location (Eq. 3). Two other errors are determined by comparing FTIR (for HCHO) and MAX-DOAS (for tropospheric NO_2) with TROPOMI data (Section 4.3.3). Detailed explanation of how these datasets are paired can be found in Vigouroux et al. (2020) and Verhoelst et al. (2021). Both datasets cover the period of 2018-2023.

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 FTIR or MAX-DOAS), respectively; i is the subscript refers to i-th observation point, and f is the proposed linear fit subject to optimization. In terms of TROPOMI NO_2 and HCHO, the errors are populated based on the L2 information. According to Verhoelst et al. (2021), a fixed error of 30% is assumed for MAX-DOAS NO_2 observations whose values are above 1.4×10^{15}molec/cm^2. Because of the detection limit of MAX-DOAS NO_2, we set errors for values below that threshold to 1.4×10^{15}molec/cm^2. The FTIR retrieval errors described in Vigouroux et al. (2020) were used to populate the errors associated with this benchmark. The minimization is performed 10000 times, each with a set of random perturbations of x and y within their respective prescribed errors. This approach allows us to assess the robustness of the estimates across the range of errors associated with each data point.

The offset (a uniform additive term) and the slope (multiplicative error) drawn from the ground validation are used to correct the biases associated with TROPOMI via:

$$VCD_{bias-corrected} = \frac{VCD_{original} - offset}{slope} \tag{1}$$

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:

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

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:

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

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

2.3. TROPOMI Surface Albedo

To account for the effect of surface albedo on photolysis rates (Section 2.5), we use a newly developed algorithm based on the directionally dependent Lambertian-equivalent reflectivity (DLER) UV surface albedo climatology made from TROPOMI radiance (Tilstra et al., 2024). This new database leverages 60 months of TROPOMI reprocessed radiance and is produced at the grid resolution of $0.125^o \times 0.125^o$. 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).

2.4. MERRA2-GMI

To convert vertical column densities of HCHO and NO_2 from TROPOMI to their volume mixing ratios in the PBL region, we use the MERRA2-GMI (M2GMI) model (https://acd-ext.gsfc.nasa.gov/Projects/GEOSCCM/MERRA2GMI/, last access: 10 Sep 2023). This model is NASA's Goddard Earth Observing System (GEOS) Chemistry-Climate Model (CCM) run spanning for the period of 1980-2019, exploiting MERRA2 (Modern Era Retrospective analysis for Research and Applications) to constrain meteorological fields (Orbe et al., 2017). The model uses the Global Modeling Initiative (GMI) chemical mechanism (Duncan et al., 2007; Strahan et al., 2007), which involves over 120 species and 400 reactions. It has a resolution of approximately 0.625° longitude by 0.5° latitude with 72 vertical layers stretching from the surface up to 0.1 hPa. Additional information about the configuration of this model can be found in Strode et al. (2019). To carry out the conversion, we apply the following conversion factor (γ) to the TROPOMI VCDs:

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

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.

2.5. TUV NCAR Photolysis Rates Look-up Table

To estimate photolysis rates, JNO_2 (NO_2+hv) and JO^1D (O_3+hv), we use a comprehensive look-up table provided by the F0AM model (Section 3.2) created for clear-sky conditions. This look-up table is based on the calculation of more than 20,064 solar spectra over a wide range of solar zenith angle (SZA) (the range [0, 90] in steps of 5°), altitude (the range [0, 15] in steps of 1 km), overhead total ozone column (the range [100, 600] in steps of 50 DU), and surface UV albedo (the range [0, 1] in steps of 0.2) using NCAR's Tropospheric Ultraviolent and Visible radiation model (TUV v5.2) and cross sections and quantum yields from IUPAC and JPL (Wolfe et al., 2016). The L2 TROPOMI granule information populates SZA, surface elevation, and surface UV albedo, while overhead total ozone columns are obtained from MERRA2-GMI (Section 2.4) which is found to agree well with satellite observations (Souri et al., 2024). Any values between these tables are bilinearly interpolated for a smoother result.

3. Methods

In this section, we begin by discussing a robust regression model specifically developed for feature selection in the parameterization of PO_3. We then describe the training dataset created for this purpose. Following that, we introduce a clustering technique utilized to organize the training data, which enables us to identify the key drivers of PO_3 variability. Finally, we provide a comprehensive overview of the PO_3 estimates algorithm by integrating data from the TROPOMI retrievals, ground-based remote sensing, and various models.

3.1. LASSO

Through the use of multi-linear regression models, it is possible to establish a simple but robust relationship between multiple variables and a target. However, when dealing with a large number of variables, there is a chance of introducing overfitting issues. This can lead to predictions that are either overly optimistic or unrealistic for values outside of the training dataset. To avoid this, it is recommended to simplify the model by removing variables that are loosely connected with the target or highly correlated with others. This process is known as "model shrinkage" and can narrow down the number of possible solutions (i.e., variance) at the cost of increasing the biases between the observed target and predictions. Ideally, we want a model that minimizes the sum of the bias and the variance. To achieve this, we can use LASSO (least absolute shrinkage and selection operator) (Tibshirani, 1996). They consider a regression,

$$Y = X\beta + \alpha + \varepsilon \tag{5}$$

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:

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

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. The first term on the right side of Eq.6 minimizes the squares of the residuals, whereas the second term reduces the sum of absolute value of coefficients resulting in a simpler model with fewer parameters. Without the second term, the regression model becomes an ordinary least-squares estimation. The most critical element here is λ. A large λ results in more aggressive regularization leading to more model shrinkage, whereas a small value preserves a high dimensional model. 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), determine the average of cross-validated error prediction among all folds, and find λ that yields the smallest error. The final solution ensures a balanced model with respect to model parsimony and bias. All explanatory variables are standardized during the regularization procedure such that their mean becomes zero and their standard deviation one.

3.2. Photochemical box modeling

To produce training data sets for LASSO-based PO_3 estimation, we use the Framework for 0-D Atmospheric Modeling (F0AM) v4 box model (Wolfe et al., 2016), constrained by a wide range of observations. These observations ensure that the model achieves a realistic range of values found in the atmosphere. We follow past setups which apply the Carbon Bond 6 (CB06, r2) chemical mechanism in F0AM (Souri et al., 2020a; Souri et al., 2023). The model is constrained by aircraft data, including meteorology, photolysis rates, and trace gas concentrations. The model configuration and observations used are listed in Table S2.

Once the model is initialized and held constant with respect to a wide range of constraining quantities, it runs at 30 minutes integration time cycling for five days to approach a steady-state environment. Several key compounds including OH, HO_2, HCHO, PAN, NO, and NO_2 are initialized with aircraft observations but they are left free to cycle with incoming solar radiation variability. These compounds play a crucial role in validating the efficacy of model performance as well as the adequacy of observations used as constraints. In particular, allowing HCHO to vary freely enables us to assess whether our mechanism for VOC treatment, steady-state, and the number of measured VOCs suffice to reproduce its concentrations reasonably. Although the individual concentration of NO_2 and NO are not constrained, we constrain total NO_x ($NO+NO_2$). Not all aircraft campaigns measured all photolysis rates included in the chemical mechanism. We first initialize the photolysis rates included in CB06 using the look-up-tables described in Section 2.5. If any photolysis reaction rates in CB06 were measured, we replace the initial guess with the observed values. For those reactions with photolysis rates not been measured, we apply a scaling factor made of the average of the ratio of the observed J-values to the modeled J-values. This approach is a sensible choice for accounting for large particles such as clouds, as their extinction coefficient is somewhat non-selective in the UV-Vis range; however, applying a wavelength-independent scaling factor may introduce some biases for optically complex environments introduced by aerosols.

It is essential to acknowledge the inherent limitations of a box model in our research. The model does not consider the diverse physical loss pathways that trace gases may undergo, including deposition and transport. As a result, we have simplified the physical loss by employing a first-order dilution rate set to 1/86400 s^{-1}, equivalent to a lifetime of 24 hours. This approach ensures that unconstrained trace gases that take longer to break down do not accumulate over time. Exact knowledge of dilution factors requires knowing molecular and turbulent diffusion, entrainment and detrainment, and deposition rates, all of which are unknown at the micro-scale level of aircraft observations. Nonetheless, studies of Brune et al. (2022) and Souri et al. (2023) showed that HO_2, OH, NO_x, and HCHO are relatively immune to the choice of the dilution factor, whereas RO_2 mixing ratios can depart introducing some biases in PO_3 estimates.

We determine simulated PO_3 by:

$$PO_3 = FO_3 - LO_3 \tag{7}$$

where LO_3 is all possible chemical loss pathways of ozone (negative stoichiometric multiplier matrix) and FO_3 is all possible chemical pathways producing ozone molecules (positive stoichiometric multiplier matrix). This calculation is theoretically equivalent to a value obtained from a chemical solver quantifying the number of ozone molecules produced/lost for each model timestep. The adoption of Eq.7 facilitates the direct comparison of PO_3 estimations with those derived from other models, including CTM-based results (see Figure 10 in Souri et al., 2021). Furthermore, it allows for a seamless integration of these estimates into Lagrangian transport models for ozone forecasting purposes.

3.3. Clustering

The aim of using a classifier to group the large quantity and types of aircraft data into similar features is to allow us to study the primary contributors to PO_3 under different chemical, solar, and meteorological conditions. Additionally, this approach will help us understand the range of atmospheric conditions included in the training dataset. To accomplish this, we employ a widely-used technique known as *k*-means, which has been used in a variety of applications (e.g., Beddows et al., 2009; Souri et al., 2016b; Govender and Sivakumar, 2020). In this approach, centroids are distributed randomly throughout a multi-dimensional dataset, with each centroid representing a distinct class. The algorithm proceeds to assign a label to each data point by identifying its closest Euclidean distance to the centroids. Following the labeling of all data points, the algorithm updates the centroids based on the means of the newly-labeled group. This process continues iteratively until there is minimal change in the location of the centroids. It is worth noting that *k*-means does not guarantee an optimal solution, so we reinitialize the classification 1000 times with a new set of initial centroids. We select the result with the lowest value for the sum of the Euclidean distance among data points and centroids to ensure the outcomes are not influenced by random seeding.

Redundant features in the input can significantly compromise the effectiveness of the classification, so we apply principal component analysis (PCA) to the matrix of datasets (Z) with n data points and p features to reduce the dimension to a PCA-transformed matrix of Z (Z') with the dimension $n \times q$, where $q<p$. Despite this reduction in dimension, Z' preserves a significant variance in Z, helping us to overcome the issues of dimensionality or overfitting.

We select 11 features simulated by the F0AM model, many of which are set to the observed values, or their precursors are observationally-constrained. These features are SZA, $HCHO/NO_2$, $HCHO \times NO_2$, HCHO, NO_2, pressure, temperature, jNO_2, jO^1D, H_2O, and NO_2/NO_y (NO_y=NO+NO_2+PAN+HNO_3+alkyl nitrate +N_2O_5). There are indeed correlations among these features such as SZA and jNO_2, or HCHO and HCHO×NO_2; nonetheless, we have used PCA to eliminate the possibility of these correlated factors causing overfitting issues.

3.4. The estimation of PO_3

In order to predict PO_3, we have developed empirical equations using LASSO to link PO_3 with various relevant prognostic candidates related to ozone chemistry. A schematic presentation on how this estimation can be done to provide daily PO_3 maps at the TROPOMI revisit time across the globe is shown in Figure 2. It is important to note that relying solely on linear regressions for a non-linear problem is not a viable approach. To address this, we have divided the data points into four distinct groups based on FNR values, meaning we divide a non-linear realm into smaller linear segments (i.e., an empirical linearization). In a study by Souri et al. (2023), a wide range of aircraft observations and box model results were used to determine that FNR~1.7 was a universal threshold for separating NOx-sensitive from VOC-sensitive regimes. We have found that by breaking down the datapoints into slightly weaker or stronger variations of the regimes, we can improve the accuracy of our results. As a result, we have established four distinct groups: VOC-sensitive (FNR<1.5), transitions (1.5<FNR<2.5 and 2.5<FNR<3.5), and NOx-sensitive (FNR>3.5). The coefficients and intercepts based on the LASSO regressions for each group were computed separately. From a long list of explanatory parameters, we selected SZA, temperature, pressure, H_2O, jNO_2, jO^1D, HCHO, and NO_2 as the most sensible candidates. The reasoning behind this selection will be discussed in Section 4.2.

Once the LASSO parameters are determined, we apply the linear functions to variables modeled/observed in the PBL region. We show that the LASSO method votes for dropping SZA, temperature, water vapor, and pressure as they do not provide significant information on PO_3 compared to the rest. As for jNO_2 and jO^1D, we use the TUV NCAR's LUT described in Section 2.5. HCHO and NO_2 are derived by converting the bias-corrected TROPOMI VCDs into PBL mixing ratios using MERRA2-GMI described in Section 2.4. To carry out the conversion, we multiply the satellite VCDs by the ratio of averaged modeled mixing ratios of a target gas (i.e., NO_2 or HCHO) in the PBL region divided by modeled VCDs (Section 2.4). The PBL field also comes from MERRA2-GMI.

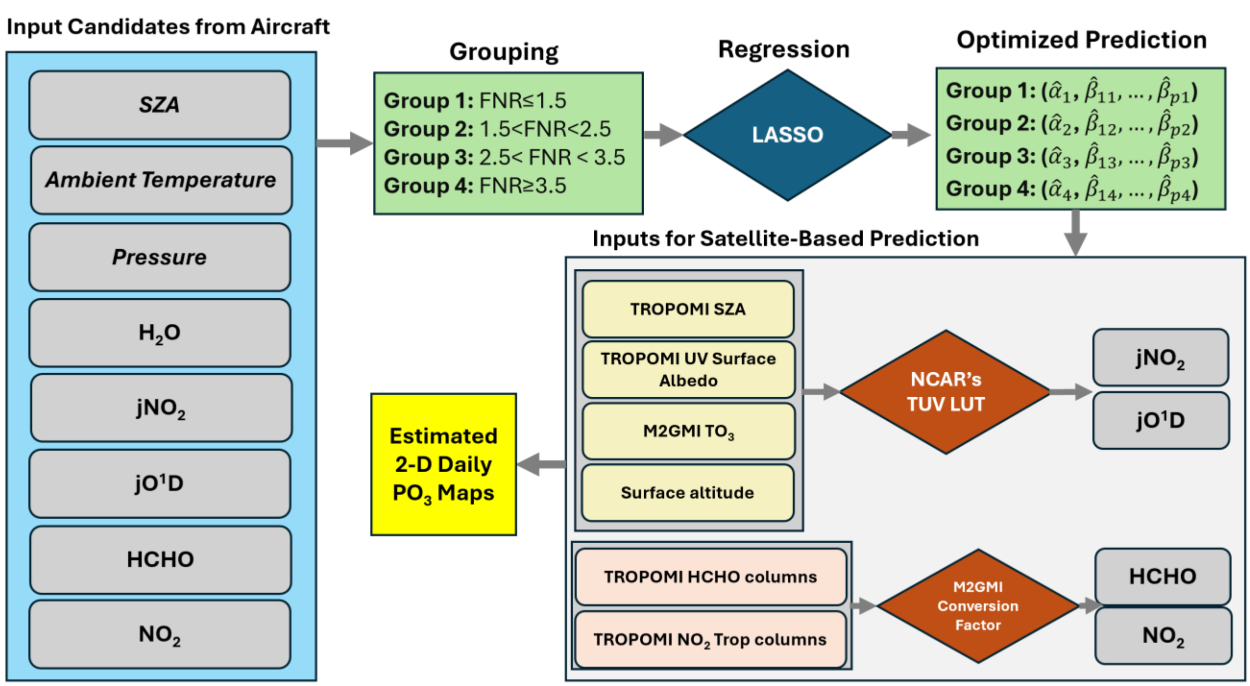

Figure 2. Schematic illustration of daily PO_3 estimation calculated in this study. This process consists of two major steps: formulating PO_3 as a function of various prognostic inputs derived from the box model results, and predicting PO_3 based on optimized features/coefficients suggested by LASSO and using information obtained from TROPOMI, TUV, and M2GMI.

4. Results and Discussion

4.1. Box Model Validation

In order to assess the accuracy of the assumptions used in the box model's setup, which involves factors such as chemical mechanism, dilution rate, and photolysis rate correction, we will compare the simulated values of HCHO, NO_2, NO, PAN, HO_2, and OH with their actual measured values. This comparison will help us determine if our model falls within an acceptable range of errors as seen in other reputable photochemical box modeling studies. This comparison is represented in Figure 3, which displays a scatterplot of the data collected from all seven aircraft campaigns. A discussion on each parameter follows:

HCHO – The box model is proficient in capturing over 77% of variance in observations with less than 15% absolute bias. While many box modeling studies prefer to have this compound constrained to potentially enhance the representation of HOx, it comes with the trade-off of hindering us from validating the number/quality of observed HCHO precursors and/or the VOC treatment. Besides the study of Souri et al. (2023), Marvin et al. (2017) is one of the few studies that did not constrain this compound to verify the efficacy of different pathways involved in HCHO formation and loss simulated by various chemical mechanisms. Marvin et al. (2017) reproduced HCHO formation during the SENEX campaign using the CB06 mechanism with a R^2=0.66 and a bias of 32% at 1-min averaged samples. Compared to that study, we recreate 86% variance in observed HCHO during the same campaign with a bias of 23% (Figure S1) at 10-sec averaged samples. The remaining unresolved variance can be attributed to an incomplete list of VOC measurements for several campaigns including DISCOVER-AQs and errors of VOCs measurements. It is unlikely for the chemical mechanism to be reason for this, as Marvin et al. (2017) did not observe substantial differences in R^2 values among various chemical mechanisms including the near-explicit MCM. A mild underestimation of HCHO could be likely due to the steady-state assumption, fixed arbitrary dilution factor, or uncertain isoprene chemistry (Archibald et al., 2000; Wolfe et al., 2016; Marvin et al., 2017).

NO_2 and NO – Comparisons for both species demonstrate a high degree of correspondence for values above 0.1 ppbv. Nonetheless, we have noted a substantial amount of fluctuation in the simulations in clean regions, particularly for NO. While we cannot rule out the possibility of chemical mechanism uncertainty contributing to this deviation, the reported measurement errors for NO_2 and NO are usually ±0.05 ppbv and ±0.1 ppbv, respectively. 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.

PAN – Our model reproduced 61% of the variance observed in PAN with a marginal absolute bias. According to Xu et al. (2021), the presence of oxygenated VOCs, particularly acetaldehyde, and the NO/NO_2 ratio are key factors controlling PAN levels. While we have constrained acetaldehyde, variations in the NO/NO_2 ratio in heavily polluted regions (where NO_x levels exceed 1 ppbv) could potentially lead to biases in PAN simulations. Furthermore, our model's dilution factor has been arbitrarily set, and it is possible that any bias caused by this factor has been canceled out by other effects, leading to seemingly bias-free performance. However, Souri et al. (2023) showed that an incorrect dilution factor can significantly impact PAN performance, causing a sharp decline in R^2 resulting in a value below 30%. Therefore, the fact that our box model has performed well with respect to PAN could be an indication that our choice of the dilution factor is reasonable.

HO_2 and OH – Based on our analysis of HO_2 and OH simulations during KORUS-AQ, INTEX-B, and ATOMs, we have found a reasonable level of correspondence (R^2>0.6) with the performance in previous studies conducted by Souri et al. (2020), Brune et al. (2022), Miller and Brune (2022), and Souri et al. (2023) that focused on some of these campaigns. Although the box model OH simulations reported in Brune et al. (2019) during ATOMs seemed to be better than ours (R^2~0.8 vs R^2~0.6), it is important to consider that their observations were averaged over 1-minute intervals as opposed to our 30-second intervals. It should also be noted that there can be large errors in ATHOS HO_x measurements of up to ±40% (Miller et al., 2022), so recreating the exact variance in the observations should not be the main objective. Nonetheless, the performance of our simulations in terms of HO_x compared to observations suggests that the number of measured compounds and chemical mechanisms used in the model was effective. Our model's performance with respect to HO_x is comparable to more sophisticated mechanisms that encompass a larger number of measured species (Brune et al., 2022; Miller and Brune, 2022).

Overall, while there are inevitably some differences between the box model results and observations, they are consistent with what other studies have found in similar aircraft campaigns. Our extensive box model results, which consider a variety of meteorological, chemical, and photolysis rates, demonstrate satisfactory results for unconstrained compounds across a wide range of atmospheric conditions. This suggests that our training dataset from the box model is a reliable source for understanding local PO_3.

It is important to note that even if a simulated data point does not match up perfectly with actual observations, it still plays a role in establishing PO_3 and other explanatory variables. Hypothetically, one can generate synthetic training data points by running the box model under random numbers for the inputs; but only a fraction of those can be truly observed in nature. Therefore, a mild outlier in our training dataset should be viewed as less likely to occur in nature (presuming that these campaigns could represent all conditions happening in nature), but still a valuable data point drawn from a physical model that can be used to bridge PO_3 with explanatory variables.

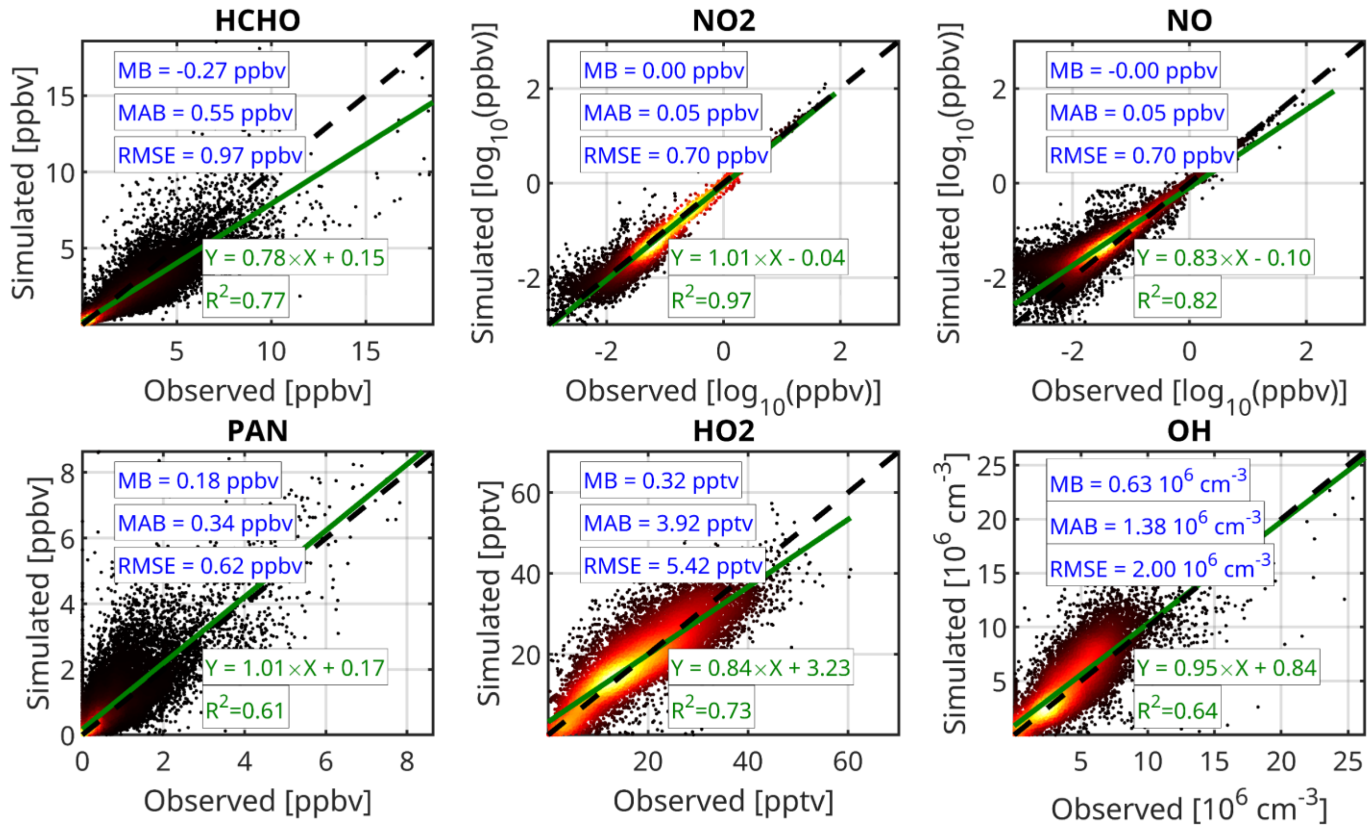

Figure 3. The scatterplot comparison of simulations with observed concentrations for six unconstrained species. More than ~133,000 observations are used for HCHO, NO_2, NO, and PAN. HO_x data points are limited to ~55,000 observations. Heat maps show the density of the data. Linear fits are calculated using the ordinary least squares method.

4.2. Classification of aircraft data

Following the method described in Section 3.3, we cluster the cloud of aircraft data (~ 133k points) into seven distinct classes. We describe them using three categories: pollution level, altitude, and SZA. Figure 4 illustrates the violin plot of these classes for various chemical, solar, and meteorological conditions. Figure 5 shows their corresponding violin plot of simulated PO_3. A discussion of each class and their relationship to PO_3 follows:

C1 (clean, high altitude, high SZA) – Characterized by high altitude flights, cold ambient temperature, and negligible water vapor content, this class consists of observations that were typically taken during relatively high SZA with a median of 50°. While high altitude observations in clear-sky conditions often should have large photolysis rates due to reduced overhead ozone, the relatively high SZA of this class leads to low photolysis rates. FNRs tend to be large in this class due to a higher amount of HCHO over NO_2, and FNP ($HCHO \times NO_2$) and NO_2/NO_y ratios are low due to the pristine conditions. The lack of sufficient ozone precursors and reduced photochemistry make this class undergo the lowest PO_3 rates with a median of 0.11 ppbv/hr.

C2 (clean, high altitude, low SZA) - This category represents samples collected in low SZA conditions, resulting in the highest photolysis rates among all classes. The mass of ozone precursors and the ozone sensitivity condition are similar to those in C1. However, C2 PO_3 rates are approximately 60% higher than C1 due to increased photochemistry.

C3 (moderately clean, medium altitude, high SZA) - This class is characterized by observations collected in mid-altitudes and high SZA. Airsheds in C3 experienced relatively more polluted air compared to C1 and C2 due to being closer to the surface. Photolysis rates are smaller than C1 possibly because of higher ozone overhead, although we cannot rule out the varying surface albedo between the classes. Despite the lower photolysis rates, C3 PO_3 (0.28 ppbv/hr) is larger than that of C2 and C1, indicating that pollution levels can have a more significant impact than favorable conditions for photochemistry.

C4 (moderately clean, medium altitude, low SZA) - This category is distinct from C3 in terms of lower SZA (resulting in more photochemistry) and a slightly smaller number of ozone precursors. As a result of the lower ozone precursor concentration, not only is C4 PO_3 (0.19 ppbv/hr) lower than C3, but also is not different from C2. This again implies that the amount of ozone precursors is more important than the photochemistry for these conditions.

C5 (extremely polluted, low altitude, low SZA) - This class features the highest amount of ozone precursors (median FNP ~ 58 $ppbv^2$) among all classes. Furthermore, it is characterized by low photolysis rates due to its proximity to the surface, and high NO_2/NO_y indicative of localized polluted airshed. Unlike the previous classes, this class has the lowest FNR, indicating that it is mainly located in the VOC-sensitive regime. C5 PO_3 values are much higher than the previous classes, with a value of 3.0 ppbv/hr.

C6 (polluted, low altitude, low SZA) - While this class shares similar features with C5 in terms of altitude, photolysis rates, and meteorology, it experiences a lower FNP (median of 8 $ppbv^2$). Despite the lower FNP, C6 has the highest amount of PO_3 (5.2 ppbv/hr) among all classes. This is a result of reduced non-linearities, as this class does not often fall into an extreme VOC-sensitive regime (median FNR ~ 1.0) where nitrogen oxides (NO_x) can hamper ozone production. This tendency coincides with Souri et al. (2023) which also found that the highest amount of PO_3, lied between the transitional regimes, gravitated towards VOC-sensitive because of abundant ozone precursors and reduced negative chemical feedback of NO_x.

C7 (moderately polluted, low altitude, high SZA) - C7 is characterized by aged air close to the surface with slightly higher photolysis rates than C5 and C6. C7 PO_3 is 2.5 ppbv/hr, only slightly smaller than C5 despite much lower FNP (median of 0.9 $ppbv^2$). This could be caused by the combined effect of higher photolysis rates and reduced non-linear ozone chemistry.

The analysis of aircraft data has revealed that the levels of HCHO and NO_2, as well as the rates of jNO_2 and jO^1D photolysis, play an important role in influencing PO_3. Additionally, FNRs can offer insights into the sensitivity of PO_3 to its main precursors. These findings align with numerous other studies that have examined the factors driving PO_3 (e.g., Duncan and Chameides, 1998; Thornton et al., 2002; Kleiman et al., 2002; Gerasopoulos et al., 2006; Chatfield et al., 2010; Baylon et al., 2018; Wang et al., 2020; Souri et al., 2023). Consequently, our PO_3 estimates will incorporate HCHO, NO_2, jNO_2, jO^1D, and FNR. While the cluster analysis did not definitively indicate whether meteorological conditions impact PO_3, we will also include ambient temperature, water vapor, pressure, and SZA to determine if they provide any additional insights into PO_3 estimates.

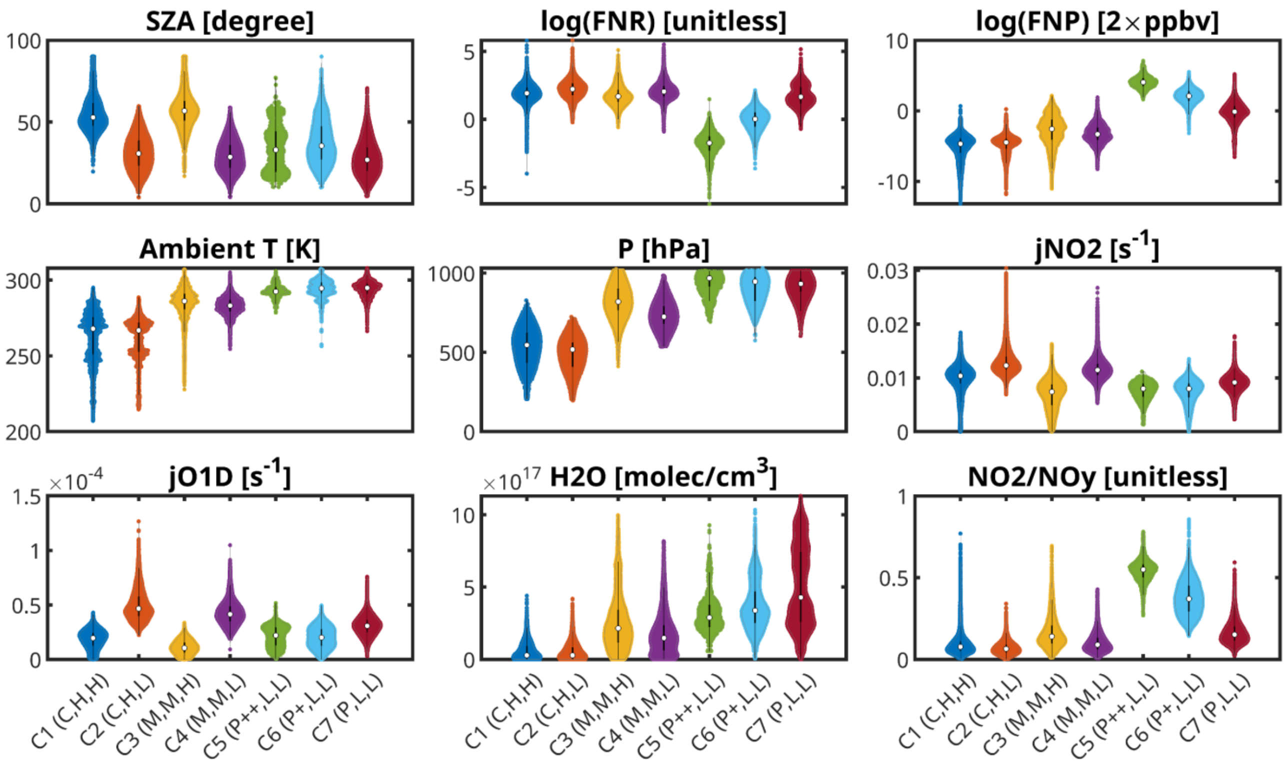

Figure 4. The violin plots of six different parameters coming from the box model clustered into seven distinct categories. Each cluster is described by three labels: air pollution levels (C: clean, M: moderately clean, P: moderately polluted, P+: polluted, P++: extremely polluted), altitude (H: high, M: medium, L: low), and SZA (H: high, L: low). The white dot is the median and the bars explain the 75th and 25th percentiles. Both FNR and FNP are scaled using the logarithmic function to enable the simultaneous visualization of low and high values within a single plot.

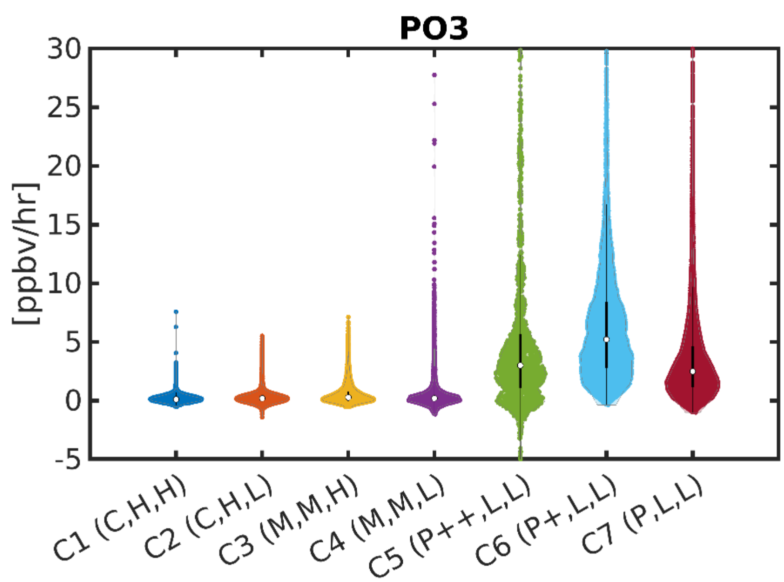

Figure 5. The corresponding violin plots of simulated PO_3 for the seven clusters described in Figure 4. The lowest PO_3 is seen in remote regions (C-M) where ozone precursors are minimal. The highest PO_3 does not happen in the most polluted region (P++) resulting from the non-linear ozone chemistry.

4.3. Estimates of PO_3

4.3.1. LASSO coefficients

Armed with a procedure that finds the important features in a linear model (Section 3.1), we now explore using LASSO for PO_3 estimation. We make use of all data points generated by the observationally-constrained box model from various atmospheric composition campaigns. Among the selected variables shown in Figure 2, the LASSO algorithm assigns zero coefficients to SZA, pressure, temperature, and water vapor, indicating that they offer less valuable information compared to other variables. This decision was made by systematically adjusting the regularization factor within a 10-fold cross-validation framework to identify the optimal factor that strikes a balance between solution variance and prediction bias. As a result, the LASSO algorithm suggests that HCHO, NO_2, jNO_2, and jO^1D contain sufficient information to accurately predict PO_3 for the most part.

Table 1 provides the intercepts and the corresponding coefficients for four different regions separated by FNR. While we do not expect for a statistical model to fully single out the "cause and effect" relationship between explanatory variables and the target, we note that it has some basic understanding of ozone chemistry; the HCHO coefficients increase as moving towards smaller FNRs (i.e., more VOC-sensitive). The same tendency is evident with respect to NO_2 and larger FNRs (i.e., more NO_x-sensitive). The negative coefficient of NO_2 in regions having FNR≤1.5, implies some levels of non-linear feedback embedded in this parameterization. Both jNO_2 and JO^1D have positive coefficients throughout the chemical conditions, suggesting that higher photolysis rates accelerate PO_3. JO^1D has a smaller effect than jNO_2 on PO_3 over remote regions (FNR≥3.5) perhaps because of redundant information available compared to jNO_2.

Table 1. Calibrated coefficients derived from the LASSO estimator using seven atmospheric composition aircraft campaigns.

Group	Criteria for FNR	Intercept	HCHO [ppbv]	NO_2 [ppbv]	$jNO_2\times10^3$ [s^{-1}]	$jO^1D\times10^6$ [s^{-1}]
1	FNR≤1.5	-1.98	1.85	-0.14	0.12	0.09
2	1.5<FNR<2.5	-3.38	1.79	0.98	0.19	0.07
3	2.5<FNR<3.5	-3.27	1.07	3.48	0.21	0.03
4	FNR≥3.5	-1.63	0.41	6.54	0.11	0.01

4.3.2. Validation of PO_3 predictions

The validation of PO_3 prediction against the box model results is performed in threefold with an increasing stringency order: i) using all data points used in the LASSO algorithm, ii) by randomly dropping data points, and iii) by dropping each air quality campaign from the LASSO estimation and using its data as benchmark.

Figure 6a shows the scatterplot of predicted PO_3 against the box model for all data points used to estimate the coefficients described in Section 4.3.1. Despite the algorithm's simplicity, we can recreate more than 88% of the variance in PO_3 with negligible absolute bias. This has an important indication that our scientific problem is not overly complex. There is less than 30% bias with respect to the mean absolute bias of the prediction. The positive offset and a slope smaller than one indicate a mild underestimation (overestimation) of PO_3 in polluted (clean) regions. Figure 6b shows the same analysis for 20,000 randomly chosen data points (~15% of the total) that we purposefully dropped from the LASSO estimation to gauge if the predictor model can replicate numbers for points not used during the training. We find almost identical statistics for these points, suggesting that the prediction stays robust for points outside the training data set.

However, the most stringent method is to drop each campaign data set entirely to understand where the prediction model struggles most.

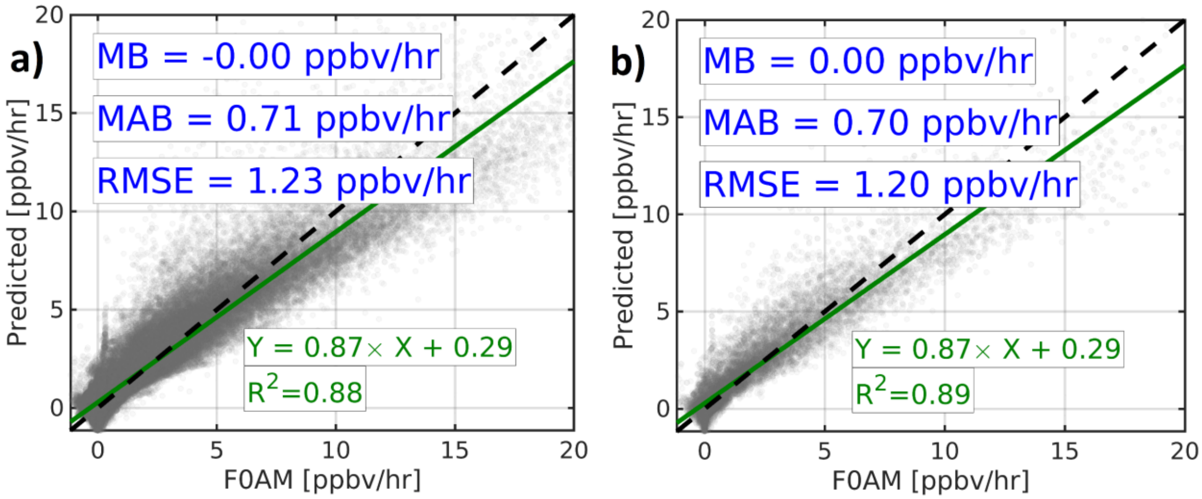

Figure 6. Scatterplots comparing observationally-constrained F0AM model PO_3 and the predictions based on the proposed algorithm for (a) all data points and (b) 20,000 randomly-dropped data points as benchmarks. Despite the simplicity of the algorithm, we can reproduce a large variance in PO_3 using only four explanatory variables.

Figure 7 shows several subplots pertaining to dropped campaigns from the analysis. Immediately evident is that our PO_3 estimation has considerable skills at capturing PO_3 for most polluted cases, including DISCOVER-AQs, KORUS-AQ, and SENEX without using their individual datasets. This provides convincing evidence about a high degree of generalizability of the predictor. However, the model has a reduction in performance in INTEX-B for PO_3 <1 ppbv/h. Moreover, the model prediction power is consistently poor for ATOMs where a significant fraction of airsheds were samples in pristine areas. We see such poor performance for PO_3<1 ppbv/hr for other campaigns such as KORUS-AQ. Therefore, it is difficult to have confidence in the predictive power of the model in remote regions, which may be caused by the lack of inclusion of HOx, halogens, and H_2O in the fit, as they can become an important sink for tropospheric ozone in those areas (Simpson et al., 2015). Nonetheless, while our predictive accuracy remains poor for this specific subset of the data, the practical utility and significance of this specific region (i.e., pristine areas) for air quality applications are notably limited. Given these results, we limit our predictions to PO_3>1 ppbv/hr for the subsequent analyses.

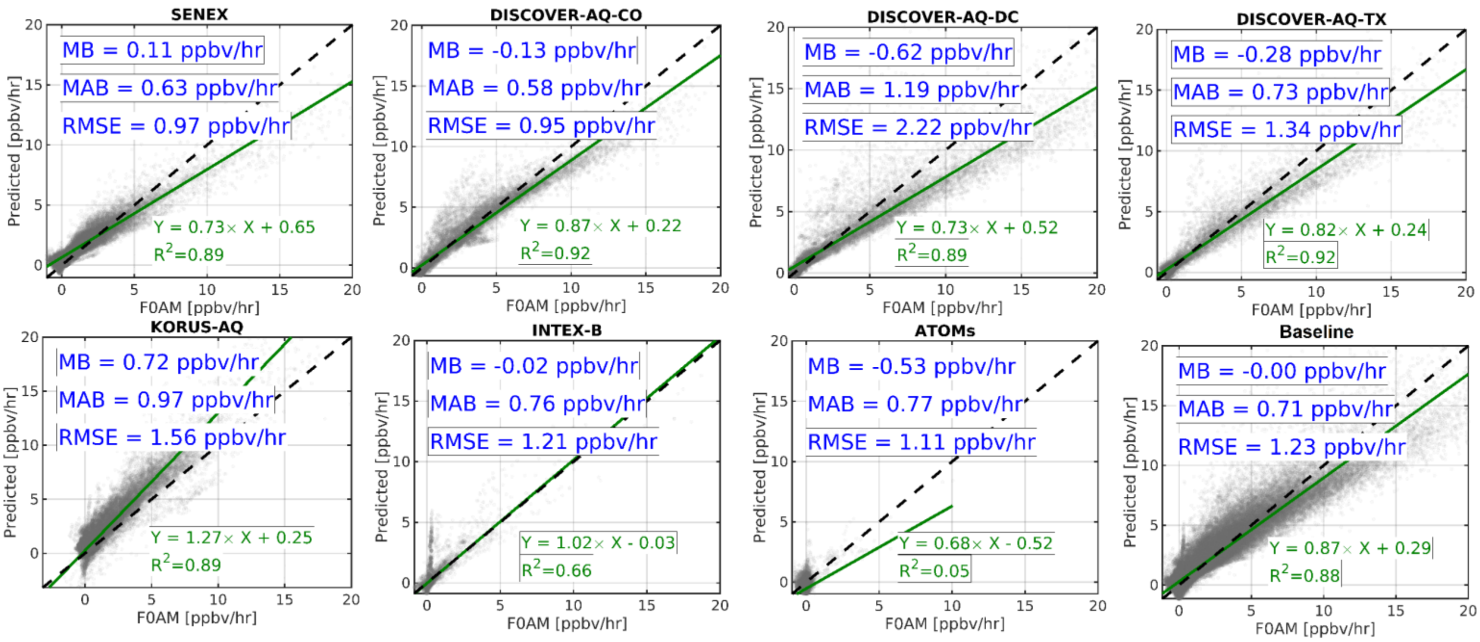

Figure 7. Same as Figure 6b, but each campaign is dropped from the LASSO estimation and subsequently used as an independent benchmark. The designed algorithm has shown a high degree of skill at predicting PO_3 in polluted regions; however, it performs poorly in pristine areas.

4.3.3. TROPOMI NO_2 and HCHO validation

To build confidence in our quantitative application of TROPOMI data for PO_3 estimates, we validate the daily tropospheric NO_2 and total HCHO columns against MAX-DOAS and FTIR observations based upon the validation framework outlined in Vigouroux et al. (2020) and Verheolst et al. (2021). Both paired datasets have been expanded to late 2023 showing a fuller picture of TROPOMI error characterization compared to former studies. Figure 8 shows the comparison of daily TROPOMI, the benchmarks and the optimal fit associated with their errors for the period of 2018-2023.

In the context of tropospheric NO_2 comparison, we observe a slope smaller than one (~ 0.66) with a positive offset (0.32×10^{15} molec/cm^2). This tendency has been repeatedly documented in various studies for various satellites or benchmarks (e.g., Griffin et al., 2019; Choi et al., 2020; Verhoelst et al. 2021; van Geffen et al., 2022). A slope smaller than one, originating from unresolved systematic biases, implies that TROPOMI is biased-low in polluted regions. A slight positive offset suggests that TROPOMI NO_2 is biased-high in remote regions. The errors of slope and the offset are relatively small, evidence of the robustness of the optimal fit against the dataset variance. Nonetheless, we will incorporate them into Eqs 2 and 3 to take the adjustment error into consideration.

Despite the inherent difficulty in obtaining HCHO observations from the UV-Vis imagery (González Abad et al., 2019), the HCHO comparison exhibits a good alignment with benchmarks. Like the previous comparison, the slope is smaller than one (~0.59) and the offset is positive (~0.9×10^{15}molec/cm^2) agreeing within 10% with studies done by Vigouroux et al. (2020) and De Smedt et al. (2021). Consequently, we will consider the fit errors and adjust all VCDs based on the slope and the offset obtained from this comparison.

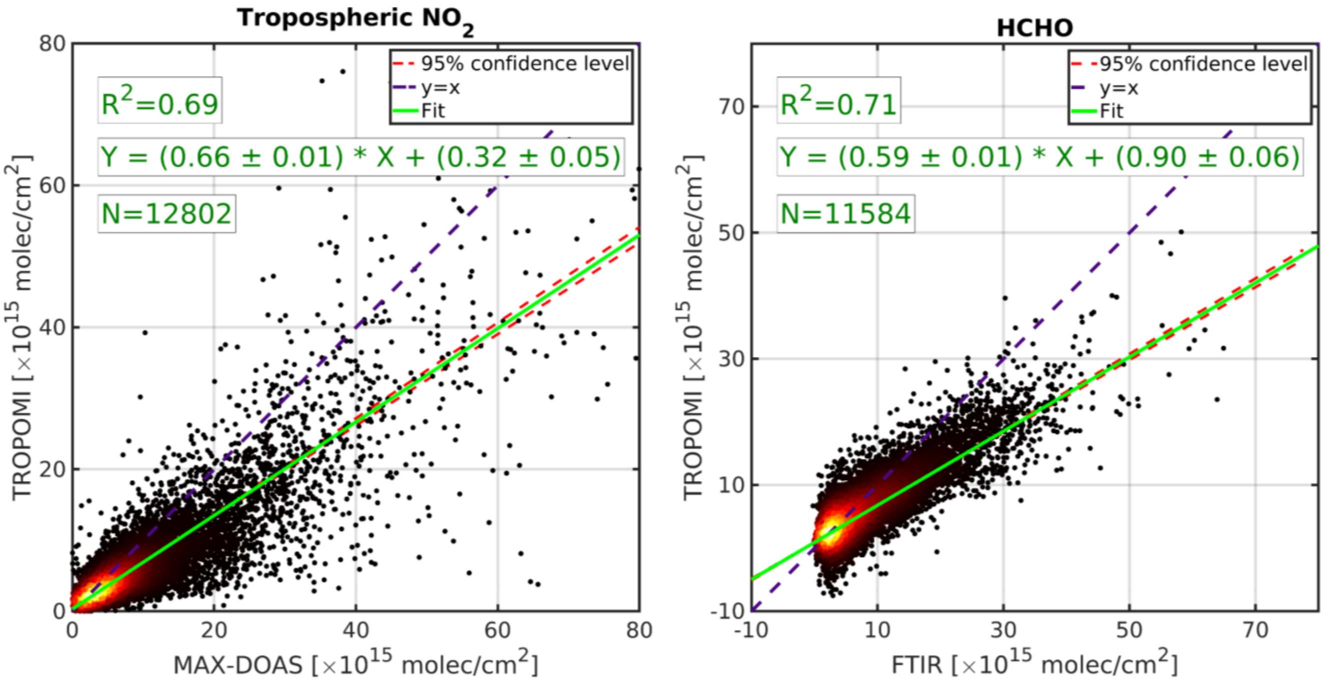

Figure 8. The comparison of TROPOMI tropospheric NO_2 and MAX-DOAS (left) and TROPOMI HCHO and FTIR (right). The data points cover the period of 2018-2023. Both errors of in-situ measurements and TROPOMI are considered in the fit. The data curation procedure has been discussed in Verhoelst et al. (2021) and Vigouroux et al. (2020). The slope smaller than one suggests that both HCHO and NO_2 retrievals are underestimated in polluted regions.

4.3.4. Maps of PO_3 across various regions and qualitative description

Taking advantage of the wealth of bias-corrected TROPOMI observations, we present the first-ever reported PO_3 maps at 0.1×0.1 degrees in the PBL in July 2019 across various geographic regions. Moreover, because of the explicit nature of our algorithm, it is straightforward to break down the contributors of PO_3 to gather insights into how each driver has shaped the distribution of PO_3. Therefore, in addition to PO_3 maps, we will show the magnitudes of various drivers of PO_3 including NO_2, HCHO, and FNR concentrations in the PBL region, the sum of scaled jO^1D and jNO_2 values, along with their contributions to PO_3. It is worth noting that these maps are only a snapshot of PO_3 whose precursors can have large interannual and interdecadal variability caused by meteorology, chemistry, and emissions. A discussion on each region follows:

Africa and the Middle East – Figure 9 illustrates the accelerated rates of PO_3 over the region, particularly concentrated over major cities such as Tehran (Iran), Cairo (Egypt), Riyadh (Saudi), Baghdad (Iraq), Algiers (Algeria), and Johannesburg (South Africa). These urban areas consistently experience poor air quality episodes (e.g., Chaichan et al., 2016; Belhout et al., 2018; Yousefian et al., 2020; Thompson et al., 2014; Boraiy et al., 2023; Choi and Souri et al. 2015a). The biomass burning activities in Africa (see Figure 1 in Roberts et al., 2009) significantly contribute to the high rates of PO_3. Moreover, we see accelerated PO_3 over the Persian Gulf, a region housing oil and gas production facilities, leading to high PO_3 in the region (Lelieveld et al., 2009; Choi and Souri et al. 2015a). Figure 10 shows NO_2 and HCHO concentrations are highly correlated in the Middle East (r=0.82) due to co-emitted NO_x and VOC emissions, predominantly from anthropogenic sources. Over the whole region, HCHO and NO_2 concentrations are only moderately correlated (r=0.61). This is because there is strong spatial heterogeneity associated with NO_x and VOC emissions over Africa that are not spatially correlated. One possible explanation for this could be the emission dependence on the type of fire combustion in Africa (van der Velde et al., 2021) and the location of biogenic isoprene emissions (Marais et al., 2014). For the most part, FNRs tend to fall in ranges above >3.5 (LASSO group 4, highly NO_x-sensitive). However, lower FNRs are prevalent in the core of cities due to elevated NO_x emissions. The contributions of HCHO to PO_3 occur predominantly over areas with low FNRs. These results suggest that NO_x emissions dictate the location of maximum VOC contributions to PO_3. The contribution of NO_2 to PO_3 behaves non-linearly with negative values at the core of cities such as Johannesburg and Tehran (Figure S2). Photolysis rates are high over low SZA and bright surface albedo (i.e., arid land). Accordingly, photolysis rates exhibit a latitudinal gradient in response to changes in SZA. Greater contributions of photolysis rates to PO_3 are observed in areas with low FNRs, as determined by the LASSO estimator (Table 1).

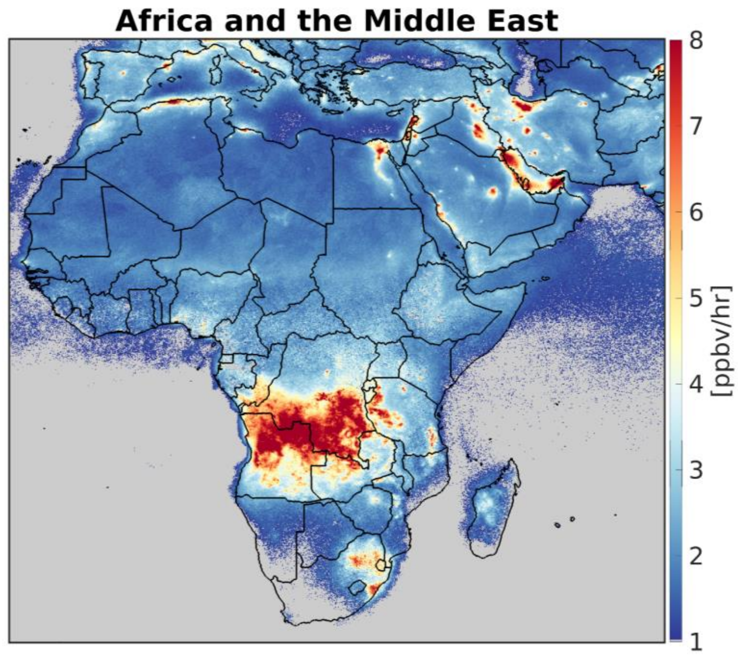

Figure 9. The spatial distribution of PO_3 within the PBL region averaged over July 2019 in Africa and the Middle East. PO_3<1 ppbv/hr is masked due to the algorithm deficiencies. Accelerated PO_3 can be seen over major cities and biomass burning activities in Africa.

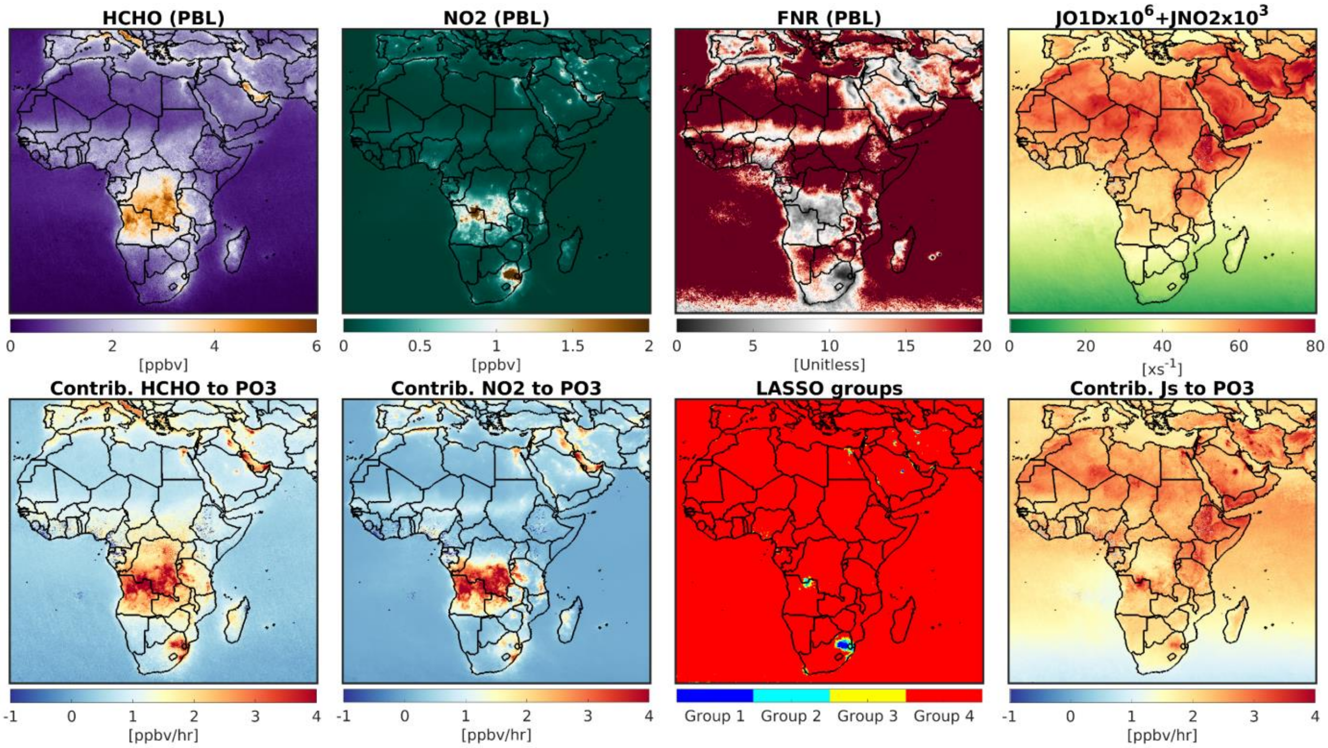

Figure 10. (first row) PBL concentrations of HCHO, NO_2, FNR and sum of scaled jO^1D and jNO_2 derived from TROPOMI and models in July 2019; (second row) the contributions of HCHO, NO_2, and photolysis rates to PO_3, along with the defined LASSO ozone production sensitivity regimes for PO_3 estimates.

Contiguous United States (CONUS) – New York City, Los Angeles (LA), the San Francisco Bay area, and Lake Michigan areas all experience accelerated PO_3 in July 2019, as shown in Figure 11. All the regions fall into non-attainment regions (marginal to extreme) with respect to ozone standards and have been immensely studied (Wu et al., 2024; Kim et al., 2022; Stainer et al., 2021). A robust relationship between PO_3 and ozone concentrations can only be established by factoring in physical processes such as horizontal and vertical transport, dry deposition rates, and background values. In regions with high background ozone concentrations, for example in mountainous areas, even a moderate level of PO_3 can elevate ozone concentration to unhealthy levels. Conversely, if there is a strong correlation between PO_3 and frequent ozone exceedances, such as those observed in the mentioned U.S. cities, it indicates that locally produced ozone through chemical reactions is the primary factor contributing to those events. Except for LA, the vast majority of CONUS fall into large FNRs (>3.5), meaning NO_2 levels largely shape the spatial distribution of PO_3 (Figure 12). HCHO levels are found to be relatively large over LA, causing PO_3 to increase due to its greater sensitivity to VOCs. In addition to high levels of HCHO and NO_2 in several Californian regions, accelerated photochemistry caused by the bright surface albedo enhances PO_3.

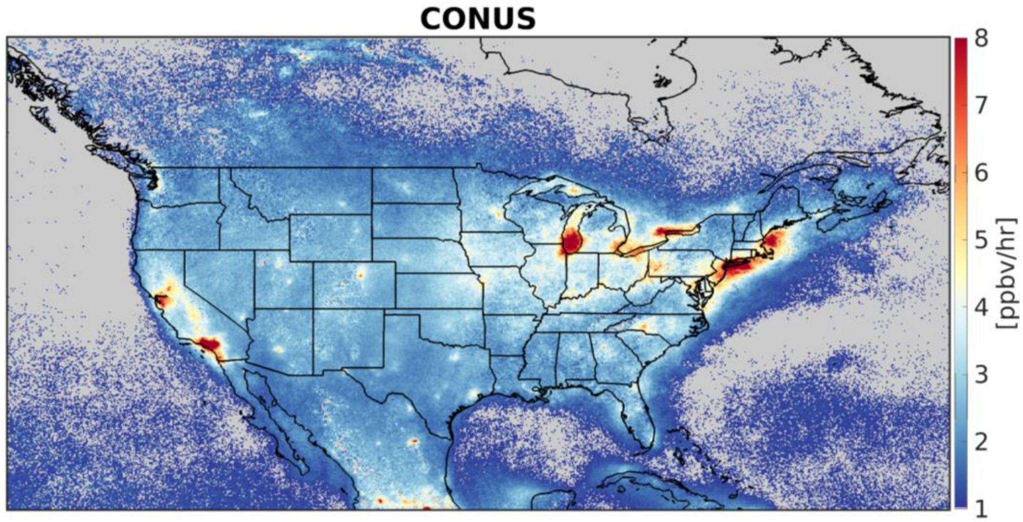

Figure 11. Same as Figure 9 but for CONUS. Elevated PO_3 prevails over various areas such as New York City, Los Angeles, San Francisco Bay area, and Lake Michigan.

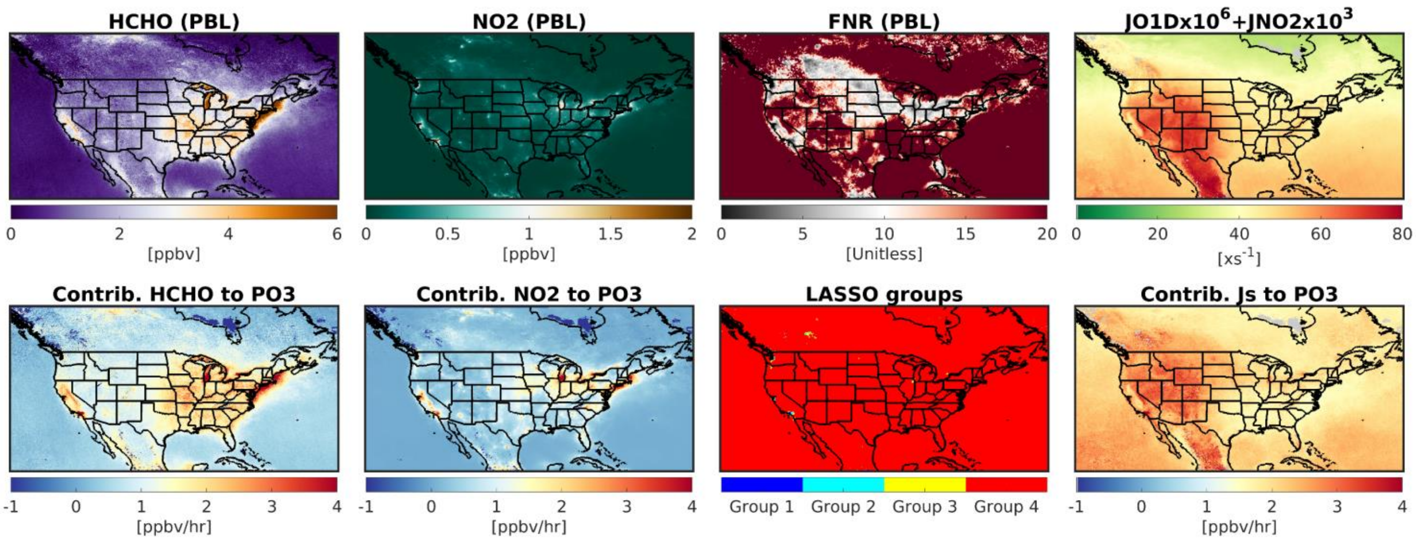

Figure 12. Same as Figure 10 but for CONUS.

East and Southeast Asia – Figure 13 shows extremely accelerated PO_3 values over East Asia, particularly over North China Plain, Yangtze River Delta, Pearl River Delta, and Seoul. These regions have experienced severely degraded air quality with respect to ozone (Souri et al., 2020a,b; Li et al., 2019; Colombi et al., 2023; Schroeder et al., 2020; Wang et al., 2017; Zhang et al., 2007). In southeast Asia, Hanoi (Vietnam), Kuala Lumpur (Malaysia), and Jakarta (Indonesia), undergoing heightened PO_3 as well, have received less attention in literature (Ahamad et al., 2020; Kusumaningtyas et al., 2024; Sakamoto et al., 2018). Figure 14 suggests that the chemical condition of many regions in China and South Korea, falling within the transitional regimes (LASSO group 2 and 3, 1.5<FNR<3.5), has made them susceptible to high PO_3 levels due to concurrent high concentrations of HCHO and NO_2. Moreover, photochemistry appears to be active throughout the region.

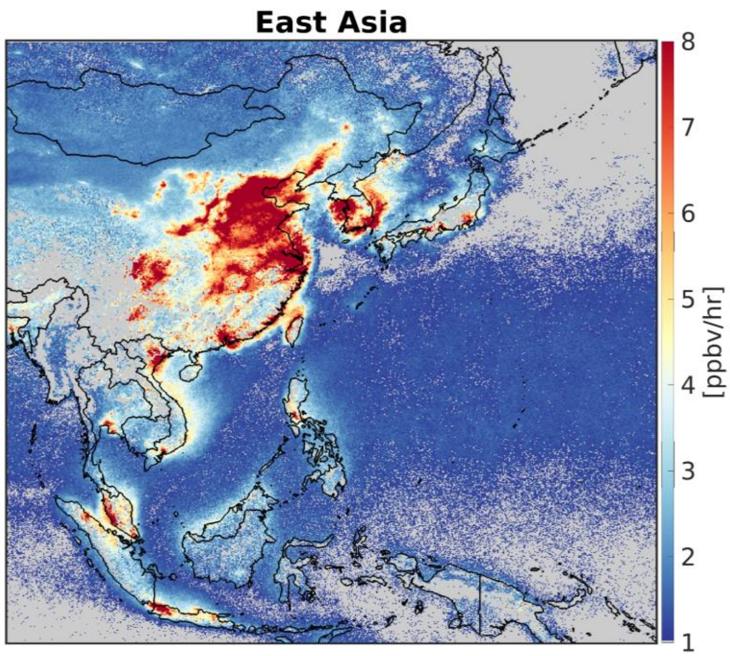

Figure 13. Same as Figure 9 but for east and southeast Asia. Because of heightened amount of photochemistry, NO_2, and HCHO, we observe accelerated PO_3 throughout the majority of the cities in East and Southeast Asia.

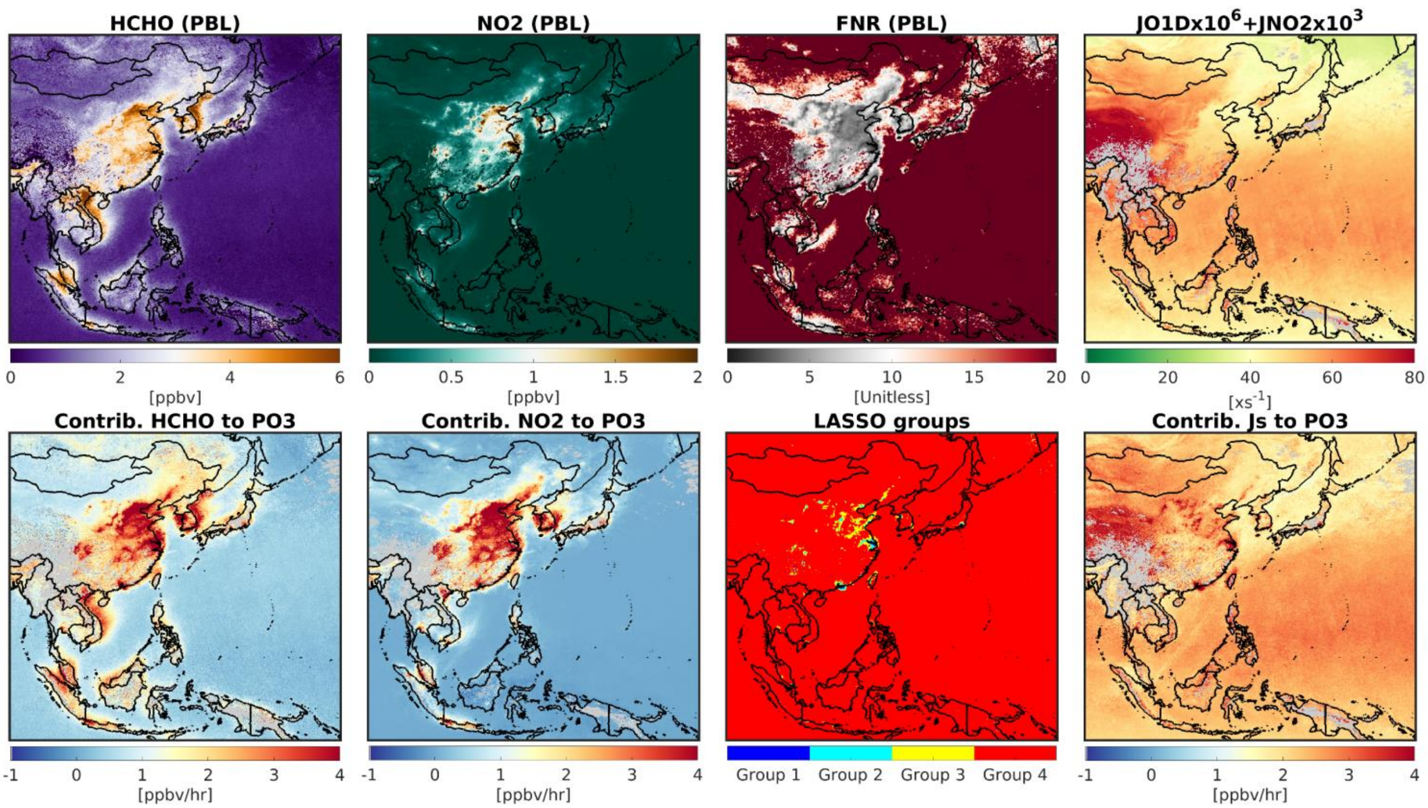

Figure 14. Same as Figure 10 but for east and southeast Asia.

Europe – Figure 15 reveals high PO_3 over Benelux (Belgium, The Netherlands, and Luxembourg), Po Valley (Italy), and several major cities such as Barcelona (Spain) and Rome (Italy). Benelux has the largest hotspot of PO_3 in the region (e.g., Zara et al., 2021). A significant portion of England, Benelux, fall into VOC-sensitive, or the transitional regime (FNR<2.5) shown in Figure 16. Because of diminished photochemistry in these high latitude regions, we do not see significant PBL concentrations of HCHO in order for PO_3 to be as high as the previous areas; moreover, the non-linear NO_x feedback has led to negative contributions of NO_2 to PO_3 in several cities such as London. In general, low photolysis rates compared to the previous regions have made most of Europe less prone to elevated PO_3.

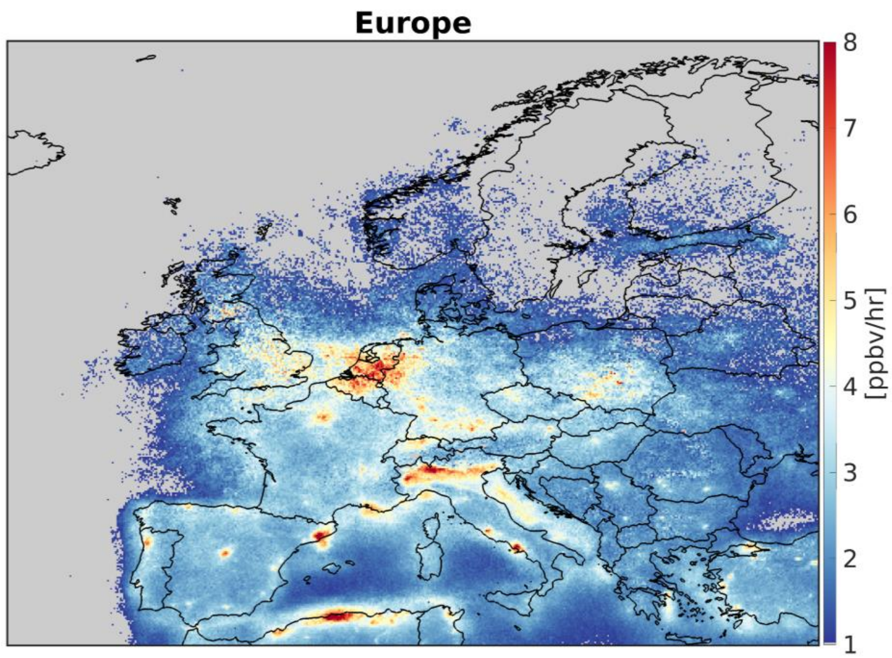

Figure 15. Same as Figure 9 but for Europe. Because of reduced photochemistry, PO_3 values tend to be smaller than the previous cases. Benelux has experienced the highest PO_3 in this region.

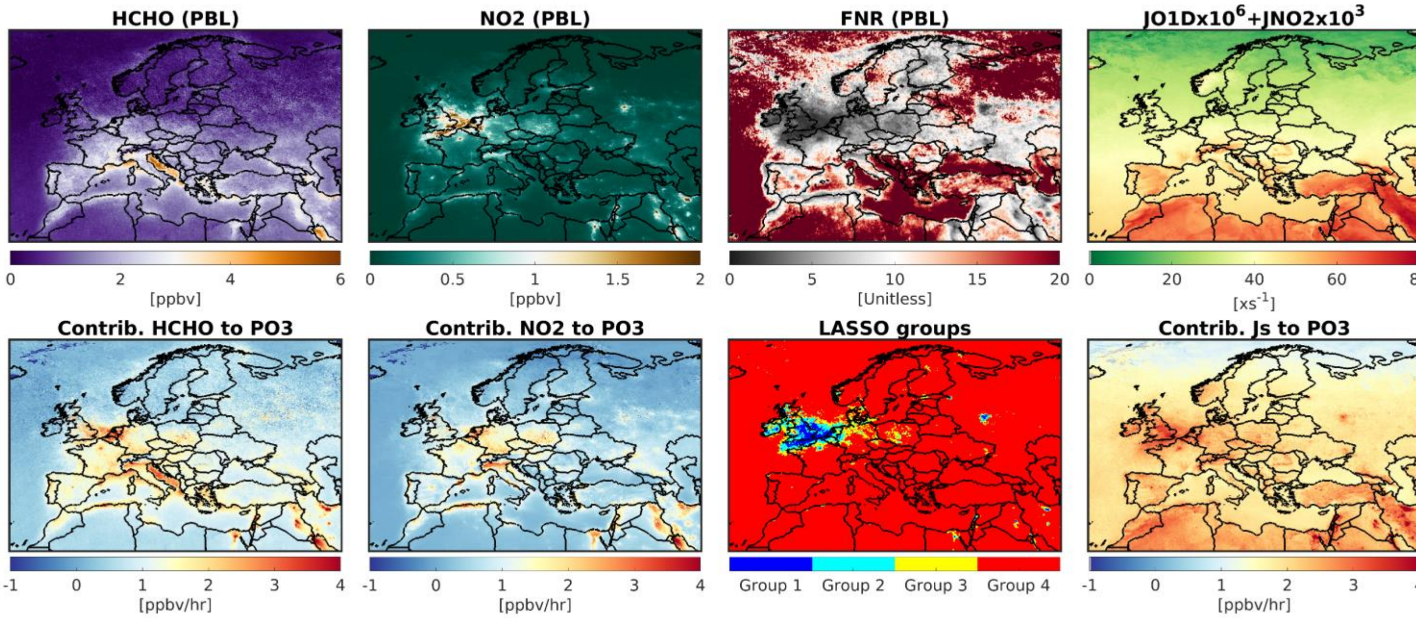

Figure 16. Same as Figure 10 but for Europe.

4.3.5. Seasonality of PO_3 over the Middle East

It is attractive to study the seasonal variations in the contributors to PO_3 over several major cities because the PO_3 drivers' seasonality can vary from location to location. We decide to focus on several Middle Eastern countries that have experienced rapid growth and degraded air quality: Cairo (Egypt), Ghaza (Palestine), Baghdad (Iraq), Riyadh (Saudi Arabia), Tehran (Iran), and the Persian Gulf region. We illustrate the seasonality of four major contributors to PO_3 including NO_2, HCHO, jNO_2, and jO^1D in 2019 in Figure 17.

The levels of HCHO (a proxy for VOCs) consistently have the greatest impact on PO_3 throughout the year in these regions. Specifically, both Baghdad and Tehran experience high levels of HCHO even during colder months, which can be observed using TROPOMI. This suggests that regulations targeting the reduction of man-made VOC emissions should be prioritized in this region. PO_3 levels over Cairo, Gaza, Baghdad, and the Persian Gulf peak during summertime, while Tehran experiences a comparable peak in the autumn due to increased VOC emissions. Additionally, we notice a decrease in PO_3 levels over the Persian Gulf and Riyadh in July, possibly due to a decline in HCHO contributions caused by meteorological factors. Even though NO_2 concentrations decline in summertime due to shorter lifetime against OH, the higher amount of HCHO makes PO_3 more sensitive to NO_2 in this season. Ghaza shows the least seasonal variation among these regions, likely due to consistently active photochemistry throughout the year.

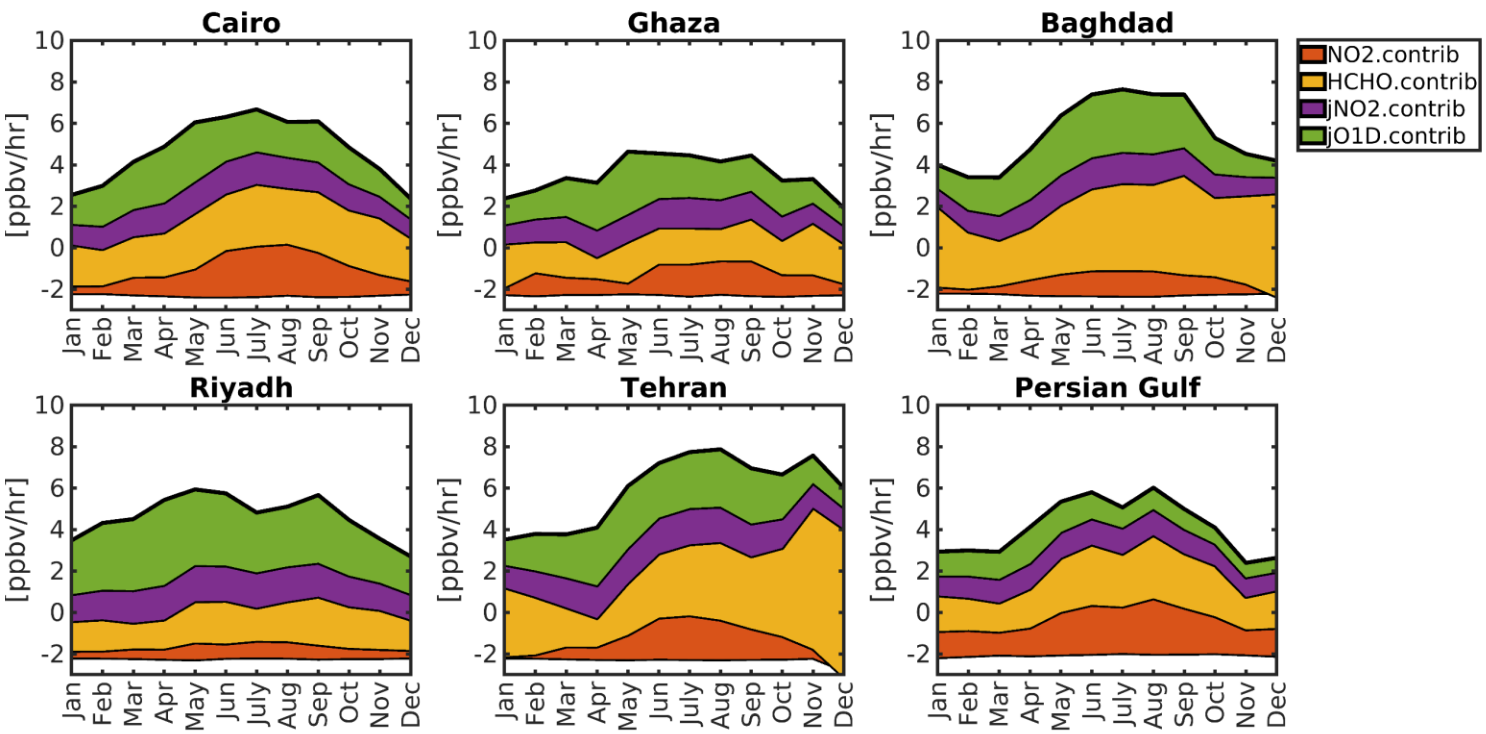

Figure 17. The contributions of NO_2, HCHO, jNO_2, and jO^1D to the PBL PO_3 for several major regions in the Middle East. These estimates are based on the proposed algorithm integrating TROPOMI, ground-based remote sensing, and atmospheric models, to estimate PO_3 based upon a statistical approach. PO_3 tends to spike around the summer due to increased HCHO, higher sensitivity of PO_3 to NO_x, and enhanced photochemistry. However, Tehran shows a second peak in autumn due to unusual high values of HCHO.

4.3.6. The effect of satellite errors on PO_3

Satellite retrieval errors have been identified as the primary obstacle to achieving a robust understanding of ozone chemistry using HCHO and NO_2 data (Souri et al., 2023; Johnson et al., 2023); therefore, generating uncertainty maps is crucial for informing the scientific community about the credibility of our PO_3 estimates. In this study, we utilize the equations outlined in Section 2.2.1 to propagate the errors of HCHO and NO_2 retrievals to the final PO_3 estimates. We achieve this by recalculating the PO_3 value for a given pixel 10,000 times, with each recalculation based on a sample drawn from a normal distribution with a standard deviation equal to the satellite total error. The standard deviation of these samples offers a good approximation of the impact of satellite errors on PO_3 estimates.

Figure 18 illustrates the maps of PO_3 absolute and relative errors over the targeted regions in the course of the month of July. The errors of PO_3 estimates tend to be high (> 50%) in remote regions where the trace gas signals are small. However, the PO_3 errors are within 10-20% in polluted regions where the signals are larger. Currently, the absence of absolute measurements of PO_3 at this vast spatial coverage makes it challenging to judge the severity of these errors for PO_3 applications. Nonetheless, any application based on this product should be recalculated within the reported errors through a Monte-Carlo to gauge the significance of the outcome.

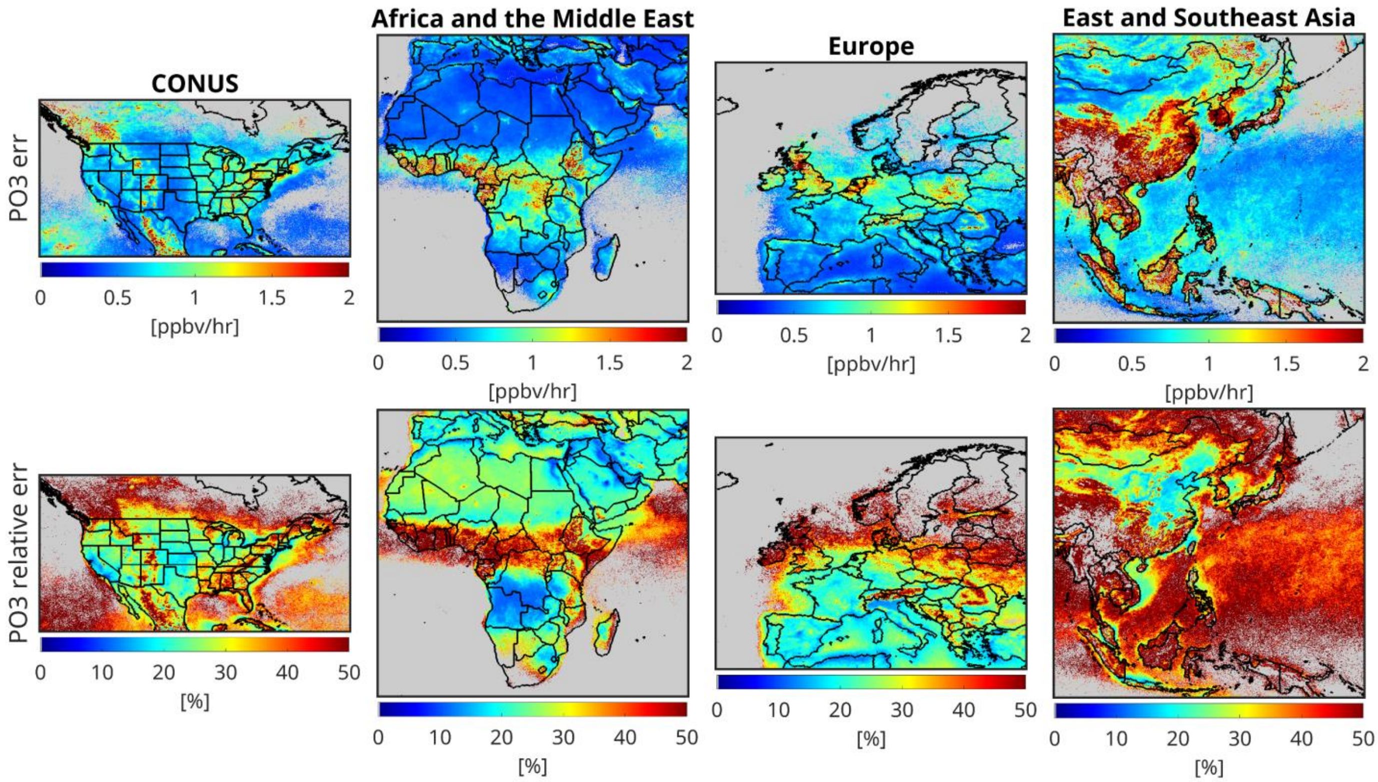

Figure 18. The influence of the satellite errors on PO_3 estimates (absolute and relative) over four major regions tackled in this work. The errors are based on monthly-averaged TROPOMI errors. The errors tend to be mild over polluted regions (10-20%) but they can exceed above 50% over pristine ones.

5. Conclusion

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. These data have indeed extended the use of formaldehyde (HCHO) over nitrogen dioxide (NO_2) ratios (FNR) beyond their current role. Through this product, we can shed light on the effects of emission regulations, wildfires, widespread lockdown, wars, and economic recessions on PO_3 levels. Furthermore, given the long-term records of satellite observations (e.g., OMI since 2005 and TROPOMI since 2018), this product can inform emission regulators about locally-produced ozone hotspots, and ultimately, enhance our understanding of the spatiotemporal variability of ozone formation for over two decades.

In this study, we generated PO_3 maps within the planetary boundary layer (PBL), constrained by bias-corrected TROPOspheric Monitoring Instrument (TROPOMI) observations, using a piecewise regularized regression model. This model was calibrated using a blend of data from a comprehensive suite of aircraft observations and a well-characterized box model. These maps, produced for various regions, allowed us to identify hotspots of locally-produced ozone pollution with unprecedented resolution. Our findings indicated that numerous urban areas in the Middle East, East Asia, and Southeast Asia exhibit accelerated PO_3 rates (>8 ppbv/hr), attributed to high levels of anthropogenic nitrogen oxides ($NO_x = NO + NO_2$), volatile organic compounds (VOCs), and active photochemistry. In contrast, such elevated PO_3 levels were less prevalent in the United States and Europe, with exceptions including Los Angeles, New

York City, and the entire region of the Benelux. Additionally, biomass burning activities in Africa contributed to high PO_3 rates across extensive areas. Seasonality of PO_3 peaked around the summer for several regions in the Middle East because of active photochemistry and concurrent large HCHO and NO_2 levels; however, Tehran experienced elevated PO_3 in the autumn due to large HCHO values possibly produced from anthropogenic emissions.

The production of these maps relied heavily on a robust training dataset. To this end, we incorporated an extensive array of aircraft observations from multiple atmospheric composition campaigns, including DISCOVER-AQ, KORUS-AQ, INTEX-B, ATOM, and SENEX, into the Framework for 0-D Atmospheric Modeling (F0AM) photochemical box model. The box model demonstrated a high level of correspondence ($R^2 > 0.6$, with minimal biases) between several unconstrained compounds (e.g., HCHO, OH, HO_2, PAN, NO, and NO_2) and their observed counterparts, indicating its effectiveness in understanding local ozone chemistry. Utilizing a classification algorithm applied to the data obtained from the constrained box model, we identified HCHO, NO_2, their ratio (known as FNR), photolysis rates, and, to some extent, meteorological factors as good candidates for reproducing PO_3 variability and magnitudes.

Subsequently, we employed a piecewise linear model known as LASSO, which is capable of feature selection by eliminating unimportant inputs, to parameterize PO_3. A key component of this parameterization was the use of FNR to empirically linearize the non-linear ozone chemistry. The LASSO algorithm indicated that more than 88% of the variance in PO_3 could be reproduced with low bias using only five parameters: FNR, HCHO, NO_2, jNO_2 (photolysis rates for NO_2 + hv), and jO^1D (photolysis rates for O_3 + hv). This parameterization demonstrated remarkable performance for the majority of air parcels collected in moderately to extremely polluted regions (PO_3 > 1 ppbv/hr). However, it performed poorly in pristine regions due to the exclusion of certain ozone loss pathways, such as HO_x ($OH+HO_2$), which are more challenging to predict.

Fortunately, TROPOMI provided critical data to enhance the representation of FNR, HCHO, NO_2, jNO_2, and jO^1D. We utilized TROPOMI's viewing geometry, UV surface albedo, and total ozone overhead from a model to predict jNO_2 and jO^1D using look-up tables derived from NCAR's TUV model. To convert TROPOMI tropospheric NO_2 and HCHO columns to their PBL mixing ratios, we employed the MERRA2GMI global transport model, extensively used in various studies. However, the coarse resolution of this model might have introduced underrepresentation issues, which could be mitigated by using higher spatial resolution models in future research.

To address the biases associated with TROPOMI observations, we updated comparisons from Verhoelst et al. (2021) and Vigouroux et al. (2020) with a larger dataset of paired TROPOMI and FTIR/MAX-DOAS measurements. TROPOMI retrievals significantly underestimated HCHO and NO_2 magnitudes in polluted regions (slope ~0.6 - 0.7) and moderately overestimated them in pristine areas. These biases were corrected using regression lines, enabling a relatively unbiased application of the data.

To build confidence in our product, we propagated TROPOMI HCHO and NO_2 errors to PO_3 estimates using a Monte Carlo approach. Results indicated that PO_3 estimates were uncertain (>50%) in clean regions due to a low trace gas signal in TROPOMI retrievals. However, in polluted regions, the errors were more moderate (10-20%) due to the stronger signal.

Over the years, extensive efforts have been devoted to measuring various critical atmospheric compounds globally, developing robust atmospheric models, and enhancing satellite retrievals along with their benchmarks. These advancements have enabled us to estimate PO_3 maps within the PBL. Nonetheless, it is crucial to acknowledge some limitations of our work, many of which are the focus of ongoing research within our team:

i) The direct measurement of PO_3 using specialized instruments (Cazorla and Brune, 2010; Sadanaga et al., 2017; Sklaveniti et al., 2018) is lacking in most atmospheric composition datasets, limiting our ability to fully understand the effects of assumptions (such as the exclusion of heterogeneous chemistry) made in the box model on PO_3.

ii) There is potential for improvement in the parameterization process by employing more sophisticated algorithms, such as neural networks, which could increase the variance explained in the predicted PO_3.

iii) The conversion of satellite column data to PBL mixing ratios requires error characterization and the use of finer-resolution models that are comparable in size to the PO_3 grid boxes.

iv) Partially cloudy pixels and aerosols can affect photolysis rates, which should be considered in future parameterization efforts.

It is important to recognize that PO_3 maps are just one piece of the puzzle when it comes to determining ozone concentrations. Several studies have indicated that accurately representing surface ozone is challenging due to difficulties in representing background ozone, transport, and dry deposition rates. Therefore, we advise against directly linking high PO_3 rates from our product to increased unhealthy ozone exposure. However, our product does provide indications as to whether heightened ozone concentrations are associated with chemistry contributions as opposed to other processes (e.g., meteorology or dry deposition rates). Further investigation using additional tools/data is necessary to gather a full picture of these processes.

Despite these limitations, our novel product offers an asset to the atmospheric science community. It provides a more comprehensive understanding of the complexities associated with spatiotemporal variability associated with the non-linear ozone chemistry at a large domain and enhances confidence in high-resolution maps of chemically-produced ozone hotspots.

Financial Support

This study is funded by NASA's ACMAP/Aura project (grant no. 80NSSC23K1250). The measurements at Paramaribo have been supported by the BMBF (German Ministry of Education and Research) in project ROMIC-II's subproject TroStra (01LG1904A). The NDACC FTIR stations Bremen, Garmisch, Izaña, Ny-Ålesund, Paramaribo, and Karlsruhe have been supported by the German Bundesministerium für Wirtschaft und Energie (BMWi) via DLR5 under grants 50EE1711A, B, and D. The measurements and data analysis at Bremen are supported by the Senate of Bremen. The NCAR FTS observation programs at Thule, GR, Boulder, CO, and Mauna Loa, HI, are supported under contract by the National Aeronautics and Space Administration (NASA). The National Center for Atmospheric Research is sponsored by the National Science Foundation. The Thule effort is also supported by the NSF Office of Polar Programs (OPP). Operations at the Rikubetsu and Tsukuba FTIR sites are supported in part by the GOSAT series project. The Paris TCCON site has received funding from Sorbonne Université, the French research center CNRS, and the French space agency CNES. The Jungfraujoch FTIR data are primarily available thanks to the support provided by the F.R.S. FNRS (Brussels), the GAW-CH program of MeteoSwiss (Zürich), and the HFSJG.ch Foundation (Bern). IUP-Bremen ground-based measurements are funded by DLR-Bonn and received through project 50EE1709A. KNMI ground-based measurements in De Bilt and Cabauw are partly supported by the Ruisdael Observatory project, Dutch Research Council (NWO) contract 184.034.015, by the Netherlands Space Office (NSO) for Sentinel-5p/TROPOMI validation, and by ESA via the EU CAMS project.

Data Availability

TROPOMI satellite data are derived from copernicus Sentinel-5P (processed by ESA), 2021, TROPOMI Level 2 Nitrogen Dioxide total column products. Version 02. European Space Agency. https://doi.org/10.5270/S5P-9bnp8q8, and copernicus Sentinel-5P (processed by ESA), 2020, TROPOMI Level 2 Formaldehyde Total Column products. Version 02. European Space Agency. https://doi.org/10.5270/S5P-vg1i7t0. The FTIR and MAX-DOAS observations were partly obtained from the Network for the Detection of Atmospheric Composition Change (NDACC) and are available through the NDACC website at http://www.ndacc.org. The box model can be obtained from https://github.com/AirChem/F0AM (last access: 10 Nov, 2024). The TROPOMI UV DLER can be obtained from https://www.temis.nl/surface/albedo/tropomi_ler.php (last access: 10 Nov 2024).

Competing interests

Bryan N. Duncan is a member of the editorial board of Atmospheric Chemistry and Physics

Acknowledgements

We thank all principal investigators, pilots, and managers who collected the aircraft data used in our research and made them publicly available. We thank the FTIR HCHO measurement team of Thomas Blumenstock, Martine De Mazière, Michel Grutter, James W. Hannigan, Nicholas Jones, Rigel Kivi, Erik Lutsch, Emmanuel Mahieu, Maria Makarova, Isamu Morino, Isao Murata, Tomoo Nagahama, Justus Notholt, Ivan Ortega, Mathias Palm, Amelie Röhling, Matthias Schneider, Dan Smale, Wolfgang Stremme, Kim String, Youwen Sun, Ralf Sussmann, Yao Té, and Pucai Wang. We thank the Meteorological Service Suriname and Cornelis Becker for their support. The MAX-DOAS data used in this publication were obtained from Alkis Bais, John Burrows, Ka Lok Chan, Michel Grutter, Cheng Liu, Hitoshi Irie, Vinod Kumar, Yugo Kanaya, Ankie Piters, Claudia Rivera-Cárdenas, Andreas Richter, Michel Van Roozendael, Robert Ryan, Vinayak Sinha, and Thomas Wagner. Fast delivery of MAX-DOAS data tailored to the S5P validation was organized through S5PVT AO project NID-FORVAL. We thank the IISER Mohali atmospheric chemistry facility for supporting the MAX-DOAS measurements at Mohali, India. We thank Julie M. Nicely for providing merged ATOMs observations.

Authors' contributions

AHS designed and implemented the research idea, analyzed the data, made all figures, and wrote the manuscript. TV, CV, GP, SC, and BL provided the paired TROPOMI and benchmark data. Other authors helped with the analysis, the model setup, and interpretation.

References:

Ahamad, F., Griffiths, P. T., Latif, M. T., Juneng, L., and Xiang, C. J.: Ozone Trends from Two Decades of Ground Level Observation in Malaysia, Atmosphere, 11, 755, https://doi.org/10.3390/atmos11070755, 2020.

Anderson, D. C., Follette-Cook, M. B., Strode, S. A., Nicely, J. M., Liu, J., Ivatt, P. D., and Duncan, B. N.: A machine learning methodology for the generation of a parameterization of the hydroxyl radical, Geosci. Model Dev., 15, 6341–6358, https://doi.org/10.5194/gmd-15-6341-2022, 2022.

Archibald, A. T., Jenkin, M. E., and Shallcross, D. E.: An isoprene mechanism intercomparison, Atmos. Environ., 44, 5356–5364, https://doi.org/10.1016/j.atmosenv.2009.09.016, 2010.

Baylon, P., Jaffe, D. A., Hall, S. R., Ullmann, K., Alvarado, M. J., and Lefer, B. L.: Impact of Biomass Burning Plumes on Photolysis Rates and Ozone Formation at the Mount Bachelor Observatory, J. Geophys. Res. Atmos., 123, 2272–2284, https://doi.org/10.1002/2017JD027341, 2018.

Beddows, D. C. S., Dall'Osto, M., and Harrison, R. M.: Cluster Analysis of Rural, Urban, and Curbside Atmospheric Particle Size Data, Environ. Sci. Technol., 43, 4694–4700, https://doi.org/10.1021/es803121t, 2009.

Belhout, D., Kerbachi, R., Relvas, H., and Miranda, A. I.: Air quality assessment in Algiers city, Air. Qual. Atmos. Health., 11, 897–906, https://doi.org/10.1007/s11869-018-0589-x, 2018.

Boraiy, M., El-Metwally, M., Wheida, A., El-Nazer, M., Hassan, S. K., El-Sanabary, F. F., Alfaro, S. C., Abdelwahab, M., and Borbon, A.: Statistical analysis of the variability of reactive trace gases (SO2, NO2 and ozone) in Greater Cairo during dust storm events, J. Atmos. Chem., 80, 227–250, https://doi.org/10.1007/s10874-023-09449-4, 2023.

Bottorff, B., Lew, M. M., Woo, Y., Rickly, P., Rollings, M. D., Deming, B., Anderson, D. C., Wood, E., Alwe, H. D., Millet, D. B., Weinheimer, A., Tyndall, G., Ortega, J., Dusanter, S., Leonardis, T., Flynn, J., Erickson, M., Alvarez, S., Rivera-Rios, J. C., Shutter, J. D., Keutsch, F., Helmig, D., Wang, W., Allen, H. M., Slade, J. H., Shepson, P. B., Bertman, S., and Stevens, P. S.: OH, HO2, and RO2 radical chemistry in a rural forest environment: measurements, model comparisons, and evidence of a missing radical sink, Atmos. Chem. Phys., 23, 10287–10311, https://doi.org/10.5194/acp-23-10287-2023, 2023.

Brune, W. H., Miller, D. O., Thames, A. B., Allen, H. M., Apel, E. C., Blake, D. R., Bui, T. P., Commane, R., Crounse, J. D., Daube, B. C., Diskin, G. S., DiGangi, J. P., Elkins, J. W., Hall, S. R., Hanisco, T. F., Hannun, R. A., Hintsa, E. J., Hornbrook, R. S., Kim, M. J., McKain, K., Moore, F. L., Neuman, J. A., Nicely, J. M., Peischl, J., Ryerson, T. B., St. Clair, J. M., Sweeney, C., Teng, A. P., Thompson, C., Ullmann, K., Veres, P. R., Wennberg, P. O., and Wolfe, G. M.: Exploring Oxidation in the Remote Free Troposphere: Insights From Atmospheric Tomography (ATom), J. Geophys. Res. Atmos., 125, e2019JD031685, https://doi.org/10.1029/2019JD031685, 2020.

Brune, W. H., Miller, D. O., Thames, A. B., Brosius, A. L., Barletta, B., Blake, D. R., Blake, N. J., Chen, G., Choi, Y., Crawford, J. H., Digangi, J. P., Diskin, G., Fried, A., Hall, S. R., Hanisco, T. F., Huey, G. L., Hughes, S. C., Kim, M., Meinardi, S., Montzka, D. D., Pusede, S. E., Schroeder, J. R., Teng, A., Tanner, D. J., Ullmann, K., Walega, J., Weinheimer, A., Wisthaler, A., and Wennberg, P. O.: Observations of atmospheric oxidation and ozone production in South Korea, Atmos. Environ., 269, 118854, https://doi.org/10.1016/j.atmosenv.2021.118854, 2022.

Cazorla, M. and Brune, W. H.: Measurement of Ozone Production Sensor, Atmos. Meas. Tech., 3, 545–555, https://doi.org/10.5194/amt-3-545-2010, 2010.

Cazorla, M., Brune, W. H., Ren, X., and Lefer, B.: Direct measurement of ozone production rates in Houston in 2009 and comparison with two estimation methods, Atmos. Chem. Phys., 12, 1203–1212, https://doi.org/10.5194/acp-12-1203-2012, 2012.

Chaichan, M. T., Kazem, H. A., and Abed, T. A.: Traffic and outdoor air pollution levels near highways in Baghdad, Iraq, Environ. Dev. Sustain., 20, 589–603, https://doi.org/10.1007/s10668-016-9900-x, 2018.

Chatfield, R. B., Ren, X., Brune, W., and Schwab, J.: Controls on urban ozone production rate as indicated by formaldehyde oxidation rate and nitric oxide, Atmos. Environ., 44, 5395–5406, https://doi.org/10.1016/j.atmosenv.2010.08.056, 2010.

Choi, J., Henze, D. K., Cao, H., Nowlan, C. R., González Abad, G., Kwon, H.-A., Lee, H.-M., Oak, Y. J., Park, R. J., Bates, K. H., Maasakkers, J. D., Wisthaler, A., and Weinheimer, A. J.: An Inversion Framework for Optimizing Non-Methane VOC Emissions Using Remote Sensing and Airborne

Observations in Northeast Asia During the KORUS-AQ Field Campaign, J. Geophys. Res. Atmos., 127, e2021JD035844, https://doi.org/10.1029/2021JD035844, 2022.

Choi, S., Lamsal, L. N., Follette-Cook, M., Joiner, J., Krotkov, N. A., Swartz, W. H., Pickering, K. E., Loughner, C. P., Appel, W., Pfister, G., Saide, P. E., Cohen, R. C., Weinheimer, A. J., and Herman, J. R.: Assessment of NO2 observations during DISCOVER-AQ and KORUS-AQ field campaigns, Atmos. Meas. Tech., 13, 2523–2546, https://doi.org/10.5194/amt-13-2523-2020, 2020.

Choi, Y. and Souri, A. H.: Chemical condition and surface ozone in large cities of Texas during the last decade: Observational evidence from OMI, CAMS, and model analysis, Remote Sens. Environ., 168, 90–101, https://doi.org/10.1016/j.rse.2015.06.026, 2015a.

Choi, Y. and Souri, A. H.: Seasonal behavior and long-term trends of tropospheric ozone, its precursors and chemical conditions over Iran: A view from space, Atmos. Environ., 106, 232–240, https://doi.org/10.1016/j.atmosenv.2015.02.012, 2015b.

Choi, Y., Kim, H., Tong, D., and Lee, P.: Summertime weekly cycles of observed and modeled NOx and O3 concentrations as a function of satellite-derived ozone production sensitivity and land use types over the Continental United States, Atmos. Chem. Phys., 12, 6291–6307, https://doi.org/10.5194/acp-12-6291-2012, 2012.

Colombi, N. K., Jacob, D. J., Yang, L. H., Zhai, S., Shah, V., Grange, S. K., Yantosca, R. M., Kim, S., and Liao, H.: Why is ozone in South Korea and the Seoul metropolitan area so high and increasing?, Atmos. Chem. Phys., 23, 4031–4044, https://doi.org/10.5194/acp-23-4031-2023, 2023.

Crawford, J. H., Ahn, J.-Y., Al-Saadi, J., Chang, L., Emmons, L. K., Kim, J., Lee, G., Park, J.-H., Park, R. J., Woo, J. H., Song, C.-K., Hong, J.-H., Hong, Y.-D., Lefer, B. L., Lee, M., Lee, T., Kim, S., Min, K.-E., Yum, S. S., Shin, H. J., Kim, Y.-W., Choi, J.-S., Park, J.-S., Szykman, J. J., Long, R. W., Jordan, C. E., Simpson, I. J., Fried, A., Dibb, J. E., Cho, S., and Kim, Y. P.: The Korea–United States Air Quality (KORUS-AQ) field study, Elem. Sci. Anth., 9, 00163, https://doi.org/10.1525/elementa.2020.00163, 2021.

De Smedt, I., Pinardi, G., Vigouroux, C., Compernolle, S., Bais, A., Benavent, N., Boersma, F., Chan, K.-L., Donner, S., Eichmann, K.-U., Hedelt, P., Hendrick, F., Irie, H., Kumar, V., Lambert, J.-C., Langerock, B., Lerot, C., Liu, C., Loyola, D., Piters, A., Richter, A., Rivera Cárdenas, C., Romahn, F., Ryan, R. G., Sinha, V., Theys, N., Vlietinck, J., Wagner, T., Wang, T., Yu, H., and Van Roozendael, M.: Comparative assessment of TROPOMI and OMI formaldehyde observations and validation against MAX-DOAS network column measurements, Atmos. Chem. Phys., 21, 12561–12593, https://doi.org/10.5194/acp-21-12561-2021, 2021.

DiMaria, C. A., Jones, D. B. A., Worden, H., Bloom, A. A., Bowman, K., Stavrakou, T., Miyazaki, K., Worden, J., Guenther, A., Sarkar, C., Seco, R., Park, J.-H., Tota, J., Alves, E. G., and Ferracci, V.: Optimizing the Isoprene Emission Model MEGAN With Satellite and Ground-Based Observational Constraints, J. Geophys. Res. Atmos., 128, e2022JD037822, https://doi.org/10.1029/2022JD037822, 2023.

Duncan, B. N. and Chameides, W. L.: Effects of urban emission control strategies on the export of ozone and ozone precursors from the urban atmosphere to the troposphere, J. Geophys. Res. Atmos., 103, 28159–28179, https://doi.org/10.1029/98JD02145, 1998.

Duncan, B. N., Strahan, S. E., Yoshida, Y., Steenrod, S. D., and Livesey, N.: Model study of the cross-tropopause transport of biomass burning pollution, Atmos. Chem. Phys., 7, 3713–3736, https://doi.org/10.5194/acp-7-3713-2007, 2007.

Duncan, B. N., Yoshida, Y., Olson, J. R., Sillman, S., Martin, R. V., Lamsal, L., Hu, Y., Pickering, K. E., Retscher, C., Allen, D. J., and Crawford, J. H.: Application of OMI observations to a space-based indicator of NOx and VOC controls on surface ozone formation, Atmos. Environ., 44, 2213–2223, https://doi.org/10.1016/j.atmosenv.2010.03.010, 2010.

Fleming, Z. L., Doherty, R. M., von Schneidemesser, E., Malley, C. S., Cooper, O. R., Pinto, J. P., Colette, A., Xu, X., Simpson, D., Schultz, M. G., Lefohn, A. S., Hamad, S., Moolla, R., Solberg, S., and Feng, Z.: Tropospheric Ozone Assessment Report: Present-day ozone distribution and trends relevant to human health, Elem. Sci. Anth., 6, 12, https://doi.org/10.1525/elementa.273, 2018.

Gaudel, A., Cooper, O. R., Ancellet, G., Barret, B., Boynard, A., Burrows, J. P., Clerbaux, C., Coheur, P.-F., Cuesta, J., Cuevas, E., Doniki, S., Dufour, G., Ebojie, F., Foret, G., Garcia, O., Granados-Muñoz, M. J., Hannigan, J. W., Hase, F., Hassler, B., Huang, G., Hurtmans, D., Jaffe, D., Jones, N., Kalabokas, P., Kerridge, B., Kulawik, S., Latter, B., Leblanc, T., Le Flochmoën, E., Lin, W., Liu, J., Liu, X., Mahieu, E., McClure-Begley, A., Neu, J. L., Osman, M., Palm, M., Petetin, H., Petropavlovskikh, I., Querel, R., Rahpoe, N., Rozanov, A., Schultz, M. G., Schwab, J., Siddans, R., Smale, D., Steinbacher, M., Tanimoto, H., Tarasick, D. W., Thouret, V., Thompson, A. M., Trickl, T., Weatherhead, E., Wespes, C., Worden, H. M., Vigouroux, C., Xu, X., Zeng, G., and Ziemke, J.: Tropospheric Ozone Assessment Report: Present-day distribution and trends of tropospheric ozone relevant to climate and global atmospheric chemistry model evaluation, Elem. Sci. Anth., 6, 39, https://doi.org/10.1525/elementa.291, 2018.

Gerasopoulos, E., Kouvarakis, G., Vrekoussis, M., Donoussis, C., Mihalopoulos, N., and Kanakidou, M.: Photochemical ozone production in the Eastern Mediterranean, Atmos. Environ., 40, 3057–3069, https://doi.org/10.1016/j.atmosenv.2005.12.061, 2006.

Govender, P. and Sivakumar, V.: Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019), Atmos. Pollut. Res., 11, 40–56, https://doi.org/10.1016/j.apr.2019.09.009, 2020.

Jeon, W., Choi, Y., Souri, A. H., Roy, A., Diao, L., Pan, S., Lee, H. W., and Lee, S.-H.: Identification of chemical fingerprints in long-range transport of burning induced upper tropospheric ozone from Colorado to the North Atlantic Ocean, Sci. Total Environ., 613–614, 820–828, https://doi.org/10.1016/j.scitotenv.2017.09.177, 2018.

Jin, X., Fiore, A. M., Murray, L. T., Valin, L. C., Lamsal, L. N., Duncan, B., Folkert Boersma, K., De Smedt, I., Abad, G. G., Chance, K., and Tonnesen, G. S.: Evaluating a Space-Based Indicator of Surface Ozone-NOx-VOC Sensitivity Over Midlatitude Source Regions and Application to Decadal Trends, J. Geophys. Res. Atmos., 122, 10,439-10,461, https://doi.org/10.1002/2017JD026720, 2017.

Johnson, M. S., Souri, A. H., Philip, S., Kumar, R., Naeger, A., Geddes, J., Judd, L., Janz, S., Chong, H., and Sullivan, J.: Satellite remote-sensing capability to assess tropospheric-column ratios of formaldehyde and nitrogen dioxide: case study during the Long Island Sound Tropospheric Ozone Study 2018 (LISTOS 2018) field campaign, Atmos. Meas. Tech., 16, 2431–2454, https://doi.org/10.5194/amt-16-2431-2023, 2023.

Kim, S.-W., McDonald, B. C., Seo, S., Kim, K.-M., and Trainer, M.: Understanding the Paths of Surface Ozone Abatement in the Los Angeles Basin, J. Geophys. Res. Atmos., 127, e2021JD035606, https://doi.org/10.1029/2021JD035606, 2022.

Kleinman, L. I., Daum, P. H., Imre, D., Lee, Y.-N., Nunnermacker, L. J., Springston, S. R., Weinstein-Lloyd, J., and Rudolph, J.: Ozone production rate and hydrocarbon reactivity in 5 urban areas: A cause of high ozone concentration in Houston, Geophys. Res. Lett., 29, 105-1-105–4, https://doi.org/10.1029/2001GL014569, 2002.

Kusumaningtyas, S. D. A., Tonokura, K., Muharsyah, R., Gunawan, D., Sopaheluwakan, A., Iriana, W., Lestari, P., Permadi, D. A., Rahmawati, R., and Samputra, N. A. R.: Comprehensive analysis of long-term trends, meteorological influences, and ozone formation sensitivity in the Jakarta Greater Area, Sci. Rep., 14, 9605, https://doi.org/10.1038/s41598-024-60374-2, 2024.

Lelieveld, J., Hoor, P., Jöckel, P., Pozzer, A., Hadjinicolaou, P., Cammas, J.-P., and Beirle, S.: Severe ozone air pollution in the Persian Gulf region, Atmos. Chem. Phys., 9, 1393–1406, https://doi.org/10.5194/acp-9-1393-2009, 2009.

Li, K., Jacob, D. J., Liao, H., Shen, L., Zhang, Q., and Bates, K. H.: Anthropogenic drivers of 2013–2017 trends in summer surface ozone in China, Proc. Natl. Acad. Sci., 116, 422–427, https://doi.org/10.1073/pnas.1812168116, 2019.

Marais, E. A., Jacob, D. J., Guenther, A., Chance, K., Kurosu, T. P., Murphy, J. G., Reeves, C. E., and Pye, H. O. T.: Improved model of isoprene emissions in Africa using Ozone Monitoring Instrument (OMI) satellite observations of formaldehyde: implications for oxidants and particulate matter, Atmos. Chem. Phys., 14, 7693–7703, https://doi.org/10.5194/acp-14-7693-2014, 2014.

Martin, R. V., Fiore, A. M., and Van Donkelaar, A.: Space-based diagnosis of surface ozone sensitivity to anthropogenic emissions, Geophys. Res. Lett., 31, https://doi.org/10.1029/2004GL019416, 2004.

Marvin, M. R., Wolfe, G. M., Salawitch, R. J., Canty, T. P., Roberts, S. J., Travis, K. R., Aikin, K. C., de Gouw, J. A., Graus, M., Hanisco, T. F., Holloway, J. S., Hübler, G., Kaiser, J., Keutsch, F. N., Peischl, J., Pollack, I. B., Roberts, J. M., Ryerson, T. B., Veres, P. R., and Warneke, C.: Impact of evolving isoprene mechanisms on simulated formaldehyde: An inter-comparison supported by in situ observations from SENEX, Atmos. Environ., 164, 325–336, https://doi.org/10.1016/j.atmosenv.2017.05.049, 2017.

Mazzuca, G. M., Ren, X., Loughner, C. P., Estes, M., Crawford, J. H., Pickering, K. E., Weinheimer, A. J., and Dickerson, R. R.: Ozone production and its sensitivity to NOx and VOCs: results from the DISCOVER-AQ field experiment, Houston 2013, Atmos. Chem. Phys., 16, 14463–14474, https://doi.org/10.5194/acp-16-14463-2016, 2016.

Miller, D. O. and Brune, W. H.: Investigating the Understanding of Oxidation Chemistry Using 20 Years of Airborne OH and HO2 Observations, J. Geophys. Res. Atmos., 127, e2021JD035368, https://doi.org/10.1029/2021JD035368, 2022.

Mills, G., Pleijel, H., Malley, C. S., Sinha, B., Cooper, O. R., Schultz, M. G., Neufeld, H. S., Simpson, D., Sharps, K., Feng, Z., Gerosa, G., Harmens, H., Kobayashi, K., Saxena, P., Paoletti, E., Sinha, V., and Xu, X.: Tropospheric Ozone Assessment Report: Present-day tropospheric ozone distribution and trends relevant to vegetation, Elem. Sci. Anth., 6, 47, https://doi.org/10.1525/elementa.302, 2018.

Miyazaki, K., Eskes, H., Sudo, K., Boersma, K. F., Bowman, K., and Kanaya, Y.: Decadal changes in global surface NOx emissions from multi-constituent satellite data assimilation, Atmos. Chem. Phys., 17, 807–837, https://doi.org/10.5194/acp-17-807-2017, 2017.

Orbe, C., Oman, L. D., Strahan, S. E., Waugh, D. W., Pawson, S., Takacs, L. L., and Molod, A. M.: Large-Scale Atmospheric Transport in GEOS Replay Simulations, J. Adv. Model. Earth Syst., 9, 2545–2560, https://doi.org/10.1002/2017MS001053, 2017.

Pan, S., Roy, A., Choi, Y., Eslami, E., Thomas, S., Jiang, X., and Gao, H. O.: Potential impacts of electric vehicles on air quality and health endpoints in the Greater Houston Area in 2040, Atmos. Environ., 207, 38–51, https://doi.org/10.1016/j.atmosenv.2019.03.022, 2019.

Ren, X., van Duin, D., Cazorla, M., Chen, S., Mao, J., Zhang, L., Brune, W. H., Flynn, J. H., Grossberg, N., Lefer, B. L., Rappenglück, B., Wong, K. W., Tsai, C., Stutz, J., Dibb, J. E., Thomas Jobson, B., Luke, W. T., and Kelley, P.: Atmospheric oxidation chemistry and ozone production: Results from SHARP 2009 in Houston, Texas, J. Geophys. Res. Atmos., 118, 5770–5780, https://doi.org/10.1002/jgrd.50342, 2013.

Roberts, G., Wooster, M. J., and Lagoudakis, E.: Annual and diurnal african biomass burning temporal dynamics, Biogeosciences, 6, 849–866, https://doi.org/10.5194/bg-6-849-2009, 2009.

Sadanaga, Y., Kawasaki, S., Tanaka, Y., Kajii, Y., and Bandow, H.: New System for Measuring the Photochemical Ozone Production Rate in the Atmosphere, Environ. Sci. Technol., 51, 2871–2878, https://doi.org/10.1021/acs.est.6b04639, 2017.

Sakamoto, Y., Shoji, K., Bui, M. T., Phạm, T. H., Vu, T. A., Ly, B. T., and Kajii, Y.: Air quality study in Hanoi, Vietnam in 2015–2016 based on a one-year observation of NOx, O3, CO and a one-week observation of VOCs, Atmos. Pollut. Res., 9, 544–551, https://doi.org/10.1016/j.apr.2017.12.001, 2018.

Schroeder, J. R., Crawford, J. H., Ahn, J.-Y., Chang, L., Fried, A., Walega, J., Weinheimer, A., Montzka, D. D., Hall, S. R., Ullmann, K., Wisthaler, A., Mikoviny, T., Chen, G., Blake, D. R., Blake, N. J., Hughes, S. C., Meinardi, S., Diskin, G., Digangi, J. P., Choi, Y., Pusede, S. E., Huey, G. L., Tanner, D. J., Kim, M., and Wennberg, P.: Observation-based modeling of ozone chemistry in the Seoul metropolitan area during the Korea-United States Air Quality Study (KORUS-AQ), Elem. Sci. Anth., 8, 3, https://doi.org/10.1525/elementa.400, 2020.

Schroeder, J. R., Crawford, J. H., Fried, A., Walega, J., Weinheimer, A., Wisthaler, A., Müller, M., Mikoviny, T., Chen, G., Shook, M., Blake, D. R., and Tonnesen, G. S.: New insights into the column CH2O/NO2 ratio as an indicator of near-surface ozone sensitivity, J. Geophys. Res. Atmos., 122, 8885–8907, https://doi.org/10.1002/2017JD026781, 2017.

Sillman, S. and He, D.: Some theoretical results concerning O3-NOx-VOC chemistry and NOx-VOC indicators, J. Geophys. Res. Atmos., 107, ACH 26-1-ACH 26-15, https://doi.org/10.1029/2001JD001123, 2002.

Silva, S. J., Heald, C. L., Ravela, S., Mammarella, I., and Munger, J. W.: A Deep Learning Parameterization for Ozone Dry Deposition Velocities, Geophys. Res. Lett., 46, 983–989, https://doi.org/10.1029/2018GL081049, 2019.

Simpson, W. R., Brown, S. S., Saiz-Lopez, A., Thornton, J. A., and von Glasow, R.: Tropospheric Halogen Chemistry: Sources, Cycling, and Impacts, Chem. Rev., 115, 4035–4062, https://doi.org/10.1021/cr5006638, 2015.

Singh, H. B., Brune, W. H., Crawford, J. H., Flocke, F., and Jacob, D. J.: Chemistry and transport of pollution over the Gulf of Mexico and the Pacific: spring 2006 INTEX-B campaign overview and first results, Atmos. Chem. Phys., 9, 2301–2318, https://doi.org/10.5194/acp-9-2301-2009, 2009.

Sklaveniti, S., Locoge, N., Stevens, P. S., Wood, E., Kundu, S., and Dusanter, S.: Development of an instrument for direct ozone production rate measurements: measurement reliability and current limitations, Atmos. Meas. Tech., 11, 741–761, https://doi.org/10.5194/amt-11-741-2018, 2018.

Souri, A. H., Chance, K., Bak, J., Nowlan, C. R., González Abad, G., Jung, Y., Wong, D. C., Mao, J., and Liu, X.: Unraveling pathways of elevated ozone induced by the 2020 lockdown in Europe by an observationally constrained regional model using TROPOMI, Atmos. Chem. Phys., 21, 18227–18245, https://doi.org/10.5194/acp-21-18227-2021, 2021.

Souri, A. H., Choi, Y., Jeon, W., Li, X., Pan, S., Diao, L., and Westenbarger, D. A.: Constraining NOx emissions using satellite NO_2 measurements during 2013 DISCOVER-AQ Texas campaign, Atmos. Environ., 131, 371–381, https://doi.org/10.1016/j.atmosenv.2016.02.020, 2016a.

Souri, A. H., Choi, Y., Li, X., Kotsakis, A., and Jiang, X.: A 15-year climatology of wind pattern impacts on surface ozone in Houston, Texas, Atmos. Res., 174–175, 124–134, https://doi.org/10.1016/j.atmosres.2016.02.007, 2016b.

Souri, A. H., Duncan, B. N., Strode, S. A., Anderson, D. C., Manyin, M. E., Liu, J., Oman, L. D., Zhang, Z., and Weir, B.: Enhancing Long-Term Trend Simulation of Global Tropospheric OH and Its Drivers from 2005-2019: A Synergistic Integration of Model Simulations and Satellite Observations, EGUsphere, 1–37, https://doi.org/10.5194/egusphere-2024-410, 2024.

Souri, A. H., Johnson, M. S., Wolfe, G. M., Crawford, J. H., Fried, A., Wisthaler, A., Brune, W. H., Blake, D. R., Weinheimer, A. J., Verhoelst, T., Compernolle, S., Pinardi, G., Vigouroux, C., Langerock, B., Choi, S., Lamsal, L., Zhu, L., Sun, S., Cohen, R. C., Min, K.-E., Cho, C., Philip, S., Liu, X., and Chance, K.: Characterization of errors in satellite-based HCHO/NO_2 tropospheric column ratios with respect to chemistry, column-to-PBL translation, spatial representation, and retrieval uncertainties, Atmos. Chem. Phys., 23, 1963–1986, https://doi.org/10.5194/acp-23-1963-2023, 2023.

Souri, A. H., Nowlan, C. R., González Abad, G., Zhu, L., Blake, D. R., Fried, A., Weinheimer, A. J., Wisthaler, A., Woo, J.-H., Zhang, Q., Chan Miller, C. E., Liu, X., and Chance, K.: An inversion of NOx and non-methane volatile organic compound (NMVOC) emissions using satellite observations during the KORUS-AQ campaign and implications for surface ozone over East Asia, Atmos. Chem. Phys., 20, 9837–9854, https://doi.org/10.5194/acp-20-9837-2020, 2020b.

Souri, A. H., Nowlan, C. R., Wolfe, G. M., Lamsal, L. N., Chan Miller, C. E., Abad, G. G., Janz, S. J., Fried, A., Blake, D. R., Weinheimer, A. J., Diskin, G. S., Liu, X., and Chance, K.: Revisiting the effectiveness of HCHO/NO_2 ratios for inferring ozone sensitivity to its precursors using high resolution airborne remote sensing observations in a high ozone episode during the KORUS-AQ campaign, Atmos. Environ., 224, 117341, https://doi.org/10.1016/j.atmosenv.2020.117341, 2020a.

Stanier, C. O., Pierce, R. B., Abdi-Oskouei, M., Adelman, Z. E., Al-Saadi, J., Alwe, H. D., Bertram, T. H., Carmichael, G. R., Christiansen, M. B., Cleary, P. A., Czarnetzki, A. C., Dickens, A. F., Fuoco, M. A., Hughes, D. D., Hupy, J. P., Janz, S. J., Judd, L. M., Kenski, D., Kowalewski, M. G., Long, R. W., Millet, D. B., Novak, G., Roozitalab, B., Shaw, S. L., Stone, E. A., Szykman, J., Valin, L., Vermeuel, M., Wagner, T. J., Whitehill, A. R., and Williams, D. J.: Overview of the Lake Michigan Ozone Study 2017, Bull. Am. Meteorol. Soc., 102, E2207–E2225, https://doi.org/10.1175/BAMS-D-20-0061.1, 2021.

Stavrakou, T., Müller, J.-F., Bauwens, M., De Smedt, I., Lerot, C., Van Roozendael, M., Coheur, P.-F., Clerbaux, C., Boersma, K. F., van der A, R., and Song, Y.: Substantial Underestimation of Post-Harvest Burning Emissions in the North China Plain Revealed by Multi-Species Space Observations, Sci. Rep., 6, 32307, https://doi.org/10.1038/srep32307, 2016.

Strode, S. A., Ziemke, J. R., Oman, L. D., Lamsal, L. N., Olsen, M. A., and Liu, J.: Global changes in the diurnal cycle of surface ozone, Atmos. Environ., 199, 323–333, https://doi.org/10.1016/j.atmosenv.2018.11.028, 2019.

Tao, M., Fiore, A. M., Jin, X., Schiferl, L. D., Commane, R., Judd, L. M., Janz, S., Sullivan, J. T., Miller, P. J., Karambelas, A., Davis, S., Tzortziou, M., Valin, L., Whitehill, A., Civerolo, K., and Tian, Y.: Investigating Changes in Ozone Formation Chemistry during Summertime Pollution Events over the Northeastern United States, Environ. Sci. Technol., 56, 15312–15327, https://doi.org/10.1021/acs.est.2c02972, 2022.

Thompson, A. M., Balashov, N. V., Witte, J. C., Coetzee, J. G. R., Thouret, V., and Posny, F.: Tropospheric ozone increases over the southern Africa region: bellwether for rapid growth in Southern Hemisphere pollution?, Atmos. Chem. Phys., 14, 9855–9869, https://doi.org/10.5194/acp-14-9855-2014, 2014.

Thompson, C. R., Wofsy, S. C., Prather, M. J., Newman, P. A., Hanisco, T. F., Ryerson, T. B., Fahey, D. W., Apel, E. C., Brock, C. A., Brune, W. H., Froyd, K., Katich, J. M., Nicely, J. M., Peischl, J., Ray, E., Veres, P. R., Wang, S., Allen, H. M., Asher, E., Bian, H., Blake, D., Bourgeois, I., Budney, J., Bui, T. P., Butler, A., Campuzano-Jost, P., Chang, C., Chin, M., Commane, R., Correa, G., Crounse, J. D., Daube, B., Dibb, J. E., DiGangi, J. P., Diskin, G. S., Dollner, M., Elkins, J. W., Fiore, A. M., Flynn, C. M., Guo, H., Hall, S. R., Hannun, R. A., Hills, A., Hintsa, E. J., Hodzic, A., Hornbrook, R. S., Huey, L. G., Jimenez, J. L., Keeling, R. F., Kim, M. J., Kupc, A., Lacey, F., Lait, L. R., Lamarque, J.-F., Liu, J., McKain, K., Meinardi, S., Miller, D. O., Montzka, S. A., Moore, F. L., Morgan, E. J., Murphy, D. M., Murray, L. T., Nault, B. A., Neuman, J. A., Nguyen, L., González, Y., Rollins, A., Rosenlof, K., Sargent, M., Schill, G., Schwarz, J. P., Clair, J. M. S., Steenrod, S. D., Stephens, B. B., Strahan, S. E., Strode, S. A., Sweeney, C., Thames, A. B., Ullmann, K., Wagner, N., Weber, R., Weinzierl, B., Wennberg, P. O., Williamson, C. J., Wolfe, G. M., and Zeng, L.: The NASA Atmospheric Tomography (ATom) Mission: Imaging the Chemistry of the Global Atmosphere, Bull. Am. Meteorol. Soc., 103, E761–E790, https://doi.org/10.1175/BAMS-D-20-0315.1, 2022.

Thornton, J. A., Wooldridge, P. J., Cohen, R. C., Martinez, M., Harder, H., Brune, W. H., Williams, E. J., Roberts, J. M., Fehsenfeld, F. C., Hall, S. R., Shetter, R. E., Wert, B. P., and Fried, A.: Ozone production rates as a function of NOx abundances and HOx production rates in the Nashville urban plume, J. Geophys. Res. Atmos., 107, ACH 7-1-ACH 7-17, https://doi.org/10.1029/2001JD000932, 2002.

Tibshirani, R.: Regression Shrinkage and Selection via the Lasso, J. R. Stat. B (Methodological), 58, 267–288, 1996.

Tilstra, L. G., de Graaf, M., Trees, V., Litvinov, P., Dubovik, O., and Stammes, P.: A directional surface reflectance climatology determined from TROPOMI observations, Atmos. Meas. Tech. Discussions, 1–29, https://doi.org/10.5194/amt-2023-222, 2023.

Tonnesen, G. S. and Dennis, R. L.: Analysis of radical propagation efficiency to assess ozone sensitivity to hydrocarbons and NO x : 1. Local indicators of instantaneous odd oxygen production sensitivity, J. Geophys. Res., 105, 9213–9225, https://doi.org/10.1029/1999JD900371, 2000a.

Tonnesen, G. S. and Dennis, R. L.: Analysis of radical propagation efficiency to assess ozone sensitivity to hydrocarbons and NO x : 2. Long-lived species as indicators of ozone concentration sensitivity, J. Geophys. Res., 105, 9227–9241, https://doi.org/10.1029/1999JD900372, 2000b.

van der Velde, I. R., van der Werf, G. R., Houweling, S., Eskes, H. J., Veefkind, J. P., Borsdorff, T., and Aben, I.: Biomass burning combustion efficiency observed from space using measurements of CO and NO2 by the TROPOspheric Monitoring Instrument (TROPOMI), Atmos. Chem. Phys., 21, 597–616, https://doi.org/10.5194/acp-21-597-2021, 2021.

van Geffen, J., Eskes, H., Compernolle, S., Pinardi, G., Verhoelst, T., Lambert, J.-C., Sneep, M., ter Linden, M., Ludewig, A., Boersma, K. F., and Veefkind, J. P.: Sentinel-5P TROPOMI NO2 retrieval: impact of version v2.2 improvements and comparisons with OMI and ground-based data, Atmos. Meas. Tech., 15, 2037–2060, https://doi.org/10.5194/amt-15-2037-2022, 2022.

Veefkind, J. P., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H. J., de Haan, J. F., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., and Levelt, P. F.: TROPOMI on the ESA

Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications, Remote Sens. Environ., 120, 70–83, https://doi.org/10.1016/j.rse.2011.09.027, 2012.

Verhoelst, T., Compernolle, S., Pinardi, G., Lambert, J.-C., Eskes, H. J., Eichmann, K.-U., Fjæraa, A. M., Granville, J., Niemeijer, S., Cede, A., Tiefengraber, M., Hendrick, F., Pazmiño, A., Bais, A., Bazureau, A., Boersma, K. F., Bognar, K., Dehn, A., Donner, S., Elokhov, A., Gebetsberger, M., Goutail, F., Grutter de la Mora, M., Gruzdev, A., Gratsea, M., Hansen, G. H., Irie, H., Jepsen, N., Kanaya, Y., Karagkiozidis, D., Kivi, R., Kreher, K., Levelt, P. F., Liu, C., Müller, M., Navarro Comas, M., Piters, A. J. M., Pommereau, J.-P., Portafaix, T., Prados-Roman, C., Puentedura, O., Querel, R., Remmers, J., Richter, A., Rimmer, J., Rivera Cárdenas, C., Saavedra de Miguel, L., Sinyakov, V. P., Stremme, W., Strong, K., Van Roozendael, M., Veefkind, J. P., Wagner, T., Wittrock, F., Yela González, M., and Zehner, C.: Ground-based validation of the Copernicus Sentinel-5P TROPOMI NO2 measurements with the NDACC ZSL-DOAS, MAX-DOAS and Pandonia global networks, Atmos. Meas. Tech., 14, 481–510, https://doi.org/10.5194/amt-14-481-2021, 2021.

Vigouroux, C., Langerock, B., Bauer Aquino, C. A., Blumenstock, T., Cheng, Z., De Mazière, M., De Smedt, I., Grutter, M., Hannigan, J. W., Jones, N., Kivi, R., Loyola, D., Lutsch, E., Mahieu, E., Makarova, M., Metzger, J.-M., Morino, I., Murata, I., Nagahama, T., Notholt, J., Ortega, I., Palm, M., Pinardi, G., Röhling, A., Smale, D., Stremme, W., Strong, K., Sussmann, R., Té, Y., van Roozendael, M., Wang, P., and Winkler, H.: TROPOMI–Sentinel-5 Precursor formaldehyde validation using an extensive network of ground-based Fourier-transform infrared stations, Atmos. Meas. Tech., 13, 3751–3767, https://doi.org/10.5194/amt-13-3751-2020, 2020.

Wang, T., Xue, L., Brimblecombe, P., Lam, Y. F., Li, L., and Zhang, L.: Ozone pollution in China: A review of concentrations, meteorological influences, chemical precursors, and effects, Sci. Total Environ., 575, 1582–1596, https://doi.org/10.1016/j.scitotenv.2016.10.081, 2017.

Wang, W., Parrish, D. D., Li, X., Shao, M., Liu, Y., Mo, Z., Lu, S., Hu, M., Fang, X., Wu, Y., Zeng, L., and Zhang, Y.: Exploring the drivers of the increased ozone production in Beijing in summertime during 2005–2016, Atmos. Chem. Phys., 20, 15617–15633, https://doi.org/10.5194/acp-20-15617-2020, 2020.

Warneke, C., Trainer, M., de Gouw, J. A., Parrish, D. D., Fahey, D. W., Ravishankara, A. R., Middlebrook, A. M., Brock, C. A., Roberts, J. M., Brown, S. S., Neuman, J. A., Lerner, B. M., Lack, D., Law, D., Hübler, G., Pollack, I., Sjostedt, S., Ryerson, T. B., Gilman, J. B., Liao, J., Holloway, J., Peischl, J., Nowak, J. B., Aikin, K. C., Min, K.-E., Washenfelder, R. A., Graus, M. G., Richardson, M., Markovic, M. Z., Wagner, N. L., Welti, A., Veres, P. R., Edwards, P., Schwarz, J. P., Gordon, T., Dube, W. P., McKeen, S. A., Brioude, J., Ahmadov, R., Bougiatioti, A., Lin, J. J., Nenes, A., Wolfe, G. M., Hanisco, T. F., Lee, B. H., Lopez-Hilfiker, F. D., Thornton, J. A., Keutsch, F. N., Kaiser, J., Mao, J., and Hatch, C. D.: Instrumentation and measurement strategy for the NOAA SENEX aircraft campaign as part of the Southeast Atmosphere Study 2013, Atmos. Meas. Tech., 9, 3063–3093, https://doi.org/10.5194/amt-9-3063-2016, 2016.

Wolfe, G. M., Hanisco, T. F., Arkinson, H. L., Blake, D. R., Wisthaler, A., Mikoviny, T., Ryerson, T. B., Pollack, I., Peischl, J., Wennberg, P. O., Crounse, J. D., St. Clair, J. M., Teng, A., Huey, L. G., Liu, X., Fried, A., Weibring, P., Richter, D., Walega, J., Hall, S. R., Ullmann, K., Jimenez, J. L., Campuzano-Jost, P., Bui, T. P., Diskin, G., Podolske, J. R., Sachse, G., and Cohen, R. C.: Photochemical evolution of the 2013 California Rim Fire: synergistic impacts of reactive hydrocarbons and enhanced oxidants, Atmos. Chem. Phys., 22, 4253–4275, https://doi.org/10.5194/acp-22-4253-2022, 2022.

Wolfe, G. M., Marvin, M. R., Roberts, S. J., Travis, K. R., and Liao, J.: The Framework for 0-D Atmospheric Modeling (F0AM) v3.1, Geosci. Model Dev., 9, 3309–3319, https://doi.org/10.5194/gmd-9-3309-2016, 2016.

Wu, Y., Zhao, K., Ren, X., Dickerson, R. R., Huang, J., Schwab, M. J., Stratton, P. R., Daley, H., Li, D., and Moshary, F.: Ozone pollution episodes and PBL height variation in the NYC urban and coastal areas during LISTOS 2019, Atmos. Environ., 320, 120317, https://doi.org/10.1016/j.atmosenv.2023.120317, 2024.

Xu, W., Zhang, G., Wang, Y., Tong, S., Zhang, W., Ma, Z., Lin, W., Kuang, Y., Yin, L., and Xu, X.: Aerosol Promotes Peroxyacetyl Nitrate Formation During Winter in the North China Plain, Environ. Sci. Technol., 55, 3568–3581, https://doi.org/10.1021/acs.est.0c08157, 2021.

Yousefian, F., Faridi, S., Azimi, F., Aghaei, M., Shamsipour, M., Yaghmaeian, K., and Hassanvand, M. S.: Temporal variations of ambient air pollutants and meteorological influences on their concentrations in Tehran during 2012–2017, Sci. Rep., 10, 292, https://doi.org/10.1038/s41598-019-56578-6, 2020.

Zara, M., Boersma, K. F., Eskes, H., Denier van der Gon, H., Vilà-Guerau de Arellano, J., Krol, M., van der Swaluw, E., Schuch, W., and Velders, G. J. M.: Reductions in nitrogen oxides over the Netherlands between 2005 and 2018 observed from space and on the ground: Decreasing emissions and increasing O3 indicate changing NOx chemistry, Atmos. Environ.: X, 9, 100104, https://doi.org/10.1016/j.aeaoa.2021.100104, 2021.

Zhang, J., Wang, T., Chameides, W. L., Cardelino, C., Kwok, J., Blake, D. R., Ding, A., and So, K. L.: Ozone production and hydrocarbon reactivity in Hong Kong, Southern China, Atmos. Chem. Phys., 7, 557–573, https://doi.org/10.5194/acp-7-557-2007, 2007.