Technical Note: Constraining the hydroxyl (OH) radical in the tropics with satellite observations of its drivers: First steps toward assessing the feasibility of a global observation strategy

Daniel C. Anderson[1,2], Bryan N. Duncan[2], Julie M. Nicely[2,3], Junhua Liu[2,4], Sarah A. Strode[2,4], Melanie B. Follette-Cook[5]

1. GESTAR II, University of Maryland Baltimore County, Baltimore, MD, USA
2. Atmospheric Chemistry and Dynamics Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA
3. Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA
4. GESTAR II, Morgan State University, Baltimore, MD, USA
5. Mesoscale Atmospheric Processes Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA

Correspondence to: Daniel C. Anderson (daniel.c.anderson@nasa.gov)

Abstract

Despite its importance in controlling the abundance of methane (CH_4) and a myriad of other tropospheric species, the hydroxyl radical (OH) is poorly constrained due to its large spatial heterogeneity and the inability to measure tropospheric OH with satellites. Here, we present a methodology to infer tropospheric column OH (TCOH) in the tropics over the open oceans using a combination of a machine learning model, output from a simulation of the GEOS model, and satellite observations. Our overall goals are to assess the feasibility of our methodology, to identify potential limitations, and to suggest areas of improvement in the current observational network. The methodology reproduces the variability of TCOH from independent 3D model output and of observations from the Atmospheric Tomography mission (ATom). While the methodology also reproduces the magnitude of the 3D model validation set, the accuracy of the magnitude when applied to observations is uncertain because current observations are insufficient to fully evaluate the machine learning model. Despite large uncertainties in some of the satellite retrievals necessary to infer OH, particularly for NO_2 and HCHO, current satellite observations are of sufficient quality to apply the machine learning methodology, resulting in an error comparable to that of *in situ* OH observations. Finally, the methodology is not limited to a specific suite of satellite retrievals. Comparison of TCOH determined from two sets of retrievals does show, however, that systematic biases in NO_2, resulting both from retrieval algorithm and instrumental differences, lead to relative biases in the calculated TCOH. Further evaluation of NO_2 retrievals in the remote atmosphere is needed to determine their accuracy. With slight modifications, a similar methodology could likely be expanded to the extra-tropics and over land, with the benefits of increasing our understanding of the atmospheric oxidation capacity and, for instance, informing understanding of recent CH_4 trends.

1 Introduction

The hydroxyl radical (OH) dictates the lifetime of many tropospheric species, including carbon monoxide (CO), methane (CH_4), and numerous volatile organic compounds (VOCs). Knowledge of OH is therefore necessary to understand the abundance, distribution, and variability of these species. For instance, Rigby et al. (2017) and Laughner et al. (2021) attribute recent trends and increases in CH_4 at least partially to changes in OH abundance. Current constraints on OH are insufficient, however, to assess its relative importance in controlling these trends (Turner et al., 2017).

Differences in OH distributions among chemistry transport (CTM) and chemistry climate models (CCM) suggest that these models are insufficient to inform understanding of OH abundance and variability without further observational constraints. OH abundance can differ by up to 80% among models constrained with identical emissions in intercomparison projects (Voulgarakis et al., 2013;Nicely et al.;Zhao et al., 2019;Murray et al., 2021), with modeled trends disagreeing with those derived from observationally constrained methods (Stevenson et al., 2020). Variables such as the photolysis frequency of O_3 (JO^1D) (Nicely et al., 2020), the NO_X lifetime ($NO_X = NO + NO_2$), and the oxidation efficiency of VOCs (Murray et al., 2021) contribute to these inter-model variations in OH. Using Gaussian emulation, Wild et al. (2020) found that the relative importance of drivers of OH variability differed widely among three CTMs. Likewise, the response of OH to the El Niño Southern Oscillation (ENSO), the dominant mode of OH variability on monthly and seasonal timescales (e.g. Anderson et al., 2021;Turner et al., 2018), and other modes of internal climate variability can vary widely among models (Anderson et al., 2021).

Despite this need for better constraints, observations of tropospheric OH are limited. The hydroxyl radical has a lifetime of approximately 1s (Mao et al., 2009), resulting in large spatial heterogeneity in both the horizontal and vertical. This spatial heterogeneity is further caused by the large variation in the relative importance of drivers of OH loss and production in different regions of the atmosphere (e.g. Spivakovsky et al., 1990;Lelieveld et al., 2016). A strategic, representative in situ observational network is therefore unfeasible. As a result, observations of OH are generally limited to intensive field campaigns (Miller and Brune, 2022) that have narrow spatial and temporal coverage. While remotely-sensed OH observations are available, those from satellites are limited to the stratosphere (e.g., Pickett et al., 2008), while ground-based observations of total column OH are dominated by the stratospheric contribution (e.g., Burnett and Minschwaner, 1998).

Reference gases with well-characterized sources and an OH sink, such as methyl chloroform (MCF), can be used to infer OH abundance (Lovelock, 1977). This methodology, however, generally yields no information on spatial heterogeneity beyond the hemispheric scale (e.g., Montzka et al., 2011;Rigby et al., 2017;Naus et al., 2019), although there has been recent success when using three dimensional inversion techniques (Naus et al., 2021). For MCF in particular, recent declines in tropospheric abundance will soon dictate the need for a new reference species (Liang et al., 2017).

Multiple studies have attempted to constrain OH through the creation of proxies and the application of satellite retrievals of OH drivers. Murray et al. (2014) showed that global OH strongly correlated with a combination of JO^1D, water vapor ($H_2O_{(v)}$), and the tropospheric sources of reactive nitrogen and carbon in the GEOS-Chem model. Murray et al. (2021) demonstrated that OH correlated with this proxy in multiple CTMs, although the relationship differs strongly among models. Miyazaki et al. (2020) created a data assimilation framework that ingested satellite observations of CO, NO_2, O_3, and HNO_3 (nitric acid) into multiple CTMs. The data assimilation reduced the spread in average OH among the models and brought the interhemispheric ratio closer to unity, in line with values suggested by MCF observations (e.g. Patra et al., 2014). These results demonstrate that the incorporation of satellite observations into a modeling framework can improve the representation of OH. Wolfe et al. (2019) developed a proxy for OH based on formaldehyde (HCHO) production and loss rates. They applied that proxy to satellite HCHO observations to estimate OH columns in the remote troposphere, a region where HCHO abundance is low and the satellite retrievals are reflective of the *a priori* (Zhu et al., 2016). Using machine learning, chemical transport model output, and retrievals of NO_2 and HCHO, Zhu et al. (2022b) developed a method to estimate surface OH in North American urban areas. Finally, Pimlott et al. (2022) used a steady state approximation of OH, including primary production from H_2O and O_3 and loss from CO, CH_4,

and O_3, to estimate OH between 600 and 700 hPa using observations from IASI (Infrared Atmospheric Sounding Interferometer). A logical next step, building on the results of these studies, is the development of a methodology to constrain OH that ingests multiple satellite retrievals, encompasses the breadth of OH chemical and dynamical drivers, and spans a significant enough portion of the globe to inform variability and trends in CH_4 and CO loss.

Combining machine learning, chemical transport model (CTM) output, and satellite data has the potential to constrain tropospheric column OH (TCOH). A variety of machine learning techniques, such as neural networks (Nicely et al., 2017;Nicely et al., 2020;Kelp et al., 2020), self-organizing maps (Stauffer et al., 2016), random forest regression (Keller and Evans, 2019), and gradient boosted regression trees (GBRTs) (Ivatt and Evans, 2020;Zhu et al., 2022b;Anderson et al., 2022) show promise in helping to solve problems in atmospheric chemistry. In particular, Zhu et al. (2022b) and Anderson et al. (2022) demonstrated the ability of GBRTs to predict OH from a chemical transport model with reasonable accuracy. GBRT models (Elith et al., 2008;Chen and Guestrin, 2016) use an ensemble of decision trees to predict the value of a target based on multiple inputs, even for targets with highly non-linear dependencies on the inputs.

Here, we present a methodology to infer clear sky TCOH in the tropics from space-based observations of its chemical and dynamical drivers with the goal of assessing the feasibility of our methodology, identifying potential limitations, and suggesting areas of improvement in the current observational network. We train a GBRT model using output from a simulation of the NASA GEOS (Goddard Earth Observing System) model, and then estimate TCOH in the actual atmosphere at the satellite overpass time using inputs from a suite of satellite retrievals. In Section 2, we describe the methodology for generating the machine learning model as well as the satellite retrievals used to constrain TCOH. We then evaluate the suitability of MERRA2 GMI as a training dataset (Sect. 3) and, in Section 4, present a satellite-constrained OH product for one month from each season. Finally, in Section 5, we explore potential methodological limitations and benefits, including lack of validation data, the impacts of observational uncertainties, and the ability to use different satellites and retrievals as inputs to the GBRT model.

2 Description of the methodology to generate the GBRT model and of the associated datasets

Our overall aim is to demonstrate the feasibility of our approach to constrain TCOH with satellite-based observations over broad regional scales. As a first step, we restrict our analysis to latitudes equatorward of 25° and regions over water. We chose to focus initially on this domain as it has appreciable OH concentrations and simplified chemistry, as compared to regions with large biogenic and anthropogenic VOC emissions. Nevertheless, this portion of the atmosphere accounts for 50 – 60% of global CO and CH_4 loss. In this section, we describe the creation of the machine learning model used to predict TCOH (Sect. 2.1) for this region as well as the satellite products used as inputs to the machine learning model (Sect. 2.2).

2.1 Creation of the TCOH model

2.1.1 Creation of the GBRT training dataset

For the machine learning model training dataset, we use a subset of output from the MERRA2 GMI simulation (https://acd-ext.gsfc.nasa.gov/Projects/GEOSCCM/MERRA2GMI/). MERRA2 GMI is a 40 year (1980 – 2019) simulation of the NASA GEOS model run in replay mode (Orbe et al., 2017) with MERRA2 (Modern Era Retrospective analysis for Research and Applications, version 2) meteorology (Gelaro et al., 2017). The simulation has a resolution of c180 on the cubed sphere (approximately 0.625° longitude by 0.5° latitude) with 72 vertical layers and uses the Global Modeling Initiative (GMI) chemical mechanism

(Duncan et al., 2007;Strahan et al., 2007). Output is available at daily- and monthly-averaged resolution, as well as instantaneous values at 10:00 and 14:00 LST. These times are within approximately 30 minutes of the overpass times of the satellites described in Section 2.2. Anderson et al. (2021) and Strode et al. (2019) provide detailed information about the simulation, including emissions.

The training target for the machine learning model is TCOH. In Anderson et al. (2022), we developed a GBRT parameterization trained on MERRA2 GMI output to predict in situ OH concentrations using 27 inputs, only a small fraction of which are observable from space. That parameterization, designed to be integrated into the GEOS modeling framework, performed better when there was a separate model for each month as opposed to one model for all months. While that GBRT model is not appropriate for the application described here, we employ a similar approach, creating a separate set of TCOH training targets for each month. We use instantaneous OH output from MERRA2 GMI at 14:00 local time for each day of a given month across the years 2005 to 2019, a timeframe that maximizes overlap between the operational lifetime of the satellites listed in Table 1 and the period of the MERRA2 GMI simulation. We omitted data from 2017 to evaluate model performance. For a given month and year, we calculate daily tropospheric column values across the grid, filtering out columns where the maximum cloud fraction in that column was greater than 30% in order to align the training targets more closely with satellite data, where retrievals of some species are often filtered for cloud cover. This yields approximately 43,000 valid grid boxes per day. For each year, we then average these values to monthly resolution. This results in approximately 600,000 total training targets for each month over the 15-year period.

Table 1: Input variables to the machine learning model and the corresponding satellite retrieval used to create the satellite OH product. Overpass times are ~13:30 LST for all satellites except MOPITT, which has a 10:30 LST overpass.

Variable	Satellite retrieval	Original horizontal and temporal resolution	Reference
Total O_3 column	OMI TOMS-Like L3 version 3	0.25° × 0.25°, daily	McPeters et al. (2015)
Tropospheric NO_2 column	OMI GSFC L3 version 4	0.25° × 0.25°, daily	Lamsal et al. (2021)
CO column	MOPITT L3 version 8	1.0° × 1.0°, monthly	Deeter et al. (2019)
HCHO column	OMI SAO L3 version 3	0.1° × 0.1°, daily	González Abad et al. (2015)
$H_2O_{(v)}$ column	AIRS L3 version 6	1.0° × 1.0°, monthly	Susskind et al. (2014)
Sea surface temperature	MUR L4 version 4.2	0.25° × 0.25°, daily	Chin et al. (2017)
Aerosol optical depth at 550 nm	MODIS Aqua L3 collection 6	0.5° × 0.5°, daily	Levy et al. (2013)
$H_2O_{(v)}$ layers: 925 – 850 hPa, 850 – 700 hPa, 700 – 600 hPa, 600 – 500 hPa, 500 – 400 hPa, 400 – 300 hPa, and 300 - 250 hPa	AIRS L3 version 6	1.0° × 1.0°, monthly	Susskind et al. (2014)
Solar zenith angle	N/A		
Latitude	N/A		

We selected the input variables for the machine learning model (Table 1) based on their relevance to OH chemistry and variability as well as our current ability to observe the variable with satellites. Performance was similar for a model including total column ozone only and for a model also including the tropospheric column. We therefore use total column ozone because of the uncertainties inherent in

separating the column into two parts in the satellite retrieval. We chose the water vapor layers to correspond with the Atmospheric Infrared Sounder (AIRS) layers product. Layers are averages over the indicated pressure range, and we denote the layer names by the highest pressure in that range. We include sea surface temperatures (SST) as a proxy for the Indian Ocean Dipole and ENSO, which has a strong impact on OH variability in the tropics (Anderson et al., 2021;Turner et al., 2018;Naus et al., 2021). In addition, we include latitude and solar zenith angle as previous work has shown that these variables can explain a large fraction of the spatial OH variability (Duncan et al., 2000;Anderson et al., 2022).

We sampled the MERRA2 GMI output to create the training dataset in the same manner as for the TCOH targets. The inputs to the machine learning model each correspond to the same model column as the OH target. All column values are instantaneous and taken from 14:00 to correspond with satellite overpass times, except for CO, which is for 10:00, near the Measurement of Pollution in the Troposphere (MOPITT) overpass time. Model performance was similar when using CO output at 14:00 and 10:00, likely because of limited diurnal variability in CO column in the study region. SSTs are monthly averages of 24-hour averaged values, and we calculated solar zenith angle at the surface for noon on the 15th of a given month.

2.1.2 Creation and tuning of the GBRT model

We used the XGBoost package (Chen and Guestrin, 2016) version 0.81 in Python version 3.6 to create a GBRT model of TCOH for each month using the training datasets from MERRA2 GMI. For each month, we used 90% of the dataset for model training and the remainder for model validation. As mentioned in Section 2.1.1, we also used MERRA2 GMI output from 2017, which was omitted from the training dataset, as further validation.

To maximize parameterization performance while also balancing the potential of overfitting, we tuned hyperparameters, including the learning rate, the maximum tree depth, and the number of trees. We chose hyperparameter values that minimized the parameterization root mean square error (RMSE) of the training dataset. We set the learning rate, which controls the magnitude of change when adding a new tree, to 0.1, while we varied the maximum tree depth and number of trees from 6 to 22 and from 10 to 150, respectively. For both maximum tree depth and number of trees, RMSE initially dropped significantly with increasing value, representing sharp improvement in parameterization performance. RMSE values eventually plateaued, increasing parameterization runtime without noticeably improving performance. A combination of a maximum tree depth of 18 and 100 trees balanced performance with model training and run time.

To determine whether the inputs to the machine learning model improved or hindered performance, we performed a “leave one out” analysis. Using 5-fold cross validation, we retrained the model, individually omitting each of the inputs, to determine the percent difference between the mean RMSE of the 5 folds for the model without a specific input and one including all inputs. Omitting the inputs listed in Table 1 lead to increases in the RMSE, suggesting that each is necessary for improved model performance. As a result of this analysis, we do not use water vapor layers for pressures less than 300 hPa because these decreased model performance.

Finally, we found that it was not necessary to apply satellite averaging kernels and shape factors to the training dataset. Of the satellite retrievals used in this work (discussed in Sect. 2.2 and listed in Table 1), only CO, HCHO, and NO_2 could require convolving the model with the averaging kernel. Shape factors for the OMI NO_2 retrieval are determined from a similar setup of the GEOS model, also employing the

GMI chemical mechanism and MERRA2 meteorology. Applying the satellite shape factors to the simulation discussed here would therefore not result in significant changes in the modeled NO_2 (Anderson et al., 2021). To test whether it is necessary to apply the averaging kernels for CO and HCHO, we created a separate training dataset, where we convolved the daily MERRA2 GMI output with the averaging kernel and a priori from the level 2 data for both species for February 2005 - 2019. All other inputs were kept the same. We then retrained the model with these adjusted CO and HCHO variables. When we applied the satellite data to the model for February 2017, as described in Section 4, the resulting TCOH differed by less than 1% on average from the model that did not include averaging kernel information. This level of uncertainty is significantly smaller than the other uncertainties discussed in Section 5, so we do not include averaging kernels in our analysis.

2.2 Description of satellite products

To create the observationally-constrained OH product, we use multiple satellite retrievals, listed in Table 1 and briefly described here. Each instrument is located onboard a polar orbiting satellite that provides near global coverage daily. For each satellite retrieval, we use the level 3 gridded product, with the exception of SST which is level 4. Where necessary, we regridded the retrieval to a common horizontal grid with a resolution of $1.0° \times 1.0°$ and averaged to the monthly scale.

We use these resolutions because, in the study domain, individual pixel retrievals, particularly of NO_2 and HCHO, are frequently at or below detection limits (González Abad et al., 2015;Lamsal et al., 2021), necessitating averaging to relatively coarse temporal and spatial scales. Missing data due to cloud cover and the Ozone Monitoring Instrument (OMI) row anomaly further increase the need for monthly-scale averaging. While other satellites, such as OMPS (Ozone Mapping and Profiler Suite) and TROPOMI (Tropospheric Monitoring Instrument), provide retrievals with increased signal to noise ratios and more complete data coverage, the satellites used here cover a far longer time period. The $1.0° \times 1.0°$ and monthly resolutions, in combination with the long data record, are sufficient to understand regional trends in TCOH and some aspects of TCOH temporal and spatial variability.

We use retrievals of three species – HCHO, O_3, and NO_2 – from OMI, an ultraviolet-visible spectrometer located onboard the Aura satellite, which has an overpass of approximately 13:30 local solar time (LST). We use the Smithsonian Astrophysical Observatory (SAO) version 3 HCHO retrieval (González Abad et al., 2015). Wolfe et al. (2019) found that this retrieval captured the variability of the HCHO columns in the remote atmosphere observed during the Atmospheric Tomography (ATom) campaign with little bias. For total column O_3, we use the TOMS-like (Total Ozone Mapping Spectrometer) retrieval version 3 (McPeters et al., 2015), which agrees with ground-based and other satellite observations within approximately 1% (Labow et al., 2013). Finally, we use the Goddard Space Flight Center version 4 NO_2 tropospheric column retrieval (Lamsal et al., 2021). While previous studies have thoroughly evaluated this retrieval in more polluted atmospheres (e.g., Lamsal et al., 2014;Choi et al., 2020), evaluation in the remote tropical atmosphere, as defined in this study, is limited.

For water vapor and aerosol optical depth (AOD) at 550 nm, we use retrievals from AIRS and the Moderate Resolution Imaging Spectroradiometer (MODIS) instruments, respectively, both located onboard the Aqua satellite with an overpass of approximately 13:30 LST. We use the total column water vapor standard physical retrieval as well as the 7 water vapor layers listed in Table 1 (Susskind et al., 2014). Multiple studies have evaluated the accuracy of the AIRS $H_2O_{(v)}$ column and layers retrievals in the remote tropical atmosphere, finding bias of 5% or less and high correlation against both remote and *in situ* observations (Bedka et al., 2010;Anderson et al., 2016;Pérez-Ramírez et al., 2019). We use

collection 6 of the dark target MODIS AOD retrieval at 550 nm, which is highly correlated with observations from the AERONET network over the ocean (Levy et al., 2013).

We also use retrievals of CO from MOPITT, which is onboard the Terra satellite with an overpass of 10:30 LST. We use the version 8 retrieval that includes both near and thermal infrared radiances (Deeter et al., 2019). CO retrievals from MOPITT in the remote tropics generally agree with ground-based remotely-sensed observations within 10% (Hedelius et al., 2019;Buchholz et al., 2017).

Finally, we use SSTs from the Multi-scale Ultra-high Resolution (MUR) analysis, which combines nighttime SST observations from multiple satellite platforms, including MODIS, as well as *in situ* observations and agrees with other SST analyses within 0.36° C (Chin et al., 2017).

3 Evaluating the Suitability of the MERRA2 GMI Simulation as a Training Dataset

Before generating the GBRT model to predict TCOH, we first demonstrate that the MERRA2 GMI simulation is suitable to use as a training dataset. Because of the paucity of in situ observations of OH over most of the globe, we necessarily use output from an atmospheric chemistry model to train the machine learning model. The atmospheric chemistry model output must reasonably capture the distribution, magnitude, and ENSO-related variability of OH and the drivers listed in Table 1, as GBRT models are unable to extrapolate beyond the photochemical environments on which they are trained (Anderson et al., 2022).

3.1 Comparison of the Distribution and Magnitude of Simulated OH Drivers to Observations

Simulated OH from MERRA2 GMI agrees with observations over the remote ocean within the instrumental uncertainty. Anderson et al. (2021) compared MERRA2 GMI output to in situ observations from the first two deployments of ATom, finding modest correlation (r^2 values between 0.3 and 0.78 depending on the hemisphere and season) between observations and the model. The average normalized mean bias was on the order of 20%, a slight high bias but within the 2σ observational uncertainty of 35%. Agreement was highest in the remote atmosphere, whereas the largest error was in regions of fresh, continental outflow off the coasts of South America and New Zealand.

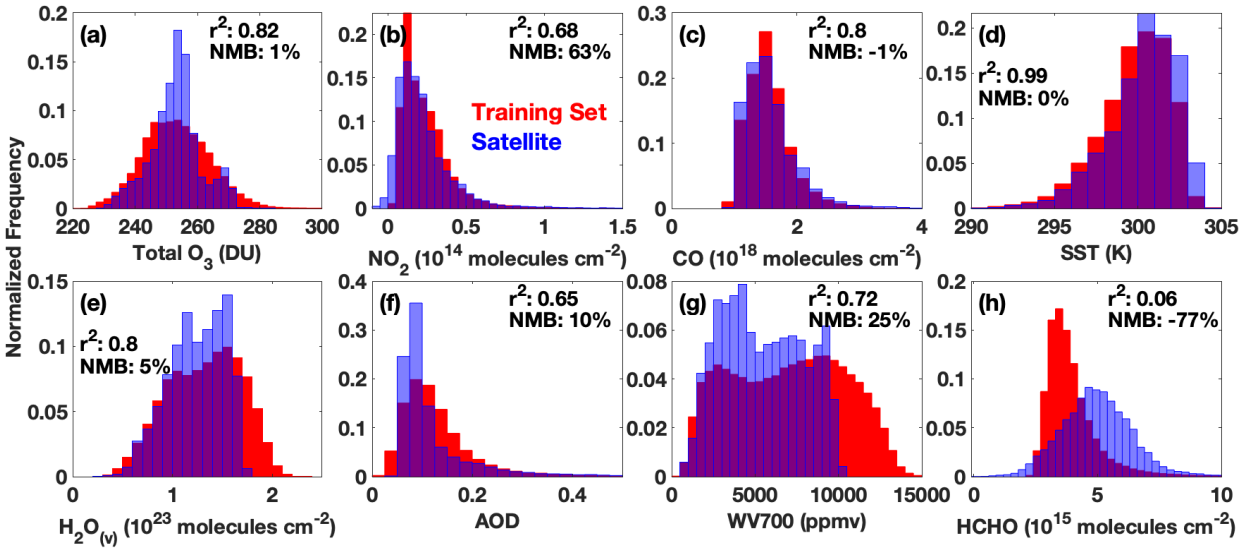

Figure 1: Comparison of the normalized distributions of the training dataset (red) for the February model and satellite observations of the indicated species for February 2017 (blue). Purple indicates regions of overlap. We

use $H_2O_{(v)}$ at 700 hPa as an example for all $H_2O_{(v)}$ layers. Distributions of the other $H_2O_{(v)}$ layers are shown in Figure S1. We also indicate the r^2 of the correlation between MERRA2 GMI output for February 2017 and the corresponding satellite retrieval as well as the normalized mean bias of that output.

The simulation captures both the observed variability and the magnitude of the majority of GBRT model inputs with reasonable fidelity, suggesting that the satellite retrievals highlighted in Section 2.2 are suitable inputs for a machine learning model trained on MERRA2 GMI output (Fig.1). Figure 1 compares the distribution of the February training dataset created from the MERRA2 GMI simulation for 2005 – 2019 to the satellite observations of the indicated species for February 2017, a month omitted from the training dataset. Distributions of the remaining water vapor layers are shown in Figure S1. In addition, correlations between observations and MERRA2 GMI output for February 2017 are shown, as an example, in Figures S2 and S3. With the exception of HCHO, distributions of the species are similar between the observations and MERRA2 GMI, with the training dataset encompassing the full range of almost all species. A GBRT model trained on MERRA2 GMI will therefore likely not have to extrapolate to photochemical environments on which it was not trained when applied to the satellite data. Further, MERRA2 GMI total column O_3, $H_2O_{(v)}$ column, AOD, CO, and SSTs are all highly correlated (r^2 of 0.65 or higher) with their respective satellite observations, and biases are within 10%, on average. Anderson et al. (2021) did show that MERRA2 GMI CO columns demonstrate biases of opposite sign in the Northern and Southern Hemispheres, however.

Agreement between MERRA2 GMI and satellite observations for NO_2, HCHO, and the $H_2O_{(v)}$ layers is more variable than for the other species. While modeled NO_2 is moderately correlated with observations (r^2 = 0.68) with relatively similar distributions, MERRA2 GMI has a NMB of 63%. This disagreement is most pronounced at low column values, however, where observational uncertainty is large. Further, Anderson et al. (2021) demonstrated distinct regions of bias in NO_2 related to biomass burning and lightning emissions. Modeled HCHO, on the other hand, is not correlated with observations and is biased low by -77%. Modeled water vapor layers are all modestly correlated with observations (r^2 of 0.64 or greater) but vary in their bias, with the 925, 850, 700, and 300 hPa layers biased within 30% and the remaining layers biased up to 71%.

The satellite product is insensitive to the differences between the HCHO distribution of the satellite and training dataset highlighted in Figure 1. To determine the effects of the difference in HCHO distribution, we extended the training dataset to cover the full time period of the MERRA2 GMI simulation (1980 – 2019) and then subsampled the resultant data to match the satellite HCHO distribution. Extending the training dataset to 1980 allows for the subsampled training dataset to have a similar size (~600,000 points) as the original training set. We then created a new machine learning model using this sub-sampled dataset and calculated OH fields for Feb. 2017 using the satellite inputs from Table 1. We compared this to the TCOH field calculated from a model using the original training dataset, finding agreement within 5%. Similarly, the satellite-constrained TCOH product discussed in Section 4.2 differs by only 3% on average from one determined with a GBRT model that excludes HCHO as an input, suggesting the limited impact of potential errors in the MERRA2 GMI HCHO distribution on model performance. These uncertainties are small in comparison to that resulting from uncertainties in the NO_2 and HCHO satellite retrievals discussed in Section 5.2. If the uncertainty of the satellite inputs decreases, as retrievals and instruments improve, then it will become necessary to more closely align the training and observed HCHO distributions.

Finally, because NO_2 and HCHO have the largest differences between satellite observations and the training dataset, we trained a separate machine learning model to predict TCOH, omitting these two

species as inputs. When this model was evaluated using the independent MERRA2 GMI output described in Section 4.1, the NRMSE was 10.1%, a more than factor of 2 degradation in performance as compared to the baseline model. This suggests that omitting these species from the machine learning model would result in a greater uncertainty in the final TCOH product than that which results from the retrieval uncertainties and the potential discrepancies between observations and the training dataset.

3.2 Evaluation of the simulated ENSO-related variability of OH drivers

Because ENSO is the dominant mode of OH variability (Anderson et al., 2021;Turner et al., 2018), the training dataset must also capture the ENSO-related variability of the GBRT model inputs. Anderson et al. (2021) demonstrated that the correlation of columns of CO, $H_2O_{(v)}$, and to a lesser extent NO_2, from the MERRA2 GMI simulation with the Multivariate ENSO Index (MEI) (Wolter and Timlin, 2011) agreed closely with correlations of the corresponding species for observations from MOPITT, AIRS, and OMI. Unsurprisingly, based on the strong correlation and low bias of MERRA2 GMI SSTs with observations, the simulation also captures the relationship between SSTs and ENSO. The simulation therefore sufficiently captures the ENSO-related variability of these species to act as training data for the GBRT model. We now evaluate this relationship for the remaining GBRT model inputs.

The MERRA2 GMI-simulated ENSO-related variability of AOD and the various water vapor layers also agrees well with observations. Figures 2 and S4 show the correlation of AOD, HCHO, and the various $H_2O_{(v)}$ layers with the MEI for the satellite retrievals and MERRA2 GMI. MERRA2 GMI captures the general distribution and magnitude of correlation between AOD and ENSO, despite the low optical depths over much of the domain. There are some regional differences, however, particularly in the eastern Southern Hemispheric Pacific. For the $H_2O_{(v)}$ layers, the simulation underestimates the magnitude of the correlation in some areas, but in general, there is excellent agreement for all layers throughout the troposphere. This suggests that, despite the high bias discussed above, including the $H_2O_{(v)}$ layers could provide important, vertically-resolved information to the machine learning model.

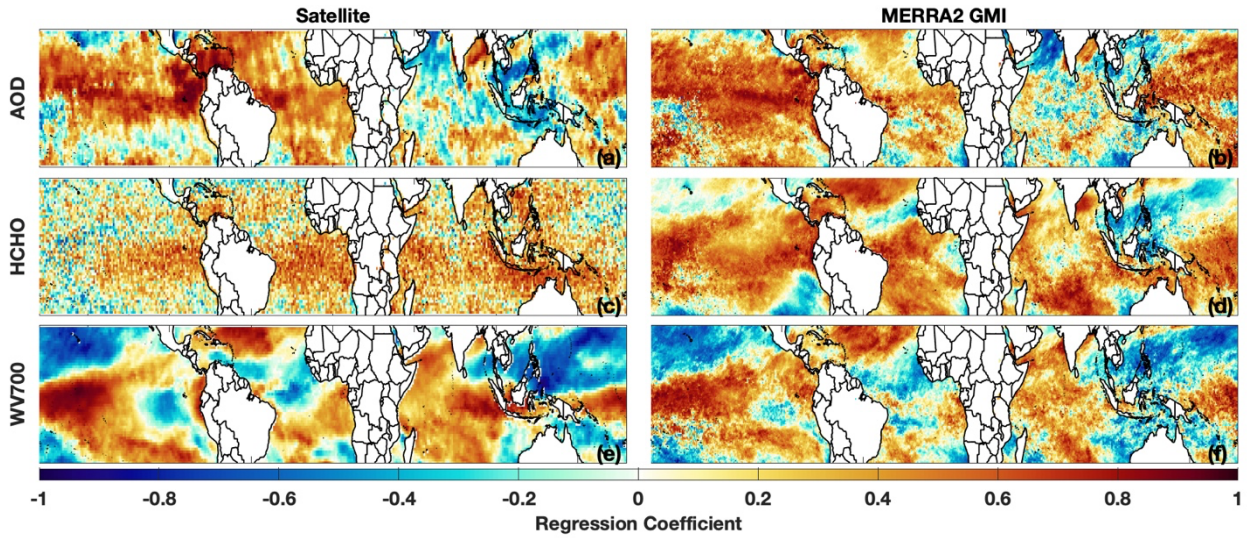

Figure 2: Distribution of the regression coefficient of a linear least squares fit of the indicated variable against the MEI for the respective satellite retrieval (a, c, and e) and MERRA2 GMI (b,d, and f) for February. Regressions of AOD are for 2010 to 2019, the years for which we have a one-degree, gridded satellite product, while HCHO and

water vapor 700 hPa are for 2005 to 2019. Satellite data are on a 1°×1° grid while model output is at the native model resolution.

Modeled accuracy of the HCHO-ENSO relationship is more difficult to assess. While both the OMI retrieval and MERRA2 GMI demonstrate broad regions of anti-correlation between HCHO and ENSO, the correlations with OMI HCHO are weaker and noisier than for the other satellite retrievals. Over much of the domain, HCHO abundance is low, often at or below the retrieval detection limit, suggesting that the HCHO retrieval might not be of sufficient quality to capture ENSO-related variability. We investigate the impacts of the HCHO observational uncertainty in Section 5.

Finally, because we use total column O_3 as an input to the GBRT model, we do not evaluate the relationship between ENSO and O_3, as the stratosphere dominates the O_3 column and the ENSO-related variability is mostly confined to the troposphere. Oman et al. (2013) found that a GEOS CCM simulation and a combination of O_3 retrievals from the Microwave Limb Sounder (MLS) and the Tropospheric Emission Spectrometer (TES) exhibited similar ENSO-related variability in the middle and upper troposphere, demonstrating that simulations in the GEOS framework can capture this relationship. If a TES-like satellite retrieval were currently available, it could be a valuable contributor to the GBRT model described here, as it would provide vertically-resolved information about one of the primary drivers of OH production.

4 Tropical tropospheric column OH constrained with observations of its drivers

We now demonstrate the ability of the GBRT model to determine TCOH. First, we show that the GBRT model can reproduce MERRA2 GMI modeled TCOH from a year independent of the training dataset, a so-called "hold out set" (Sect. 4.1). We then input satellite data from one month from each season into the GBRT model to evaluate the realism of the calculated TCOH fields (Sect 4.2).

4.1 Evaluation with an independent year from MERRA2 GMI

The machine learning model is able to capture both the magnitude and the variability of TCOH across each season when applied to MERRA2 GMI output from 2017, a year independent of the training dataset. For August 2017 (Fig. 3b), the predicted TCOH is highly correlated with MERRA2 GMI (r^2 of 0.98). TCOH from the machine learning model agrees with the CTM simulation within 4.8% on average. The overall normalized mean bias (NMB) is negligible (-0.1%), although there are some regions of coherent bias (Fig. 3a). Results are similar for February, May, and October 2017 (Fig. S5). The normalized root mean square error for each of these months is comparable to that found for a GBRT parameterization of OH created with a similar methodology that included 27 inputs (Anderson et al., 2022). This suggests that limiting inputs to model variables observable from space does not degrade the ability of the machine learning model to predict TCOH. The low bias and high correlation between the GBRT and MERRA2 GMI TCOH for all four months examined here also suggests that any potential overfitting by the GBRT model is minimal.

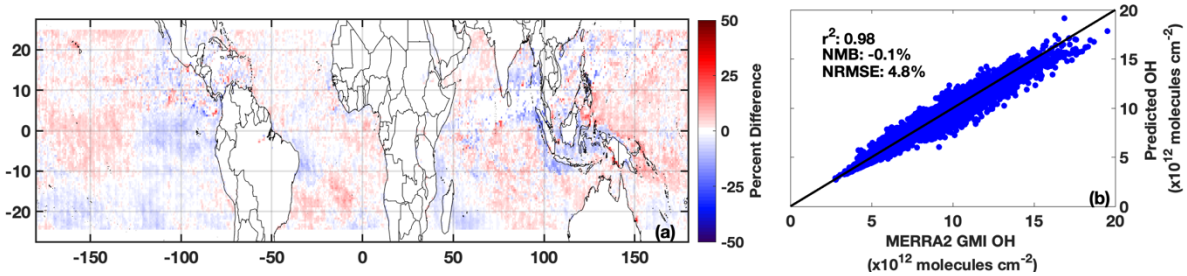

Figure 3: Percent difference between TCOH predicted by the machine learning model and that from MERRA2 GMI for August 2017, a month and year omitted from the training dataset **(a)**. A regression of the machine learning TCOH

against MERRA2 GMI for the same month **(b)**. The r^2 of a linear, least squares regression, along with the normalized mean bias (NMB) and normalized root mean square error (NRMSE), are also indicated.

4.2 TCOH from satellite observations of its drivers

We now apply satellite data from the four months corresponding to the ATom campaign (Aug. 2016, Feb. 2017, Oct. 2017, and May 2018) to the GBRT model to determine TCOH fields across the tropics. More details about ATom as well as evaluation of the GBRT model with ATom observations are in Section 5. We use the satellite observations listed in Table 1, all of which have been averaged to the monthly scale and to a $1° \times 1°$ horizontal resolution. We include only grid boxes with observations for all GBRT model inputs and where those observations are within the range of the corresponding inputs from the training dataset. Because the satellite inputs for most species exclude grid boxes with a cloud fraction greater than approximately 30%, the product presented here represents predominantly clear sky conditions.

The GBRT model and multi-satellite inputs yield TCOH fields that are geophysically credible based on our current understanding of OH photochemistry. Although the domain-wide average changes little with season, with a minimum of 5.84×10^{12} molecules/cm^2 in May 2018 and a maximum of 6.35×10^{12} molecules/cm^2 in August 2016, the spatial distribution varies widely among the four months (Fig. 4). In both Feb. 2017 and Aug. 2016, TCOH minimizes in the winter hemisphere, consistent with lower OH production due to low insolation. The reverse is true for the summer hemisphere. In addition, TCOH maximizes in regions with strong continental outflow and along coastlines, regions likely to be impacted by anthropogenic and biomass burning emissions of OH drivers.

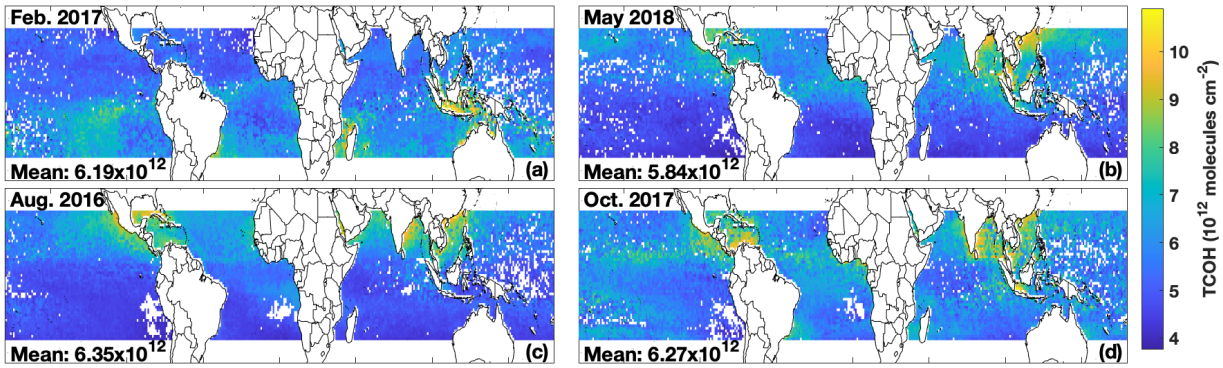

Figure 4: TCOH calculated with the machine learning model using satellite inputs for the months of each ATom deployment: Feb. 2017 (a), May 2018 (b), Aug. 2016 (c), and Oct. 2017 (d). The mean, domain-wide TCOH value in molecules/cm^2 for each month is also indicated.

In general, TCOH from the multi-satellite product differs in both magnitude and distribution from the MERRA2 GMI simulation. For example, for Feb. 2017, mean MERRA2 GMI TCOH is 6.96×10^{12} molecules/cm^2, 12% higher than the satellite product (Fig. S6). This is consistent with the comparison to *in situ* observations discussed in Section 3.1 where MERRA2 GMI overestimates ATom observations by ~20% and underestimates CH_4 lifetime, suggesting that the satellite product is again of reasonable magnitude. While understanding the satellite/model differences in TCOH is beyond the scope of this work, we consider the variety in TCOH spatial distributions generated by the GBRT model to be promising. The difference between the satellite-constrained product and MERRA2 GMI lends some confidence that the GBRT model is not overfit or "tied" to geographic determiners in the training dataset, but rather, is sensitive to variations in the chemical and dynamical drivers of OH. These results all suggest that the methodology presented here can produce a reasonable satellite TCOH product in the

tropics, with values and distributions independent of the chemistry model used to create the GBRT model.

5 Understanding and mitigating potential challenges in using this methodology to constrain TCOH

In this section, we outline possible limitations of the machine learning methodology and the current observational network of the GBRT model inputs and provide potential means to mitigate these limitations where necessary. In section 5.1, we discuss the current lack of sufficient in situ observations to thoroughly evaluate the methodology, highlighting this point by validating the GBRT model with data from the ATom campaign. In section 5.2, we investigate the impacts of random retrieval errors in satellite retrievals on the TCOH product, while in section 5.3, we evaluate the impacts on TCOH when using different satellite retrievals as inputs.

5.1 Insufficient in situ observations for thorough independent evaluation

While we demonstrated in Section 4.1 that TCOH calculated with the GBRT model agrees closely with a hold-out set from MERRA2 GMI, it is also important to demonstrate that the GBRT model can replicate observed TCOH from the actual atmosphere. Because the satellite TCOH product shown in Figure 4 is monthly and at a $1^\circ \times 1^\circ$ resolution, however, there are no observations with which to evaluate the product. We can test the ability of the GBRT model to reproduce observed TCOH from field campaigns, however, assuming there are concomitant observations of the input species listed in Table 1. The additional need for tropospheric column values of many of these species severely limits the datasets available for validation. To our knowledge, the ATom campaign is the only source of the required inputs with enough observations to attempt a limited validation.

During ATom (Thompson et al., 2022), scientists measured a suite of air quality and climate relevant trace gases and aerosols throughout the atmosphere above the remote Pacific and Atlantic. ATom took place in four parts: ATom 1 (July – August 2016), ATom 2 (January – February 2017), ATom 3 (September – October 2017), and ATom 4 (April – May 2018). During each deployment, flights consisted of a series of ascents and descents across all tropical latitudes over the Pacific and Atlantic Oceans. This allows for the calculation of tropospheric column content of the observed species and evaluation of the machine learning model across most latitudes of our study domain and across all seasons.

To evaluate the GBRT model performance, we calculated TCOH using a modified GBRT model and observations from the ATom deployments as inputs. We then compared the values to the observed OH columns. To calculate the column values from the observations, we averaged data into 25 hPa pressure bins for each ATom profile. We filled in missing data using a log-linear interpolation and then integrated the column. Our analysis here includes only profiles with observations of all necessary species, that spanned at least 700 hPa, and where less than 25% of the pressure bin values were interpolated. We also omitted any profiles that had pressure bins with negative OH values. In addition, we restrict our analysis to latitudes within 25° of the equator and profiles conducted between 12:00 and 15:00 LST. Values for total column O_3, AOD, and SSTs, for which there were no observations during ATom, were taken from the MERRA2 GMI simulation from the grid box closest to the center of the respective profile. Because ATom profiles did not span the entire tropospheric column, we trained a separate GBRT model where OH and all tropospheric column input variables were substituted with columns spanning 990 – 250 hPa, the median range of ATom profiles. This allows for a more direct comparison between observed and modeled TCOH. The spatial distribution of the valid ATom columns and the corresponding columns calculated with the GBRT model are shown in Figure S7.

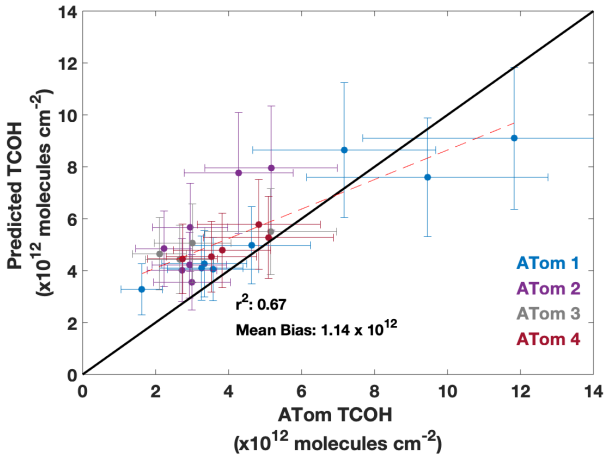

Figure 5: Regression of TCOH observed from the ATom deployments against that predicted from the GBRT model. Error bars represent the 2σ observational uncertainty as reported in Brune et al. (2020) and the GBRT uncertainty described in Section 5.2. The r^2 of a linear least squares fit and the mean bias are also shown.

The GBRT model captures the variability of the observed TCOH, and, while there is a modest overall high bias, the median normalized absolute error of 28.3% is within observational uncertainty. When applied to all ATom deployments, predicted TCOH is correlated with the observations with an r^2 of 0.67 and a mean bias of 1.14×10^{12} molecules/cm^2 (Fig.5). Many of the data points agree within the combined modeled and observational uncertainty. The r^2 values for individual deployments are 0.88 for ATom 1, 0.73 for ATom2, and 0.78 for ATom 3 and 4. The level of agreement between observed and predicted OH is comparable or better than that of other methods to infer OH from space. For example, Pimlott et al. (2022) found an r of 0.78 (r^2 = 0.61) when estimating ATom OH using a steady state approach, with r values ranging from 0.51 to 0.85 (r^2 of 0.26 to 0.72) for the different deployments. The level of agreement we show here therefore demonstrates the validity of the machine learning method to capture the variability of OH.

The source of the model/measurement disagreement, with over- and underprediction at low and high column content respectively, is unclear, although there are multiple potential error sources. For example, a typical profile taken during ATom spanned 300 – 400 km in latitude, disconnecting the top and bottom of the profile in space. This is in contrast to the data used to train the model, which were vertical columns over one location. This could lead to a degradation in model performance when applied to ATom, since the columns are not directly analogous to the training dataset. These effects are likely limited because ATom observations are in the remote atmosphere, where the spatial distribution of relevant species is likely to be more homogeneous than over land.

Further, there is a known interference with the ATom NO_2 observations, suggesting another possible contributor to disagreement between measured and modeled OH. Because of thermal degradation of NO_2 reservoir species, such as organic nitrates and peroxyacetyl nitrate, in the instrument inlet, ATom NO_2 observations are likely biased high (Silvern et al., 2018;Shah et al., 2023;Nault et al., 2015). To test the potential impact of NO_2 on the predicted OH columns, we applied the ATom observations to a model that omits NO_2 as an input. Removing NO_2 increases the r^2 to 0.74, decreases the mean bias to 0.82×10^{12} molecules/cm^2, and decreases the median normalized absolute error slightly to 25.7% (Fig. S8). These improvements in performance suggest that errors in NO_2 could be contributing to the measurement/model differences. Omitting NO_2 does, however, likely introduce additional errors as NO_X compounds are essential to OH production in some regions of the atmosphere. When we apply the hold

out set from MERRA2 GMI to this model, for example, the NRMSE increases by approximately 50%, highlighting the importance of keeping NO_2 as an input variable.

For more certain evaluation of the GBRT model with observations, greater certainty in the in situ NO_2 observations is needed. Although the in situ observations are insufficient to evaluate the absolute accuracy of the product, the results presented here demonstrate that a machine learning model trained on data from a CTM simulation can capture TCOH variability in the actual atmosphere and suggest that predicted OH columns agree with observations within instrumental uncertainty.

5.2 Impacts of uncertainties in the satellite retrievals on TCOH

In the remote atmosphere where HCHO and NO_2 abundances are low, retrieval uncertainty of an individual pixel for both species can be on the order of 100% and is often reflective of the *a priori* (González Abad et al., 2015;Lamsal et al., 2021). Given the importance of these species to the GBRT model as well as to OH chemistry, it is necessary to determine how the propagation of the retrieval uncertainties from these and other model inputs impacts the predicted TCOH.

We determined the total uncertainty in TCOH from all inputs as well as the resultant uncertainty from each individual input for Feb. 2017. First, we estimated an average retrieval uncertainty for each input based on reported values in the retrieval files or from the literature (Table S1). We note that for NO_2 and HCHO we use a fit uncertainty for a single retrieval. Because we are using monthly-averaged data at $1^\circ \times 1^\circ$ horizontal resolution, this likely significantly overestimates the actual uncertainty in these retrievals as the random error from individual pixels will tend to cancel when averaged over such large spatial and temporal scales. Our results are therefore an upper bound on the estimated TCOH uncertainty.

Next, for each grid box and model input, we created a Gaussian distribution of 2000 values with the modeled value for Feb. 2017 as the mean and the estimated uncertainty as the standard deviation. For each input, we then ran the GBRT model 2000 times to create a distribution of predicted TCOH values for each grid box. The normalized uncertainty in TCOH attributable to a given input is the ratio of the standard deviation of the resultant distribution divided by the mean value. We repeated this process individually for all inputs. In addition, to estimate a total uncertainty in TCOH, we varied all inputs simultaneously with the same Gaussian distributions described above.

Uncertainty from the NO_2 retrieval, and to a lesser extent HCHO, dominates the total uncertainty in the TCOH product but is of a magnitude comparable to that of in situ OH observations. Median TCOH 1σ uncertainty resulting from NO_2 is 16.5%, with maxima in the remote atmosphere in regions where NO_2 columns are low. Median uncertainty in TCOH resulting from HCHO is 7%, averaged over the study domain, despite the large uncertainty in the HCHO retrieval itself. In contrast to NO_2, uncertainties in TCOH resulting from HCHO maximize in regions with higher HCHO columns (Fig. 6). The magnitude of that uncertainty is likely an overestimate as the actual retrieval uncertainty for HCHO in these regions is significantly lower than the value assumed for the error analysis. In comparison, median TCOH uncertainties resulting from other inputs are 2.9% or less (Figs. S9 and S10). Total TCOH uncertainty is 16.6% and is dominated by the NO_2 uncertainty. This uncertainty analysis is in general agreement with the model feature importance (Supplementary Fig. 11), a measure of the relative importance of GBRT model inputs, where HCHO and NO_2 consistently have the largest values of the satellite inputs.

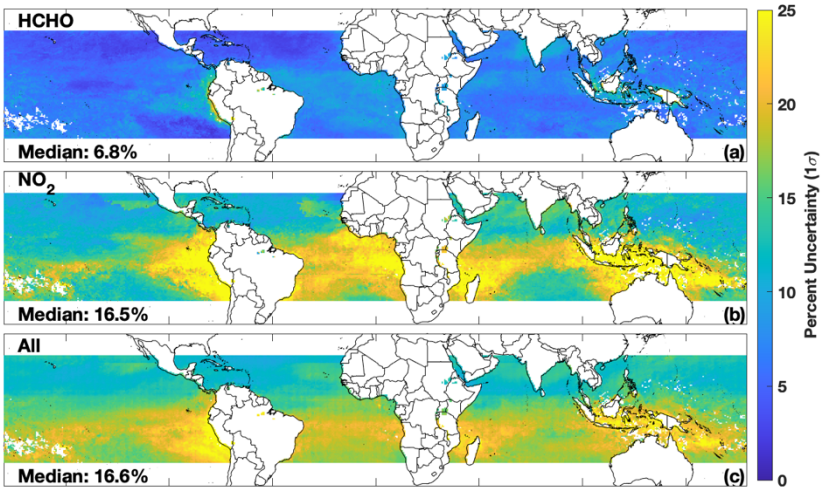

Figure 6: Normalized 1σ uncertainty in the satellite TCOH product due to uncertainties in the HCHO (a) and NO2 (b) retrievals. The combined uncertainty from all input species is shown in panel c.

These results demonstrate that the satellite retrieval inputs to the machine learning model are of sufficient quality to produce a meaningful TCOH data product when averaged over large spatial and temporal scales. The 2σ uncertainty in TCOH resulting from the uncertainties in these retrievals is on the order of that reported for in situ OH observations (Brune et al., 2020). As discussed earlier, this is also likely an upper bound on the uncertainty from random retrieval errors, and uncertainties could be reduced through further averaging, although at the expense of reduced spatial and temporal resolution. Improving the satellite retrievals of NO_2 and HCHO in the remote atmosphere, using retrievals with less noise over the remote atmosphere such as HCHO from OMPS (González Abad et al., 2016), or incorporating data from satellites with higher resolution, such as TROPOMI, could also reduce the uncertainty in their retrievals and thus in TCOH. As discussed in the next section, however, systematic biases between satellite retrievals can also lead to uncertainties in the TCOH.

5.3 Sensitivity of TCOH to different satellite retrievals of GBRT inputs

The satellite retrievals listed in Table 1 provide the benefit of a long record, with data from most retrievals available from at least 2005 to the present. Such a rich dataset would allow for long-term trend analysis of TCOH. These instruments are near the end of their life cycle, however, so it is instructive to see how retrievals from newer satellites impact the predicted TCOH from the GBRT model. In addition, although these newer satellites, such as TROPOMI, have a significantly shorter observational record than those in Table 1, TROPOMI also has finer spatial resolution and the added advantage of providing retrievals for CO, NO_2, O_3, HCHO, and $H_2O_{(v)}$. Using retrievals of multiple species from the same instrument could negate errors resulting from differences in viewing geometry as well as from overpass time. Here, we investigate the effects of applying retrievals from TROPOMI to the machine learning model and compare them to the results from the product described in Section 4, highlighting potential impacts resulting from instrumental differences as well as those resulting from differences in retrieval algorithms. The results emphasize the need for thorough retrieval validation in the remote atmosphere, particularly of NO_2.

5.3.1 Description of TROPOMI and a modified GBRT model

TROPOMI, a successor instrument to OMI, is a spectrometer covering portions of the ultraviolet, visible, and infrared spectrum (Veefkind et al., 2012). It is located onboard the Sentinel 5 Precursor satellite, which is polar orbiting and has a local overpass time of approximately 13:30. Horizontal resolution for

the month examined here (May 2018) is as high as 7 km × 3.5 km at nadir. We have gridded the Level 2 product for each species to a 1° × 1° resolution and averaged the data to the monthly scale, applying the recommended quality flags and filtering for cloud fraction greater than 30%.

We use two different retrievals of TROPOMI NO_2 for this analysis. First, we use the KNMI (Royal Netherlands Meteorological Institute) NO_2 retrieval (van Geffen et al., 2020), which is based on the DOMINO (Dutch OMI NO_2 product) retrieval developed for the OMI instrument. Wang et al. (2020) found that this retrieval was biased high when compared to ship-based observations from a MAX-DOAS instrument over the remote oceans, while Verhoelst et al. (2021) found good agreement between the retrieval and ground-based observations in Reunion. In addition, we use the MINDS (Multi-Decadal Nitrogen Dioxide and Derived Products from Satellites) retrieval, which uses the same algorithm as for the OMI product described in Section 2 (Lamsal et al., 2022). This retrieval has not been evaluated in the remote tropics.

We also use TROPOMI retrievals of HCHO, $H_2O_{(v)}$ column, total column O_3, and CO. The HCHO retrieval (De Smedt et al., 2018) was found to have a 30% low bias with respect to an OMI retrieval using the same algorithm due to differences in cloud processing (De Smedt et al., 2021). While evaluation in the remote tropics is limited, the TROPOMI retrieval does overestimate HCHO in polluted regions (De Smedt et al., 2021) when compared to ground-based observations. The TROPOMI $H_2O_{(v)}$ (Chan et al., 2022) retrieval has a slight dry bias with comparison to other satellite products, while the total column O_3 retrieval (Garane et al., 2019) agrees within 0 – 1.5% with ground-based observations. Finally, the CO retrieval (Borsdorff et al., 2019) agrees with MOPITT over the oceans within 3% on average (Martínez-Alonso et al., 2020). TROPOMI does not have an equivalent retrieval of the AIRS $H_2O_{(v)}$ layers.

To calculate TCOH using TROPOMI data, we trained a separate machine learning model using all inputs from Table 1 except the water vapor layers, for which there are no TROPOMI retrievals. Removal of the layers from the machine learning model does not significantly degrade performance. For example, for May 2017, removing the $H_2O_{(v)}$ layers from the model, increases the NRMSE from 5.34% to 5.73% when applying the GBRT model to the hold out set. For this new model, we then calculate TCOH using TROPOMI data, including the KNMI NO_2 retrieval. For SSTs and AOD, we use the MUR and MODIS products respectively. While TROPOMI does have an aerosol product, the UV aerosol index, the corresponding output from the MERRA2 GMI simulation is unavailable. We refer to this TCOH as the TROPOMI-KNMI product. We have also calculated TCOH using the satellite retrievals in Table 1, except for the water vapor layers, using this GBRT model, and refer to that as the OMI/MOPITT/AIRS product. We restrict our analysis to May 2018, the only month for which we have TROPOMI water vapor data.

5.3.2 TROPOMI data applied to the GBRT model

TCOH from the TROPOMI-KNMI product is higher than that from the OMI/MOPITT/AIRS product for May 2018. Figure 7 shows TCOH calculated from the TROPOMI-KNMI product as well as the percent difference between the two products. While there is modest correlation between the two (r^2 = 0.63), the TROPOMI product is 27.6% higher than the OMI/MOPITT/AIRS product, with higher values across almost the entire domain. Differences between the products are most pronounced in the Indian Ocean and off the coasts of Indonesia and the Philippines.

In general, observations from TROPOMI agree with those from the satellites in Table 1, with the exception of NO_2 and HCHO. Ozone, $H_2O_{(v)}$, and CO from TROPOMI are highly correlated (r^2 of 0.85 or higher) and agree within 10% on average (Fig. S12) with their respective retrievals from OMI, MOPITT, and AIRS. On the other hand, TROPOMI KNMI-NO_2 is systematically higher (145% on average), and

TROPOMI HCHO is 20% lower than their corresponding OMI retrievals. The higher TCOH from the TROPOMI product is consistent with the increase in NO_2, which would lead to higher secondary production of OH. Further, while TROPOMI KNMI-NO_2 is modestly correlated with OMI NO_2 (r^2 = 0.61), TROPOMI and OMI HCHO are not correlated (r^2 = 0.23), highlighting the difficulty of the HCHO retrieval. Note that we are not seeking to determine which retrieval, if any, is more accurate. We are highlighting the differences to emphasize the impact that systematic differences in retrieval magnitudes of GBRT model inputs can have on the resultant TCOH.

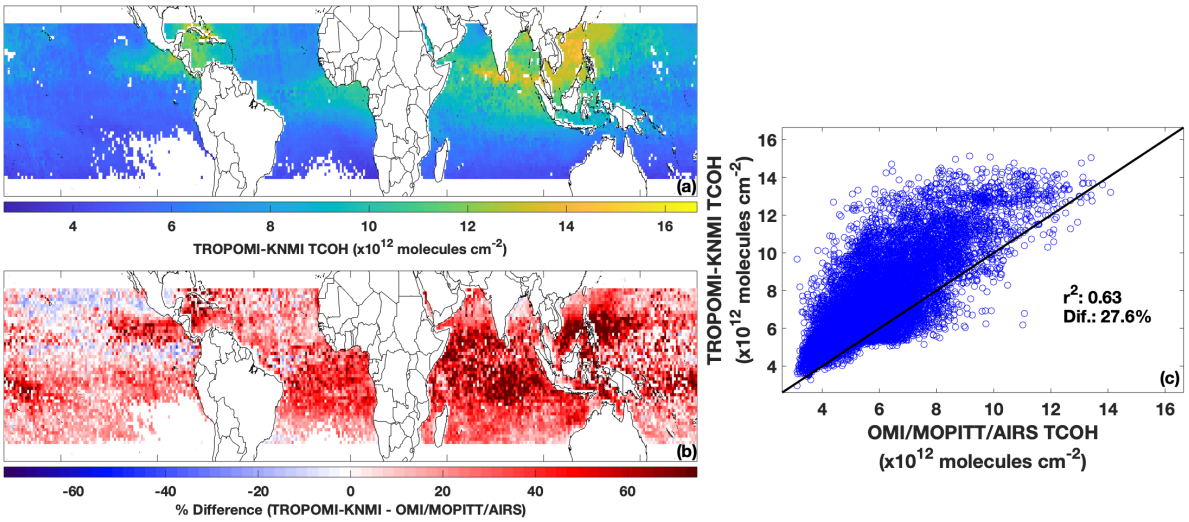

Figure 7: TCOH for May 2018 determined using TROPOMI inputs, including the KNMI NO_2 retrieval (a). The difference between the TROPOMI and multi-satellite product is shown in (b). Panel (c) shows the regression of TCOH calculated from TROPOMI against that calculated from retrievals from MOPITT, OMI, and AIRS as well as the percent difference between the two TCOH products.

NO_2 drives the differences between the two TCOH products. To determine the impacts of the different TROPOMI inputs on the TCOH product, we individually swapped each TROPOMI input into the OMI/MOPITT/AIRS product, replacing the corresponding input from Table 1. We then determined the difference in TCOH from the OMI/MOPITT/AIRS product that does not include TROPOMI. While this method will not yield the exact contribution from a particular retrieval because of the non-linear nature of OH chemistry, it does yield information about the relative importance of each species. Swapping in TROPOMI CO, $H_2O_{(v)}$, and O_3 changed TCOH by less than 2%, while using TROPOMI HCHO increased TCOH by 3%. In contrast, TROPOMI NO_2 increased TCOH by 29%, showing that the higher TCOH in the TROPOMI product is driven by differences in NO_2.

The increased TCOH in the TROPOMI product likely results from a combination of differences in the NO_2 retrieval algorithm as well as instrumental differences. Comparison of the KNMI and MINDS retrievals illustrate this point. When compared to OMI, the MINDS NO_2 retrieval is 58% higher for May 2018, as compared to 145% higher for the KNMI retrieval. The closer agreement is unsurprising since the MINDS NO_2 uses the same retrieval algorithm as for OMI. Substituting the MINDS NO_2 as an input to the TROPOMI product (TROPOMI-MINDS product) reduces the difference with respect to the OMI/MOPITT/AIRS product to 18% (Fig. S13). While this is an improvement in agreement, the differences in TCOH as well as the lack of change in r^2 value still suggest that differences between OMI and TROPOMI unrelated to the retrieval algorithm account for some of the discrepancy. In addition, the training dataset does not take TROPOMI averaging kernels and shape factors into account, which could also contribute to the observed differences.

The results here demonstrate the sensitivity of the methodology to any systematic bias in the input retrievals. As with the random error analysis, the level of uncertainty introduced by these biases is low enough to allow for a meaningful OH product. Despite these differences, the methodology to determine TCOH using machine learning that we have presented here still captures the variability in TCOH, consistent with the ATom evaluation outlined in Section 5.1. To reduce the uncertainty of TCOH, better evaluation of NO_2 in the remote atmosphere is needed to determine which retrievals, if any, are accurate.

6 Discussion and recommendations for future observations

The method of estimating clear-sky TCOH presented here has the potential to increase our understanding of the atmospheric oxidation capacity. Because of the long record of observations from MOPITT, OMI, AIRS, and MODIS, we can calculate tropical TCOH from 2005 to the present, and since the methodology is not constrained to a particular satellite, newer satellite missions could extend the dataset beyond the end of these instruments' lifetimes. In addition, this methodology will provide sub-hemispheric information on OH variability, supplementing information available from MCF inversions.

The methodology could be expanded to the extra-tropics and over land, allowing for global constraints on OH. Expansion over land will likely require additional satellite retrievals, like that of isoprene (Wells et al., 2020), in regions with more complex VOC chemistry than in the remote atmosphere. Expanding this product beyond the tropics could increase understanding of global CH_4, CO, and VOC trends and variability and allow for a wider range of satellite retrievals as inputs. For example, current and upcoming geostationary air quality satellites such as Sentinel 4, TEMPO (Tropospheric Emissions: Monitoring of Pollution), and GEMS (Geostationary Environment Monitoring Spectrometer) could provide retrievals of most of the necessary inputs to the machine learning model, allowing for the understanding of diurnal variability in TCOH and potentially in the diurnal variability of ozone production (Zhu et al., 2022a).

A similar methodology could likely be used to determine OH at different layers of the atmosphere. Because CH_4 loss is not evenly distributed throughout the tropospheric column, vertically resolved OH would better help inform this process. Vertically-resolved OH could also help understand differences in OH drivers in the upper and lower troposphere (Spivakovsky et al., 1990;Lelieveld et al., 2016), which can often be decoupled from the column. While column inputs, such as those discussed here, could be used, the inclusion of vertically resolved satellite retrievals, such as the AIRS $H_2O_{(v)}$ layers, would provide additional information. Tropospheric O_3 at different atmospheric layers, such as that previously provided by the TES satellite, could also be invaluable here, as O_3 is a large driver of primary OH production.

Satellite-derived OH would also a provide a much-needed, observational constraint on OH variability in global chemistry models. Because the methodology can capture variability in TCOH of both observations and 3-dimensional model output, TCOH trends from a satellite-constrained product could be used to evaluate modeled trends and as well as the spatial variability resulting from events like ENSO. While the satellite-derived OH could not explicitly indicate the cause of differences, the spatial distribution of the differences as well as differences in observed and modeled machine learning model inputs could indicate potential dynamical or emission sources of error in the 3D model.

Further, the combination of the satellite-derived OH and the machine learning model could help identify the impacts of any diagnosed errors in emissions inventories as well as the impacts of unexpected

events, such as COVID-19-related shutdowns, on TCOH. For example, if there are significant discrepancies between observed and modeled NO_2 in a specific region of the atmosphere, the satellite NO_2 could be scaled to more closely match the 3D model values and then be input into the machine learning model. The difference in TCOH would then indicate the relative impact of the model error. This would serve as a computationally efficient complement to other methodologies constraining models with observations (e.g. Miyazaki et al., 2020;Miyazaki et al., 2021) to identify the impacts of these errors on the atmospheric oxidation capacity. A similar methodology could be used for unexpected events that significantly impact emissions of OH drivers, allowing for quick determination of their potential impacts on the atmospheric oxidation capacity before emissions inventories could be revised.

While we have shown that the methodology captures the variability of observed OH and generally agrees with observations within measurement uncertainty, it is unclear whether differences result from GBRT model deficiencies or structural differences between the *in situ* observations and the training dataset. Additional field campaigns with observations of OH and the GBRT model inputs would allow for a more thorough evaluation of both the OH product and the methodology itself. Such a field campaign would need to provide complete tropospheric columns of all species and cover less horizontal distance than the ATom profiles (e.g. from spiral flight patterns). In situ observations of NO_2 without significant interference from NO_X reservoir species are also needed to reduce uncertainty. Alternatively, NO_2 and other species could be measured through aircraft-based remote sensing. Finally, repeated sampling over the same locations for multiple days within a defined area would allow for meaningful statistical analysis while also allowing for the comparison of TCOH columns calculated from satellite observations.

Finally, accuracy of the TCOH product is dependent on the accuracy of the satellite retrievals input into the machine learning model, with the NO_2 retrieval having the largest effect. To reduce the uncertainty of the TCOH product, more information about the accuracy of individual NO_2 retrievals is required. Currently, there is little validation of OMI and TROPOMI NO_2 retrievals in the remote, tropical atmosphere, so it is difficult to assess which retrievals, if any, are correct. Recent efforts, such as the QA4ECV (Quality Assurance for the Essential Climate Variables), to improve NO_2 retrieval algorithms have reduced uncertainty, particularly over land (Boersma et al., 2018), although it is unclear how the accuracy of these retrievals translates to the remote tropics as validation data are still extremely limited. Even retrievals of TROPOMI and OMI made with the same algorithm show differences, suggesting that instrumental differences could also affect the results. Future satellite missions should focus on trying to reduce the uncertainty in NO_2 retrievals, particularly in the remote atmosphere, both through improvements in instrument design and algorithm development.

7 Data Availability

Output from the MERRA2 GMI simulation are publicly available at https://acd-ext.gsfc.nasa.gov/Projects/GEOSCCM/MERRA2GMI/ (NASA Goddard Space Flight Center, 2023). All satellite products, except for TROPOMI water vapor, are available at https://disc.gsfc.nasa.gov (GES DISC, 2023). Data from the ATom campaign are located at https://daac.ornl.gov (Wofsy et al., 2021).

8 Author contributions

DCA wrote the manuscript, performed the data analysis, and created the GBRT model. DCA, BND, JMN, and MBFC developed the idea for the methodology. SAS performed three-dimensional modeling for the work. JMN provided advice on machine learning. JL helped perform data analysis. All authors helped develop ideas for the analysis and contributed to the manuscript.

9 Competing Interests

BND is a member of the editorial board of Atmospheric Chemistry and Physics. The peer-review process was guided by an independent editor, and the authors also have no other competing interests to declare.

10 Financial support

This research has been supported by the National Aeronautics and Space Administration (NASA) Atmospheric Composition Campaign Data Analysis and Modeling (ACCDAM) program (grant no. 80NSSC21K1440).

11 Acknowledgements

The authors wish to thank Lok Chan and Diego Loyola for use of the TROPOMI water vapor product.

10. References

Anderson, D. C., Nicely, J. M., Salawitch, R. J., Canty, T. P., Dickerson, R. R., Hanisco, T. F., Wolfe, G. M., Apel, E. C., Atlas, E., Bannan, T., Bauguitte, S., Blake, N. J., Bresch, J. F., Campos, T. L., Carpenter, L. J., Cohen, M. D., Evans, M., Fernandez, R. P., Kahn, B. H., Kinnison, D. E., Hall, S. R., Harris, N. R., Hornbrook, R. S., Lamarque, J. F., Le Breton, M., Lee, J. D., Percival, C., Pfister, L., Pierce, R. B., Riemer, D. D., Saiz-Lopez, A., Stunder, B. J., Thompson, A. M., Ullmann, K., Vaughan, A., and Weinheimer, A. J.: A pervasive role for biomass burning in tropical high ozone/low water structures, Nat Commun, 7, 10267, 10.1038/ncomms10267, 2016.

Anderson, D. C., Duncan, B. N., Fiore, A. M., Baublitz, C. B., Follette-Cook, M. B., Nicely, J. M., and Wolfe, G. M.: Spatial and temporal variability in the hydroxyl (OH) radical: understanding the role of large-scale climate features and their influence on OH through its dynamical and photochemical drivers, Atmospheric Chemistry and Physics, 21, 6481-6508, 10.5194/acp-21-6481-2021, 2021.

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, 10.5194/gmd-15-6341-2022, 2022.

Bedka, S., Knuteson, R., Revercomb, H., Tobin, D., and Turner, D.: An assessment of the absolute accuracy of the Atmospheric Infrared Sounder v5 precipitable water vapor product at tropical, midlatitude, and arctic ground-truth sites: September 2002 through August 2008, Journal of Geophysical Research: Atmospheres, 115, D17310, https://doi.org/10.1029/2009JD013139, 2010.

Boersma, K. F., Eskes, H. J., Richter, A., De Smedt, I., Lorente, A., Beirle, S., van Geffen, J. H. G. M., Zara, M., Peters, E., Van Roozendael, M., Wagner, T., Maasakkers, J. D., van der A, R. J., Nightingale, J., De Rudder, A., Irie, H., Pinardi, G., Lambert, J. C., and Compernolle, S. C.: Improving algorithms and uncertainty estimates for satellite NO2 retrievals: results from the quality assurance for the essential climate variables (QA4ECV) project, Atmos. Meas. Tech., 11, 6651-6678, 10.5194/amt-11-6651-2018, 2018.

Borsdorff, T., aan de Brugh, J., Schneider, A., Lorente, A., Birk, M., Wagner, G., Kivi, R., Hase, F., Feist, D. G., Sussmann, R., Rettinger, M., Wunch, D., Warneke, T., and Landgraf, J.: Improving the TROPOMI CO data product: update of the spectroscopic database and destriping of single orbits, Atmos. Meas. Tech., 12, 5443-5455, 10.5194/amt-12-5443-2019, 2019.

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), Journal of Geophysical Research: Atmospheres, 125, e1019JD031685, 10.1029/2019jd031685, 2020.

Buchholz, R. R., Deeter, M. N., Worden, H. M., Gille, J., Edwards, D. P., Hannigan, J. W., Jones, N. B., Paton-Walsh, C., Griffith, D. W. T., Smale, D., Robinson, J., Strong, K., Conway, S., Sussmann, R., Hase, F., Blumenstock, T., Mahieu, E., and Langerock, B.: Validation of MOPITT carbon monoxide using ground-based Fourier transform infrared spectrometer data from NDACC, Atmos. Meas. Tech., 10, 1927-1956, 10.5194/amt-10-1927-2017, 2017.

Burnett, C. R., and Minschwaner, K.: Continuing development in the regime of decreased atmospheric column OH at Fritz Peak, Colorado, Geophysical Research Letters, 25, 1313-1316, https://doi.org/10.1029/98GL01062, 1998.

Chan, K. L., Xu, J., Slijkhuis, S., Valks, P., and Loyola, D.: TROPOspheric Monitoring Instrument observations of total column water vapour: Algorithm and validation, Science of The Total Environment, 821, 153232, https://doi.org/10.1016/j.scitotenv.2022.153232, 2022.

Chen, T., and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13 - 17 Aug. 2016, 785-794, San Francsisco, CA, USA, https://doi.org/10.1145/2939672.2939785, 2016.

Chin, T. M., Vazquez-Cuervo, J., and Armstrong, E. M.: A multi-scale high-resolution analysis of global sea surface temperature, Remote Sensing of Environment, 200, 154-169, https://doi.org/10.1016/j.rse.2017.07.029, 2017.

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, 10.5194/amt-13-2523-2020, 2020.

De Smedt, I., Theys, N., Yu, H., Danckaert, T., Lerot, C., Compernolle, S., Van Roozendael, M., Richter, A., Hilboll, A., Peters, E., Pedergnana, M., Loyola, D., Beirle, S., Wagner, T., Eskes, H., van Geffen, J., Boersma, K. F., and Veefkind, P.: Algorithm theoretical baseline for formaldehyde retrievals from S5P TROPOMI and from the QA4ECV project, Atmospheric Measurement Techniques, 11, 2395-2426, 10.5194/amt-11-2395-2018, 2018.

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, 10.5194/acp-21-12561-2021, 2021.

879 Deeter, M. N., Edwards, D. P., Francis, G. L., Gille, J. C., Mao, D., Martínez-Alonso, S., Worden, H. M.,
880 Ziskin, D., and Andreae, M. O.: Radiance-based retrieval bias mitigation for the MOPITT instrument: the
881 version 8 product, Atmospheric Measurement Techniques, 12, 4561-4580, 10.5194/amt-12-4561-2019,
882 2019.

883 Duncan, B., Portman, D., Bey, I., and Spivakovsky, C.: Parameterization of OH for efficient computation in
884 chemical tracer models, Journal of Geophysical Research: Atmospheres, 105, 12259-12262,
885 10.1029/1999JD901141, 2000.

886 Duncan, B. N., Strahan, S. E., Yoshida, Y., Steenrod, S. D., and Livesey, N.: Model study of the cross-
887 tropopause transport of biomass burning pollution, Atmos. Chem. Phys., 7, 3713-3736, 10.5194/acp-7-
888 3713-2007, 2007.

889 Elith, J., Leathwick, J. R., and Hastie, T.: A working guide to boosted regression trees, J Anim Ecol, 77,
890 802-813, 10.1111/j.1365-2656.2008.01390.x, 2008.

891 Garane, K., Koukouli, M. E., Verhoelst, T., Lerot, C., Heue, K. P., Fioletov, V., Balis, D., Bais, A., Bazureau,
892 A., Dehn, A., Goutail, F., Granville, J., Griffin, D., Hubert, D., Keppens, A., Lambert, J. C., Loyola, D.,
893 McLinden, C., Pazmino, A., Pommereau, J. P., Redondas, A., Romahn, F., Valks, P., Van Roozendael, M.,
894 Xu, J., Zehner, C., Zerefos, C., and Zimmer, W.: TROPOMI/S5P total ozone column data: global ground-
895 based validation and consistency with other satellite missions, Atmos. Meas. Tech., 12, 5263-5287,
896 10.5194/amt-12-5263-2019, 2019.

897 Gelaro, R., McCarty, W., Suarez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C., Darmenov, A.,
898 Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty,
899 A., da Silva, A., Gu, W., Kim, G. K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G.,
900 Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era
901 Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), J Clim, Volume 30, 5419-
902 5454, 10.1175/JCLI-D-16-0758.1, 2017.

903 GES DISC: Earth Science Data at NASA, available at https://disc.gsfc.nasa.gov, last access: 6 Mar. 2023.

904 González Abad, G., Liu, X., Chance, K., Wang, H., Kurosu, T. P., and Suleiman, R.: Updated Smithsonian
905 Astrophysical Observatory Ozone Monitoring Instrument (SAO OMI) formaldehyde retrieval,
906 Atmospheric Measurement Techniques, 8, 19-32, 10.5194/amt-8-19-2015, 2015.

907 González Abad, G., Vasilkov, A., Seftor, C., Liu, X., and Chance, K.: Smithsonian Astrophysical Observatory
908 Ozone Mapping and Profiler Suite (SAO OMPS) formaldehyde retrieval, Atmospheric Measurement
909 Techniques, 9, 2797-2812, 10.5194/amt-9-2797-2016, 2016.

910 Hedelius, J. K., He, T. L., Jones, D. B. A., Baier, B. C., Buchholz, R. R., De Mazière, M., Deutscher, N. M.,
911 Dubey, M. K., Feist, D. G., Griffith, D. W. T., Hase, F., Iraci, L. T., Jeseck, P., Kiel, M., Kivi, R., Liu, C.,
912 Morino, I., Notholt, J., Oh, Y. S., Ohyama, H., Pollard, D. F., Rettinger, M., Roche, S., Roehl, C. M.,
913 Schneider, M., Shiomi, K., Strong, K., Sussmann, R., Sweeney, C., Té, Y., Uchino, O., Velazco, V. A., Wang,
914 W., Warneke, T., Wennberg, P. O., Worden, H. M., and Wunch, D.: Evaluation of MOPITT Version 7 joint
915 TIR–NIR XCO retrievals with TCCON, Atmos. Meas. Tech., 12, 5547-5572, 10.5194/amt-12-5547-2019,
916 2019.

Ivatt, P. D., and Evans, M. J.: Improving the prediction of an atmospheric chemistry transport model using gradient-boosted regression trees, Atmospheric Chemistry and Physics, 20, 8063-8082, 10.5194/acp-20-8063-2020, 2020.

Keller, C. A., and Evans, M. J.: Application of random forest regression to the calculation of gas-phase chemistry within the GEOS-Chem chemistry model v10, Geoscientific Model Development, 12, 1209-1225, 10.5194/gmd-12-1209-2019, 2019.

Kelp, M. M., Jacob, D. J., Kutz, J. N., Marshall, J. D., and Tessum, C. W.: Toward Stable, General Machine-Learned Models of the Atmospheric Chemical System, Journal of Geophysical Research: Atmospheres, 125, e2020JD032759, https://doi.org/10.1029/2020JD032759, 2020.

Labow, G. J., McPeters, R. D., Bhartia, P. K., and Kramarova, N.: A comparison of 40 years of SBUV measurements of column ozone with data from the Dobson/Brewer network, Journal of Geophysical Research: Atmospheres, 118, 7370-7378, https://doi.org/10.1002/jgrd.50503, 2013.

Lamsal, L. N., Krotkov, N. A., Celarier, E. A., Swartz, W. H., Pickering, K. E., Bucsela, E. J., Gleason, J. F., Martin, R. V., Philip, S., Irie, H., Cede, A., Herman, J., Weinheimer, A., Szykman, J. J., and Knepp, T. N.: Evaluation of OMI operational standard NO2 column retrievals using in situ and surface-based NO2 observations, Atmospheric Chemistry and Physics, 14, 11587-11609, 10.5194/acp-14-11587-2014, 2014.

Lamsal, L. N., Krotkov, N. A., Vasilkov, A., Marchenko, S., Qin, W., Yang, E. S., Fasnacht, Z., Joiner, J., Choi, S., Haffner, D., Swartz, W. H., Fisher, B., and Bucsela, E.: Ozone Monitoring Instrument (OMI) Aura nitrogen dioxide standard product version 4.0 with improved surface and cloud treatments, Atmos. Meas. Tech., 14, 455-479, 10.5194/amt-14-455-2021, 2021.

Lamsal, L. N., Krotkov, N. A., Marchenko, S. V., Joiner, J., Oman, L., Vasilkov, A., Fisher, B., Qin, W., Yang, E.-S., Fasnacht, Z., Choi, S., Leonard, P., and Haffner, D.: TROPOMI/S5P NO_2 Tropospheric, Stratospheric and Total Columns MINDS 1-Orbit L2 Swath 5.5 km x 3.5 km, in, NASA Goddard Earth Sciences Data and Information Services Center (GES DISC) [data set], https://10.5067/MEASURES/MINDS/Data203, 2022.

Laughner, J. L., Neu, J. L., Schimel, D., Wennberg, P. O., Barsanti, K., Bowman, K. W., Chatterjee, A., Croes, B. E., Fitzmaurice, H. L., Henze, D. K., Kim, J., Kort, E. A., Liu, Z., Miyazaki, K., Turner, A. J., Anenberg, S., Avise, J., Cao, H., Crisp, D., de Gouw, J., Eldering, A., Fyfe, J. C., Goldberg, D. L., Gurney, K. R., Hasheminassab, S., Hopkins, F., Ivey, C. E., Jones, D. B. A., Liu, J., Lovenduski, N. S., Martin, R. V., McKinley, G. A., Ott, L., Poulter, B., Ru, M., Sander, S. P., Swart, N., Yung, Y. L., and Zeng, Z. C.: Societal shifts due to COVID-19 reveal large-scale complexities and feedbacks between atmospheric chemistry and climate change, Proc Natl Acad Sci U S A, 118, e2109481118, 10.1073/pnas.2109481118, 2021.

Lelieveld, J., Gromov, S., Pozzer, A., and Taraborrelli, D.: Global tropospheric hydroxyl distribution, budget and reactivity, Atmos. Chem. Phys., 16, 12477-12493, 10.5194/acp-16-12477-2016, 2016.

Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., and Hsu, N. C.: The Collection 6 MODIS aerosol products over land and ocean, Atmos. Meas. Tech., 6, 2989-3034, 10.5194/amt-6-2989-2013, 2013.

Liang, Q., Chipperfield, M. P., Fleming, E. L., Abraham, N. L., Braesicke, P., Burkholder, J. B., Daniel, J. S., Dhomse, S., Fraser, P. J., Hardiman, S. C., Jackman, C. H., Kinnison, D. E., Krummel, P. B., Montzka, S. A., Morgenstern, O., McCulloch, A., Mühle, J., Newman, P. A., Orkin, V. L., Pitari, G., Prinn, R. G., Rigby, M.,

Rozanov, E., Stenke, A., Tummon, F., Velders, G. J. M., Visioni, D., and Weiss, R. F.: Deriving Global OH Abundance and Atmospheric Lifetimes for Long-Lived Gases: A Search for CH3CCl3 Alternatives, Journal of Geophysical Research: Atmospheres, 122, 11,914-911,933, 10.1002/2017JD026926, 2017.

Lovelock, J. E.: Methyl chloroform in the troposphere as an indicator of OH radical abundance, Nature, 267, 32-32, 10.1038/267032a0, 1977.

Mao, J., Ren, X., Brune, W. H., Olson, J. R., Crawford, J. H., Fried, A., Huey, L. G., Cohen, R. C., Heikes, B., Singh, H. B., Blake, D. R., Sachse, G. W., Diskin, G. S., Hall, S. R., and Shetter, R. E.: Airborne measurement of OH reactivity during INTEX-B, Atmos. Chem. Phys., 9, 163-173, 10.5194/acp-9-163-2009, 2009.

Martínez-Alonso, S., Deeter, M., Worden, H., Borsdorff, T., Aben, I., Commane, R., Daube, B., Francis, G., George, M., Landgraf, J., Mao, D., McKain, K., and Wofsy, S.: 1.5 years of TROPOMI CO measurements: comparisons to MOPITT and ATom, Atmos. Meas. Tech., 13, 4841-4864, 10.5194/amt-13-4841-2020, 2020.

McPeters, R. D., Frith, S., and Labow, G. J.: OMI total column ozone: extending the long-term data record, Atmos. Meas. Tech., 8, 4845-4850, 10.5194/amt-8-4845-2015, 2015.

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

Miyazaki, K., Bowman, K. W., Yumimoto, K., Walker, T., and Sudo, K.: Evaluation of a multi-model, multi-constituent assimilation framework for tropospheric chemical reanalysis, Atmos. Chem. Phys., 20, 931-967, 10.5194/acp-20-931-2020, 2020.

Miyazaki, K., Bowman, K., Sekiya, T., Takigawa, M., Neu, J. L., Sudo, K., Osterman, G., and Eskes, H.: Global tropospheric ozone responses to reduced NOx emissions linked to the COVID-19 worldwide lockdowns, Science Advances, 7, eabf7460, 10.1126/sciadv.abf7460, 2021.

Montzka, S. A., Krol, M., Dlugokencky, E., Hall, B., Jockel, P., and Lelieveld, J.: Small Interannual Variability of Global Atmospheric Hydroxyl, Science, 331, 67-69, 10.1126/science.1197640, 2011.

Murray, L. T., Mickley, L. J., Kaplan, J. O., Sofen, E. D., Pfeiffer, M., and Alexander, B.: Factors controlling variability in the oxidative capacity of the troposphere since the Last Glacial Maximum, Atmospheric Chemistry and Physics, 14, 3589-3622, 10.5194/acp-14-3589-2014, 2014.

Murray, L. T., Fiore, A. M., Shindell, D. T., Naik, V., and Horowitz, L. W.: Large uncertainties in global hydroxyl projections tied to fate of reactive nitrogen and carbon, Proceedings of the National Academy of Sciences, 118, e2115204118, 10.1073/pnas.2115204118, 2021.

NASA Goddard Space Flight Center. MERRA2 GMI, NASA [data set], https://acd-ext.gsfc.nasa.gov/Projects/GEOSCCM/MERRA2GMI/, last access: 6 March 2023.

Nault, B. A., Garland, C., Pusede, S. E., Wooldridge, P. J., Ullmann, K., Hall, S. R., and Cohen, R. C.: Measurements of $CH_3O_2NO_2$ in the upper troposphere, Atmospheric Measurement Techniques, 8, 987-997, 10.5194/amt-8-987-2015, 2015.

992 Naus, S., Montzka, S. A., Pandey, S., Basu, S., Dlugokencky, E. J., and Krol, M.: Constraints and biases in a
993 tropospheric two-box model of OH, Atmos. Chem. Phys., 19, 407-424, 10.5194/acp-19-407-2019, 2019.

994 Naus, S., Montzka, S. A., Patra, P. K., and Krol, M. C.: A three-dimensional-model inversion of methyl
995 chloroform to constrain the atmospheric oxidative capacity, Atmospheric Chemistry and Physics, 21,
996 4809-4824, 10.5194/acp-21-4809-2021, 2021.

997 Nicely, J. M., Salawitch, R. J., Canty, T., Anderson, D. C., Arnold, S. R., Chipperfield, M. P., Emmons, L. K.,
998 Flemming, J., Huijnen, V., Kinnison, D. E., Lamarque, J.-F., Mao, J., Monks, S. A., Steenrod, S. D., Tilmes,
999 S., and Turquety, S.: Quantifying the causes of differences in tropospheric OH within global models,
1000 Journal of Geophysical Research: Atmospheres, JD026239, 10.1002/2016JD026239, 2017.

1001 Nicely, J. M., Duncan, B. N., Hanisco, T. F., Wolfe, G. M., Salawitch, R. J., Deushi, M., Haslerud, A. S.,
1002 Jöckel, P., Josse, B., Kinnison, D. E., Klekociuk, A., Manyin, M. E., Marécal, V., Morgenstern, O., Murray, L.
1003 T., Myhre, G., Oman, L. D., Pitari, G., Pozzer, A., Quaglia, I., Revell, L. E., Rozanov, E., Stenke, A., Stone, K.,
1004 Strahan, S., Tilmes, S., Tost, H., Westervelt, D. M., and Zeng, G.: A machine learning examination of
1005 hydroxyl radical differences among model simulations for CCMI-1, Atmospheric Chemistry and Physics,
1006 20, 1341-1361, 10.5194/acp-20-1341-2020, 2020.

1007 Oman, L. D., Douglass, A. R., Ziemke, J. R., Rodriguez, J. M., Waugh, D. W., and Nielsen, J. E.: The ozone
1008 response to ENSO in Aura satellite measurements and a chemistry-climate simulation, Journal of
1009 Geophysical Research-Atmospheres, 118, 965-976, 10.1029/2012jd018546, 2013.

1010 Orbe, C., Oman, L. D., Strahan, S. E., Waugh, D. W., Pawson, S., Takacs, L. L., and Molod, A. M.: Large-
1011 Scale Atmospheric Transport in GEOS Replay Simulations, Journal of Advances in Modeling Earth
1012 Systems, 9, 2545-2560, 10.1002/2017ms001053, 2017.

1013 Patra, P. K., Krol, M. C., Montzka, S. A., Arnold, T., Atlas, E. L., Lintner, B. R., Stephens, B. B., Xiang, B.,
1014 Elkins, J. W., Fraser, P. J., Ghosh, A., Hintsa, E. J., Hurst, D. F., Ishijima, K., Krummel, P. B., Miller, B. R.,
1015 Miyazaki, K., Moore, F. L., Muhle, J., O'Doherty, S., Prinn, R. G., Steele, L. P., Takigawa, M., Wang, H. J.,
1016 Weiss, R. F., Wofsy, S. C., and Young, D.: Observational evidence for interhemispheric hydroxyl-radical
1017 parity, Nature, 513, 219-223, 10.1038/nature13721, 2014.

1018 Pérez-Ramírez, D., Smirnov, A., Pinker, R. T., Petrenko, M., Román, R., Chen, W., Ichoku, C., Noël, S.,
1019 Abad, G. G., Lyamani, H., and Holben, B. N.: Precipitable water vapor over oceans from the Maritime
1020 Aerosol Network: Evaluation of global models and satellite products under clear sky conditions,
1021 Atmospheric Research, 215, 294-304, https://doi.org/10.1016/j.atmosres.2018.09.007, 2019.

1022 Pickett, H. M., Drouin, B. J., Canty, T., Salawitch, R. J., Fuller, R. A., Perun, V. S., Livesey, N. J., Waters, J.
1023 W., Stachnik, R. A., Sander, S. P., Traub, W. A., Jucks, K. W., and Minschwaner, K.: Validation of Aura
1024 Microwave Limb Sounder OH and HO_2 measurements, Journal of Geophysical Research: Atmospheres,
1025 113, D16S30, https://doi.org/10.1029/2007JD008775, 2008.

1026 Pimlott, M. A., Pope, R. J., Kerridge, B. J., Latter, B. G., Knappett, D. S., Heard, D. E., Ventress, L. J.,
1027 Siddans, R., Feng, W., and Chipperfield, M. P.: Investigating the global OH radical distribution using
1028 steady-state approximations and satellite data, Atmos. Chem. Phys., 22, 10467-10488, 10.5194/acp-22-
1029 10467-2022, 2022.

Rigby, M., Montzka, S. A., Prinn, R. G., White, J. W. C., Young, D., O'Doherty, S., Lunt, M. F., Ganesan, A. L., Manning, A. J., Simmonds, P. G., Salameh, P. K., Harth, C. M., Muhle, J., Weiss, R. F., Fraser, P. J., Steele, L. P., Krummel, P. B., McCulloch, A., and Park, S.: Role of atmospheric oxidation in recent methane growth, Proc Natl Acad Sci U S A, 114, 5373-5377, 10.1073/pnas.1616426114, 2017.

Shah, V., Jacob, D. J., Dang, R., Lamsal, L. N., Strode, S. A., Steenrod, S. D., Boersma, K. F., Eastham, S. D., Fritz, T. M., Thompson, C., Peischl, J., Bourgeois, I., Pollack, I. B., Nault, B. A., Cohen, R. C., Campuzano-Jost, P., Jimenez, J. L., Andersen, S. T., Carpenter, L. J., Sherwen, T., and Evans, M. J.: Nitrogen oxides in the free troposphere: implications for tropospheric oxidants and the interpretation of satellite NO_2 measurements, Atmos. Chem. Phys., 23, 1227-1257, 10.5194/acp-23-1227-2023, 2023.

Silvern, R. F., Jacob, D. J., Travis, K. R., Sherwen, T., Evans, M. J., Cohen, R. C., Laughner, J. L., Hall, S. R., Ullmann, K., Crounse, J. D., Wennberg, P. O., Peischl, J., and Pollack, I. B.: Observed NO/NO2 Ratios in the Upper Troposphere Imply Errors in $NO\text{-}NO_2\text{-}O_3$ Cycling Kinetics or an Unaccounted NOx Reservoir, Geophysical Research Letters, 45, 4466-4474, https://doi.org/10.1029/2018GL077728, 2018.

Spivakovsky, C. M., Yevich, R., Logan, J. A., Wofsy, S. C., McElroy, M. B., and Prather, M. J.: Tropospheric OH in a three-dimensional chemical tracer model: An assessment based on observations of CH_3CCl_3, Journal of Geophysical Research: Atmospheres, 95, 18441-18471, 10.1029/JD095iD11p18441, 1990.

Stauffer, R. M., Thompson, A. M., and Young, G. S.: Tropospheric ozonesonde profiles at long-term U.S. monitoring sites: 1. A climatology based on self-organizing maps, Journal of Geophysical Research: Atmospheres, 121, 1320-1339, https://doi.org/10.1002/2015JD023641, 2016.

Stevenson, D. S., Zhao, A., Naik, V., amp, apos, Connor, F. M., Tilmes, S., Zeng, G., Murray, L. T., Collins, W. J., Griffiths, P., Shim, S., Horowitz, L. W., Sentman, L., and Emmons, L.: Trends in global tropospheric hydroxyl radical and methane lifetime since 1850 from AerChemMIP, Atmos. Chem. Phys., 20, 12905-12920, https://doi.org/10.5194/acp-20-12905-2020, 2020.

Strahan, S. E., Duncan, B. N., and Hoor, P.: Observationally derived transport diagnostics for the lowermost stratosphere and their application to the GMI chemistry and transport model, Atmos. Chem. Phys., 7, 2435-2445, 10.5194/acp-7-2435-2007, 2007.

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, Atmospheric Environment, 199, 323-333, 10.1016/j.atmosenv.2018.11.028, 2019.

Susskind, J., Blaisdell, J. M., and Iredell, L.: Improved methodology for surface and atmospheric soundings, error estimates, and quality control procedures: the atmospheric infrared sounder science team version-6 retrieval algorithm, Journal of Applied Remote Sensing, 8, 084994, 10.1117/1.Jrs.8.084994, 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., Gonzalez, 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, Bulletin of the American Meteorological Society, 103, E761-E790, 10.1175/BAMS-D-20-0315.1, 2022.

Turner, A. J., Frankenberg, C., Wennberg, P. O., and Jacob, D. J.: Ambiguity in the causes for decadal trends in atmospheric methane and hydroxyl, Proceedings of the National Academy of Sciences, 114, 5367, 10.1073/pnas.1616020114, 2017.

Turner, A. J., Fung, I., Naik, V., Horowitz, L. W., and Cohen, R. C.: Modulation of hydroxyl variability by ENSO in the absence of external forcing, Proc Natl Acad Sci U S A, 115, 8931-8936, 10.1073/pnas.1807532115, 2018.

van Geffen, J., Boersma, K. F., Eskes, H., Sneep, M., ter Linden, M., Zara, M., and Veefkind, J. P.: S5P TROPOMI NO2 slant column retrieval: method, stability, uncertainties and comparisons with OMI, Atmos. Meas. Tech., 13, 1315-1335, 10.5194/amt-13-1315-2020, 2020.

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 Sensing of Environment, 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 NO_2 measurements with the NDACC ZSL-DOAS, MAX-DOAS and Pandonia global networks, Atmos. Meas. Tech., 14, 481-510, 10.5194/amt-14-481-2021, 2021.

Voulgarakis, A., Naik, V., Lamarque, J. F., Shindell, D. T., Young, P. J., Prather, M. J., Wild, O., Field, R. D., Bergmann, D., Cameron-Smith, P., Cionni, I., Collins, W. J., Dalsøren, S. B., Doherty, R. M., Eyring, V., Faluvegi, G., Folberth, G. A., Horowitz, L. W., Josse, B., MacKenzie, I. A., Nagashima, T., Plummer, D. A., Righi, M., Rumbold, S. T., Stevenson, D. S., Strode, S. A., Sudo, K., Szopa, S., and Zeng, G.: Analysis of present day and future OH and methane lifetime in the ACCMIP simulations, Atmospheric Chemistry and Physics, 13, 2563-2587, 10.5194/acp-13-2563-2013, 2013.

Wang, P., Piters, A., van Geffen, J., Tuinder, O., Stammes, P., and Kinne, S.: Shipborne MAX-DOAS measurements for validation of TROPOMI NO_2 products, Atmos. Meas. Tech., 13, 1413-1426, 10.5194/amt-13-1413-2020, 2020.

Wells, K. C., Millet, D. B., Payne, V. H., Deventer, M. J., Bates, K. H., de Gouw, J. A., Graus, M., Warneke, C., Wisthaler, A., and Fuentes, J. D.: Satellite isoprene retrievals constrain emissions and atmospheric oxidation, Nature, 585, 225-233, 10.1038/s41586-020-2664-3, 2020.

Wild, O., Voulgarakis, A., amp, apos, Connor, F., Lamarque, J.-F., Ryan, E. M., and Lee, L.: Global sensitivity analysis of chemistry-climate model budgets of tropospheric ozone and OH: exploring model diversity, Atmospheric Chemistry and Physics, 20, 4047-4058, 10.5194/acp-20-4047-2020, 2020.

Wolfe, G. M., Nicely, J. M., St Clair, J. M., Hanisco, T. F., Liao, J., Oman, L. D., Brune, W. B., Miller, D., Thames, A., Gonzalez Abad, G., Ryerson, T. B., Thompson, C. R., Peischl, J., McCain, K., Sweeney, C., Wennberg, P. O., Kim, M., Crounse, J. D., Hall, S. R., Ullmann, K., Diskin, G., Bui, P., Chang, C., and Dean-Day, J.: Mapping hydroxyl variability throughout the global remote troposphere via synthesis of airborne and satellite formaldehyde observations, Proc Natl Acad Sci U S A, 116, 11171-11180, 10.1073/pnas.1821661116, 2019.

Wolter, K., and Timlin, M. S.: El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext), International Journal of Climatology, 31, 1074-1087, 10.1002/joc.2336, 2011.

Zhao, Y., Saunois, M., Bousquet, P., Lin, X., Berchet, A., Hegglin, M. I., Canadell, J. G., Jackson, R. B., Hauglustaine, D. A., Szopa, S., Stavert, A. R., Abraham, N. L., Archibald, A. T., Bekki, S., Deushi, M., Jöckel, P., Josse, B., Kinnison, D., Kirner, O., Marécal, V., Connor, F. M., Plummer, D. A., Revell, L. E., Rozanov, E., Stenke, A., Strode, S., Tilmes, S., Dlugokencky, E. J., and Zheng, B.: Inter-model comparison of global hydroxyl radical (OH) distributions and their impact on atmospheric methane over the 2000-2016 period, Atmospheric Chemistry and Physics, 19, 13701-13723, 10.5194/acp-19-13701-2019, 2019.

Zhu, L., Jacob, D. J., Kim, P. S., Fisher, J. A., Yu, K., Travis, K. R., Mickley, L. J., Yantosca, R. M., Sulprizio, M. P., De Smedt, I., González Abad, G., Chance, K., Li, C., Ferrare, R., Fried, A., Hair, J. W., Hanisco, T. F., Richter, D., Jo Scarino, A., Walega, J., Weibring, P., and Wolfe, G. M.: Observing atmospheric formaldehyde (HCHO) from space: validation and intercomparison of six retrievals from four satellites (OMI, GOME2A, GOME2B, OMPS) with SEAC^{4}RS aircraft observations over the southeast US, Atmospheric Chemistry and Physics, 16, 13477-13490, 10.5194/acp-16-13477-2016, 2016.

Zhu, Q., Laughner, J. L., and Cohen, R. C.: Estimate of OH trends over one decade in North American cities, Proc Natl Acad Sci U S A, 119, e2117399119, https://doi.org/10.1073/pnas.2117399119, 2022a.

Zhu, Q., Laughner, J. L., and Cohen, R. C.: Combining Machine Learning and Satellite Observations to Predict Spatial and Temporal Variation of near Surface OH in North American Cities, Environmental Science & Technology, 7362 - 7371, 10.1021/acs.est.1c05636, 2022b.