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

4   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.
5   Follette-Cook[5]
6
7   1. GESTAR II, University of Maryland Baltimore County, Baltimore, MD, USA
8   2. Atmospheric Chemistry and Dynamics Laboratory, NASA Goddard Space Flight Center, Greenbelt,
9   MD, USA
10   3. Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA
11   4. GESTAR II, Morgan State University, Baltimore, MD, USA
12   5. Mesoscale Atmospheric Processes Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD,
13   USA
14
15   *Correspondence to:* Daniel C. Anderson (daniel.c.anderson@nasa.gov)

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

39

40   **1   Introduction**
41   The hydroxyl radical (OH) dictates the lifetime of many tropospheric species, including carbon monoxide
42   (CO), methane (CH_4), and numerous volatile organic compounds (VOCs). Knowledge of OH is therefore
43   necessary to understand the abundance, distribution, and variability of these species. For instance,
44   Rigby et al. (2017) and Laughner et al. (2021) attribute recent trends and increases in CH_4 at least
45   partially to changes in OH abundance. Current constraints on OH are insufficient, however, to assess its
46   relative importance in controlling these trends (Turner et al., 2017).
47

48   Differences in OH distributions among chemistry transport (CTM) and chemistry climate models (CCM)
49   suggest that these models are insufficient to inform understanding of OH abundance and variability
50   without further observational constraints. OH abundance can differ by up to 80% among models
51   constrained with identical emissions in intercomparison projects (Voulgarakis et al., 2013;Nicely et
52   al.;Zhao et al., 2019;Murray et al., 2021). Variables such as the photolysis frequency of O_3 (JO^1D) (Nicely
53   et al., 2020), the NO_X lifetime ($NO_X = NO + NO_2$), and the oxidation efficiency of VOCs (Murray et al.,
54   2021) contribute to these inter-model variations in OH. Likewise, the response of OH to the El Niño
55   Southern Oscillation (ENSO), the dominant mode of OH variability on monthly and seasonal timescales
56   (Anderson et al., 2021;Turner et al., 2018), and other modes of internal climate variability can vary
57   widely among models (Anderson et al., 2021).

59   Despite this need for better constraints, observations of tropospheric OH are limited. The hydroxyl
60   radical has a lifetime of approximately 1s (Mao et al., 2009), resulting in large spatial heterogeneity in
61   both the horizontal and vertical. A strategic, representative in situ observational network is therefore
62   unfeasible. As a result, observations of OH are generally limited to intensive field campaigns (Miller and
63   Brune, 2022) that have narrow spatial and temporal coverage. While remotely-sensed OH observations
64   are available, those from satellites are limited to the stratosphere (e.g., Pickett et al., 2008), while
65   ground-based observations of total column OH are dominated by the stratospheric contribution (e.g.,
66   Burnett and Minschwaner, 1998).

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

74   Multiple studies have attempted to constrain OH through the creation of proxies and the application of
75   satellite retrievals of OH drivers. Murray et al. (2014) showed that global OH strongly correlated with a
76   combination of JO^1D, water vapor ($H_2O_{(v)}$), and the tropospheric sources of reactive nitrogen and carbon
77   in the GEOS-Chem model. Murray et al. (2021) demonstrated that OH correlated with this proxy in
78   multiple CTMs, although the relationship differs strongly among models. Wolfe et al. (2019) developed
79   a proxy for OH based on formaldehyde (HCHO) production and loss rates. They applied that proxy to
80   satellite HCHO observations to estimate OH columns in the remote troposphere, a region where HCHO
81   abundance is low and the satellite retrievals are reflective of the *a priori* (Zhu et al., 2016). Using
82   machine learning, chemical transport model output, and retrievals of NO_2 and HCHO, Zhu et al. (2022a)
83   developed a method to estimate surface OH in North American urban areas. Finally, Pimlott et al.
84   (2022) used a steady state approximation of OH, including primary production from H_2O and O_3 and loss
85   from CO, CH_4, and O_3, to estimate OH between 600 and 700 hPa using observations from IASI (Infrared
86   Atmospheric Sounding Interferometer). A logical next step, building on the results of these studies, is
87   the development of a methodology to constrain OH that ingests multiple satellite retrievals,
88   encompasses the breadth of OH chemical and dynamical drivers, and spans a significant enough portion
89   of the globe to inform variability and trends in CH_4 and CO loss.

91   Combining machine learning, chemical transport model (CTM) output, and satellite data has the
92   potential to constrain tropospheric column OH (TCOH). A variety of machine learning techniques, such
93   as neural networks (Nicely et al., 2017;Nicely et al., 2020;Kelp et al., 2020), self-organizing maps
94   (Stauffer et al., 2016), random forest regression (Keller and Evans, 2019), and gradient boosted
95   regression trees (GBRTs) (Ivatt and Evans, 2020;Zhu et al., 2022a;Anderson et al., 2022) show promise in

96 helping to solve problems in atmospheric chemistry. In particular, Zhu et al. (2022a) and Anderson et al.
97 (2022) demonstrated the ability of GBRTs to predict OH from a chemical transport model with
98 reasonable accuracy. GBRT models (Elith et al., 2008;Chen and Guestrin, 2016) use an ensemble of
99 decision trees to predict the value of a target based on multiple inputs, even for targets with highly non-
100 linear dependencies on the inputs.

101

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

114

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

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

124

2.1 Creation of the TCOH model
2.1.1 Creation of the GBRT training dataset

127 For the machine learning model training dataset, we use a subset of output from the MERRA2 GMI
128 simulation (https://acd-ext.gsfc.nasa.gov/Projects/GEOSCCM/MERRA2GMI/). MERRA2 GMI is a 40 year
129 (1980 – 2019) simulation of the NASA GEOS model run in replay mode (Orbe et al., 2017) with MERRA2
130 (Modern Era Retrospective analysis for Research and Applications, version 2) meteorology (Gelaro et al.,
131 2017). The simulation has a resolution of c180 on the cubed sphere (approximately $0.625°$ longitude by
132 $0.5°$ latitude) with 72 vertical layers and uses the Global Modeling Initiative (GMI) chemical mechanism
133 (Duncan et al., 2007;Strahan et al., 2007). Output is available at daily- and monthly-averaged resolution,
134 as well as instantaneous values corresponding with the overpass times of the satellites described in
135 Section 2.2. Anderson et al. (2021) and Strode et al. (2019) provide detailed information about the
136 simulation, including emissions.

137

138 The training target for the machine learning model is TCOH. In Anderson et al. (2022), we developed a
139 GBRT parameterization trained on MERRA2 GMI output to predict in situ OH concentrations using 27
140 inputs, only a small fraction of which are observable from space. That parameterization, designed to be
141 integrated into the GEOS modeling framework, performed better when there was a separate model for
142 each month as opposed to one model for all months. While that GBRT model is not appropriate for the
143 application described here, we employ a similar approach, creating a separate set of TCOH training

144   targets for each month. We use instantaneous OH output from MERRA2 GMI at 14:00 local time for
145   each day of a given month across the years 2005 to 2019, to correspond with the satellite record. We
146   omitted data from 2017 to evaluate model performance. For a given month and year, we calculate daily
147   tropospheric column values across the grid, filtering out columns where the maximum cloud fraction in
148   that column was greater than 30%, yielding approximately 43,000 valid grid boxes per day. For each
149   year, we then average these values to monthly resolution. This results in approximately 600,000 total
150   training targets for each month over the 15-year period.

151

152   *Table 1*: Input variables to the machine learning model and the corresponding satellite retrieval used to create the
153   satellite OH product.

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

154

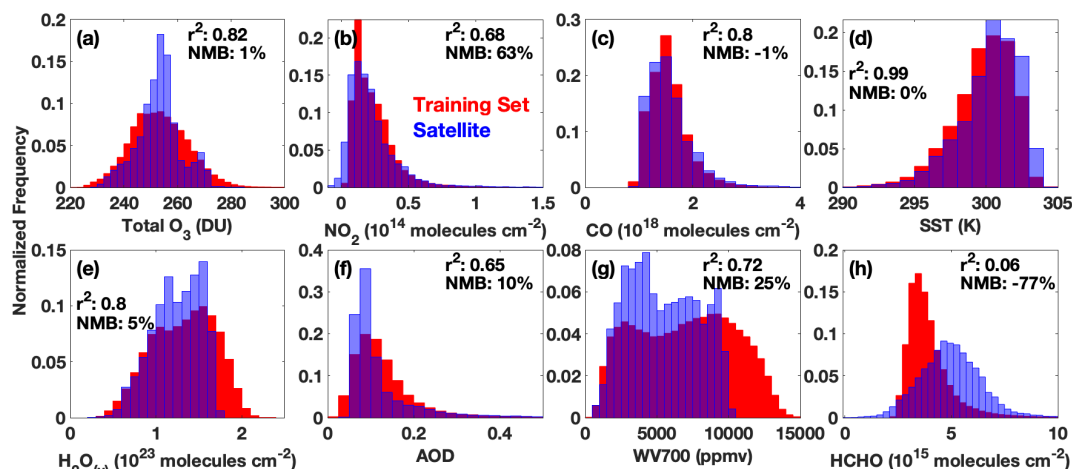
155   We selected the input variables for the machine learning model (Table 1) based on their relevance to OH
156   chemistry and variability as well as our current ability to observe the variable with satellites.
157   Performance was similar for a model including total column ozone only and for a model also including
158   the tropospheric column. We therefore use total column ozone because of the uncertainties inherent in
159   separating the column into two parts in the satellite retrieval. We chose the water vapor layers to
160   correspond with the Atmospheric Infrared Sounder (AIRS) layers product. Layers are averages over the
161   indicated pressure range, and we denote the layer names by the highest pressure in that range. We
162   include sea surface temperatures (SST) as a proxy for the Indian Ocean Dipole and ENSO, which has a
163   strong impact on OH variability in the tropics (Anderson et al., 2021;Turner et al., 2018). In addition, we
164   include latitude and solar zenith angle as previous work has shown that these variables can explain a
165   large fraction of the spatial OH variability (Duncan et al., 2000;Anderson et al., 2022).

166

167   We sampled the MERRA2 GMI output to create the training dataset in the same manner as for the TCOH
168   targets. The inputs to the machine learning model each correspond to the same model column as the
169   OH target. All column values are instantaneous and taken from 14:00 to correspond with satellite
170   overpass times, except for CO, which is for 10:00, corresponding with the Measurement of Pollution in
171   the Troposphere (MOPITT) overpass time. Model performance was similar when using CO output at
172   14:00 and 10:00, likely because of limited diurnal variability in CO column in the study region. SSTs are

173   monthly averages of 24-hour averaged values, and we calculated solar zenith angle at the surface for
174   noon on the 15th of a given month.
175
176   **2.1.2 Creation and tuning of the GBRT model**
177   We used the XGBoost package (Chen and Guestrin, 2016) version 0.81 in Python version 3.6 to create a
178   GBRT model of TCOH for each month using the training datasets from MERRA2 GMI. For each month,
179   we used 90% of the dataset for model training and the remainder for model validation. As mentioned in
180   Section 2.1.1, we also used MERRA2 GMI output from 2017, which was omitted from the training
181   dataset, as further validation.
182
183   To maximize parameterization performance while also balancing the potential of overfitting, we tuned
184   hyperparameters, including the learning rate, the maximum tree depth, and the number of trees. We
185   chose hyperparameter values that minimized the parameterization root mean square error (RMSE) of
186   the training dataset. We set the learning rate, which controls the magnitude of change when adding a
187   new tree, to 0.1, while we varied the maximum tree depth and number of trees from 6 to 22 and from
188   10 to 150, respectively. For both maximum tree depth and number of trees, RMSE initially dropped
189   significantly with increasing value, representing sharp improvement in parameterization performance.
190   RMSE values eventually plateaued, increasing parameterization runtime without noticeably improving
191   performance. A combination of a maximum tree depth of 18 and 100 trees balanced performance with
192   model training and run time.
193
194   To determine whether the inputs to the machine learning model improved or hindered performance, we
195   performed a "leave one out" analysis. Using 5-fold cross validation, we retrained the model, individually
196   omitting each of the inputs, to determine the percent difference between the mean RMSE of the 5 folds
197   for the model without a specific input and one including all inputs. Omitting the inputs listed in Table 1
198   lead to increases in the RMSE, suggesting that each is necessary for improved model performance. As a
199   result of this analysis, we do not use water vapor layers for pressures less than 300 hPa because these
200   decreased model performance.
201
202   Finally, we found that it was not necessary to apply satellite averaging kernels and shape factors to the
203   training dataset. Of the satellite retrievals used in this work (discussed in Sect. 2.2 and listed in Table 1),
204   only CO, HCHO, and NO_2 could require convolving the model with the averaging kernel. Shape factors
205   for the OMI NO_2 retrieval are determined from a similar setup of the GEOS model, also employing the
206   GMI chemical mechanism and MERRA2 meteorology. Applying the satellite shape factors to the
207   simulation discussed here would therefore not result in significant changes in the modeled NO_2
208   (Anderson et al., 2021). To test whether it is necessary to apply the averaging kernels for CO and HCHO,
209   we created a separate training dataset, where we convolved the daily MERRA2 GMI output with the
210   averaging kernel and a priori from the level 2 data for both species for February 2005 - 2019. All other
211   inputs were kept the same. We then retrained the model with these adjusted CO and HCHO variables.
212   When we applied the satellite data to the model for February 2017, as described in Section 4, the
213   resulting TCOH differed by less than 1% on average from the model that did not include averaging kernel
214   information. This level of uncertainty is significantly smaller than the other uncertainties discussed in
215   Section 5, so we do not include averaging kernels in our analysis.
216
217   **2.2 Description of satellite products**
218   To create the observationally-constrained OH product, we use multiple satellite retrievals, listed in Table
219   1 and briefly described here. For each satellite retrieval, we use the level 3 gridded product, with the
220   exception of SST which is level 4. Where necessary, we regridded the retrieval to a common horizontal

221 grid with a resolution of $1.0° \times 1.0°$ and averaged to the monthly scale. Each instrument is located
222 onboard a polar orbiting satellite that provides near global coverage daily.
223
224 We use retrievals of three species – HCHO, O_3, and NO_2 – from the Ozone Monitoring Instrument (OMI),
225 an ultraviolet-visible spectrometer located onboard the Aura satellite, which has an overpass of
226 approximately 13:30 local solar time (LST). We use the Smithsonian Astrophysical Observatory (SAO)
227 version 3 HCHO retrieval (González Abad et al., 2015). Wolfe et al. (2019) found that this retrieval
228 captured the variability of the HCHO columns in the remote atmosphere observed during the
229 Atmospheric Tomography (ATom) campaign with little bias. For total column O_3, we use the TOMS-like
230 (Total Ozone Mapping Spectrometer) retrieval version 3 (McPeters et al., 2015), which agrees with
231 ground-based and other satellite observations within approximately 1% (Labow et al., 2013). Finally, we
232 use the Goddard Space Flight Center version 4 NO_2 retrieval (Lamsal et al., 2021). While previous
233 studies have thoroughly evaluated this retrieval in more polluted atmospheres (e.g., Lamsal et al.,
234 2014;Choi et al., 2020), evaluation in the remote tropical atmosphere, as defined in this study, is limited.
235
236 For water vapor and aerosol optical depth (AOD) at 550 nm, we use retrievals from AIRS and the
237 Moderate Resolution Imaging Spectroradiometer (MODIS) instruments, respectively, both located
238 onboard the Aqua satellite with an overpass of approximately 13:30 LST. We use the total column water
239 vapor standard physical retrieval as well as the 7 water vapor layers listed in Table 1 (Susskind et al.,
240 2014). Multiple studies have evaluated the accuracy of the AIRS $H_2O_{(v)}$ column and layers retrievals in
241 the remote tropical atmosphere, finding bias of 5% or less and high correlation against both remote and
242 *in situ* observations (Bedka et al., 2010;Anderson et al., 2016;Pérez-Ramírez et al., 2019). We use
243 collection 6 of the dark target MODIS AOD retrieval at 550 nm, which is highly correlated with
244 observations from the AERONET network over the ocean (Levy et al., 2013).
245
246 We also use retrievals of CO from MOPITT, which is onboard the Terra satellite with an overpass of
247 10:30 LST. We use the version 8 retrieval that includes both near and thermal infrared radiances (Deeter
248 et al., 2019). CO retrievals from MOPITT in the remote tropics generally agree with ground-based
249 remotely-sensed observations within 10% (Hedelius et al., 2019;Buchholz et al., 2017).
250
251 Finally, we use SSTs from the Multi-scale Ultra-high Resolution (MUR) analysis, which combines
252 nighttime SST observations from multiple satellite platforms, including MODIS, as well as *in situ*
253 observations and agrees with other SST analyses within 0.36° C (Chin et al., 2017).
254
255 **3   Evaluating the Suitability of the MERRA2 GMI Simulation as a Training Dataset**
256 Before generating the GBRT model to predict TCOH, we first demonstrate that the MERRA2 GMI
257 simulation is suitable to use as a training dataset. Because of the paucity of in situ observations of OH
258 over most of the globe, we necessarily use output from an atmospheric chemistry model to train the
259 machine learning model. The atmospheric chemistry model output must reasonably capture the
260 distribution, magnitude, and ENSO-related variability of OH and the drivers listed in Table 1, as GBRT
261 models are unable to extrapolate beyond the photochemical environments on which they are trained
262 (Anderson et al., 2022).
263
264 **3.1 Comparison of the Distribution and Magnitude of Simulated OH Drivers to Observations**
265 Simulated OH from MERRA2 GMI agrees with observations over the remote ocean within the
266 instrumental uncertainty. Anderson et al. (2021) compared MERRA2 GMI output to in situ observations
267 from the first two deployments of the Atmospheric Tomography mission (ATom), finding modest
268 correlation (r^2 values between 0.3 and 0.78 depending on the hemisphere and season) between

269   observations and the model. The average normalized mean bias was on the order of 20%, a slight high
270   bias but within the 2σ observational uncertainty of 35%. Agreement was highest in the remote
271   atmosphere, whereas the largest error was in regions of fresh, continental outflow off the coasts of
272   South America and New Zealand.
273

274
275   *Figure 1:* Comparison of the normalized distributions of the training dataset (red) for the February model and
276   satellite observations of the indicated species for February 2017 (blue). Purple indicates regions of overlap. We
277   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
278   S1. We also indicate the r^2 of the correlation between MERRA2 GMI output for February 2017 and the
279   corresponding satellite retrieval as well as the normalized mean bias of that output.

280   The simulation captures both the observed variability and the magnitude of the majority of GBRT model
281   inputs with reasonable fidelity, suggesting that the satellite retrievals highlighted in Section 2.2 are
282   suitable inputs for a machine learning model trained on MERRA2 GMI output (Fig.1). Figure 1 compares
283   the distribution of the February training dataset created from the MERRA2 GMI simulation for 2005 –
284   2019 to the satellite observations of the indicated species for February 2017, a month omitted from the
285   training dataset. Distributions of the remaining water vapor layers are shown in Figure S1. In addition,
286   correlations between observations and MERRA2 GMI output for February 2017 are shown, as an
287   example, in Figures S2 and S3. With the exception of HCHO, distributions of the species are similar
288   between the observations and MERRA2 GMI, with the training dataset encompassing the full range of
289   almost all species. A GBRT model trained on MERRA2 GMI will therefore likely not have to extrapolate
290   to photochemical environments on which it was not trained when applied to the satellite data. Further,
291   MERRA2 GMI total column O_3, $H_2O_{(v)}$ column, AOD, CO, and SSTs are all highly correlated (r^2 of 0.65 or
292   higher) with their respective satellite observations, and biases are within 10%, on average. Anderson et
293   al. (2021) did show that MERRA2 GMI CO columns demonstrate biases of opposite sign in the Northern
294   and Southern Hemispheres, however.
295
296   Agreement between MERRA2 GMI and satellite observations for NO_2, HCHO, and the $H_2O_{(v)}$ layers is
297   more variable than for the other species. While modeled NO_2 is moderately correlated with
298   observations ($r^2 = 0.68$) with relatively similar distributions, MERRA2 GMI has a NMB of 63%. This
299   disagreement is most pronounced at low column values, however, where observational uncertainty is
300   large. Further, Anderson et al. (2021) demonstrated distinct regions of bias in NO_2 related to biomass
301   burning and lightning emissions. Modeled HCHO, on the other hand, is not correlated with observations

302   and is biased low by -77%. Modeled water vapor layers are all modestly correlated with observations (r^2
303   of 0.64 or greater) but vary in their bias, with the 925, 850, 700, and 300 hPa layers biased within 30%
304   and the remaining layers biased up to 71%.

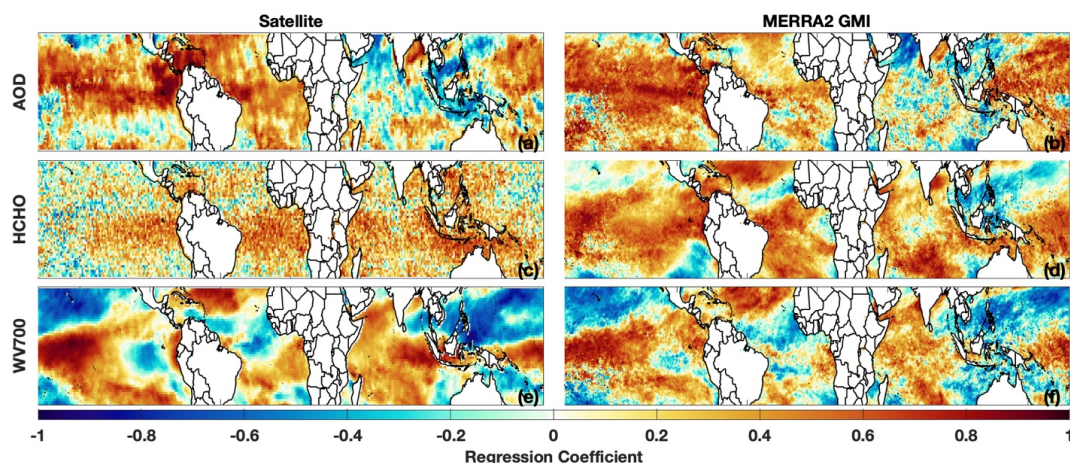
306   The satellite product is insensitive to the differences between the HCHO distribution of the satellite and
307   training dataset highlighted in Figure 1. To determine the effects of the difference in HCHO distribution,
308   we extended the training dataset to cover the full time period of the MERRA2 GMI simulation (1980 –
309   2019) and then subsampled the resultant data to match the satellite HCHO distribution. Extending the
310   training dataset to 1980 allows for the subsampled training dataset to have a similar size (~600,000
311   points) as the original training set. We then created a new machine learning model using this sub-
312   sampled dataset and calculated OH fields for Feb. 2017 using the satellite inputs from Table 1. We
313   compared this to the TCOH field calculated from a model using the original training dataset, finding
314   agreement within 5%. This uncertainty is small in comparison to that resulting from uncertainties in the
315   NO_2 and HCHO satellite retrievals discussed in Section 5.2. If the uncertainty of the satellite inputs
316   decreases, as retrievals and instruments improve, then it will become necessary to more closely align
317   the training and observed HCHO distributions.

319   Finally, because NO_2 and HCHO have the largest differences between satellite observations and the
320   training dataset, we trained a separate machine learning model to predict TCOH, omitting these two
321   species as inputs. When this model was evaluated using the independent MERRA2 GMI output
322   described in Section 4.1, the NRMSE was 10.1%, a more than factor of 2 degradation in performance as
323   compared to the baseline model. This suggests that omitting these species from the machine learning
324   model would result in a greater uncertainty in the final TCOH product than that which results from the
325   retrieval uncertainties and the potential discrepancies between observations and the training dataset.

327   **3.2 Evaluation of the simulated ENSO-related variability of OH drivers**
328   Because ENSO is the dominant mode of OH variability (Anderson et al., 2021;Turner et al., 2018), the
329   training dataset must also capture the ENSO-related variability of the GBRT model inputs. Anderson et
330   al. (2021) demonstrated that the correlation of columns of CO, $H_2O_{(v)}$, and to a lesser extent NO_2, from
331   the MERRA2 GMI simulation with the Multivariate ENSO Index (MEI) (Wolter and Timlin, 2011) agreed
332   closely with correlations of the corresponding species for observations from MOPITT, AIRS, and OMI.
333   Unsurprisingly, based on the strong correlation and low bias of MERRA2 GMI SSTs with observations, the
334   simulation also captures the relationship between SSTs and ENSO. The simulation therefore sufficiently
335   captures the ENSO-related variability of these species to act as training data for the GBRT model. We
336   now evaluate this relationship for the remaining GBRT model inputs.

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

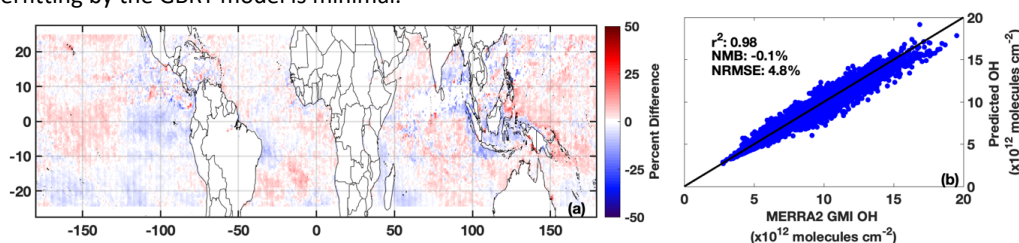
348   Modeled accuracy of the HCHO-ENSO relationship is more difficult to assess. While both the OMI
349   retrieval and MERRA2 GMI demonstrate broad regions of anti-correlation between HCHO and ENSO, the

350   correlations with OMI HCHO are weaker and noisier than for the other satellite retrievals. Over much of
351   the domain, HCHO abundance is low, often at or below the retrieval detection limit, suggesting that the
352   HCHO retrieval might not be of sufficient quality to capture ENSO-related variability. We investigate the
353   impacts of the HCHO observational uncertainty in Section 5.
354

355
356   *Figure 2:* Distribution of the regression coefficient of a linear least squares fit of the indicated variable against the
357   MEI for the respective satellite retrieval (a, c, and e) and MERRA2 GMI (b,d, and f) for February. Regressions of
358   AOD are for 2010 to 2019, the years for which we have a one-degree, gridded satellite product, while HCHO and
359   water vapor 700 hPa are for 2005 to 2019. Satellite data are on a $1° \times 1°$ grid while model output is at the native
360   model resolution.

361   Finally, because we use total column O_3 as an input to the GBRT model, we do not evaluate the
362   relationship between ENSO and O_3, as the stratosphere dominates the O_3 column and the ENSO-related
363   variability is mostly confined to the troposphere. Oman et al. (2013) found that a GEOS CCM simulation
364   and a combination of O_3 retrievals from the Microwave Limb Sounder (MLS) and the Tropospheric
365   Emission Spectrometer (TES) exhibited similar ENSO-related variability in the middle and upper
366   troposphere, demonstrating that simulations in the GEOS framework can capture this relationship. If a
367   TES-like satellite retrieval were currently available, it could be a valuable contributor to the GBRT model
368   described here, as it would provide vertically-resolved information about one of the primary drivers of
369   OH production.
370

4   Tropical tropospheric column OH constrained with observations of its drivers
372   We now demonstrate the ability of the GBRT model to determine TCOH. First, we show that the GBRT
373   model can reproduce MERRA2 GMI modeled TCOH from a year independent of the training dataset, a
374   so-called "hold out set" (Sect. 4.1). We then input satellite data from one month from each season into
375   the GBRT model to evaluate the realism of the calculated TCOH fields (Sect 4.2).
376

4.1 Evaluation with an independent year from MERRA2 GMI
378   The machine learning model is able to capture both the magnitude and the variability of TCOH across
379   each season when applied to MERRA2 GMI output from 2017, a year independent of the training
380   dataset. For August 2017 (Fig. 3b), the predicted TCOH is highly correlated with MERRA2 GMI (r^2 of
381   0.98). TCOH from the machine learning model agrees with the CTM simulation within 4.8% on average.
382   The overall normalized mean bias (NMB) is negligible (-0.1%), although there are some regions of

383   coherent bias (Fig. 3a). Results are similar for February, May, and October 2017 (Fig. S5). The
384   normalized root mean square error for each of these months is comparable to that found for a GBRT
385   parameterization of OH created with a similar methodology that included 27 inputs (Anderson et al.,
386   2022). This suggests that limiting inputs to model variables observable from space does not degrade the
387   ability of the machine learning model to predict TCOH. The low bias and high correlation between the
388   GBRT and MERRA2 GMI TCOH for all four months examined here also suggests that any potential
389   overfitting by the GBRT model is minimal.

390
391   **Figure 3:** Percent difference between TCOH predicted by the machine learning model and that from MERRA2 GMI
392   for August 2017, a month and year omitted from the training dataset **(a)**. A regression of the machine learning TCOH
393   against MERRA2 GMI for the same month **(b)**. The r^2 of a linear, least squares regression, along with the normalized
394   mean bias (NMB) and normalized root mean square error (NRMSE), are also indicated.

4.2 TCOH from satellite observations of its drivers

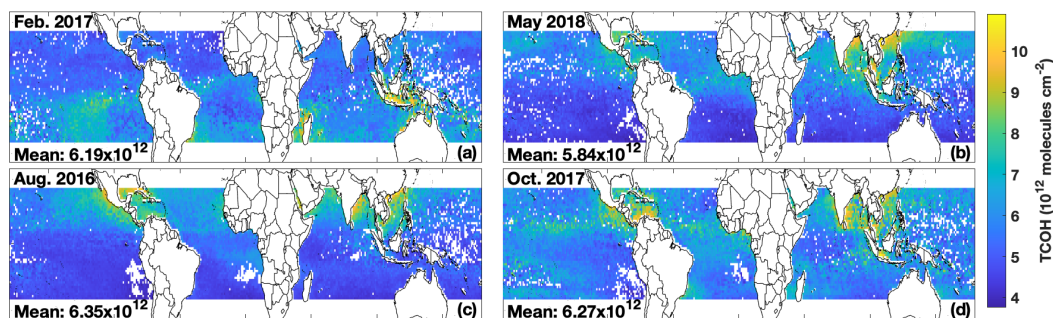
396   We now apply satellite data from the four months corresponding to the ATom campaign (Aug. 2016,
397   Feb. 2017, Oct. 2017, and May 2018) to the GBRT model to determine TCOH fields across the tropics.
398   More details about ATom as well as evaluation of the GBRT model with ATom observations are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

405

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

414

415   In general, TCOH from the multi-satellite product differs in both magnitude and distribution from the
416   MERRA2 GMI simulation. For example, for Feb. 2017, mean MERRA2 GMI TCOH is 6.96×10^{12}
417   molecules/cm^2, 12% higher than the satellite product (Fig. S6). This is consistent with the comparison to
418   *in situ* observations discussed in Section 3.1 where MERRA2 GMI overestimates ATom observations by
419   ~20% and underestimates CH$_4$ lifetime, suggesting that the satellite product is again of reasonable
420   magnitude. While understanding the satellite/model differences is beyond the scope of this work, we
421   consider the variety in TCOH spatial distributions generated by the GBRT model to be promising. The
422   difference between the satellite-constrained product and MERRA2 GMI lends some confidence that the
423   GBRT model is not overfit or "tied" to geographic determiners in the training dataset, but rather, is

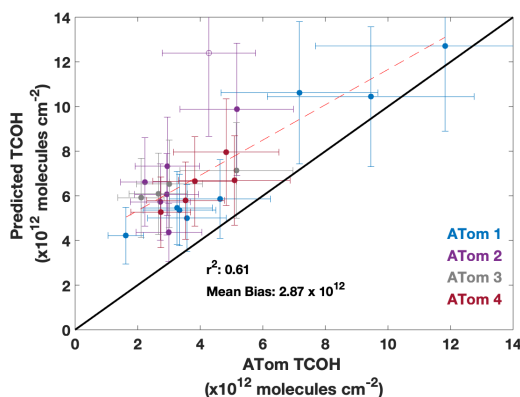
424 sensitive to variations in the chemical and dynamical drivers of OH. These results all suggest that the
425 methodology presented here can produce a reasonable satellite TCOH product in the tropics, with
426 values and distributions independent of the chemistry model used to create the GBRT model.
427

429 **Figure 4:** TCOH calculated with the machine learning model using satellite inputs for the months of each ATom
430 deployment: Feb. 2017 (a), May 2018 (b), Aug. 2016 (c), and Oct. 2017 (d). The mean, domain-wide TCOH value in
431 molecules/cm² for each month is also indicated.

5   Understanding and mitigating potential challenges in using this methodology to constrain TCOH
433 In this section, we outline possible limitations of the machine learning methodology and the current
434 observational network of the GBRT model inputs and provide potential means to mitigate these
435 limitations where necessary. In section 5.1, we discuss the current lack of sufficient in situ observations
436 to thoroughly evaluate the methodology, highlighting this point by validating the GBRT model with data
437 from the ATom campaign. In section 5.2, we investigate the impacts of random retrieval errors in
438 satellite retrievals on the TCOH product, while in section 5.3, we evaluate the impacts on TCOH when
439 using different satellite retrievals as inputs.
440

5.1 Insufficient in situ observations for thorough independent evaluation
442 While we demonstrated in Section 4.1 that TCOH calculated with the GBRT model agrees closely with a
443 hold-out set from MERRA2 GMI, it is also important to demonstrate that the GBRT model can replicate
444 observed TCOH from the actual atmosphere. Because the satellite TCOH product shown in Figure 4 is
445 monthly and at a 1° × 1° resolution, however, there are no observations with which to evaluate the
446 product. We can test the ability of the GBRT model to reproduce observed TCOH from field campaigns,
447 however, assuming there are concomitant observations of the input species listed in Table 1. The
448 additional need for tropospheric column values of many of these species severely limits the datasets
449 available for validation. To our knowledge, the ATom campaign is the only source of the required inputs
450 with enough observations to attempt a limited validation.
451

452 During ATom (Thompson et al., 2022), scientists measured a suite of air quality and climate relevant
453 trace gases and aerosols throughout the atmosphere above the remote Pacific and Atlantic. ATom took
454 place in four parts: ATom 1 (July – August 2016), ATom 2 (January – February 2017), ATom 3 (September
455 – October 2017), and ATom 4 (April – May 2018). During each deployment, flights consisted of a series
456 of ascents and descents across all tropical latitudes over the Pacific and Atlantic Oceans. This allows for
457 the calculation of tropospheric column content of the observed species and evaluation of the machine
458 learning model across most latitudes of our study domain and across all seasons.
459

460 To evaluate the GBRT model performance, we calculated TCOH using the GBRT model and observations
461 from the ATom deployments as inputs. We then compared the values to the observed OH columns. To

462   calculate the column values from the observations, we averaged data into 25 hPa pressure bins for each
463   ATom profile. We filled in missing data using a log-linear interpolation and then integrated the column.
464   Our analysis here includes only profiles with observations of all necessary species, that spanned at least
465   700 hPa, and where less than 25% of the pressure bin values were interpolated. We also omitted any
466   profiles that had pressure bins with negative OH values. In addition, we restrict our analysis to latitudes
467   within 25° of the equator and profiles conducted between 12:00 and 15:00 LST. Values for total column
468   O_3, AOD, and SSTs, for which there were no observations during ATom, were taken from the MERRA2
469   GMI simulation from the grid box closest to the center of the respective profile. The spatial distribution
470   of the valid ATom columns and the corresponding columns calculated with the GBRT model are shown
471   in Figure S7.
472

473
474   ***Figure 5***: Regression of TCOH observed from the ATom deployments against that predicted from the GBRT model.
475   Error bars represent the 2σ observational uncertainty as reported in Brune et al. (2020) and the GBRT uncertainty
476   described in Section 5.2. The r^2 of a linear least squares fit and the mean bias is also shown.

477   The GBRT model captures the variability of the observed TCOH although there is a consistent high bias.
478   When applied to all ATom deployments, predicted TCOH is correlated with the observations with an r^2
479   of 0.61 and a mean bias of 2.67×10^{12} molecules/cm^2 (Fig.5). Despite the high bias, many of the data
480   points agree within the combined modeled and observational uncertainty. The r^2 value is heavily driven
481   by one outlier from the ATom 2 campaign (open circle), which, when removed, increases the r^2 to 0.78.
482   Likewise, there is variability in the correlation among the deployments, ranging from 0.94 for ATom 1
483   and 3 to 0.57 for ATom 2. The level of agreement between observed and predicted OH, particularly for
484   ATom1 and 3, is comparable or better than that of other methods to infer OH from space. For example,
485   Pimlott et al. (2022) found an r of 0.78 (r^2 = 0.61) when estimating ATom OH using a steady state
486   approach, with r values ranging from 0.51 to 0.85 (r^2 of 0.26 to 0.72) for the different deployments. The
487   level of agreement we show here therefore demonstrates the validity of the machine learning method
488   to capture the variability of OH.
489
490   The source of the constant high bias, equal almost to a factor of 2 overestimate at the lowest column
491   values, is unclear, although there are multiple potential error sources. For example, a typical profile
492   taken during ATom spanned 300 – 400 km in latitude, disconnecting the top and bottom of the profile in
493   space. This is in contrast to the data used to train the model, which were vertical columns over one
494   location. This could lead to a degradation in model performance when applied to ATom, since the
495   columns are not directly analogous to the training dataset. These effects are likely limited because
496   ATom observations are in the remote atmosphere, where the spatial distribution of relevant species is

497   likely to be more homogeneous than over land. Likewise, in most cases, ATom columns do not span the
498   entire troposphere. While this is unlikely to consistently result in large errors, if a large fraction of the
499   tropospheric column of one input was outside the range of the ATom profile, this would likely cause
500   large errors in calculated TCOH.

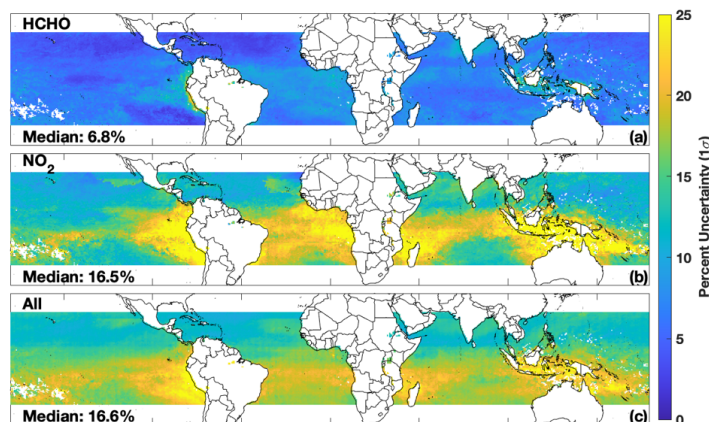
502   Further, there is a known interference with the ATom NO_2 observations. Because of thermal
503   degradation of NO_2 reservoir species, such as organic nitrates and peroxyacetyl nitrate, in the
504   instrument inlet, NO_2 observations are likely biased high (Shah et al., 2022;Thompson et al.,
505   2022;Silvern et al., 2018). Recalculating the TCOH from ATom with NO_2 from a box model constrained
506   with NO observations with a setup similar to that described in Nicely et al. (2016) yields a similar
507   correlation (Fig. S8) as that shown in Figure 5 but with a lower slope, leading to underestimates at
508   higher column values. The uncertainty in observed NO_2 could explain some of the scatter seen in Figure
509   5. For more certain evaluation with observations, greater certainty in the in situ NO_2 observations is
510   needed. Although the in situ observations are insufficient to evaluate the absolute accuracy of the
511   product, the results presented here demonstrate that a machine learning model trained on data from a
512   CTM simulation can capture TCOH variability in the actual atmosphere.

514   **5.2 Impacts of uncertainties in the satellite retrievals on TCOH**
515   In the remote atmosphere where HCHO and NO_2 abundances are low, retrieval uncertainty of an
516   individual pixel for both species can be on the order of 100% and is often reflective of the *a priori*
517   (González Abad et al., 2015;Lamsal et al., 2021). Given the importance of these species to the GBRT
518   model as well as to OH chemistry, it is necessary to determine how the propagation of the retrieval
519   uncertainties from these and other model inputs impacts the predicted TCOH.

521   We determined the total uncertainty in TCOH from all inputs as well as the resultant uncertainty from
522   each individual input for Feb. 2017. First, we estimated an average retrieval uncertainty for each input
523   based on reported values in the retrieval files or from the literature (Table S1). We note that for NO_2
524   and HCHO we use a fit uncertainty for a single retrieval. Because we are using monthly-averaged data at
525    $1° \times 1°$ horizontal resolution, this likely significantly overestimates the actual uncertainty in these
526   retrievals as the random error from individual pixels will tend to cancel when averaged over such large
527   and spatial temporal scales. Our results are therefore an upper bound on the estimated TCOH
528   uncertainty.

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

538   Uncertainty from the NO_2 retrieval, and to a lesser extent HCHO, dominates the total uncertainty in the
539   TCOH product but is of a magnitude comparable to that of in situ OH observations. Median TCOH 1σ
540   uncertainty resulting from NO_2 is 16.5%, with maxima in the remote atmosphere in regions where NO_2
541   columns are low. Median uncertainty in TCOH resulting from HCHO is 7%, averaged over the study
542   domain, despite the large uncertainty in the HCHO retrieval itself. In contrast to NO_2, uncertainties
543   maximize in regions with higher HCHO columns (Fig. 6), although this is likely an overestimate as the
544   retrieval uncertainty for these columns is lower than in the more remote atmosphere. In comparison,

545 median TCOH uncertainties resulting from other inputs are 2.9% or less (Figs. S9 and S10). Total TCOH
546 uncertainty is 16.6% and is dominated by the NO$_2$ uncertainty.
547

548
549 *Figure 6:* Normalized 1σ uncertainty in the satellite TCOH product due to uncertainties in the HCHO (a) and NO2
550 (b) retrievals. The combined uncertainty from all input species is shown in panel c.

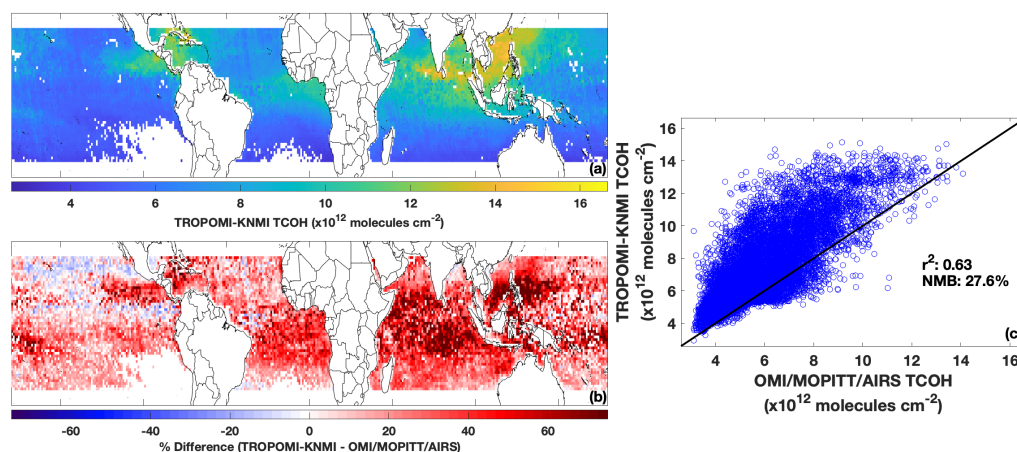
551 These results demonstrate that the satellite retrieval inputs to the machine learning model are of
552 sufficient quality to produce a meaningful TCOH data product when averaged over large spatial and
553 temporal scales. The 2σ uncertainty in TCOH resulting from the uncertainties in these retrievals is on
554 the order of that reported for in situ OH observations (Brune et al., 2020). As discussed earlier, this is
555 also likely an upper bound on the uncertainty from random retrieval errors, and uncertainties could be
556 reduced through further averaging, although at the expense of reduced spatial and temporal resolution.
557 Improving the satellite retrievals of NO$_2$ and HCHO in the remote atmosphere, using retrievals with less
558 noise over the remote atmosphere such as HCHO from the Ozone Mapping and Profiler Suite (OMPS)
559 (González Abad et al., 2016), or incorporating data from satellites with higher resolution, such as
560 TROPOMI (Tropospheric Monitoring Instrument), could also reduce the uncertainty in their retrievals
561 and thus in TCOH. As discussed in the next section, however, systematic biases between satellite
562 retrievals can also lead to uncertainties in the TCOH.
563
564 **5.3   Sensitivity of TCOH to different satellite retrievals of GBRT inputs**
565 The satellite retrievals listed in Table 1 provide the benefit of a long record, with data from most
566 retrievals available from at least 2005 to the present. Such a rich dataset would allow for long-term
567 trend analysis of TCOH. These instruments are near the end of their life cycle, however, so it is
568 instructive to see how retrievals from newer satellites impact the predicted TCOH from the GBRT model.
569 In addition, although these newer satellites, such as TROPOMI, have a significantly shorter observational
570 record than those in Table 1, TROPOMI also has finer spatial resolution and the added advantage of
571 providing retrievals for CO, NO$_2$, O$_3$, HCHO, and H$_2$O$_{(v)}$. Using retrievals of multiple species from the
572 same instrument could negate errors resulting from differences in viewing geometry as well as from
573 overpass time. Here, we investigate the effects of applying retrievals from TROPOMI to the machine
574 learning model and compare them to the results from the product described in Section 4, highlighting
575 potential impacts resulting from instrumental differences as well as those resulting from differences in
576 retrieval algorithms. The results emphasize the need for thorough retrieval validation in the remote
577 atmosphere, particularly of NO$_2$.
578

5.3.1   Description of TROPOMI

579
580   TROPOMI, a successor instrument to OMI, is a spectrometer covering portions of the ultraviolet, visible,
581   and infrared spectrum (Veefkind et al., 2012). It is located onboard the Sentinel 5 Precursor satellite,
582   which is polar orbiting and has a local overpass time of approximately 13:30. Horizontal resolution for
583   the month examined here (May 2018) is 7 km × 3.5 km at nadir. We have gridded the Level 2 product
584   for each species to a 1° × 1° resolution and averaged the data to the monthly scale, applying the
585   recommended quality flags and filtering for cloud fraction greater than 30%.
586
587   We use two different retrievals of TROPOMI NO_2 for this analysis. First, we use the KNMI (Royal
588   Netherlands Meteorological Institute) NO_2 retrieval (van Geffen et al., 2020), which is based on the
589   DOMINO (Dutch OMI NO_2 product) retrieval developed for the OMI instrument. Wang et al. (2020)
590   found that this retrieval was biased high when compared to ship-based observations from a MAX-DOAS
591   instrument over the remote oceans. In addition, we use the MINDS (Multi-Decadal Nitrogen Dioxide
592   and Derived Products from Satellites) retrieval, which uses the same algorithm as for the OMI product
593   described in Section 2 (Lamsal et al., 2022). This retrieval has not been evaluated in the remote tropics.
594
595   We also use TROPOMI retrievals of HCHO, $H_2O_{(v)}$, total column O_3, and CO. The HCHO retrieval (De
596   Smedt et al., 2018) was found to have a 30% low bias with respect to an OMI retrieval using the same
597   algorithm due to differences in cloud processing (De Smedt et al., 2021). While evaluation in the remote
598   tropics is limited, the TROPOMI retrieval does overestimate HCHO in polluted regions (De Smedt et al.,
599   2021) when compared to ground-based observations. The TROPOMI $H_2O_{(v)}$ (Chan et al., 2022) retrieval
600   has a slight dry bias with comparison to other satellite products, while the total column O_3 retrieval
601   (Garane et al., 2019) agrees within $0 – 1.5\%$ with ground-based observations. Finally, the CO retrieval
602   (Borsdorff et al., 2019) agrees with MOPITT over the oceans within 3% on average (Martínez-Alonso et
603   al., 2020). TROPOMI does not have an equivalent retrieval of the AIRS $H_2O_{(v)}$ layers.
604

5.3.2   TROPOMI data applied to the GBRT model

606   First, we calculate TCOH for May 2018 using inputs from TROPOMI. We trained a separate machine
607   learning model using all inputs from Table 1 except the water vapor layers, for which there are no
608   TROPOMI retrievals. Removal of the layers from the machine learning model does not significantly
609   degrade performance. For example, for May 2017, removing the $H_2O_{(v)}$ layers from the model, increases
610   the NRMSE from 5.34% to 5.73% when applying the GBRT model to the hold out set. For this new
611   model, we then calculate TCOH using TROPOMI data, including the KNMI NO_2 retrieval. For SSTs and
612   AOD, we use the MUR and MODIS products respectively. While TROPOMI does have an aerosol
613   product, the UV aerosol index, the corresponding output from the MERRA2 GMI simulation is
614   unavailable. We refer to this as the TROPOMI-KNMI product. We have also calculated TCOH using the
615   satellite retrievals in Table 1, except for the water vapor layers, using this GBRT model, and refer to that
616   as the OMI/MOPITT/AIRS product.
617
618   TCOH from the TROPOMI-KNMI product are biased high with respect to those from the
619   OMI/MOPITT/AIRS product. Figure 7 shows TCOH calculated from the TROPOMI-KNMI product as well
620   as the percent difference between the two products. While there is modest correlation between the
621   two ($r^2 = 0.63$), the TROPOMI product is 27.6% higher than the multi-satellite product, with
622   overestimates across almost the entire domain. Differences between the products are most
623   pronounced in the Indian Ocean and off the coasts of Indonesia and the Philippines.
624
625   In general, observations from TROPOMI agree with those from the satellites in Table 1, with the
626   exception of NO_2 and HCHO. Ozone, $H_2O_{(v)}$, and CO from TROPOMI are highly correlated (r^2 of 0.85 or

627 higher) with their respective retrievals from the OMI/MOPITT/AIRS product, with a modest high bias
628 (less than 10%) in each case (Fig. S11). Note that we do not seek to identify the more accurate retrievals
629 here; we are reporting the bias with respect to the satellite products listed in Table 1 because those
630 retrievals underlie the TCOH product. TROPOMI NO_2 and HCHO, on the other hand, are systematically
631 biased with respect to their respective OMI products, with a distinct high bias in KNMI-NO_2 (145%) and
632 low bias in HCHO (20%). The higher TCOH from the TROPOMI product is consistent with the increase in
633 NO_2, which would lead to higher secondary production of OH. Further, while TROPOMI KNMI-NO_2 is
634 modestly correlated with OMI NO_2 (r^2 = 0.61), TROPOMI and OMI HCHO are not correlated (r^2 = 0.23),
635 highlighting the difficulty of the HCHO retrieval.
636

637
638 *Figure 7:* TCOH for May 2018 determined using TROPOMI inputs, including the KNMI NO_2 retrieval (a). The
639 difference between the TROPOMI and multi-satellite product is shown in (b). Panel (c) shows the regression of
640 TCOH calculated from TROPOMI against that calculated from retrievals from MOPITT, OMI, and AIRS.

641 NO_2 drives the differences between the two TCOH products. To determine the impacts of the different
642 TROPOMI inputs on the TCOH product, we individually swapped each TROPOMI input into the
643 OMI/MOPITT/AIRS product, replacing the corresponding input from Table 1. We then determined the
644 difference in TCOH from the OMI/MOPITT/AIRS product that does not include TROPOMI. While this
645 method will not yield the exact contribution from a particular retrieval because of the non-linear nature
646 of OH chemistry, it does yield information about the relative importance of each species. Swapping in
647 TROPOMI CO, $H_2O_{(v)}$, and O_3 changed TCOH by less than 2%, while using TROPOMI HCHO increased
648 TCOH by 3%. In contrast, TROPOMI NO_2 increased TCOH by 29%, showing that the high bias in the
649 TROPOMI TCOH product is dominated by differences in NO_2.
650
651 The increased TCOH in the TROPOMI product likely results from a combination of differences in the NO_2
652 retrieval algorithm as well as instrumental differences. Comparison of the KNMI and MINDS retrievals
653 illustrate this point. When compared to OMI, the MINDS NO_2 retrieval has a NMB of 58% for May 2018,
654 as compared to a NMB of 145% for the KNMI retrieval. The closer agreement is unsurprising since the
655 MINDS NO_2 uses the same retrieval algorithm as for OMI. Substituting the MINDS NO_2 as an input to the
656 TROPOMI product (TROPOMI-MINDS product) reduces the bias with respect to the OMI/MOPITT/AIRS
657 product to 18% (Fig. S12). While this is an improvement in agreement, the differences in TCOH as well
658 as the lack of change in r^2 value still suggest that differences between OMI and TROPOMI unrelated to
659 the retrieval algorithm account for some of the discrepancy.

660
661   The results here demonstrate the sensitivity of the methodology to any systematic bias in the input
662   retrievals. As with the random error analysis, the level of uncertainty introduced by these biases is low
663   enough to allow for a meaningful OH product. Despite these differences, the methodology to determine
664   TCOH using machine learning that we have presented here still captures the variability in TCOH,
665   consistent with the ATom evaluation outlined in Section 5.1. To reduce the uncertainty of TCOH, better
666   evaluation of NO_2 in the remote atmosphere is needed to determine which retrievals, if any, are
667   accurate.
668
669   **6   Discussion and recommendations for future observations**
670   The method of estimating clear-sky TCOH presented here has the potential to increase our
671   understanding of the atmospheric oxidation capacity. Because of the long record of observations from
672   MOPITT, OMI, AIRS, and MODIS, we can calculate tropical TCOH from 2005 to the present, and since the
673   methodology is not constrained to a particular satellite, newer satellite missions could extend the
674   dataset beyond the end of these instruments' lifetimes. In addition, this methodology will provide sub-
675   hemispheric information on OH variability, supplementing information available from MCF inversions.
676
677   The methodology could be expanded to the extra-tropics and over land, allowing for global constraints
678   on OH. This will likely require additional satellite retrievals, like that of isoprene (Wells et al., 2020), in
679   regions with more complex VOC chemistry than in the remote atmosphere. Expanding this product
680   beyond the tropics could increase understanding of global CH_4, CO, and VOC trends and variability and
681   allow for a wider range of satellite retrievals as inputs. For example, geostationary air quality satellites
682   such as TEMPO (Tropospheric Emissions: Monitoring of Pollution), scheduled to be launched in 2023,
683   will provide retrievals of most of the necessary inputs to the machine learning model, allowing for the
684   understanding of diurnal variability in TCOH and potentially in the diurnal variability of ozone production
685   (Zhu et al., 2022b).
686
687   A similar methodology could likely be used to determine OH at different layers of the atmosphere.
688   Because CH_4 loss is not evenly distributed throughout the tropospheric column, vertically resolved OH
689   would better help inform this process. While column inputs, such as those discussed here, could be
690   used, the inclusion of vertically resolved satellite retrievals, such as the AIRS $H_2O_{(v)}$ layers, would provide
691   additional information. Tropospheric O_3 at different atmospheric layers, such as that previously
692   provided by the TES satellite, could also be invaluable here, as O_3 is a large driver of primary OH
693   production.
694
695   Satellite-derived OH would also a provide a much-needed, observational constraint on OH variability in
696   global chemistry models. Because the methodology can capture variability in TCOH of both
697   observations and 3-dimensional model output, TCOH trends from a satellite-constrained product could
698   be used to evaluate modeled trends and as well as the spatial variability resulting from events like ENSO.
699   While the satellite-derived OH could not explicitly indicate the cause of differences, the spatial
700   distribution of the differences as well as differences in observed and modeled machine learning model
701   inputs could indicate potential dynamical or emission sources of error in the 3D model.
702
703   Further, the combination of the satellite-derived OH and the machine learning model could help identify
704   the impacts of any diagnosed errors in emissions inventories as well as the impacts of unexpected
705   events, such as COVID-19-related shutdowns, on TCOH. For example, if there are significant
706   discrepancies between observed and modeled NO_2 in a specific region of the atmosphere, the satellite
707   NO_2 could be scaled to more closely match the 3D model values and then be input into the machine

708 learning model. The difference in TCOH would then indicate the relative impact of the model error,
709 serving as a computationally efficient way to identify the impacts of these errors on the atmospheric
710 oxidation capacity. A similar methodology could be used for unexpected events that significantly impact
711 emissions of OH drivers, allowing for quick determination of their potential impacts on the atmospheric
712 oxidation capacity before emissions inventories could be revised.
713
714 While we have shown that the methodology captures the variability of OH, it is unclear whether a bias in
715 magnitude when compared to observations of OH results from GBRT model deficiencies or structural
716 differences between the *in situ* observations and the training dataset. Additional field campaigns with
717 observations of OH and the GBRT model inputs would allow for a more thorough evaluation of both the
718 OH product and the methodology itself. Such a field campaign would need to provide complete
719 tropospheric columns of all species and cover less horizontal distance than the ATom profiles (e.g. from
720 spiral flight patterns). In situ observations of NO_2 without significant interference from NO_x reservoir
721 species are also needed to reduce uncertainty. Alternatively, NO_2 and other species could be measured
722 through aircraft-based remote sensing. Finally, repeated sampling over the same locations for multiple
723 days within a defined area would allow for meaningful statistical analysis while also allowing for the
724 comparison of TCOH columns calculated from satellite observations.
725
726 Finally, accuracy of the TCOH product is dependent on the accuracy of the satellite retrievals input into
727 the machine learning model, with the NO_2 retrieval having the largest effect. To reduce the uncertainty
728 of the TCOH product, more information about the accuracy of individual NO_2 retrievals is required.
729 Currently, there is little validation of OMI and TROPOMI NO_2 retrievals in the remote, tropical
730 atmosphere, so it is difficult to assess, which retrievals, if any, are correct. Even retrievals of TROPOMI
731 and OMI made with the same algorithm show differences, suggesting that instrumental differences
732 could also affect the results. Future satellite missions should focus on trying to reduce the uncertainty
733 in NO_2 retrievals, both through improvements in instrument design and algorithm development.
734

7 Data Availability

736 Output from the MERRA2 GMI simulation are publicly available at https://acd-
737 ext.gsfc.nasa.gov/Projects/GEOSCCM/MERRA2GMI/ (NASA Goddard Space Flight Center, 2022). All
738 satellite products, except for TROPOMI water vapor, are available at https://disc.gsfc.nasa.gov (GES
739 DISC, 2022). Data from the ATom campaign are located at https://daac.ornl.gov (Wofsy et al., 2021).
740

8 Author contributions

742 DCA wrote the manuscript, performed the data analysis, and created the GBRT model. DCA, BND, JMN,
743 and MBFC developed the idea for the methodology. SAS performed three-dimensional modeling for the
744 work. JMN provided advice on machine learning. JL helped perform data analysis. All authors helped
745 develop ideas for the analysis and contributed to the manuscript.
746

9 Competing Interests

748 BND is a member of the editorial board of Atmospheric Chemistry and Physics. The peer-review process
749 was guided by an independent editor, and the authors also have no other competing interests to
750 declare.
751

10 Financial support

753 This research has been supported by the National Aeronautics and Space Administration (NASA)
754 Atmospheric Composition Campaign Data Analysis and Modeling (ACCDAM) program (grant no.
755 80NSSC21K1440).

11 Acknowledgements

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

12 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, https://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, https://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, https://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.

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, https://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, e2019JD031685, https://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, https://10.5194/amt-10-1927-2017, 2017.

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

799   Chan, K. L., Xu, J., Slijkhuis, S., Valks, P., and Loyola, D.: TROPOspheric Monitoring Instrument
800   observations of total column water vapour: Algorithm and validation, Science of The Total Environment,
801   821, 153232, https://doi.org/10.1016/j.scitotenv.2022.153232, 2022.

802   Chen, T., and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, KDD '16: Proceedings of the 22nd
803   ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13 - 17 Aug. 2016,
804   785-794, San Francsisco, CA, USA, https://doi.org/10.1145/2939672.2939785, 2016.

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

808   Choi, S., Lamsal, L. N., Follette-Cook, M., Joiner, J., Krotkov, N. A., Swartz, W. H., Pickering, K. E.,
809   Loughner, C. P., Appel, W., Pfister, G., Saide, P. E., Cohen, R. C., Weinheimer, A. J., and Herman, J. R.:
810   Assessment of NO_2 observations during DISCOVER-AQ and KORUS-AQ field campaigns, Atmos. Meas.
811   Tech., 13, 2523-2546, https://10.5194/amt-13-2523-2020, 2020.

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

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

823   Deeter, M. N., Edwards, D. P., Francis, G. L., Gille, J. C., Mao, D., Martínez-Alonso, S., Worden, H. M.,
824   Ziskin, D., and Andreae, M. O.: Radiance-based retrieval bias mitigation for the MOPITT instrument: the
825   version 8 product, Atmospheric Measurement Techniques, 12, 4561-4580, https://10.5194/amt-12-
826   4561-2019, 2019.

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

830   Duncan, B. N., Strahan, S. E., Yoshida, Y., Steenrod, S. D., and Livesey, N.: Model study of the cross-
831   tropopause transport of biomass burning pollution, Atmos. Chem. Phys., 7, 3713-3736, 10.5194/acp-7-
832   3713-2007, 2007.

833   Elith, J., Leathwick, J. R., and Hastie, T.: A working guide to boosted regression trees, J Anim Ecol, 77,
834   802-813, https://10.1111/j.1365-2656.2008.01390.x, 2008.

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

841   Gelaro, R., McCarty, W., Suarez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C., Darmenov, A.,
842   Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty,
843   A., da Silva, A., Gu, W., Kim, G. K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G.,
844   Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era
845   Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), J Clim, Volume 30, 5419-
846   5454, https://10.1175/JCLI-D-16-0758.1, 2017.

847   GES DISC: Earth Science Data at NASA, available at https://disc.gsfc.nasa.gov, last access: 1 Nov. 2022.

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

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

854   Hedelius, J. K., He, T. L., Jones, D. B. A., Baier, B. C., Buchholz, R. R., De Mazière, M., Deutscher, N. M.,
855   Dubey, M. K., Feist, D. G., Griffith, D. W. T., Hase, F., Iraci, L. T., Jeseck, P., Kiel, M., Kivi, R., Liu, C.,
856   Morino, I., Notholt, J., Oh, Y. S., Ohyama, H., Pollard, D. F., Rettinger, M., Roche, S., Roehl, C. M.,
857   Schneider, M., Shiomi, K., Strong, K., Sussmann, R., Sweeney, C., Té, Y., Uchino, O., Velazco, V. A., Wang,
858   W., Warneke, T., Wennberg, P. O., Worden, H. M., and Wunch, D.: Evaluation of MOPITT Version 7 joint
859   TIR–NIR XCO retrievals with TCCON, Atmos. Meas. Tech., 12, 5547-5572, https://10.5194/amt-12-5547-
860   2019, 2019.

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

864   Keller, C. A., and Evans, M. J.: Application of random forest regression to the calculation of gas-phase
865   chemistry within the GEOS-Chem chemistry model v10, Geoscientific Model Development, 12, 1209-
866   1225, https://10.5194/gmd-12-1209-2019, 2019.

867   Kelp, M. M., Jacob, D. J., Kutz, J. N., Marshall, J. D., and Tessum, C. W.: Toward Stable, General Machine-
868   Learned Models of the Atmospheric Chemical System, Journal of Geophysical Research: Atmospheres,
869   125, e2020JD032759, https://doi.org/10.1029/2020JD032759, 2020.

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

873 Lamsal, L. N., Krotkov, N. A., Celarier, E. A., Swartz, W. H., Pickering, K. E., Bucsela, E. J., Gleason, J. F.,
874 Martin, R. V., Philip, S., Irie, H., Cede, A., Herman, J., Weinheimer, A., Szykman, J. J., and Knepp, T. N.:
875 Evaluation of OMI operational standard NO_2 column retrievals using in situ and surface-based NO_2
876 observations, Atmospheric Chemistry and Physics, 14, 11587-11609, https://10.5194/acp-14-11587-
877 2014, 2014.

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

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

886 Laughner, J. L., Neu, J. L., Schimel, D., Wennberg, P. O., Barsanti, K., Bowman, K. W., Chatterjee, A.,
887 Croes, B. E., Fitzmaurice, H. L., Henze, D. K., Kim, J., Kort, E. A., Liu, Z., Miyazaki, K., Turner, A. J.,
888 Anenberg, S., Avise, J., Cao, H., Crisp, D., de Gouw, J., Eldering, A., Fyfe, J. C., Goldberg, D. L., Gurney, K.
889 R., Hasheminassab, S., Hopkins, F., Ivey, C. E., Jones, D. B. A., Liu, J., Lovenduski, N. S., Martin, R. V.,
890 McKinley, G. A., Ott, L., Poulter, B., Ru, M., Sander, S. P., Swart, N., Yung, Y. L., and Zeng, Z. C.: Societal
891 shifts due to COVID-19 reveal large-scale complexities and feedbacks between atmospheric chemistry
892 and climate change, Proc Natl Acad Sci U S A, 118, e2109481118, https://10.1073/pnas.2109481118,
893 2021.

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

897 Liang, Q., Chipperfield, M. P., Fleming, E. L., Abraham, N. L., Braesicke, P., Burkholder, J. B., Daniel, J. S.,
898 Dhomse, S., Fraser, P. J., Hardiman, S. C., Jackman, C. H., Kinnison, D. E., Krummel, P. B., Montzka, S. A.,
899 Morgenstern, O., McCulloch, A., Mühle, J., Newman, P. A., Orkin, V. L., Pitari, G., Prinn, R. G., Rigby, M.,
900 Rozanov, E., Stenke, A., Tummon, F., Velders, G. J. M., Visioni, D., and Weiss, R. F.: Deriving Global OH
901 Abundance and Atmospheric Lifetimes for Long-Lived Gases: A Search for CH_3CCl_3 Alternatives, Journal
902 of Geophysical Research: Atmospheres, 122, 11,914-11,933, https://10.1002/2017JD026926, 2017.

903 Lovelock, J. E.: Methyl chloroform in the troposphere as an indicator of OH radical abundance, Nature,
904 267, 32-32, https://10.1038/267032a0, 1977.

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

908 Martínez-Alonso, S., Deeter, M., Worden, H., Borsdorff, T., Aben, I., Commane, R., Daube, B., Francis, G.,
909 George, M., Landgraf, J., Mao, D., McKain, K., and Wofsy, S.: 1.5 years of TROPOMI CO measurements:
910 comparisons to MOPITT and ATom, Atmos. Meas. Tech., 13, 4841-4864, https://10.5194/amt-13-4841-
911 2020, 2020.

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

914   Miller, D. O., and Brune, W. H.: Investigating the Understanding of Oxidation Chemistry Using 20 Years
915   of Airborne OH and HO_2 Observations, Journal of Geophysical Research: Atmospheres, 127,
916   e2021JD035368, https://doi.org/10.1029/2021JD035368, 2022.

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

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

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

925   NASA Goddard Space Flight Center. MERRA2 GMI, NASA [data set], https://acd-
926   ext.gsfc.nasa.gov/Projects/GEOSCCM/MERRA2GMI/, last axccess: 1 November 2022.

927   Naus, S., Montzka, S. A., Pandey, S., Basu, S., Dlugokencky, E. J., and Krol, M.: Constraints and biases in a
928   tropospheric two-box model of OH, Atmos. Chem. Phys., 19, 407-424, https://10.5194/acp-19-407-2019,
929   2019.

930   Nicely, J. M., Anderson, D. C., Canty, T. P., Salawitch, R. J., Wolfe, G. M., Apel, E. C., Arnold, S. R., Atlas, E.
931   L., Blake, N. J., Bresch, J. F., Campos, T. L., Fernandez, R. P., Flemming, J., Hall, S. R., Hanisco, T. F.,
932   Honomichl, S. B., Hornbrook, R. S., Huijnen, V., Kaser, L., Kinnison, D. E., Pan, L. L., Riemer, D. D., Saiz-
933   Lopez, A., Steenrod, S. D., Stell, M. H., Tilmes, S., Turquety, S., Ullmann, K., and Weinheimer, A. J.: An
934   observationally constrained evaluation of the oxidative capacity in the tropical western Pacific
935   troposphere, Journal of Geophysical Research: Atmospheres, 121, 7461-7488,
936   https://10.1002/2016JD025067, 2016.

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

941   Nicely, J. M., Duncan, B. N., Hanisco, T. F., Wolfe, G. M., Salawitch, R. J., Deushi, M., Haslerud, A. S.,
942   Jöckel, P., Josse, B., Kinnison, D. E., Klekociuk, A., Manyin, M. E., Marécal, V., Morgenstern, O., Murray, L.
943   T., Myhre, G., Oman, L. D., Pitari, G., Pozzer, A., Quaglia, I., Revell, L. E., Rozanov, E., Stenke, A., Stone, K.,
944   Strahan, S., Tilmes, S., Tost, H., Westervelt, D. M., and Zeng, G.: A machine learning examination of
945   hydroxyl radical differences among model simulations for CCMI-1, Atmospheric Chemistry and Physics,
946   20, 1341-1361, https://10.5194/acp-20-1341-2020, 2020.

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

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

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

957 Pickett, H. M., Drouin, B. J., Canty, T., Salawitch, R. J., Fuller, R. A., Perun, V. S., Livesey, N. J., Waters, J.
958 W., Stachnik, R. A., Sander, S. P., Traub, W. A., Jucks, K. W., and Minschwaner, K.: Validation of Aura
959 Microwave Limb Sounder OH and HO_2 measurements, Journal of Geophysical Research: Atmospheres,
960 113, D16S30, https://doi.org/10.1029/2007JD008775, 2008.

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

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

969 Shah, V., Jacob, D. J., Dang, R., Lamsal, L. N., Strode, S. A., Steenrod, S. D., et al. (2022). Nitrogen oxides
970 in the free troposphere: Implications for tropospheric oxidants and the interpretation of satellite NO2
971 measurements. *EGUsphere, 2022*, 1-47. https://egusphere.copernicus.org/preprints/egusphere-2022-
972 656/.

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

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

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

983 Strode, S. A., Ziemke, J. R., Oman, L. D., Lamsal, L. N., Olsen, M. A., and Liu, J.: Global changes in the
984 diurnal cycle of surface ozone, Atmospheric Environment, 199, 323-333,
985 https://10.1016/j.atmosenv.2018.11.028, 2019.

986 Susskind, J., Blaisdell, J. M., and Iredell, L.: Improved methodology for surface and atmospheric
987 soundings, error estimates, and quality control procedures: the atmospheric infrared sounder science

988   team version-6 retrieval algorithm, Journal of Applied Remote Sensing, 8, 084994,
989   https://10.1117/1.Jrs.8.084994, 2014.

990   Thompson, C. R., Wofsy, S. C., Prather, M. J., Newman, P. A., Hanisco, T. F., Ryerson, T. B., Fahey, D. W.,
991   Apel, E. C., Brock, C. A., Brune, W. H., Froyd, K., Katich, J. M., Nicely, J. M., Peischl, J., Ray, E., Veres, P. R.,
992   Wang, S., Allen, H. M., Asher, E., Bian, H., Blake, D., Bourgeois, I., Budney, J., Bui, T. P., Butler, A.,
993   Campuzano-Jost, P., Chang, C., Chin, M., Commane, R., Correa, G., Crounse, J. D., Daube, B., Dibb, J. E.,
994   DiGangi, J. P., Diskin, G. S., Dollner, M., Elkins, J. W., Fiore, A. M., Flynn, C. M., Guo, H., Hall, S. R.,
995   Hannun, R. A., Hills, A., Hintsa, E. J., Hodzic, A., Hornbrook, R. S., Huey, L. G., Jimenez, J. L., Keeling, R. F.,
996   Kim, M. J., Kupc, A., Lacey, F., Lait, L. R., Lamarque, J.-F., Liu, J., McKain, K., Meinardi, S., Miller, D. O.,
997   Montzka, S. A., Moore, F. L., Morgan, E. J., Murphy, D. M., Murray, L. T., Nault, B. A., Neuman, J. A.,
998   Nguyen, L., Gonzalez, Y., Rollins, A., Rosenlof, K., Sargent, M., Schill, G., Schwarz, J. P., Clair, J. M. S.,
999   Steenrod, S. D., Stephens, B. B., Strahan, S. E., Strode, S. A., Sweeney, C., Thames, A. B., Ullmann, K.,
1000 Wagner, N., Weber, R., Weinzierl, B., Wennberg, P. O., Williamson, C. J., Wolfe, G. M., and Zeng, L.: The
1001 NASA Atmospheric Tomography (ATom) Mission: Imaging the Chemistry of the Global Atmosphere,
1002 Bulletin of the American Meteorological Society, 103, E761-E790, https://10.1175/BAMS-D-20-0315.1,
1003 2022.

1004 Turner, A. J., Frankenberg, C., Wennberg, P. O., and Jacob, D. J.: Ambiguity in the causes for decadal
1005 trends in atmospheric methane and hydroxyl, Proceedings of the National Academy of Sciences, 114,
1006 5367, https://10.1073/pnas.1616020114, 2017.

1007 Turner, A. J., Fung, I., Naik, V., Horowitz, L. W., and Cohen, R. C.: Modulation of hydroxyl variability by
1008 ENSO in the absence of external forcing, Proc Natl Acad Sci U S A, 115, 8931-8936,
1009 https://10.1073/pnas.1807532115, 2018.

1010 van Geffen, J., Boersma, K. F., Eskes, H., Sneep, M., ter Linden, M., Zara, M., and Veefkind, J. P.: S5P
1011 TROPOMI NO2 slant column retrieval: method, stability, uncertainties and comparisons with OMI,
1012 Atmos. Meas. Tech., 13, 1315-1335, https://10.5194/amt-13-1315-2020, 2020.

1013 Veefkind, J. P., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H. J., de Haan, J.
1014 F., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P.,
1015 Voors, R., Kruizinga, B., Vink, R., Visser, H., and Levelt, P. F.: TROPOMI on the ESA Sentinel-5 Precursor: A
1016 GMES mission for global observations of the atmospheric composition for climate, air quality and ozone
1017 layer applications, Remote Sensing of Environment, 120, 70-83,
1018 https://doi.org/10.1016/j.rse.2011.09.027, 2012.

1019 Voulgarakis, A., Naik, V., Lamarque, J. F., Shindell, D. T., Young, P. J., Prather, M. J., Wild, O., Field, R. D.,
1020 Bergmann, D., Cameron-Smith, P., Cionni, I., Collins, W. J., Dalsøren, S. B., Doherty, R. M., Eyring, V.,
1021 Faluvegi, G., Folberth, G. A., Horowitz, L. W., Josse, B., MacKenzie, I. A., Nagashima, T., Plummer, D. A.,
1022 Righi, M., Rumbold, S. T., Stevenson, D. S., Strode, S. A., Sudo, K., Szopa, S., and Zeng, G.: Analysis of
1023 present day and future OH and methane lifetime in the ACCMIP simulations, Atmospheric Chemistry and
1024 Physics, 13, 2563-2587, https://10.5194/acp-13-2563-2013, 2013.

1025 Wang, P., Piters, A., van Geffen, J., Tuinder, O., Stammes, P., and Kinne, S.: Shipborne MAX-DOAS
1026 measurements for validation of TROPOMI NO2 products, Atmos. Meas. Tech., 13, 1413-1426,
1027 https://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, https://10.1038/s41586-020-2664-3, 2020.

Wofsy, S. C., Afshar, S., Allen, H. M., Apel, E. C., Asher, E. C., Barletta, B., Bent, J., Bian, H., Biggs, B. C., Blake, D. R., Blake, N., Bourgeois, I., Brock, C. A., Brune, W. H., Budney, J. W., Bui, T. P., Butler, A., Campuzano-Jost, P., Chang, C. S., Chin, M., Commane, R., Correa, G., Crounse, J. D., Cullis, P. D., Daube, B. C., Day, D. A., Dean-Day, J. M., Dibb, J. E., DiGangi, J. P., Diskin, G. S., Dollner, M., Elkins, J. W., Erdesz, F., Fiore, A. M., Flynn, C. M., Froyd, K. D., Gesler, D. W., Hall, S. R., Hanisco, T. F., Hannun, R. A., Hills, A. J., Hintsa, E. J., Hoffman, A., Hornbrook, R. S., Huey, L. G., Hughes, S., Jimenez, J. L., Johnson, B. J., Katich, J. M., Keeling, R. F., Kim, M. J., Kupc, A., Lait, L. R., McKain, K., McLaughlin, R. J., 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., Newman, P. A., Nicely, J. M., Pan, X., Paplawsky, W., Peischl, J., Prather, M. J., Price, D. J., Ray, E. A., Reeves, J. M., Richardson, M., Rollins, A. W., Rosenlof, K. H., Ryerson, T. B., Scheuer, E., Schill, G. P., Schroder, J. C., Schwarz, J. P., St.Clair, J. M., Steenrod, S. D., Stephens, B. B., Strode, S. A., Sweeney, C., Tanner, D., Teng, A. P., Thames, A. B., Thompson, C. R., Ullmann, K., Veres, P. R., Wagner, N. L., Watt, A., Weber, R., Weinzierl, B. B., Wennberg, P. O., Williamson, C. J., Wilson, J. C., Wolfe, G. M., Woods, C. T., Zeng, L. H., and Vieznor, N.: ATom: Merged Atmospheric Chemistry, Trace Gases, and Aerosols, Version 2, in, ORNL Distributed Active Archive Center, 2021. https://10.3334/ORNLDAAC/1925.

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 USA, 116, 11171-11180, https://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, https://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 SEACRS 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.: 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, https://10.1021/acs.est.1c05636, 2022a.

1071

1072   Zhu, Q., Laughner, J. L., and Cohen, R. C.: Estimate of OH trends over one decade in North American
1073   cities, Proc Natl Acad Sci U S A, 119, e2117399119, https://doi.org/10.1073/pnas.2117399119, 2022b.

1074

1075