

Supplementary Information for Regional oxidation in the remote troposphere explained by an observation-based reduced-form model

Colleen B. Baublitz, Arlene M. Fiore, Sarah M. Ludwig, Julie M. Nicely, Glenn M. Wolfe, Lee T. Murray, Róisín Commane, Michael J. Prather, Daniel C. Anderson, Gustavo Correa, Bryan N. Duncan, Melanie Follette-Cook, Daniel M. Westervelt, Ilann Bourgeois, William H. Brune, T. Paul Bui, Joshua P. DiGangi, Glenn S. Diskin, Samuel R. Hall, Kathryn McKain, David O. Miller, Jeff Peischl, Alexander Thames, Chelsea Thompson, Kirk Ullmann, Steven C. Wofsy

*Colleen B. Baublitz
Email: cbb2158@columbia.edu

This PDF file includes:

- Supplementary text
- Figures S1 to S18
- Tables S1 to S7
- SI References

Supplementary Information Text

Text S1. Assessing the role of spatial variations in the $\text{Proxy}_{\text{OH}}[\text{OH}]$ slope and intercept distributions

We assess the extent to which the 2 km altitude by 20° latitude bin-specific regression parameters are statistically unique from the other bins in the same latitude, altitude or oceanic basin, respectively, using the distribution of parameters from the Bayesian regression with Markov Chain Monte Carlo simulations (Methods 8.2). We count instances when the bin-specific 75th percentile is lower than the 25th percentile of the other bins, or when the bin-specific 25th percentile is higher than the 75th percentile of the other bins. We then sum across the probability distribution for each dimension and divide by the total number of bins (60) to calculate the mean probability.

Applying Proxy_{OH} to infer $[\text{OH}]$ spatial variations will require accounting for vertical and latitudinal patterns in the $\text{Proxy}_{\text{OH}}[\text{OH}]$ relationship because the slopes and intercepts vary more by altitude and latitude than by ocean basin (Figure 2, S18). The intercept maximizes while the slope minimizes in the lower tropical troposphere. This could potentially result from chemistry downwind of biogenic or fire emissions that leads the Proxy_{OH} to underestimate OH loss to heterogeneous chemistry or reactive carbon oxidation. Toward the poles, $\text{HO}_2 + \text{O}_3$ is relatively more influential over total OH production (1, 2), while it could be more important to include peroxide or acetone photolysis in a proxy formulation for the upper troposphere (3, 4).

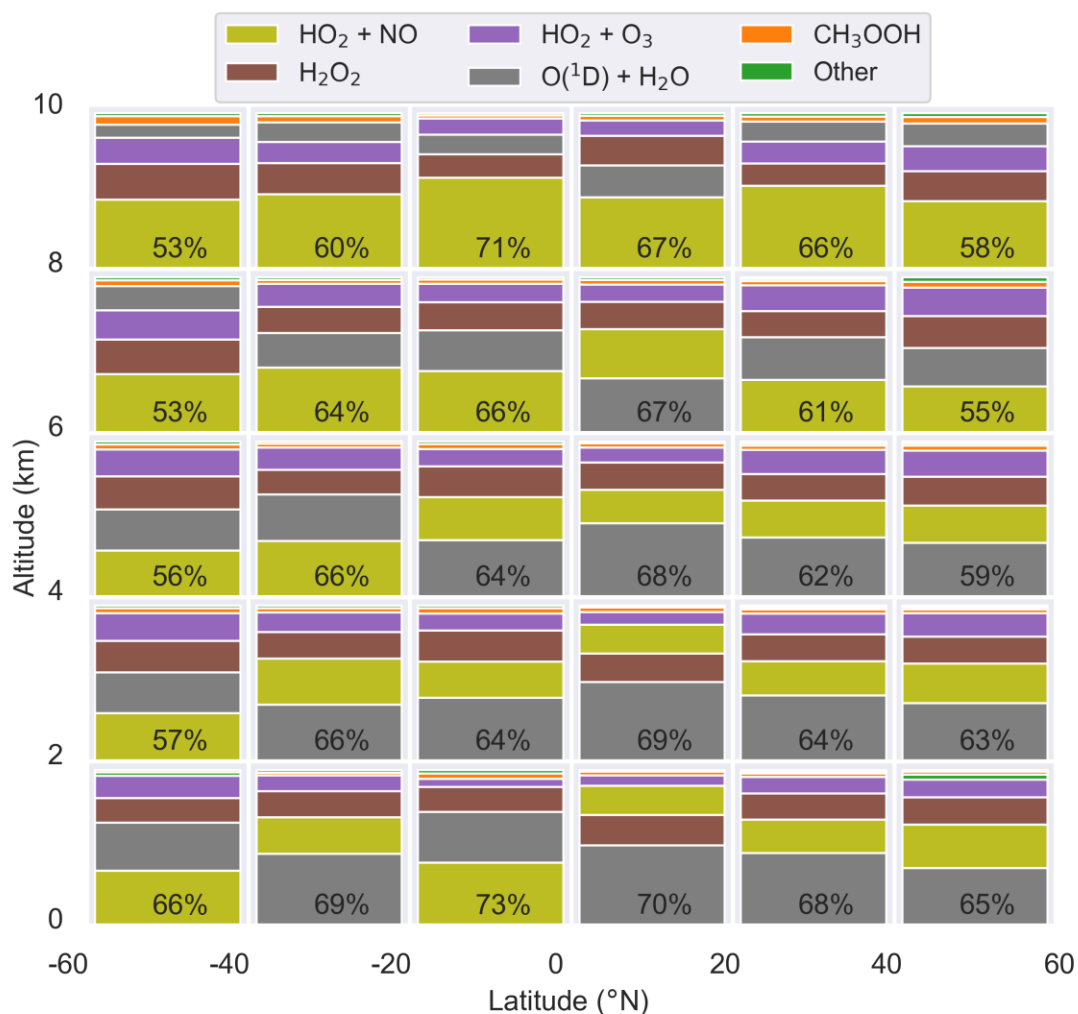
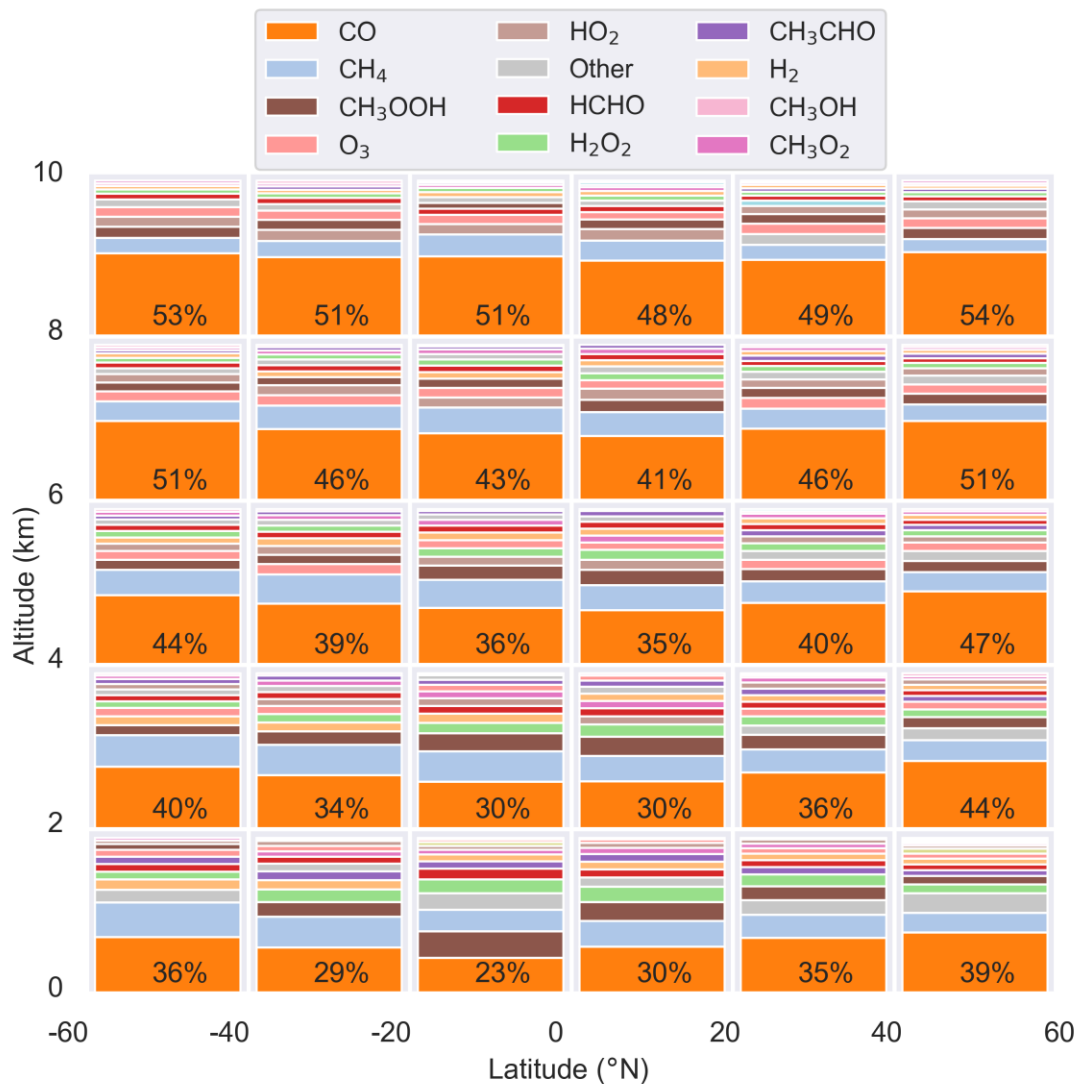


Figure S1. The Proxy_{OH} production processes comprise most of the total OH production on regional scales across the remote marine troposphere, according to F0AM box model simulations of HO_x constrained with the ATom observations. Stacked bars within each 2 km by 20° latitudinal bin describe the main contributors to total OH production in each bin, with percentages detailing the summed contribution among the reaction rates included in Proxy_{OH} , $\text{O}(^1\text{D}) + \text{H}_2\text{O}$ and $\text{HO}_2 + \text{NO}$.



61

62 **Figure S2.** Reaction with carbon monoxide (CO) is the largest sink for OH in all regional bins (2
 63 km altitude by 20° latitude). Stacked bars within each bin describe the main contributors to total
 64 OH reactivity in each bin, according to FOAM box model simulations of HO_x constrained with the
 65 ATom observations with percentages detailing the contribution of the OH lifetime against CO to
 66 the total, bin-specific OH reactivity.

67

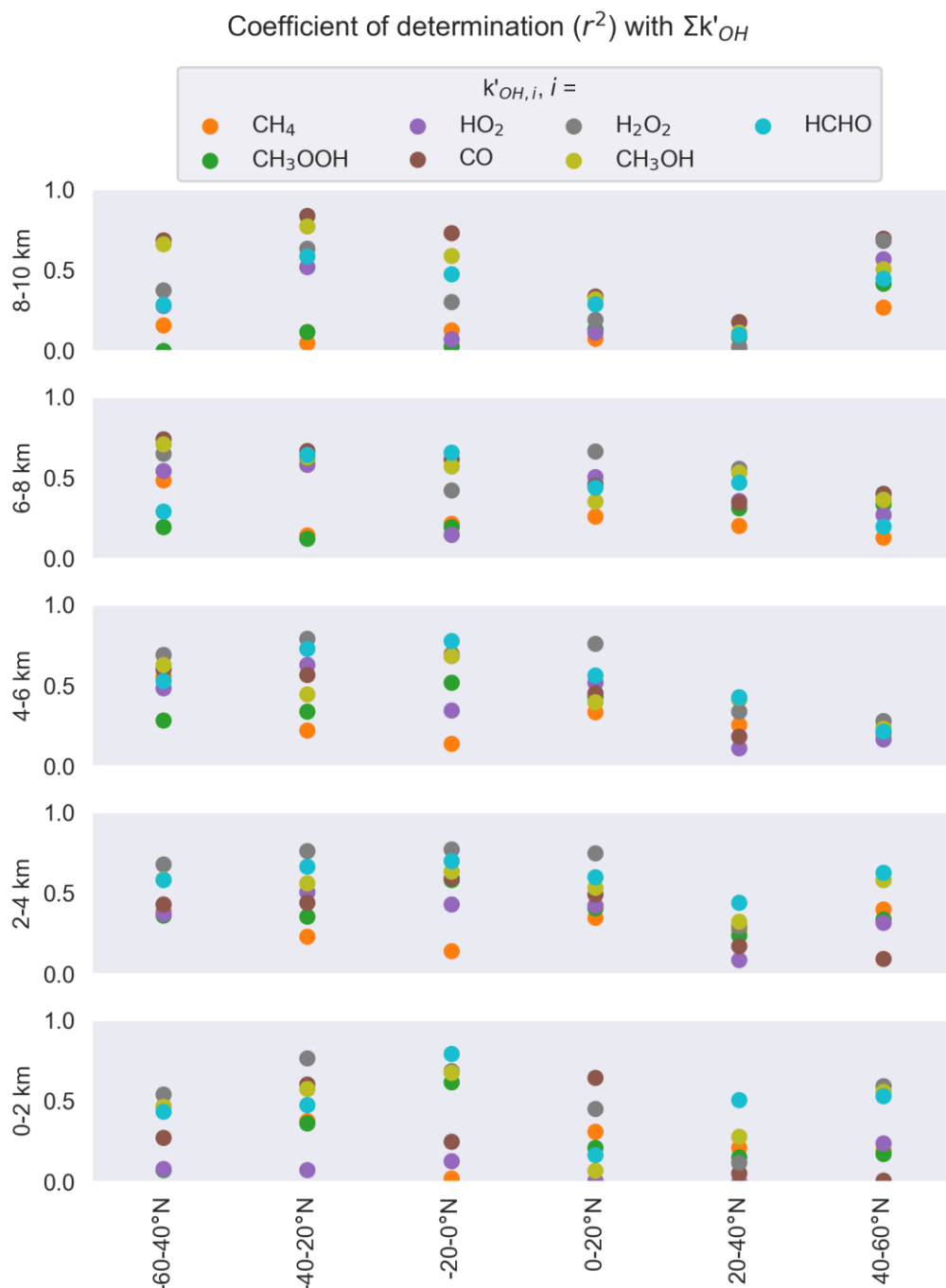


Figure S3. Among all OH loss processes, the OH reaction with the hydroperoxyl radical (HO₂), carbon monoxide (CO), hydrogen peroxide (H₂O₂), methanol (CH₃OH) and formaldehyde (HCHO) show the strongest correlation with the total OH reactivity ($\Sigma k'_{OH}$) along the ATom flight tracks subset into 2 km altitude by 20° latitude spatial bin. The OH reactivity for each of these molecules has a larger coefficient of determination with $\Sigma k'_{OH}$ across more of the spatial bins of our domain than any other individual sink ($r^2 > 0.4$ in at least 10 bins) according to FOAM box model simulations constrained with the ATom observations. Colors indicate the molecule.

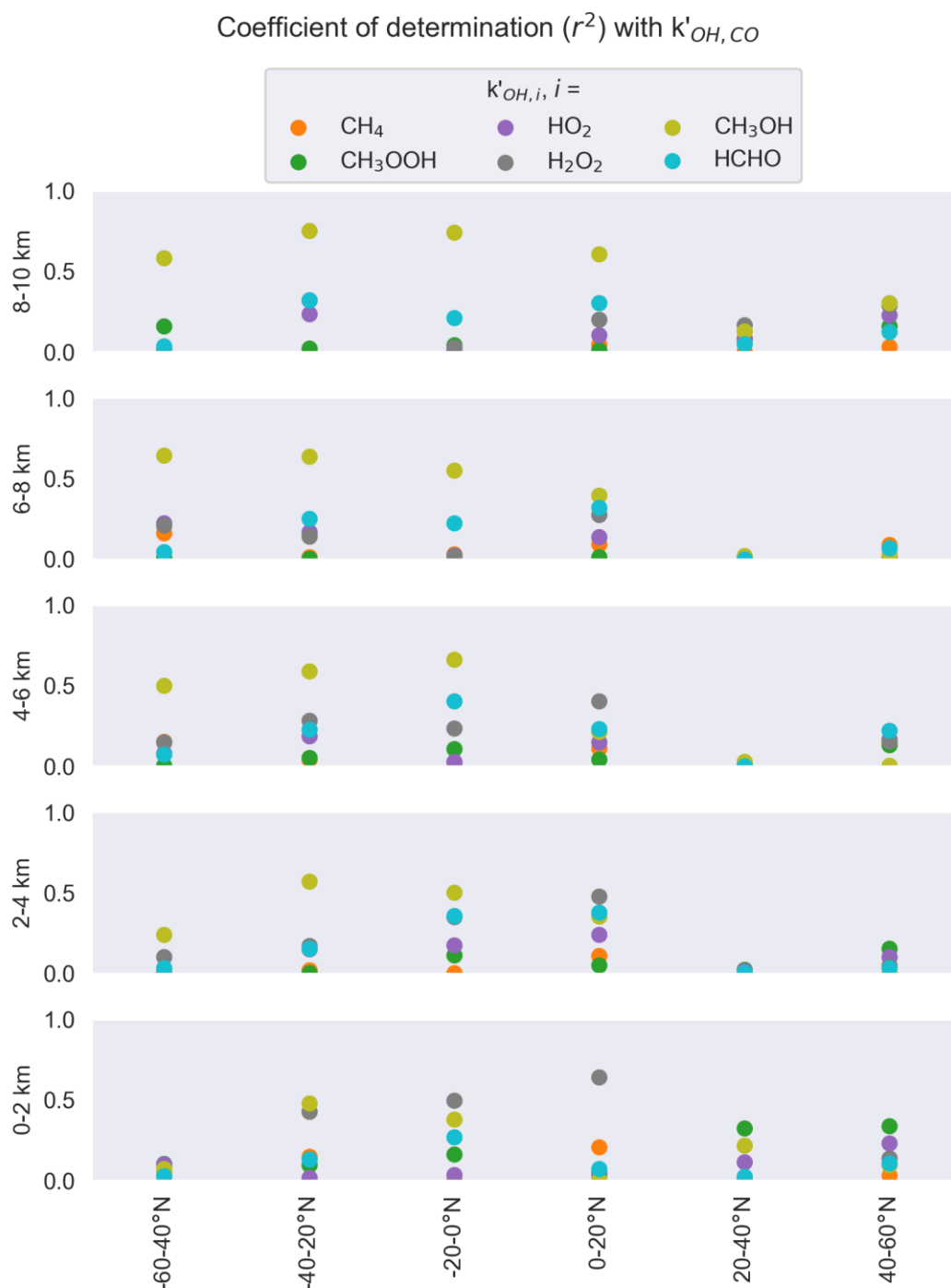


Figure S4. The OH reactivity against carbon monoxide ($k'_{OH,CO}$) covaries with other OH loss pathways that also co-vary with total OH reactivity along the ATom flight tracks (see Figure S3) over regional (2 km altitude by 20° latitudinal bin) scales, according to FOAM simulations of HO_x (OH + HO₂) constrained with ATom observations.

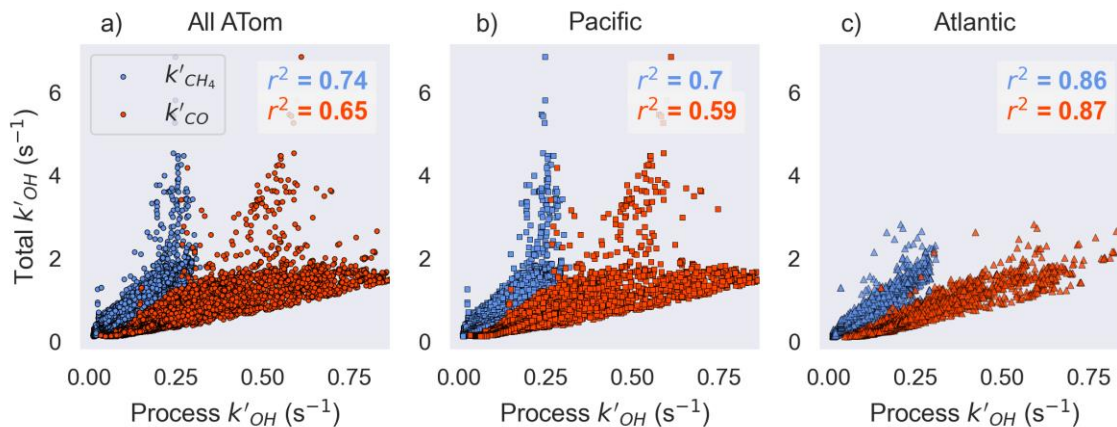


Figure S5. F0AM box model simulations of OH constrained with the ATom observations suggest that the methane (CH_4 , blue) reactivity ($k'_{\text{CH}_4} = k'_{\text{CH}_4+\text{OH}}[\text{CH}_4]$) best reflects variations in the total OH sink (interpreted via the coefficient of determination, r^2) over larger spatial scales (a, full ATom dataset) or in the Pacific (b), while carbon monoxide reactivity ($k'_{\text{CO}} = k'_{\text{CO}+\text{OH}}[\text{CO}]$, orange-red) reflects variability in the total OH reactivity over the Atlantic (c).

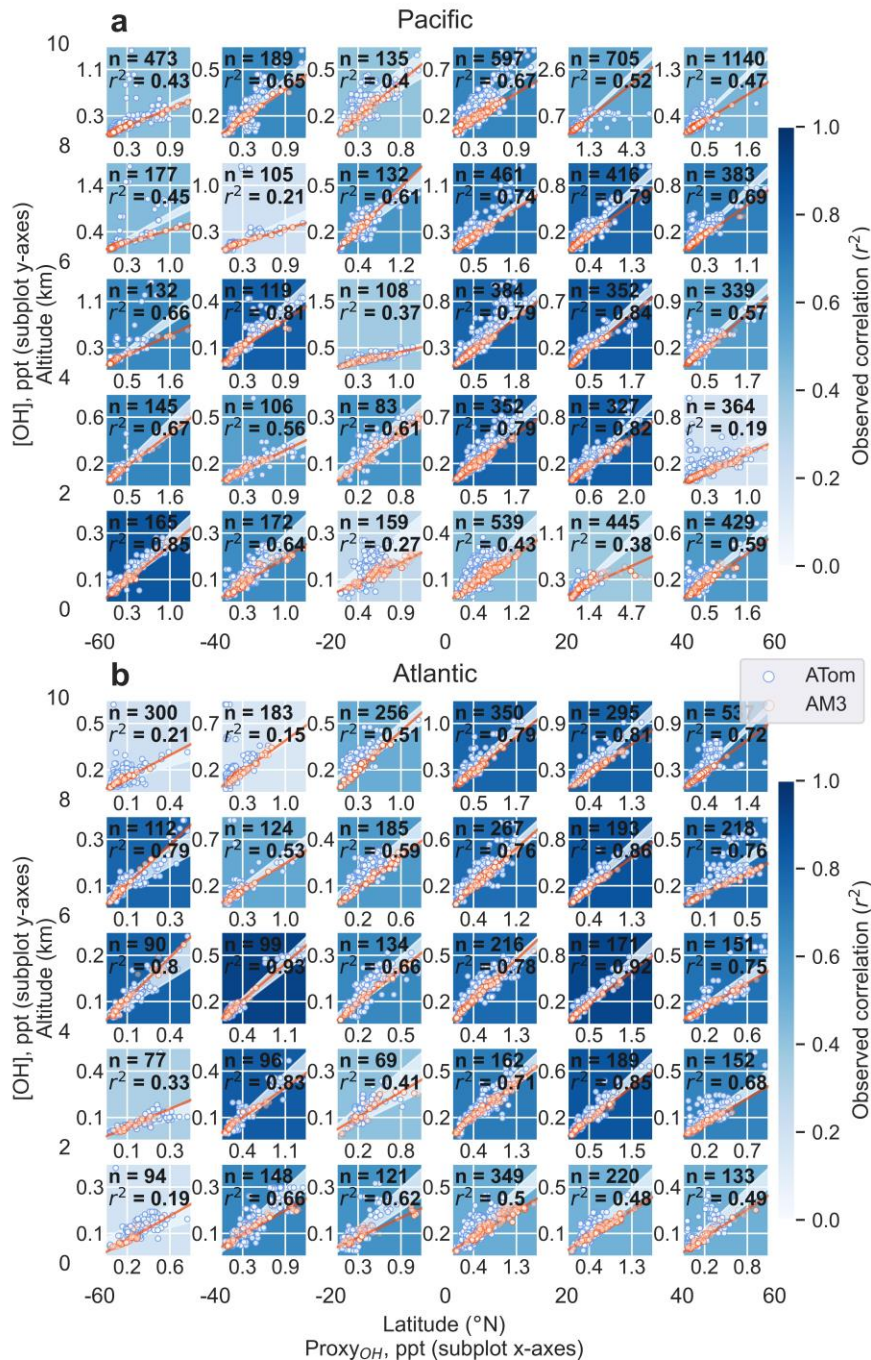


Fig. S6. As in Figure 1, but the background color reflects the $\text{Proxy}_{\text{OH}}[\text{OH}]$ coefficient of determination (r^2) for the ordinary least square regression of the ATom data in each bin. The $\text{Proxy}_{\text{OH}}[\text{OH}]$ mean variance explained ($r^2_{\text{Mdn}} = 0.66$) using the raw ATom data is lower than that using the posterior distributions that account for the instrumental uncertainty in Proxy_{OH} (compare subplot background colors in Figure S6 to Figure 1). For this figure, each subplot includes the sample size and r^2 value for the ATom data in that spatial bin (black text at top) as well as the Proxy_{OH} and $[\text{OH}]$ range on the x- and y-axes, respectively.

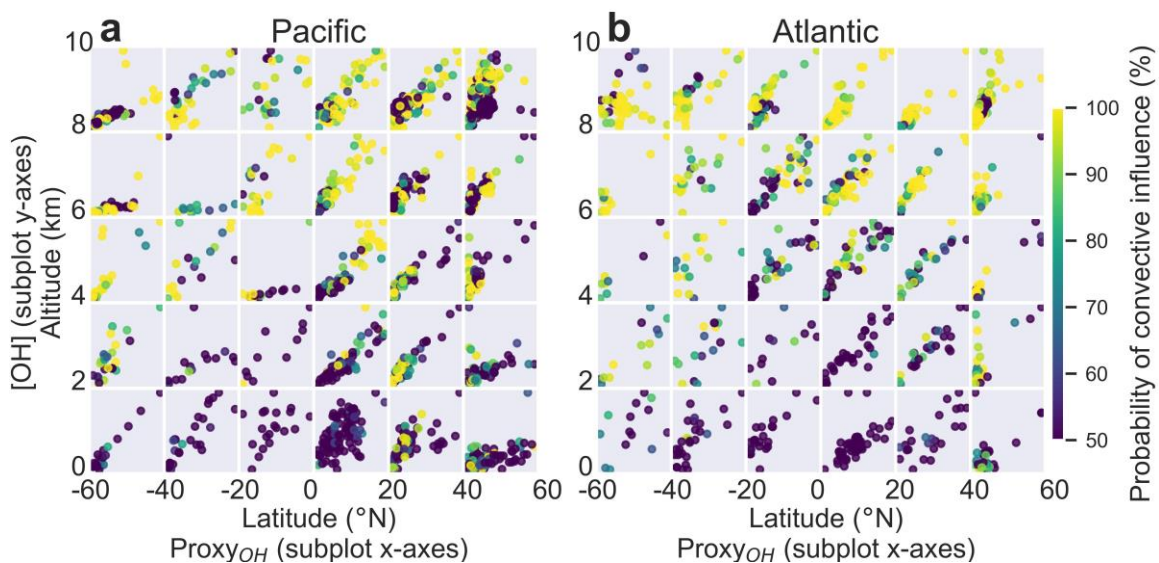


Fig. S7. ATom-observed Proxy_{OH}[OH], as in the blue points in Figures 1 and S6, color coded with the probability of convective influence. The probability of convective influence was provided as part of the 10 s ATom merge dataset (5), and was developed with 10 day back trajectories that use the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) meteorology, as well as NASA Langley cloud products (6).

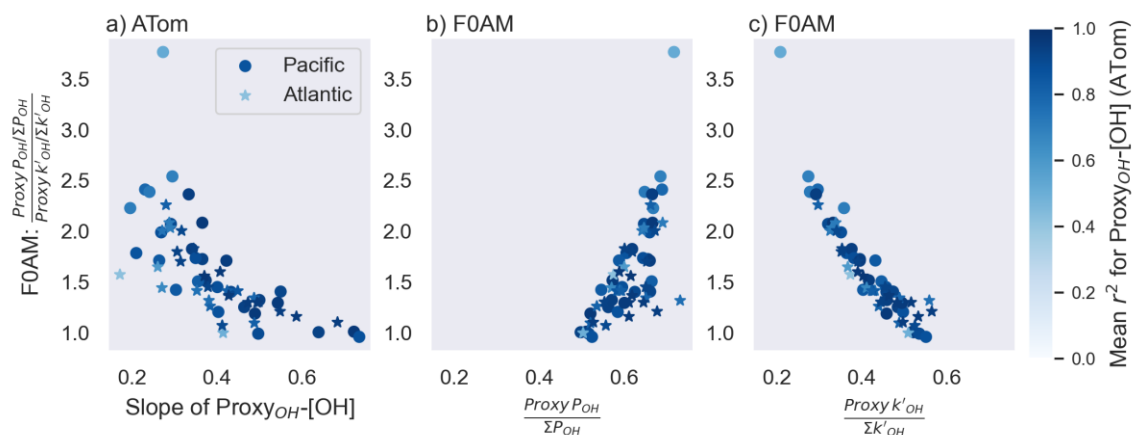


Figure S8. The ATom-observed Proxy_{OH}-[OH] regression slope reflects the extent to which the proxy processes capture the total magnitude of OH production and loss (a). According to F0AM, Proxy_{OH} underestimates the total OH sink more than its source, as indicated by the x-axis range of subplots c) and b), respectively, that show the fractional contribution of Proxy_{OH} processes to the total. This suggest that the sink component of Proxy_{OH} modulates the spatial variations in slope. Each point represents a 2 km by 20° latitudinal bin in either the Pacific (circle) or Atlantic (star) oceanic basin along the ATom flight tracks over the remote marine troposphere.

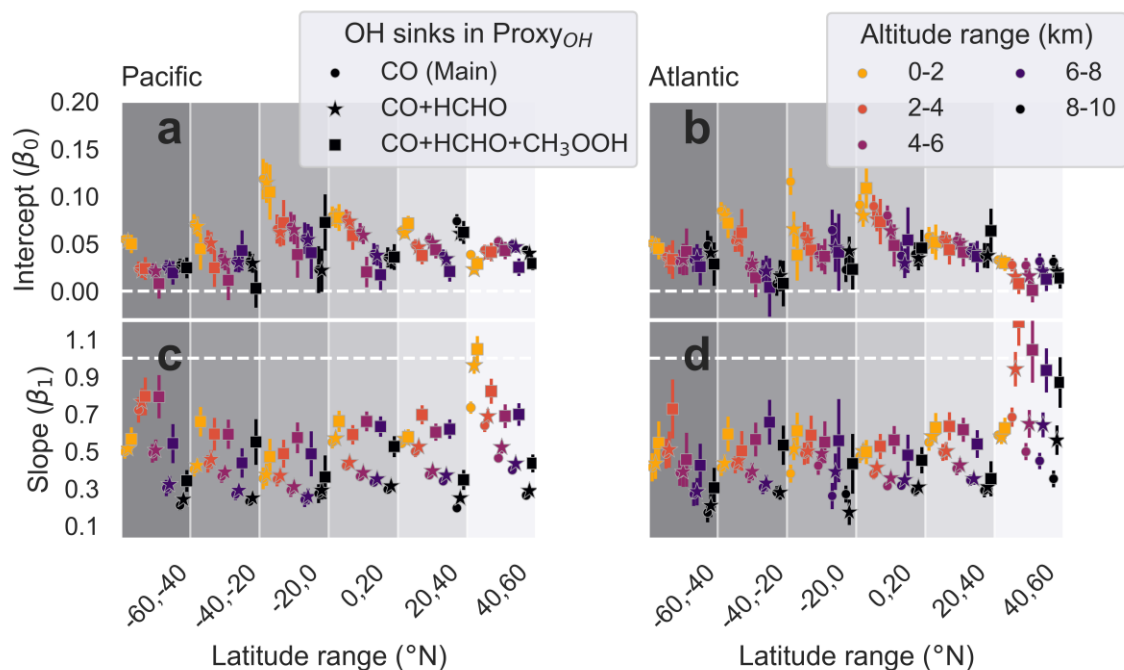


Figure S9. As in Figure 2, but here including results for several potential formulations of Proxy_{OH} in which the sink term is represented by the OH lifetime against: carbon monoxide (CO, circles; reproduced from the main text), CO and formaldehyde (HCHO, stars), or CO, HCHO and methyl hydroperoxide (MHP, squares).

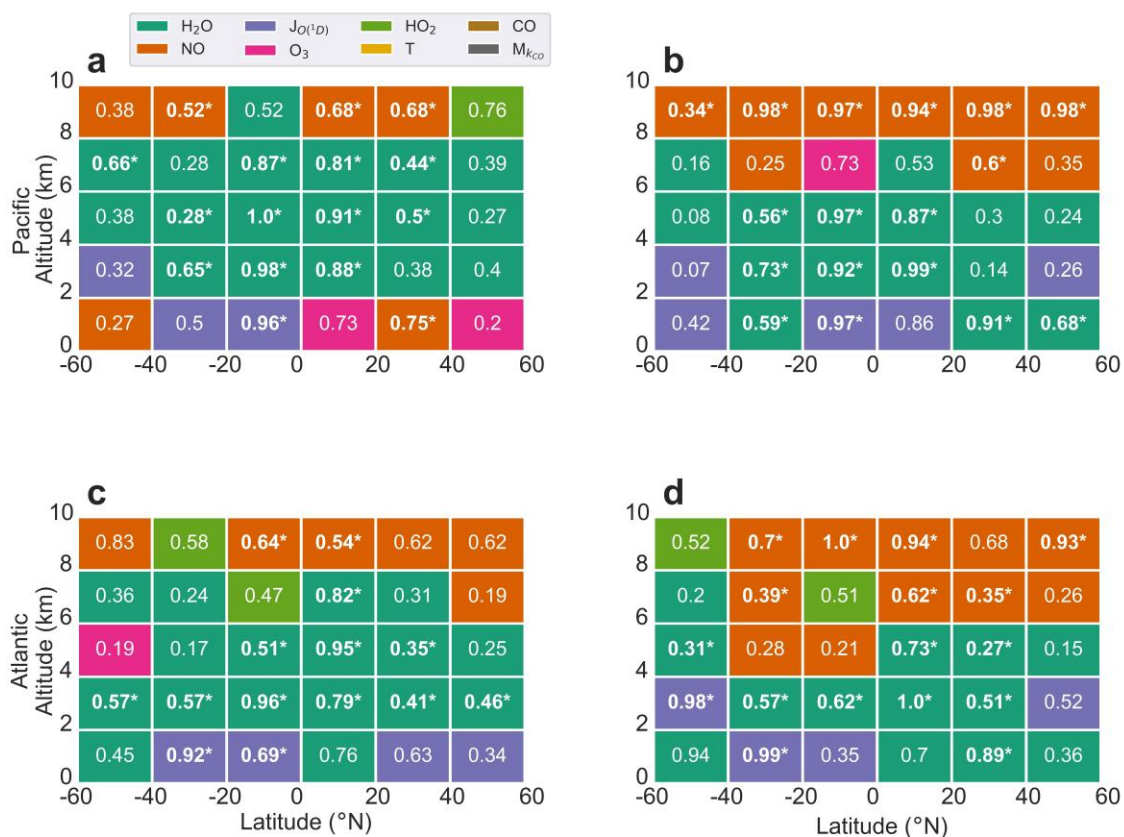


Fig. S10. As in Figure 3, but separated among the ATom observations sampled within the Pacific (top row) or Atlantic (bottom) oceanic basins.

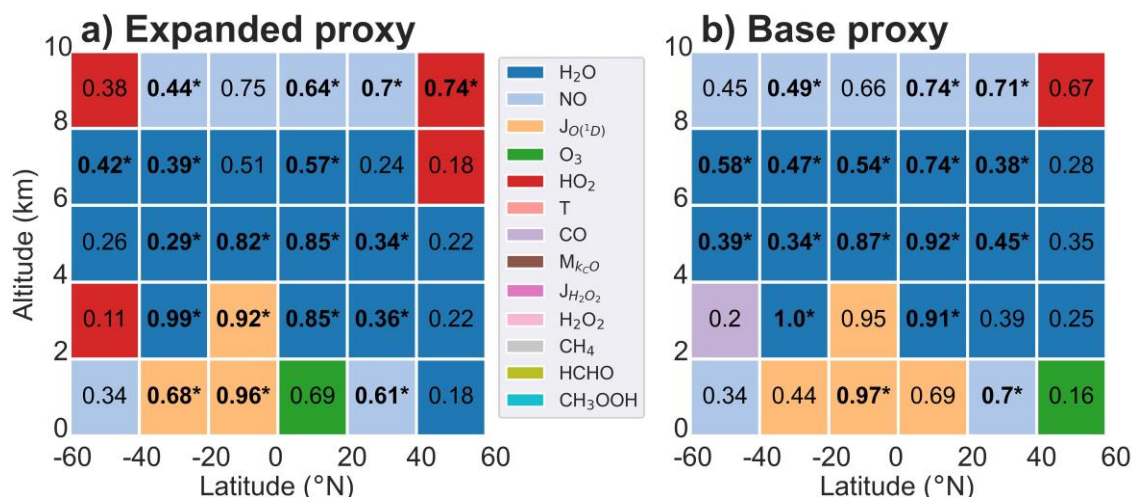


Figure S11. As in Figure 3 (all estimates are directly from ATom data), but here showing the proxy sensitivity to each of the physical or chemical terms for an expanded form of the proxy (a). Panel (b) shows that the proxy sensitivity using the same data sampling as in (a), but only including the terms used in ProxyOH as in Figure 3. Differences between (b) and Figure 3 emerge because the sample size of the expanded proxy analysis is smaller due to missing data associated with including more instruments and/or measured terms. The main conclusions regarding the widespread sensitivity to water vapor and NO hold. This includes each spatial bin where Figure 3 indicates the term contributes $r^2 > 0.2$ units over the next most influential term (denoted with asterisks), except for the 2-4 km 0-20°S bin, which switches from water vapor (Figure 3) to $J_{O(^1D)}$ (b). Inclusion of the additional terms in panel (a) also does not change the variable identified for any spatial bins denoted with asterisks in panel (b). Overall, both potential proxy forms identify the terms included in the base proxy as the primary contributors to proxy spatial variations.

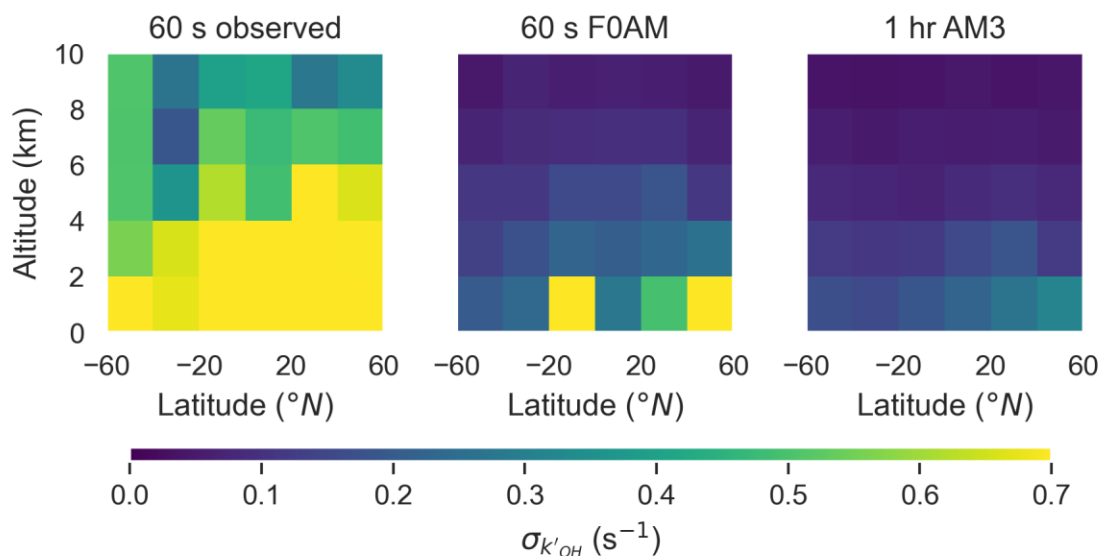


Fig. S12. ATHOS-observed variability in OH reactivity (k'_{OH}) (a) is underestimated by the F0AM box model (b) and AM3 global chemistry-climate model (c) over the remote marine troposphere, as indicated by the k'_{OH} standard deviation ($\sigma_{k'_{OH}}$) for each 2 km by 20° latitudinal bin along the ATom flight tracks. In this plot, the observations are screened for OH reactivity below the limit of detection ($< 0.3 \text{ s}^{-1}$, 7).

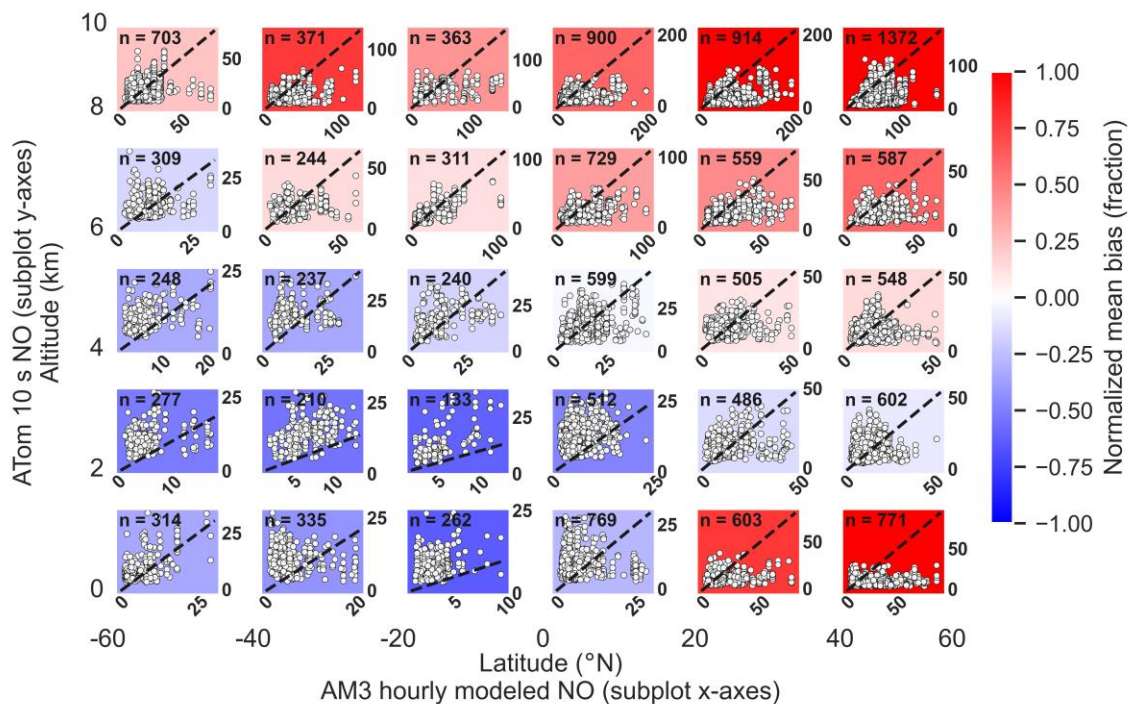


Fig. S13. Spatial distribution of the normalized mean bias of NO (subplot background color) when comparing 10 s average ATom observations (y-axis) to hourly AM3 simulations sampled along the aircraft flight tracks (x-axis). The 1:1 ratio, which would indicate a perfect observation-model match, is shown as a black dashed line. Note that this plot does not screen for [OH] below the limit-of-detection. The text at the upper left of each subplot describes the sample size in that bin. [NO] is in ppt.

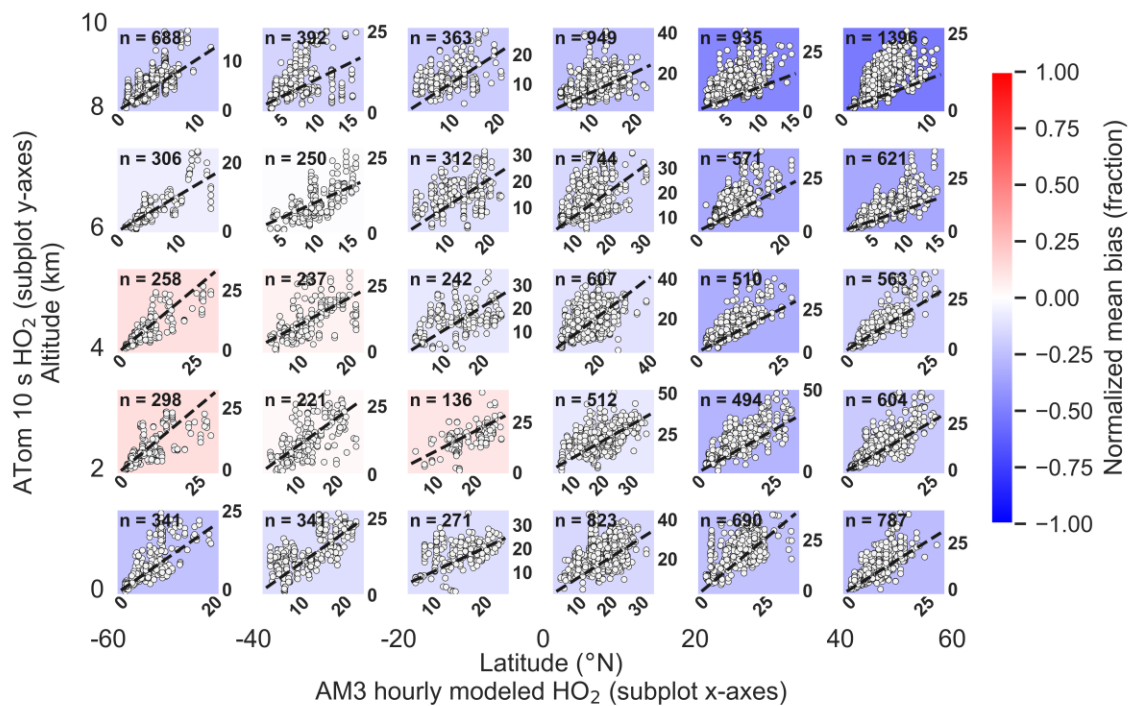


Figure S14. As in Figure S13, but for HO₂ (ppt).

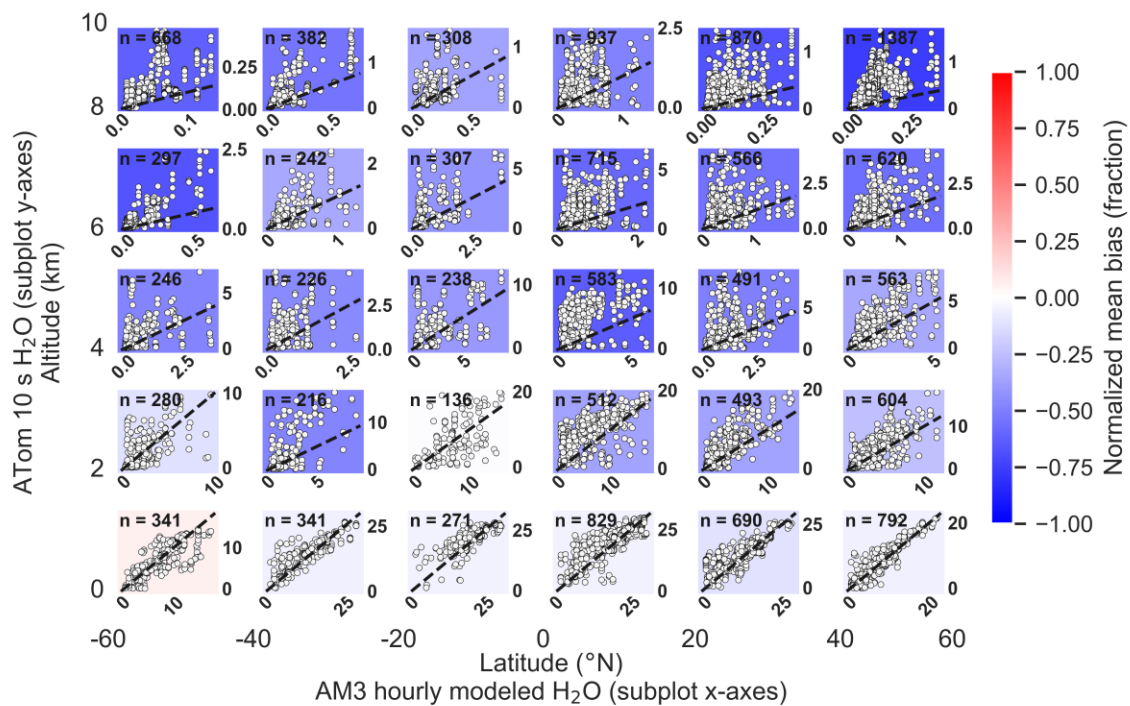


Fig. S15. As in Figure S13, but for H₂O (ppth).

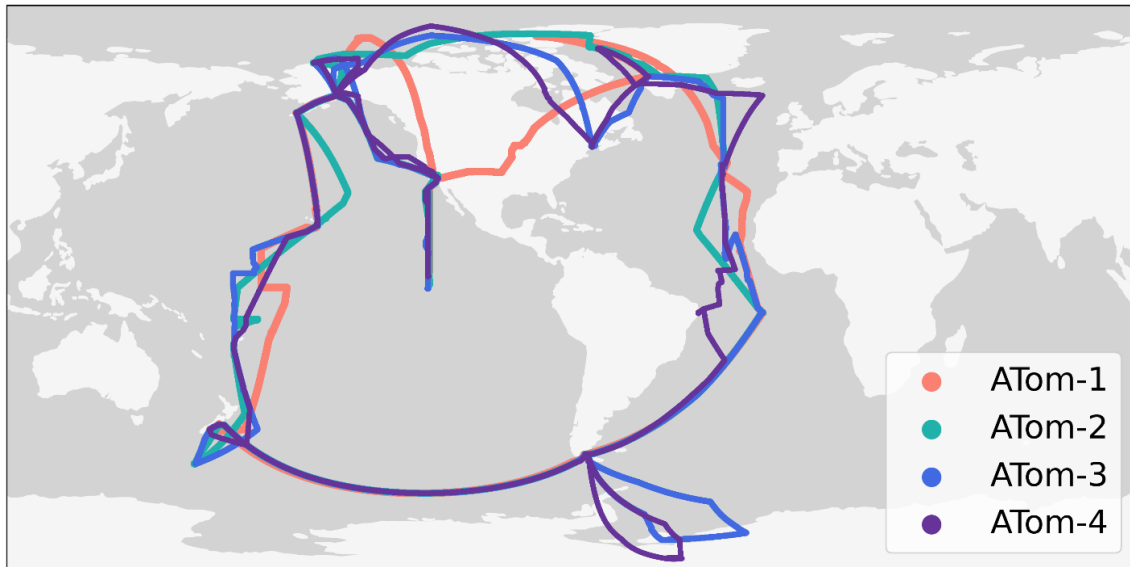


Fig. S16. Map of the flight tracks for ATom-1 (August, 2016), ATom-2 (February, 2017), ATom-3 (October, 2017), ATom-4 (May, 2018).

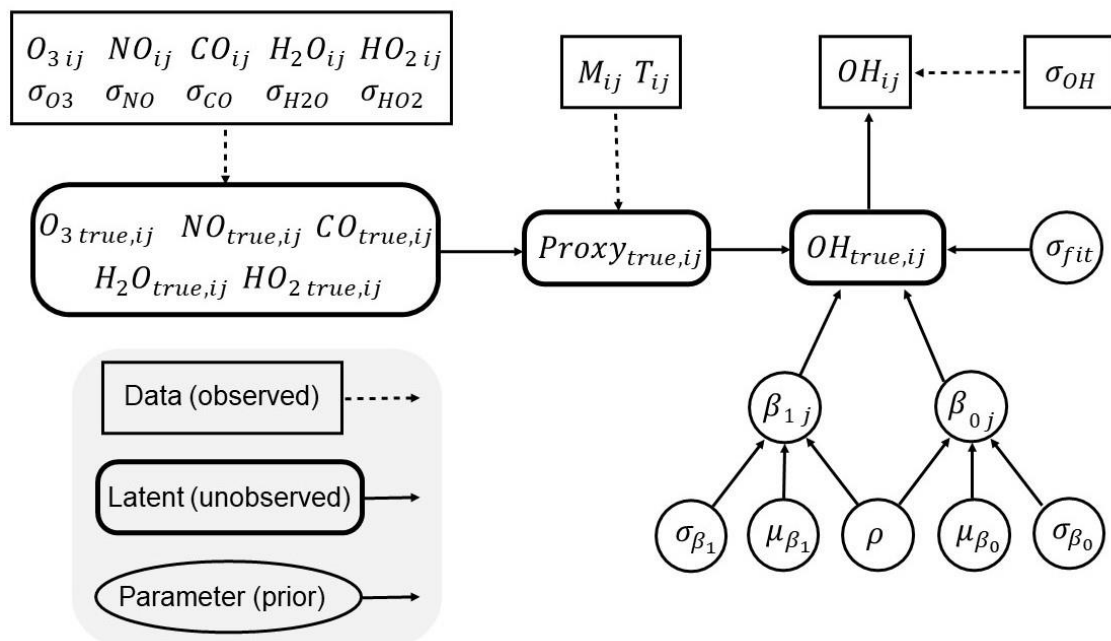
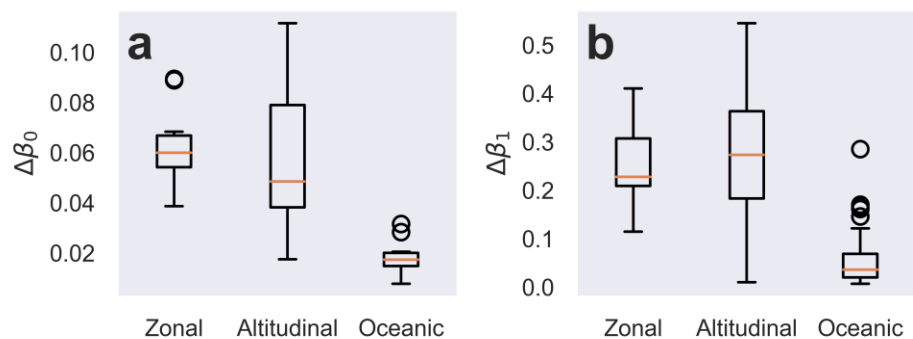


Fig. S17. Directed acyclic graph demonstrating the relationship between the prior parameters and observed and unobserved variables in the Proxy_{OH}-[OH] Bayesian regression. The explanatory variables for this model are the observed (“Obs”) mixing ratios of OH and each of the proxy terms: H₂O, O₃, NO, HO₂ and CO, $J_{O(^1D)}$, temperature (T), and the pressure dependence of the CO and OH reaction (M). The Proxy_{OH} and [OH] values sampled across the the individual measurements’ uncertainty distributions are designated by the subscript “true.” ρ defines the necessary relationship between intercept (β_0) and slope (β_1) priors, and the parameter σ_{fit} describes the overall variance in OH observations unexplained by the proxy relationship. The shape of the variable enclosure details whether it is observed data (rectangle), unobserved/latent (rounded corner rectangle) or a regression parameter (circles). The i and j subscripts indicate that the variable is grouped by ATom 10 s timestep (i) and/or spatial bin (j).

190



191

192 **Figure S18.** Difference in the a) intercept (β_0) and b) slope (β_1) of the Proxy_{OH}-[OH] relationship
193 between the maximum 75th percentile and minimum 25th percentile when sampled across latitude
194 (10 bins), altitude (12 bins) and ocean (30 bins). Boxes span the 25th (lower) to 75th (upper)
195 percentile range of differences, while whiskers span 1.5 times the interquartile range either below
196 the 25th or above the 75th percentile values. Circles fall outside of this range.

Table S1. The main reaction rates that contribute to total OH production (ΣP_{OH}) generally covary with either the reaction of water vapor and singlet-d atomic oxygen ($H_2O + O(^1D)$) or the hydroperoxyl radical and nitric oxide ($HO_2 + NO$), such that summing across additional reaction rates does not offer more information about spatial variations in ΣP_{OH} . Reaction rates are drawn from 60 s simulations of HO_x ($HO_x = OH + HO_2$) along the ATom flight tracks, constrained with the aircraft observations (Methods 7.1.2). The FOAM simulations are subset into 2 km altitude by 20° latitude bins over the remote marine troposphere. Among the main contributors to ΣP_{OH} , the table details their spatial median fractional contribution across all bins, as well as the spatial median coefficient of determination (r^2) between each term and either $H_2O + O(^1D)$ or $H_2O + NO$.

	Median contribution to ΣP_{OH}	Median r^2	
		$H_2O + O(^1D)$	$HO_2 + NO$
Hydrogen peroxide photolysis $J_{H_2O_2}[H_2O_2]$	0.18	0.82	0.12
HO_2 reaction with ozone $HO_2 + O_3$	0.14	0.46	0.40
Methyl hydroperoxide photolysis $J_{CH_3OOH}[CH_3OOH]$	0.03	0.36	0.04

Table S2. The $Proxy_{OH}[OH]$ residuals do not show a systematic observed relationship with other atmospheric variables that may affect OH over 2 km altitude by 20° latitude bins over the remote marine troposphere (middle column). The within-bin mean for these processes also do not spatially correlate (across bins) with the within-bin mean intercept (right column). $Proxy_{OH}[OH]$ residuals are calculated as the difference between $[OH]$ inferred using the $Proxy_{OH}$ regression and ATom-observed $[OH]$, when implementing the bin-specific mean intercept and slope across the Markov Chain Monte Carlo simulations (Methods 7.2) to infer $[OH]$.

ATom-observed variable or calculated reaction rate	Regression residual		Mean intercept
	r^2_{Mdn}	r^2 IQR	r^2
Solar zenith angle	0.03	0.01 – 0.07	0.11
Formaldehyde	0.02	0.01 – 0.06	0.08
Hydrogen peroxide photolysis ($J_{H_2O_2}[H_2O_2]$)	0.02	0.00 – 0.05	0.01
P_{OH} via $HO_2 + O_3$	0.02	0.00 – 0.05	0.01
Methyl hydroperoxide	0.01	0.00 – 0.03	0.14
Hydrogen peroxide (H_2O_2)	0.01	0.00 – 0.05	0.05
Particle number concentration	0.01	0.00 – 0.04	0.01
Methane	0.01	0.00 – 0.02	0.01
Probability of convective influence	0.00	0.00 – 0.02	0.13
Bromine oxide	0.00	0.00 – 0.01	0.05

Table S3. Alternative formulations for the sink component of Proxy_{OH} show a weaker observed relationship than when Proxy_{OH} employs only the OH lifetime against CO using ATom measurements of the remote marine troposphere. Potential sink formulations for Proxy_{OH} were identified using FOAM simulations of HO_x constrained using ATom data (section 2). The individual OH reactivities ($k'_{OH,i}$) that best reflect variations in the total reactivity are implemented in the Proxy_{OH} denominator, separately or together (when summed). Methanol (CH₃OH) is excluded from this analysis because the ATom observations of CH₃OH are not available on the same timestep. The table includes the median coefficient of determination (r^2_{Mdn}) among the bin-specific mean r^2 values and overall mean intercept (β_0) and slope (β_1) for the proxy-[OH] regression across 1,000 Markov Chain Monte Carlo simulations (Methods 7.2) for each of the 2 km altitude by 20° latitude regional domain.

Proxy sink term (denominator): $k'_{OH,i}$, where $i =$	r^2_{Mdn}	r^2 IQR	β_0	β_1
Methyl hydroperoxide (CH ₃ OOH)	0.51	0.25 – 0.72	0.09	0.07
Formaldehyde (HCHO)	0.77	0.59 – 0.85	0.03	0.05
Carbon monoxide (CO; base)	0.90	0.80 – 0.94	0.05	0.39
CO + HCHO	0.91	0.82 – 0.95	0.04	0.48
CO + HCHO + CH ₃ OOH	0.91	0.81 – 0.95	0.04	0.64
CO + HCHO + CH ₃ OOH + CH ₄	0.78	0.62 – 0.89	0.06	0.69

Table S4. Water vapor modulates OH variations over large swaths of the remote marine troposphere. The sensitivity of OH variations is inferred using the difference in ordinary least squares fit between Proxy_{OH} and a sensitivity version that calculates Proxy_{OH} while holding an individual term constant at its bin-specific mean value. A larger $1 - r^2$ indicates greater influence over OH variations. Temperature is not found to be influential ($1 - r^2 < 0.01$).

Sensitivity term	$1 - r^2$		
	Global	Pacific	Atlantic
Carbon monoxide (CO)	0.13	0.13	0.10
Water vapor (H ₂ O)	0.91	0.96	0.79
Hydroperoxyl (HO ₂)	0.20	0.30	0.28
$J_{O(^1D)}$	0.23	0.20	0.27
Air density (M)	0.01	0.01	0.02
Nitric oxide (NO)	0.26	0.28	0.22
Ozone (O ₃)	0.50	0.46	0.25

Table S5. Measurements from the NASA ATom airborne campaign used in this analysis (5); Proxy_{OH} varies with H₂O, O₃, NO, HO₂, CO, $J_{O(^1D)}$, temperature and pressure.

Measurement	Instrument
OH, Hydroperoxyl (HO ₂), OH reactivity	Airborne Tropospheric Hydrogen Oxides Sensor (ATHOS)
Photolysis frequencies (J)	Charged-coupled device Actinic Flux Spectroradiometers (CAFS)
Ozone, Nitric oxide (NO)	NO _y O ₃ 4-channel chemiluminescence
Water vapor (H ₂ O)	Diode Laser Hygrometer (DLH)
Carbon monoxide (CO)	Merge ("CO.X") of the NOAA Picarro cavity ring down spectrometer and the Quantum Cascade Laser Spectrometer (QCLS)
Meteorology	Meteorological Measurement System (MMS)
Positioning	Navigational data
Methane (CH ₄)	QCLS
Formaldehyde (HCHO)	In Situ Airborne Formaldehyde (ISAF)
Methyl hydroperoxide (CH ₃ OOH)	California Institute of Technology – Chemical Ionization Mass Spectrometer (CIT-CIMS)

Table S6. Measurement uncertainty associated with each individual proxy variable. Each MCMC simulation draws a sample of the individual proxy variables from a normal distribution that has a standard deviation equal to the instrumental uncertainty shown here.

Explanatory variable	Uncertainty (1 σ)
OH	$\pm 16\%$
H ₂ O	$\pm 5\%$
$J_{O(^1D)}$	$\pm 20\%$
O ₃	Assigned dynamically through dataset variable O3_CL_2sigma
NO	Assigned dynamically through dataset variable NO_CL_2sigma
HO ₂	$\pm 16\%$
CO	3.5 ppb
HCHO	$\pm 10\%$
CH ₃ OOH	$\pm (30\% + 25 \text{ ppt})$

Table S7. Initial values for each of three parallel chains for the slope (β_0) and intercept (β_1) parameters of the Bayesian linear fit between the steady-state proxy and [OH].

Regression parameter	Chain 1	Chain 2	Chain 3
β_0	0.5	0.75	1
β_1	1.5	2.25	3
$\beta_{0,\mu}$	0.01	0.05	0.1
$\beta_{1,\mu}$	0.2	0.4	0.6
$\beta_{0,\sigma}$	0.2	0.02	0.1
$\beta_{1,\sigma}$	0.5	0.3	0.1
$\rho = \frac{\sigma_{\beta_{0,1}}}{\beta_{0,\sigma} \times \beta_{1,\sigma}}$	0.1	-0.6	-0.2
σ_{fit}	0.01	0.2	0.05

SI References

1. D. M. Westervelt, A. M. Fiore, C. B. Baublitz, G. Correa, Impact of regional Northern Hemisphere mid-latitude anthropogenic sulfur dioxide emissions on local and remote tropospheric oxidants. *Atmospheric Chem. Phys.* **21**, 6799–6810 (2021).
2. C. M. Spivakovsky, *et al.*, Three-dimensional climatological distribution of tropospheric OH: Update and evaluation. *J. Geophys. Res. Atmospheres* **105**, 8931–8980 (2000).
3. S. A. McKeen, *et al.*, The photochemistry of acetone in the upper troposphere: A source of odd-hydrogen radicals. *Geophys. Res. Lett.* **24**, 3177–3180 (1997).
4. S. Wang, *et al.*, Global Atmospheric Budget of Acetone: Air-Sea Exchange and the Contribution to Hydroxyl Radicals. *J. Geophys. Res. Atmospheres* **125**, e2020JD032553 (2020).
5. S. C. Wofsy, ATom Science Team, ATom: Aircraft Flight Track and Navigational Data (2018) <https://doi.org/10.3334/ORNLDAAAC/1613>.
6. Ray, E.A., Atmospheric Tomography Mission (ATom)ATom: Back Trajectories and Influences of Air Parcels Along Flight Track, 2016-2018. 0 MB (2022).
7. A. B. Thames, *et al.*, Missing OH reactivity in the global marine boundary layer. *Atmospheric Chem. Phys.* **20**, 4013–4029 (2020).
8. A. Gelman, D. B. Rubin, Inference from Iterative Simulation Using Multiple Sequences. *Stat. Sci.* **7**, 457–472 (1992).