Supplementary Information for

Heterogeneity and Chemical Reactivity of the Remote Troposphere defined by Aircraft Measurements

Results from the NASA Atmospheric Tomography mission (ATom)

Hao Guo[1], Clare M. Flynn[2], Michael J. Prather[1], Sarah A. Strode[3], Stephen D. Steenrod[3], Louisa Emmons[4], Forrest Lacey[4,5], Jean-Francois Lamarque[4], Arlene M. Fiore[6], Gus Correa[6], Lee T. Murray[7], Glenn M. Wolfe[3,8], Jason M. St. Clair[3,8], Michelle Kim[9], John Crounse[10], Glenn Diskin[10], Joshua DiGangi[10], Bruce C. Daube[11,12], Roisin Commane[11,12], Kathryn McKain[13,14], Jeff Peischl[14,15], Thomas B. Ryerson[13,15], Chelsea Thompson[13], Thomas F. Hanisco[3], Donald Blake[16], Nicola J. Blake[16], Eric C. Apel[4], Rebecca S. Hornbrook[4], James W. Elkins[14], Eric J. Hintsa[13,14], Fred L. Moore[13,14], Steven Wofsy[11]

[1] Department of Earth System Science, University of California, Irvine, CA 92697
[2] Department of Meteorology, Stockholm University, Stockholm SE-106 91, Sweden
[3] Atmospheric Chemistry and Dynamics Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD 20771
[4] Atmospheric Chemistry Observations and Modeling Laboratory, National Center for Atmospheric Research, Boulder, CO 80301
[5] Department of Mechanical Engineering, University of Colorado, Boulder, CO 80309
[6] Department of Earth and Environmental Sciences and Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY 10964
[7] Department of Earth and Environmental Sciences, University of Rochester, Rochester, NY 14611
[8] Joint Center for Earth Systems Technology, University of Maryland, Baltimore County, Baltimore, MD 21228
[9] Department of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125
[10] Atmospheric Composition, NASA Langley Research Center, Hampton VA 23666
[11] John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138
[12] Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02138
[13] Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309
[14] Global Monitoring Division, Earth System Research Laboratory, NOAA, Boulder, CO 80305
[15] Chemical Sciences Division, National Oceanic and Atmospheric Administration Earth System Research Laboratory, Boulder, CO 80305
[16] Department of Chemistry, University of California, Irvine, CA 92697

Correspondence to: Hao Guo (haog2@uci.edu) and Michael J. Prather (mprather@uci.edu).

S.1. The Modeling Data Stream

The ATom mission was designed to collect a multi-species, detailed chemical climatology that documents the patterns of physical and chemical heterogeneity throughout the remote troposphere. The work here requires a complete set of key species in each air parcel to initialize global 3D chemistry models to be able to calculate the CH_4 and O_3 reactivities over a 24 hour cycle. The ATom Modeling Data Stream (MDS) provides a semi-continuous set of 10 s air parcels with a full set of values for the key chemical reactants and conditions. We choose 10 s averages for our air parcels as a compromise to include most of the instruments, and because the 10 s merged data is a standard product (Wofsy et al., 2018). Some of our core species are measured with gas chromatographs or flask samples with longer integrations times (30-90 sec), but these can be mapped onto the 10 s parcels with loss of the higher frequency variability found in the 10 s measurements. The frequent profiling of the DC-8 gives us both vertical and horizontal scales: the vertical extent of a 10 s parcel is 50 - 110 m (55%-95% of all parcels, with <50% having near level flight) and the horizontal extent is typically 1.4 - 2.5 km (10%-90% of all parcels).

ATom completed its four deployments: ATom-1 starting 20160729 (YYYMMDD), ATom-2 starting 20170126, ATom-3 starting 20170928, and ATom-4 starting 20180424. ATom targets the remote troposphere by sampling over the middle of the Pacific and Atlantic Ocean basins. The DC-8 aircraft performed in situ profiling of the atmosphere from 0.2 km to 12 km along each flight segment as often as possible. Each deployment lasted about 4 weeks and contained 11 to 13 research flights (RF). Figure S1 maps the 48 RF, and the Table S1 summarizes each flight in terms of airports, starting date (UT), and number of 10 s parcels. For convenience, we designate the RF across the 4 deployments as ATom flights (AF) 1 through 48. The MDS data reported here consisit of 149,133 air parcels over 4 deployments with a total of 48 research flights. AF 46 is a short ferry flight from Kangerlussuaq, Greenland to Bangor, Maine with many instruments turned off and no profiling, thus these 1,106 parcels contain only flight data (MDS variables 1:11) and no chemical data.

ATom sampling of the troposphere is more uniform than most aircraft missions, but still contains some biases that can be adjusted by weighting each air parcel. Due to the typical profiling sequence (level at cruising attitude for 10 min, descent for 20 min, level flight about 160 m above the sea level for 5 min, and a 20-min climb back to cruising altitude) and to the occasional requirements of weather or air traffic control, the sampling is skewed towards the uppermost troposphere (P < 300 hPa) and, secondly, the marine boundary layer. We designate a weight for each MDS air parcel to achieve a more uniform sampling of the troposphere by mass: data are binned into 100 hPa-wide pressure bins and 10°-wide latitude bins, and each point is assigned a weight equal to the inverse of the number of points in the bin times cosine of the latitude. There are very few measurements for pressures <200 hPa and so these points are included in the uppermost 200-300 hPa bin. This ATom-1 analysis has three study domains: Global includes all parcels (32,383) weighted as above; Pacific considers all measurements (11,486) over the Pacific Ocean from 54°S to 60°N (research flights RF 1,3,4,5,6); and the Atlantic,

likewise, from 54°S to 60°N (RF 7, 8, 9) over the Atlantic basin (7,501). The ATom-1 flight tracks shown in Figure S1 identify the Pacific and Atlantic domains with very thick lines. Also shown are the regional blocks used to calculate the model climatologies for those domains.

Figure S2a shows the time series of O_3 and H_2O measured during one of the profiles of RF 3. The 1 s data is plotted along with the 10 s averages. Most of the heterogeneity including correlated variability is caught with the 10 s parcels. For RF #3. the root mean square error (RMSE) of the 10 s averages linearly interpolated to 1 sec is 6% for H_2O and 3% for O_3. For comparison the short-gap interpolation described below has an RMSE twice as large for these species.

The problem in developing the MDS from the 10 s merge files is that gaps occur in individual species on a range of times scales due to calibration cycles, sampling rates, and instrument malfunction. For the chemistry modeling of an air parcel, we need complete chemical specification and thus data gaps in individual species must be filled where we have adequate information. Early versions of the ATom-1 MDS were generated and used in modeling studies that are included here, but we found several problems with the approaches used for gap filling and had to entirely redo the method. MDS_R0 adopted early recommendations for use of a photo-stationary steady-state value for NO_x (PSS), which was later rejected by the ATom science team as flawed. MDS_R1 reverted to the observed NO_x values but had problems when using flask sample data with a lower limit of detection. Both of these MDS versions used CO and other species as a proxy for gap filling, but closer examination showed that this method lacks skill.

The MDS R2 method for gap filling is fully documented in this Supporting Information. MDS_R2 defined the core reactive species (H_2O, O_3, CO, CH_4, NO_x, NO_xPSS, HNO_3, HNO_4, PAN, CH_2O, H_2O_2, CH_3OOH, acetone, acetaldehyde, C_2H_6, C_3H_8, i-C_4H_{10}, n-C_4H_{10}, alkanes, C_2H_4, alkenes, C_2H_2, C_5H_8, benzene, toluene, xylene, CH_3ONO_2, $C_2H_5ONO_2$, $RONO_2$, CH_3OH) and corollary species indicative of pollution or processing (HCN, CH_3CN, SF_6, relative humidity, aerosol surface area (4 modes), and cloud indicator), see Table S2. Every species in each air parcel is now flagged so that the instrument is clearly identified (in the case that two instruments measure the same species) and the type of the gap filling (dependent on the length of the gap) is denoted so that the users can develop their own criteria for including, or not including, the gap-filled species. Flags 1 & 2 indicate a reported measurement from a primary (1) or secondary (2) instrument. Flag 3 means short-gap filling. Flags 4 & 6 indicate log-gap filling for tropospheric and stratospheric parcels, respectively. Flag 5 applies to missing flights with no data from that instrument(s), and these were filled by a multiple linear regression from the parallel flights. Flag 0 indicates not a number (NaN), which only occurs for AF 46.

S.1.1. Primary ATom data sets

This section describes the creation of MDS revision R2; the early-release revisions R0 and R1 used a different algorithm and flagging system and should no longer be used. The 'Mor' data sets created by Wofsy et al. (2018) contain merges of the ATom 10 s data

(Mor.all), the WAS flask data analyzed post-flight (Mor.WAS.all) and the in-flight TOGA chromatograph-mass spectrometer data (Mor.TOGA.all). These data sets are released in a gzip file with the YYYY-MM-DD of their creation. For this MDS version (2020-05-27), we use the following 3 data sets:

'Mor.all.at1234.2020-05-27.tbl' (653,494,900 bytes)
'Mor.WAS.all.at1234.2020-05-27.tbl' (49,091,169 bytes)
'Mor.TOGA.all.at1234.2020-05-27.tbl' (80,579,206 bytes)

The Mor data are ASCII text files with extremely long records and difficult to read, containing a mix of comma-separated floating point, integer and character strings. For Mor.all, the 149133 records contain 675 comma-separated variables (but this can change with different releases). Some of the floating point variables are longer than 20 characters due to excess precision in the scientific notation. We pre-process these with a Fortran generic read(5,*) using the comma separation to generate character strings. The code searches the title (first) record of the Mor…tbl to identify the specific columns that we need for MDS (in this case 39 out of 675). The 39 key data from each record are rewritten in formatted form (39a40, because some floating point variables were excessively long and 39a20 was inadequate) with comma separation. All numerical values are copied verbatim, but the text 'NA' is replaced by 'NaN'. This new file can be simply imported into Matlab or more easily read by other software. Further, this approach ensures that the correct quantities are pulled from the Mor…tbl file, even if the column order changes due to addition or removal of data. The WAS and TOGA observations have separate files with the start and end times of the observed air mass, which is greater than the 10s interval in the regular file. Both WAS & TOGA Mor…data sets have a large number of data columns (729 & 727) with fewer records (6,991 & 12,168, respectively).

The 3 Fortran output files are imported into Matlab (using 'Import Data') and then processed as described below. The instructions and Matlab code are included in text files containing Matlab commands: 'Pmat-Mor1.txt', 'Pmat-WAS+TOGA.txt', 'Pmat-MDS0n.txt').

S.1.2. Preliminary processing and identifying gaps

In terms of critical flight data (time, latitude, longitude, altitude), there are no gaps in the record. UTC_stop has a gap, but this variable is not used in the MDS (10s intervals are assumed).

The Mor.all.at1234.2020-05-27 data set of 149,133 10s parcels was sorted into deployments and research flights. The beginning and end points of each research flight (RF) along with the deployment and starting date of each flight are given in Table S1. All together there are 48 flights, but AF 46 contains only flight data. All three types of Mor data include some measurements close to the airports, which often have ground-level pollution. We remove these data by including only measurements at altitudes of 900 m or more above the takeoff/landing airport. The record collapses to 146,494 parcels, also shown in Table S1.

The list of MDS R2 variables, their MDS identifiers (ending in _M) and the sources in standard ATom nomenclature are given in Table S2. The flag variables (0 to 6) are also explained there. Information about each research flight is summarized in **Table S3abcd**, including the average latitude, longitude and altitude of the 10s parcels (all equally weighted here). The abcd sub-tables correspond to the 4 deployments. For each of the MDS variables 12 to 50, The % of non-NaN values with flags = 1, 2 or 3, is shown (the remaining % has flags = 4, 5 or 6). These data correspond to a primary or secondary direct measurement (1 or 2) or else short-gap interpolation (3, see text below). Missing data for an entire flight (0%) has shaded cells.

Mor.all combined species and fixes. The primary MDS NOx values were created by simply summing NO_CL + NO2_CL before any attempt to deal with the negative values. The number (27071) of NOx NaNs coincides with those of NO2_CL. The alternative photostationary state NOx values (NOxPSS) were calculated from O_3, NO and J-values and was originally proposed as a more accurate value for NOx. Subsequent analysis has shown this approach is biased, and it is included here only for ATom-1 because some early model studies used it in the MDS R0 version. A small number (22) of CH_4_QCLS values have unrealistic abundances <1000 ppb and these are converted to NaNs. The NaNs in these cases were filled using the algorithm below.

TOGA and WAS combined species and immediate fixes. Methyl and ethyl nitrate (WAS only) are kept separately but the 6 higher organo-nitrates are combined into $RONO_2$; the limited TOGA organo-nitrates are not used. For both WAS and TOGA, toluene and ethylbenzene are combined into toluene, and the two forms of xylene are combined. Both forms of butane are kept, but higher alkanes are combined into 'Alkanes' for both TOGA and WAS. TOGA and WAS use -888 flags for LLOD and these are converted to 0.001 ppt because the LLOD values for these species (e.g., 3 ppt) are much higher than remote background values and setting them to the LLOD level would be misleading. TOGA's toluene has some mistaken values of 888 and 999 instead of -888 and -999 and these are corrected. All -999 values, as well as all gaps in either TOGA or WAS measurement intervals are converted to NaNs. The WAS and TOGA data have time stamps (stop minus start) much longer than the 10 sec parcels in the Mor.all data sets, and their values are mapped onto the 164,494 parcels whenever their start or stop time falls within the 10s start-stop range, else they are filled with NaNs. The WAS and TOGA instruments sample air averaged over typically 30 to 90 sec, and then have a gap before the next measurement, varying from 30 to 300 sec. The TOGA length-of-measurement is regular with the 10%-90%ile range being about 35 sec and the same percentile length-of-gaps being about 85 sec. The WAS data comes from flasks filled in flight, and the time to fill a flask depends on the pressure, and the gap depends on the operator decision: the 10%-90%ile length-of-measurement is 32 to 90 sec, and the corresponding gaps are 33 to 285 sec.

S.1.3. Interpolation and fill of data gaps

The actions here are arbitrary but judicious, and every attempt was made to avoid introducing spurious data. There are a number of negative values for chemical variables that are intrinsically positive definite. Instrument reporting of a negative value is expected when the concentration is near the limit of detection or within the instrumental noise range. The MDS choice is simply to take all such values less than or equal to 0 and convert to 0.001. Since these negative values usually represent a small concentration close to the detection limit, they have little impact on the chemistry calculations using the MDS. If analyzing statistics near this range, the original Mor data sets should be used.

Pressure and temperature. P and T have 5 very small gaps of length ~6 (# of 10s parcels missing) plus a longer gap of length 28. All gaps occurred during smooth descent or ascent and so were filled using linear interpolation. These are denoted by flag_M(:,10) = flag_M(:,11) = 3.

H_2O and relative humidity. There are a number of short gaps in the record of H_2O_DLH and RHw_DLH, and only 2 longer gaps (length = 83 and 87). One of the long gaps occurs during descent as H_2O jumps from 240 to 18,000 ppm. Thus we choose a linear in the log method for all H_2O gaps, while a simply linear method is used to fill RHw gaps. These are denoted by flag_M(:,12) = flag_M(:,13) = 3.

CO. One task is the creation of a continuous CO record since that species has two well calibrated, nearly continuous measurements. CO can be used to check for unusual or polluted air during the gaps in other species. The primary CO data are from QCLS because it has higher precision and the secondary are from NOAA with a more continuous record but greater noise. This processing of the CO data was done with the full 149,133-parcel dataset, and not the airport-collapsed data set. For the MDS airport-truncated data set, the number of NaN points in CO_NOAA is 8463; that in CO_QCLS is 30,233.

1. Modify CO_QCLS: interpolate short gaps in the CO_QCLS record (≤10 parcels = 100s ~ 1000 m vertically)
2. Create a continuous CO_N record.
    a. Start with CO_NOAA and locate all the NaN gaps.
    b. Fill gaps with modified CO_QCLS where available and locate new NaN gaps.
    c. Average CO for 5 points on either side of gap, interpolate linearly across the gaps.
3. Smooth the CO_N record, which is visibly noisy at 10 s with 11-point running average (~ 1000 m in vertical).
4. Create a continuous CO record.
    a. Define CO = modified CO_QCLS (step 1).
    b. Fill the gaps in CO with CO_N (step 3).
    c. Define CO flags:
        1 = primary, QCLS (116,261 parcels);
        2 = secondary, smoothed CO_N (29428);
        3 = modified, short-interpolated QCLS (80);
        4 = interpolated CO_N (725).

Two samples of this CO interpolation method are shown in Figure S3. The frequency of occurrence of all flags for this new CO_M variable, along with the other MDS chemical variables are given in Table S4.

Short-gap simple interpolation for remaining species. It was decided that the least intrusive method for filling short data gaps was to simply interpolate using only the instrument data. In MDS revisions R0 and R1, CO was used as a proxy to fill these gaps, but later analysis showed little correlation with absolute CO or even the short-term variability in CO. We examined the typical size of gaps and their frequency. For the Mor.all species we selected gaps of $\leq$13 for short-gap interpolation; for WAS the gap frequency peaked about 10 (100 s) and we selected gaps of $\leq$10; for TOGA there was a strong peak at gap length of 7-8 (instrument cycle time) and we also selected $\leq$10 as the criterion. These gaps correspond to about 1000 m or less in the vertical during ascent or descent. For most Mor.all variables this adds about 10% (absolute) to the number of non-NaN parcels, but for WAS and TOGA with many smaller gaps it greatly enhances the coverage. WAS coverage goes from 28% to 41%, while TOGA jumps from 31% to 93% because most gaps are 85 sec. For all short-gap interpolation, the parcel data for that species is tagged with flag = 3.

Long-gap interpolation for remaining species - Troposphere. We choose a robust and minimally intrusive method for filling gaps >10 (100 s) based upon the average tropospheric profile for that flight, using eight 100-hPa-wide bins (<300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, >900 hPa). The gap value is replaced by the appropriate bin value. If any bins have no measured values, we use the nearest bin or average of the nearest bins. It is important not to confuse stratospheric and tropospheric air when gap filling. From our analysis, a number of key reactive species (e.g., CH_2O, HOOH, NOx) show distinctly different patterns as ATom crosses into the stratosphere.
Long-gap interpolation - Stratosphere. We find the most robust definition of stratospheric-like air to be based primarily on H_2O rather than O_3, because O_3 abundances >200 ppb are often seen in large, clearly tropospheric air masses with H_2O > 50 ppm. Based on percentiles of O_3 at different values of H_2O (see Figure S4a) we pick <30 ppm as the criteria for being stratospheric, with the secondary requirements that O_3 > 80 ppb and CO < 120 ppb (see Figure S4b). For the stratospheric air we create mean 'profiles' in terms of 6 O_3 bins (<200, 200-300, 300-400, 400-500, 500-700, >700 ppb) use this as a lookup table for gap filling. There are many fewer stratospheric parcels, and the stratosphere tends to be similar across latitudes, and so we create a single lookup tables from all research flights at all latitudes. In general, these near tropopause air parcels are cold and dry and not highly reactive; however when partitioning the chemistry model calculated reactivities between stratosphere and troposphere, these criteria may need to be re-investigated.

As a measure of the error in this long-gap interpolation, we randomly select 10% of the air parcels from data stream before calculating the long-gap interpolation, interpolate those 10% points, and calculate the mean bias and root-mean-square error (rmse). This is repeated 10 times and we show the average results in Table S5 below. We find these results acceptable, and better than the multiple linear regressions we tried. There may be

a better way to do this in future versions beyond R2, perhaps with a machine-learning approach. Gaps interpolated in this way are given flag = 4 (troposphere tables) and flag = 6 (stratosphere tables).

Missing data for an entire flight. For **ATom-1 RF-5**, an instrument failed and we lost all data for H2O2_M, HNO3_M, and HNO4_M. This flight was from American Samoa to Christ Church. We fill these species using a multiple linear regression from the parallel flights ATom-1 RF-4 and ATom-2/3/4 RF-4/5. The independent (explanatory) variables for the multiple linear regression for these missing flights are chosen to be pressure, noontime solar zenith angle and latitude (in that order). For H2O2_M and HNO3_M, we calculate the missing ATom-1 RF-5 data using the full set of parallel flights, but for but for HNO4_M, we can use only ATom-1/2 flights (see Table S3 & S6). Data filled for missing flights are given flag = 5. For **ATom-2 RF-2**, we also lost all data for H2O2_M, HNO3_M, and HNO4_M. In this case the regression is based on parallel flights ATom-2 RF-3 and ATom-1/3/4 RF-2/3 for H2O2_M and HNO3_M, but only ATom-2 RF-3 and ATom-1 RF-2/3 for HNO4_M. For **ATom-3 RF-1**, we lost all data for NOx_M. A multiple linear regression is based on parallel flights ATom-3 RF-2 and ATom-1/2/4 RF-1/2. For **ATom-3/4 all,** we lost all data for HNO4_M, and the best we can do is base the regression on all HNO4_M measurements (not filled as noted above) from ATom-1/2. This is clearly one of the weakest gap filled here, and users should be careful if key results depend HNO4_M values for ATom-3/4. For **ATom-4 RF-5/6/7/8/9/12/13**, we lost all data for CH3OOH_M. A multiple linear regression approach was based on data from the preceding RF-4 as well as the parallel research flights from the other 3 deployments (i.e., ATom-1/2 RF-5 to 11, ATom-3 RF-5 to 13, ATom-4 RF-4). For **ATom-4 RF-11** (AF 46), all chemical data have flag = 0, value = NaN. A summary of the missing flights and species along with estimated error in our gap filling is given in Table S6.

From the reactivity results for ATom-1 shown in this paper, the lack of ATom-3 NOx observations in the Eastern Pacific (RF 1) mean that the P-O3 statistics there (not calculated in this paper) will not be useful.

S.1.4. Species measured by two instruments

Several species have redundant measurements and these are identified by the duplicate sources in Table S2. The choice of primary (flag = 1) and secondary (flag = 2) are chosen based on continuity of record or coverage of related species, or our estimate of the higher precision measurement. Primary data sources usually have a better data coverage.

CH_4: (1) CH_4_NOAA, (2) CH_4_QCLS. The primary has more data and does not have spurious anomalies (see previous). There is no evident bias, but some scatter, and so the NaNs in the primary record (which first has had short-gap interpolation as noted above) is simply filled with the secondary record (also with short-gap interpolation).

CH_2O: (1) CH_2O[ISAF], (2) CH_2O_TOGA. Formaldehyde is a key reactive species and TOGA provides a secondary record for the 2nd half of ATom-4 when ISAF failed. The

overlapping data with both instruments is plotted in below (Figure S5). The mean difference in overlapping observations is very small (-1 out of a mean of 134 ppt), but the rms is larger (75 ppt). ISAF has a number of values > 1000 ppt, while TOGA has none. A linear fit gives a slope of 0.8 with $R^2 = 0.59$, but a 1:1 slope has only slightly smaller $R^2 = 0.55$. We do not attempt to rescale the TOGA data in this case and just replace any NaNs remaining in the short-gap-interpolated ISAF record (particularly flights 42:48) with TOGA data (also short-gap interpolated).

PAN: (1) PAN_GTCIMS, (2) PAN_PECD*. The GTCIMS joined the mission at ATom-2. The overlap period shows a clear bias between the GTCIMS and PECD observations. A linear fit is clear ($R^2 = 0.84$), and we rescale the secondary PECD* = (PECD + 0.45)/1.18.

C_3H_8: (1) Propane_WAS, (2) Propane_TOGA. No obvious bias is found. A linear fit gives an $R^2 = 0.90$, but the 1:1 line has an $R^2 = 0.85$, so we just use the TOGA data directly as the secondary observation.

iC_4H_{10}: (1) iButane_WAS, (2) iButane_TOGA. No obvious bias is found. A linear fit gives an $R^2 = 0.955$, but the 1:1 line has an $R^2 = 0.947$, so we just use the TOGA data directly as the secondary observation.

nC_4H_{10}: (1) nButane_WAS, (2) nButane_TOGA. No obvious bias is found. A linear fit gives an $R^2 = 0.962$, but the 1:1 line has an $R^2 = 0.942$, so we just use the TOGA data directly as the secondary observation.

C_5H_8: (1) Isoprene_TOGA, (2) Isoprene_WAS. No obvious bias is found. A linear fit gives an $R^2 = 0.938$, but the 1:1 line has an $R^2 = 0.904$, so we just use the WAS data directly as the secondary observation.

benzene: (1) Benzene_TOGA, (2) Benzene_WAS. The is some systematic difference between WAS and TOGA (TOGA = ~0.75 x WAS), but the contribution of WAS to the aromatics is small (see flag=2 is <3% in **Table S4**) and so we did not scale WAS.

toluene: (1) Toluene_TOGA + EthBenzene_TOGA, (2) Toluene_WAS + EthBenzene_WAS. No obvious bias is found in spite of the large scatter. A linear fit gives an $R^2 = 0.75$, but the 1:1 line has an $R^2 = 0.74$, so we just use the TOGA data directly as the secondary observation.

xylene: (1) mpXylene_TOGA + oXylene_TOGA, (2) mpXylene_WAS + oXylene_WAS. No obvious bias is found in spite of the very large scatter. A linear fit gives an $R^2 = 0.3$, so we just use the WAS data directly as the secondary observation.

HCN: (1) HCN_CIT, (2) HCN_TOGA. The CIT observation is chosen as primary because of its more continuous, 10s record. In spite of the large scatter, a linear fit with a slope of 0.8 does not greatly reduce the variance ($R^2 = 0.74$ vs 0.65 for 1:1), so we just use the TOGA data directly as the secondary observation.

SF6: (1) SF6_PECD, (2) SF6_UCATS. The scatter seems large, but the relationship is mostly 1:1 with $R^2 = 0.90$. A linear fit gives a slope of 0.99, and so we just use the UCATS data directly as the secondary observation. Both data sets are sparse.

S.1.5. Seasonaillty of ATom 1-4 data

As a quick look at the opportunities provided by the ATom data, we consider two examples. Figure S6 shows that the 4-season transects of ATom (deployments 1, 2, 3 and 4) produced remarkably similar patterns of the 2-D PDs for HOOH versus NO_x in Central Pacific, providing a useful benchmark for the modeling community. The ellipse fits show almost identical overlap for ATom 2-3-4 and overlap with ATom-1 except for regions of very low (2-10 ppt) NO_x. The implication is that the chemical patterns of the tropical Central Pacific are represented by a single transect and do not change much seasonally.

S.2. The Reactivity Data Stream

In this paper, we use the 6 models for their August chemical statistics, and use 5 of them plus a box model to calculate the reactivities (i.e., chemical tendencies) from the ATom-1 MDS, see Table S7.

In the models, a grid cell is initialized with all the core reactive species needed for a regular chemistry simulation. The model is then integrated over 24 hours without transport or mixing, without scavenging, and without emissions. Each model uses its own varying cloud fields for the period to calculate photolysis rates; and the reactions rates are integrated. The F0AM box model simply takes the instant J-values as measured on the flight and applies a diurnal scaling. We calculate the three major reactivities (Rs) from the rates.

L-CH4:	$CH_4 + OH \rightarrow CH_3 + H_2O$	(R1)
P-O3:	$HO_2 + NO \rightarrow NO_2 + OH$	(R2a)
	$RO_2 + NO \rightarrow NO_2 + RO$	(R2b)
L-O3:	$O_3 + OH \rightarrow O_2 + HO_2$	(R3a)
	$O_3 + HO_2 \rightarrow HO + O_2 + O_2$	(R3b)
	$O(^1D) + H_2O \rightarrow OH + OH$	(R3c)

The peroxy radical RO_2 is used to represent all species like CH_3O_2, $C_2H_5O_2$, and more complex organic species. While the net P-O3 minus L-O3 can involve other reactions, particularly with NOx and unsaturated organics in polluted environments, Prather et al. (2017) showed that the difference in these two reactivities accurately described the O_3 tendencies in most regions, including aged pollution plumes over the ocean basins. Two photolysis rates (J-values) are linked strongly with the reactivities, and their 24-hr

averages are also included in the model-derived RDS because they provide useful diagnostics:

J-O1D: $O_3 + h\nu \rightarrow O(^1D) + O_2$ (R4a)

J-NO2: $NO_2 + h\nu \rightarrow NO + O$ (R4b)

Prather et al. (2018) showed that we can initialize with the core species and let the radicals (OH, HO_2, RO_2) come into photochemical balance. Prather et al. (2018) also found that the 24-hour integration using the synthetic MDS along the Dateline was not overly sensitive to the initialization time (at most 4% in P-O3, 1% in L-O3 and L-CH4), and thus models do not have to synchronize with the local time of observation (see their Figure S8 and their Table S8). The summary statistics of these reactivities and J-values from the six models using MDS_R0, from 3 different years with the UCI model, from GMI using MDS_R1, and from UCI using MDS_R2 are given in Table S8. The ATom-1 data are sorted for the Pacific and Atlanitc basisn as well as global. We show: means, medians and mean of the top 10% of reactive parcels; percent of total reactivity in the top 50%, top 10% and top 3%; and mean J-values.

Variations in reactivities dues to clouds is an irreducible source of uncertainty: predicting the cloud-driven photolysis rates that a shearing air parcel will experience over 24 hours is not possible here. The protocol asks models to sample 5 separated days during the deployment month (e.g., August 1, 6, 11, 16, 21 for ATom-1) to average over synoptically varying cloud conditions. The averaged standard deviation (σ) of reactivity over the 5 days is calculated in % of the mean. In Table S9, it is about 10% for the 3 similar CTMs (GC, GMI, UCI) but twice as large for the 2 CCMs (GISS, NCAR). For the 3 CTMs, the σ of the J-values is similar (~10%), as expected because variations in the J's via clouds drives the reactivities. Results from the UCI CTM running different years show the same 10% level. For GISS and NCAR, the results are more cryptic: J-value σ are larger (12 – 17%); but the reactivity σ are larger still (14 – 32%). We have not resolved these differences. The F0AM box model does not include clouds directly but calculates its own clear-sky J-values and scales the diurnal cycle to the single observed J-value at the time of measurement, thus maintaining the same proportional cloud effects over all daytime hours.

One-dimensional probability density functions (PDs) for the 3 Rs along the Pacific and Atlantic transects of ATom-1 are presented in **Figure S7**. These are the 54°S-60°N oceanic measurements. The mean values for the six models (colors, "a day in mid-August" for the blocks shown in **Figure S1**) and the ATom-1 flights (black, UCI model, MDS_R2) are shown in the figure legend. For the Pacific, the high reactivities in the Eastern Pacific (10°N-30°N) do not substantially alter the PD as seen by comparing the whole Pacific (solid black) with just the Central Pacific (dashed gray), but they do shift the mean reactivities upward by 10-20%. Occurrences beyond the uppermost bin shown (6 ppb/d for P-O3, L-O3; 3 ppb/d for L-CH4) are included in that bin and result in an obvious uptick for P-O3 and L-O3. This uptick is not seen for the Central Pacific alone (gray circles), indicating that the extreme reactivities occur in the Eastern Pacific (Figure 2 in main text).

For P-O3, the difference between the five of the six global chemistry models and ATom is distinct in both basins. ATom shows a large occurrence of moderately reactive air with P-O3 > 2 ppb/d in both Pacific (48%) and Atlantic (54%). The models show much less of such air in both Pacific (6-10%) and Atlantic (17-30%). Thus, the ATom chemical mix of reactive species is far more effective in producing O_3 than that in most all CTM/CCMs. This result holds for both Pacific and Atlantic transects and thus is unlikely a result of biased ATom sampling. We conclude that our models have a fundamental flaw in O_3 production over the major ocean basins. For both L-O3 and L-CH4, both ATom and models have similar mean values and their PDs are remarkably similar. For L-CH4, one can see a slight systematic shift between the two, with ATom having lower occurrence in the 0.5-1.5 ppb/d range, but greater in the 1.5-2.5 ppb/d range. Generally, five models (GC, GFDL, GMI, NCAR, UCI) show a similar pattern and mean values for all 3 Rs and both ocean basins, while the sixth model (GISS) is inexplicable as noted before.

Also, the ability to test the model's reactivity statistics with the ATom 10 s data is not obvious. For example, in Figure S8, we take the Pacific P-O3 frequency (black), generate a random set of points from that distribution, and then start averaging adjacent points in groups (2, 4, 8; denoted as 4 km, 8km, 16 km). The resulting statistics rapidly evolve into a Gaussian-like distribution about the mean value. Thus, the ability to nearly match the ATom-1 statistics with our global chemistry models is significant, and we cannot explain the P-O3 discrepancy as a model-averaging problem.

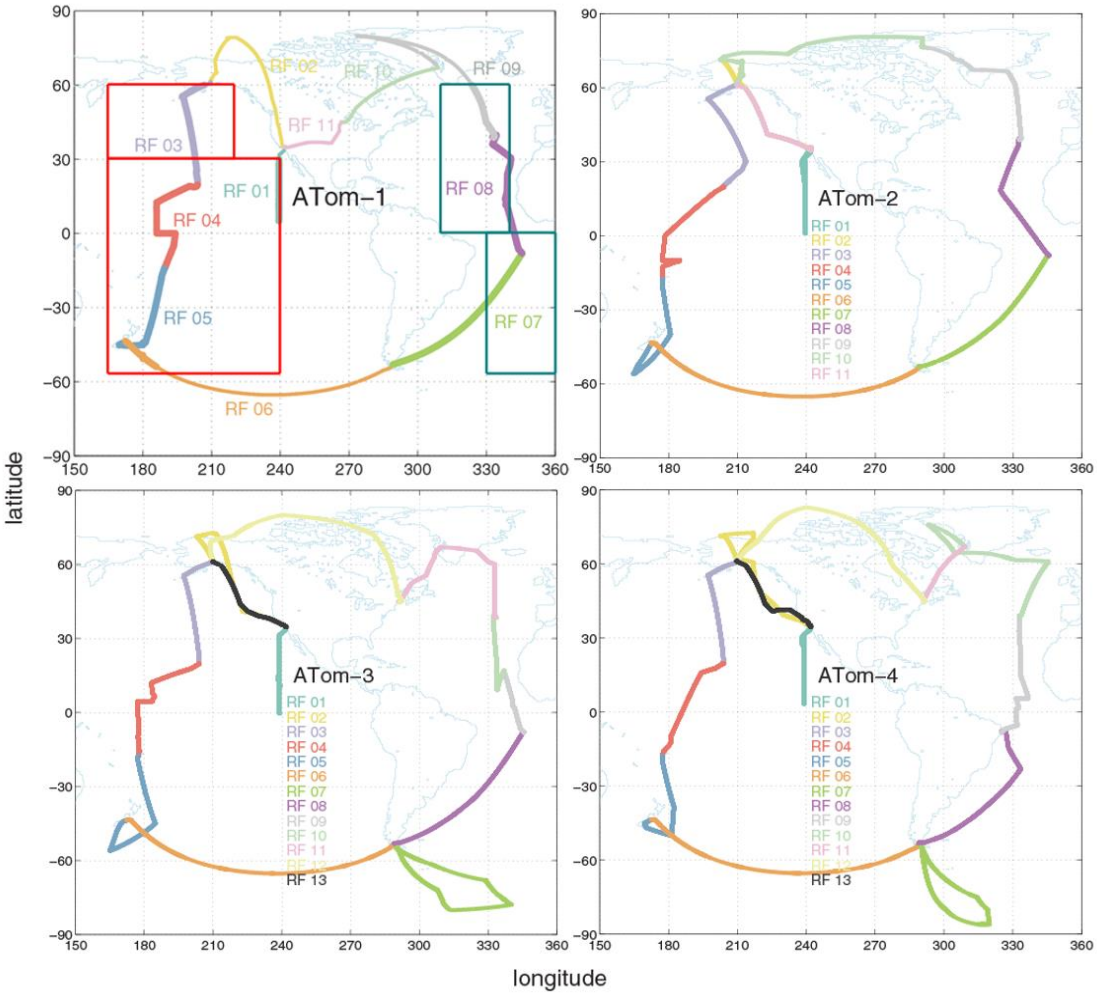

Figure S1. Flight tracks for the 4 ATom deployments. For ATom-1, the flight segments considered Pacific and Atlantic domains are shown with very thick lines. The corresponding blocks used for model climatologies are outlined with rectangles: Pacific, red; Atlantic, blue-green.

530

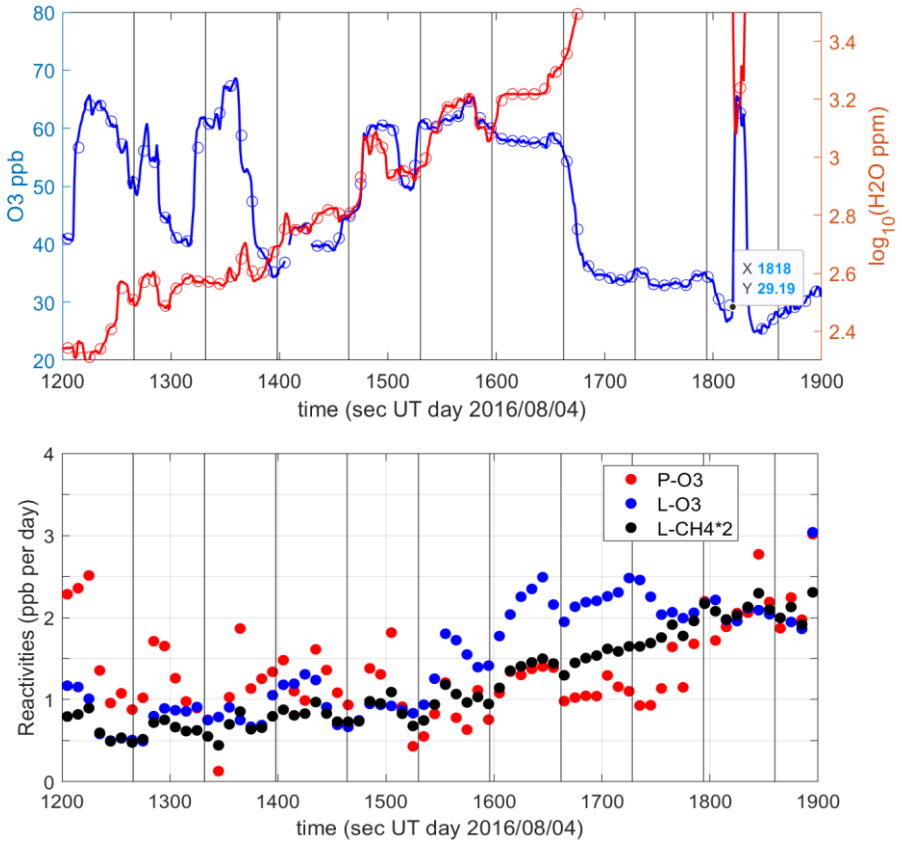

Figure S2. Profile during a descent on the Anchorage-Kona flight (ATom-1, RF-3, 31°N). The profile here begins at 7.2 km (1200 s) and ends at 2.1 km (1900 s, H_2O is cut off). (a) Fine structure in O_3 (ppb) and H_2O (log_{10}, ppm) at 1-sec (solid line) and 10 s (open circles) resolution. (b) Reactivities for the 10 s parcels calculated with the UCI CTM. Descent rate averaged 7.5 m/s, and vertical lines indicate 500 m thick layers.

531
532
533
534

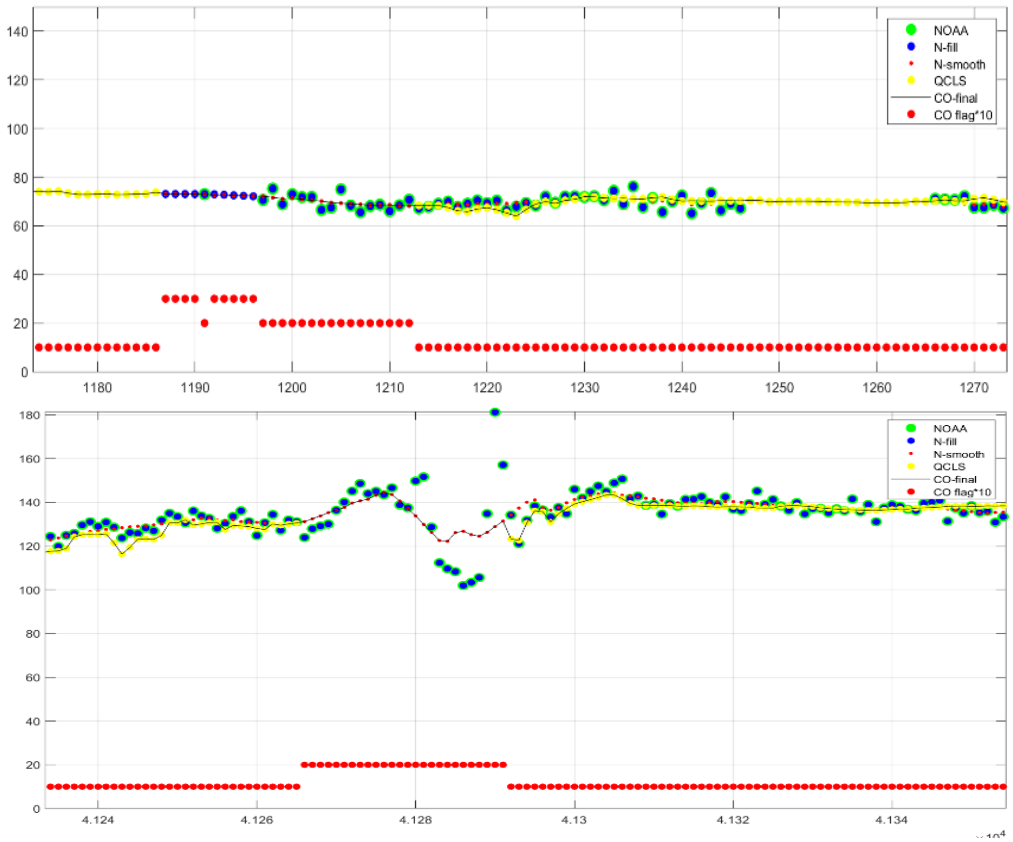

Figure S3. Example of CO time series showing all the intermediate CO products and flags. See legend and text.

539

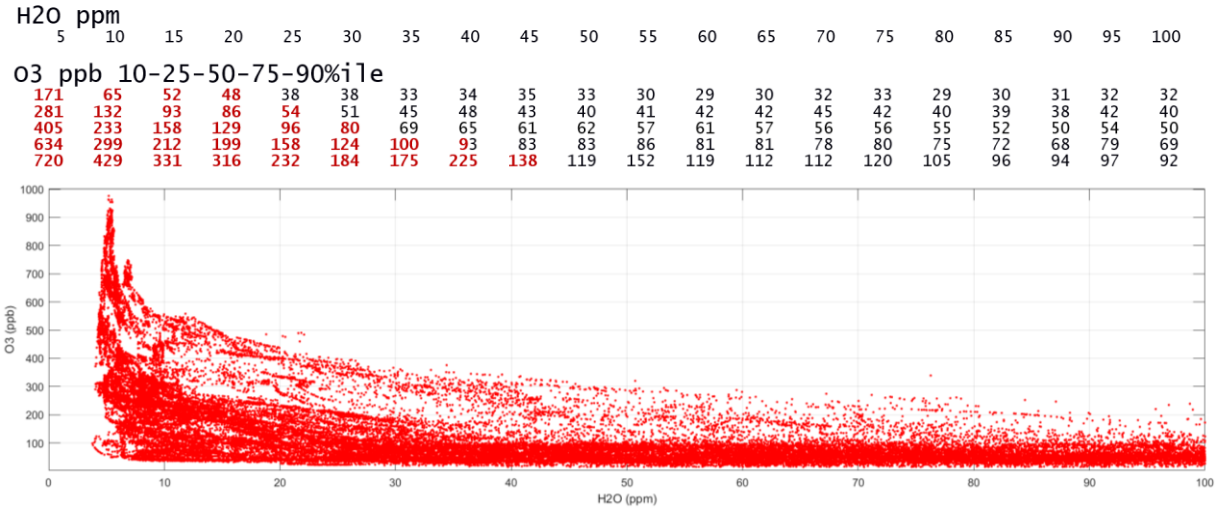

Figure S4a. Scatter plot of O_3 (ppb) and H_2O (ppm) for all ATom deployments, filtered by H_2O < 100 ppm. The percentiles (10-25-50-75-90 %ile) of O_3 in each 5-ppm-wide bin starting at 5 ppm (= 2.5–7.5 ppm) ending at 100 pm in in the table at the top of this figure. Stratospheric influence (red) is clearly seen in the median for <30 ppm.

540
541
542

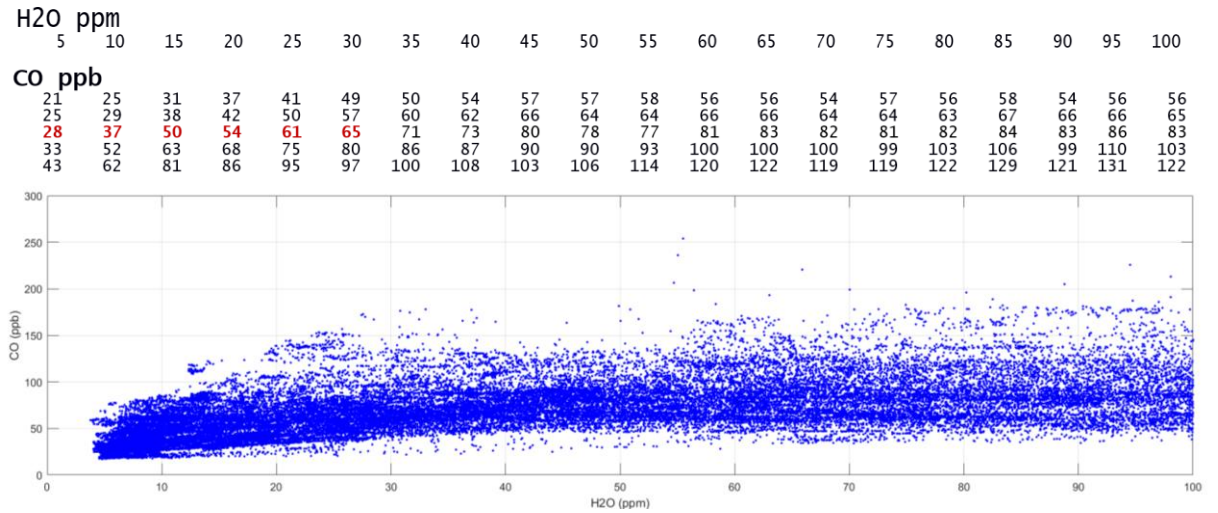

Figure S4b. Scatter plot of CO (ppb) and H_2O (ppm) for all ATom deployments, filtered by H_2O < 100 ppm. See Figure S4a.

543
544
545

546

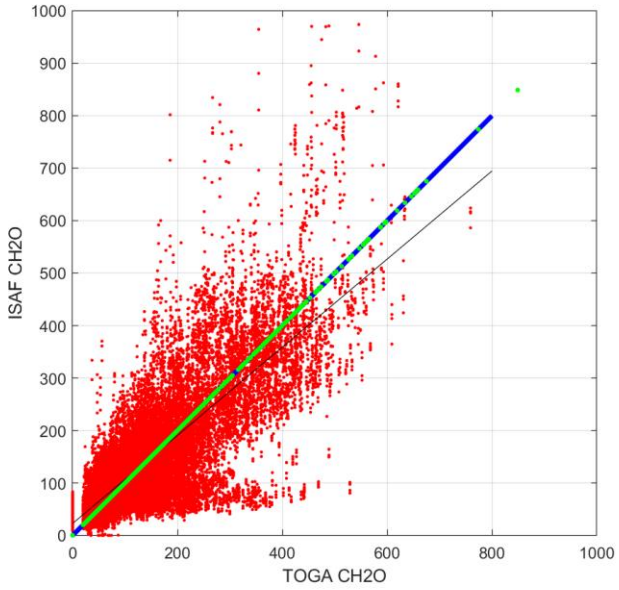

Figure S5. Scatter plot of coincident HCHO measurements (ppb) from ISAF and TOGA for all ATom deployments. The thick blue-green line is the 1:1 relationship and the thin black line shows a linear regression of ISAF vs. TOGA. Notably, ISAF has more frequent high values >600 ppb, with some above 1000 ppb (not shown).

547
548
549

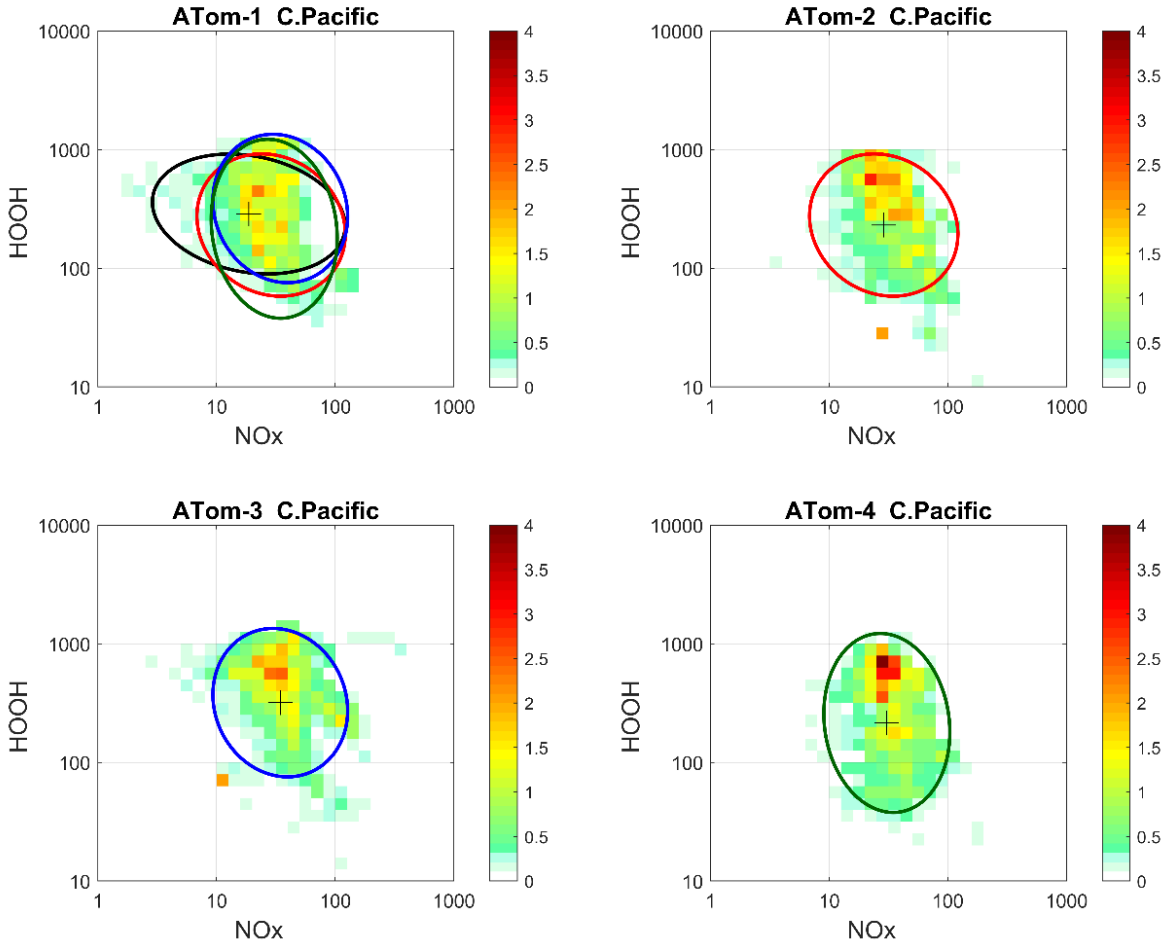

Figure S6. 2D frequency of occurrence (PDs in log ppt mole fraction) of HOOH vs. NOx for the tropical Central Pacific for all 4 ATom deployments. The cross marks the mean (in log space), and the ellipse is fitted to the rotated PD to have the smallest semi-minor axis. The semi-minor and semi-major axes are 2 standard deviations of PD in that direction. The ellipses from ATom-2 (red), ATom-3 (blue), and ATom-4 (dark green) are also plotted in the ATom-1 quadrant.

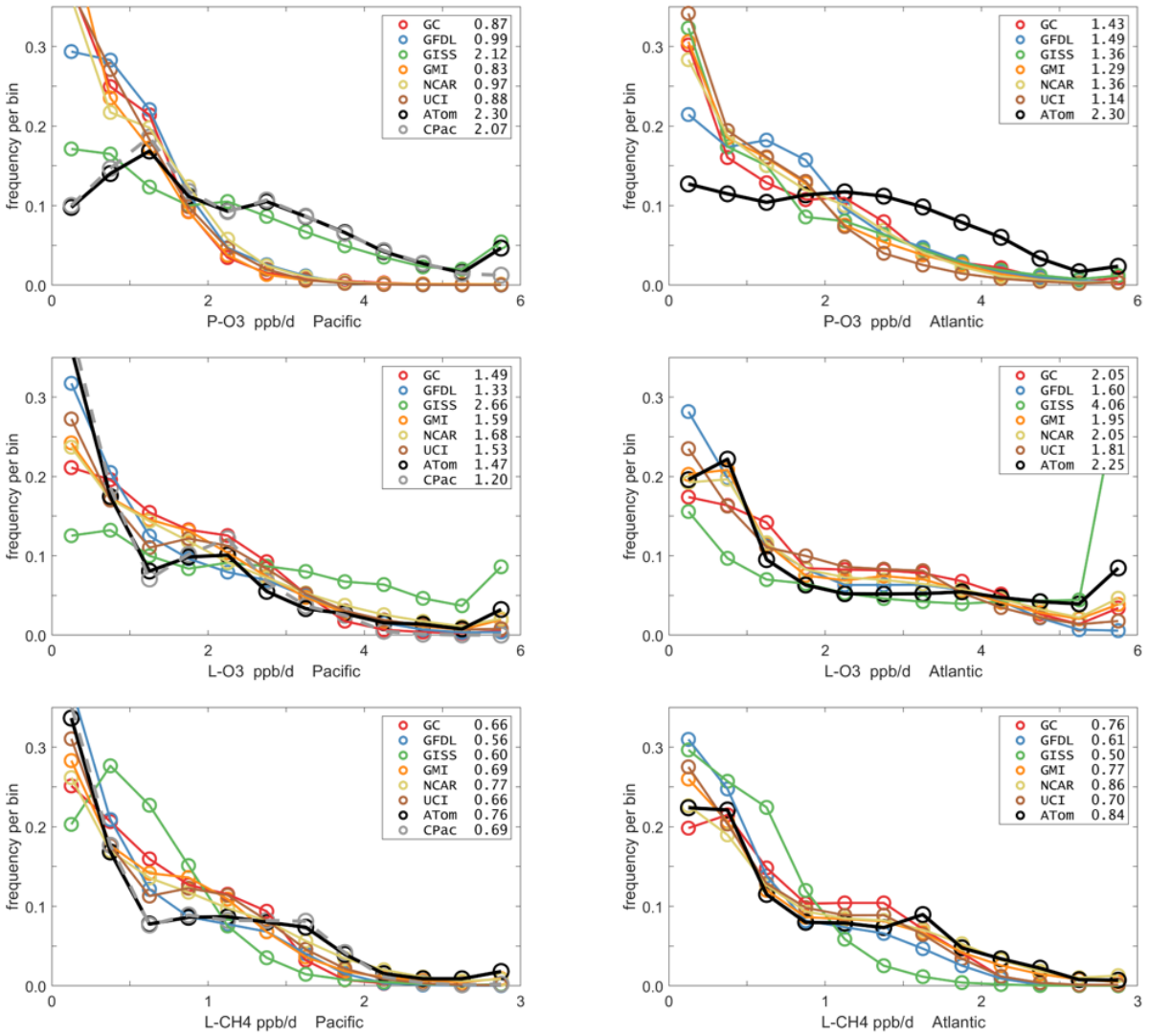

Figure S7. Probability distributions (frequency of occurrence) for the ATom-1 3 reactivities (rows: P-O3, L-O3, L-CH4 in ppb/day) and for the Pacific and Atlantic from 54°S to 60°N (columns left and right). Each air parcel is weighted as described in the text for equal frequency in large latitude-pressure bins, and also by cosine(latitude). The ATom statistics are for UCI model and ATom-1 MDS_R2. The full Pacific results (solid black) also include just the Central Pacific (dashed gray). The 6 models' values for a day in mid-August are averaged over longitude for the domains shown in Figure S1, and then cosine(latitude) weighted. Mean values (ppb/day) are shown in the legend.

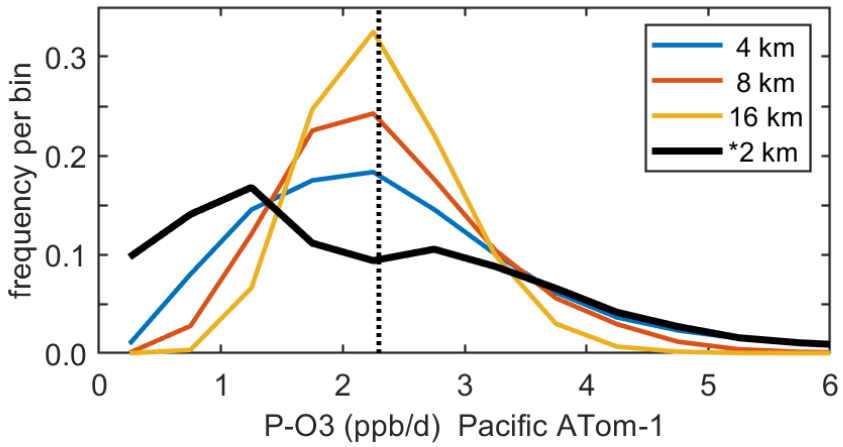

Figure S8. Frequency of occurrence of P-O3 (ppb/d) per 0.5 ppb/d bins. The black curve shows the ATom-1 results for the Pacific Ocean (54°S-60°N) with each air parcel weighted as described in the text. The 10 s parcels have a resolution of about 2 km horizontally. A large random time series (10^5 parcels) was generated from the ATom frequency curve, and then smoothed over 2, 4, and 8 sequential points. If the successive ATom parcels were randomly distributed, then smoothing over scales of ~10 km rapidly changes the frequency into a Gaussian centered on the mean value (vertical dotted line). The heterogeneity scales in P-O3 are clearly much larger than 2 km as seen in **Figure S2**. Further, the ability of models with 100 km scales to reproduce the frequency of occurrence seen in the ATom parcels indicates that the heterogeneity is large scale and being resolved by the models, not average over.

562

Table S1. ATom flight data								
ATom research flights in the Mor.2020-05-27…tbl (149,133 parcels)							Airport removed (146,494 parcels)	
ATom deployment	Research Flight no.	ATom flight	Airports	parcel begin	parcel end	YYYYMMDD	parcel begin	parcel end
1	1	1	PMD PMD*	1	3380	20160729	1	3333
1	2	2	PMD ANC	3381	7038	20160801	3334	6939
1	3	3	ANC KOA	7039	9658	20160803	6940	9526
1	4	4	KOA PPG	9659	12760	20160806	9527	12583
1	5	5	PPG CHC	12761	15141	20160808	12584	14917
1	6	6	CHC PUQ	15142	18976	20160812	14918	18692
1	7	7	PUQ ASI	18977	22355	20160815	18693	21998
1	8	8	ASI TER	22356	25431	20160817	21999	25040
1	9	9	TER SFJ	25432	28976	20160820	25041	28544
1	10	10	SFJ MSP	28977	31127	20160822	28545	30663
1	11	11	MSP PMD	31128	32899	20160823	30664	32383
2	1	12	PMD PMD*	32900	36621	20170126	32384	36061
2	2	13	PMD ANC	36622	40115	20170129	36062	39480
2	3	14	ANC KOA	40116	43062	20170201	39481	42360
2	4	15	KOA NAN	43063	46470	20170203	42361	45717
2	5	16	NAN CHC	46471	49562	20170205	45718	48774
2	6	17	CHC PUQ	49563	53116	20170210	48775	52267
2	7	18	PUQ ASI	53117	56358	20170213	52268	55390
2	8	19	ASI TER	56359	59468	20170215	55391	58446
2	9	20	TER THU	59469	62151	20170218	58447	61088
2	10	21	THU ANC	62152	64893	20170219	61089	63762
2	11	22	ANC PMD	64894	66978	20170221	63763	65807
3	1	23	PMD PMD*	66979	70683	20170928	65808	69465
3	2	24	PMD ANC	70684	74281	20171001	69466	73001
3	3	25	ANC KOA	74282	76949	20171004	73002	75608
3	4	26	KOA NAN	76950	80163	20171006	75609	78754
3	5	27	NAN CHC	80164	83472	20171008	78755	82000
3	6	28	CHC PUQ	83473	87028	20171011	82001	85462
3	7	29	PUQ PUQ^	87029	90872	20171014	85463	89225
3	8	30	PUQ ASI	90873	94279	20171017	89226	92576
3	9	31	ASI SID	94280	95928	20171019	92577	94191
3	10	32	SID TER	95929	98695	20171020	94192	96916
3	11	33	TER BGR	98696	102094	20171023	96917	100272
3	12	34	BGR ANC	102095	105540	20171025	100273	103677
3	13	35	ANC PMD	105541	107873	20171027	103678	105983
4	1	36	PMD PMD*	107874	111294	20180424	105984	109357
4	2	37	PMD ANC	111295	115012	20180427	109358	113028
4	3	38	ANC KOA	115013	117934	20180429	113029	115847
4	4	39	KOA NAN	117935	120880	20180501	115848	118741
4	5	40	NAN CHC	120881	123717	20180503	118742	121542
4	6	41	CHC PUQ	123718	127370	20180506	121543	125122
4	7	42	PUQ PUQ^	127371	131238	20180509	125123	128934
4	8	43	PUQ REC	131239	134829	20180512	128935	132463
4	9	44	REC TER	134830	138214	20180514	132464	135770
4	10	45	TER SFJ	138215	141697	20180517	135771	139210
4	11	46	SFJ BGR	141698	142846	20180518	139211	140316
4	12	47	BGR ANC	142847	146670	20180519	140317	144095
4	13	48	ANC PMD	146671	149133	20180521	144096	146494
* 4 flights to equator following 120W. ^ 2 flights to 80S and 86S over Antarctica.								

563
564

565
566

Table S2. MDS data and source			
id#	**MDS data designation**	**Description**	**ATom source name**
1	parcel_M	Unique sequential parcel number for all MDS 10s data, beginning 1,000,001	
2	ATno	ATom deployment number (1:4)	A.no
3	RFno	Research Flight number (1:11, 1:11, 1:13, 1:13)	RF
4	RRno	RF number across all of ATom (1:48)	
5	YYYMMDD	Date (UT) of the start of each RF	YYYYMMDD
6	UTC_M	Start time in sec relative to Date for each 10s parcel	UTC_Start
7	Lat_M	Latitude (-90:+90)	G_LAT
8	Lng_M	Longitude (-180:+180)	G_LONG
9	Alt_M	Altitude (m above mean sea level)	G_ALT
10	P_M	Pressure (hPa)	P
11	T_M	Temperture (K)	T
12	H2O_M	water, ppm (all dry air mole fraction)	H2O_DLH
13	RHw_M	relative humidity over liquid water (%)	RHw_DLH
14	O3_M	ozone, ppb	O3_CL
15	CO_M	carbon monoxide, ppb	(1) CO_QCLS, (2) CO_NOAA
16	CH4_M	methane, ppb	(1) CH4_NOAA, (2) CH4_QCLS
17	NOx_M	odd-nitrogen, NO+NO2, ppt	NO_CL + NO2_CL
18	NOxPSS_M	odd-nitrogen, with photo-stationary state NO2, ppt	NOx_PSS
19	HNO3_M	nitric acid, HONO2, ppt	HNO3_CIT
20	HNO4_M	pernitric acid, HO2NO2, ppt	PNA_CIT
21	PAN_M	peroxyacetyl nitrate, C2H3NO5 - CH3C(O)OONO2, ppt	(1) PAN_GTCIMS, (2) PAN_PECD*
22	CH2O_M	formaldehyde, HCHO, ppt	(1) CH2O[_ISAF], (2) CH2O_TOGA
23	H2O2_M	hydrogen peroxide, HOOH, ppt	H2O2_CIT
24	CH3OOH_M	methyl hydrogen peroxide, ppt	MHP_CIT
25	Acetone_M	acetone, CH3C(O)CH3, ppt	Acetone_TOGA
26	Acetald_M	acetaldehyde, CH3C(O)H, ppt	CH3CHO_TOGA
27	C2H6_M	ethane, C2H6, ppt	Ethane_WAS
28	C3H8_M	propane, C3H8, ppt	(1) Propane_WAS, (2) Propane_TOGA
29	iC4H10_M	iso-butane, iC4H10, ppt	(1) iButane_WAS, (2) iButane_TOGA
30	nC4H10_M	n-butane, nC4H10, ppt	(1) nButane_WAS, (2) nButane_TOGA
31	Alkanes_M	pentane (C5H12) and higher, ppt	iPentane_WAS + nPentane_WAS + nHexane_WAS + nHeptane_WAS + x2MePentane_WAS + x3MePentane_WAS
32	C2H4_M	ethene, C2H4, ppt	Ethene_WAS
33	Alkenes_M	propene (C3H6) and higher, ppt	Propene_WAS
34	C2H2_M	acetylene (ethyne), C2H2, ppt	Ethyne_WAS
35	C5H8_M	isoprene, C5H8, ppt	(1) Isoprene_TOGA, (2) Isoprene_WAS

36	Benzene_M	benzene, C6H6, ppt	(1) Benzene_TOGA, (2) Benzene_WAS*
37	Toluene_M	methylbenzene, C7H8, ppt	(1) Toluene_TOGA+EthBenzene_TOGA, (2) Toluene_WAS + EthBenzene_WAS
38	Xylene_M	dimethylbenzene, C8H10, ppt	(1) mpXylene_TOGA+oXylene_TOGA, (2) mpXylene_WAS+oXylene_WAS
39	MeONO2_M	methyl nitrate, CH3ONO2, ppt	MeONO2_WAS
40	EtONO2_M	ethyl nitrate, CH3ONO2, ppt	EthONO2_WAS
41	RONO2_M	higher organo nitrates, R=C3+, ppt	iPropONO2_WAS + nPropONO2_WAS + x2ButONO2_WAS + x3PentONO2_WAS + x2PentONO2_WAS + x3Me2ButONO2_WAS
42	MeOH_M	methanol, CH3OH, ppt	CH3OH_TOGA
43	HCN_M	hydrogen cyanide, ppt	(1) HCN_CIT, (2) HCN_TOGA
44	CH3CN_M	acetonitrile (methyl cyanide), CH3CN, ppt	CH3CN_TOGA
45	SF6_M	sulfure hexafluoride, ppt	(1) SF6_PECD, (2) SF6_UCATS
46	S_nuc_M	particle surface area (um^2/cm^3), nucleation: 0.0027 < Dp <= 0.012 um	S_nucl_AMP
47	S_atk_M	particle surface area (um^2/cm^3), Aitken: 0.012 < Dp <=0.06 um	S_aitken_AMP
48	S_acc_M	particle surface area (um^2/cm^3), accumulation: 0.06 < Dp <=0.50 um	S_accum_AMP
49	S_crs_M	particle surface area (um^2/cm^3), coarse: 0.50 < Dp <=4.8 um	S_coarse_AMP
50	CloudInd_M	cloud indicator (0:4), dimensionless	cloudindicator_CAPS
Note: The flag value, flag_M(:,1:50) is indexed to the 50 variables above. Only flag_M(:,10:50) have meaningful values. The flag values are: 0 (NaNs, only in research flight 46), 1 (primary data), 2 (secondary data), 3 (short-gap interpolation), 4 (long-gap interpolation for troposphere), 5 (missing flight filled) and 6 (long-gap interpolation for stratosphere) are described in text.			

567
568
569

570

Table S3a. ATom-1, % of non-NaNs after short-gap interpolation

RRno	1	2	3	4	5	6	7	8	9	10	11
<Lat> (deg)	20	62	42	4	-34	-58	-32	18	65	55	38
<Lng> (deg)	-120	-133	-158	-169	-83	-87	-37	-21	-49	-78	-104
<Alt> (m)	7055	8092	7118	6143	6634	7034	6761	6494	6930	6090	7736
# parcels	**3333**	**3606**	**2587**	**3057**	**2334**	**3775**	**3306**	**3042**	**3504**	**2119**	**1720**
H2O_M	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
RHw_M	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
O3_M	99%	99%	100%	99%	100%	100%	99%	100%	100%	100%	100%
CO_M	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
CH4_M	54%	95%	95%	94%	86%	93%	94%	92%	95%	95%	93%
NOx_M	90%	94%	91%	84%	91%	85%	96%	98%	89%	95%	94%
NOxPSS_M	94%	91%	91%	86%	88%	28%	67%	95%	88%	95%	92%
HNO3_M	92%	96%	97%	92%	**0%**	95%	95%	97%	96%	97%	97%
HNO4_M	59%	87%	74%	67%	**0%**	90%	85%	67%	88%	73%	66%
PAN_M	78%	67%	48%	90%	40%	87%	97%	93%	98%	92%	95%
CH2O_M	99%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
H2O2_M	92%	96%	97%	92%	**0%**	95%	95%	97%	96%	97%	97%
CH3OOH_M	56%	69%	81%	83%	84%	79%	81%	82%	82%	79%	79%
Acetone_M	89%	92%	88%	98%	92%	90%	93%	94%	94%	94%	94%
Acetald_M	89%	92%	88%	98%	90%	90%	90%	94%	93%	94%	94%
C2H6_M	50%	32%	43%	44%	62%	37%	39%	43%	40%	46%	45%
C3H8_M	90%	95%	92%	97%	97%	95%	96%	97%	96%	98%	95%
iC4H10_M	95%	95%	92%	98%	97%	95%	96%	97%	98%	98%	96%
nC4H10_M	95%	95%	92%	98%	97%	95%	96%	97%	98%	98%	96%
Alkanes_M	50%	32%	43%	44%	62%	37%	39%	43%	40%	46%	45%
C2H4_M	50%	32%	43%	44%	62%	37%	39%	43%	40%	46%	45%
Alkenes_M	50%	32%	43%	44%	62%	37%	39%	43%	40%	46%	45%
C2H2_M	50%	32%	43%	44%	62%	37%	39%	43%	40%	46%	45%
C5H8_M	95%	95%	92%	98%	97%	95%	96%	97%	98%	98%	96%
Benzene_M	95%	95%	92%	98%	97%	95%	96%	97%	98%	98%	96%
Toluene_M	100%	99%	94%	98%	98%	99%	100%	99%	100%	100%	99%
Xylene_M	100%	99%	94%	98%	98%	99%	100%	99%	100%	100%	99%
MeONO2_M	50%	32%	43%	44%	55%	37%	39%	43%	33%	43%	43%
EtONO2_M	50%	31%	40%	43%	47%	28%	34%	42%	31%	39%	39%
RONO2_M	50%	32%	43%	44%	62%	37%	39%	43%	40%	46%	45%
MeOH_M	89%	92%	88%	98%	92%	90%	92%	92%	92%	94%	94%
HCN_M	98%	100%	100%	100%	92%	100%	100%	100%	100%	100%	100%
CH3CN_M	89%	92%	88%	98%	92%	90%	93%	94%	94%	94%	91%
SF6_M	90%	88%	98%	92%	91%	80%	96%	79%	99%	90%	84%
S_nuc_M	95%	92%	93%	99%	92%	87%	91%	94%	91%	88%	93%
S_atk_M	95%	92%	93%	99%	92%	87%	91%	94%	91%	88%	93%
S_acc_M	95%	92%	93%	99%	92%	87%	91%	93%	91%	88%	93%
S_crs_M	95%	92%	93%	99%	92%	87%	91%	93%	91%	88%	93%
CloudInd_M	100%	100%	100%	100%	99%	100%	100%	100%	100%	100%	100%

571

572

Table S3b. ATom-2, % of non-NaNs after short-gap interpolation											
RRno	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**
<Lat> (deg)	18	55	40	0	-41	-58	-32	15	60	73	45
<Lng> (deg)	-120	-142	-154	-46	138	-89	-37	-28	-38	-129	-135
<Alt> (m)	8477	6915	5726	7514	7233	7629	8835	6832	5869	5553	6969
# parcels	**3678**	**3419**	**2880**	**3357**	**3057**	**3493**	**3123**	**3056**	**2642**	**2674**	**2045**
H2O_M	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
RHw_M	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
O3_M	99%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
CO_M	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
CH4_M	100%	100%	100%	99%	98%	99%	100%	100%	100%	99%	100%
NOx_M	85%	89%	100%	95%	82%	82%	87%	80%	82%	100%	96%
NOxPSS_M											
HNO3_M	90%	**0%**	91%	95%	96%	92%	97%	97%	97%	93%	98%
HNO4_M	82%	**0%**	77%	70%	77%	81%	87%	77%	87%	93%	94%
PAN_M	84%	100%	100%	95%	100%	100%	99%	97%	94%	100%	94%
CH2O_M	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
H2O2_M	90%	**0%**	91%	95%	96%	92%	97%	97%	97%	93%	98%
CH3OOH_M	67%	62%	71%	67%	65%	58%	58%	59%	58%	60%	56%
Acetone_M	91%	92%	85%	97%	96%	93%	95%	96%	89%	91%	94%
Acetald_M	91%	92%	85%	97%	96%	93%	95%	97%	89%	91%	94%
C2H6_M	38%	28%	45%	36%	42%	43%	40%	47%	56%	58%	61%
C3H8_M	95%	88%	81%	94%	94%	93%	87%	87%	87%	58%	88%
iC4H10_M	97%	94%	91%	97%	97%	95%	95%	97%	94%	95%	97%
nC4H10_M	97%	94%	91%	97%	97%	95%	95%	97%	94%	95%	97%
Alkanes_M	38%	28%	45%	36%	42%	43%	40%	47%	56%	58%	61%
C2H4_M	38%	28%	45%	36%	42%	43%	40%	47%	56%	58%	61%
Alkenes_M	38%	28%	45%	36%	42%	43%	40%	47%	56%	58%	61%
C2H2_M	38%	28%	45%	36%	42%	43%	40%	47%	56%	58%	61%
C5H8_M	97%	94%	93%	97%	97%	95%	96%	98%	94%	96%	97%
Benzene_M	97%	94%	93%	97%	97%	95%	96%	98%	94%	96%	97%
Toluene_M	100%	96%	96%	100%	100%	100%	100%	100%	98%	100%	100%
Xylene_M	100%	96%	96%	100%	100%	100%	100%	100%	98%	100%	100%
MeONO2_M	37%	26%	45%	36%	38%	35%	40%	47%	53%	54%	51%
EtONO2_M	37%	26%	45%	36%	38%	35%	40%	47%	52%	54%	50%
RONO2_M	38%	28%	45%	36%	42%	43%	40%	47%	56%	58%	61%
MeOH_M	90%	92%	83%	97%	92%	93%	95%	97%	89%	91%	94%
HCN_M	99%	89%	100%	100%	100%	98%	100%	100%	100%	93%	100%
CH3CN_M	91%	92%	85%	97%	96%	93%	95%	97%	89%	87%	94%
SF6_M	87%	97%	96%	88%	98%	99%	98%	99%	99%	99%	69%
S_nuc_M	86%	81%	98%	95%	85%	95%	85%	98%	75%	89%	91%
S_atk_M	86%	81%	98%	95%	85%	95%	85%	98%	75%	89%	91%
S_acc_M	86%	81%	97%	95%	84%	95%	85%	98%	75%	88%	91%
S_crs_M	86%	81%	97%	95%	84%	95%	85%	98%	75%	88%	91%
CloudInd_M	100%	100%	100%	97%	100%	100%	100%	100%	100%	100%	100%

573

574

Table S3c. ATom-3, % of non-NaNs after short-gap interpolation													
RRno	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**
<Lat> (deg)	18	55	42	4	-41	-58	-67	-32	4	22	55	67	46
<Lng> (deg)	-121	-141	-158	-14	63	-91	-50	-36	-19	-26	-43	-105	-136
<Alt> (m)	8988	7623	6720	6781	6844	6836	7263	8169	6678	6329	5522	6231	6033
# parcels	**3658**	**3536**	**2607**	**3146**	**3246**	**3462**	**3763**	**3351**	**1615**	**2725**	**3356**	**3405**	**2306**
H2O_M	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
RHw_M	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
O3_M	99%	100%	100%	100%	100%	100%	100%	100%	89%	99%	100%	100%	100%
CO_M	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
CH4_M	100%	98%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
NOx_M	**0%**	98%	100%	100%	97%	100%	87%	94%	89%	94%	99%	100%	100%
NOxPSS_M													
HNO3_M	96%	96%	96%	95%	97%	91%	94%	96%	91%	85%	97%	90%	66%
HNO4_M	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**
PAN_M	100%	100%	100%	98%	100%	100%	99%	99%	100%	98%	100%	98%	100%
CH2O_M	100%	100%	100%	100%	100%	100%	98%	100%	100%	100%	100%	100%	100%
H2O2_M	96%	96%	96%	95%	97%	91%	94%	96%	91%	85%	97%	90%	95%
CH3OOH_M	61%	59%	59%	60%	58%	58%	59%	61%	58%	53%	67%	60%	64%
Acetone_M	94%	95%	87%	95%	96%	97%	92%	96%	86%	93%	94%	98%	98%
Acetald_M	94%	95%	87%	97%	97%	97%	92%	96%	86%	96%	94%	98%	98%
C2H6_M	46%	47%	61%	57%	52%	48%	33%	33%	36%	33%	40%	39%	50%
C3H8_M	95%	97%	94%	98%	98%	98%	95%	97%	92%	96%	94%	98%	98%
iC4H10_M	95%	97%	94%	99%	98%	98%	95%	97%	91%	96%	94%	98%	98%
nC4H10_M	95%	97%	94%	99%	98%	98%	95%	97%	91%	96%	94%	98%	98%
Alkanes_M	46%	47%	61%	57%	52%	48%	34%	34%	39%	33%	40%	39%	50%
C2H4_M	46%	47%	61%	57%	52%	48%	34%	34%	39%	33%	40%	39%	50%
Alkenes_M	46%	47%	61%	57%	52%	48%	34%	34%	39%	33%	40%	39%	50%
C2H2_M	46%	47%	61%	57%	46%	46%	34%	33%	39%	33%	40%	39%	50%
C5H8_M	95%	97%	94%	99%	98%	98%	95%	97%	92%	96%	94%	98%	98%
Benzene_M	95%	97%	94%	99%	98%	98%	95%	97%	92%	96%	94%	98%	98%
Toluene_M	100%	100%	95%	100%	100%	100%	99%	100%	95%	100%	100%	100%	100%
Xylene_M	100%	100%	95%	100%	100%	100%	99%	100%	95%	100%	100%	100%	100%
MeONO2_M	46%	47%	61%	57%	52%	48%	34%	34%	39%	33%	40%	39%	50%
EtONO2_M	46%	47%	61%	57%	52%	48%	34%	34%	39%	33%	40%	39%	50%
RONO2_M	46%	47%	61%	57%	52%	48%	34%	34%	39%	33%	40%	39%	50%
MeOH_M	94%	95%	87%	97%	97%	97%	92%	96%	86%	96%	94%	98%	98%
HCN_M	100%	99%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
CH3CN_M	94%	95%	87%	96%	95%	97%	92%	96%	86%	95%	94%	98%	98%
SF6_M	77%	100%	76%	84%	60%	96%	95%	83%	91%	99%	97%	82%	92%
S_nuc_M	92%	77%	74%	94%	91%	86%	92%	91%	99%	88%	91%	81%	92%
S_atk_M	92%	77%	74%	94%	91%	86%	92%	91%	99%	88%	91%	81%	92%
S_acc_M	92%	77%	67%	94%	91%	86%	91%	91%	99%	88%	91%	81%	91%
S_crs_M	92%	77%	67%	94%	91%	86%	91%	91%	99%	88%	91%	81%	91%
CloudInd_M	98%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	99%	100%

575
576

577

Table S3d. ATom-4, % of non-NaNs after short-gap interpolation

RRno	36	37	38	39	40	41	42	43	44	45	46	47	48
<Lat> (deg)	19	56	42	3	-38	-59	-70	-32	13	60	56	67	46
<Lng> (deg)	-121	-141	-158	-132	10	-93	-59	-41	-27	-37	-62	-105	-135
<Alt> (m)	8278	6678	6123	6419	5922	6843	7197	6672	6729	7019	9678	6759	5935
# parcels	**3374**	**3671**	**2819**	**2894**	**2801**	**3580**	**3812**	**3529**	**3307**	**3440**	**1106**	**3779**	**2399**
H2O_M	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
RHw_M	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
O3_M	100%	100%	100%	100%	100%	100%	99%	100%	100%	100%	**0%**	100%	100%
CO_M	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
CH4_M	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
NOx_M	62%	77%	93%	84%	99%	100%	89%	100%	100%	99%	100%	100%	100%
NOxPSS_M													
HNO3_M	93%	94%	98%	75%	95%	96%	96%	96%	96%	97%	96%	96%	98%
HNO4_M	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**
PAN_M	99%	92%	100%	100%	99%	100%	100%	100%	100%	100%	83%	100%	100%
CH2O_M	100%	82%	100%	100%	98%	100%	98%	98%	98%	96%	**0%**	95%	93%
H2O2_M	94%	94%	98%	75%	95%	96%	96%	96%	96%	97%	96%	96%	98%
CH3OOH_M	43%	59%	59%	59%	**0%**	**0%**	**0%**	**0%**	**0%**	69%	67%	**0%**	**0%**
Acetone_M	96%	98%	98%	88%	98%	96%	98%	98%	98%	97%	**0%**	95%	93%
Acetald_M	96%	87%	97%	88%	92%	91%	94%	97%	93%	89%	**0%**	95%	92%
C2H6_M	26%	35%	40%	40%	46%	34%	31%	28%	31%	29%	**0%**	27%	31%
C3H8_M	96%	99%	99%	94%	100%	97%	98%	98%	98%	97%	**0%**	96%	95%
iC4H10_M	96%	99%	99%	94%	100%	97%	98%	98%	98%	97%	**0%**	96%	95%
nC4H10_M	96%	99%	99%	94%	100%	97%	98%	98%	98%	97%	**0%**	96%	95%
Alkanes_M	26%	35%	42%	43%	46%	34%	33%	28%	31%	29%	**0%**	27%	31%
C2H4_M	26%	35%	42%	43%	46%	34%	33%	28%	31%	29%	**0%**	27%	31%
Alkenes_M	26%	35%	42%	43%	46%	34%	33%	28%	31%	29%	**0%**	27%	31%
C2H2_M	26%	35%	42%	43%	46%	34%	33%	28%	31%	29%	**0%**	27%	31%
C5H8_M	96%	99%	99%	94%	100%	97%	98%	98%	98%	97%	**0%**	96%	95%
Benzene_M	96%	99%	99%	94%	100%	97%	98%	98%	98%	97%	**0%**	96%	95%
Toluene_M	100%	100%	99%	95%	100%	100%	100%	100%	100%	100%	**0%**	100%	100%
Xylene_M	100%	100%	99%	95%	100%	100%	100%	100%	100%	100%	**0%**	100%	100%
MeONO2_M	26%	35%	42%	43%	46%	34%	33%	28%	31%	29%	**0%**	27%	31%
EtONO2_M	26%	35%	42%	43%	46%	34%	33%	28%	31%	29%	**0%**	27%	31%
RONO2_M	26%	35%	42%	43%	46%	34%	33%	28%	31%	29%	**0%**	27%	31%
MeOH_M	96%	98%	98%	88%	98%	96%	98%	98%	98%	97%	**0%**	95%	93%
HCN_M	99%	100%	100%	95%	99%	100%	100%	99%	99%	100%	96%	100%	100%
CH3CN_M	96%	98%	98%	88%	98%	96%	98%	98%	98%	97%	**0%**	95%	93%
SF6_M	76%	92%	97%	95%	97%	85%	90%	98%	88%	85%	94%	97%	94%
S_nuc_M	94%	99%	89%	94%	96%	82%	81%	96%	98%	65%	85%	93%	94%
S_atk_M	94%	99%	89%	94%	96%	82%	81%	96%	98%	65%	85%	93%	94%
S_acc_M	94%	99%	88%	94%	96%	82%	81%	95%	98%	65%	85%	92%	94%
S_crs_M	94%	99%	88%	94%	96%	82%	81%	95%	98%	65%	85%	92%	94%
CloudInd_M	100%	100%	100%	100%	100%	100%	99%	94%	100%	100%	100%	100%	99%

578
579

Table S4. ATom, % of records by flag							
Flags	**0***	**1**	**2**	**3**	**4**	**5**	**6**
H2O_M	0.8%	99.0%	0.0%	0.3%	0.0%	0.0%	0.0%
RHw_M	0.8%	99.0%	0.0%	0.3%	0.0%	0.0%	0.0%
O3_M	0.8%	98.6%	0.0%	0.3%	0.3%	0.0%	0.0%
CO_M	0.8%	79.4%	19.4%	0.1%	0.5%	0.0%	0.0%
CH4_M	0.8%	93.5%	1.3%	1.9%	2.5%	0.0%	0.0%
NOx_M	0.8%	80.8%	0.0%	8.3%	7.6%	2.5%	0.0%
NOxPSS_M	0.8%	82.4%	0.0%	11.8%	5.1%	0.0%	0.0%
HNO3_M	0.8%	78.0%	0.0%	11.6%	5.7%	3.9%	0.0%
HNO4_M	0.8%	28.5%	0.0%	4.0%	8.5%	58.3%	0.0%
PAN_M	0.8%	58.0%	28.4%	7.5%	5.4%	0.0%	0.0%
CH2O_M	0.8%	82.9%	14.9%	0.3%	1.1%	0.0%	0.0%
H2O2_M	0.8%	78.5%	0.0%	11.6%	5.3%	3.9%	0.0%
CH3OOH_M	0.8%	42.0%	0.0%	12.0%	29.4%	15.8%	0.0%
Acetone_M	0.8%	31.7%	0.0%	61.6%	6.0%	0.0%	0.0%
Acetald_M	0.8%	31.4%	0.0%	60.9%	6.9%	0.0%	0.0%
C2H6_M	0.8%	28.0%	0.0%	12.4%	56.3%	0.0%	2.5%
C3H8_M	0.8%	28.0%	53.1%	12.5%	5.1%	0.0%	0.7%
iC4H10_M	0.8%	28.1%	54.9%	12.5%	3.2%	0.0%	0.5%
nC4H10_M	0.8%	28.1%	54.9%	12.5%	3.2%	0.0%	0.5%
Alkanes_M	0.8%	28.1%	0.0%	12.5%	56.0%	0.0%	2.6%
C2H4_M	0.8%	28.1%	0.0%	12.5%	56.0%	0.0%	2.6%
Alkenes_M	0.8%	28.1%	0.0%	12.5%	56.0%	0.0%	2.6%
C2H2_M	0.8%	28.0%	0.0%	12.5%	56.2%	0.0%	2.6%
C5H8_M	0.8%	31.8%	2.3%	61.7%	3.1%	0.0%	0.5%
Benzene_M	0.8%	31.8%	2.3%	61.7%	3.1%	0.0%	0.5%
Toluene_M	0.8%	33.0%	0.6%	64.8%	0.6%	0.0%	0.2%
Xylene_M	0.8%	33.0%	0.6%	64.8%	0.6%	0.0%	0.2%
MeONO2_M	0.8%	27.4%	0.0%	12.3%	57.0%	0.0%	2.6%
EtONO2_M	0.8%	26.8%	0.0%	12.1%	57.8%	0.0%	2.6%
RONO2_M	0.8%	28.1%	0.0%	12.5%	56.0%	0.0%	2.6%
MeOH_M	0.8%	31.7%	0.0%	61.5%	6.0%	0.0%	0.0%
HCN_M	0.8%	78.5%	8.3%	11.6%	0.8%	0.0%	0.0%
CH3CN_M	0.8%	31.7%	0.0%	61.5%	6.0%	0.0%	0.0%
SF6_M	0.8%	10.4%	5.8%	79.2%	3.8%	0.0%	0.0%
S_nuc_M	0.8%	84.6%	0.0%	4.4%	10.3%	0.0%	0.0%
S_atk_M	0.8%	84.6%	0.0%	4.4%	10.3%	0.0%	0.0%
S_acc_M	0.8%	84.1%	0.0%	4.6%	10.6%	0.0%	0.0%
S_crs_M	0.8%	84.1%	0.0%	4.6%	10.6%	0.0%	0.0%
CloudInd_M	0.8%	98.7%	0.0%	0.2%	0.3%	0.0%	0.0%
* The 0.8% flag=0 corresponds to the short flight RF #46, for which we NaN'd all chemical data.							

582

Table S5. Test of long-gap interpolation method				
Species	All parcels	Long-gap interpolated parcels		Short-gap fill
(ppt unless noted)	mean	bias (% of mean)	RMSE (% of mean)	RMSE (% of mean)
H2O_M (ppm)	336			16%
RHw_M (%)	40			12%
O3_M (ppb)	80	3%	12%	6%
CO_M (ppb)	80	1%	8%	3%
CH4_M (ppb)	1850	<1%	<1%	<1%
NOx_M	64	-8%	44%	22%
NOxPSS_M	46	-17%	70%	25%
HNO3_M	162	-6%	22%	12%
HNO4_M	26	-7%	54%	28%
PAN_M	87	6%	25%	14%
CH2O_M	140	6%	22%	11%
H2O2_M	250	9%	30%	16%
CH3OOH_M	381	12%	45%	21%
Acetone_M	351	3%	18%	
Acetald_M	56	3%	19%	
C2H6_M	644	2%	16%	
C3H8_M	109	3%	16%	
iC4H10_M	11	6%	29%	
nC4H10_M	21	5%	29%	
Alkanes_M	16	3%	33%	
C2H4_M	6	28%	94%	
Alkenes_M	0.2	17%	78%	
C2H2_M	97	10%	42%	
C5H8_M	0.5	16%	70%	
Benzene_M	15	-12%	33%	
Toluene_M	1	4%	28%	
Xylene_M	0.1	33%	97%	
MeONO2_M	9	-11%	29%	
EtONO2_M	2	-11%	33%	
RONO2_M	5	-5%	37%	
MeOH_M	590	3%	38%	
HCN_M	185	5%	31%	10%
CH3CN_M	114	11%	44%	
SF6_M	9	<1%	1%	<1%

583
584

585
586

Table S6. Test of missing flight data			
Missing data	All parcels	Interpolated	Flights used
(ppt unless noted)	Mean (ppt)	RMSE (% of mean)	
ATom-1 RF-5			
H2O2_M	392	24%	AT-1 RF4, AT-2/3/4 RF-4/5
HNO3_M	139	58%	AT-1 RF4, AT-2/3/4 RF-4/5
HNO4_M	30.2	66%	AT-1 RF4, AT-2 RF-4/5
ATom-2 RF-2			
H2O2_M	125	23%	AT-2 RF-3, AT-1/3/4 RF-2/3
HNO3_M	30.9	52%	AT-2 RF-3, AT-1/3/4 RF-2/3
HNO4_M	14.3	63%	AT-2 RF-3, AT-1 RF-2/3
ATom-3 RF-1			
NOx_M	80.9	55%	AT-3 RF-2, AT-1/2/4 RF-1/2
ATom-3/4 all			
HNO4_M	26.1	105%	AT-1/2 all
ATom-4 RF-5/6/7/8/9/12/13			
CH3OOH_M	336	72%	AT-1/2 RF-5:11, AT-3 RF-5:13, AT-4 RF-4
Notes: Missing flight data are filled using a multiple linear regression from other flights based on the explanatory variables: pressure, noontime solar zenith angle, and latitude (in that order). RMSE is calculated from the residuals of this fit for the flights used in the regression.			

587
588

589

Table S7. Chemistry models

Used in	ID	Model name	Type	Meteorology	Model Grid	References	Point of Contact
clim, R0	GFDL	GFDL-AM3	CCM	NCEP (nudged)	C180 x L48	Horowitz et al., 2003; Li et al. 2017	amfiore @ldeo.columbia.edu
clim, R0	GISS	GISS-E2.1	CCM	Daily SSTs, nudged to MERRA	2° x 2.5° x 40L	Rienecker et al.,	lee.murray @rochester.edu
clim, R0, R1	GMI	GMI-CTM	CTM	MERRA	1° x 1.25° x 72L	Strahan et al., 2013; Duncan et al., 2007	Sarah.A.Strode @nasa.gov
clim, R0	GC	GEOS-Chem	CTM	MERRA-2	2° x 2.5° x 72L	Gelaro et al., 2017	lee.murray @rochester.edu
clim, R0	NCAR	CAM4-Chem	CCM	MERRA	0.47° x 0.625° x 52L	Tilmes et al., 2016	emmons @ucar.edu
clim, R0, R1, R2	UCI	UCI-CTM	CTM	ECMWF IFS Cy38r1	T159N80 x L60	Holmes et al., 2017; Prather 2015	mprather @uci.edu
R0	F0AM	F0AM	box	MDS+scaled ATom Js	N/A	Wolfe et al., 2016	glenn.m.wolfe @nasa.gov

590

591

Table S8. Reactivity statistics and mean J-values for the 3 large domains (Global, Pacific, Atlantic).

Table S7a. Average Reactivity: mean, median, mean of top 10%

Value	Region	Models with R0								Models w/ R1/R2	
		F0AM	GC	GISS	GMI	NCAR	UCI	U15	U97	GMI1	UCI2
P-O3, mean, ppb/d											
	Global	1.83	1.58	1.98	1.53	1.64	1.75	1.75	1.75	1.75	1.83
	Pacific	1.96	1.97	2.00	1.92	1.99	2.13	2.09	2.10	2.19	2.35
	Atlantic	2.11	2.29	3.73	2.38	2.57	2.61	2.60	2.61	2.74	2.63
P-O3, median, ppb/d											
	Global	1.15	0.95	1.23	0.85	0.95	0.96	0.96	0.97	1.05	1.13
	Pacific	1.31	1.64	1.62	1.48	1.63	1.68	1.66	1.67	1.76	1.97
	Atlantic	2.00	2.26	3.66	2.31	2.48	2.45	2.45	2.46	2.64	2.59
P-O3, mean of top 10%, ppb/d											
	Global	7.04	6.06	7.05	5.84	6.02	7.07	7.07	7.02	6.64	6.84
	Pacific	6.80	5.48	5.44	5.60	5.64	6.33	6.12	6.17	6.11	6.13
	Atlantic	4.53	4.92	7.50	5.10	5.36	5.92	5.82	5.85	5.74	5.46
L-O3, mean, ppb/d											
	Global	1.55	1.17	1.39	1.22	1.24	1.27	1.27	1.27	1.23	1.29
	Pacific	1.62	1.46	1.69	1.49	1.52	1.53	1.49	1.51	1.48	1.53
	Atlantic	2.32	2.17	2.57	2.36	2.43	2.48	2.47	2.49	2.39	2.51
L-O3, median, ppb/d											
	Global	0.67	0.43	0.55	0.43	0.49	0.47	0.47	0.47	0.43	0.48
	Pacific	0.93	0.88	1.10	0.88	0.99	0.94	0.94	0.94	0.85	0.94
	Atlantic	1.98	1.76	2.28	2.01	1.98	2.04	2.05	2.04	2.05	2.14
L-O3, mean of top 10%, ppb/d											
	Global	5.99	5.10	5.75	5.32	5.37	5.64	5.64	5.66	5.53	5.67
	Pacific	5.88	5.00	5.41	5.17	5.04	5.24	5.09	5.18	5.26	5.12
	Atlantic	5.90	5.42	6.35	5.93	6.74	6.35	6.32	6.37	5.98	6.35
L-CH4, mean, ppb/d											
	Global	0.68	0.53	0.32	0.52	0.51	0.55	0.55	0.55	0.55	0.56
	Pacific	0.81	0.76	0.39	0.74	0.75	0.77	0.75	0.76	0.77	0.78
	Atlantic	0.90	0.88	0.57	0.92	0.90	0.95	0.95	0.95	0.96	0.97
L-CH4, median, ppb/d											
	Global	0.30	0.19	0.20	0.17	0.21	0.19	0.19	0.19	0.20	0.22
	Pacific	0.44	0.49	0.33	0.45	0.49	0.48	0.48	0.48	0.50	0.53
	Atlantic	0.80	0.81	0.58	0.80	0.79	0.81	0.81	0.82	0.84	0.88
L-CH4, mean of top 10%, ppb/d											
	Global	2.52	2.10	1.04	2.11	1.99	2.28	2.27	2.28	2.25	2.21
	Pacific	2.79	2.31	1.01	2.35	2.27	2.41	2.33	2.37	2.42	2.24
	Atlantic	2.21	1.96	1.10	2.14	2.17	2.30	2.25	2.30	2.26	2.33

Table S7b. Percent of total Reactivity in the top 50%, top 10%, top 3% of parcels

P-O3, % of total R in top 50%	F0AM	GC	GISS	GMI	NCAR	UCI	U15	U97	GMI1	UCI2
Global	87%	92%	88%	92%	91%	92%	92%	92%	91%	90%
Pacific	84%	79%	77%	81%	80%	80%	80%	80%	78%	78%
Atlantic	71%	72%	69%	71%	70%	72%	72%	72%	71%	72%
P-O3, %of total R in top 10%										
Global	38%	38%	36%	38%	37%	40%	40%	40%	38%	37%
Pacific	35%	28%	27%	29%	28%	30%	29%	29%	28%	26%
Atlantic	22%	22%	20%	22%	21%	23%	23%	23%	21%	21%

P-O3, %of total R in top 3%										
Global	19%	18%	17%	17%	17%	19%	19%	19%	18%	17%
Pacific	16%	12%	12%	12%	12%	13%	13%	13%	11%	11%
Atlantic	9%	9%	8%	9%	8%	10%	9%	9%	8%	8%
L-O3, % of total R in top 50%										
Global	92%	94%	93%	94%	93%	94%	94%	94%	94%	94%
Pacific	89%	87%	86%	88%	86%	87%	87%	87%	87%	87%
Atlantic	83%	83%	82%	84%	82%	83%	83%	84%	84%	83%
L-O3, %of total R in top 10%										
Global	39%	44%	41%	44%	43%	45%	45%	45%	45%	44%
Pacific	36%	34%	32%	35%	33%	34%	34%	34%	36%	34%
Atlantic	26%	25%	25%	25%	28%	26%	26%	26%	25%	26%
L-O3, %of total R in top 3%										
Global	15%	17%	17%	17%	19%	18%	18%	18%	18%	18%
Pacific	15%	14%	13%	15%	14%	14%	14%	15%	15%	14%
Atlantic	9%	9%	9%	9%	13%	10%	10%	9%	9%	9%
L-CH4, % of total R in top 50%										
Global	92%	94%	93%	94%	93%	94%	94%	94%	94%	94%
Pacific	89%	87%	86%	88%	86%	87%	87%	87%	87%	87%
Atlantic	83%	83%	82%	84%	82%	83%	83%	84%	84%	83%
L-CH4, %of total R in top 10%										
Global	37%	39%	33%	40%	39%	41%	41%	41%	41%	39%
Pacific	35%	30%	26%	32%	30%	31%	31%	31%	31%	29%
Atlantic	25%	23%	19%	23%	24%	24%	24%	24%	24%	24%
L-CH4, %of total R in top 3%										
Global	15%	15%	14%	15%	15%	16%	16%	16%	16%	16%
Pacific	15%	11%	10%	12%	11%	13%	12%	13%	12%	11%
Atlantic	9%	8%	7%	8%	9%	9%	8%	8%	8%	8%

Table S7c. Mean J-values

J-O1D, mean, e-5 /s	F0AM	GC	GISS	GMI	NCAR	UCI	U15	U97
Global	1.18	0.87	1.29	0.93	1.00	0.91	0.91	0.91
Pacific	1.48	1.34	1.95	1.42	1.47	1.40	1.39	1.40
Atlantic	1.44	1.33	1.76	1.42	1.56	1.42	1.42	1.43
J-NO2, mean, e-3 /s								
Global	4.32	3.60	4.69	3.62	3.81	3.97	3.99	3.97
Pacific	4.84	4.67	5.75	4.62	4.86	5.13	5.09	5.11
Atlantic	4.89	4.84	5.66	4.82	5.00	5.40	5.40	5.42

Global includes all ATom-1 parcels (32,383), Pacific considers all measurements (11,486) over the Pacific Ocean from 54°S to 60°N, and Atlantic uses parcels from 54°S to 60°N over the Atlantic basin (7,501). All parcels are weighted at described in the text, including a cosine(latitude) factor.

592
593

Table S9. Standard deviation across 5 separated days in August (% of mean reactivity or J-value) using MDS_R0.

	P-O3	L-O3	L-CH4	J-O1D	J-NO2
GC	11%	9%	10%	9%	9%
GISS	22%	14%	17%	14%	12%
GMI	10%	9%	10%	10%	10%
NCAR	23%	32%	28%	17%	16%
UCI	10%	10%	11%	10%	11%

SI References

Duncan, B.N., Logan, J.A., Bey, I., Megretskaia, I.A., Yantosca, R.M., Novelli, P.C., Jones, N.B. and Rinsland, C.P., 2007. Global budget of CO, 1988–1997: Source estimates and validation with a global model. Journal of Geophysical Research: Atmospheres, 112(D22).

Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R. and Wargan, K., 2017. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). Journal of climate, 30(14), pp.5419-5454.

Holmes, C.D. and Prather, M.J., 2017. An atmospheric definition of the equator and its implications for atmospheric chemistry and climate. Nature Geoscience.

Horowitz, L.W., Walters, S., Mauzerall, D.L., Emmons, L.K., Rasch, P.J., Granier, C., Tie, X., Lamarque, J.F., Schultz, M.G., Tyndall, G.S. and Orlando, J.J., 2003. A global simulation of tropospheric ozone and related tracers: Description and evaluation of MOZART, version 2. Journal of geophysical research: Atmospheres, 108(D24).

Li, D., Zhang, R. and Knutson, T.R., 2017. On the discrepancy between observed and CMIP5 multi-model simulated Barents Sea winter sea ice decline. Nature Communications, 8(1), pp.1-7.

Prather, M.J., Zhu, X., Flynn, C.M., Strode, S.A., Rodriguez, J.M., Steenrod, S.D., Liu, J., Lamarque, J.F., Fiore, A.M., Horowitz, L.W. and Mao, J., 2017. Global atmospheric chemistry–which air matters. Atmospheric Chemistry and Physics, 17(14), pp.9081-9102.

Prather, M.J., Flynn, C.M., Zhu, X., Steenrod, S.D., Strode, S.A., Fiore, A.M., Correa, G., Murray, L.T. and Lamarque, J.F., 2018. How well can global chemistry models calculate the reactivity of short-lived greenhouse gases in the remote troposphere, knowing the chemical composition. Atmospheric Measurement Techniques, 11(5), pp.2653-2668.

Rienecker, M.M., Suarez, M.J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M.G., Schubert, S.D., Takacs, L., Kim, G.K. and Bloom, S., 2011. MERRA: NASA's modern-era retrospective analysis for research and applications. Journal of climate, 24(14), pp.3624-3648.

Strahan, S.E., Douglass, A.R. and Steenrod, S.D., 2016. Chemical and dynamical impacts of stratospheric sudden warmings on Arctic ozone variability. Journal of Geophysical Research: Atmospheres, 121(19), pp.11-836.

Tilmes, S., Sanderson, B.M. and O'Neill, B.C., 2016. Climate impacts of geoengineering in a delayed mitigation scenario. Geophysical Research Letters, 43(15), pp.8222-8229.

642 Wofsy, S.C., Afshar, S., Allen, H.M., Apel, E.C., Asher, E.C., Barletta, B., Bent, J., Bian,
643 H., Biggs, B.C., Blake, D.R. and Blake, N., 2018. ATom: Merged atmospheric chemistry,
644 trace gases, and aerosols. ORNL DAAC Oak Ridge, Tennessee, USA.
645
646 Wolfe, G.M., Marvin, M.R., Roberts, S.J., Travis, K.R. and Liao, J., 2016. The
647 framework for 0-D atmospheric modeling (F0AM) v3. 1. Geoscientific Model
648 Development, 9(9), pp.3309-3319.