Quantifying biases in TROPESS AIRS, CrIS, and joint AIRS+OMI tropospheric ozone products using ozonesondes

Elyse A. Pennington[1], Gregory B. Osterman[1], Vivienne H. Payne[1], Kazuyuki Miyazaki[1], Kevin W. Bowman[1], Jessica L. Neu[1]

[1]Jet Propulsion Laboratory, California Institute of Technology

Text: S1
Tables: S1-S4
Figures: S1-S12

Section S1: Sonde Quality Control Methods

The ozonesonde data used in this study are provided by the Harmonization and Evaluation of Ground-based Instruments for Free Tropospheric Ozone Measurements (HEGIFTOM) Working Group of TOAR-II, as described in the main text in Sect. 2.2. This group used state-of-the-art techniques to reduce the uncertainty of sonde measurements across measurement sites and over time. However, the data is not fully quality controlled due to the large quantity of measured profiles and different management at each sonde launch site. Thus, users of the data products must perform their own quality control (QC), resulting in a lack of consistency in the sonde dataset used across studies. In this section, we investigate the impact that different QC methods have on the resulting sonde dataset and satellite-sonde comparisons. We use the joint AIRS+OMI and matched sonde data to test/demonstrate QC methods, and present results for all 3 satellite products (CrIS, AIRS, and joint AIRS+OMI) at the end of this section.

Quality controls can be applied to the sonde data at two points in the analysis. They can be applied to the raw sonde profiles or to the sonde profiles after the satellite operator has been applied. The joint AIRS+OMI data that has been colocated with sonde data are shown in Fig. S1b. The raw sonde profiles that have been colocated with joint AIRS+OMI data are shown in Fig. S1c and the same profiles that have had the satellite operator applied are shown in Fig. S1a. There are features in some of the raw sonde profiles that are physically unrealistic, including anomalously high ozone at low altitudes and anomalously low ozone at high altitudes. In some of the sonde profiles that have had the satellite operator applied, these anomalous features persist. The application of the satellite operator generally retains the unphysical features seen in the raw sonde profiles and can, in rarer cases, generate unphysical profiles from sonde profiles that do not have obvious quality issues.

To investigate the sources of poor quality data, we test multiple QC methods.
Some methods are applied to the raw sonde profiles -- Step 1 – and other methods are applied to the sonde + operator profiles – Step 2. The methods in both steps are described in Table S1. Some of our Step 1 methods address many of the physical impossibilities listed above by removing profiles that display erroneously high and/or low values at different pressure levels (i.e., 50_p01, 50_p1, 600_p3, 300_1, max16). Some of our other Step 1 methods instead remove profiles in which the sonde stopped measuring at low altitude (i.e., minP_70, minP_60, minP_50, minP_TP), so they do not provide sufficient information in the stratosphere. Lastly, some Step 1 methods compare the sonde profiles to the distribution of satellite profiles and remove any sonde profiles that fall outside different ranges of the variance in the satellite profiles (i.e., 3_sigma, 4_sigma, 5_sigma). After applying the satellite operator, the noise in the data has been smoothed and/or removed, so the QC techniques used here are slightly different. Simple statistics can again be used to filter out profiles that fall far from the bulk of the satellite profiles (i.e., 3_sigma_op, 4_sigma_op, 5_sigma_op). To reduce the circularity of our approach, we also tried utilizing an independent climatology of ozone in the stratosphere from the Aura Microwave Limb Sounder to filter out profiles with unrealistic stratospheric concentrations (MLS4sigma, MLS5sigma, MLS6sigma, MLS7sigma, MLSminmax, MLS199, MLS01999). Because MLS only measures in the stratosphere, we compare the sonde profiles in the troposphere to the distribution of satellite profiles in the troposphere and remove any sonde profiles that fall outside different ranges of the variance in the satellite profiles (trop3sigma, trop4sigma, trop5sigma). This method is similar to some of the methods used in Step 1, but now are only applied in the troposphere to supplement the MLS data in the stratosphere. For all Step 1-2 methods, we remove an entire profile if it contains any data that does not pass the quality check. If an alternative method is used which removes the single data point failing the quality check – but

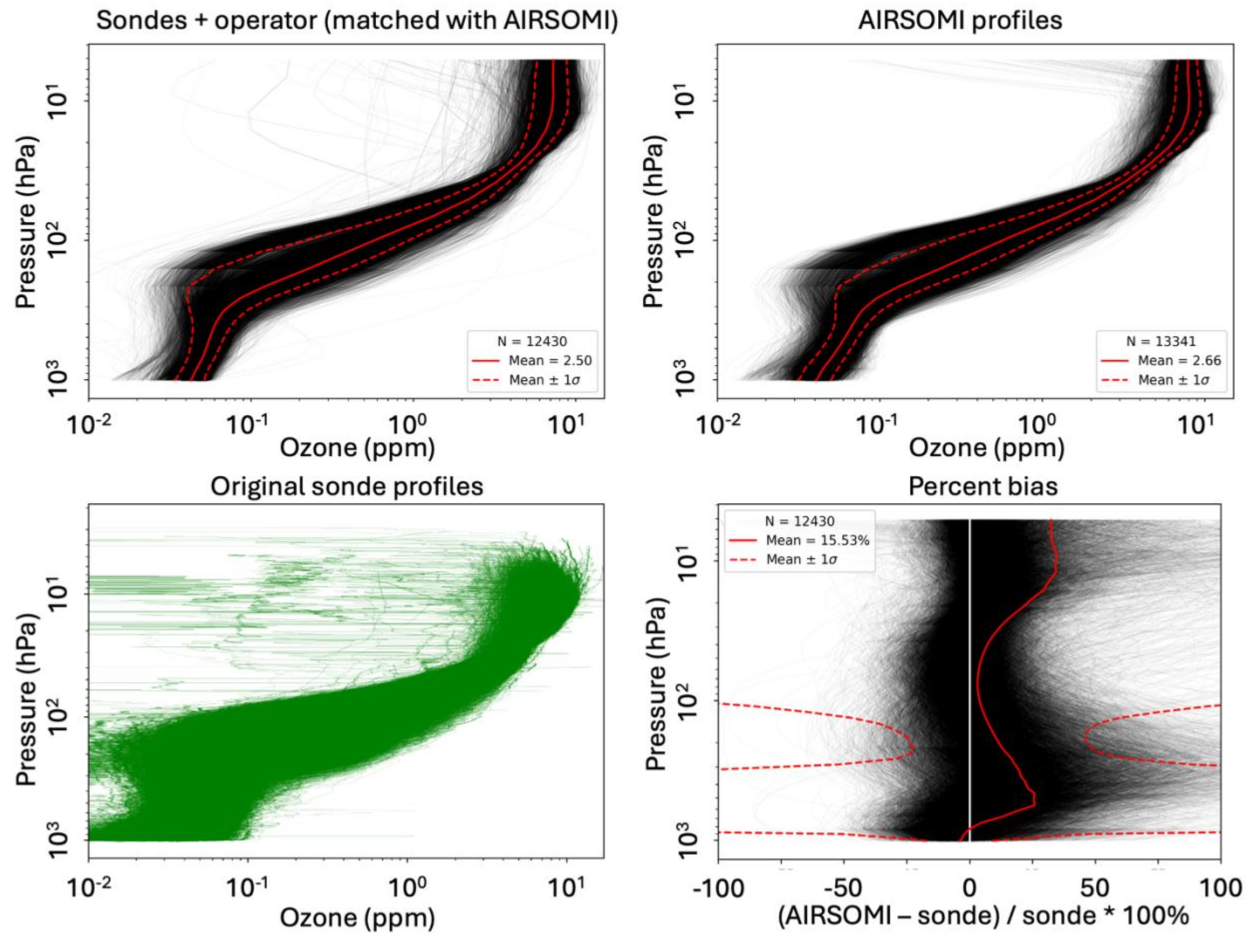

Figure S1: All ozone profiles used in the joint AIRS+OMI-sonde comparison, with no QC applied. (a) Sonde profiles with satellite operators applied. (b) Joint AIRS+OMI profiles. (c) Sonde profiles without satellite operators applied. (d) Percent difference between profiles in (a) and (b). The solid red line displays the mean difference profile and the dashed red lines display one standard deviation outside of the mean.

saves the remainder of the profile – the results are minimally changed relative to removing the whole profile, suggesting that poor-quality profiles have many erroneous data points that impact the entire profile and not only a small portion of the profile.

Three metrics are used to determine the impact of each QC method: the percentage of profiles removed, the median satellite-sonde percent bias between matched profiles, and the trend of satellite-sonde percent bias between matched profiles (Table S2). Ideally, the percentage of profiles removed should not be so large that good-quality profiles are removed and create an unrepresentative sample. There are no benchmark, "goal" values to obtain for the bias or bias trend, but the variation in these values between metrics can provide information about the relative impact of each QC method compared to other QC methods. Additionally, the filtered satellite-sonde bias profiles along with profiles of their mean and standard deviation can provide information about the spread of sonde profiles. Without quality controls, the standard deviation in the satellite-sonde percent bias is very large (Fig. S1d) and impacted by outliers.

Table S1: The quality control methods tested on the original sonde data (left 2 columns) and the sonde data with the satellite operator applied (right 2 columns). The 1st and 3rd columns list the labels used in testing and the 2nd and 4th columns describe those methods.

QC Step 1 Label	**Description**	**QC Step 2 Label**	**Description**
none	No QC applied	none_op	No QC applied
50_p01	Removed if VMR(<50 hPa) < 0.01 ppm	3_sigma_op	Removed if the percent bias btwn sonde and sat was > 3 std. dev. outside the mean bias profile
50_p1	Removed if VMR(<50 hPa) < 0.1 ppm	4_sigma_op	Same as above but 4 std. dev.
600_p3	Removed if VMR(>600 hPa) > 0.3 ppm	5_sigma_op	Same as above but 5 std. dev.
300_1	Removed if VMR(>300 hPa) > 1 ppm	MLS4sigma	Removed if value in profile was outside MLS mean +/- 4 std. dev.
max16	Removed if any concentration > 16 ppm	MLS5sigma	Same as above but 5 std. dev.
minP_70	Removed if measurements stop below reaching 70 hPa	MLS6sigma	Same as above but 6 std. dev.
minP_60	Same as above but 60 hPa	MLS7sigma	Same as above but 7 std. dev.
minP_50	Same as above but 50 hPa	MLSminmax	Removed if value in profile was outside MLS min/max
minP_TP	Removed if measurements stop below reaching the tropopause	MLS199	Removed if value in profile was outside MLS 1st / 99th percentile
3_sigma	Removed if the percent bias btwn sonde and sat was > 3 std. dev. outside the mean bias profile	MLS01999	Removed if value in profile was outside MLS 0.1 / 99.9th percentile
4_sigma	Same as above but 4 std. dev.	trop3sigma	Removed if the percent bias btwn sonde and sat was > 3 std. dev. outside the mean bias profile in the troposphere
5_sigma	Same as above but 5 std. dev.	trop4sigma	Same as above but 4 std. dev.
		trop5sigma	Same as above but 5 std. dev.

Many of the Step 1 methods do not sufficiently QC the sonde data, necessitating the use of Step 2 filters. The Step 1 methods that filter the sonde profiles using strict concentration and pressure cutoffs (i.e., 50_p01, 50_p1, 600_p3, 300_1, max16, minP_70, minP_60, minP_50, minP_TP) do not substantially impact the standard deviation of the satellite-sonde bias profiles (Fig. S2A-I) or the tropospheric column percent bias (Table S2), making them insufficient to completely filter the sonde data. The Step 1 methods that filter sonde profiles via comparison to the satellite data (i.e., 3_sigma, 4_sigma, 5_sigma) have an improved impact on the bias profile standard deviations (Fig. S2J-L), but 4_sigma and 5_sigma leave large standard deviations in the troposphere and 3_sigma

Table S2: Metrics describing the outcome of using each QC method on the matched joint AIRS+OMI-sonde data as described in Table S1.

QC method	**No. of matched profiles**	**No. of profiles removed**	**Percent profiles removed (%)**	**Trop. column percent bias (%)**	**Trop. column percent bias trend (%/ decade)**
none, none_op	12430	0	0	6.6	0.11
50_p01	12425	5	0.04	6.6	0.11
50_p1	12407	23	0.19	6.58	0.09
600_p3	12428	2	0.02	6.6	0.12
300_1	12428	2	0.02	6.6	0.12
max16	12428	2	0.02	6.6	0.12
minP_70	12131	299	2.41	6.57	0.04
minP_60	12090	340	2.74	6.57	0.02
minP_50	12038	392	3.15	6.57	0.02
minP_TP	12322	108	0.87	6.59	0.12
3_sigma	10348	2082	16.7	7.69	-0.06
4_sigma	11681	749	6.03	7.18	0.1
5_sigma	12086	344	2.77	6.92	0.04
3_sigma_op	12257	173	1.39	6.46	0.21
4_sigma_op	12354	76	0.61	6.54	0.12
5_sigma_op	12385	45	0.36	6.56	0.1
MLS4sigma	10963	1467	11.8	6.44	-0.05
MLS5sigma	11731	699	5.62	6.6	0.12
MLS6sigma	12033	397	3.19	6.57	0.15
MLS7sigma	12181	249	2	6.58	0.1
MLSminmax	10365	2065	16.6	6.47	-0.24
MLS199	-	-	> 30	-	-
MLS01999	-	-	> 30	-	-
trop3sigma	12324	106	0.85	6.49	0.18
trop4sigma	12392	38	0.31	6.56	0.14
trop5sigma	12414	16	0.13	6.57	0.13

removes too many profiles (Table S2). The Step 2 filters tend to have a larger impact on satellite-sonde bias. The Step 2 methods that compare the smoothed sonde profiles to the spread in the satellite profiles (i.e., 3_sigma_op, 4_sigma_op, 5_sigma_op) have a noticeable impact on the spread of bias in the stratosphere (Fig. S2M-O) as well as the trend in bias (Table S2). We choose to focus on the methods that use MLS climatology so that we compare to an independent dataset.

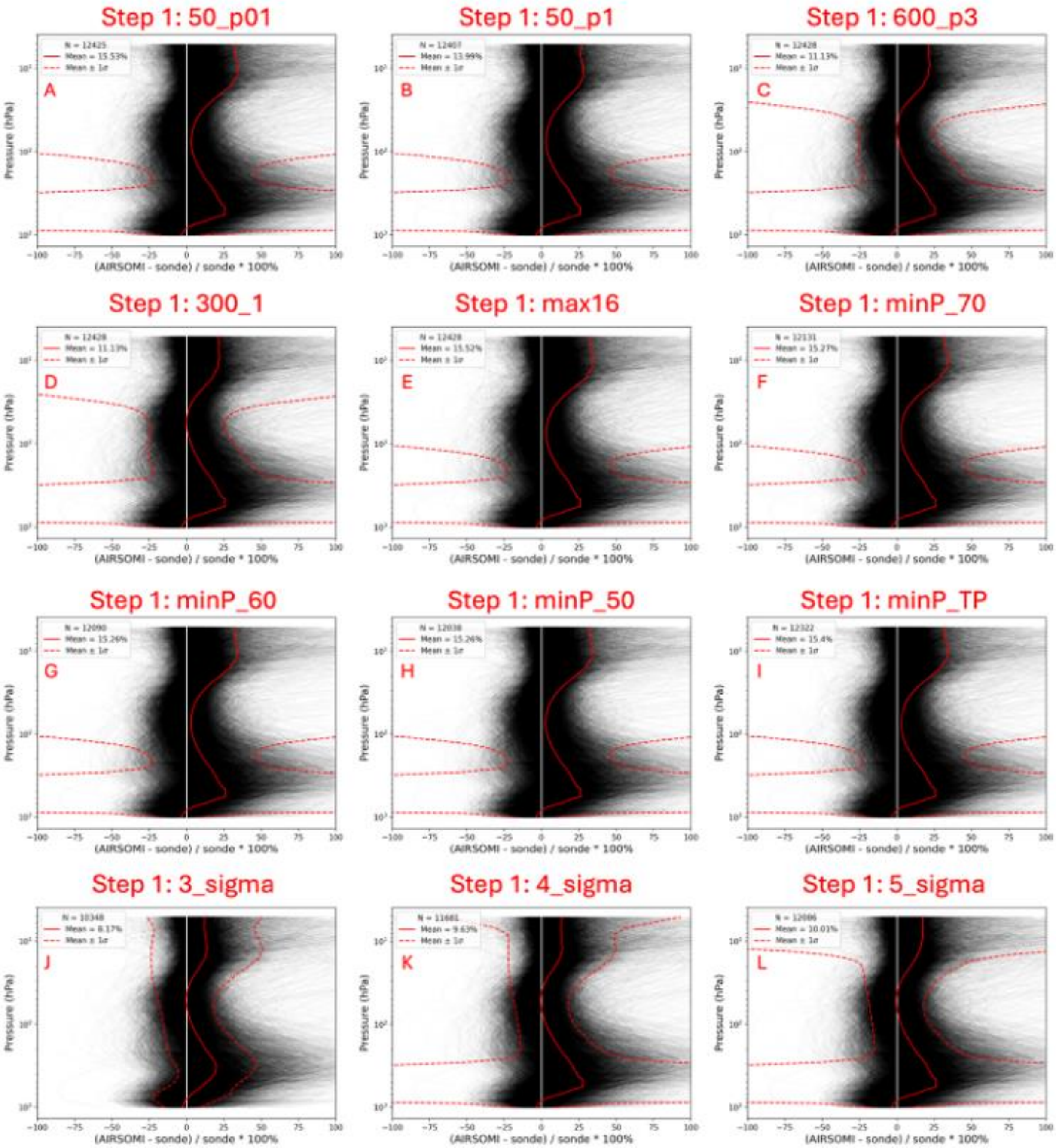

Figure S2: Same as Fig. S1d using different QC methods described in Table S1.

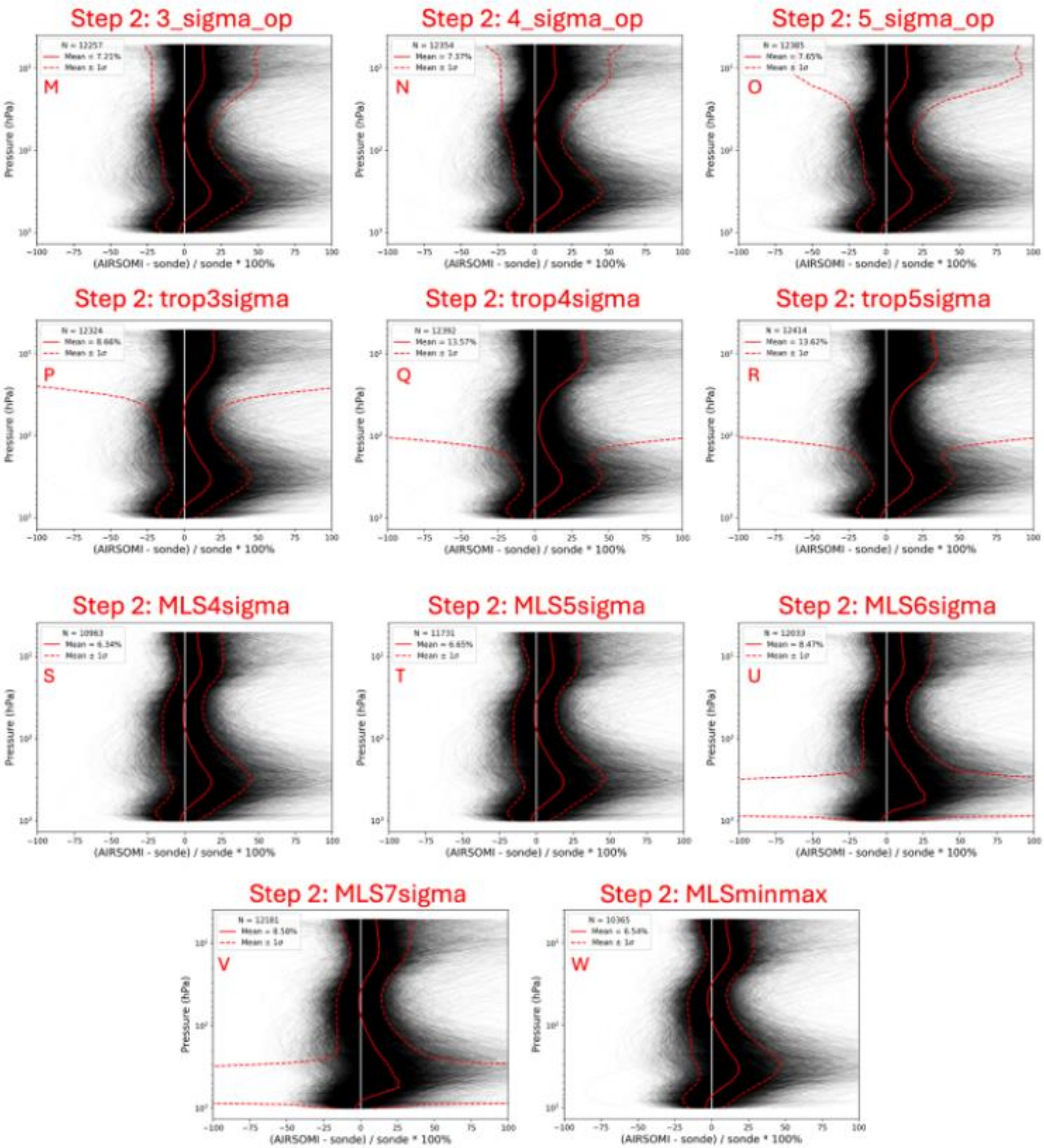

Figure S2, cont.

MLSminmax, MLS199, and MLS01999 led to the removal of too many profiles (Table S2), while MLS6sigma and MLS7sigma distorted the shape of the mean bias profile (Fig. S2U-V). Despite similar mean and standard deviation profiles for MLS4sigma and MLS5sigma (Fig. S2S-T), MLS4sigma removed many more profiles and altered the tropospheric column bias and bias trend (Table S2). Therefore, MLS5sigma is the most reliable method for filtering sondes in the stratosphere. Given the lack of an independent tropospheric ozone climatology, we focus on QC methods that use the spread of the tropospheric profiles from each satellite product to constrain the smoothed sonde profiles in the troposphere. While filtering the troposphere may not be strictly necessary given the good performance of the MLS5sigma method to filter full atmospheric profiles, the examples given here are for joint AIRS+OMI which has the lowest retrieval throughput. Given tests with our other satellite products, and to create a method that can be extended to datasets that do not extend into the stratosphere, we investigate the trop3sigma, trop4sigma, and trop5sigma methods. Their impact on the standard deviation profiles in Fig. S2P-R appears consistent, but their impact on the quantities in Table S2 is varied. The trop5sigma method is selected given the low percentage of profiles removed and its low bias trend.

Finally, we test the combination of methods across and within Steps 1 and 2. Because many of the Step 1 methods did not have large impacts on resulting satellite-sonde metrics, we apply some of them together to test their combined efficacy (i.e., 50_p1, 600_p3, 300_1, max16). We also test two of the Step 2 methods together to cover both the stratosphere and troposphere (i.e., MLS5sigma, trop5sigma). Third, we test some of the Step 1 and Step 2 methods together (i.e., 50_p1, 600_p3, 300_1, max16, MLS5sigma, trop5sigma). These three scenarios listed in Table S3. The Step 1 methods (i.e., 50_p1, 600_p3, 300_1, max16) removed few profiles and had minimal improvement on the standard deviation of the bias (Fig. S3a). In contrast, the combination of the two Step 2 methods (i.e., MLS5sigma, trop5sigma) improved the standard deviation profiles by removing outliers that made up 5.7% of all profiles (Fig. S3b). Combining the Step 1 and Step 2 methods (i.e., 50_p1, 600_p3, 300_1, max16, MLS5sigma, trop5sigma) had little impact on the percentage of profiles removed, standard deviation of the bias, tropospheric column bias, or trend in the bias (Fig. S3c, Table S3). The fact that the number of profiles removed by the combined Step 1+2 test is lower than the sum of the Step 1 or Step 2 methods alone means that the Step 2 removal process removes the profiles of some poor-quality raw profiles that would have been flagged in the Step 1 filtering process.

Given the above results, we chose to move forward with the MLS5sigma_trop5sigma sonde QC method. While this method does not QC the original sonde data directly, it constrains the sondes to physically realistic values through the use of MLS climatology data in the stratosphere and the distributions of each TROPESS dataset in the troposphere. Using this method of QC, the percentage of profiles removed from CrIS, AIRS, and joint AIRS+OMI comparisons are 4.5%, 3.5%, and 5.7%, respectively.

Table S3: Same as Table S2, but for combined QC methods.

QC method	**No. of matched profiles**	**No. of profiles removed**	**Percent profiles removed (%)**	**Trop. column percent bias (%)**	**Trop. column percent bias trend (%/ decade)**
50_p1, 600_p3, 300_1, max16	12403	27	0.22	6.58	0.1
MLS5sigma, trop5sigma	11722	708	5.7	6.59	0.12
50_p1, 600_p3, 300_1, max16, MLS5sigma, trop5sigma	11715	715	5.75	6.59	0.12

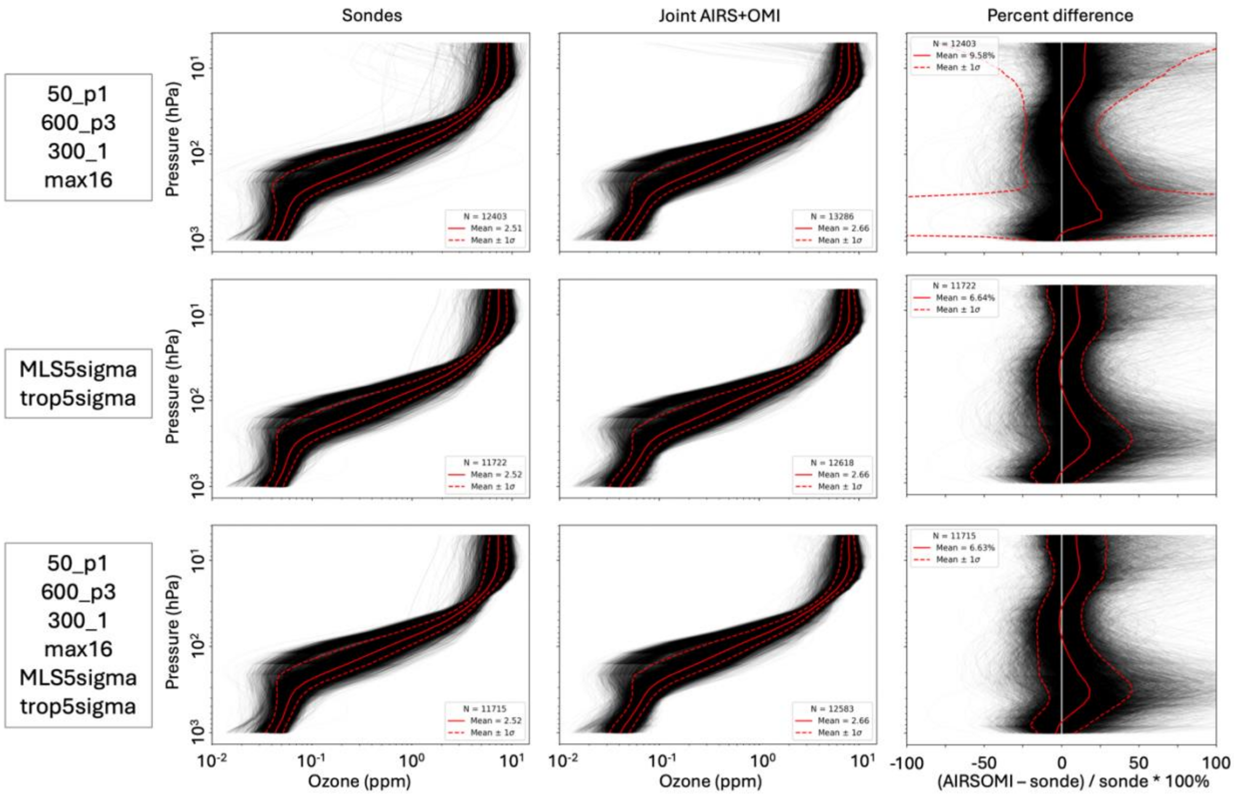

Figure S3: Ozone profiles and percent difference profiles using three QC methods (boxed on the left-hand side). First column: Sonde profiles with satellite priors applied. Middle column: Joint AIRS+OMI profiles. Right column: Percent difference profiles.

Table S4: The p values corresponding to the trends calculated from the monthly-averaged percent bias in Fig. 5.

Latitude Band	**CrIS**	**AIRS**	**Joint AIRS+OMI**
60° → 90°	0.295	0.002	0.529
30° → 60°	0.864	0.091	0.435
0° → 30°	0.405	0.955	0.479
-15° → 15°	0.054	0.006	0.028
-30° → 0°	0.333	0.161	0.09
-60° → -30°	0.957	0.491	0.431

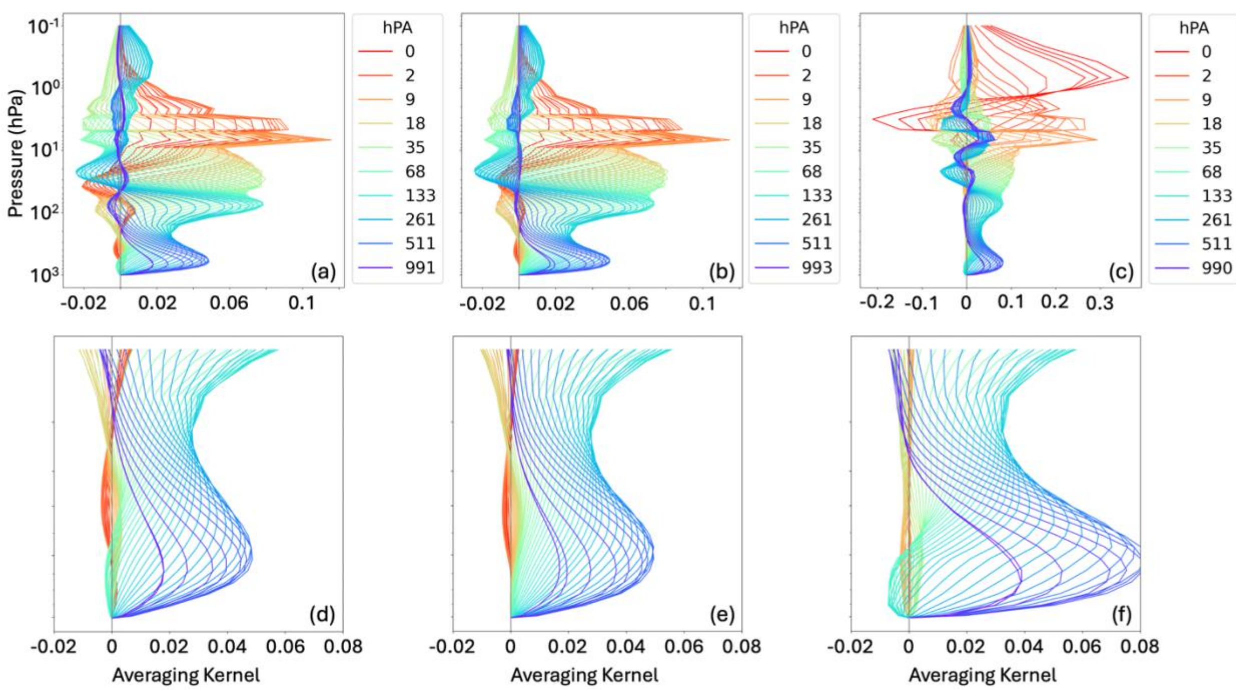

Figure S4: Representative averaging kernels for CrIS (left), AIRS (middle), and joint AIRS+OMI (right) averaged over satellite retrievals that were matched with sonde observations in SON (September/October/November) over Irene, South Africa. Top: Complete atmospheric profile. Bottom: Troposphere only.

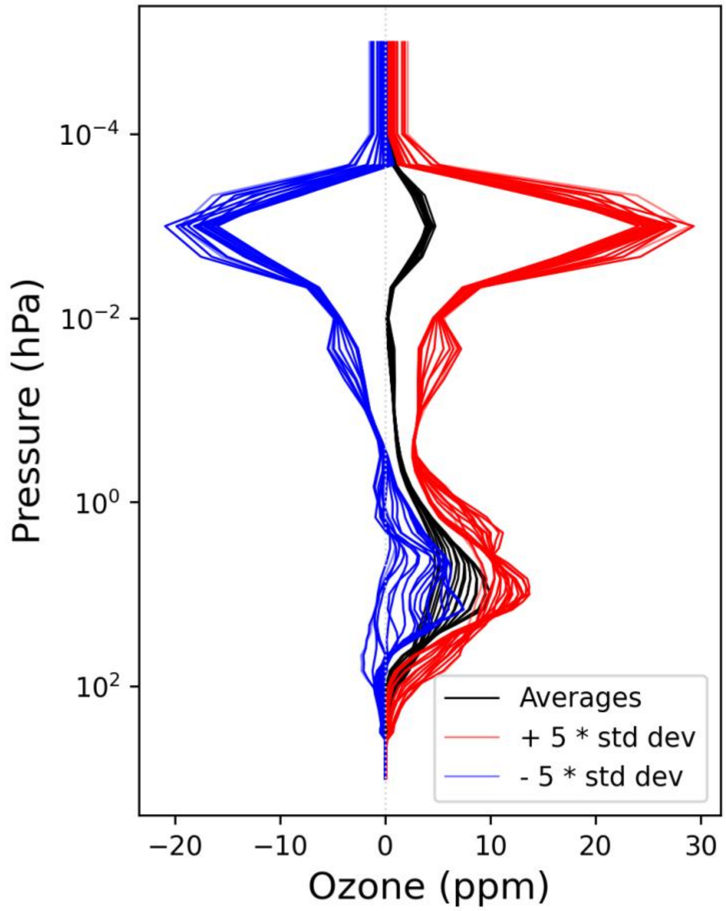

Figure S5: MLS climatology derived from data in Werner et al. (2023). The averages (black) +/- 5 * standard deviation (red and blue) are plotted for each latitude band spaced evenly every 10°.

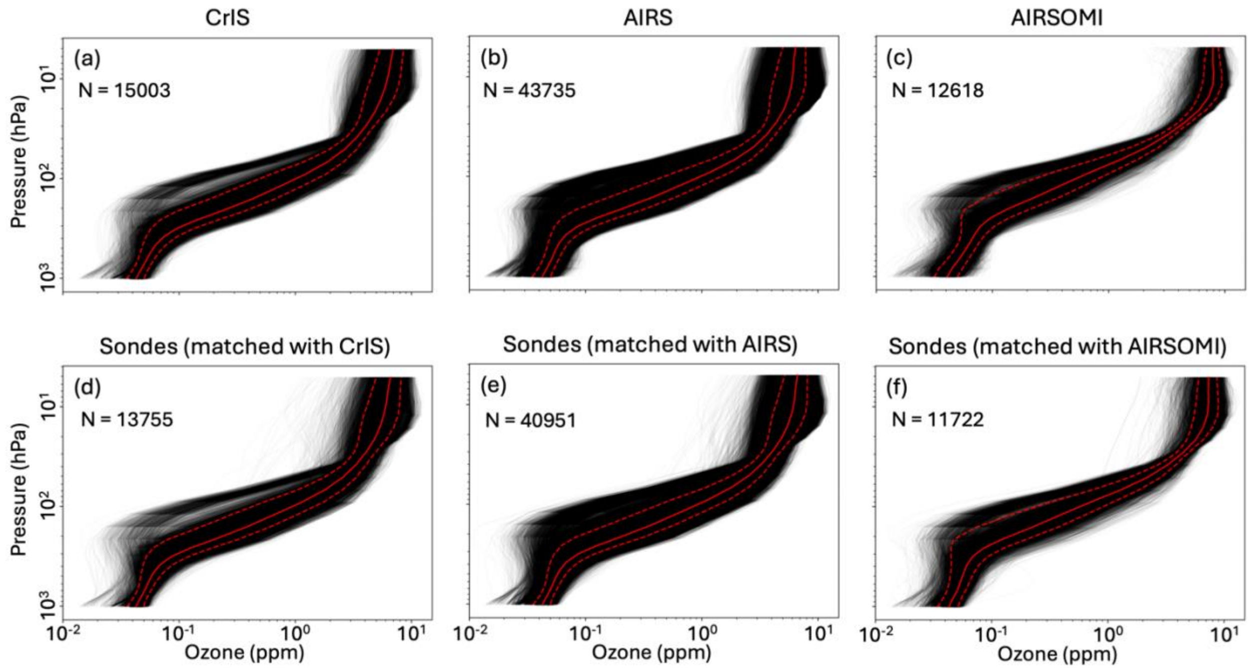

Figure S6: Satellite (top) and sonde (with satellite operators applied; bottom) for coincident measurements of CrIS (left), AIRS (middle), and joint AIRS+OMI (right) at all locations over the entirety of the data period for each product.

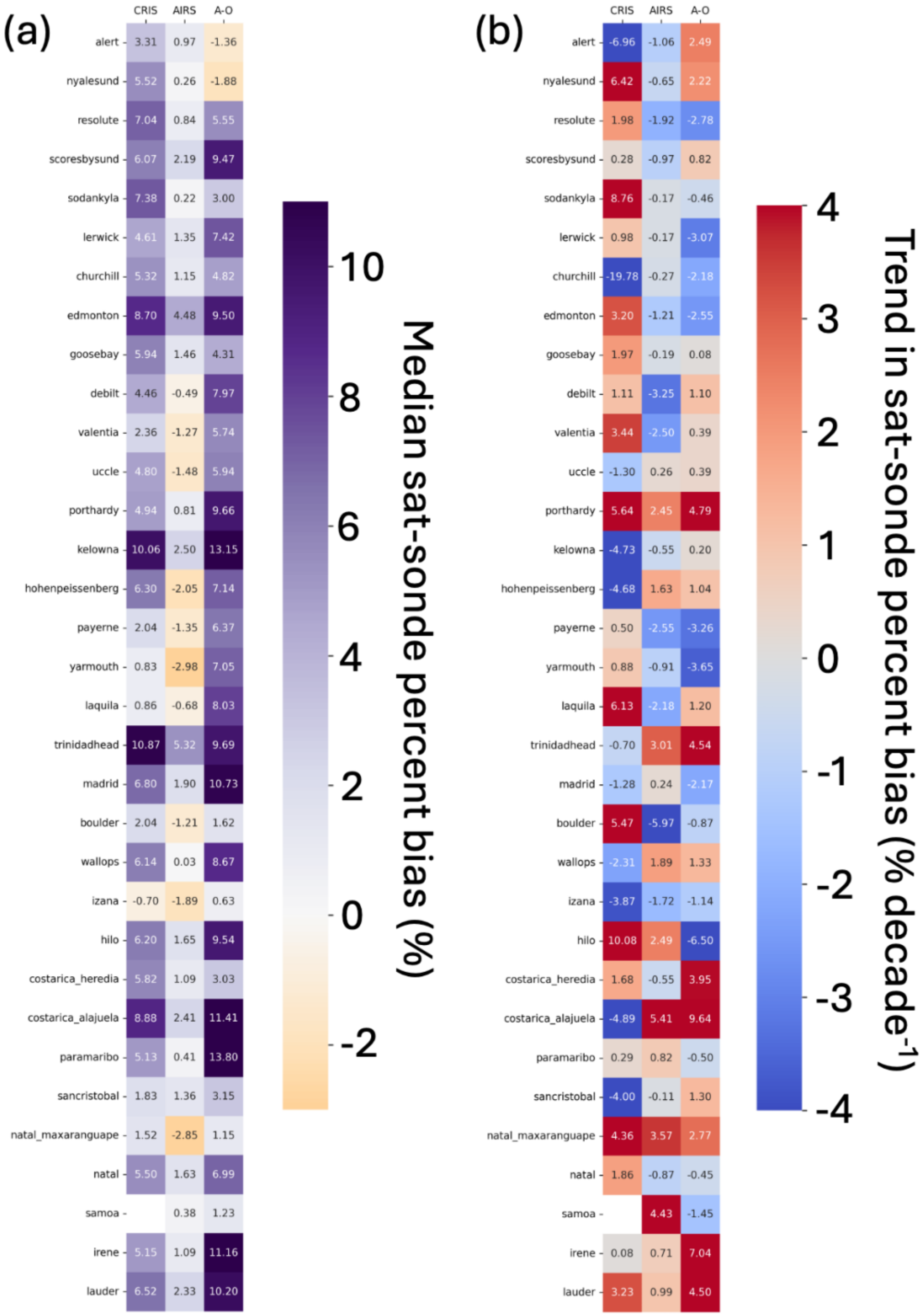

Figure S7: Left: Median satellite-sonde percent bias averaged over the complete time period for each satellite product at each sonde launch site. Right: Trend in satellite-sonde percent bias calculated over the complete time period for each satellite product at each sonde launch site.

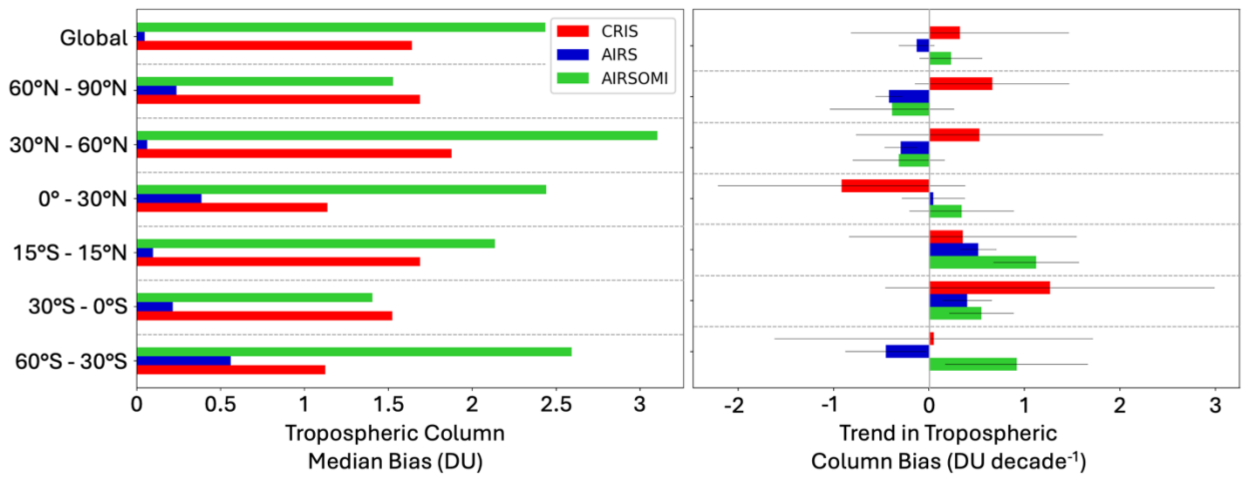

Figure S8: Left: The median tropospheric column absolute bias averaged over sites in each latitude band. Bottom right: The trend in tropospheric column absolute bias in each latitude band.

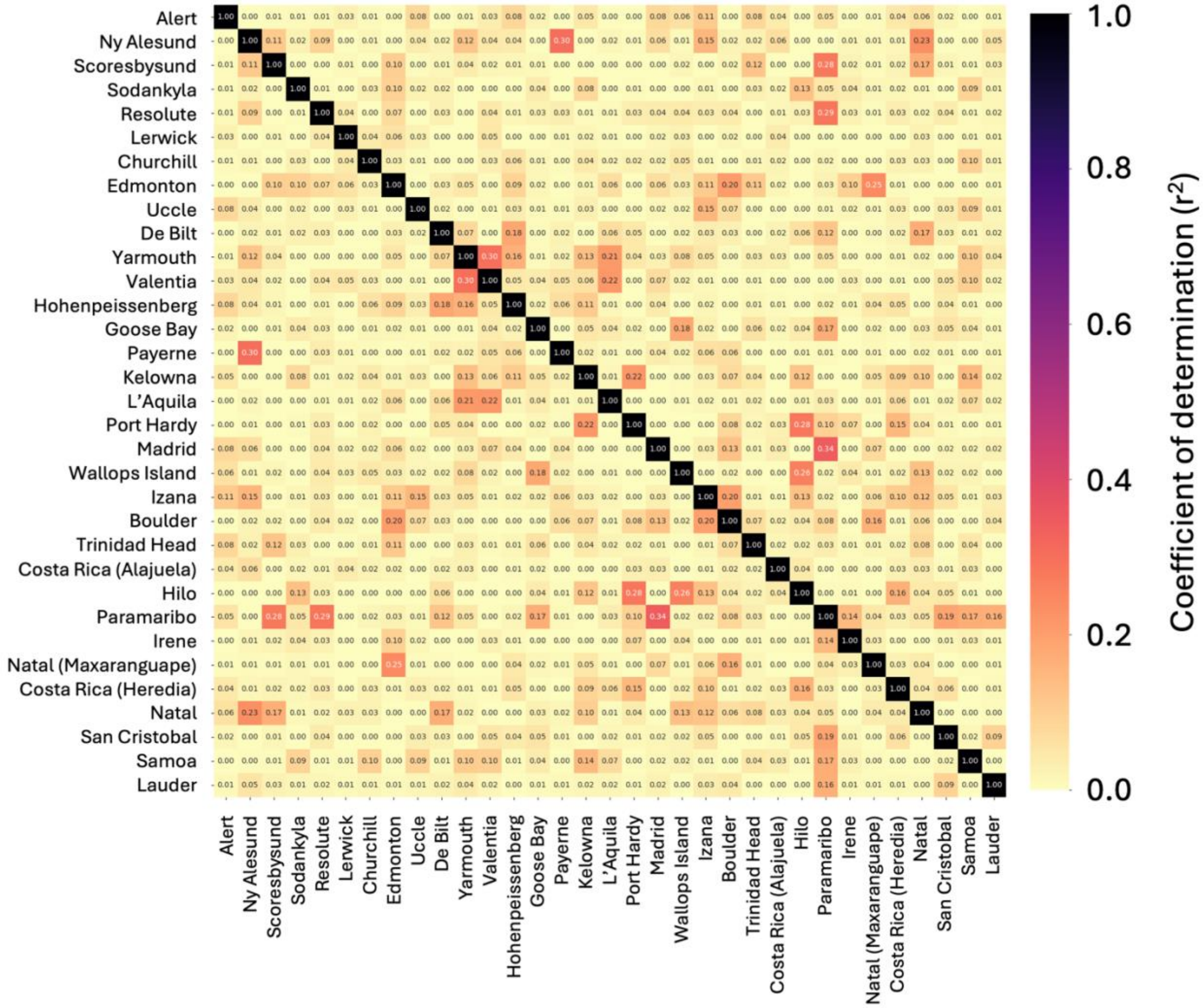

Figure S9: Coefficient of determination (r^2) between the monthly-averaged time series of tropospheric column percent bias at each sonde launch site.

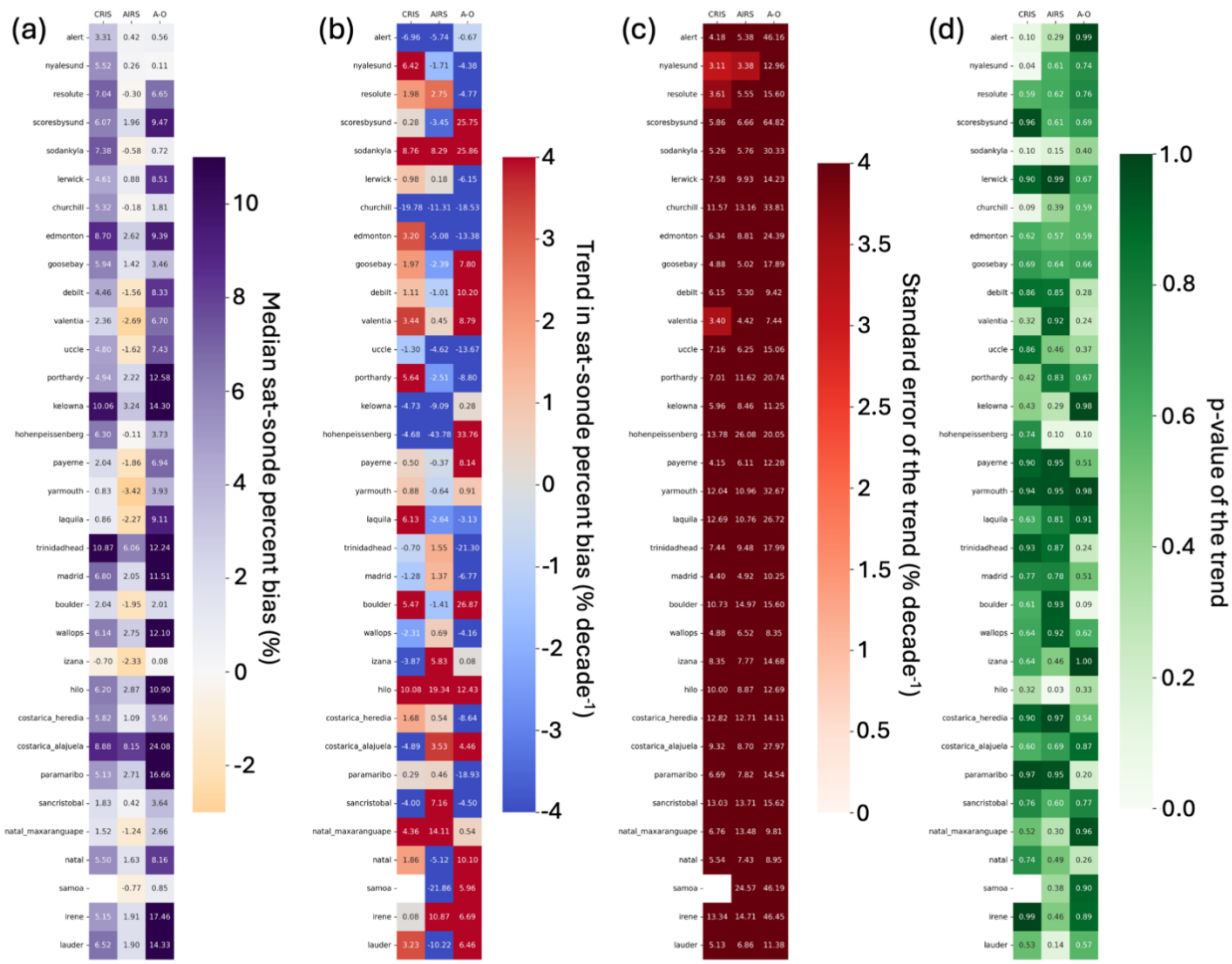

Figure S10: First column: Median satellite-sonde percent bias averaged over the CrIS time period for each satellite product at each sonde launch site. Second column: Trend in satellite-sonde percent bias calculated over the CrIS time period for each satellite product at each sonde launch site. Third column: standard errors associated with the reported trends in the second column. Fourth column: p values associated with the reported trends in the second column.

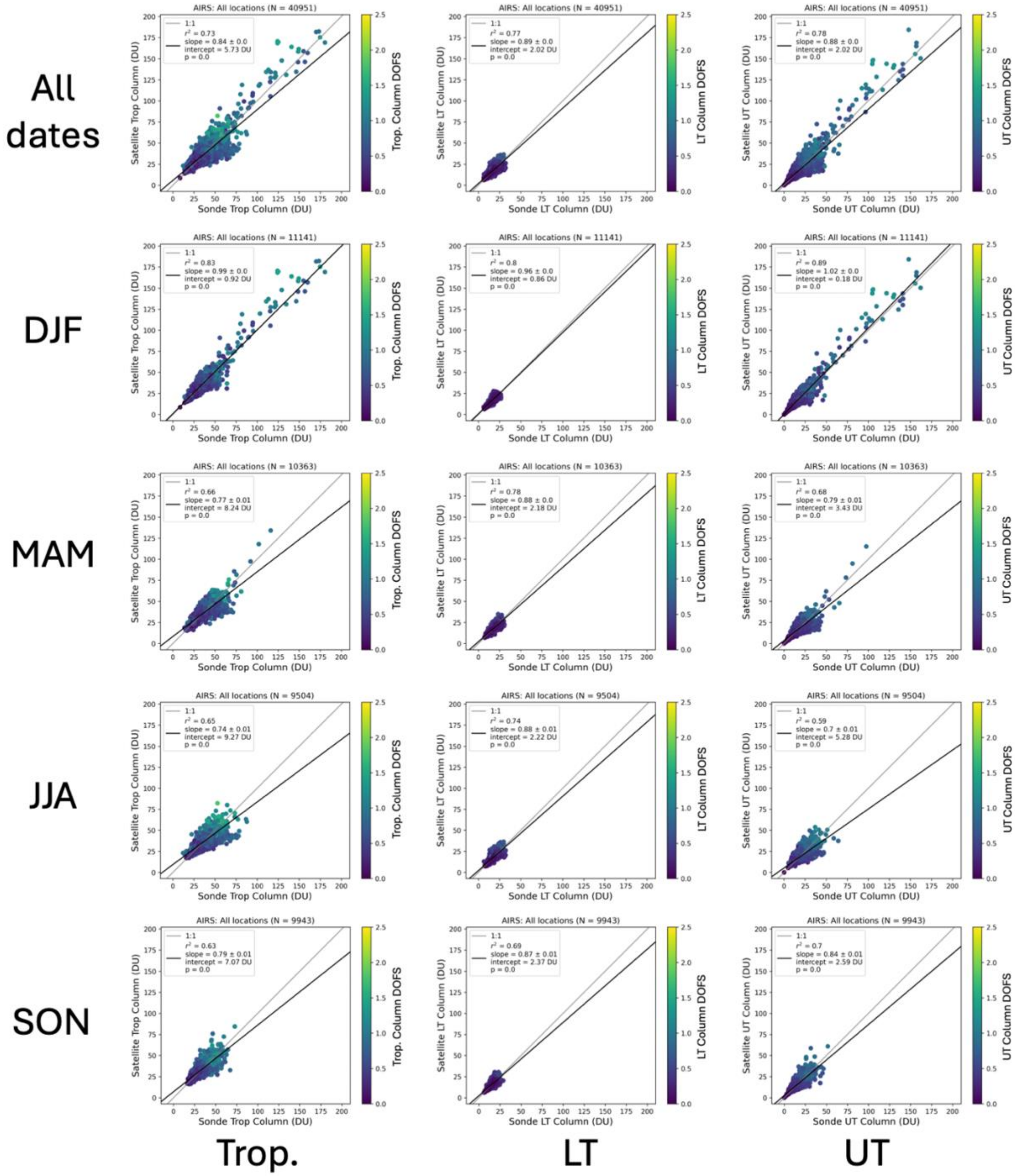

Figure S11: AIRS versus sonde tropospheric (left), LT (middle), and UT (right) ozone columns for all dates (top), December/January/February (second row), March/April/May (third row), June/July/August (fourth row), and September/October/November (bottom row) colored by the degrees of freedom for signal (DOFS) in each subcolumn. The gray line is the 1:1 line and the black line is a simple linear regression fit.

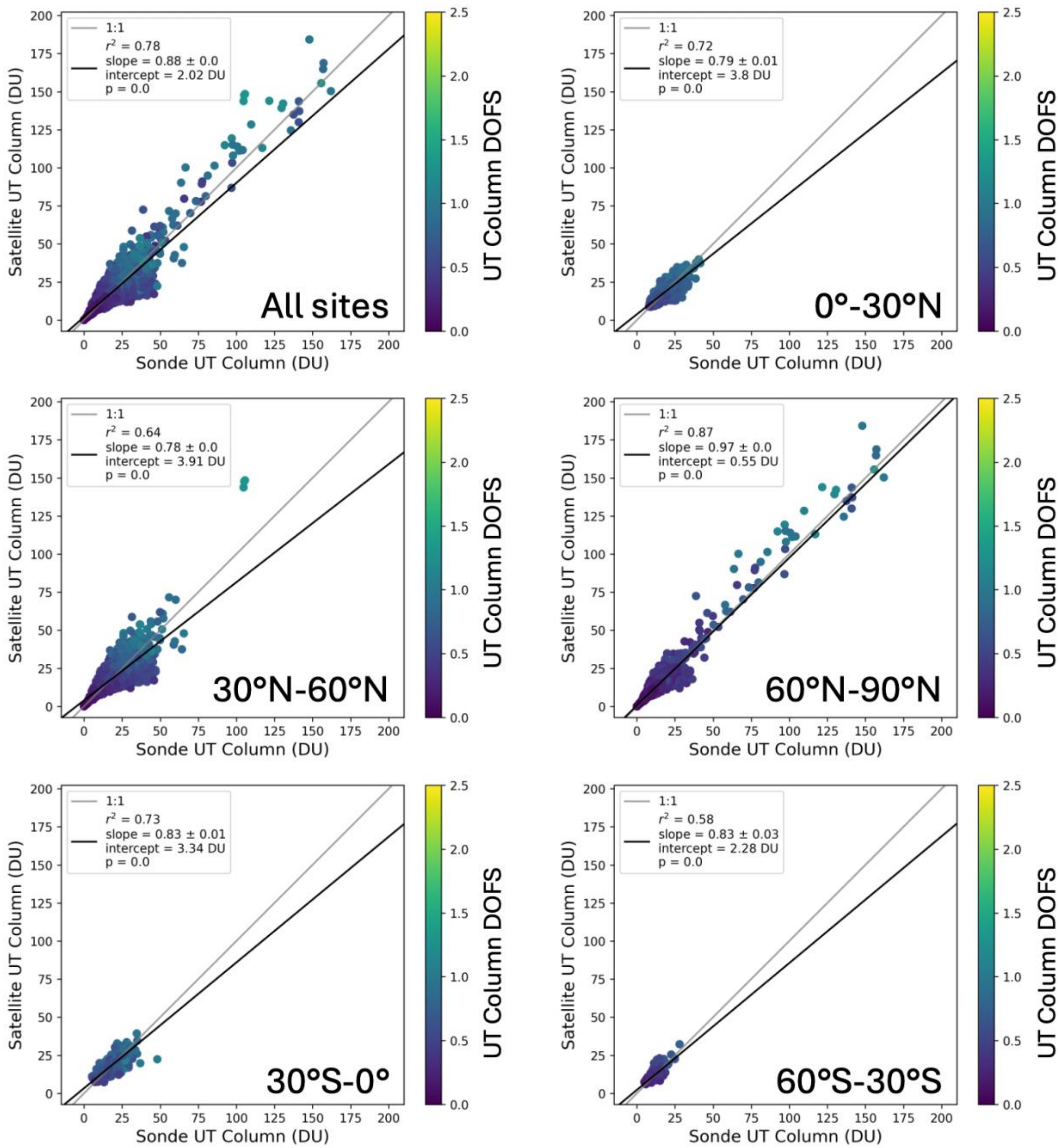

Figure S12: AIRS versus sonde UT ozone columns for all dates globally (top left) and all other latitude bands (other subpanels) colored by the degrees of freedom for signal (DOFS) in each subcolumn. The gray line is the 1:1 line and the black line is a simple linear regression fit.