Supplement of Atmos. Meas. Tech., 12, 5547–5572, 2019
https://doi.org/10.5194/amt-12-5547-2019-supplement

Atmospheric Measurement Techniques
Open Access
EGU

Supplement of

Evaluation of MOPITT Version 7 joint TIR–NIR X_{CO} retrievals with TCCON

Jacob K. Hedelius et al.

Correspondence to: Jacob K. Hedelius (jacob.hedelius@atmosp.physics.utoronto.ca)

S1 Stratospheric contribution to MOPITT priors

From 50–0 hPa there is insufficient information in MOPITT data to retrieve a layer value. Instead this layer is taken from the a priori when creating the column product. The mass ratio of CO in this layer compared to the total column of CO for 2017 is shown in Fig. S1a. We also use the blended tropopause height from the The Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) to calculate the stratospheric fraction to the total (Fig. S1b). In general the contribution in the NH is lower than in the SH, and the contribution in the tropics where the tropopause is higher is lower than the rest of the world.

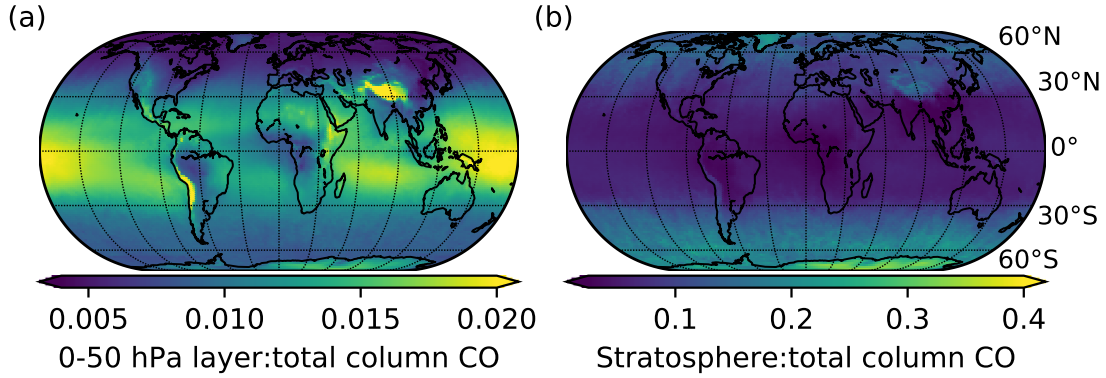

Figure S1. Mass fractions of CO from the stratosphere compared to the total column in the MOPITT priors for 2017. (a) The fraction of the uppermost layer to the total. (b) Fraction of the stratosphere as defined by the MERRA-2 blended tropopause height to the total.

S2 TCCON CO scaling to WMO

Tying column measurements from TCCON to the WMO scale from aircraft measurements is a challenge due to the heterogeneity of atmospheric CO, and lack of stability of CO in cylinders. The heterogeneity of CO means assumptions with large uncertainty may need to be made about CO volume mixing ratios to fill gaps in atmospheric profiles. Usually these gaps are near the surface or in the upper atmosphere beyond the aircraft ceiling. In cylinders CO may change with time. The WMO scale has been periodically updated to tie to gravimetric standards and to account for drift. In the early 1990s measurements were tied to the WMO 88 scale, also known as the CMDL (Climate Monitoring and Diagnostics Laboratory) scale (Novelli et al., 2003). This scale has been updated several times (e.g, WMO-2000, WMO-CO-X2004 (GAW, 2010), WMO-CO-X2014, WMO-CO-X2014A (Zellweger et al., 2017)), which has made small but important improvements. There are also other independent scales that in situ or flask measurements may be tied to such as the Commonwealth Science and Industrial Research Organization (CSIRO) carbon monoxide calibration scale. The next WMO-CO scale update is expected in late 2019 (https://www.esrl.noaa.gov/gmd/ccl/co_scale_update.html, last accessed: 10 December 2018).

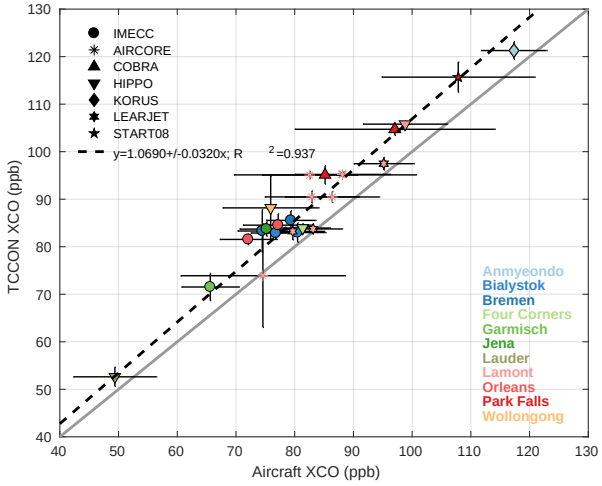

Figure S2. One-to-one plot comparing TCCON to aircraft measurements on the WMO. Before scaling to create the official TCCON product, column measurements are about 7 % higher than WMO scale. Courtesy of Matthäus Kiel.

Scaling the retrievals by $\sim 7\,\%$ seems large given the uncertainly on spectroscopy is likely smaller. This scaling is well known within the TCCON community and has been investigated and discussed recently. This of course can lead to an observed bias between X_{CO} data sets (Zhou et al., 2019).

S3 Small region approximation bin sizes

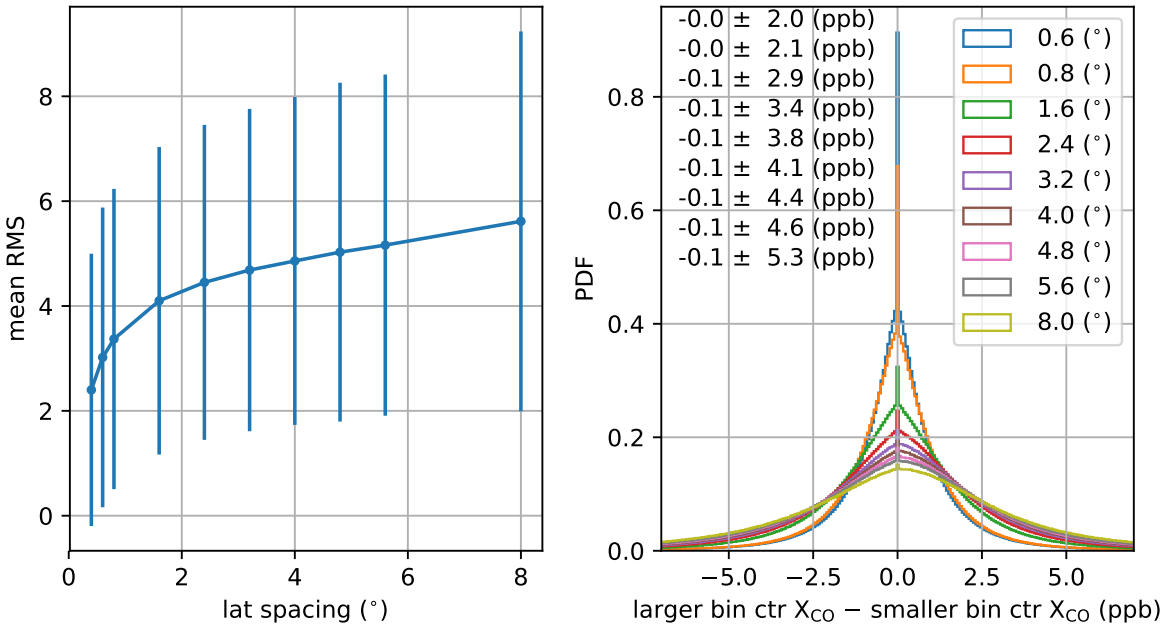

Figure S3. Results from using different SRA sizes. Left: The mean RMS as a function of bin size (latitude height). For larger bin sizes the mean RMS increases. Error bars are 1 σ. Right: Histograms of the difference in the center (median) X_{CO} values between larger bins and the smallest (0.4 °) bin. Text values are the mean deviations $\pm 1\sigma$. Smaller bins tend to have a median X_{CO} that is closer to the median X_{CO} of the (0.4 °) bin.

We examine the effects of the bin size on the small region approximation (SRA) using the MOPITT observations. Ten different bin sizes are compared all with a latitude:longitude ratio of 2:3 near the equator. Moving towards the poles, the latitude spacing is kept constant and the longitudinal width is increased to approximately maintain the same area size as at the equator. We examined areas with latitudinal spacings of 0.4, 0.6, 0.8, 1.6, 2.4, 3.2, 4.0, 4.8, 5.6, and 8.0 degrees. These were required to have a minimum of 6, 8, 10, 10, 10, 10, 10, 12, 15, and 25 points respectively for further analysis. We examine 1) the average root mean squared error compared with the center (median) X_{CO} point, and 2) the difference between the median X_{CO} for larger areas with the median X_{CO} for the 0.4° latitude areas with their center latitude and longitude inside the larger area (Fig. S3). For larger areas, the overall mean RMS increases. The spread of the difference between the median X_{CO} values between the larger and 0.4° bin increases with larger area bin size, but has little bias.

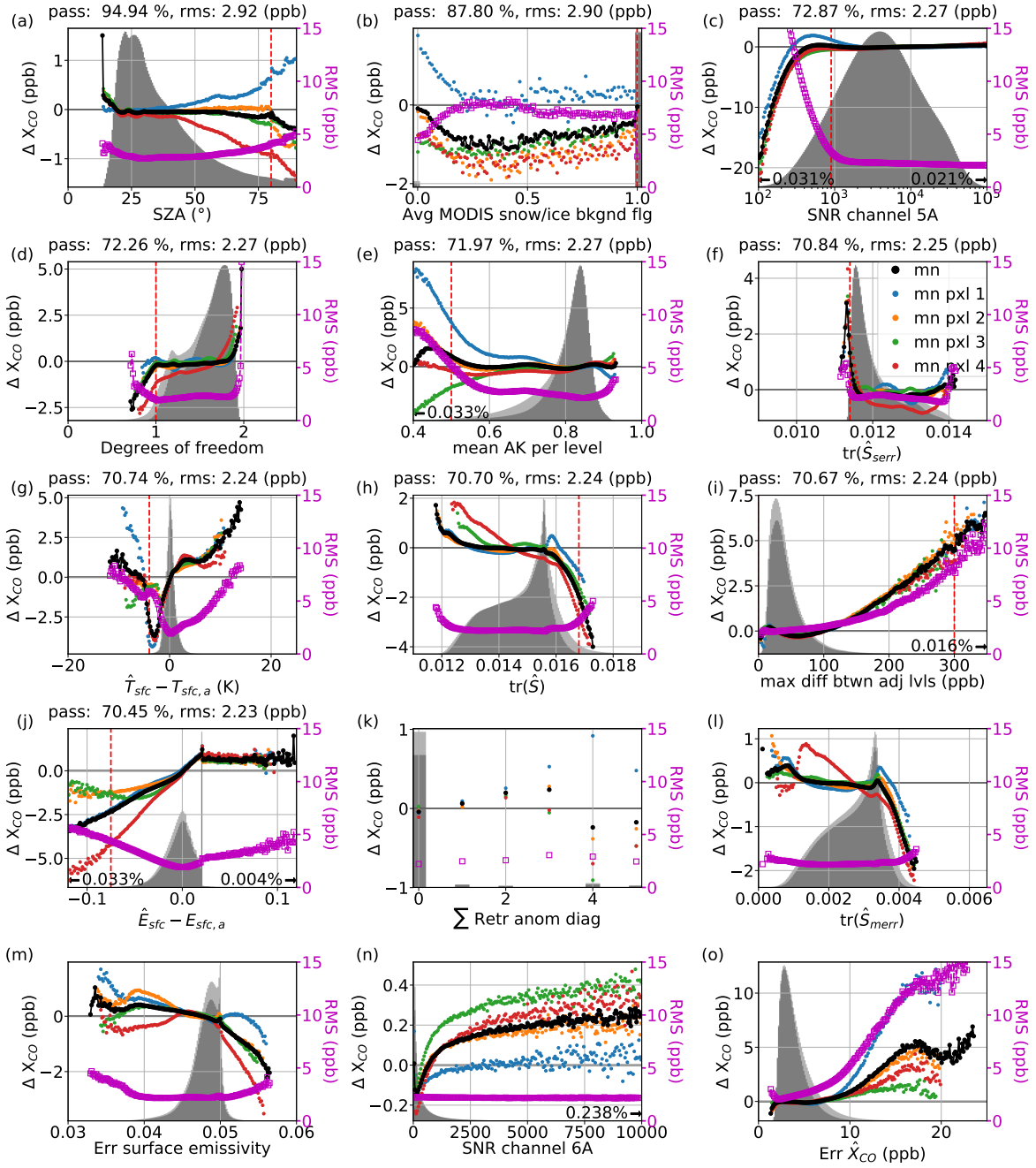

Figure S4. Diagrams showing the SRA bias as a function of one of the parameters in Table 3 in the main text for water. The black points show the overall mean bias (minimum 2000 points), the magenta points show the RMS, and the other points show the mean bias for the individual pixels (minimum 300 points). Red dashed lines indicate filter limits. (Continued on next page).

Figure S4. (Continued.) Lighter histograms are of all the data. Darker histograms are of data not excluded by the any previous filter. Percentages with arrows at the bottom are for data outside the shown range. Data percentages at the top are for data that remain after current filter is taken into account. QC filters applied to the first 10 plots. Plots are shown left to right and top to bottom in terms of more stringent or important filtering. Without any filtering and up to 90° SZA $n = 3.69 \times 10^8$ RMS=2.93 ppb.

S4 SRA relationships plots

We consider a variety of features for finding filter limits and checking for feature related bias based on the small region approximation (SRA). Figures S4 and S5 show fields that we opted to filter on, along with a few extra fields (such as SNR 6A) which we did not. Fields that were filtered on have text at the top. Various features of these plots are analyzed to determine if and where cutoffs should be made including the RMS, the overall mean bias, and the spread in bias among pixels. The underlying histograms give us a sense of how many observations will be removed at a certain cutoff. The mean SRA (black) also may be used to find systematic bias.

Figures S6, S7, S8, and S9 contains additional plots like those in Fig. S4–S5. The parameters along the x-axes in these plots were decided to be improper, or unfit to use as filter criteria.

The histograms in these figures allow us to check for the overall distribution of the data and find interesting features, of which we name a few. The difference in retrieved to surface emissivity over water ($\hat{E}_{\mathrm{sfc}} - E_{\mathrm{sfc,a}}$) is often 0.021 (Fig. S4j). This is because the a priori surface emissivity is 0.979 for all ocean scenes and the upper bound for $\hat{E}_{\mathrm{sfc}}$ is 1.000 as higher values are unphysical (Merritt Deeter, personal communications August 22, 2018). We can see an average offset of about 0.007 from zero for this parameter over land (Fig. S5f). We see a similar offset of about 6 K in $\hat{T}_{\mathrm{sfc}} - T_{\mathrm{sfc,a}}$ over land (Fig. S5m). Finally, the values of the histograms may be useful in themselves to get a sense of probability distributions. For example, distributions of degrees of freedom (Fig. S4d, S5k) and information bits (Fig. S6a, S8a) give a sense of the range of information content.

Mean bias and pixel biases have had a preliminary temporal bias correction applied to bring the values from individual pixels into better agreement. Without this, the truth proxy (local median) would often be a systematically biased value from pixel 1, leading to bimodal distributions for each pixel SRA anomaly. A feature-dependent bias correction was not considered until after generating these plots. The mean bias is only shown for at least 2000 points, and pixel biases are only shown if there are at least 300 points.

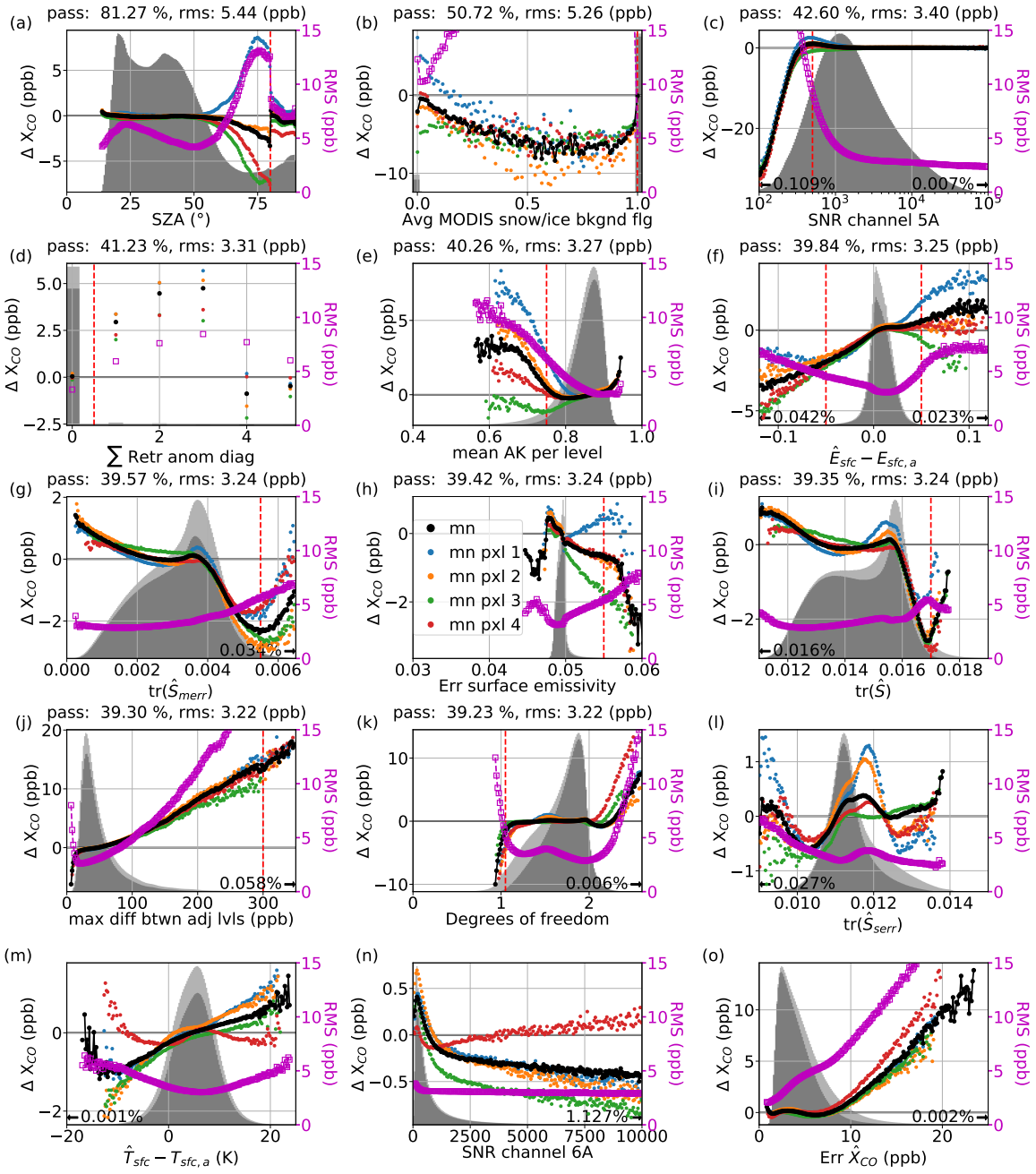

Figure S5. Similar to Fig. S4 but for land. QC filters applied to the first 11 plots. Without any filtering and up to 90° SZA $n = 2.31 \times 10^8$, RMS=5.46 ppb.

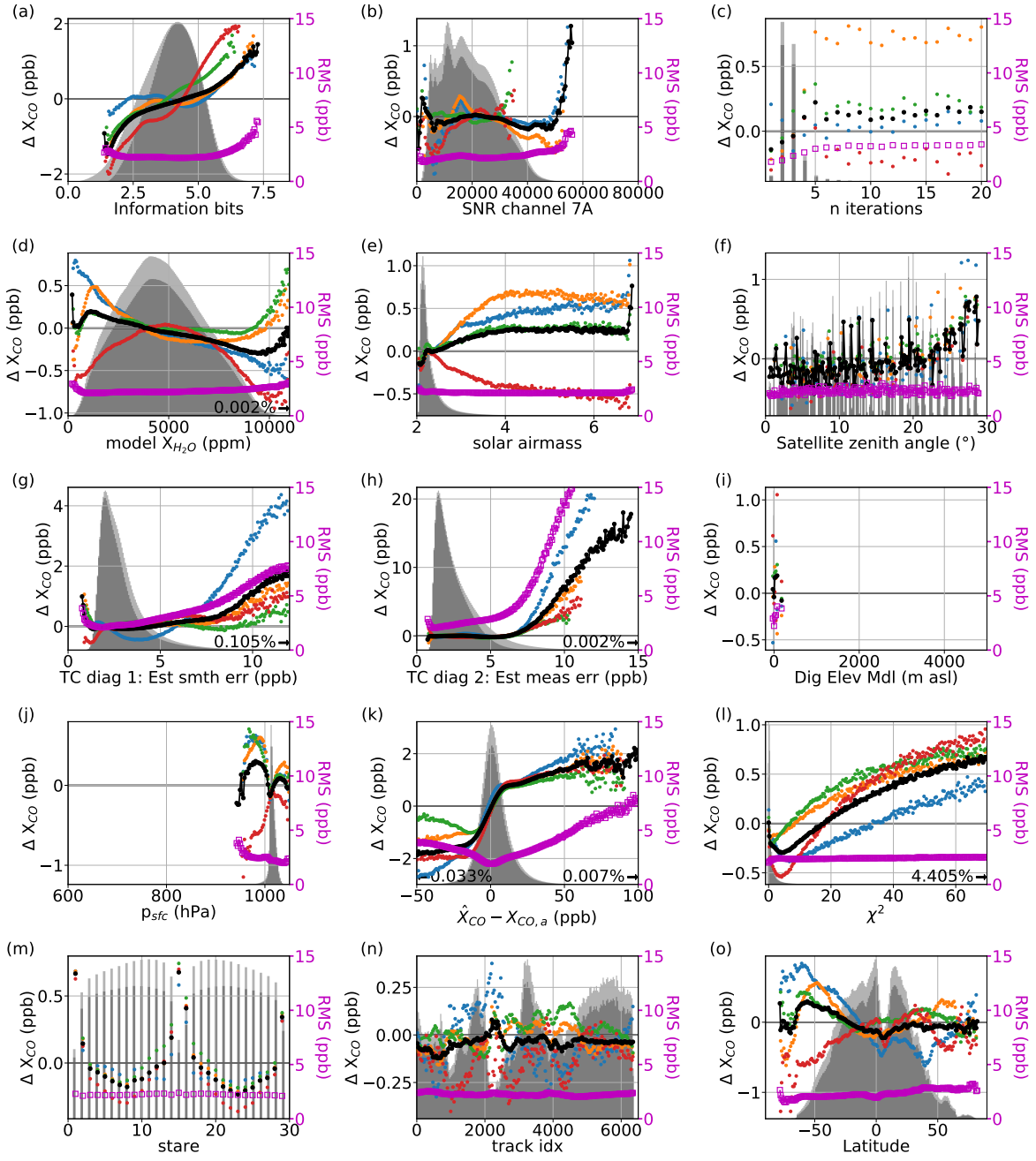

Figure S6. Additional fields considered for filters over water part I. Diagrams showing the SRA bias as a function of one of various parameters either included in the MOPITT files, derived from fields in the files, or linked separately. Some of these parameters are listed in Table 3 in the main text. See also the caption to main Fig. 3 and Fig. S4.

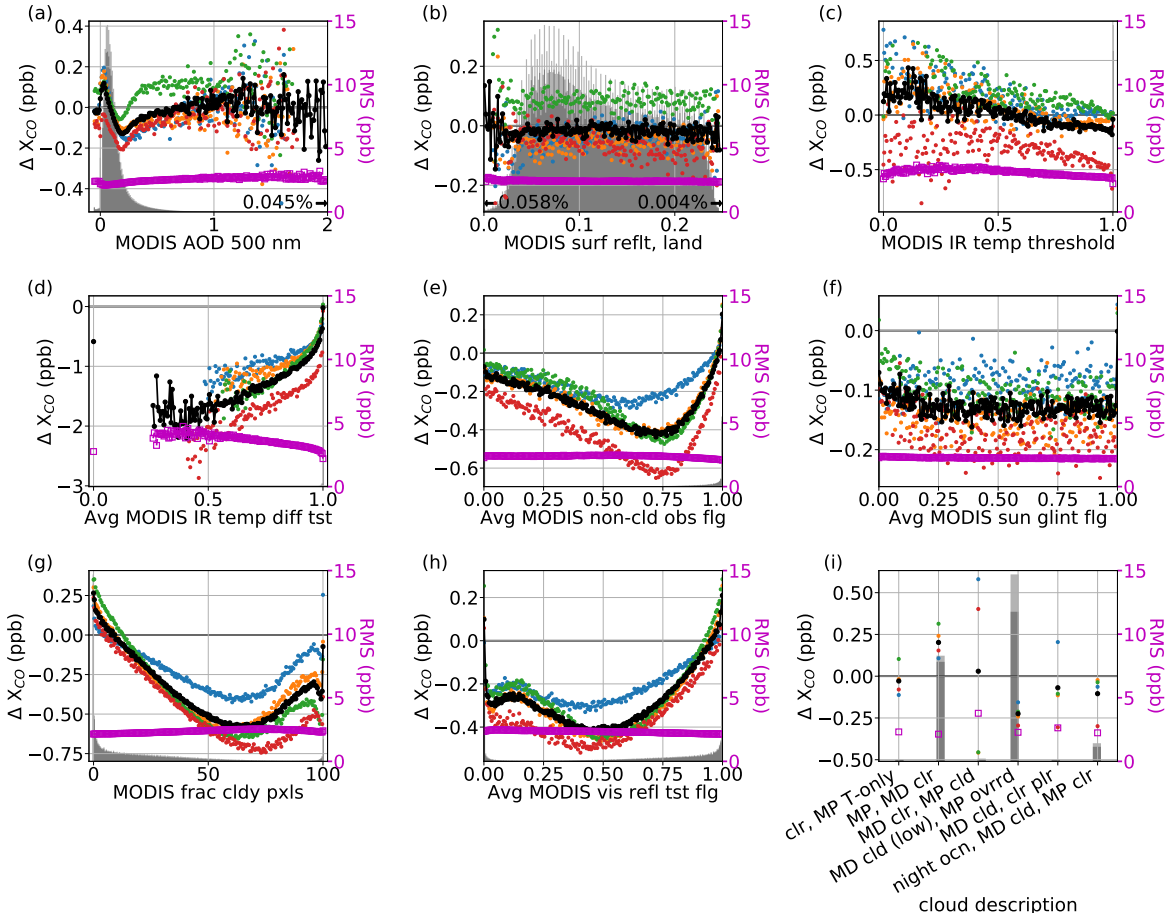

Figure S7. Additional fields considered for filters over water part II. These fields have some relation to MODIS. See also the caption to Figure S6.

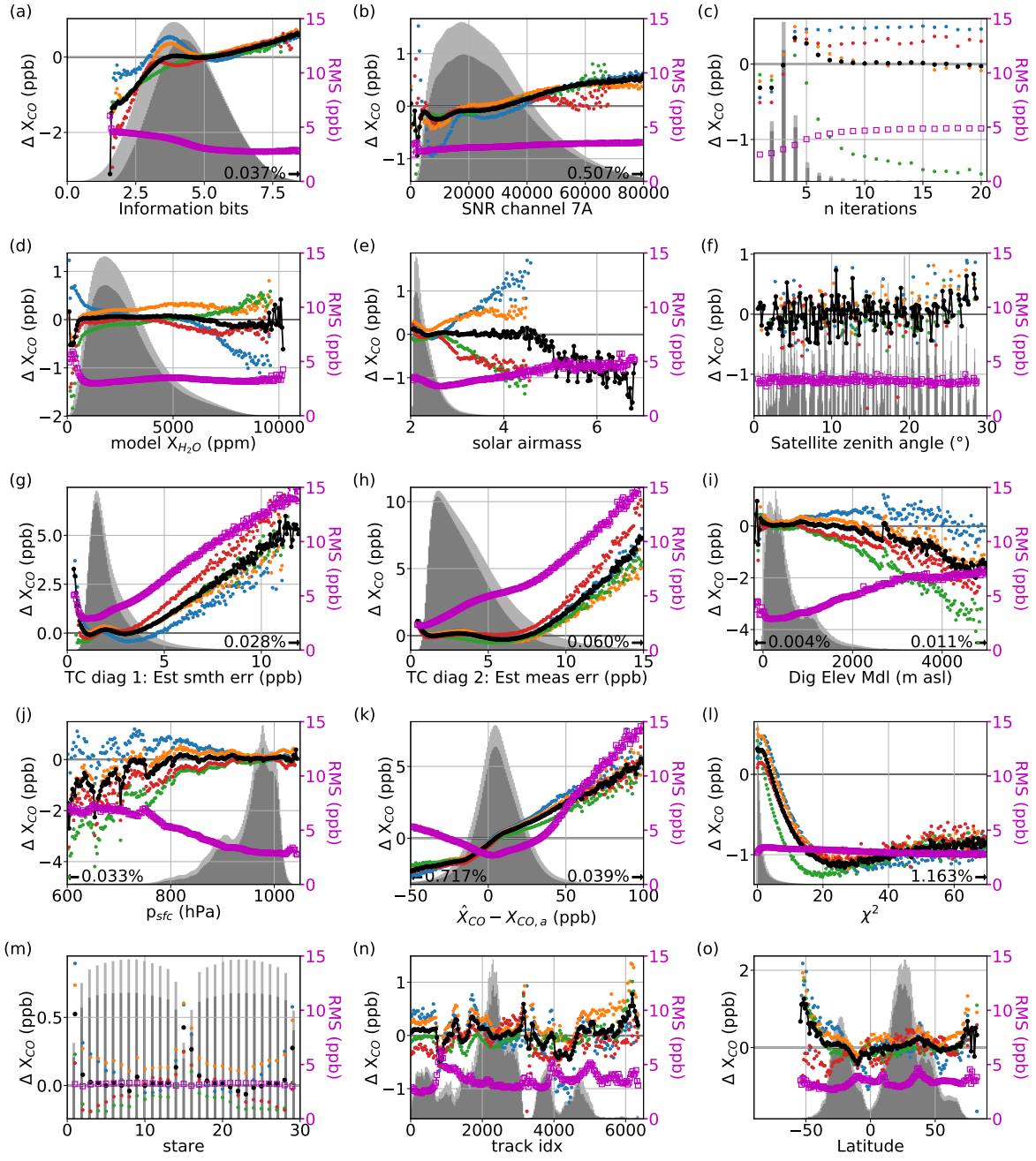

Figure S8. Additional fields considered for filters over land part I. See also the caption to Figure S6.

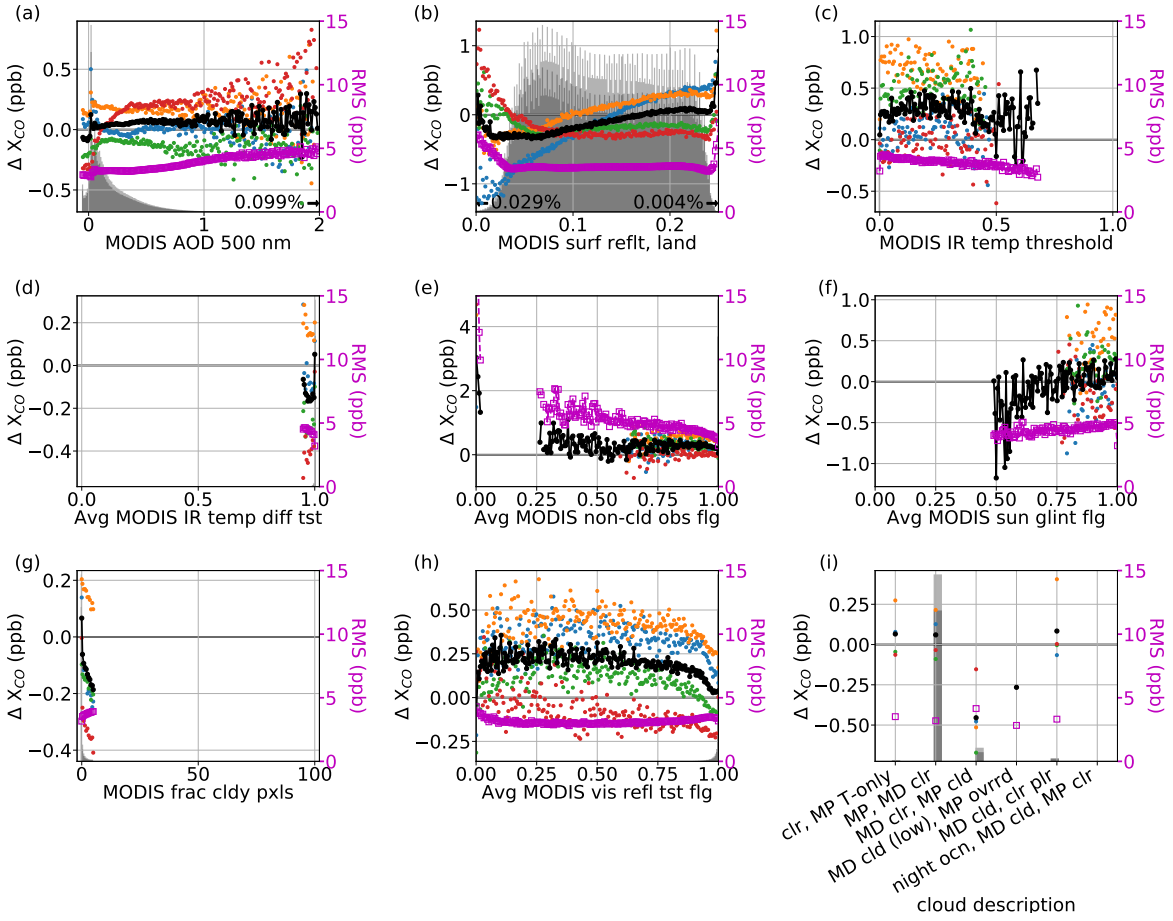

Figure S9. Additional fields considered for filters over land part II. These fields have some relation to MODIS. See also the caption to Figure S7.

S5 Spatial locations of filtered data

We plot the spatial variability of where data are filtered based on our criteria in Sec. 3.3 in the main text. Generally it is desirable for data to be filtered uniformly globally unless there is a known geographical dependent feature that could cause significant uncertainty or bias, such as snow or ice. Figures S10 and S11 show the geographical locations of filters over water and land respectively. There is some geographical dependence, including low averaging kernels near the North pole, low degrees of freedom near the South Pole, and large deviations of surface temperature from the prior around western coastlines. Over land, data from the Tibetan Plateau are screened by several filters. Filtering on the maximum difference between adjacent levels feature leads to a loss of some data over western China.

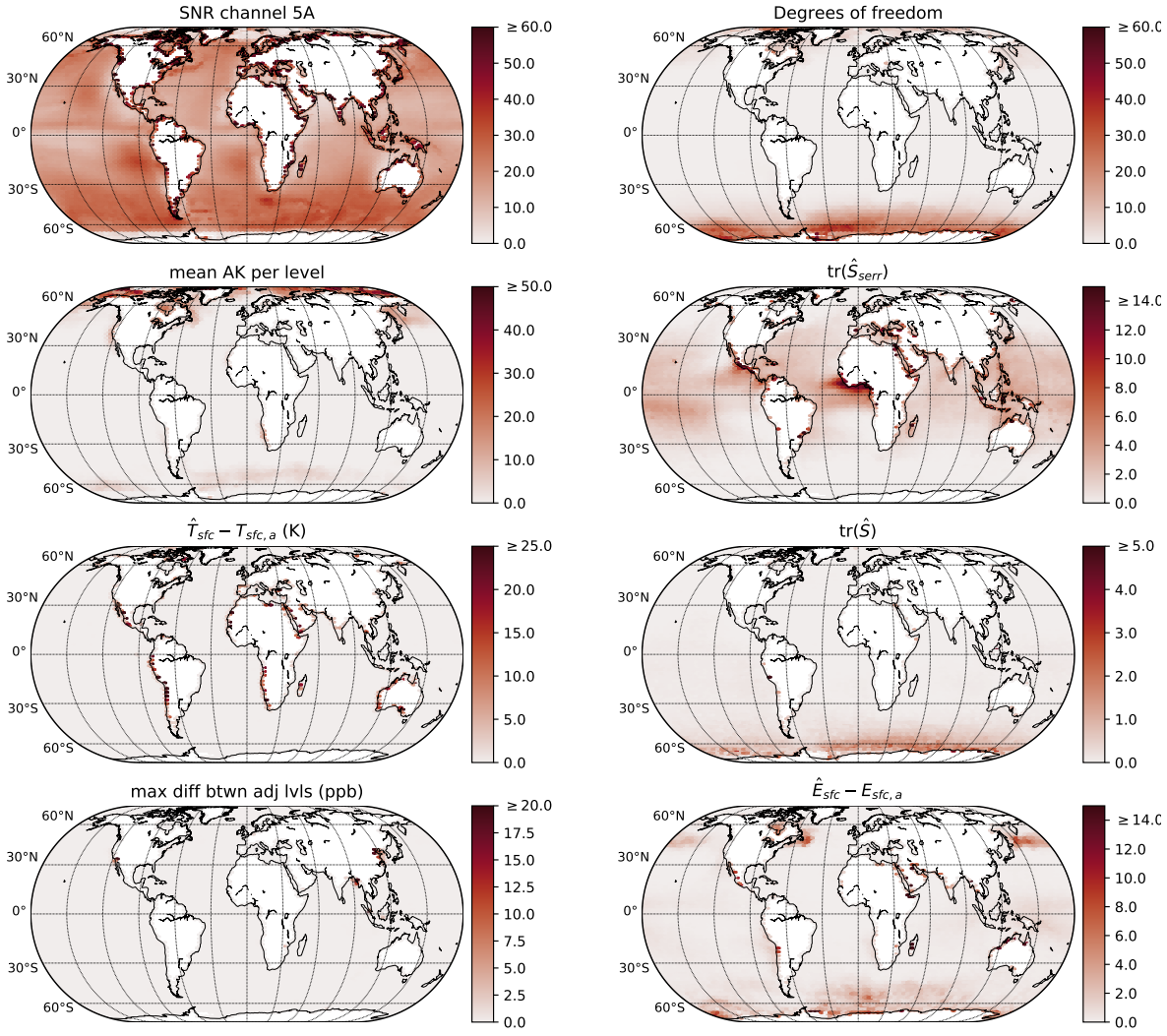

Figure S10. Maps of where data are filtered over water. Plots are of the percentage of data removed after all previous filters are applied. Order is first left to right, then top to bottom.

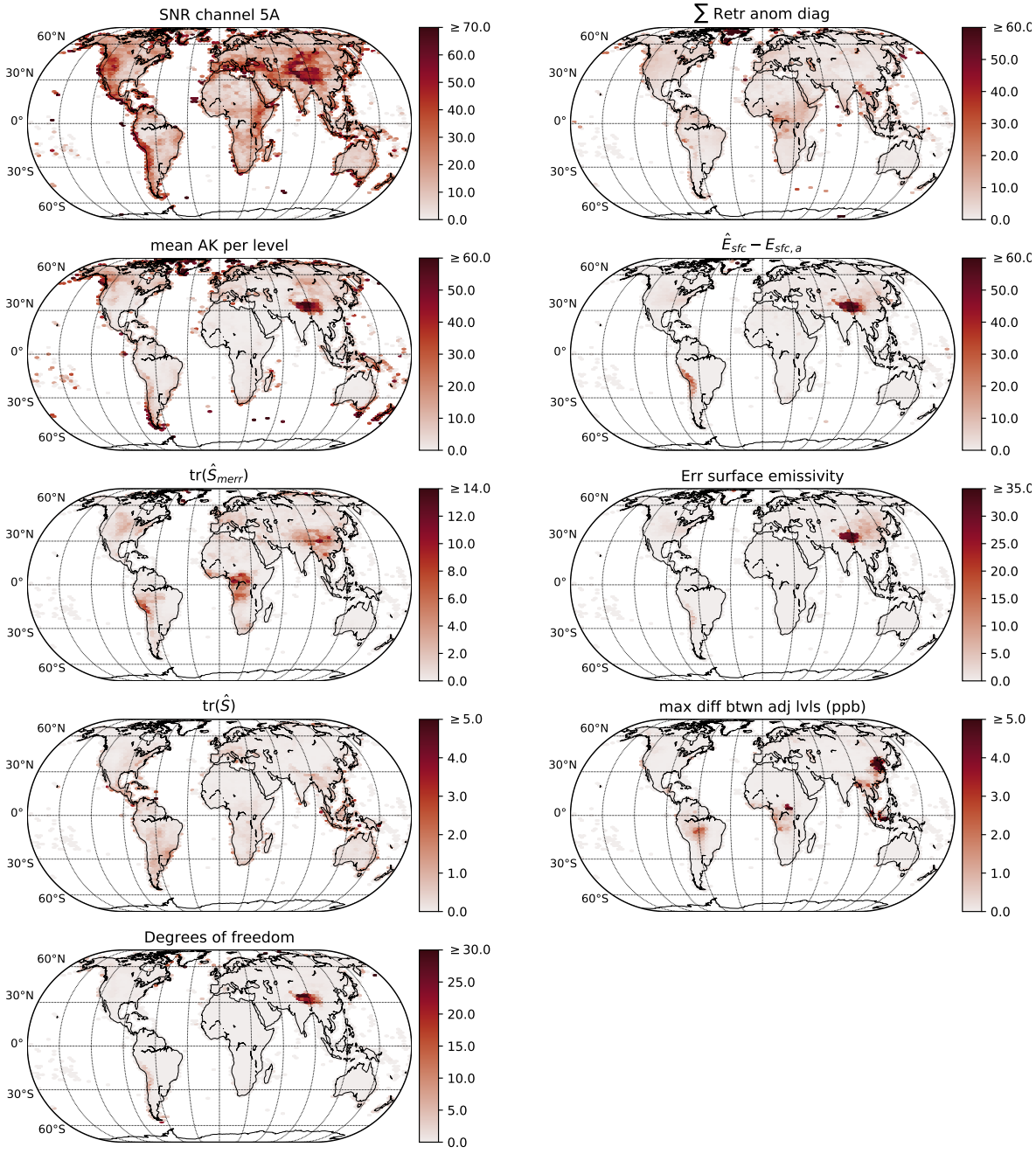

Figure S11. Maps of where data are filtered over land. Plots are of the percentage of data removed after all previous filters are applied. Order is first left to right, then top to bottom.

S6 Accounting for different averaging kernels and priors when comparing MOPITT and TCCON

While it is common to account for priors and averaging kernels when comparing measurements from different remote sounders, there is not universal agreement for how these differences should be taken into account. Here we formally describe different ways to do this for the MOPITT TCCON comparison. In Sect. S6.3 we show overall comparisons for different methods as well as other sensitivity tests. In Sect. S7 we show how different methods compare for each site.

S6.1 Adjust to a common prior and apply AKs from one system to other

Measurements from different remote sounders (e.g., MOPITT and TCCON) are not directly comparable due to their different averaging kernels and priors. There are various methods to account for these differences, for example the commonly used methods of Rodgers and Connor (2003), and further detailed by Wunch et al. (2011). Briefly, these methods are based on choosing a comparison ensemble profile ($\mathbf{x}_c$), which is often the prior of one of the sounders. Then, one of the retrieved profiles (which we call system 1) is taken and the averaging kernels of the other (system 2) are applied in the manner of

$$\hat{\mathbf{x}} = \mathbf{x}_a + \mathbf{A}\left(\mathbf{x} - \mathbf{x}_a\right) + \boldsymbol{\epsilon}_x\left(\mathbf{b}, \mathbf{c}\right). \tag{1}$$

These adjusted results are then compared with the retrievals from system 2.

This method raises the questions, which system priors should we use, and which system averaging kernels should be applied? Arguments for using TCCON are that the profile is finer at all levels except the surface, the algorithm has less of a tendency to overfit the radiances, and it has often been used in previous studies. An argument for using MOPITT is that the retrieval is a profile retrieval rather than column scaling. For the choice of $\mathbf{x}_c$, the MOPITT prior is 4D, while the TCCON is only 3D. However, if TCCON averaging kernels are applied then there is an argument for using it as $\mathbf{x}_c$. For MOPITT and TCCON these choices are formally listed in Table S1.

S6.2 Use an ensemble of "true" profiles

A difficulty to the above method is that multiple adjustments are required. For example, Method II in Table S1 involves adjusting both the MOPITT product and the TCCON product to be as if it were MOPITT, but using the TCCON a priori. The method also requires choosing which system to apply AKs to, which can have an effect on the comparison. Depending on $\mathbf{A}$, the $\hat{\mathbf{x}}$ profiles may be biased at certain levels. For example, consider the following 2-level example for 3 systems using the same prior where the true state is known:

$$\mathbf{x}_a = \begin{bmatrix} 1 \\ 1 \end{bmatrix},\ \mathbf{x} = \begin{bmatrix} 1 \\ 1.1 \end{bmatrix},\ \mathbf{A}_1 = \begin{bmatrix} 1.1 & 0 \\ 0 & 0.9 \end{bmatrix},\ \mathbf{A}_2 = \begin{bmatrix} 1.2 & 0 \\ 0 & 0.8 \end{bmatrix},\ \mathbf{A}_3 = \begin{bmatrix} 0.9 & 0 \\ 0 & 1.1 \end{bmatrix}. \tag{2}$$

Table S1. Equations to account for different priors and averaging kernels when comparing MOPITT and TCCON soundings.

Method	$\mathbf{x}_c$	$\mathbf{x}$ (assumed)	Comparison 1	Comparison 2
0			$\hat{c}_T$	$\hat{c}_M$
I	$\mathbf{x}_{T,\mathrm{a}}$	$\hat{\mathbf{x}}'_M = \hat{\mathbf{x}}_M + \mathbf{x}_{T,\mathrm{a}} - \mathbf{x}_{M,\mathrm{a}} + \mathbf{A}_M\left(\mathbf{x}_{M,\mathrm{a}} - \mathbf{x}_{T,\mathrm{a}}\right)$	$\hat{c}_T$	$\hat{c}_{T\leftarrow M'} = \hat{c}_{T,\mathrm{a}} + \sum_j^{70} h_j a_{T,j}\left(\hat{x}'_M - x_{T,\mathrm{a}}\right)_j$
II	$\mathbf{x}_{T,\mathrm{a}}$	$\hat{\mathbf{x}}_T = \gamma_T \mathbf{x}_{T,\mathrm{a}}$	$\hat{c}'_M = \hat{c}_M + c_{T,\mathrm{a}} - c_{M,\mathrm{a}} + \sum_j^{10} a_{M,j}\left(\log_{10} x_{M,\mathrm{a}} - \log_{10} x_{T,\mathrm{a}}\right)_j$	$\hat{c}_{M\leftarrow T} = c_{T,\mathrm{a}} + \sum_j^{10} a_{M,j}\left(\log_{10}\hat{x}_T - \log_{10} x_{T,\mathrm{a}}\right)_j$
III	$\mathbf{x}_{M,\mathrm{a}}$	$\hat{\mathbf{x}}_M$	$\hat{c}'_T = \hat{c}_T + c_{M,\mathrm{a}} - c_{T,\mathrm{a}} + \sum_j^{70} h_j a_{T,j}\left(x_{T,\mathrm{a}} - x_{M,\mathrm{a}}\right)_j$	$\hat{c}_{T\leftarrow M} = c_{M,\mathrm{a}} + \sum_j^{70} h_j a_{T,j}\left(\hat{x}_M - x_{M,\mathrm{a}}\right)_j$
IV	$\mathbf{x}_{M,\mathrm{a}}$	$\hat{\mathbf{x}}'_T = \frac{\hat{c}'_T}{c_{M,\mathrm{a}}}\mathbf{x}_{M,\mathrm{a}}$	$\hat{c}_M$	$\hat{c}_{M\leftarrow T'} = c_{M,\mathrm{a}} + \sum_j^{10} a_{M,j}\left(\log_{10}\hat{x}'_T - \log_{10} x_{M,\mathrm{a}}\right)_j$

For method 0, priors, averaging kernels, and surface pressures are not considered. Subscript T=TCCON, M=MOPITT, a=a priori, the prime symbol represents an adjustment if another prior were used, and the left arrow indicates what would be observed from one system in the absence of error, if the profile from another was taken as truth. The MOPITT column averaging kernel is defined in Eq. **??**, the TCCON column averaging kernel is $a_{T,j} = \frac{1}{h_j}\frac{\partial \hat{c}_T}{\partial x_j}$, and the TCCON retrieval scaling factor is $\gamma_T = \frac{\hat{c}_T}{c_{T,a}}$.

Then, if $\epsilon_x = 0$ for all three systems

$$\hat{\mathbf{x}}_1 = \begin{bmatrix} 1 \\ 1.09 \end{bmatrix},\ \hat{\mathbf{x}}_2 = \begin{bmatrix} 1 \\ 1.08 \end{bmatrix},\ \hat{\mathbf{x}}_3 = \begin{bmatrix} 1 \\ 1.11 \end{bmatrix},\ \hat{\mathbf{x}}_{12} = \hat{\mathbf{x}}_{21} = \begin{bmatrix} 1 \\ 1.072 \end{bmatrix},\ \hat{\mathbf{x}}_{13} = \hat{\mathbf{x}}_{31} = \begin{bmatrix} 1 \\ 1.099 \end{bmatrix},$$
$$\hat{\mathbf{x}}_{23} = \hat{\mathbf{x}}_{32} = \begin{bmatrix} 1 \\ 1.088 \end{bmatrix},\ \hat{\mathbf{x}}_{11} = \begin{bmatrix} 1 \\ 1.081 \end{bmatrix},\ \hat{\mathbf{x}}_{22} = \begin{bmatrix} 1 \\ 1.064 \end{bmatrix},\ \hat{\mathbf{x}}_{33} = \begin{bmatrix} 1 \\ 1.121 \end{bmatrix}. \tag{3}$$

where $\hat{\mathbf{x}}_{12}$ is retrieval 2 smoothed with the averaging kernel of 1. In this example $\hat{\mathbf{x}}_1$ is less comparable to $\hat{\mathbf{x}}_{12}$ than to $\hat{\mathbf{x}}_2$, but is more comparable to $\hat{\mathbf{x}}_{13}$ than $\hat{\mathbf{x}}_3$. In all cases, when the system is compared with itself (e.g., $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_{11}$) the comparison becomes worse. The choice of which profile to apply the AKs to can affect the comparison, and the differences are summarized in Table S2.

In some cases it may be straightforward to make an assumption of a set of true profiles corresponding to soundings to be intercompared. For this project we can readily obtain a set of eight including: $\mathbf{x}_{T,\mathrm{a}}$, $\mathbf{x}_{M,\mathrm{a}}$, $\hat{\mathbf{x}}_T$, $\hat{\mathbf{x}}_M$, $\frac{\hat{c}_M}{c_{T,a}}\mathbf{x}_{T,\mathrm{a}}$, $\frac{\hat{c}_T}{c_{M,a}}\mathbf{x}_{M,\mathrm{a}}$, $\frac{\hat{c}_T}{\hat{c}_M}\hat{\mathbf{x}}_M$, and $\frac{\hat{c}_M}{c_{M,a}}\mathbf{x}_{M,\mathrm{a}}$. In the case of 2 profile retrievals:

$$\boldsymbol{\epsilon}_{x,1} - \boldsymbol{\epsilon}_{x,2} = (\hat{\mathbf{x}_1} - \mathbf{x}_{a,1} + \mathbf{A}_1\mathbf{x}_{a,1} - \hat{\mathbf{x}}_2 + \mathbf{x}_{a,2} - \mathbf{A}_2\mathbf{x}_{a,2}) - \mathbf{A}_1\mathbf{x} + \mathbf{A}_2\mathbf{x} \tag{4}$$

Table S2. Differences between retrieved and smoothed profiles (ppb).

	$\mathbf{x}_{i1}$	$\mathbf{x}_{i2}$	$\mathbf{x}_{i3}$
$\mathbf{x}_1 - \mathbf{x}_{1j}$	0.009	0.018	−0.009
$\mathbf{x}_2 - \mathbf{x}_{2j}$	0.008	0.016	−0.008
$\mathbf{x}_3 - \mathbf{x}_{3j}$	0.011	0.022	0.011

Differences between systems (second row) with AKs applied. Rows represent system 1 (or i), columns represent system 2 (or j). System 2 is smoothed with the AKs of system 1 and compared to system 1 retrievals.

Table S3. Simulated differences between retrievals assuming a true profile (ppb).

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_a$
$\boldsymbol{\epsilon}_1 - \boldsymbol{\epsilon}_2$	0.001	0.002	−0.001	0.010
$\boldsymbol{\epsilon}_1 - \boldsymbol{\epsilon}_3$	−0.002	−0.004	0.002	−0.020
$\boldsymbol{\epsilon}_2 - \boldsymbol{\epsilon}_3$	−0.003	−0.006	0.003	−0.030

Differences in error between pairs of observing systems (row) using a variety of assumed truth profiles (columns). Differences in error between systems with the same averaging kernels and prior (e.g., $\boldsymbol{\epsilon}_1 - \boldsymbol{\epsilon}_1$) is always zero.

If the 2 retrievals are not on the same levels then regridding is necessary, for example using a method that conserves layer mass (Appendix C in main text). More specifically for MOPITT and TCCON column retrievals

$$\boldsymbol{\epsilon}_{x,M} - \boldsymbol{\epsilon}_{x,T} = \left(\hat{c}_M - c_{a,M} + \mathbf{a}_M^T \log_{10}\mathbf{x}_{a,M} - \hat{c}_T + c_{a,T} - \sum_j^{70} h_j a_{T,j} x_{T,j} \right) - \mathbf{a}_M^T \log_{10}\mathbf{x} + \sum_j^{70} h_j a_{T,j} x_j, \tag{5}$$

and all the terms relevant to the TCCON retrieval use the TCCON surface pressure, and the terms relevant to the MOPITT retrieval use the MOPITT surface pressure. This way biases from different surface pressures cancel out. The terms in the parenthesis only need to be computed once per pair of compared soundings. For the column retrieval case, regridding is only necessary for the set of profiles treated as truth. Table S3 lists the differences in retrieval errors between systems for a variety of truth profiles. The magnitude of the error depends on the assumption of true profile and may be greater or less than the differences in Table S2.

S6.3 MOPITT and TCCON comparisons for different methods and sensitivity tests

In addition to the main comparison (Fig. 5 in main text), we perform a variety of sensitivity tests shown in Fig. S12. Results are described in Sect. 4.2 in the main text.

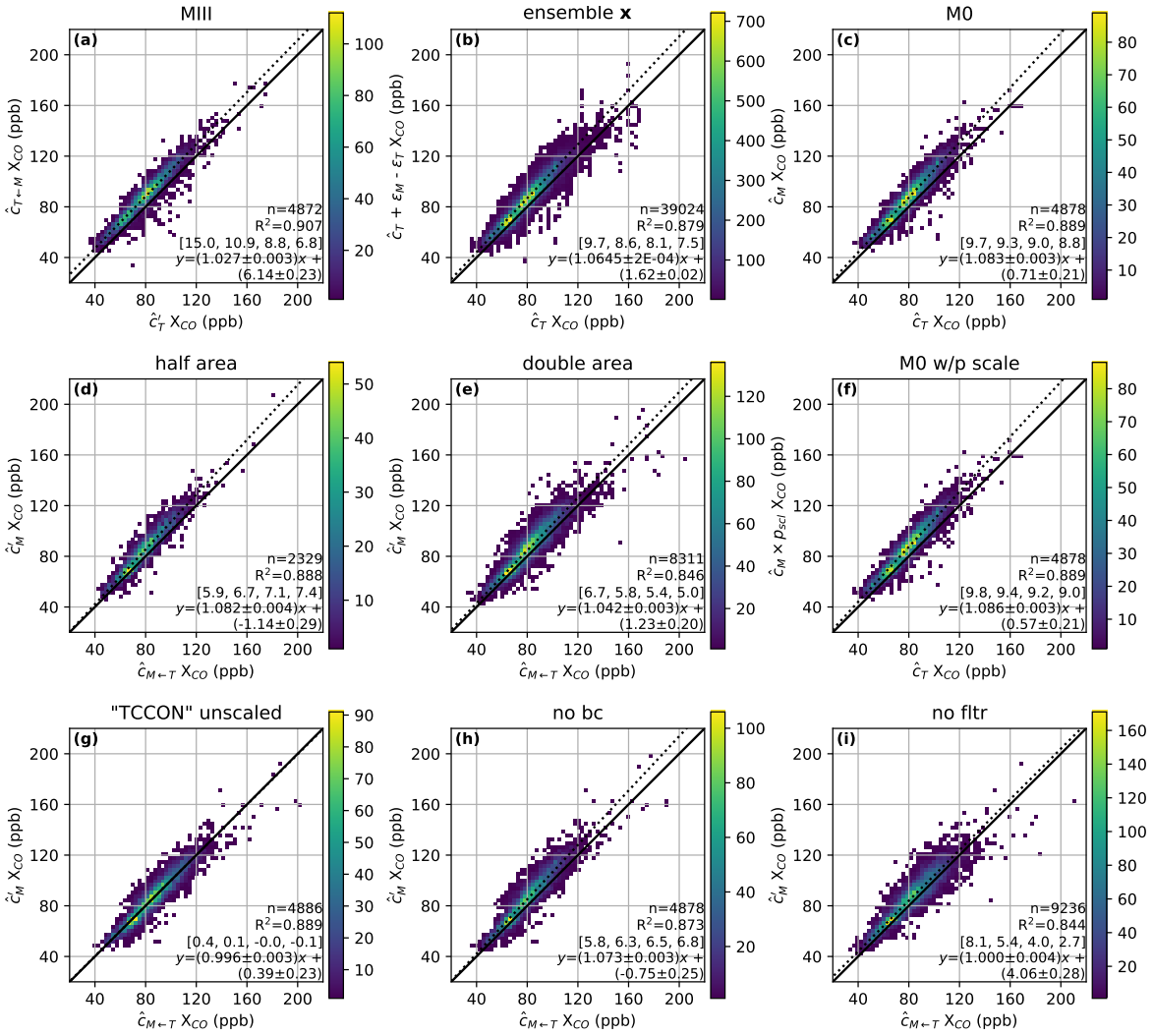

Figure S12. See Table S1 for a list of different methods with TCCON data on the x-axis, and MOPITT on the y-axis. Text is number of points or days n, coefficient of determination for ordinary least squares regression R^2, bias (in %) at 50, 75, 100, and 150 ppb using the shown fit, and equation for the shown fit using the methods of York et al. (2004). Unless specified, derived filters and bias corrections for MOPITT are applied. Colorbars indicate number of points in heatmap bins. **(a)** Using method III. **(b)** Using a variety of simulated profiles at "truth" to estimate TCCON and MOPITT retrieval errors (see Sect. S6.2). **(c)** Using method 0 (i.e., not accounting for averaging kernels). **(d)** Same as main Fig. 5 but with half-sized spatial bins. **(e)** Same as main Fig. 5 but with double-sized spatial bins. **(f)** Same as (c), but scaling MOPITT data to account for differences in surface pressure compared with the TCCON site (Fig. 4). **(g)** Using TCCON data without an overall scaling factor to tie to the WMO scale, and without empirical corrections for airmass. **(h)** Without bias corrections for MOPITT observations. **(i)** Without filtering MOPITT observations.

S7 Comparisons for each TCCON site

In Fig. S13 are the mean biases and standard deviation for each site using a variety of different methods to account for the different a priori profiles and averaging kernels. In general, there is not a clear and persistent difference from the various methods, possibly with the exception of the first and fourth columns (methods 0 and III), which on average appear larger.

Because the MOPITT to TCCON comparison can depend on the site, we include individual plots for sites here. In Fig. S14 the data are shown as annual plots with boxplots for each month. There is a wide variety of behavior depending on site. Many of the sites appear to have a seasonal cycle. However, we do not find a consistent pattern in these so we do not devise a bias correction based on a seasonally varying parameter.

S8 Bias compared with other data

We also check for any systematic trends in the MOPITT to TCCON difference with several co-observed fields of interest. A systematic trend could occur for example in the case of interfering species. Results are shown in Fig. S15. The interference of N_2O on the MOPITT retrieval has been discussed by Deeter et al. (2017). In the V7 algorithm a global linear trend with time is used for N_2O levels. When we subtract similar values ($y = 0.06082x + 293.5$, where x is months starting from January 1970) first the correlation increases slightly to R^2=0.09, and the magnitude of the slope increases to $-0.47\frac{X_{\mathrm{CO\,ppb}}}{X_{\mathrm{N_2O\,ppb}}}$. This is much larger than an estimate of $-0.13\frac{X_{\mathrm{CO\,ppb}}}{X_{\mathrm{N_2O\,ppb}}}$ based on the sensitivity test by Deeter et al. (2017) where the N_2O was increased in the V6 retrieval by 3 % which led to an decrease in the total column of about 1.2 ppb (based on Fig. 1 therein).

S9 Comparisons of retrieved to prior values

The MOPITT prior is based on monthly averages from a Community Atmospheric Model with Chemistry (CAM-Chem) run from 2000–2009 (Buchholz et al., 2017). The model was gridded at 1° resolution. Assuming a linear change in emissions over model duration, 2004–2005 are the center years. If emissions have decreased from the 2004–2005 model years we would expect to see, to a first order, smaller retrieved than prior values. This is complicated by the averaging kernel matrix, as a change in the mixing ratio at one level can significantly affect the retrieved amount at other levels. For MOPITT profile retrievals, it is not uncommon to have larger off-diagonal elements than those on diagonal. A comparison of the retrieved to prior ratios (on a $\log_2$ scale) is shown in Fig. S16.

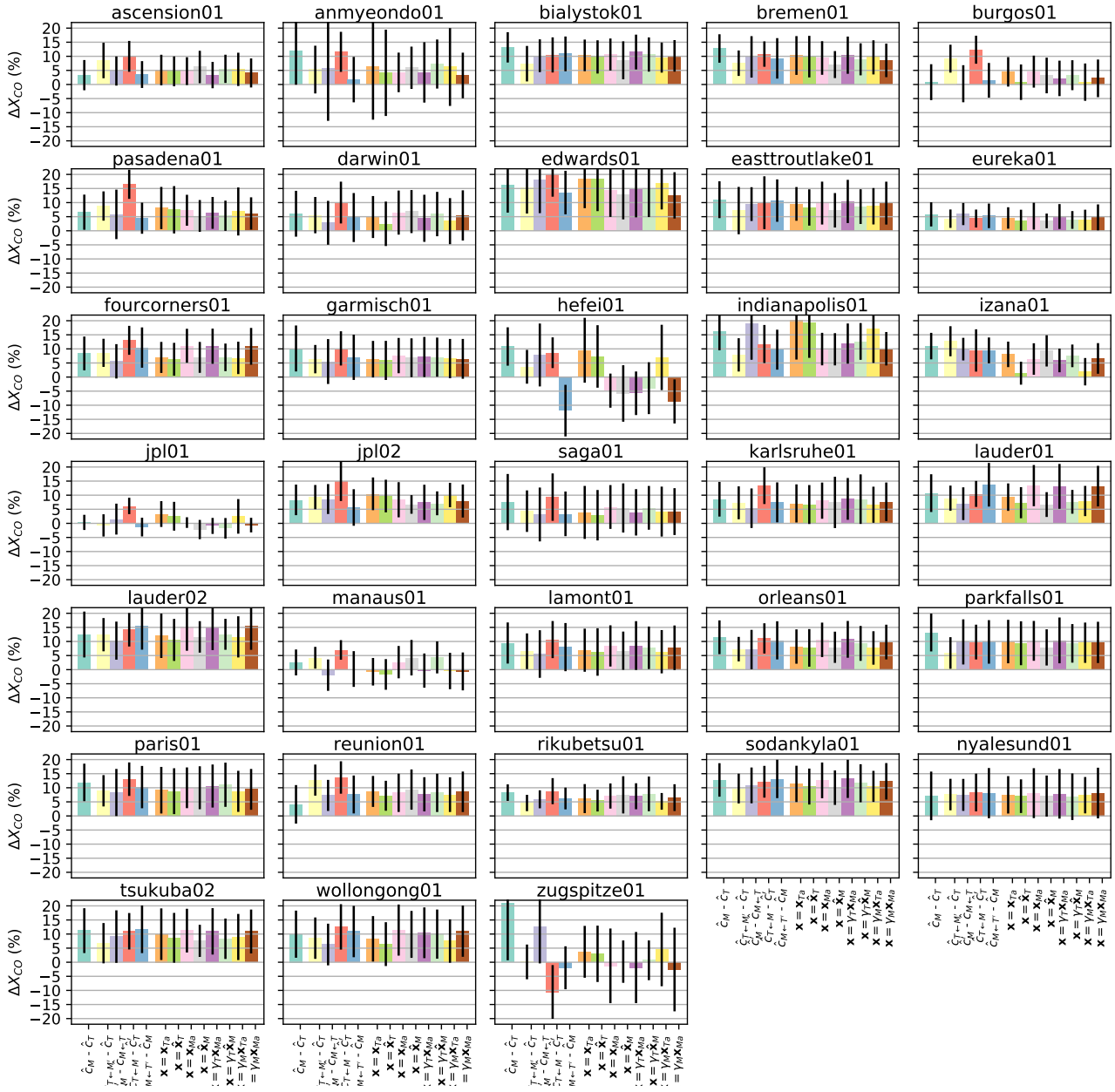

Figure S13. Average bias and standard deviation (black lines) for each site using the different methods described in Sect. S6. Differences are converted to percent by dividing by the mean TCCON values. The bars on the far left are from a direct comparison of TCCON to MOPITT. The next four bars are from using a variety of different adjustments described as methods I–IV in Table S1. The last eight bars are from treating a variety of different as "truth" and simulating the difference in errors ($\epsilon_M - \epsilon_T$). Sorted alphabetically.

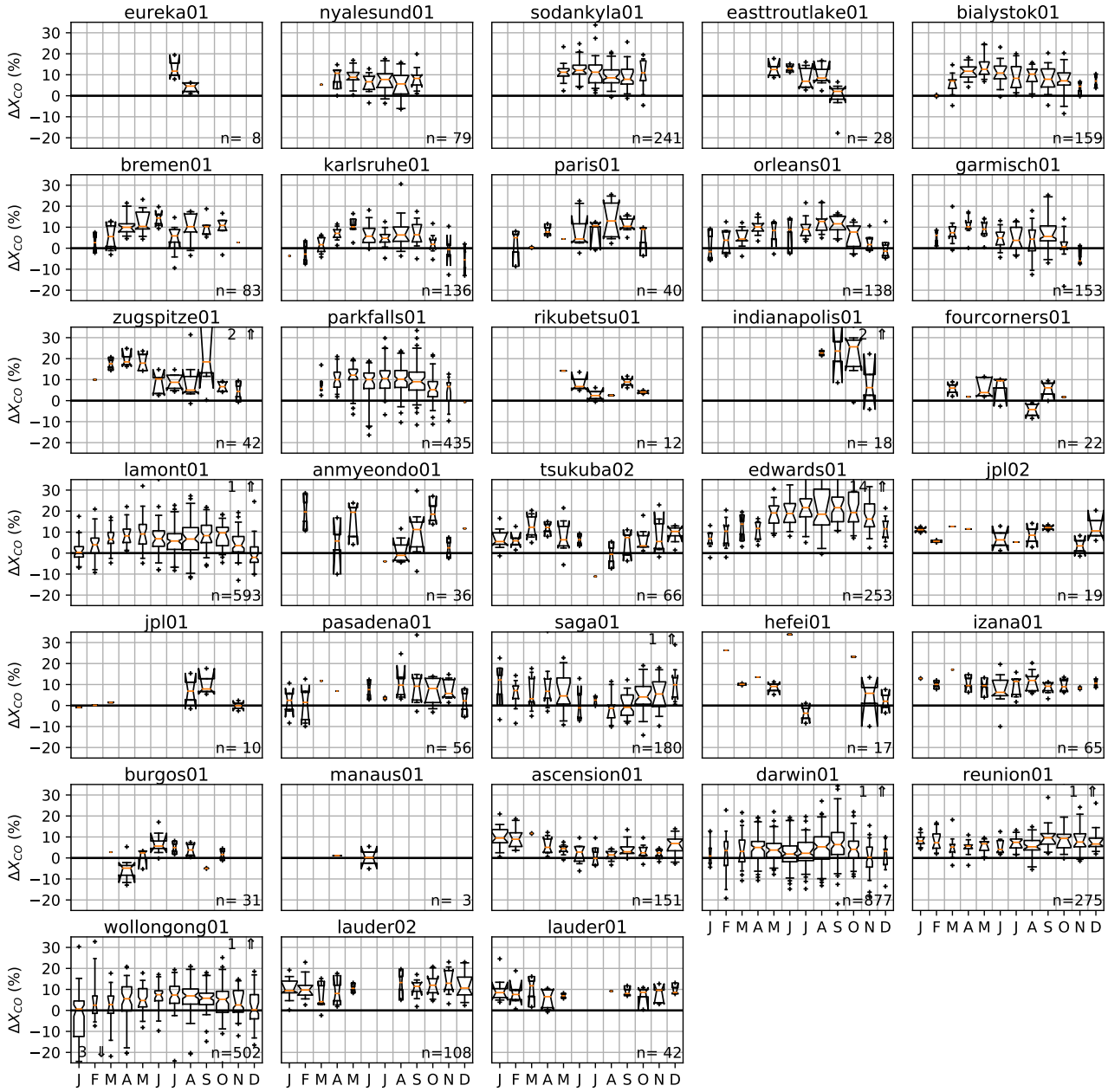

Figure S14. Seasonal variability in the difference between MOPITT and TCCON for all sites using method II. Data from all years are averaged for each month and shown as a box and whisker plot. Orange bars represent medians, and whiskers represent the central 95 % of data. Widths represent the relative amount of observations compared to other months for a given site. Notches represent the 95 % confidence interval on the median, as determined using the bootstrap method ($n = 5000$). Ordinate limits are fixed, so outliers are noted with numbers. Sorted by latitude.

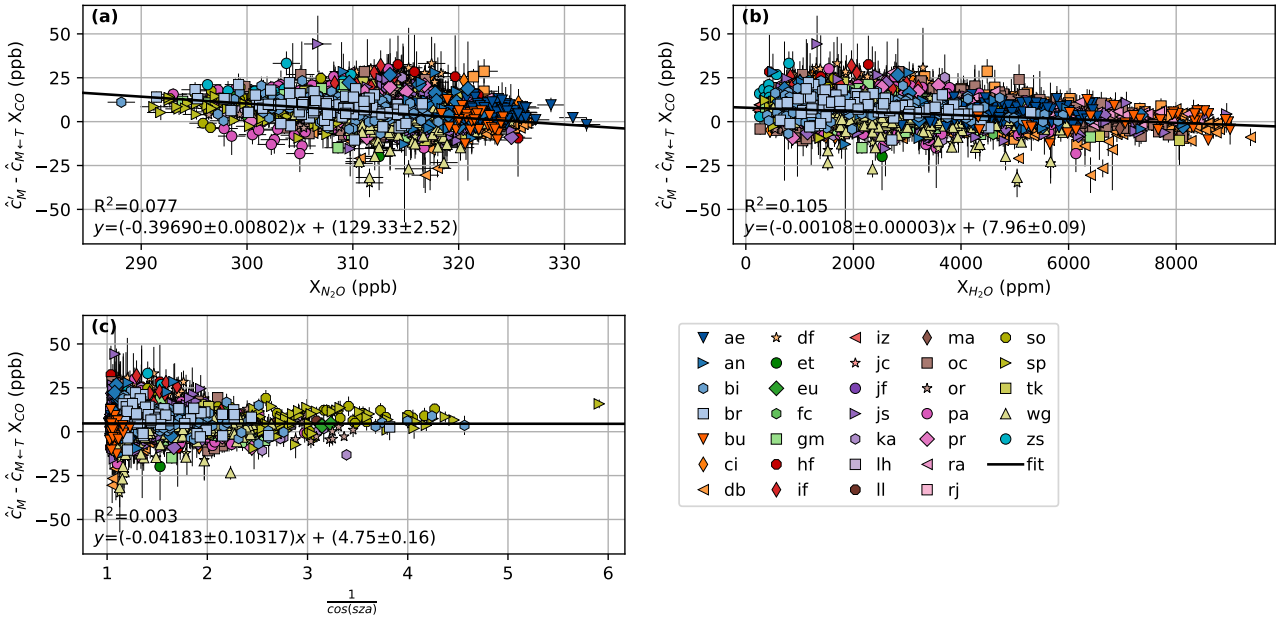

Figure S15. MOPITT−TCCON differences using method II compared with fields of interest from the TCCON. R^2 values are for an ordinary least squares regression. Fits shown are from using the methods of York et al. (2004). **(a)** Versus X_{N_2O} retrieved from the TCCON. **(b)** Versus X_{H_2O} retrieved from the TCCON. **(c)** Difference data are plotted against an approximation of airmass.

References

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

Deeter, M. N., Edwards, D. P., Francis, G. L., Gille, J. C., Martínez-Alonso, S., Worden, H. M., and Sweeney, C.: A climate-scale satellite record for carbon monoxide: The MOPITT Version 7 product, Atmospheric Measurement Techniques, 10, 2533–2555, https://doi.org/10.5194/amt-10-2533-2017, 2017.

GAW: GAW Report No . 192 Guidelines for the Measurement of Atmospheric Carbon Monoxide, Tech. Rep. 192, World Meteorological Organization, Global Atmospheric Watch, 2010.

Novelli, P. C., Masarie, K. A., Lang, P. M., Hall, B. D., Myers, R. C., and Elkins, J. W.: Reanalysis of tropospheric CO trends: Effects of the 1997–1998 wildfires, Journal of Geophysical Research, 108, 4464, https://doi.org/10.1029/2002JD003031, 2003.

Rodgers, C. D. and Connor, B. J.: Intercomparison of remote sounding instruments, Journal of Geophysical Research, 108, 4116, https://doi.org/10.1029/2002JD002299, 2003.

Wunch, D., Wennberg, P. O., Toon, G. C., Connor, B. J., Fisher, B., Osterman, G. B., Frankenberg, C., Mandrake, L., O'Dell, C., Ahonen, P., Biraud, S. C., Castano, R., Cressie, N., Crisp, D., Deutscher, N. M., Eldering, A., Fisher, M. L., Griffith, D. W. T., Gunson, M., Heikkinen, P., Keppel-Aleks, G., Kyrö, E., Lindenmaier, R., Macatangay, R., Mendonca, J., Messerschmidt, J., Miller, C. E., Morino, I., Notholt, J.,

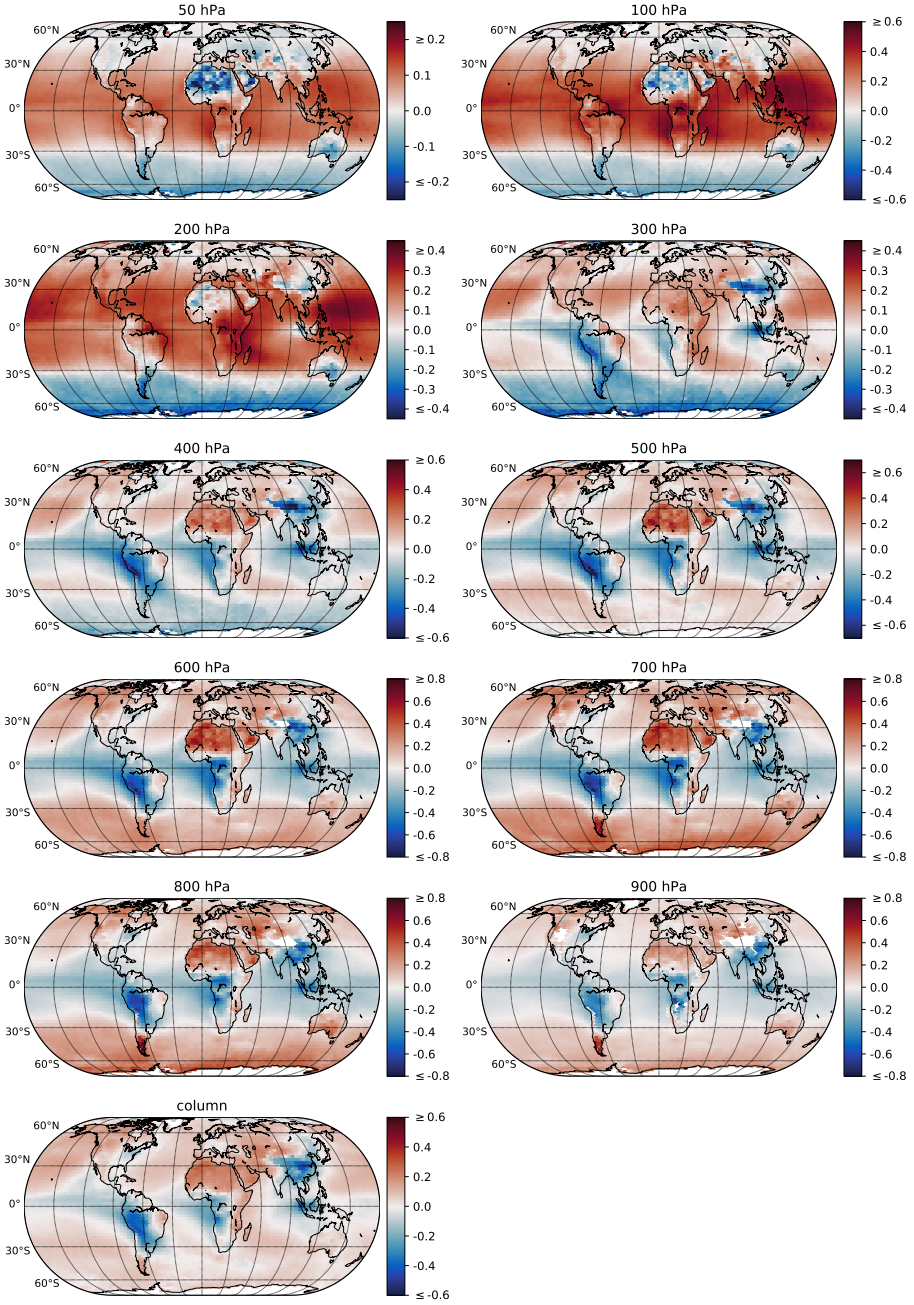

Figure S16. Ratios of retrieved to prior values averaged over all years (2002–2017). Labels for the first 10 plots are the top of the level (e.g, 300 hPa represents 300–400 hPa). Ratios are plotted on a $\log_2$ scale, i.e., 0 indicates $\hat{c} = c_a$, 1 indicates $\hat{c} = 2c_a$, and -1 indicates $\hat{c} = 0.5c_a$.

Oyafuso, F. A., Rettinger, M., Robinson, J., Roehl, C. M., Salawitch, R. J., Sherlock, V., Strong, K., Sussmann, R., Tanaka, T., Thompson, D. R., Uchino, O., Warneke, T., and Wofsy, S. C.: A method for evaluating bias in global measurements of CO2 total columns from space, Atmospheric Chemistry and Physics, 11, 12 317–12 337, https://doi.org/10.5194/acp-11-12317-2011, 2011.

York, D., Evensen, N. M., Martinez, M. L., and De Basabe Delgado, J.: Unified equations for the slope, intercept, and standard errors of the best straight line, American Journal of Physics, 72, 367–375, https://doi.org/10.1119/1.1632486, 2004.

Zellweger, C., Buchmann, B., and Steinbrecher, R.: System and performance audit of surface ozone, carbon monoxide, methane and nitrous oxide at the global GAW station Jungfraujoch Switzerland, March 2015, Tech. rep., World Meteorological Organization, https://www.wmo.int/pages/prog/arep/gaw/documents/JFJ_2015.pdf%0A%0A, 2017.

Zhou, M., Langerock, B., Vigouroux, C., Sha, M. K., Hermans, C., Metzger, J.-M., Chen, H., Ramonet, M., Kivi, R., Heikkinen, P., Smale, D., Pollard, D. F., Jones, N., Velazco, V. A., García, O. E., Schneider, M., Palm, M., Warneke, T., and De Mazière, M.: TCCON and NDACC XCO measurements: difference, discussion and application, Atmospheric Measurement Techniques Discussions, pp. 1–24, https://doi.org/10.5194/amt-2019-266, 2019.