We thank the reviewers for providing feedback on our manuscript. We address their comments below.

Reviewer comments are in bold-type.

Our responses are italicized.

We use tracked changes to show ~~changes~~ edits to the manuscript.

Jacob Hedelius (on behalf of all coauthors)

Reviewer #1

R1-1: **However, to some degree this paper is extremely detail-heavy and short on high-level take-aways. The paper could be enhanced by adding some additional high-level conclusions in terms of a) which areas the filters and bias correction help most in, and (b) the real impact of the proposed filters and bias correction in terms of inverse flux modelers at both local scales, as well as regional/global scales. (Ie, do they matter more in the tropics, or high latitudes; more over land vs. ocean, etc).**

We agree the previous conclusion needed to be updated to include take-aways. We have rewritten it and moved the previous text to a new section. Our inversion was too coarse to quantify impact at local scales, but seems to be acting similarly on all regions.

In this study quality filtered and bias corrected MOPITT data are compared with TCCON data. We first derive filters using only the MOPITT data assuming homogeneity over small regions. These filters have the largest affect over snow/ice scenes and over high terrain. They reduce the overall RMS from 3.84 ppb to 2.55 ppb. We find and correct a bias among the four pixels, which we confirmed exists using AirCore. We also find and correct a feature bias.

After the filtering and bias correction, we compare with TCCON data. Using a method (method II, see SM Sect. 6.1) similar to Wunch et al., (2011) to account for differences in priors and AKs we find MOPITT data are biased high by about 6% compared with TCCON, but it is not clear which of MOPITT or TCCON is biased. We also test different methods which all lead to a bias of about 6—10%. There is a trend of -0.06±0.06% yr^{-1} in the MOPITT-TCCON difference. The bias also appears to depend on site and latitude, but the scatter is not consistent enough to derive a correction. We also compared AKs and information content from the different retrievals. TCCON AKs are more sensitive to changes in the stratosphere. MOPITT AKs peak in the mid-troposphere and can vary significantly among locations.

After applying filtering, and an overall scaling to match the TCCON, we assimilate the data into GEOS-Chem. Filtering and bias correction are uniform enough to not make a large difference among regional fluxes. When data are also scaled down to TCCON before implementing into

GEOS-Chem, fluxes were lower in all regions. However, because of bottom-up uncertainties in global CO fluxes these experiments were inconclusive. Additional work is needed to understand the relatively large (~6%) difference between MOPITT and the TCCON.

R1-2: **Sect 2.2 (TCCON), page 6 : Please state the estimated possible mean biases per site (in ppb), and variable errors (noise plus faster systematic errors). You currently only list this single 4% number, which presumably corresponds to the latter.**
Empirical estimates of TCCON XCO site-to-site bias are underway from a campaign in 2018 (Pak et al., 2019 TCCON meeting, New Zealand). The systematic 4% number is also made up of longer-term systematic errors that could vary site-to-site due to e.g., misalignment.
GGG2014 XCO systematic errors for TCCON are below 4% (Wunch et al., 2015). The uncertainty on the scaling slope is 6% (2σ).

R1-3: **Do you consider colocation errors with TCCON, or their possible size? For CO from local emission sources, it seems like colocation errors could be large, depending on your colocation criteria.**
Yes, but only in an overall global bias sense. See former Fig. 6 d-e. In general changing the colocation criteria size (doubling or halving) increased the intercept. While the bias for the base colocation criteria was around 6.4%, halving led to a bias of ~5.9–7.4%, and doubling led to a bias of 5–6.7%. Based on R2-1 we've moved this figure to the supplement but increased discussion in the main text.
We also examine how the scaling changes for different colocation criteria in SM Fig. S12d--e by halving and doubling the coincidence areas. We find MOPITT is biased higher than TCCON in these tests by 5--7.4%. Doubling the area decreases R^2 for the global comparison.

R1-4: **Section 2.3: Please state something about the accuracy and precision of AirCore CO measurements.**
AirCore CO is still a developmental product, so we made the following addition with the caveat that a full uncertainty budget has not yet been quantified.
AirCore CO is still a developmental product with a sample measurement precision typically less than 5 ppb (Engel et al., 2017). However, stratospheric AirCore CO profile comparisons have shown differences as large as 20 ppb, which could be a result of diffusion in stratospheric AirCore samples, AirCore surface effects, or incorrect AirCore sample end member assumptions. Accuracy is dependent on the quality of calibration and standards (see SM Sect. 2).

R1-5: **Phase "truth metric" seems an imperfect term. I would prefer truth estimate or truth proxy. "metric" doesn't convey the imperfection of the actual data sources you will use.**
Agreed. We've changed to "proxy" throughout.

R1-6: **S3, SM, and associated text: What was the source of data you used to construct these plots? Presumably some high-resolution model?**

These are generated from the MOPITT data itself. Unfortunately we do not have access to a high-resolution XCO model.

We examine the effects of the bin size on the small region approximation (SRA) using the MOPITT observations.

R1-7: **TCCON-MOPITT trends over the 13 years?**

See response to R1-12.

R1-8: **Figure S4 (filter plot):**

- **R1-8A: How can 108.13% of points pass? This problem also exists in S5.**
  Before this percentage was compared to SZA and the snow/ice masks determined before the SAA filtering, so in this case the percentage was more than 100 because we were disregarding the snow/ice mask. We've updated the percentages to be compared with all observations up to 90 degrees so they are now under 100%.

- **R1-8B: Why do you show the MODIS snow/ice flag if it passes everything and appears to do nothing?**
  This was because the histogram bars were too narrow to notice, and the distribution is nearly bimodal. We've increased the bin width on the ends to improve visibility.

- **R1-8C: SNR, plot c, consider plotting SNR on a log scale so we can see the behavior at low SNR more clearly.**
  This is an excellent idea, we've updated this subplot.

- **R1-8D: Plot g, why not also have an upper limit on Tsfc – Tsfc,a ?**
  Limits are all somewhat subjective, and reconsidering it does seem reasonable to have an upper limit here. Because the rest of the analysis was without it we don't include it at this stage, but data users may consider one.

- **R1-8E: Also in Table 3. Please state what the three tr(S) values are, physically, including units.**
  Interpretation physically is more obscure than linear covariance matrices, but is related to uncertainty. We have described these in the table
  $\text{tr}(\widehat{S})$ unitless due to log scale, related to overall uncert. of retrieval combination of meas & smooth matrices
  $\text{tr}(\widehat{S}_{merr})$ from uncert. in prior & weighting
  $\text{tr}(\widehat{S}_{smth})$ from uncert. in measured radiances

- **R1-8F: Sum(Retr anom diag) – why is there no histogram?**
  We have increased bin widths to improve visibility.

○ **R1-8G: Do these filter plots include any bias correction (pixel and/or feature)? Why or why not?**
They include a pixel BC, but not a feature BC. We've added the following.
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.

R1-9: **Figure S5**
○ **R1-9A: Please fix/add histograms for panels b and d.**
Done.
○ **R1-9B: Given the strong bias and scatter driven by "max diff btwn adj levels", panel j, why not more strongly filter on this variable? Based on the criteria you set in section 3.3, (2 ppb bias, 6 ppb scatter), it seems like a reasonable cut-off would be more like 150 ppb (over land), rather than 300 ppb.**
These are somewhat subjective, and there was internal discussion over this. If readers decided on a 150 ppb cutoff we would not disagree. The hope was that because the relationship looked linear that data could be bias corrected instead of filtered out. In practice the scatter was still large after the bias correction.

R1-10: **Page 14, end of section 3: please describe the "maximum difference between adjacent levels" variable physically, and why it might be correlated with bias.**
We've added the following:
This feature is larger for strong gradients between levels which can appear when there are strong surface fluxes, or when the retrieval is unstable and oscillates. This instability may be caused by bias related to e.g., spectroscopic errors.

R1-11: **Section 4.1 - please state if there a noticeable improvement when using the altitude adjustment in comparing to TCCON, in particular for the highest altitude sites, or sites in mountainous regions.**
In aggregate there is no improvement, as enough high and low values balance out. However, there may be large differences for individual sites.
In this intercomparison, we implicitly account for differences in surface pressure using the **h** vector. ̶T̶his can make a difference for individual sites by as much as -10.5±4.1 (1σ) ppb (for Zugspitze). However, we have found in practice that accounting for differences in surface pressure makes little difference here on the overall comparison (compare SM Fig S12c and S12f). In aggregate the difference is only -0.2±1.5 (1σ) ppb.

R1-12: **Page 19, please state at what level you can rule out a long-term trend between MOPITT & TCCON, at for example 95% confidence (2 sigma). 0.1 %/yr ? 0.01 %/yr? Etc.**

We thank the reviewer for pointing out this oversight. Our previous conclusion of no trend was based only on a visual comparison. When we fit a line despite the scatter, there is a slight negative trend similar to the one found in Deeter et al., 2017.

~~We do not find a clear long term trend in the MOPITT$ $TCCON difference.~~ Though there is significant scatter among individual comparisons, we find a long-term trend of -0.06±0.06% yr^{-1} in the MOPITT-TCCON difference using the Thiel-Sen estimator.

R1-13: **Page 20, top: Are any of the observed seasonal biases mitigated using your 1-parameter "feature" bias correction?**

There does not appear to be a significant change in seasonal biases from before and after this correction.

R1-14: **Page 21 bottom to page 22, top: I'm not sure what the DOF and information content analysis is adding. You state numbers but not something that we qualitatively learned. Suggesting removing, or add a sentence or two to true to put the numbers into some kind of context for what we actually learn from them.**

We've added the following:

DOF gives an indication of how many independent parameters can be improved compared with the prior. MOPITT DOFs are between 1 and 2, which indicates total column measurements may be reasonable, but individual layer measurements may not always be accurate.

R1-15: **Page 23, top: Because you only assimilate MOPITT, it's not clear what the difference is between assimilations 2 and 3. Are the bias corrections different between the two, i.e. one is still biased high by 6% relative to (bias corrected) TCCON, and one is not? Please state what the purpose is to have three tests rather than just two. Ie, you are proposing a new filtering and bias correction relative to the standard MOPITT product, so the logical thing to do here would be to test those two things. You are testing three things, so please state clearly why.**

Yes, assimilation 3 is still biased high relative to the bias corrected TCCON. While it still has the footprint/pixel BC and single "feature" BC, it can be thought of as nearly being not bias corrected compared to TCCON. We've updated the explanation.

We conducted three experiments in which we assimilated the following datasets: 1) the original MOPITT data, 2) the filtered and bias-corrected data with a scaling down of about 6% to match the standard TCCON data (Fig. 5) (referred to as Assim. 2), 3) the filtered and bias-corrected data with a scaling of less than 0.5% to the TCCON-based data not tied to the WMO scale and without the empirical airmass correction (SM Fig. S12g~~Fig. 6g~~) (referred to as Assim. 3).

R1-16: **"Assimilation 2 results are lower." Lower in terms of what, inferred CO fluxes? Please clarify.**

Here, we meant mole fractions. We've modified this sentence.

Assimilation 2 ~~results~~ mole fractions are lower compared to HIPPO than the other assimilations.

R1-17: **Page 24: Finally, in your comparison to HIPPO, it's not clear if any assimilation is clearly "better" in terms of comparison to HIPPO to any other. Please discuss this somewhere in the paper. Currently, your readers cannot tell if your new filtering and bias correction made any real improvement to the MOPITT data in terms of inferred fluxes.**

Unfortunately the results are mixed and hence inconclusive. It's difficult to say which of assimilations are "better" as the comparisons are imperfect. We've added the following paragraph.

These results are inconclusive as to which of the assimilations is best. Comparisons with HIPPO mole fractions are mixed, and uncertainties in the assimilated prior fluxes prevent us from making definitive conclusions from the flux comparison. It is unclear if the filtering and bias corrections improved fluxes in these experiments.

R1-18: **Appendix D & Section 4.4: In terms of the need to adjust to common priors, and correct for AK differences you never state (1) the general size of these adjustments (in % or ppb), or (2) the effect on the apparent quality of the TCCON vs. MOPITT comparison. Does the comparison improve when you employ one or both of these corrections? You have gobs of equations and figures but not just a sentence or two saying something like "In general, these corrections are critically important in order to obtain meaningful results" or "In general, these corrections are small and may be in practice ignored", or something in between. Sometimes readers want to skip the equations and numbers and get to the main point!**

We agree that these sections did not have enough high level conclusions. Per R2-1 comments, we moved Appendix D to the supplement. We have added the following to Section 4.4

We take differences in priors and averaging kernels into account following method II described in the SM Sect. S6.1. Corrections are applied to each MOPITT retrieval and to daily averages of TCCON retrievals within coincidence criteria. We find in practice that corrections change the comparison by about 3%. TCCON data are adjusted by 0.7±1.8 (1σ) ppb, and MOPITT data are adjusted by -1.0±3.1 (1σ) ppb.

R1-19: **Table D3: what are the units? Ppb? % ?**

Units are ppb or whatever units the state vector is in. We've added to the table title (same for Table D2).

Grammar/Technical:
Page 8: "As a second example, empirical corrections" - please add comma after the leading clause.
Page 9, "To examine the effects of averaging kernels" – please add comma after leading clause.
Page 12: "To reduce the likelihood of overfitting, " – please add comma after leading clause.
Fixed, thanks.

Page 13: What is the approximate size of the AK correction typically? Can you apply the AK correction for all MOPITT soundings as compared to TCCON?
See response to R1-18. We apply the AK correction to each individual MOPITT sounding.

Reviewer #2

R2-1: **The level of detail and the discussion of various methods applied for the intercomparison studies make the paper quite lengthy. I feel that the readability could be enhanced if parts of these discussions would be moved to the appendix or the supplement. In particular, a large number of correlation plots of MOPITT versus TCCON XCO are presented in Section 4.2, each of them dealing with the differences in a priori and AKs of both instruments in a different way, with the different methods summarised in Table D1 being difficult to understand. However, there is very little discussion of the impact of these different methods and, in my opinion, not much to learn for the reader. I would therefore suggest to restrict the discussion to the method that has been chosen as standard, and to move Figure 6 to the appendix or supplement.**
We agree the paper is quite lengthy, and did have some internal discussion where parts were moved to the supplement. We have moved Figure 6 and Appendix D to the supplement. These were included for completeness – when we presented this work in meetings we were often asked what would happen if we applied the AKs in the opposite way. Each of the methods has its relative merits, and we could simulate situations where they worsen the comparison. We've opted for Method II as standard because it is most similar to what has been used in the past. We have increased discussion on these in the main text.

R2-2A: **The manuscript concludes with useful technical suggestions for filtering and bias correction of MOPITT data, and with suggestions for additional information that could enhance future MOPITT data products. What is missing are quantitative statements on the level of agreement between MOPITT and TCCON CO measurements using the different methods presented,**
See response to R1-1.

R2-2B: **an assessment of which method can be recommended,**
We now mention method II in the conclusion. This is most similar to a method that has been used in previous studies.

R2-2C: **as well as a summary of the lessons learned from the model assimilation studies presented in Section 5.**
R2-2D: **Which of the different assimilations proposed here would be the 'best'?**
We refrain from stating which of the assimilations, if any, is 'best.' It is difficult to state the best assimilation because the outputs (mixing ratios and fluxes) are themselves compared only with imperfect values.

Technical Comments

P2, L21: 'It acts as an indirect greenhouse gas (GHG) as a minor source of CO2...': please check sentence structure

Modified to:

It acts as an indirect greenhouse gas (GHG) as <u>both</u> a minor source of CO_2, and by…

P4, L10: The fact that 'MOPITT data are the longest satellite record of atmospheric CO' has already been stated earlier (P3, L7) and should be removed here.

Fixed, thanks.

P5, L23: Insert 'that' between 'They noted' and 'the surface type'

Fixed.

P6, L13: I suggest to remove 'eliminates' with 'is insensitive to' (two times)

Replaced both occurrences.

P6, L15: 'The network has been around since...' -> 'The network has been established in...'

Fixed.

P6, L12: The statement that MOPITT data are subject to errors is trivial since this applies to any measurement.

Good point. We've modified to lead sentence.

Despite these advantages, <u>there are known sources of uncertainty that could be biasing the measurements</u><s>the data are still subject to errors</s>.

P6, L30: 'long length' -> 'large length'

Opted to omit 'long' instead.

P8, L18: A first order approximation is always linear (at least if it is based on a Taylor expansion, which I guess is what the authors mean).

Modified to:

<s>To a first order, the biases </s><u>Biases</u> in **b** and **c** <u>may be approximated as having a</u> <s>have a</s> linear effect on x (Rodgers, 1990).

P8, L30: Please specify what you mean with 'statistically reasonable products'.

Here we meant that while we cannot say much about bias for an individual measurement, in aggregate they may tell us about scatter and bias.

Rather than using proxies that work for each measurement, <s>statistically reasonable products can be used</s><u>we aggregate many measurements</u> to empirically identify artifacts and outliers.

P8, L32: Do you mean 1° × 1° latitude/longitude regions?

Yes, we do mean an area rather than length. Replaced with "~100 km × 100 km"

Figure 2: It is not clear to me what is shown on the y-axis. The axis label states it would be the difference between retrieved and prior XCO, whereas it is stated on P9, L17 that it would be the difference to the weighted mean from all sensors.

Great catch. The y-axis label was incorrect, we've changed to "pxl - mean X_{CO} (ppb)"

P12, L4: I have difficulties to understand this sentence. Do you mean that the bias is determined by subtracting the median from all the XCO measurements within a specific region?

Yes.

To calculate anomalies, w~~W~~e subtract the median from all the points within that region.

P12, L33: Replace '×' with 'times' (here and anywhere else).

Modified 6 occurrences where '×' was not followed by a number.

Figure 7: It is not clear which of the various methods to account for a priori and AK differences has been applied to calculate the relative differences between MOPITT and TCCON shown here.

Thanks, we've clarified in the caption and in the main text.

Figure 7 shows boxplots of the MOPITT to TCCON differences (using method II) for each site for land only and water only soundings.

Boxplots of the MOPITT-TCCON percent difference at the TCCON sites (using method II), ordered by latitude (degrees north in parenthesis).

P21, L31: I don't understand what you mean with the statement that the information content is 'a measure how accurate a measurement is to how well a value is a priori known'. Hs expresses the Shannon information content and quantifies to what extent the knowledge about the system is improved by the measurement.

I do not see a conflict between these descriptions. The extent of knowledge or uncertainty about the system can be described by the covariance matrices. The elements of the retrieved covariance matrix (accuracy of measurement) should be smaller, indicating an improvement in understanding the system, than the a priori covariance (a priori knowledge). In either case, the accompanying equation should make it more clear for readers.

Evaluation of MOPITT version 7 joint TIR-NIR X_{CO} retrievals with TCCON

Jacob K. Hedelius[1], Tai-Long He[1], Dylan B. A. Jones[1], Bianca C. Baier[21,22], Rebecca R. Buchholz[2], Martine De Mazière[3], Nicholas M. Deutscher[4], Manvendra K. Dubey[5], Dietrich G. Feist[26,6,7], David W. T. Griffith[4], Frank Hase[8], Laura T. Iraci[9], Pascal Jeseck[10], Matthäus Kiel[11,12], Rigel Kivi[13], Cheng Liu[14], Isamu Morino[15], Justus Notholt[16], Young-Suk Oh[17], Hirofumi Ohyama[15], David F. Pollard[18], Markus Rettinger[19], Sébastien Roche[1], Coleen M. Roehl[11], Matthias Schneider[8], Kei Shiomi[20], Kimberly Strong[1], Ralf Sussmann[19], Colm Sweeney[21,22], Yao Té[10], Osamu Uchino[14], Voltaire A. Velazco[4,23], Wei Wang[24], Thorsten Warneke[15], Paul O. Wennberg[11,25], Helen M. Worden[2], and Debra Wunch[1]

[1]Department of Physics, University of Toronto, Toronto, Canada
[2]Atmospheric Chemistry Observations & Modeling, National Center for Atmospheric Research, Boulder, Colorado, USA
[3]Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Brussels, 1180, Belgium
[4]Centre for Atmospheric Chemistry, School of Earth, Atmospheric and Life Sciences, University of Wollongong, Northfields Ave. Wollongong NSW 2522 Australia
[5]Los Alamos National Laboratory, Earth and Environmental Sciences, Los Alamos, NM, USA
[6]Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
[7]Max Planck Institute for Biogeochemistry, Jena, Germany
[8]Institute of Meteorology and Climate Research (IMK-ASF), Karlsruhe Institute of Technology, Karlsruhe, Germany
[9]NASA Ames Research Center, Mountain View, CA, USA
[10]LERMA-IPSL, Sorbonne Université, CNRS, Observatoire de Paris, PSL Université, 75005, Paris, France
[11]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA
[12]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
[13]Finnish Meteorological Institute, Sodankylä, Finland
[14]University of Science and Technology of China, Hefei, 230026, China
[15]National Institute for Environmental Studies (NIES), Tsukuba, Japan
[16]University of Bremen, Bremen, Germany
[17]National Institute of Meteorological Sciences 33, Seohobuk-ro, Seogwipo-si, Jeju-do 63568, Republic of Korea
[18]National Institute of Water and Atmospheric Research, Lauder, New Zealand
[19]Karlsruhe Institute of Technology (KIT), Institute of Meteorology and Climate Research (IMK-IFU), Garmisch-Partenkirchen, Germany
[20]Japan Aerospace Exploration Agency, Tsukuba, Japan
[21]Cooperative Institutes for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA
[22]NOAA Earth System Research Laboratory, Boulder, CO, USA
[23]Oscar M. Lopez Center for Climate Change Adaptation and Disaster Risk Management Foundation, Inc., Philippines.
[24]Key Laboratory of Environmental Optics and Technology, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei, China
[25]Division of Engineering and Applied Science, California Institute of Technology, Pasadena, California, USA
[26]Lehrstuhl für Physik der Atmosphäre, Ludwig-Maximilians-Universität München, Munich, Germany

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

Abstract. Observations of carbon monoxide (CO) from the Measurements Of Pollution In The Troposphere (MOPITT) instrument onboard the Terra spacecraft were expected to have an accuracy of 10 % prior to launch in 1999. Here we evaluate MOPITT version 7 joint TIR-NIR (V7J) accuracy and precision, and suggest ways to further improve the accuracy of the observations. We take five steps involving filtering or bias corrections to reduce scatter and bias in the data relative to other MOPITT soundings, and ground based measurements. 1) We apply a preliminary filtering scheme in which measurements over snow and ice are removed. 2) We find a systematic pairwise bias among the four MOPITT along-track detectors (pixels) on the order of 3–4 ppb with a small temporal trend, which we remove on a global scale using a temporally trended bias correction. 3) Using a small region approximation (SRA) a new filtering scheme is developed and applied based on additional quality indicators such as signal-to-noise. After applying these new filters, the root mean squared error computed using the local median from the SRA over 16 years of global observations decreases from 3.84 ppb to 2.55 ppb. 4) We also use the SRA to find variability in MOPITT retrieval anomalies that relates to retrieval parameters. We apply a bias correction to one parameter from this analysis. 5) After applying the previous bias corrections and filtering, we compare the MOPITT results with the GGG2014 ground-based Total Carbon Column Observing Network (TCCON) observations to obtain an overall global bias correction. These comparisons show that MOPITT V7J is biased high by about 6–8 %, which is similar to past studies using independent validation datasets on V6J. When using TCCON spectrometric column retrievals without the standard airmass correction or scaling to aircraft (WMO scale), the ground and satellite based observations overall agree to better than 0.5 %. GEOS-Chem data assimilations are used to estimate the influence of filtering and scaling to TCCON on global CO, and tend to pull concentrations away from the prior, and closer to the truth. We conclude with suggestions for further improving the MOPITT data products.

1 Introduction

Carbon monoxide (CO) is an important atmospheric trace gas. It is a tracer of pollution and atmospheric transport, and plays an important role in the atmospheric hydroxyl (OH) budget. About 2800 Tg CO yr^{-1} are emitted globally, with about 45 % of the emissions coming from oxidation of volatile organic compounds (VOCs—predominately methane and isoprene), about 25 % from biomass burning, 25 % from fossil fuel and domestic fuel burning, and the rest from vegetation, oceans, and geological activity (Seinfeld and Pandis, 2006). It acts as an indirect greenhouse gas (GHG) as both a minor source of CO_2, and by affecting OH concentrations, which in turn affects the lifetime of methane. Its per mass 100-year global warming potential is 1.9 (Forster, P. et al., 2007). The ultimate fate for 90 % of CO is oxidation by OH to form carbon dioxide, and HO_2. CO has an average global lifetime of about 1–3 months, with a shorter lifetime in the tropics and a longer lifetime in the Southern Hemisphere extratropics (Lelieveld et al., 2016). The moderate lifetime of CO makes it a good tracer for both emissions and transport of pollution.

The Measurements Of Pollution In The Troposphere (MOPITT) is a Canadian instrument onboard the Terra earth observing satellite, launched December 1999. Drummond et al. (2010) describe the instrument in more detail, but briefly it is a gas correlation radiometer with near-infrared (NIR) and thermal-infrared (TIR) channels. The primary MOPITT mission goal is to quantify CO in the Earth's atmosphere. Space-based observations of CO can provide greater spatial coverage than a few

surface observations. However, space-based observations that rely on reflected (e.g., NIR) sunlight can be influenced by surface properties, airglow, clouds, and are more strongly affected by aerosol scattering than solar-viewing instruments. For MOPITT the thermal-infrared (TIR) sensitivity depends on the strength of the temperature contrast between the surface and atmosphere, which is variable across the globe. Due to the physical limitations of passive earth nadir-viewing remote sensing, satellite

5 instruments often have lower information content per observation than ground-based instruments (e.g., Deeter et al., 2015), especially compared to the Total Carbon Column Observing Network (TCCON), which measures atmospheric absorption of the sun's radiance. Ground-based spectrometers often have higher spectral resolution and/or coverage as well as temporal resolution at an individual location. These differences between observing systems make intercomparisons useful to check for and reduce biases.

10 While MOPITT data are the longest satellite record of total column CO (Deeter et al., 2017), there are several other satellite instruments that measure column CO and we mention a few of them here. SCIAMACHY (Scanning Imaging Absorption Spectrometer for Atmospheric Chartography) onboard Envisat (Environmental Satellite) launched March 2002 was first compared with ground based observations in 2005 (Sussmann and Buchwitz, 2005), and later compared with the larger TCCON network and found to be biased about 10 ~~%~~ ppb lower (Hochstaffl et al., 2018). TROPOMI (Tropospheric Monitoring Instru-

15 ment) onboard Sentinel-5 Precursor was launched October 2017 and was found to be biased 6 ppb higher than TCCON with the difference depending on location (Borsdorff et al., 2018). GOSAT-2 (Greenhouse gases observing satellite-2) was recently launched in October 2018 and TCCON will be used for its validation.

Most intercomparisons with MOPITT have used aircraft data (e.g, Deeter et al., 2014, 2017, 2019). The first systematic validation of MOPITT CO with ground-based column measurements was by Buchholz et al. (2017), who used the Network

20 for the Detection of Atmospheric Composition Change (NDACC) mid-IR retrievals. There have been some studies to compare observations from MOPITT with data from a few (3–6) TCCON sites (e.g, Mu et al., 2011; Té et al., 2016), but this is the first to use observations from all the sites in an intercomparison with MOPITT. Continual comparisons of MOPITT observations with other systems ensures data quality, and can be used to determine areas of improvement. This intercomparison exercise uses the MOPITT version 7 joint (V7J) product, and ground-based NIR observations from the TCCON.

25 The rest of this paper is summarized as follows: Section 2 describes the different instruments, systems, and datasets used in this study. Section 3 describes our effort to derive filters for MOPITT data, and to improve the single sounding accuracy and precision using bias corrections. Section 4 describes the MOPITT and TCCON comparisons, including sensitivity tests and a comparison of averaging kernels and information content. Section 5 describes data assimilation tests where the GEOS-Chem model is used to estimate how filtering and bias correcting MOPITT data affects global fluxes. Finally we conclude in Section

30 ~~22~~ 6 with a summary of practical considerations in this study, along with suggestions on how MOPITT retrievals might be improved in future iterations and summarize our work in Section 7.

2 Datasets

2.1 MOPITT

The MOPITT instrument onboard the Terra satellite launched December 1999 has been described elsewhere (Drummond et al., 2010). Briefly, it is a gas filter correlation radiometer with eight optical channels of which three have been used since August 2001 for CO observations, two in the TIR band (channels #5 and #7, $4.617 \pm 0.055 \, \mu$m) and one in the NIR band (channel #6, $2.334 \pm 0.011 \, \mu$m). Each channel produces an "average" (A) and "difference" (D) radiance measurement. A linear detector array in each channel allows MOPITT to make observations at four different sounding locations simultaneously. The ground field of view is approximately 22 km × 22 km for each sounding. Retrievals from among these four "footprints" or pixels were previously shown to have a bias compared to ground-based column measurements from the NDACC-IRWG (Buchholz et al., 2017). A moving mirror scans cross track for 29 "stares" in each direction for a swath that is approximately 650 km wide, and one back and forth sweep takes approximately 26 s.

Terra is in a daytime descending (nighttime ascending) sun-synchronous orbit at an altitude of about 700 km with a local equator crossing time around 10:30 (22:30 nighttime) and an inclination angle of 98.4°. Terra makes 14–15 orbits daily with an exact repeat time of 16 days. However, with its wide swath-width MOPITT is able to achieve near global coverage every 3–4 days. ~~MOPITT data are the longest satellite record of atmospheric CO (Deeter et al., 2017).~~ The redundancies built in to the MOPITT mission allowed for continued measurements after a cooler failure in May 2001 eliminated one of the two optical boards and the usefulness of channels 1–4, leaving channels 5–8 (Drummond et al., 2010). The impact of other early anomalies is minor. No abrupt changes since 2001 are expected to impact the retrievals, with the possible exception of annual hot calibrations the latest of which was in March 2019, and a separate temporary cooler malfunction in July 2009. Due to the different instrument configuration from the early record, we only include MOPITT data from 2002–2017 (inclusive) in this study.

There are different retrieval products corresponding to TIR-only (T) retrievals, NIR-only (N) retrievals, and TIR-NIR (J, joint) retrievals. We chose to make comparisons with the level 2 version 7 joint (L2, V7J) product because it should theoretically contain the most information. Deeter et al. (2014) noted that the V6 TIR-NIR product has the greatest vertical resolution, but has large retrieval errors and bias drift. The TIR-only product has the highest stability, and the NIR-only is best at CO total column retrievals. The MOPITT retrievals are performed on a logarithmic scale due to the large variability of CO in the atmosphere (~an order of magnitude). The state vector includes up to 10 vertical layers of $\log_{10}(VMR_{CO})$ (dry volume mixing ratios), surface temperature, and surface emissivity. Retrievals are performed on a grid of 100 hPa spaced layers up to 100 hPa (e.g., surface–900 hPa, 900–800 hPa, etc. Deeter, 2017). The top layer retrieved is 100–50 hPa, and above that the prior is used due to low sensitivity. Above 50 hPa represents $1.2 \pm 0.4\%$ (1σ) of the a priori CO column (1.3 % in SH, 1.0 % in NH). Fractions of CO in this layer compared to the total column are shown in the Supplemental Material (SM) Fig. S1. The a priori is from climatological output from the Community Atmosphere Model with Chemistry (CAM-chem, Lamarque et al., 2012) and is described by Deeter et al. (2014). The a priori covariance matrix is described by Deeter et al. (2010). A total column is obtained by a weighted average of the layers, and this can be converted to a column-average dry-air mole

4

fraction (denoted X_{CO}) by dividing by the model total column of dry air included in the MOPITT V7 product that takes into account surface pressure and water content. We focus on only daytime soundings, which are defined as those with a solar zenith angle less than 80° in the retrieval. In the V7J data product the 100–0 hPa layer is an average of the 100–50 hPa and 50–0 hPa layers, and we use the 100–0 hPa values for our 100–50 hPa layer but use values that are 48 % as much for 50–0 hPa based on recommendations of the MOPITT V5 users guide.

There are a number of previous studies that have compared MOPITT with different observing systems. Because the algorithm has been improved several times since the start of the mission, here we only list validation studies on versions 6 (released in 2013), and 7 (released in 2016, used in this study). Recently version 8 was released (December 2018). Versions prior to 6 are no longer available (https://www2.acom.ucar.edu/mopitt/products, last access: 2 August 2018). Deeter et al. (2014) noticed a bias between the MOPITT V6J column retrievals and aircraft observations of +4.3 ppb (assuming an average total air column density of 2.1×10^{25} molec cm^{-2}, roughly 5 % for a global 80 ppb average). They noted a correlation of $r =0.89$ between the systems, and a drift of only 0.15 ± 0.1 ppb yr^{-1} ($\sim0.18\pm0.12$ %). The V6J retrievals had an overall positive bias at the surface and 800 hPa layers, a negative bias at the 600 and 400 hPa layers, and a positive bias again at the 200 hPa layer. The bias, drift, and correlation all depended on what data products were compared. Later, the V6J profiles were compared with aircraft measurements over the Amazon basin (Deeter et al., 2016). Limited maximum aircraft altitudes precluded column retrieval comparisons, but Deeter et al. (2016) noted maximum biases at the 800 hPa of -27 %.

Three studies compared ground-based remote sensing observations with those from MOPITT (Rakitin et al., 2015; Té et al., 2016; Buchholz et al., 2017). Rakitin et al. (2015) made comparisons between MOPITT V6J L3 and various ground-based remote sensing sites in Eurasia. There is significant variability in the unadjusted comparisons for different sites in their study, which could be from the influence of averaging kernels (Rodgers and Connor, 2003), but in general MOPITT observations were larger than ground-based observations. Té et al. (2016) compared MOPITT V6J and IASI (Infrared Atmospheric Sounding Interferometer) satellite observations with ground-based observations in a urban site (Paris), a high altitude site (Jungfraujoch), and a Southern Hemisphere site (Wollongong). They noted good agreement between space and ground-based observations with slopes of 0.91–0.99, satellite observations being slightly lower. Recently, Buchholz et al. (2017) compared MOPITT V6 observations with those from 14 different ground-based NDACC sites between 78°S and 80°N, and used data from August 2001–February 2012 for comparisons with V6T, V6N, and V6J. We focus on their V6J comparison results. They found MOPITT is generally biased high relative to the NDACC, and 11 sites have a bias less than 10 % over land. The all-station mean bias is 5.1 %, and the average correlation is $\bar{r} = 0.78$. They noted the that surface type (land or water) had little effect on validation statistics. However, they did note that validation results differed among pixels, and pixel 1 has the lowest correlation while pixel 3 has the highest.

Deeter et al. (2017) is the only systematic global validation study of the MOPITT V7 algorithm. They use aircraft measurements from the HIAPER Pole-to-Pole Observations (HIPPO) campaign, and National Oceanic and Atmospheric Administration (NOAA) aircraft flask samples primarily over North America for their validation dataset. They describe the improvements included to create the V7 algorithm. They find the V7J column observations have a smaller bias and larger r (1.4 ppb and 0.93 respectively) than the V6J product (3.8 ppb and 0.89).

While L1 includes radiance bias corrections, there are no empirical bias corrections to the physics-based retrieval in the L2, V7 MOPITT products. There are retrieval anomaly diagnostics included in the L2 product, but users need to define filters to use for their particular application. For L3, V7J daytime observations where both the signal to noise ratio (SNR) of channel #5A < 1000 and channel #6A SNR < 400 are excluded (Deeter, 2017). All observations from Pixel 3 are also excluded due to excessive and unstable noise from NIR measurements from that pixel (Deeter et al., 2015). In this study suggested filters are developed along with a bias correction.

2.2 TCCON

The TCCON is a global network of independently operated solar-viewing Fourier transform spectrometers (SV-FTS) operated under a common set of standards. From measurements taken by these spectrometers, retrieved estimates of X_{CO} are made (Wunch et al., 2011a, 2015). Because profiles are not a part of the TCCON data product, we focus on validating the MOPITT total columns rather than profiles. Data are quality screened by both individual site operators as well as a centralized team. From sensitivity tests perturbing the algorithm to each known source of uncertainty (e.g., priors, surface pressure, etc.), GGG2014 X_{CO} systematic errors for TCCON are below 4 % (Wunch et al., 2015). The uncertainty on the scaling slope is 6 % (2σ).

One of the primary uses of the TCCON data has been satellite validation (e.g., Inoue et al., 2016; Kulawik et al., 2016; Wunch et al., 2017b). There are several reasons why TCCON data are considered more accurate than satellite observations, and hence a good validation source. 1) Observations are directly pointed at the sun, which increases the signal-to-noise ratio (SNR), ~~eliminates~~ is insensitive to effects of surface properties, and ~~eliminates~~ is insensitive to the effects of both airglow and aerosol scattering (e.g., Zhang et al., 2015). 2) Instruments are operated at a resolution of at least $0.02\,\mathrm{cm}^{-1}$, which provides more information for spectral fitting than most satellite measurements. 3) The network ~~has been around since~~ was established in 2004 with contributions from many different institutions. This international collaboration has lead to many discoveries on how to reduce errors in X_{gas} retrievals (e.g., Kiel et al., 2016).

Despite these advantages, ~~the data are still subject to errors~~ there are known sources of uncertainty that could be biasing the measurements. For example, to tie to the World Meteorological Organization (WMO) in situ scale there is a 7 % scaling factor in GGG2014 for X_{CO} (Wunch et al., 2015). This factor is considered large compared to the current uncertainty in spectroscopy and there is an ongoing effort to determine if this factor is appropriate. In this study we use both the official TCCON X_{CO} product, as well as a derived product without the empirical scaling factor applied. For a discussion and current comparison of unscaled TCCON data to the WMO scale, see SM Sect. S2.

We compare MOPITT with TCCON from mid-2004 through 2017. Prior to 2007 there were only four TCCON sites (Table 1). During 2007 and 2008 the TCCON grew to nine sites. Table 1 also lists the site locations and number of coincidence days after MOPITT data are filtered.

2.3 AirCore

AirCore measurements are a novel way to vertically sample the atmosphere to obtain profiles of various gases and have been described elsewhere (Karion et al., 2010; Membrive et al., 2017). Briefly, a coiled tube on the order of 100–300 m long, with

Table 1. Details for TCCON sites used in this study. Occasionally one site had more than one instrument as indicated by multiple 2-letter IDs.

Site name	ID	Lat	Long	m asl	Operational dates[a]	days[b]	References
AFRC, Edwards, CA, USA	df	34.958°N	117.882°W	699	Jul 2013–present	253	Iraci et al. (2016a)
Anmyeondo, South Korea	an	36.624°N	126.331°E	30	Feb 2015–present	24	Goo et al. (2014)
Ascension Island	ae	7.917°S	14.332°W	10	May 2012–present	151	Feist et al. (2014)
Białystok, Poland	bi	53.23°N	23.025°E	180	Mar 2009–Sep 2018	159	Deutscher et al. (2014)
Bremen, Germany	br	53.104°N	8.850°E	27	Jan 2007–present	83	Notholt et al. (2014)
Burgos, Philippines	bu	18.553°N	120.650°E	35	Mar 2017–present	31	Morino et al. (2018); Velazco et al. (2017)
Caltech, Pasadena, CA, USA	ci	34.136°N	118.127°W	237	Sep 2012–present	56	Wennberg et al. (2014b)
Darwin, Australia	db	12.456°S	130.927°E	37	Aug 2005–present	599	Griffith et al. (2014a)
East Trout Lake, SK, Canada	et	54.354°N	104.987°W	502	Oct 2016–present	28	Wunch et al. (2017a)
Eureka, Nunavat, Canada	eu	80.05°N	86.42°W	610	Mar 2010–present	8	Strong et al. (2017)
Four Corners, NM, USA	fc	36.797°N	108.480°E	1643	Mar 2013–Oct 2013	22	Dubey et al. (2014b); Lindenmaier et al. (2014)
Garmisch, Germany	gm	47.476°N	11.063°E	743	Jul 2007–present	153	Sussmann and Rettinger (2014)
Hefei, China	hf	31.90°N	117.17°E	29	Sep 2015–Dec 2016	17	Liu et al. (2018)
Indianapolis, IN, USA	if	39.861°N	86.004°W	270	Aug 2012–Dec 2012	18	Iraci et al. (2016b)
Izaña, Tenerife, Spain	iz	28.3°N	16.5°W	2370	May 2007–present	65	Blumenstock et al. (2014)
JPL, Pasadena, CA, USA	jc	34.202°N	118.175°W	390	Jul 2007–Jun 2008,	10	Wennberg et al. (2014c, 2016a)
	jf				May 2011–May 2018	19	
Karlsruhe, Germany	ka	49.100°N	8.439°E	116	Apr 2010–present	136	Hase et al. (2014)
Lamont, OK, USA	oc	36.604°N	97.486°W	320	Jul 2008–present	593	Wennberg et al. (2016b)
Lauder, New Zealand	lh, ll	45.038°S	169.684°E	370	Jun 2004–present	150	Sherlock et al. (2014a, b); Pollard et al. (2017)
Manaus, Brazil	ma	3.213°S	60.598°W	50	Oct 2014–Jun 2015	3	Dubey et al. (2014a)
Ny-Ålesund, Spitsbergen, NOR	sp	78.923°N	11.923°E	20	Mar 2006–present	79	Notholt et al. (2017)
Orléans, France	or	47.97°N	2.113°E	130	Aug 2009–present	138	Warneke et al. (2014)
Paris, France	pr	48.846°N	2.356°E	60	Sep 2014–present	40	Té et al. (2014)
Park Falls, WI, USA	pa	45.945°N	90.273°W	473	Jun 2004–present	435	Wennberg et al. (2014a)
Réunion Island	ra	20.091°S	55.485°E	87	Sep 2011–present	275	De Mazière et al. (2014)
Rikubetsu, Japan	rj	43.457°N	143.766°E	380	Nov 2013–present	12	Morino et al. (2014b)
Saga, Japan	js	33.241°N	130.288°E	7	Jul 2011–present	136	Kawakami et al. (2014)
Sodankylä, Finland	so	67.367°N	26.631°E	188	May 2009–present	241	Kivi et al. (2017); Kivi and Heikkinen (2016)
Tsukuba, Japan	tk	36.051°N	140.122°E	31	Aug 2011–present	66	Morino et al. (2014a)
Wollongong, Australia	wg	34.406°N	150.879°E	30	Jun 2008–present	342	Griffith et al. (2014a) Griffith et al. (2014b)
Zugspitze, Germany	zs	47.42°N	10.98°E	2960	Apr 2015–present	42	Sussmann and Rettinger (2018)

[a]Operational dates refers to time range where public GGG2014 retrievals are available. [b]Coincidence days only, and after filtering MOPITT data.

7

an inner diameter on the order of 2–5 mm is taken to altitude. One end of the tube is sealed, so during ascent it is evacuated and on descent the tube slowly fills with ambient air. Because diffusion is slow over the ~~long~~ length of the tube, but fast across the 2–5 mm diameter of the tube, air from different altitudes does not mix significantly. ~~So upon~~ Upon landing the vertical profile of the gas is preserved along the length of the tube, with high altitudes near the closed end and low altitudes near the open end. On the ground, the AirCore is analyzed within a few hours, which minimizes molecular diffusion. By pulling the air through and measuring concentrations with a calibrated trace-gas analyzer, ~~such as a Picarro, a~~ a vertical profile can be obtained. AirCore CO is still a developmental product with a sample measurement precision typically less than 5 ppb (Engel et al., 2017). However, stratospheric AirCore CO profile comparisons have shown differences as large as 20 ppb, which could be a result of diffusion in stratospheric AirCore samples, AirCore surface effects, or incorrect AirCore sample end member assumptions. Accuracy is dependent on the quality of calibration and standards (see SM Sect. 2).

Often AirCores are flown on balloons that can reach a ceiling of around 30 km ($\sim$10 hPa), depending on the type of balloon. Once altitude is reached, the payload is cut away from the balloon. Higher altitude data (during rapid descent) often need to be discarded, hence 22 km ($\sim$40 hPa) is the median highest altitude in this dataset. The vertical resolution depends on ~~tube design~~AirCore tubing dimensions, measurement altitude, recovery time, and temperature but is on order of the 200–1000 m. From 2012–2017 there are 36 AirCore profiles available. AirCore profiles are used among other profile measurements to tie TCCON retrievals to the WMO scale (Wunch et al., 2015). Here we use them for sensitivity tests when an approximation of the true atmospheric profile is needed.

3   Quality Control Filters and Bias Correction

Typically a retrieved state vector $\hat{\mathbf{x}}$ (e.g., an atmospheric profile) is described as a linearization about the a priori state vector $\mathbf{x}_a$ (Rodgers, 2000) i.e.,

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

In this equation $\mathbf{A}$ is the averaging kernel, a matrix in this case, with elements $A_{ij} = \frac{\partial \hat{x}_i}{\partial x_j}$, and $\mathbf{x}$ is the true state vector. The term $\boldsymbol{\epsilon}_x$ is a catch-all for any remaining systematic or random uncertainties from instrument calibration or the retrieval. This term is a function of forward model parameters not perfectly known ($\mathbf{b}$), such as pressure, temperature, pointing, spectroscopy, and modeling of instrument response (e.g., the instrument line shape). $\mathbf{c}$ contains other values in the retrieval not used in the forward model, such as convergence criteria. Changes in $\hat{\mathbf{x}}$ may thus be related to changes in $\mathbf{b}$ and $\mathbf{c}$. ~~To a first order, the biases~~ Biases in $\mathbf{b}$ and $\mathbf{c}$ ~~have~~ may be approximated as having a linear effect on $\hat{\mathbf{x}}$ (Rodgers, 1990). However, these effects may not be accounted for in models so measurement teams may reduce effects of these spurious variations by filtering data empirically.

For example, empirical corrections are employed for various gases in the final TCCON products after the physics based retrievals to improve accuracy up to about 0.1 % which would otherwise be currently limited to accuracies of about of 2–3 % due to spectroscopic uncertainties, especially in O_2 (Wunch et al., 2011b, 2015). As a second example, empirical corrections to CO_2 measurements from the Orbiting Carbon Observatory-2 satellite (OCO-2, launched 2014) did not always improve data

at all scales, but did reveal areas where the algorithm could be improved (O'Dell et al., 2018; Kiel et al., 2019). Though their studies were for CO_2, we apply many of the same methods for CO including similar truth ~~metrics~~proxies.

By comparing retrieved data with a truth ~~metric~~proxy, some data may stand out as possibly biased due to the $\epsilon_x(\mathbf{b}, \mathbf{c})$ term. These may be filtered out, deweighted, or bias corrected to improve the final product. It is challenging to define a truth ~~metric~~

5 proxy because if the true state of the atmosphere was known a priori the measurement would not be needed in the first place. Rather than using ~~metrics~~ proxies that work for each measurement, ~~statistically reasonable products can be used~~ we aggregate many measurements to empirically identify artifacts and outliers. We use TCCON and a small region approximation (SRA - also ~~called~~ known as small area approximation or variation in other studies) as truth proxies. For the SRA we assume that over a sufficiently small region (e.g., $\sim$~~1°~~$100\,\mathrm{km}\times100\,\mathrm{km}$) remote from point sources the atmosphere is approximately homogeneous

10 and outliers are due to inadequacies in the retrieval.

Filter selection and biases are interdependent, thus our quality control (QC) and bias correction process was iterative.

3.1 Pixel to pixel bias

Buchholz et al. (2017) observed biases among the four MOPITT pixels. This bias significantly affects our SRA (Section 3.2) as a biased value may be chosen as the median. We spatially grid the data in $2°\times2°$ bins and average for each pixel separately over

15 monthly time-scales to evaluate variability in the bias. Here and throughout, data are averaged as described in Appendix A. We analyze multiple months, but for here show results from April and November 2016 in Fig. 1 for the difference between pixels 2 and 4. We choose these two pixels because the instrumental noise is larger for pixel 3 (Deeter et al., 2015) and pixel 1 has a known large global bias (Buchholz et al., 2017) and would therefore expect the $2-4$ difference to be a lower bound on pixel to pixel bias. We see large pixel to pixel bias polewards of $60°$. Comparing with scenes flagged as snowy or icy by retrievals from

20 MODIS (MOderate Resolution Imaging Spectroradiometer, also onboard Terra), we see there is some correlation between the bias with the snow/ice scenes. This bias can be positive or negative. For example, we see pixel 2 is lower than pixel 4 towards the north pole and is biased positive over Antarctica land. Over sea ice around Antarctica, pixel 2 is lower than pixel 4. We also compare pixel 1 and 3 to the weighted mean and find pixel 3 is biased low over land snow/ice and pixel 1 is biased high over both land and water snow/ice. These biases likely arise from the effects snow/ice have on the thermal contrast of the surface

25 and hence affect the TIR channels. For the rest of our analysis, we filter for daytime scenes and remove soundings where the MODIS diagnostics indicate the presence of any snow or ice.

We examine temporal trends in MOPITT pixel bias compared to the weighted mean from all ~~sensors~~ pixels (Fig. 2). Data are averaged globally for each pixel and surface type separately into 15-day bins. This analysis relies on the assumption that on average each pixel samples the same area. We see that the absolute bias of pixel 1 is largest. However in contrast with Buchholz

30 et al. (2017), we observe a negative rather than positive bias between pixel 1 and the mean in the TIR-NIR retrievals, which may be because their study was of V6 data. Pixel 3 has a smaller absolute bias that is positive. In 2002, the spread of the biases is larger than in 2017. On average, the land and water biases are similar (within 0.4 ppb), however there is a larger seasonal cycle ($\sim$1.5 ppb) in the bias for the land that may be an artifact of the sampling and averaging global 15-day bins.
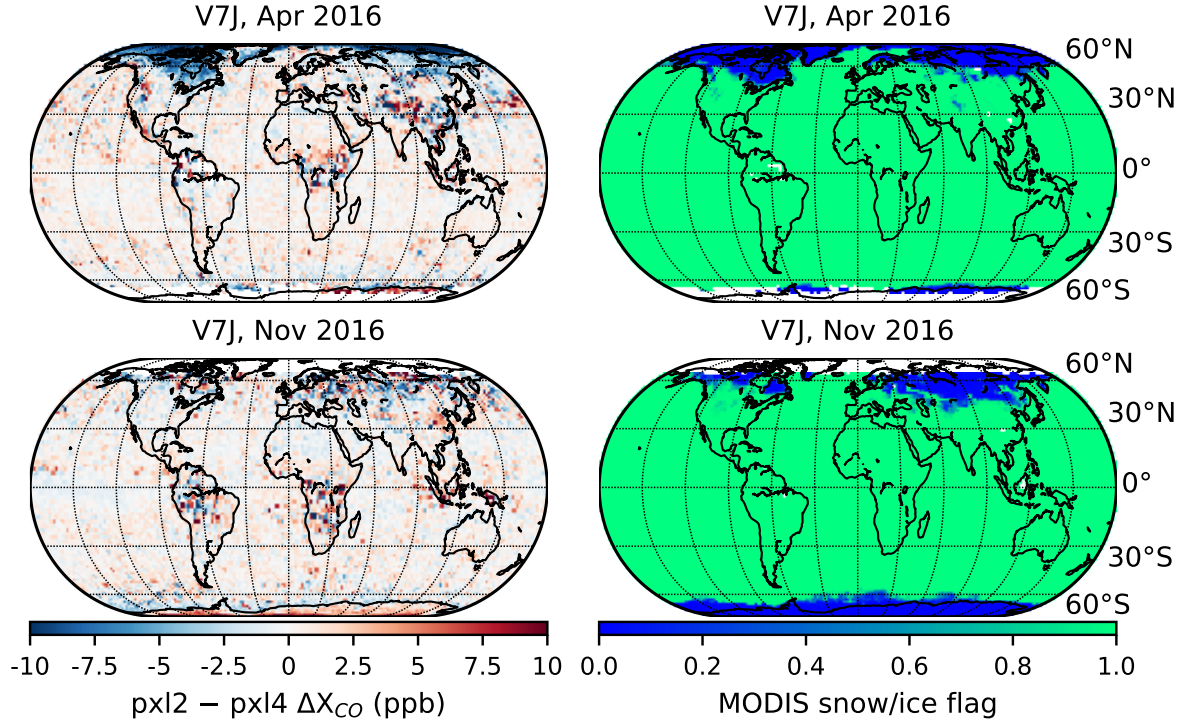
Figure 1. On the left are average differences in X_{CO} between pixels 2 and 4. On the right is the corresponding MODIS snow/ice flag where 0 indicates all snow/ice and 1 indicates the scenes were clear of snow/ice. Some correlation is observed between bins with a large pixel to pixel bias and snow/ice cover. Here and throughout we use an Eckert IV equal area projection.

One consideration for bias corrections is whether accounting for differences in averaging kernels can account for the bias. Buchholz et al. (2017) noticed a large absolute bias for pixel 1 as compared with NDACC observations even after accounting for averaging kernels. To examine the effects of averaging kernels, we find MOPITT soundings within an ellipse ($\pm1°$ latitude, $\pm1.5°$ longitude) around the center location of AirCore flights on the same day. There are 22 20 flights with coincident ob-
5 servations and 1966 1933 total corresponding MOPITT soundings. We apply averaging kernels to create simulated MOPITT column retrievals from AirCore profile measurements:

$$\hat{c} = c_a + \mathbf{a}_M^T \left(\log_{10}\mathbf{x} - \log_{10}\mathbf{x}_a \right),\qquad(2)$$

where $\hat{c}$ is the simulated X_{CO}, and c_a is the a priori column X_{CO}. $\mathbf{x}$ is the dry VMR profile (from AirCore) and should not be confused with the state vector which is $\log_{10}(\mathrm{VMR})$ for MOPITT. For this study we have defined the MOPITT column
10 averaging kernel for a pressure level i to be (Appendix B)

$$a_{M,i} = \frac{\partial \hat{c}}{\partial \log_{10} x_i} = \ln10 \sum_{j}^{n} h_j \hat{x}_j A_{ij}.\qquad(3)$$

10
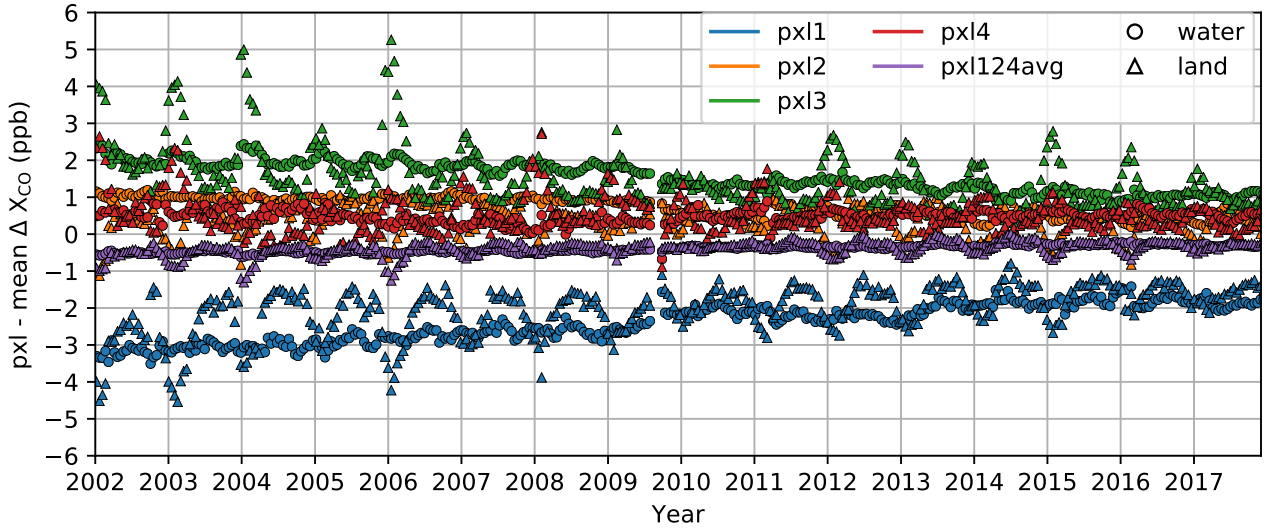
Figure 2. Pixel (pxl) biases compared to the weighted mean with time. Data are averaged into 15-day bins separated for land and water soundings. The mean of pixels 1, 2, and 4 is also shown because pixel 3 data are not included in the L3 product. The small gap in 2009 is from a temporary cooler malfunction on July 28.

Table 2. Mean values of MOPITT X_{CO} retrievals colocated with AirCore measurements and separated by pixel, compared to the simulated X_{CO} from applying MOPITT averaging kernels to AirCore profiles for 22 flights.

	Retv. mean (ppb)	Sim. mean (ppb)
Pixel 1	~~86.2~~ 86.0	~~84.0~~ 84.1
Pixel 2	~~88.0~~ 87.7	~~84.4~~ 84.3
Pixel 3	~~88.5~~ 88.3	84.5
Pixel 4	~~88.9~~ 88.6	~~84.5~~ 84.4

The pressure weighting function **h** has been described by Connor et al. (2008) and Wunch et al. (2010). We find that the maximum bias for the retrieved columns is between pixels 1 and 4 and is about ~~5~~ 8 times larger for the retrieved (~~2.7~~2.6 ppb) than for the simulated columns (~~0.5~~0.3 ppb, Table 2). For these soundings MOPITT is also biased high compared to the AirCore simulated columns by ~~3.5~~3.3 ppb, which is greater than the bias of 0.5–1.4 ppb compared to other aircraft profiles (Deeter et al., 2017).

5   We make a preliminary pixel bias correction by adjusting soundings over land and water for each pixel separately based on a linear fit to the overall timeseries shown in Fig. 2. This fit is later improved after filtering (Sect 3.4). After this adjustment we

11

noticed some residual bias among the histograms, so we also apply a year to year pixel bias correction of up to 0.4 ppb that is the same for water and land.

3.2 Small Region Approximation

We perform a SRA on the dataset with the preliminary filter for daytime and snow/ice free scenes and preliminary pixel bias correction. In a SRA, data within a specified area and time frame are assumed to be homogeneous, and variation within that area is assumed to be non-physical. There is always some real variation in the atmosphere, however, statistically for a large sample size these variations are expected to average out. If the area is too small then there will be too few points for an unbiased median. If the area is too large then true atmospheric variability will be significant. A disadvantage of using this method as a "truth" proxy is that it is insensitive to bias on larger scales related to e.g., latitude, and surface albedo (e.g., O'Dell et al., 2018, for OCO-2 and CO_2).

We use small ~~areas~~ regions that are approximately $89\,\text{km} \times 133\,\text{km}$ ($0.8° \times 1.2°$, latitude × longitude at the equator). Region size is a trade off between having sufficient points per region, while keeping regions small enough that real variations in X_{CO} are small. The effects of different region sizes are described in the SM Sect. S3. ~~We~~ To calculate anomalies, we subtract the median from all the points within that region. If the median point does not have at least 1 degree of freedom for signal (DOF) then the entire region is discarded. We also require at least 10 points in each region, which retains about 50 % of the SRA bins.

3.3 Quality Control Filters

Using the SRA "truth" proxy, we can look for correlations of differences to the local median (i.e., anomalies) with various parameters that are or may be related to the retrieval. Table 3 lists parameters we consider for filtering and bias corrections. We make plots similar to those by Wunch et al. (2011b) and O'Dell et al. (2018) (though their studies were of X_{CO_2}) of anomalies versus one of the various parameters to aid in determining filter cutoffs (e.g., Fig. 3). Such plots may reveal empirical relationships with features. Similar plots with additional parameters, including some we decided were inappropriate to use as filters, are available in the SM (Sect. S4). Several features can be examined in these plots to decide on where to set the filter limits including: the underlying histograms, systematic biases from 0 in the mean including spikes, the spread among pixels - which indicates pixel-to-pixel bias, and the root mean square (RMS) from the SRA - which includes systematic and random deviations from the truth ~~metric~~proxy. We define filters based usually on one of the following criteria 1) absolute mean bias is greater than 2 ppb, 2) RMS is greater than 6 ppb, or 3) spread of pixel-to-pixel bias is greater than 5 ppb. These criteria are not strict and we change thresholds if too few data are in a bin (due to possible sampling bias), if too many data would be removed, or if the overall trend in the mean seems like it could be corrected by a bias correction.

Several features are apparent in the SRA diagrams (Fig. 3 & SM S4–S9) that indicate data may be less reliable. For example, there is a step change in the bias for soundings over land going from day to night. The RMS is much smaller over snow/ice free scenes (flag of 1). We also note large anomalies for low channel 5A SNR which, in agreement with the L3 product filters, suggests it is a good parameter to filter on. However, the bias is small for low channel 6A SNR soundings so unlike the L3 product we do not use it as a filter criterion. We also find the sum of the retrieval anomaly diagnostics is a better indicator for
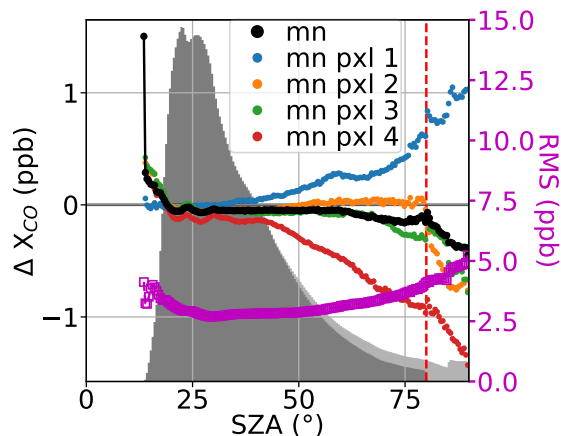
Figure 3. Example diagram showing the Small Region Approximation (SRA) bias as a function of solar zenith angle for water. The black points show the overall mean bias (minimum 2000 points), the magenta points show the RMS, and the other points show the mean bias for the individual pixels (minimum 300 points). The lighter histogram is of all the data. The darker histogram is data remaining after the SZA and snow/ice filters. Figures like this are used to make filters for and check for bias in MOPITT L2 data. The red line is the filter cutoff at $80°$. The equivalent diagram for land, along with diagrams for other features are in the SM Sect. S4.

suspicious data over land than over water. These particular tests also do not support excluding all pixel 3 soundings, though on average it does have lower and more variable DOF (Deeter et al., 2015). Maps of where data are filtered are available in the SM Fig. S10–S11. Using these filters reduces the number of daytime soundings to 3.50×10^8 (of 5.40×10^8), and reduces the RMS from 3.84 ppb to 2.55 ppb. By comparison, when we apply the L3 filters it reduces data to 3.27×10^8 daytime soundings and an RMS of 3.02 ppb.

3.4 Bias Correction

5   We observe trends in the mean bias with various parameters (e.g., Fig. 3 & SM S4–S9). To reduce the likelihood of overfitting, O'Dell et al. (2018) used linear fits as bias corrections only if they removed at least 5 % of the variance for X_{CO_2} from OCO-2. For X_{CO} from MOPITT, the ratio between the scatter (indicated by RMS) and bias is larger than for X_{CO_2} from OCO-2, however over our period of analysis there are about 400 ~~×~~ times more data over water, and about 100 ~~×~~ times more over land. Fitting concerns here are primarily over how representative the SRA is as a truth proxy, and how much the biases would already
10 be accounted for by adjusting individual soundings using averaging kernels.

    Even with a criterion of only 3 % reduction in the overall RMS, the only parameter to meet this is the maximum difference between adjacent levels over land ~~,~~ (see SM Fig. S5(j)). This feature is larger for strong gradients between levels which can appear when there are strong surface fluxes, or when the retrieval is unstable and oscillates. This instability may be caused by bias related to e.g., spectroscopic errors. Following O'Dell et al. (2018) we make piecewise linear fits to the overall mean over
15 two regimes, split at 100 ppb for a bias correction. The Multivariate Adaptive Regression Splines (MARS) algorithm could

Table 3. Parameters in or related to MOPITT retrievals that we considered for filtering and bias correction. Data are excluded if any of the criteria below are met. Mixed surface-type soundings are also excluded.

Field name	Source[a]	Limits land	Limits water	Description
χ^2	I			Goodness of fit indicator
Digital elevation model	I			Surface height, m a.s.l.
Diff. surf. emissivity (E)[b]	D	$<-0.05, >0.05$	<-0.75	$\hat{E} - E_a$, E is in state vector
DOF	I	<1.05	<1.00	tr($\mathbf{A}$)
Diff. surf. temperature (T)[b]	D		<-4	$\hat{T} - T_a$, T is in state vector
Error surface E	I	>0.055		Error on $\hat{E}$
Total error X_{CO}	R			Unused as a filter, used to weight data instead
Information content	D			$H = -\frac{1}{2}\log_2\left|\mathbf{I}_n - \mathbf{A}\right|$, unused, too similar to DOF
MODIS IR T threshold	I			MODIS cloud diagnostic 8
MODIS frac. cloudy	I			MODIS cloud diagnostic 2
MODIS snow & ice	I	<0.999	<0.999	MODIS cloud diagnostic 5 (see also Sec. 3.1)
Meas error X_{CO}	R			Unused, see Total error X_{CO}
Mean averaging kernel	D	<0.75	<0.50	$\frac{\sum \mathbf{u}^T \mathbf{A} \mathbf{u}}{n_{\mathrm{lvls}}}$, sum of all elements in $\mathbf{A}$ divided by number of profile levels ($\mathbf{u}$ is vector of ones)
AOD 500 nm	E			Average of colocated MODIS pixels (within $\pm 0.1^\circ$)
Max. diff. between adj lvls	D	>300	>300	Indicator for possibly oscillating profiles
Number of iterations	I			Until convergence, 1–20
Solar airmass	D			$\frac{1}{\cos(\mathrm{SZA})} + \frac{1}{\cos(\mathrm{satelliteZA})}$
Smth. error X_{CO}	R			Unused, see Total error X_{CO}
SNR 5A (TIR)	R	<500	<900	L1 radiance divided by error. 1000 is a threshold for L3 TIR & TIR-NIR
SNR 6A (NIR)	R			See SNR 5A. 400 is a threshold for L3 NIR & TIR-NIR
SNR 7A (TIR)	R			See SNR 5A
Sum retr anom diagnostic	D	>0.5		Sum of five flags, user's guide suggests caution/exclusion if ≥ 1
Solar zenith angle (SZA)	I	$>80^\circ$	$>80^\circ$	$>80^\circ$ is considered nighttime in the retrieval algorithm
tr total retr cov matrix	D	>0.0170	>0.0168	tr$\left(\hat{S}\right)$, unitless due to log scale, related to overall uncert. of retrieval combination of meas & smooth matrices
tr meas err retr cov matrix	D	>0.0055		tr$\left(\hat{S}_{merr}\right)$ from uncert. in prior & weighting
tr smooth err retr cov matrix	D		<0.0114	tr$\left(\hat{S}_{smth}\right)$ from uncert. in measured radiances

[a] Source I=included in L2 files, R=ratio within L2 files, D=other derivation from L2 files, E=external. [b] Difference from the prior.

AOD = aerosol optical depth. tr = matrix trace.

also be used to make a piecewise linear fit over a multidimensional dataset. However, it is more likely to overfit the data. When we applied it to the top three most variable fields the RMS for land soundings was not significantly reduced compared with our piecewise fit so we did not use those results.

In addition to the single "feature" bias correction above, we apply a pixel-to-pixel bias correction after the filtering described in Sect. 3.3. We perform a second SRA on the filtered data without a pixel bias correction. SRA data are binned separately for each pixel and land or water surface type and averaged over 10 days. On July 28, 2009 one of the coolers on MOPITT malfunctioned, which caused a 2 month instrument shut down. We separate the period before and after this event, and make 16 different linear fits of bias relative to the all-pixel mean with time (2 ~~×~~ for land and water, 4 ~~× pixel~~for pixels, 2 ~~×~~ for time) following the method of York et al. (2004). These linear fits are used to define the pixel-to-pixel bias.

4 Comparisons with TCCON

4.1 Coincidence criteria

Various coincidence criteria have been used to match MOPITT soundings with other data sets, such as aircraft measurements, other satellites, or ground-based sensors. For example, Deeter et al. (2014, 2017) used a co-location radius of 50 km for aircraft profiles primarily over North America, and a co-location radius of 200 km for aircraft profiles primarily over remote ocean. Over the Amazon, Deeter et al. (2016) also used a co-location radius of 200 km, and a co-location time of 24 hrs. Té et al. (2016) used criteria of $\pm 0.15°$ latitude and $\pm 0.23°$ longitude, corresponding to 33 km×33 km over Paris. Buchholz et al. (2017) used a $1°$ radius and ground-based measurements within the same day. Criteria could also include fields such as the temperature of the free troposphere (e.g., around 700 hPa, Wunch et al., 2011b; Nguyen et al., 2014). Sha et al. (2018) used a sampling cone based on the solar azumith angle at the time of measurement for comparing TCCON with TROPOMI. This is likely unimportant for MOPITT given the larger footprint size (22 km×22 km versus 7 km×7 km). For example, at $60°$ SZA for a MOPITT pixel centered on a TCCON site at sea level, the TCCON ray would leave the MOPITT pixel at around 11 km or above 250 hPa. For a comparison of SCIAMACHY (Scanning Imaging Absorption Spectrometer for Atmospheric Chartography) with NDACC/TCCON, Hochstaffl et al. (2018) found it was necessary to deweight observations that were further away in time and space from points of comparison. This is likely much less of an issue for this study due to differences in retrieval errors and coincidence scales. For MOPITT the median retrieval error is about 3.5 ppb versus 24.8 ppb for SCIAMACHY. For SCIAMACHY temporal averaging was on order of a month compared to this study where we only use TCCON observations within ± 30 min. We apply spatial averaging to the MOPITT data typically over areas of $2° \times 4°$ (with exceptions noted below). Spatial weighting is not as much of a concern here as for Hochstaffl et al. (2018) with SCIAMACHY because they used coincidence criteria of 500–2000 km radii, which are significantly larger in terms of area (about 8–100 ~~×~~times). However, despite using smaller areas heterogeneities in CO sources that MOPITT averages over may occasionally introduce bias for real reasons (e.g., Lindenmaier et al., 2014).

We make exceptions to the $\pm 1°$ latitude $\pm 2°$ longitude spatial coincidence criteria for several sites. For sites poleward of $60°$ (eu, sp, and so) we expand the area to $4° \times 8°$ because the atmosphere is expected to be well mixed and retrievals are more

15

sparse. For sites in the Los Angeles basin (ci, jc, and jf) we limit the area to 33.4–34.3°N, 116.7–118.8°W because we expect X_{CO} within the basin to be much larger than the surrounding area due to urban emissions. We set the minimum latitude to 34.5°N for the AFRC site to avoid the polluted Los Angeles basin. We average soundings over land and water separately.

Because of the long (13+ year) comparison between MOPITT and TCCON, random representation error is much less important than systematic error. Té et al. (2016) and Buchholz et al. (2017) noted that systematic biases can arise from comparing total column observations (in molec cm^{-2}) from MOPITT and NDACC when the surface altitudes differ significantly. This effect will be diminished in column averages (X_{CO}) in locations away from strong local surface fluxes, however different surface altitudes can lead to biases because CO profiles are not completely uniform. Between two TCCON sites only ∼10 km apart in an urban region Hedelius et al. (2017) noted an X_{CO_2} difference of nearly 1 ppm. They attributed part of this to the different site altitudes. We estimate the ratio between observations at the surface pressure of TCCON versus the surface pressure of MOPITT soundings. The total column average dry mole fraction is

$$X_{CO}, ppb = \sum h_j x_j. \tag{4}$$

The vector $\mathbf{x}$ here can be either the retrieved profile or the a priori profile. We use the MOPITT profiles because they are likely more representative of the true atmosphere than TCCON a priori profiles and apply Eq. 4 to find the retrieved and prior MOPITT X_{CO} at the MOPITT sounding surface pressure. We then recalculate $\mathbf{h}$ based on the daily average TCCON site surface pressure. When TCCON altitude is lower, the MOPITT surface level is uniformly extended. For higher altitude sites, the lowest altitude MOPITT levels are either unused ($h_j = 0$) or deweighted. We then calculate X_{CO} based on TCCON surface pressure. Figure 4 shows the ratios between the MOPITT retrieved X_{CO} using the TCCON surface pressure compared to the MOPITT sounding surface pressure for $10° \times 10°$ areas. Larger areas are used to get a larger variety in surface pressures. We see that for the high altitude Zugspitze site this scaling is particularly large (around 15 %). Over these areas the overall scaling for all sites is 0.996 ±0.023 (1σ). A scaling factor less than unity is usually due to larger CO mixing ratios near the surface than the rest of the column and a lower TCCON site pressure (Hedelius et al., 2017). In this intercomparison, we implicitly account for differences in surface pressure using the ~~h vector.~~ $\mathbf{h}$ vector. This can make a difference for individual sites by as much as $-10.5 \pm 4.1 (1σ)$ ppb (for Zugspitze). However, we have found in practice that accounting for differences in surface pressure makes little difference here on the overall comparison (compare SM Fig S12c and S12f). In aggregate the difference is only $-0.2 \pm 1.5 (1σ)$ ppb.

4.2 Overall ~~comparison using different methods~~ global scaling

MOPITT and TCCON use different priors and have different averaging kernels (AKs, Sect. 4.4) and these differences in sensitivity need to be taken into account when comparing retrievals from the different instruments. Here we account for differences in AKs and a priori profiles following the methods of Wunch et al. (2011b), which is formally described as method II in SM Sect. 6.1. Retrievals are also on different vertical grids, and regridding is described in Appendix C. Figure 5 shows the comparison for all sites. We find that MOPITT observations are higher than TCCON by about 6.4 %. This is similar to the 5.1 % positive bias between MOPITT V6J and NDACC total column observations (Buchholz et al., 2017).
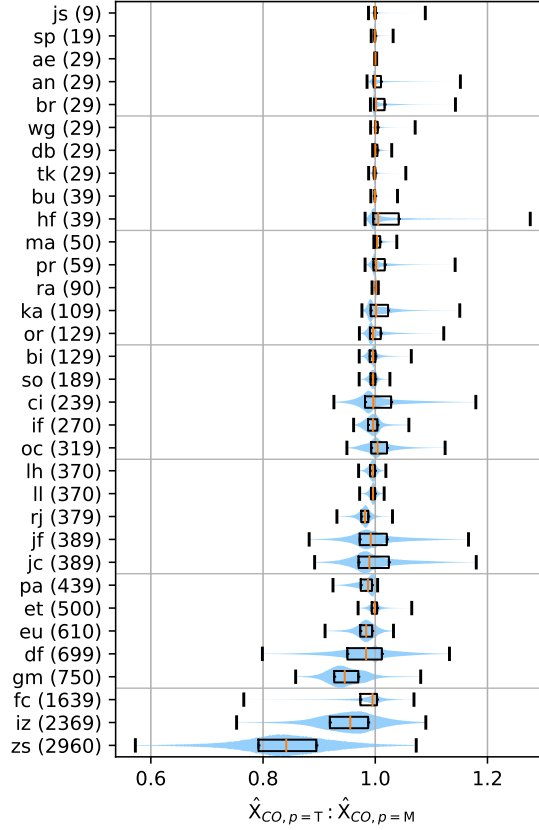
Figure 4. Scaling factors for MOPITT retrieved profiles if the surface were at the surface of the TCCON site (listed in m asl in parenthesis). Ordered by increasing site altitude. Soundings within $\pm 5°$ latitude and $\pm 5°$ longitude of TCCON site are used. The center 99 % of data are shown. Blue filled sections indicate data density, similar to a violin plot but using histograms rather than a kernel density estimation due to sufficient data. Black boxes indicate the central 50 % of data, and medians are orange.

We perform a variety of sensitivity tests on the overall global comparison. There are different approaches to account for different priors and AKs such as the choice of comparison ensemble . Here, we choose method II detailed in Appendix ?? as our standard choice, and it is equivalent to how Wunch et al. (2011b) accounted for differences in a priori profiles and AKs when comparing X_{CO_2} between GOSAT (Greenhouse Gases Observing Satellite)and TCCON. Figure ?? shows the comparison

5   using this method. (SM Sect. 6). Supplemental Fig. S12 is a series of bar-charts of how this method compares with the others for each site. Figure ?? shows the comparison for a variety of tests when AKs are applied differently or not at all, for different spatial coincidence criteria, when filtering or bias corrections are not applied, and with a "TCCON" product without empirical corrections for airmass and to the WMO scale (Wunch et al., 2015). Generally all comparisons show MOPITT is about 6–9 % higher than TCCON, with some exceptions. This is similar to the 5.1 % positive bias between MOPITT V6J and NDACC total

10   column observations (Buchholz et al., 2017). For method III, where AKs are applied in a manner opposite to method II, the

17

bias is as high as 15 % but ~~typically~~ is closer to ∼10 % or less. ~~In Fig. ??g where the TCCON~~ Supplemental Fig. S13 is a series of bar-charts of how this the different methods compare for each site. We also examine how the scaling changes for different colocation criteria in SM Fig. S12d–e by halving and doubling the coincidence areas. We find MOPITT is biased higher than TCCON in these tests by 5–7.4 %. Doubling the area decreases R^2 for the global comparison.

5   Next we test sensitivity to pressure scaling. Our vertical regridding (Appendix C) accounts for differences in surface pressure, so we use a basic comparison without AK corrections (Method 0) for this test. Between these two, the overall offset is not significantly different (9–10 %).

    Finally we test if filtering and bias corrections affect the comparison (SM Fig. S12g–i). The pixel and feature bias correction make little change in the overall global comparison (∼ 5.8–6.8 %). Without filtering, the scatter in the comparison increases

10 leading to a smaller R^2. Due to a large intercept the percent difference spans about 3–8 %. Supplemental Fig. S12g shows the comparison for a "TCCON" product without empirical corrections for airmass and to the WMO scale (Wunch et al., 2015). In this comparison the TCCON data do not have the standard scaling to aircraft~~, the bias between the datasets is less than 0.5 %~~. Due to uncertainties in the TCCON WMO scaling (SM Sect. S2), some comparisons are made without it. ~~Deeter et al. (2017) found a positive bias less than about 1~~ Here the bias between the datasets is significantly different, and is less than 0.5 %~~when~~. When

15 MOPITT V7J data were compared directly with NOAA flask measurements from aircraft, Deeter et al. (2017) found a positive bias less than about 1 %.
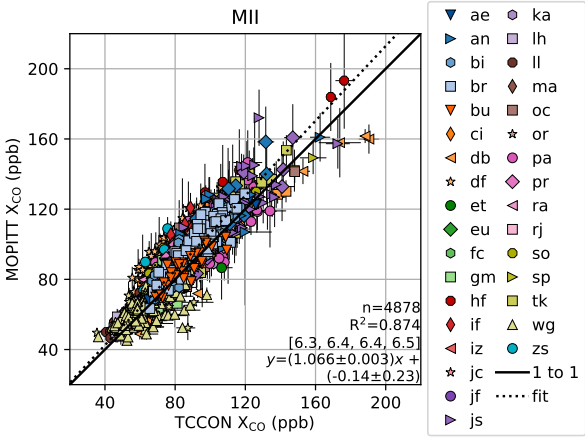
Figure 5. One-to-one plot comparing MOPITT and TCCON, following method II ~~(similar to Wunch et al., 2011b, Appendix ??)~~(similar to Wunch et al., 2 MOPITT data were adjusted to the TCCON a priori ($\hat{c}'_M$) and MOPITT averaging kernels were applied to TCCON data ($\hat{c}_{M \leftarrow T}$). Error bars represent standard deviations of the weighted averages. Triangles represent soundings over water, other shapes are over land. 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).

~~See Table ?? for a list of different methods with TCCON data on the x-axis, and MOPITT on the y-axis. See Fig. ?? for a description of the text. 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. ??). (c) Using method 0 (i.e., not accounting for averaging kernels). (d) Same as Fig. ?? but with half-sized spatial bins. (e) Same as Fig. ?? 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.~~

Figure 6 shows boxplots of the MOPITT to TCCON differences (using method II) for each site for land only and water only soundings. We do not note an overall bias between land and water. For all sites the TCCON−MOPITT bias is positive and usually on order of about 3–10 ppb with a few exceptions. For example, MOPITT observations compared to the AFRC (df) TCCON are particularly high (∼14 ppb). This could be related to the high albedo or high surface temperatures of this desert site.
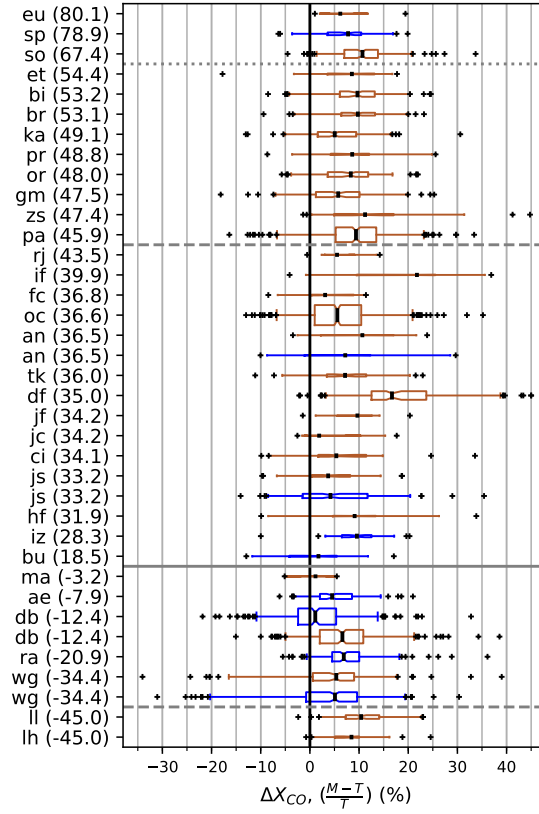
Figure 6. Boxplots of the MOPITT-TCCON percent difference at the TCCON sites (using method II), ordered by latitude (degrees north in parenthesis). Blue boxes are MOPITT soundings over water, brown boxes are over land. Whiskers represent the inner 95 % of data. Notches are 95 % confidence intervals of the median. Box heights represent the relative number of observations. The solid horizontal line is the equator, dashed lines are ±45°, and the dotted line is 60°N.

4.3 Systematic biases

A seasonal variation in bias may be indicative of differences in sensitivities between the instruments to some feature, such as airmass or water content, that varies seasonally. Figure 7 shows the timeseries of the difference averaged in 1 month 5° latitude bands. ~~We do not find a clear long term trend~~ Though there is significant scatter among individual comparisons, we find a long-term trend of $-0.06 \pm 0.06\,\%\,\mathrm{yr}^{-1}$ in the MOPITT$-$TCCON difference using the Theil-Sen estimator. Deeter et al. (2017) reported a bias drift of $-0.04 \pm 0.10\,\%\,\mathrm{yr}^{-1}$ for V7J though bias drifts for individual layers were larger. Including a correction trend to the L1 radiances significantly reduced the bias drift for the layers (Deeter et al., 2019). Seasonalities of the difference for each site are in SM Fig. ~~S13~~S14. There does not appear to be a persistent seasonal trend for all sites, though there is some seasonal variability for individual sites. For Lamont and AFRC the bias is larger in July–October, while for Białystok the bias is larger April–June. At Ascension the bias is largest January–February, while for Reunion it is largest September–November. We do not make a seasonal bias correction.

There appears to be some latitudinal dependent bias, with a larger bias in the Northern Hemisphere. Part of this could be related to stratospheric CO (SM Sect. S1). Deeter et al. (2017) also showed some latitudinal variation in MOPITT retrievals compared with aircraft. They suggested that part of the variability could arise from interfering species such as N_2O, which has spectral lines that overlap with the TIR channels. Before V7 a constant value of N_2O was assumed, which was determined to cause biases on order of a few ppb (Deeter et al., 2017). In V7 a global average is used based on a linear fit to monthly in situ observations. Supplemental Fig. S14 shows the bias as a function of column N_2O measured by the TCCON. There is a slope of $-0.40\frac{X_{\mathrm{CO\,ppb}}}{X_{\mathrm{N_2O\,ppb}}}$ though the overall correlation is small (R^2=0.08). There also appears to be a small dependence on column H_2O, which was likely reduced in V8 (Deeter et al., 2019). We do not make bias corrections for any of these systematic features.

4.4 Averaging kernels, covariance matrices, and information content

According to Rodgers and Connor (2003, Sect. 2 therein), an intercomparison of two observing systems should also include a comparison of (1) averaging kernels, (2) retrieval noise covariance, (3) degrees of freedom, and (4) the Shannon information content. In conjunction with (1), we think it is also helpful to compare a priori profiles, which is done in Appendix D. Because the MOPITT retrievals are of logarithmic profiles and the TCCON uses a linear scaling retrieval some aspects of the comparison are inherently different.

Example AKs for MOPITT and TCCON are shown in Fig. 8. Because the MOPITT retrieval is on a log-scale, we make an assumption that the a priori VMRs represent the true profile to obtain unitless AKs (Appendix B). We find the TCCON AKs are more sensitive than MOPITT. Shaded regions in Fig. 8a show a wide variability in MOPITT column AKs. In addition, the typical state significantly affects the MOPITT AKs (e.g., compare Pasadena and Lauder). TCCON CO column AKs are most sensitive to the stratosphere, and are assumed to be consistent at all sites. We make a sensitivity test where the AKs were explicitly calculated in GGG2014 for days with a wide range of X_{CO} at the East Trout Lake site. In general the difference from the standard AKs is small, on the order of 5 % at most.
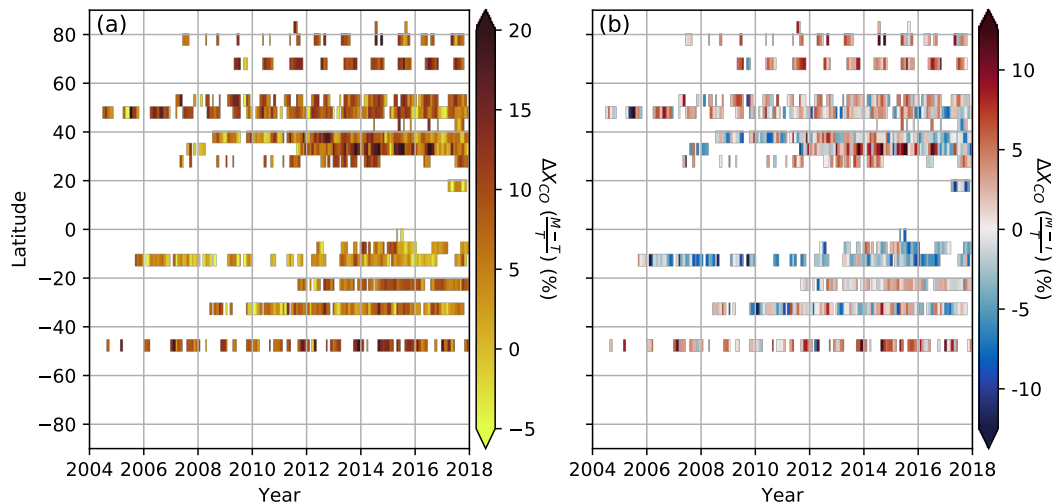
Figure 7. Rotated Hovmöller diagram of the mean percent differences between binned MOPITT and TCCON data using method II (~~Appendix ??~~SM Sect 6.1). Latitude bins are 5° zonal bands, and temporal bins are monthly. Left plot is using standard TCCON data. Right plot is without the TCCON scaling to aircraft (WMO scale - see SM Sect. S2).

Priors and MOPITT retrieved profiles along with their differences for select sites are shown in Appendix D. We compare MOPITT and TCCON priors. In general, MOPITT priors are influenced more by localized emissions as they are based on 1°

20   simulated monthly climatologies from the CAM-chem model (Deeter et al., 2014). This can be seen especially at Pasadena, and to a lesser extent at Lamont and Tsukuba. Ascension Island shows a special case where enhanced CO in the lower free troposphere is seen coming from biomass burning and rain forest VOC emissions in Africa. At sites far removed from local emissions (e.g., Ny-Ålesund and Lauder) the MOPITT and TCCON priors are in better agreement with each other (see e.g., Pollard et al., 2017).

25   We take differences in priors and averaging kernels into account following method II described in the SM Sect. 6.1. Corrections are applied to each MOPITT retrieval and to daily averages of TCCON retrievals within coincidence criteria. We find in practice that corrections change the comparison by about 3 %. TCCON data are adjusted by $0.7 \pm 1.8 \,(1\sigma)$ ppb, and MOPITT data are adjusted by $-1.0 \pm 3.1 \,(1\sigma)$ ppb.

Rather than compare the retrieval noise covariance, we compare reported errors and measures of precision and accuracy.

30   Histograms of total reported retrieval error for MOPITT are shown in SM Figures S4o and S5o. With our prescribed filtering, global mean uncertainty values are $2.60 \pm 1.27 \,(1\sigma)$ ppb for smoothing, $2.68 \pm 1.40 \,(1\sigma)$ ppb for measurement, and $3.86 \pm 1.63 \,(1\sigma)$ ppb for the total error. The average of the errors reported in the TCCON files is $0.62 \pm 0.50 \,(1\sigma)$ ppb. However, these errors are more a measure of repeatability rather than the total error or the accuracy. The 2σ uncertainty for TCCON (GGG2009) was reported as 4 ppb (Wunch et al., 2010), and the uncertainty budget from a range of sensitivity tests is less than 4 % (Wunch et al., 2015).
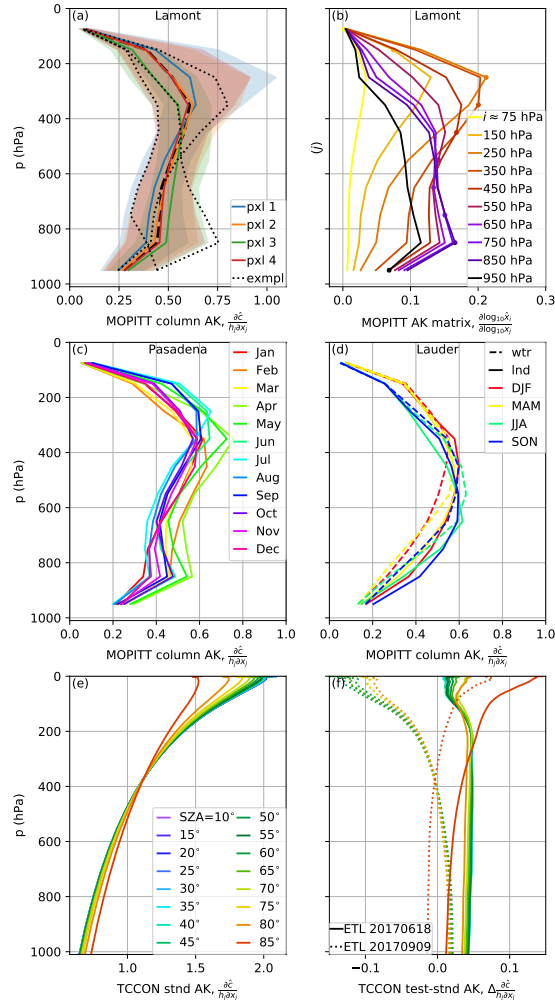
Figure 8. Examples of AKs from TCCON and MOPITT—subplots are not always related. MOPITT daytime AKs (a–d) are shown as center-point values along the y-axis for clarity, though MOPITT retrievals are layer averages. Unitless MOPITT column AKs are generated using the methods of Appendix B. (a) MOPITT column AKs around Lamont for 2012–2013 separated by pixel. Filled areas are the central 80 % and solid center lines are the medians per level. Black lines are select examples from single soundings that show wide variability from sounding to sounding. The thicker example corresponds to the full AK in subplot b. (b) Example MOPITT full AK from 16 Oct 2012. Dots highlight the ith level shown in the legend. (c) Median MOPITT column AK per level by month for Pasadena. (d) Median MOPITT column AK per level by season for land and water soundings for Lauder. (e) Standard TCCON GGG2014 AKs, which are assumed to be a function of only SZA and pressure. (f) Differences from AKs explicitly calculated at the ETL site on specific days compared to the standard AKs. For 18 Jun 2017, the mean of the ETL X_{CO} is $74.0 \pm 0.8\,(1\sigma)$ ppb. For 9 Sep 2017, the mean of the ETL X_{CO} is $169 \pm 32\,(1\sigma)$ ppb with a range of 95–225 ppb.

Histograms of the MOPITT DOF for signal for water and land are shown in Figures S4d and S5k. The DOF for signal (d_s) can be determined from

$$d_s = \mathcal{E}\left\{ (\hat{\mathbf{x}} - \mathbf{x}_a)^T \mathrm{S}_a^{-1} (\hat{\mathbf{x}} - \mathbf{x}_a) \right\}, \tag{5}$$

where $\mathcal{E}$ is the expected value operator (Rodgers, 2000, Eq 2.46 therein). However, d_s is usually determined from the trace of the averaging kernel matrix (Rodgers, 2000, Eq 2.80 therein), which is equivalent to Eq. 5 for profile retrievals. Because GGG2014 is a scaling retrieval, we treat TCCON measurements as having $d_s = 1$. With a profile retrieval we would expect $d_s > 1$, as was the case for CO_2 (Connor et al., 2016). DOF gives an indication of how many independent parameters can be improved compared with the prior. MOPITT DOFs are between 1 and 2, which indicates total column measurements may be reasonable, but individual layer measurements may not always be accurate.

Finally, the information content H_s is a measure of how accurate a measurement is to how well a value is a priori known. Rodgers (2000) expresses it on a natural log scale (Eq. 2.73 and 2.80 therein)

$$H_s = \frac{1}{2}\ln\left|\hat{\mathbf{S}}^{-1}\mathbf{S}_a\right| = -\frac{1}{2}\ln\left|\mathbf{I}_n - \mathbf{A}\right|, \tag{6}$$

where $\mathbf{I}_n$ is the identity matrix. Here we express H_s on a $\log_2$ scale instead. Histograms of H_s for MOPITT profile retrievals are shown in Figs. S6(a) and S8(a) in the SM and values are on the order of 2.5–5.5 bits over water and 2.5–7 bits over land. If model values of X_{CO} are accurate to about 32 ppb, and if the TCCON accuracy is about 4 ppb then the TCCON X_{CO} information content is about $\log_2\frac{32}{4} \approx 3$ bits.

5 Model assimilations

We assimilate MOPITT observations using the GEOS-Chem model to show how filtering and the bias correction affect estimated emissions inferred from inversion analyses. We conducted three experiments in which we assimilated the following datasets: 1) the original MOPITT data, 2) the filtered and bias-corrected data with a scaling ~~to~~ down of about 6 % to match the standard TCCON data (Fig. ~~??~~5) (referred to as Assim. 2), 3) the filtered and bias-corrected data with a scaling of less than 0.5 % to the TCCON-based data not tied to the WMO scale and without the empirical airmass correction (~~Fig. ??g~~SM Fig. S12g) (referred to as Assim. 3). The assimilation is performed using the GEOS-Chem four-dimensional variational (4D-Var) data assimilation system, employing version 35J of the adjoint model at a horizontal resolution of $4° \times 5°$. The GEOS-Chem 4D-Var system has been used in previous studies for assimilation of MOPITT data (e.g., Kopacz et al., 2010; Jiang et al., 2013, 2015, 2017). We assimilate the MOPITT data to optimize monthly average CO emissions. We assimilate daytime observations for the periods of October–December 2009, and May–July 2011 to coincide with flights from the HIAPER Pole-to-Pole Observations (HIPPO) campaign. The posterior CO fluxes are compared with the a priori, and the posterior CO concentrations are validated against CO measurements from the HIPPO 10-second merged data (Santoni et al., 2014).

The assimilation uses the offline CO simulation in GEOS-Chem with prescribed monthly mean OH fields from TransCom (Patra et al., 2011) to compute the sink of CO. The prior anthropogenic CO emissions are from the EDGAR v4.2 inventory,

which are overwritten regionally with the following inventories: the Streets 2006 emissions over China and Southeast Asia from Zhang et al. (2009), the annual Canadian anthropogenic emissions from the Criteria Air Contaminants (CAC), the National Emissions 2005 Inventory (NEI2005) from the United States Environmental Protection Agency (EPA), the Co-operative Programme for Monitoring and Evaluation of the Long-range Transmission of Air Pollutants in Europe (EMEP) inventory, and the Big Bend Regional Aerosol and Visibility Observational (BRAVO) inventory in Mexico. The Global Fire Emissions Database version 3 (GFED3) provides the biomass emissions. The biofuel emissions are the Yevich and Logan (2003) inventory. The initial condition of CO states is generated by spinning up the GEOS-Chem model from January 2009. The initial CO concentrations are not optimized in the assimilation. The prior emissions are scaled by a factor of 1.5, and the emission error is purposely set to be 500%, so that the posterior CO source estimates will be less influenced by the a priori emissions and more strongly reflect the information from the filtered MOPITT observations.

~~Compared to the~~ Using HIPPO-2 and HIPPO-4 measurements for comparison, the a priori ~~are biased~~ produces mole fractions that are low by approximately 5% (Tables 4 and 5). ~~In general~~On the other hand, the original MOPITT ~~data assimilation and~~ assimilation and the assimilation which is not tied to WMO (assimilation 3) tend to agree with each other ~~as compared with HIPPO~~ and are biased high relative to HIPPO measurements. Assimilation 2 ~~results are lower~~ mole fractions are lower compared to HIPPO than the other assimilations. This suggests that scaling down MOPITT observations to match TCCON is translated to less CO in the assimilation, as expected. However, the ~~difference depends on the specific latitude band as does the assimilation that is closest to HIPPO~~comparison with HIPPO shows mixed results with each simulation depending on which latitudinal bands is considered.

To validate the quality of the filtered and bias-corrected MOPITT observations, the prior CO fluxes are compared with the a posteriori (Fig. 9). Again we find that assimilations 1 and 3 are in general agreement and assimilation 2 produces lower fluxes. Fluxes using assimilated data are nearly always smaller than fluxes using the prior scaled up by 50%. Assimilation 2, which includes scaling MOPITT to the standard TCCON product, ~~returns results that are the most similar~~ produces fluxes that are in between the unscaled (lower) and scaled (higher) prior. Though fluxes from assimilation 2 are closest to the unscaled prior~~and is~~, they are higher by about 30% and 15% during HIPPO-2 and HIPPO-4 respectively.

These results are inconclusive as to which of the assimilations is best. Comparisons with HIPPO mole fractions are mixed, and uncertainties in the assimilated prior fluxes prevent us from making definitive conclusions from the flux comparison. It is unclear if the filtering and bias corrections improved fluxes in these experiments.

6 Discussion

6.1 Practical considerations in intercomparisons of remote sounding retrievals

In addition to the formal aspects of intercomparing retrievals from different remote sounding retrievals, there are a variety of practical aspects to consider. For several of these aspects, an entire study could be devoted to them for each intercomparison. We summarize our comparison methodology in Table 6 and give examples of other studies that provide additional details or
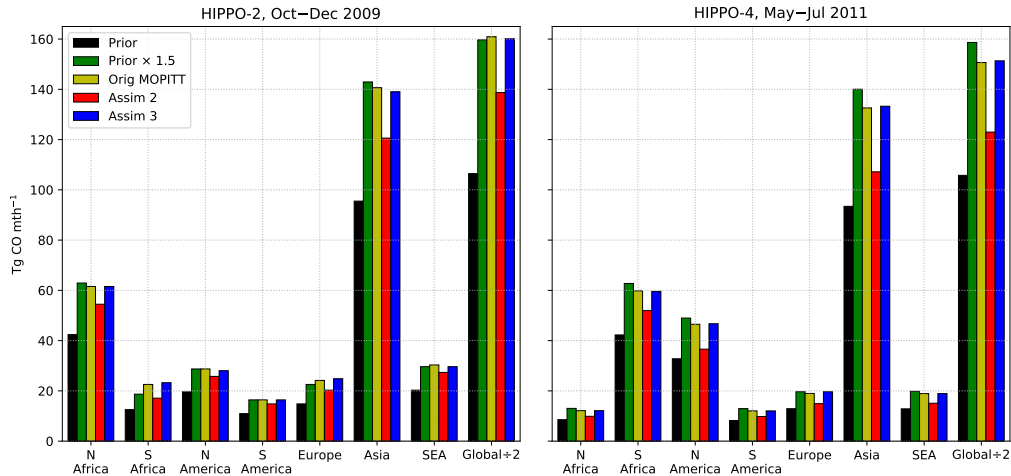
Figure 9. Emission estimates from assimilations in GEOS-Chem adjoint model for two different times. For this figure global emissions are scaled down by 2. SEA=South East Asia.

Table 4. Comparisons of GEOS-Chem model simulated mole fractions assimilating MOPITT data with HIPPO-2 observations. Uncertainties are 1σ. Units are in ppb. See text for descriptions of the assimilations.

Latitude bands	Prior	Orig. MOPITT	Assim. 2	Assim. 3
$[-40, -20]$	-4.8 ± 14.0	2.5 ± 15.3	-2.4 ± 14.9	2.1 ± 15.2
$[-20, 20]$	-5.4 ± 6.9	0.0 ± 7.3	-6.0 ± 6.9	-0.3 ± 7.3
$[20, 40]$	-5.3 ± 22.7	14.0 ± 28.3	-4.1 ± 22.9	13.2 ± 28.1
$[40, 60]$	-7.4 ± 14.1	2.9 ± 15.7	-6.2 ± 13.9	2.5 ± 15.6
$[-90, 90]$	-5.2 ± 19.0	3.3 ± 19.9	-4.2 ± 18.9	3.0 ± 19.8

alternative methods. Though it is impractical to test all combinations of different considerations, we test some as described in Sect. 4.2 such as coincidence criteria, filtering, bias corrections, and applications of averaging kernels.

6.2 Considerations for future MOPITT data use

Several lessons learned in this study may be useful for future versions of MOPITT data products or users assimilating the data. Additional fields used in the retrieval, such as the a priori mixing ratio from 50–0 hPa and the water vapor profile would be useful outputs when converting mixing ratios from whole-air to dry-air. Though the prior covariance matrix is fixed (Deeter et al., 2010), a single matrix per daily file may be helpful. The retrieved surface emissivity over land is on average about 0.007 or about 0.75 % larger than the prior, and the retrieved surface temperature is on average about 6 K larger than the prior (see histograms SM Fig. S5f, S5m)). This suggests prior surface emissivity and temperature values should perhaps be reconsidered

Table 5. Comparisons of GEOS-Chem model simulated mole fractions assimilating MOPITT data with HIPPO-4 observations. Uncertainties are 1σ. Units are in ppb. See text for descriptions of the assimilations.

Latitude bands	Prior	Orig. MOPITT	Assim. 2	Assim. 3
$[-40, -20]$	-2.0 ± 5.4	2.3 ± 5.9	-0.5 ± 5.4	2.4 ± 5.8
$[-20, 20]$	-6.8 ± 6.0	-1.5 ± 6.5	-5.5 ± 6.4	-1.4 ± 6.5
$[20, 40]$	-13.2 ± 17.9	-2.9 ± 18.9	-8.6 ± 18.2	-2.7 ± 19.0
$[40, 60]$	-3.5 ± 27.3	12.1 ± 26.6	4.2 ± 27.1	12.4 ± 26.6
$[-90, 90]$	-4.1 ± 17.9	6.0 ± 19.2	0.5 ± 18.5	6.3 ± 19.3

Table 6. Summary of practical considerations comparing MOPITT and TCCON soundings.

Practical consideration	Example	This work
Coincidence criteria	Nguyen et al. (2014)	$2° \times 4°$, ± 30 min (typically), Sect. 4.1
Weighted averaging	Hochstaffl et al. (2018) (distance & time)	On reported errors $\frac{1}{\epsilon^2}$, Apdx A
Weighted avg error	Gatz and Smith (1995)	Gatz & Smith method, Apdx. A
Representation errors (e.g., p)	Buchholz et al. (2017); Té et al. (2016)	Accounting for surface p, Sect. 4.1
Filtering	Genetic algorithms, e.g., Mandrake et al. (2013)	Chosen using SRA, Sect. 3.3
Bias corrections	Wunch et al. (2011b); O'Dell et al. (2018)	Primarily pixel-based, Sect. 3.4
Accounting for different $\mathbf{x}_a$, $\mathbf{A}$	Rodgers and Connor (2003); Wunch et al. (2011b)	~~Appendix ??~~ SM Sect 6
Pre-averaging	Buchholz et al. (2017), Apdx. B therein	Typically point-wise application of $\mathbf{x}_a$, $\mathbf{A}$
Vertical regridding	Delhez (2003)	Mass conserving, Apdx C

10   as they may be biased low over land. Further updates to prior values of CO, N_2O, and H_2O are expected to further improve the retrievals. For example, the retrieved column CO is slightly larger than the a priori globally (SM Fig. ~~S15~~ S16) but the difference depends on the level and could be related to uncertainties in model transport, sinks, and sources.

      Filtering can reduce spurious values. MOPITT files include parameters that could be used in filtering such as a Retrieval Anomaly Diagnostic, various cloud indicators, and DOF. Data users should consider creating a QC flag for their analyses or a

15   binary flag could be included in future versions, e.g. based on parameters in Table 3 or based on the recommendations of the MOPITT team (e.g., the L3 filters). Often highly deviant retrieved surface temperatures show up around coastlines (especially western coastlines, SM Fig. S10) that did not pass our quality screening. These may be related to sounding definitions of surface type. SNR 6A is used to filter MOPITT data when creating the TIR-NIR L3 product to maintain consistency with the NIR product and increase stability of the DOF, but we do not find sufficient evidence to use it as a TIR-NIR filter criterion

20   based on X_{CO} stability alone.

When biases are found in the MOPITT L2 data the strategy is to correct the L1 radiances or the retrieval algorithm (e.g., Deeter et al., 2019). MOPITT data users of L2 X_{CO} may consider implementing a bias correction before analysis or model assimilation. In terms of X_{CO}, pixel 3 data agree with pixels 2 and 4, however this agreement may not necessarily hold for retrieved profiles and pixel 3 data are excluded in the L3 product due to excessive NIR noise and to increase stability in the DOF (e.g., Deeter et al., 2015). A bias correction should be considered when assimilating pixel 1. There is a bias in the SRA for large retrieval errors on X_{CO} above about 8 ppb (SM Fig. S4o, S5o). This bias suggests that perhaps these data should be excluded or deweighted further, which we did not do here. A bias adjustment field could also be included as a field in future MOPITT files. Such an adjustment could account for empirical biases noted with various parameters, pixel-to-pixel biases (Sect. 3.4), and an overall bias compared with NDACC (Buchholz et al., 2017), aircraft flights (Deeter et al., 2017), and/or TCCON.

7 Conclusions

In this study quality filtered and bias corrected MOPITT data are compared with TCCON data. We first derive filters using only the MOPITT data assuming homogeneity over small regions. These filters have the largest affect over snow/ice scenes and over high terrain. They reduce the overall RMS from 3.84 ppb to 2.55 ppb. We find and correct a bias among the four pixels, which we confirmed exists using AirCore. We also find and correct a feature bias.

After the filtering and bias correction, we compare with TCCON data. Using a method (method II, see SM Sect. 6.1) similar to Wunch et al. (2011b) to account for differences in priors and AKs we find MOPITT data are biased high by about 6 % compared with TCCON, but it is not clear which of MOPITT or TCCON is biased. We also test different methods which all lead to a bias of about 6–10 %. There is a trend of $-0.06 \pm 0.06 \%$ yr^{-1} in the MOPITT$-$TCCON difference. The bias also appears to depend on site and latitude, but the scatter is not consistent enough to derive a correction. We also compared AKs and information content from the different retrievals. TCCON AKs are more sensitive to changes in the stratosphere. MOPITT AKs peak in the mid-troposphere and can vary significantly among locations.

After applying filtering, and an overall scaling to match the TCCON, we assimilate the data into GEOS-Chem. Filtering and bias correction are uniform enough to not make a large difference among regional fluxes. When data are also scaled down to TCCON before implementing into GEOS-Chem, fluxes were lower in all regions. However, because of bottom-up uncertainties in global CO fluxes these experiments were inconclusive. Additional work is needed to understand the relatively large ($\sim 6 \%$) difference between MOPITT and the TCCON.

5 Appendix A: Calculation of X_{CO} and weighted averaging

The MOPITT V7 data product contains fields for retrieved total column CO (in molec cm^{-2}). Unlike the TCCON, MOPITT does not retrieve a dry-air column. However, a model dry air column is provided. We obtain a dry-air mole fraction from

$$X_{CO}, \text{ppb} = \frac{retrieved\ CO\ column}{model\ dry\ air\ column} \times 10^9. \tag{A1}$$

The retrieval error in molec cm^{-2} can be converted to ppb the same way.

10   When averaging n soundings together, we use a weighted average using the inverse squared retrieval errors as weights. The average retrieved value $\bar{\hat{y}}$ is

$$\bar{\hat{y}} = \frac{\sum\limits_{i}^{n} \hat{y}_i \hat{y}_{i,\text{err}}^{-2}}{\sum\limits_{i}^{n} \hat{y}_{i,\text{err}}^{-2}}, \tag{A2}$$

where $\hat{y}_i$ denotes an individual measurement in the bin, and $\hat{y}_{i,\text{err}}$ is the corresponding error. When an average weighted error is needed, we calculate a weighted standard error of the mean (SEM) using

$$15 \quad \text{SEM} = \sqrt{\frac{n}{(n-1)\left(\sum\limits_{i}^{n} \hat{y}_{i,\text{err}}^{-2}\right)^2} \sum\limits_{i}^{n} \left(\hat{y}_{i,\text{err}}^{-2}\right)^2 \left(\hat{y}_i - \bar{\hat{y}}\right)^2}. \tag{A3}$$

In the case of uniform weights $\hat{y}_{i,\text{err}}^{-2}$ this reduces to the typical SEM equation. We also test a bootstrap analysis (Efron and Gong, 1983) on binned data for one of the parameters (DEsfc) in the bias correction analysis (Section 3.4) to evaluate Eq. A3. Data are placed into 146 bins with at least 2000 points in each. The bootstrap is ran 500 times per bin. We find, in agreement with Gatz and Smith (1995), that Eq. A3 is a reasonable approximation to the SEM determined from the bootstrap method with

20   an offset of only +0.2 % $\pm$ 3.1 % ($1\,\sigma$).

Appendix B: MOPITT column averaging kernel

We derive our own MOPITT column averaging kernel (AK) vector based on the full averaging kernel matrix. To fulfill Eq. 2 (and using Eq. 4), MOPITT AK elements a_j are:

$$a_j = \frac{\partial \hat{c}}{\partial \log_{10} x_j} = \frac{\partial \sum_i h_i \hat{x}_i}{\partial \log_{10} x_j}. \tag{B1}$$

25   Making use of $\frac{\partial \log_{10} \hat{x}_i}{\partial \hat{x}_i} = \frac{1}{\hat{x}_i \ln 10}$, and $A_{ij} = \frac{\partial \log_{10} \hat{x}_i}{\partial \log_{10} x_j}$

$$a_j = \sum_i h_i \frac{\partial \hat{x}_i}{\partial \log_{10} x_j} = \ln 10 \sum_i h_i \hat{x}_i \frac{\partial \log_{10} \hat{x}_i}{\partial \log_{10} x_j} = \ln 10 \sum_i h_i \hat{x}_i A_{ij}. \tag{B2}$$

This MOPITT column averaging kernel is not directly comparable with the TCCON column averaging kernels because of the log scale. A unitless column averaging kernel can be made, but requires an a apriori assumption about the true state of the

atmosphere. E.g.:

$$5 \quad \frac{1}{h_j}\frac{\partial \hat{c}}{\partial x_j} = \frac{1}{h_j}\frac{\partial \hat{c}}{\partial \log_{10} x_j}\frac{\partial \log_{10} x_j}{\partial x_j} = \frac{1}{h_j}\frac{a_j}{x_j \ln 10}. \quad (B3)$$

Appendix C: Vertical regridding

We find it necessary to express values from one retrieval on the vertical pressure grid of the other. MOPITT profiles are reported as layer averages, but TCCON profiles are reported as level values. TCCON profiles are converted to the MOPITT grid by linear interpolation. We divide each MOPITT layer into 500 finer equal pressure layers (about 0.4 hPa each). We interpolate

10 the TCCON profiles to these finer layers and then take the overall average to put the TCCON profile on the MOPITT pressure grid.

Basic interpolation on the midpoints should not be used to convert the MOPITT layer averages to the TCCON grid because it does not require that mass is conserved when the layers have different widths. Instead we use a mass conserving linear interpolation scheme based on the MOPITT layer ~~midpoints~~averages. This is based ~~in part~~ on the work of Hedelius and Wunch (2019),

15 and Delhez (2003).

Appendix D: ~~Accounting for different averaging kernels and priors when comparing MOPITT and TCCON~~

C1 ~~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~~

20 ~~methods of Rodgers and Connor (2003), and further detailed by Wunch et al. (2011b). 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 Eq. 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?~~

25 ~~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 ??.~~

~~=0pt 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 $\mathbf{x}_{T,a}$ $\hat{\mathbf{x}}'_M = \hat{\mathbf{x}}_M + \mathbf{x}_{T,a} - \mathbf{x}_{M,a} + \mathbf{A}_M (\mathbf{x}_{M,a} - \mathbf{x}_{T,a})\hat{c}_T$ $\hat{c}_{T \leftarrow M'} = \hat{c}_{T,a} + \sum_j^{70} h_j a_{T,j} (\hat{x}'_M - x_{T,a})_j$ II $\mathbf{x}_{T,a}$ $\hat{\mathbf{x}}_T = \gamma_T \mathbf{x}_{T,a}$ $\hat{c}'_M = \hat{c}_M + c_{T,a} - c_{M,a} + \sum_j^{10} a_{M,j} (\log_{10} x_{M,a} - \log_{10} x_{T,a})_j \hat{c}_{M \leftarrow T} =$~~

29

5 $\quad -\sum_{j}^{10} a_{M,j} \left(\log_{10}\hat{x}_T - \log_{10}x_{T,\mathrm{a}}\right)_j$ III $\mathbf{x}_{M,\mathrm{a}}\ \hat{\mathbf{x}}_M\ \hat{c}'_T = \hat{c}_T + c_{M,\mathrm{a}} - c_{T,\mathrm{a}} +\quad \sum_{j}^{70} h_j a_{T,j} \left(x_{T,\mathrm{a}} - x_{M,\mathrm{a}}\right)_j \hat{c}_{T\leftarrow M} = c_{M,\mathrm{a}} +$

$-\sum_{j}^{70} h_j a_{T,j} \left(\hat{x}_M - x_{M,\mathrm{a}}\right)_j$ IV $\mathbf{x}_{M,\mathrm{a}}\ \hat{\mathbf{x}}'_T = \dfrac{\hat{c}'_T}{c_{M,\mathrm{a}}} \mathbf{x}_{M,\mathrm{a}}\ \hat{c}_M\ \hat{c}_{M\leftarrow T'} = c_{M,\mathrm{a}} +\quad \sum_{j}^{10} a_{M,j} \left(\log_{10}\hat{x}'_T - \log_{10}x_{M,\mathrm{a}}\right)_j$ 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

10 in Eq. 3, the TCCON column averaging kernel is $a_{T,j} = \dfrac{1}{h_j}\dfrac{\partial \hat{c}_T}{\partial x_j}$, and the TCCON retrieval scaling factor is $\gamma_T = \dfrac{\hat{c}_T}{c_{T,a}}$.

C1 ~~Use an ensemble of "true" profiles~~

~~A difficulty to the above method is that multiple adjustments are required. For example, Method II in Table ?? 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~~ $\mathbf{A}$,

15 ~~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:~~

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

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

$$\hat{\mathbf{x}}_1 = \begin{bmatrix} 1 \\ 1.09 \end{bmatrix}, \ \hat{\mathbf{x}}_2 = \begin{bmatrix} 1 \\ 1.08 \end{bmatrix}, \ \hat{\mathbf{x}}_3 = \begin{bmatrix} 1 \\ 1.11 \end{bmatrix}, \ \hat{\mathbf{x}}_{12} = \hat{\mathbf{x}}_{21} = \begin{bmatrix} 1 \\ 1.072 \end{bmatrix}, \ \hat{\mathbf{x}}_{13} = \hat{\mathbf{x}}_{31} = \begin{bmatrix} 1 \\ 1.099 \end{bmatrix},$$

$$\hat{\mathbf{x}}_{23} = \hat{\mathbf{x}}_{32} = \begin{bmatrix} 1 \\ 1.088 \end{bmatrix}, \ \hat{\mathbf{x}}_{11} = \begin{bmatrix} 1 \\ 1.081 \end{bmatrix}, \ \hat{\mathbf{x}}_{22} = \begin{bmatrix} 1 \\ 1.064 \end{bmatrix}, \ \hat{\mathbf{x}}_{33} = \begin{bmatrix} 1 \\ 1.121 \end{bmatrix}.$$

20 ~~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 ??.~~

~~=0pt Differences between retrieved and smoothed profiles. x_{i1} x_{i2} x_{i3} $x_1 - x_{1j}$ 0.009 0.018 −0.009 $x_2 - x_{2j}$ 0.008 0.016~~

25 ~~−0.008 $x_3 - x_{3j}$ 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.~~

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: $\mathbf{x}_{T,\mathrm{a}}$, $\mathbf{x}_{M,\mathrm{a}}$, $\hat{\mathbf{x}}_T$, $\hat{\mathbf{x}}_M$, $\dfrac{\hat{c}_M}{c_{T,a}}\mathbf{x}_{T,\mathrm{a}}$, $\dfrac{\hat{c}_T}{c_{M,a}}\mathbf{x}_{M,\mathrm{a}}$,

$\frac{\hat{c}_T}{\hat{c}_M}\hat{\mathbf{x}}_M$, and $\frac{\hat{c}_M}{c_{M,a}}\mathbf{x}_{M,\mathrm{a}}$. In the case of 2 profile retrievals:

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

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). More specifically for MOPITT and TCCON column retrievals

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

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 ?? 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 ??.

=0pt Simulated differences between retrievals assuming a true profile. $\mathbf{x}_1$ $\mathbf{x}_2$ $\mathbf{x}_3$ $\mathbf{x}_a$ $\boldsymbol{\epsilon}_1 - \boldsymbol{\epsilon}_2$ 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.

Appendix D: Comparison of profiles

The a priori profiles differ between MOPITT and TCCON. This could lead to different intercomparison results depending on which is chosen as a comparison ensemble. In general, the TCCON a priori profiles are smooth with only 3D variation (time, latitude, altitude) that takes into account the local tropopause height. MOPITT a priori profiles are 4D (time, latitude, longitude, altitude) and hence differ in locations with strong local pollution (e.g., Pasadena). Examples of prior and retrieved profiles and their differences for several sites are in Fig. D1. Global maps of the average ratios between the retrieved and prior values for MOPITT are available in the SM Sect. S8S9.
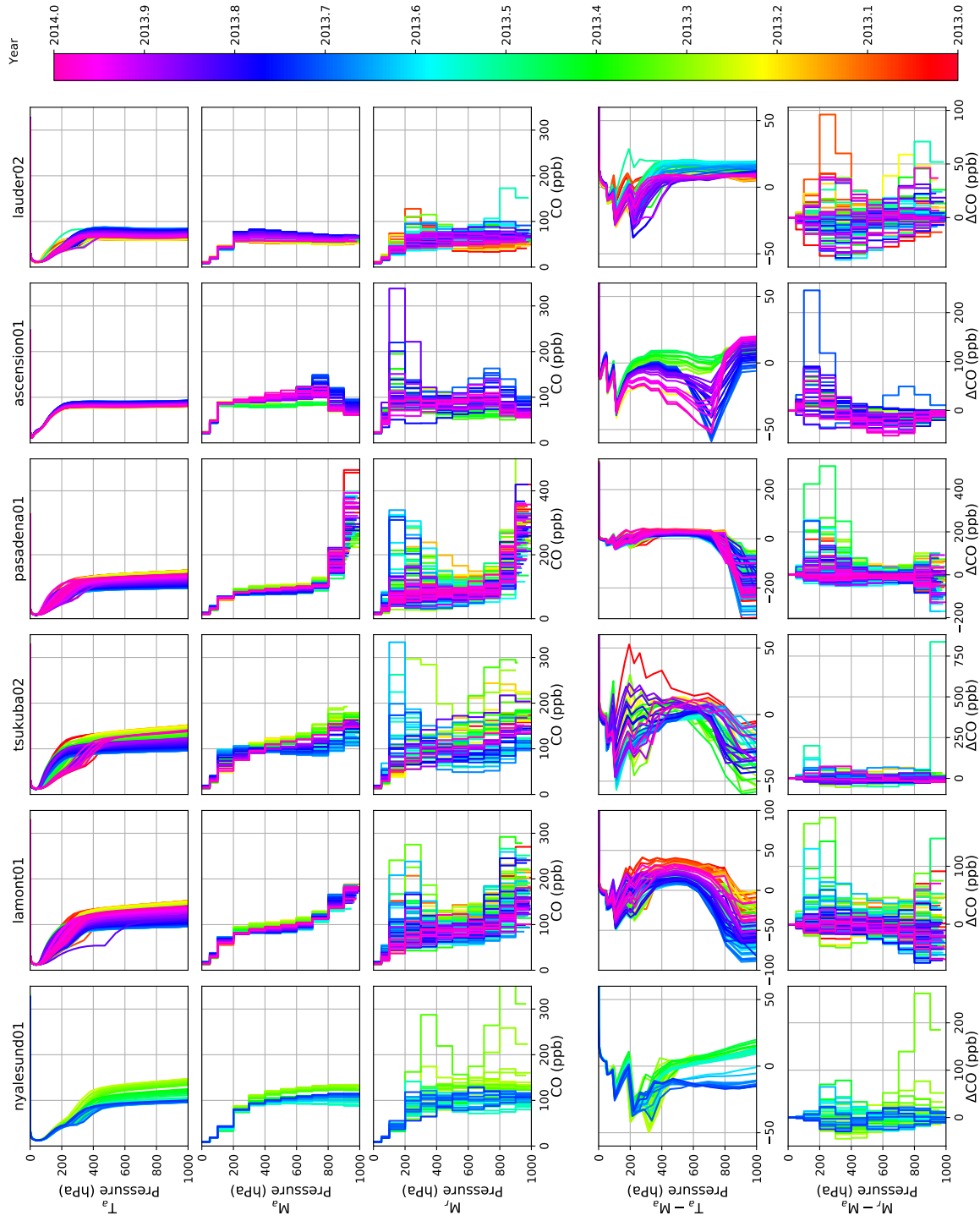
Figure D1. Profiles and profile differences between MOPITT and TCCON for six select sites and different days in 2013. For clarity, only one MOPITT profile within the coincidence criteria is selected per day. The rows are: the TCCON priors (T_a), the MOPITT priors (M_a), the MOPITT retrieved profiles (M_r), the difference between TCCON and MOPITT priors, and the difference between MOPITT priors and retrieved profiles.

JKH, DW, DBAJ, RRB, NMD, FH, MK, IM, JN, RS, POW, HMW, CS, and BCB reviewed the manuscript. JKH and DW implemented edits to the manuscript.

5 *Competing interests.* We declare that we have no competing interests.

References

Blumenstock, T., Hase, F., Schneider, M., Garcia, O. E., and Sepulveda, E.: TCCON data from Izana (ES), Release GGG2014R0, TCCON
data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.izana01.R0/1149295, 2014.

Borsdorff, T., aan de Brugh, J., Hu, H., Hasekamp, O., Sussmann, R., Rettinger, M., Hase, F., Gross, J., Schneider, M., Garcia, O., Stremme,
W., Grutter, M., Feist, D. G., Arnold, S. G., De Mazière, M., Kumar Sha, M., Pollard, D. F., Kiel, M., Roehl, C., Wennberg, P. O., Toon,
G. C., and Landgraf, J.: Mapping carbon monoxide pollution from space down to city scales with daily global coverage, Atmospheric
Measurement Techniques, 11, 5507–5518, https://doi.org/10.5194/amt-11-5507-2018, 2018.

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.

Connor, B. J., Boesch, H., Toon, G., Sen, B., Miller, C., and Crisp, D.: Orbiting Carbon Observatory: Inverse method and prospective error
analysis, Journal of Geophysical Research: Atmospheres, 113, D05 305, https://doi.org/10.1029/2006JD008336, 2008.

Connor, B. J., Sherlock, V., Toon, G., Wunch, D., and Wennberg, P. O.: GFIT2: An experimental algorithm for vertical profile retrieval from
near-IR spectra, Atmospheric Measurement Techniques, 9, 3513–3525, https://doi.org/10.5194/amt-9-3513-2016, 2016.

De Mazière, M., Sha, M. K., Desmet, F., Hermans, C., Scolas, F., Kumps, N., Metzger, J.-M., Duflot, V., and Cam-
mas, J.-P.: TCCON data from Reunion Island (RE), Release GGG2014R0, TCCON data archive, hosted by CaltechDATA,
https://doi.org/10.14291/tccon.ggg2014.reunion01.R0/1149288, 2014.

Deeter, M. N.: MOPITT (Measurements of Pollution in the Troposphere) Version 7 Product User's Guide, 2017.

Deeter, M. N., Edwards, D. P., Gille, J. C., Emmons, L. K., Francis, G., Ho, S. P., Mao, D., Masters, D., Worden, H., Drummond, J. R., and
Novelli, P. C.: The MOPITT version 4 CO product: Algorithm enhancements, validation, and long-term stability, Journal of Geophysical
Research-Atmospheres, 115, 1–13, https://doi.org/Artn D07306\nDoi 10.1029/2009jd013005, 2010.

Deeter, M. N., Martínez-Alonso, S., Edwards, D. P., Emmons, L. K., Gille, J. C., Worden, H. M., Sweeney, C., Pittman, J. V., Daube, B. C.,
and Wofsy, S. C.: The MOPITT Version 6 product: Algorithm enhancements and validation, Atmospheric Measurement Techniques, 7,
3623–3632, https://doi.org/10.5194/amt-7-3623-2014, 2014.

Deeter, M. N., Edwards, D. P., Gille, J. C., and Worden, H. M.: Information content of MOPITT CO profile retrievals: Temporal and
geographical variability, Journal of Geophysical Research, 120, 723–12, https://doi.org/10.1002/2015JD024024, 2015.

Deeter, M. N., Martínez-Alonso, S., Gatti, L. V., Gloor, M., Miller, J. B., Domingues, L. G., and Correia, C. S.: Validation and analysis of
MOPITT CO observations of the Amazon Basin, Atmospheric Measurement Techniques, 9, 3999–4012, https://doi.org/10.5194/amt-9-
3999-2016, 2016.

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.

Deeter, M. N., Edwards, D. P., Francis, G. L., Gille, J. C., Mao, D., Martinez-Alonso, S., Worden, H. M., Ziskin, D., and Andreae, M. O.:
Radiance-based Retrieval Bias Mitigation for the MOPITT Instrument: The Version 8 Product, Atmospheric Measurement Techniques
Discussions, pp. 1–30, https://doi.org/10.5194/amt-2019-41, 2019.

Delhez, E.: A spline interpolation technique that preserves mass budgets, Applied Mathematics Letters, 16, 17–26, https://doi.org/10.1016/S0893-9659(02)00139-8, 2003.

20 Deutscher, N. M., Notholt, J., Messerschmidt, J., Weinzierl, C., Warneke, T., Petri, C., Grupe, P., and Katryn-ski, K.: TCCON data from Bialystok (PL), Release GGG2014R1, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.bialystok01.R1/1183984, 2014.

Drummond, J. R., Zou, J., Nichitiu, F., Kar, J., Deschambaut, R., and Hackett, J.: A review of 9-year performance and operation of the MOPITT instrument, Advances in Space Research, 45, 760–774, https://doi.org/10.1016/j.asr.2009.11.019, 2010.

25 Dubey, M., Henderson, B., Green, D., Butterfield, Z., Keppel-Aleks, G., Allen, N., Blavier, J.-F., Roehl, C., Wunch, D., and Lindenmaier, R.: TCCON data from Manaus (BR), Release GGG2014R0, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.manaus01.R0/1149274, 2014a.

Dubey, M., Lindenmaier, R., Henderson, B., Green, D., Allen, N., Roehl, C., Blavier, J.-F., Butterfield, Z., Love, S., Hamelmann, J., and Wunch, D.: TCCON data from Four Corners (US), Release GGG2014R0, TCCON data archive, hosted by CaltechDATA,

30 https://doi.org/10.14291/tccon.ggg2014.fourcorners01.R0/1149272, 2014b.

Efron, B. and Gong, G.: A leisurely look at the bootstrap, the jackknife, and cross-validation, American Statistician, 37, 36–48, https://doi.org/10.1080/00031305.1983.10483087, 1983.

Engel, A., Bönisch, H., Ullrich, M., Sitals, R., Membrive, O., Danis, F., and Crevoisier, C.: Mean age of stratospheric air derived from AirCore observations, Atmospheric Chemistry and Physics, 17, 6825–6838, https://doi.org/10.5194/acp-17-6825-2017, 2017.

35 Feist, D. G., Arnold, S. G., John, N., and Geibel, M. C.: TCCON data from Ascension Island (SH), Release GGG2014R0, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.ascension01.R0/1149285, 2014.

Forster, P., Ramaswamy, V., Artaxo, P., Berntsen, T., Betts, R., Fahey, D., Haywood, J., Lean, J., Lowe, D., Myhre, G., Nganga, J., Prinn, R., Raga, G., Schulz, M., and Van Dorland, R.: Changes in Atmospheric Constituents and in Radiative Forcing, in: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., K.B., A., Tignor, M., and Miller, H., chap. 2, pp. 129–234, Cambridge University Press, Cambridge, United Kingdom and New York, USA, https://www.ipcc.ch/publications_and_data/ar4/wg1/en/contents.html, 2007.

5 Gatz, D. F. and Smith, L.: The standard error of a weighted mean concentration—I. Bootstrapping vs other methods, Atmospheric Environment, 29, 1185–1193, https://doi.org/10.1016/1352-2310(94)00210-C, 1995.

Goo, T.-Y., Oh, Y.-S., and Velazco, V. A.: TCCON data from Anmeyondo (KR), Release GGG2014R0, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.anmeyondo01.R0/1149284, 2014.

Griffith, D. W., Deutscher, N. M., Velazco, V. A., Wennberg, P. O., Yavin, Y., Aleks, G. K., Washenfelder, R. a., Toon, G. C.,

10 Blavier, J.-F., Murphy, C., Jones, N., Kettlewell, G., Connor, B. J., Macatangay, R., Roehl, C., Ryczek, M., Glowacki, J., Cul-gan, T., and Bryant, G.: TCCON data from Darwin (AU), Release GGG2014R0, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.darwin01.R0/1149290, 2014a.

Griffith, D. W., Velazco, V. A., Deutscher, N. M., Murphy, C., Jones, N., Wilson, S., Macatangay, R., Kettlewell, G., Buchholz, R. R., and Riggenbach, M.: TCCON data from Wollongong (AU), Release GGG2014R0, TCCON data archive, hosted by CaltechDATA,

15 https://doi.org/10.14291/tccon.ggg2014.wollongong01.R0/1149291, 2014b.

Hase, F., Blumenstock, T., Dohe, S., Gross, J., and Kiel, M.: TCCON data from Karlsruhe (DE), Release GGG2014R1, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.karlsruhe01.R1/1182416, 2014.

Hedelius, J. K. and Wunch, D.: Mass conserving quartic-spline and linear interpolation techniques over variable width partitions, Scholars Portal Dataverse, V1, https://doi.org/10.5683/SP2/57ROYV, 2019.

20 Hedelius, J. K., Feng, S., Roehl, C. M., Wunch, D., Hillyard, P. W., Podolske, J. R., Iraci, L. T., Patarasuk, R., Rao, P., O'Keeffe, D., Gurney, K. R., Lauvaux, T., and Wennberg, P. O.: Emissions and topographic effects on column CO 2 (XCO2) variations, with a focus on the Southern California Megacity, Journal of Geophysical Research: Atmospheres, 122, 7200–7215, https://doi.org/10.1002/2017JD026455, 2017.

Hochstaffl, P., Schreier, F., Lichtenberg, G., and García, S. G.: Validation of carbon monoxide total column retrievals from SCIAMACHY
25   observations with NDACC/TCCON ground-based measurements, Remote Sensing, 10, https://doi.org/10.3390/rs10020223, 2018.

Inoue, M., Morino, I., Uchino, O., Nakatsuru, T., Yoshida, Y., Yokota, T., Wunch, D., Wennberg, P. O., Roehl, C. M., Griffith, D. W. T., Velazco, V. A., Deutscher, N. M., Warneke, T., Notholt, J., Robinson, J., Sherlock, V., Hase, F., Blumenstock, T., Rettinger, M., Sussmann, R., Kyr??, E., Kivi, R., Shiomi, K., Kawakami, S., De Mazi??re, M., Arnold, S. G., Feist, D. G., Barrow, E. A., Barney, J., Dubey, M., Schneider, M., Iraci, L. T., Podolske, J. R., Hillyard, P. W., Machida, T., Sawa, Y., Tsuboi, K., Matsueda, H., Sweeney, C., Tans, P. P.,
30   Andrews, A. E., Biraud, S. C., Fukuyama, Y., Pittman, J. V., Kort, E. A., and Tanaka, T.: Bias corrections of GOSAT SWIR XCO2 and XCH4 with TCCON data and their evaluation using aircraft measurement data, Atmospheric Measurement Techniques, 9, 3491–3512, https://doi.org/10.5194/amt-9-3491-2016, 2016.

Iraci, L. T., Podolske, J., Hillyard, P. W., Roehl, C., Wennberg, P. O., Blavier, J.-F., Allen, N., Wunch, D., Osterman, G. B., and Albertson, R.: TCCON data from Edwards (US), Release GGG2014R1, TCCON data archive, hosted by CaltechDATA,
35   https://doi.org/10.14291/tccon.ggg2014.edwards01.R1/1255068, 2016a.

Iraci, L. T., Podolske, J., Hillyard, P. W., Roehl, C., Wennberg, P. O., Blavier, J.-F., Landeros, J., Allen, N., Wunch, D., Zavaleta, J., Quigley, E., Osterman, G. B., Barrow, E., and Barney, J.: TCCON data from Indianapolis (US), Release GGG2014R1, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.indianapolis01.R1/1330094, 2016b.

Jiang, Z., Jones, D. B. A., Worden, H. M., Deeter, M. N., Henze, D. K., Worden, J., Bowman, K. W., Brenninkmeijer, C. A. M., and Schuck, T. J.: Impact of model errors in convective transport on CO source estimates inferred from MOPITT CO retrievals, Journal of Geophysical
5   Research Atmospheres, 118, 2073–2083, https://doi.org/10.1002/jgrd.50216, 2013.

Jiang, Z., Jones, D. B., Worden, H. M., and Henze, D. K.: Sensitivity of top-down CO source estimates to the modeled vertical structure in atmospheric CO, Atmospheric Chemistry and Physics, 15, 1521–1537, https://doi.org/10.5194/acp-15-1521-2015, 2015.

Jiang, Z., Worden, J. R., Worden, H., Deeter, M., Jones, D. B., Arellano, A. F., and Henze, D. K.: A 15-year record of CO emissions constrained by MOPITT CO observations, Atmospheric Chemistry and Physics, 17, 4565–4583, https://doi.org/10.5194/acp-17-4565-
10   2017, 2017.

Karion, A., Sweeney, C., Tans, P., and Newberger, T.: AirCore: An innovative atmospheric sampling system, Journal of Atmospheric and Oceanic Technology, 27, 1839–1853, https://doi.org/10.1175/2010JTECHA1448.1, 2010.

Kawakami, S., Ohyama, H., Arai, K., Okumura, H., Taura, C., Fukamachi, T., and Sakashita, M.: TCCON data from Saga (JP), Release GGG2014R0, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.saga01.R0/1149283, 2014.

15 Kiel, M., Wunch, D., Wennberg, P. O., Toon, G. C., Hase, F., and Blumenstock, T.: Improved retrieval of gas abundances from near-infrared solar FTIR spectra measured at the Karlsruhe TCCON station, Atmospheric Measurement Techniques, 9, 669–682, https://doi.org/10.5194/amt-9-669-2016, 2016.

Kiel, M., O'Dell, C. W., Fisher, B., Eldering, A., Nassar, R., MacDonald, C. G., and Wennberg, P. O.: How bias correction goes wrong: measurement of XCO2 affected by erroneous surface pressure estimates, Atmospheric Measurement Techniques, 12, 2241–2259, https://doi.org/10.5194/amt-12-2241-2019, 2019.

Kivi, R. and Heikkinen, P.: Fourier transform spectrometer measurements of column CO2 at Sodankylä, Finland, Geoscientific Instrumentation, Methods and Data Systems, 5, 271–279, https://doi.org/10.5194/gi-5-271-2016, 2016.

Kivi, R., Heikkinen, P., and Kyrö, E.: TCCON data from Sodankyla (FI), Release GGG2014R0, TCCON data archive, hosted by Caltech-DATA, https://doi.org/10.14291/tccon.ggg2014.sodankyla01.R0/1149280, 2017.

Kopacz, M., Jacob, D. J., Fisher, J. A., Logan, J. A., Zhang, L., Megretskaia, I. A., Yantosca, R. M., Singh, K., Henze, D. K., Burrows, J. P., Buchwitz, M., Khlystova, I., McMillan, W. W., Gille, J. C., Edwards, D. P., Eldering, A., Thouret, V., and Nedelec, P.: Global estimates of CO sources with high resolution by adjoint inversion of multiple satellite datasets (MOPITT, AIRS, SCIAMACHY, TES), Atmospheric Chemistry and Physics, 10, 855–876, https://doi.org/10.5194/acp-10-855-2010, 2010.

Kulawik, S., Wunch, D., O'Dell, C., Frankenberg, C., Reuter, M., Oda, T., Chevallier, F., Sherlock, V., Buchwitz, M., Osterman, G., Miller, C. E., Wennberg, P. O., Griffith, D., Morino, I., Dubey, M. K., Deutscher, N. M., Notholt, J., Hase, F., Warneke, T., Sussmann, R., Robinson, J., Strong, K., Schneider, M., De Mazière, M., Shiomi, K., Feist, D. G., Iraci, L. T., and Wolf, J.: Consistent evaluation of ACOS-GOSAT, BESD-SCIAMACHY, CarbonTracker, and MACC through comparisons to TCCON, Atmospheric Measurement Techniques, 9, 683–709, https://doi.org/10.5194/amt-9-683-2016, 2016.

Lamarque, J.-F., Emmons, L. K., Hess, P. G., Kinnison, D. E., Tilmes, S., Vitt, F., Heald, C. L., Holland, E. A., Lauritzen, P. H., Neu, J., Orlando, J. J., Rasch, P. J., and Tyndall, G. K.: CAM-chem: description and evaluation of interactive atmospheric chemistry in the Community Earth System Model, Geoscientific Model Development, 5, 369–411, https://doi.org/10.5194/gmd-5-369-2012, 2012.

Lelieveld, J., Gromov, S., Pozzer, A., and Taraborrelli, D.: Global tropospheric hydroxyl distribution, budget and reactivity, Atmospheric Chemistry and Physics, 16, 12 477–12 493, https://doi.org/10.5194/acp-16-12477-2016, 2016.

Lindenmaier, R., Dubey, M. K., Henderson, B. G., Butterfield, Z. T., Herman, J. R., Rahn, T., and Lee, S.-h.: Multiscale observations of CO2, 13CO2, and pollutants at Four Corners for emission verification and attribution., Proceedings of the National Academy of Sciences of the United States of America, 111, 8386–8391, https://doi.org/10.1073/pnas.1321883111, 2014.

Liu, C., Wang, W., and Sun, Y.: TCCON data from Hefei, China, Release GGG2014.R0, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.hefei01.R0, 2018.

Mandrake, L., Frankenberg, C., O'Dell, C. W., Osterman, G., Wennberg, P., and Wunch, D.: Semi-autonomous sounding selection for OCO-2, Atmospheric Measurement Techniques, 6, 2851–2864, https://doi.org/10.1177/1756283X16649143, 2013.

Membrive, O., Crevoisier, C., Sweeney, C., Danis, F., Hertzog, A., Engel, A., Bönisch, H., and Picon, L.: AirCore-HR: a high-resolution column sampling to enhance the vertical description of CH4 and CO2, Atmospheric Measurement Techniques, 10, 2163–2181, https://doi.org/10.5194/amt-10-2163-2017, 2017.

Morino, I., Matsuzaki, T., and Shishime, A.: TCCON data from Tsukuba (JP), 125HR, Release GGG2014R1, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.tsukuba02.R1/1241486, 2014a.

Morino, I., Yokozeki, N., Matzuzaki, T., and Horikawa, M.: TCCON data from Rikubetsu (JP), Release GGG2014R1, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.rikubetsu01.R1/1242265, 2014b.

Morino, I., Velazco, V. A., Hori, A., Uchino, O., and Griffith, D. W. T.: TCCON data from Burgos, Ilocos Norte (PH), Release GGG2014.R0, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.burgos01.R0, 2018.

Mu, M., Randerson, J. T., van der Werf, G. R., Giglio, L., Kasibhatla, P., Morton, D., Collatz, G. J., DeFries, R. S., Hyer, E. J., Prins, E. M., Griffith, D. W. T., Wunch, D., Toon, G. C., Sherlock, V., and Wennberg, P. O.: Daily and 3-hourly variability in global fire emissions and consequences for atmospheric model predictions of carbon monoxide, Journal of Geophysical Research: Atmospheres, 116, D24 303, https://doi.org/10.1029/2011JD016245, 2011.

Nguyen, H., Osterman, G., Wunch, D., O'Dell, C., Mandrake, L., Wennberg, P., Fisher, B., and Castano, R.: A method for colocating satellite XCO2 data to ground-based data and its application to ACOS-GOSAT and TCCON, Atmospheric Measurement Techniques, 7, 2631–2644, https://doi.org/10.5194/amt-7-2631-2014, 2014.

Notholt, J., Petri, C., Warneke, T., Deutscher, N. M., Buschmann, M., Weinzierl, C., Macatangay, R., and Grupe, P.: TCCON data from Bremen (DE), Release GGG2014R0, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.bremen01.R0/1149275, 2014.

Notholt, J., Warneke, T., Petri, C., Deutscher, N. M., Weinzierl, C., Palm, M., and Buschmann, M.: TCCON data from Ny Ålesund, Spitsbergen (NO), Release GGG2014.R0, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.nyalesund01.R0/1149278, 2017.

O'Dell, C. W., Eldering, A., Wennberg, P. O., Crisp, D., Gunson, M. R., Fisher, B., Frankenberg, C., Kiel, M., Lindqvist, H., Mandrake, L., Merrelli, A., Natraj, V., Nelson, R. R., Osterman, G. B., Payne, V. H., Taylor, T. E., Wunch, D., Drouin, B. J., Oyafuso, F., Chang, A., McDuffie, J., Smyth, M., Baker, D. F., Basu, S., Chevallier, F., Crowell, S. M. R., Feng, L., Palmer, P. I., Dubey, M., García, O. E., Griffith, D. W. T., Hase, F., Iraci, L. T., Kivi, R., Morino, I., Notholt, J., Ohyama, H., Petri, C., Roehl, C. M., Sha, M. K., Strong, K., Sussmann, R., Te, Y., Uchino, O., and Velazco, V. A.: Improved retrievals of carbon dioxide from Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm, Atmospheric Measurement Techniques, 11, 6539–6576, https://doi.org/10.5194/amt-11-6539-2018, 2018.

Patra, P. K., Houweling, S., Krol, M., Bousquet, P., Belikov, D., Bergmann, D., Bian, H., Cameron-Smith, P., Chipperfield, M. P., Corbin, K., Fortems-Cheiney, A., Fraser, A., Gloor, E., Hess, P., Ito, A., Kawa, S. R., Law, R. M., Loh, Z., Maksyutov, S., Meng, L., Palmer, P. I., Prinn, R. G., Rigby, M., Saito, R., and Wilson, C.: TransCom model simulations of CH 4 and related species: Linking transport, surface flux and chemical loss with CH 4 variability in the troposphere and lower stratosphere, Atmospheric Chemistry and Physics, 11, 12 813–12 837, https://doi.org/10.5194/acp-11-12813-2011, 2011.

Pollard, D. F., Sherlock, V., Robinson, J., Deutscher, N. M., Connor, B., and Shiona, H.: The Total Carbon Column Observing Network site description for Lauder, New Zealand, Earth System Science Data, 9, 977–992, https://doi.org/10.5194/essd-9-977-2017, 2017.

Rakitin, V. S., Shtabkin, Y. A., Elansky, N. F., Pankratova, N. V., Skorokhod, A. I., Grechko, E. I., and Safronov, A. N.: Comparison results of satellite and ground-based spectroscopic measurements of CO, CH4, and CO2 total contents, Atmospheric and Oceanic Optics, 28, 533–542, https://doi.org/10.1134/S1024856015060135, 2015.

Rodgers, C. D.: Characterization and error analysis of profiles retrieved from remote sounding measurements, Journal of Geophysical Research, 95, 5587–5595, https://doi.org/10.1029/JD095iD05p05587, 1990.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding - Theory and Practice, vol. 2 of *Series on Atmospheric Oceanic and Planetary Physics*, World Scientific Publishing Co. Pte. Ltd., https://doi.org/10.1142/9789812813718, http://ebooks.worldscinet.com/ISBN/9789812813718/9789812813718.html, 2000.

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.

Santoni, G. W., Daube, B. C., Kort, E. A., Jiménez, R., Park, S., Pittman, J. V., Gottlieb, E., Xiang, B., Zahniser, M. S., Nelson, D. D., McManus, J. B., Peischl, J., Ryerson, T. B., Holloway, S., Andrews, A. E., Sweeney, C., Hall, B., Hintsa, E. J., Moore, F. L., Elkins,

J. W., Hurst, D. F., Stephens, B. B., Bent, J., and Wofsy, S. C.: Evaluation of the airborne quantum cascade laser spectrometer (QCLS) measurements of the carbon and greenhouse gas suite - CO2, CH4, N2O, and CO - during the CalNex and HIPPO Campaigns, Atmospheric Measurement Techniques, 7, 1509–1526, https://doi.org/10.5194/amt-7-1509-2014, 2014.

Seinfeld, J. H. and Pandis, S. N.: Atmospheric Chemistry and Physics: From Air Pollution to Climate Change, John Wiley & Sons, Hoboken, New Jersey, USA, 2006.

Sha, M. K., Langerock, B., De Mazière, M., Dils, B., Feist, D. G., Sussmann, R., Hase, F., Schneider, M., Blumenstock, T., Notholt, J., Warneke, T., Petri, C., Kivi, R., Te, Y., Wennberg, P. O., Wunch, D., Iraci, L. T., Strong, K., Griffith, D. W. T., Deutscher, N. M., Velazco, V. A., Morino, I., Ohyama, H., Uchino, O., Shiomi, K., Goo, T. Y., Pollard, D. F., Borsdorff, T., Hu, H., Hasekamp, O. P., Landgraf, J., Roehl, C. M., Kiel, M., Toon, G., and TCCON team: First results of the Sentinel-5 Precursor CO and CH4 validation using TCCON data (ESA AO project TCCON4S5P) and NDACC CO data, in: Annual Joint NDACC-IRWG & TCCON Meeting, Mexico City, 11–15 June, Mexico City, Mexico, 2018.

Sherlock, V., Connor, B. J., Robinson, J., Shiona, H., Smale, D., and Pollard, D.: TCCON data from Lauder (NZ), 120HR, Release GGG2014R0, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.lauder01.R0/1149293, 2014a.

Sherlock, V., Connor, B. J., Robinson, J., Shiona, H., Smale, D., and Pollard, D.: TCCON data from Lauder (NZ), 125HR, Release GGG2014R0, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.lauder02.R0/1149298, 2014b.

Strong, K., Roche, S., Franklin, J. E., Mendonca, J., Lutsch, E., Weaver, D., Fogal, P. F., Drummond, J. R., Batchelor, R., and Lindenmaier, R.: TCCON data from Eureka (CA), Release GGG2014.R2, https://doi.org/10.14291/tccon.ggg2014.eureka01.r2, 2017.

Sussmann, R. and Buchwitz, M.: Initial validation of ENVISAT/SCIAMACHY columnar CO by FTIR profile retrievals at the Ground-Truthing Station Zugspitze, Atmospheric Chemistry and Physics, 5, 1497–1503, https://doi.org/10.5194/acp-5-1497-2005, 2005.

Sussmann, R. and Rettinger, M.: TCCON data from Garmisch (DE), Release GGG2014R0, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.garmisch01.R0/1149299, 2014.

Sussmann, R. and Rettinger, M.: TCCON data from Zugspitze (DE), Release GGG2014R1, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.zugspitze01.R1, 2018.

Té, Y., Jeseck, P., and Janssen, C.: TCCON data from Paris (FR), Release GGG2014R0, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.paris01.R0/1149279, 2014.

Té, Y., Jeseck, P., Franco, B., Mahieu, E., Jones, N., Paton-Walsh, C., Griffith, D. W., Buchholz, R. R., Hadji-Lazaro, J., Hurtmans, D., and Janssen, C.: Seasonal variability of surface and column carbon monoxide over the megacity Paris, high-altitude Jungfraujoch and Southern Hemispheric Wollongong stations, Atmospheric Chemistry and Physics, 16, 10 911–10 925, https://doi.org/10.5194/acp-16-10911-2016, 2016.

Velazco, V., Morino, I., Uchino, O., Hori, A., Kiel, M., Bukosa, B., Deutscher, N., Sakai, T., Nagai, T., Bagtasa, G., Izumi, T., Yoshida, Y., and Griffith, D.: TCCON Philippines: First Measurement Results, Satellite Data and Model Comparisons in Southeast Asia, Remote Sensing, 9, 1228, https://doi.org/10.3390/rs9121228, 2017.

Warneke, T., Messerschmidt, J., Notholt, J., Weinzierl, C., Deutscher, N. M., Petri, C., Grupe, P., Vuillemin, C., Truong, F., Schmidt, M., Ramonet, M., and Parmentier, E.: TCCON data from Orléans (FR), Release GGG2014R0, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.orleans01.R0/1149276, 2014.

Wennberg, P. O., Roehl, C., Wunch, D., Toon, G. C., Blavier, J.-F., Washenfelder, R. a., Keppel-Aleks, G., Allen, N., and Ayers, J.: TCCON data from Park Falls (US), Release GGG2014R0, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.parkfalls01.R0/1149161, 2014a.

Wennberg, P. O., Wunch, D., Roehl, C., Blavier, J.-F., Toon, G. C., and Allen, N.: TCCON data from Caltech (US), Release GGG2014R1, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.pasadena01.R1/1182415, 2014b.

20   Wennberg, P. O., Wunch, D., Yavin, Y., Toon, G. C., Blavier, J.-F., Allen, N., and Keppel-Aleks, G.: TCCON data from Jet Propulsion Laboratory (US), 2007, Release GGG2014R0, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.jpl01.R0/1149163, 2014c.

Wennberg, P. O., Roehl, C., Blavier, J.-F., Wunch, D., Landeros, J., and Allen, N.: TCCON data from Jet Propulsion Laboratory (US), 2011, Release GGG2014R1, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.jpl02.R1/1330096, 2016a.

25   Wennberg, P. O., Wunch, D., Roehl, C., Blavier, J.-F., Toon, G. C., Allen, N., Dowell, P., Teske, K., Martin, C., and Martin., J.: TCCON data from Lamont (US), Release GGG2014R1, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.lamont01.R1/1255070, 2016b.

Wunch, D., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Stephens, B. B., Fischer, M. L., Uchino, O., Abshire, J. B., Bernath, P., Biraud, S. C., Blavier, J.-F. L., Boone, C., Bowman, K. P., Browell, E. V., Campos, T., Connor, B. J., Daube, B. C., Deutscher, N. M., Diao, M., Elkins,

30   J. W., Gerbig, C., Gottlieb, E., Griffith, D. W. T., Hurst, D. F., Jiménez, R., Keppel-Aleks, G., Kort, E. A., Macatangay, R., Machida, T., Matsueda, H., Moore, F., Morino, I., Park, S., Robinson, J., Roehl, C. M., Sawa, Y., Sherlock, V., Sweeney, C., Tanaka, T., and Zondlo, M. A.: Calibration of the Total Carbon Column Observing Network using aircraft profile data, Atmospheric Measurement Techniques, 3, 1351–1362, https://doi.org/10.5194/amt-3-1351-2010, 2010.

Wunch, D., Toon, G. C., Blavier, J.-F. L., Washenfelder, R. A., Notholt, J., Connor, B. J., Griffith, D. W. T., Sherlock, V., and

35   Wennberg, P. O.: The Total Carbon Column Observing Network, Philosophical Transactions of the Royal Society A, 369, 2087–2112, https://doi.org/10.1098/rsta.2010.0240, 2011a.

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., 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, 2011b.

5   Wunch, D., Toon, G. C., Sherlock, V., Deutscher, N. M., Liu, C., Feist, D. G., and Wennberg, P. O.: The Total Carbon Column Observing Network's GGG2014 Data Version, Tech. rep., California Institute of Technology, Pasadena, CA, https://doi.org/10.14291/tccon.ggg2014.documentation.R0/1221662, http://dx.doi.org/10.14291/tccon.ggg2014.documentation.R0/1221662, 2015.

Wunch, D., Mendonca, J., Colebatch, O., Allen, N., Blavier, J.-F. L., Roche, S., Hedelius, J. K., Neufeld, G., Springett, S., Worthy, D.

10   E. J., Kessler, R., and Strong, K.: TCCON data from East Trout Lake (CA), Release GGG2014R1, TCCON data archive, hosted by CaltechDATA, https://doi.org/10.14291/tccon.ggg2014.easttroutlake01.R1, 2017a.

Wunch, D., Wennberg, P. O., Osterman, G., Fisher, B., Naylor, B., Roehl, C. M., O'Dell, C., Mandrake, L., Viatte, C., Kiel, M., Griffith, D. W. T., Deutscher, N. M., Velazco, V. A., Notholt, J., Warneke, T., Petri, C., De Maziere, M., Sha, M. K., Sussmann, R., Rettinger, M., Pollard, D., Robinson, J., Morino, I., Uchino, O., Hase, F., Blumenstock, T., Feist, D. G., Arnold, S. G., Strong, K., Mendonca, J., Kivi,

960   R., Heikkinen, P., Iraci, L., Podolske, J., Hillyard, P. W., Kawakami, S., Dubey, M. K., Parker, H. A., Sepulveda, E., Garcia, O. E., Te, Y., Jeseck, P., Gunson, M. R., Crisp, D., and Eldering, A.: Comparisons of the Orbiting Carbon Observatory-2 (OCO-2) XCO2 measurements with TCCON, Atmospheric Measurement Techniques, 10, 2209–2238, https://doi.org/10.5194/amt-10-2209-2017, 2017b.

Yevich, R. and Logan, J. A.: An assessment of biofuel use and burning of agricultural waste in the developing world, Global Biogeochemical Cycles, 17, n/a–n/a, https://doi.org/10.1029/2002gb001952, 2003.

965 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.

Zhang, Q., Streets, D. G., Carmichael, G. R., He, K. B., Huo, H., Kannari, A., Klimont, Z., Park, I. S., Reddy, S., Fu, J. S., Chen, D., Duan, L., Lei, Y., Wang, L. T., and Yao, Z. L.: Asian emissions in 2006 for the NASA INTEX-B mission, Atmospheric Chemistry and Physics, 9, 5131–5153, https://doi.org/10.5194/acp-9-5131-2009, 2009.

970 Zhang, Q., Natraj, V., Li, K.-F., Shia, R.-L., Fu, D., Pongetti, T. J., Sander, S. P., Roehl, C. M., and Yung, Y. L.: Accounting for aerosol scattering in the CLARS retrieval of column averaged CO 2 mixing ratios, Journal of Geophysical Research: Atmospheres, 120, 7205–7218, https://doi.org/10.1002/2015JD023499, 2015.