Anonymous Referee #1

The focus of this paper is a comparison of the latest MOPITT CO V8 retrievals with aircraft measurements. The authors use aircraft profiles that have not been previously used for comparisons with MOPITT, including profiles around urban regions. A variety of sensitivity tests are also performed to see how specific parameters affect the comparison. Overall I think the paper is well written and sound, and though I have numerous comments they are mostly minor.
Response: Thank you for your time and effort in reviewing our manuscript.

General comments
G1: Even though the word "validation" or a variant has been used frequently in these contexts in the literature, "comparison" would nearly always be a better choice. "Validation" has a positive connotation which to me makes it sound like nothing new can be learned, and like the result is already known before the study. However, there is almost always something new that can be learned and new ways to improve the retrievals. In addition the aircraft measurements themselves are not perfect as the authors point out, with limited measurement altitudes and possible inhomogeneities. Please try to limit use of the word "validation" or a variant to five or fewer instances throughout the entire paper. I include some suggestions for rephrasing in the technical comments, but I do not have a technical comment for each instance.
Response: We revised the title (see response to the comment #S1) following the reviewer's comment #S1. And by addressing the specific comments and technical comments regarding the use of the term "validation", we have substantially reduced the use of the term "validation" and its variant (e.g., "validated", "validating"). Please see responses to the comments # S28, T2, T25, T27, T28, T38. In addition, we also tried to use terms such as "agreement" and "comparison" to replace "validation" when applicable, in our description and discussion of our own results in the manuscript. Note that we did not change the usage of "validation" when we refer to some previous studies because "validation" and "evaluation" are more effective to describe these studies than a substitute, for example "*the primary goal of DISCOVER-AQ was to provide aircraft observation methodologies for satellite validation*". Please see the updated manuscript for details.

G2: Sometimes it is not always clear when all data from the listed flights are used in a comparison and when it is just urban areas like the title implies. Please clarify throughout (especially Sect. 4) if comparisons are from just urban or if they include both urban and non-urban.
Response: Thank you for pointing this out. In Section 3 we analyze "all profiles", "profiles over urban regions", and "profiles over non-urban regions". While in Section 4, we use all the in-situ profiles listed in the Table 1 regardless of if the profiles are over urban or non-urban regions. We explicitly added the following statement at the beginning of Section 4:
"*While in Section 3 we compared profiles over urban regions, and profiles over non-urban regions separately to MOPITT V8T, V8N, V8J, V7T, V7N, and V7J, in this section, we compare only the MOPITT V8J product to all the in-situ profiles (both over urban and non-urban regions) described in Table 1 to test the sensitivity of comparison results to the assumptions made during the comparisons.*"
We separated Section 3.2 (Discussions on individual campaigns) into 5 paragraphs so that each paragraph discusses overall conclusions or comparisons with one specific campaign.
In addition, we explicitly listed what campaign data were used for each statement in Section 4.

Specific comments:

S1 - p1| 1: Suggest the title be changed to be more descriptive, e.g., “Comparisons of MOPITT carbon monoxide retrievals with aircraft measurements, focusing on urban regions”

Response: We changed the manuscript title to
“Assessing MOPITT carbon monoxide retrievals over urban versus non-urban regions”.

S2 - p1| 21: list date range of campaigns (2011-2016)

Response: We added date range for the campaigns “*DISCOVER-AQ (2011-2014), SEAC4RS (2013), ARIAs (2016), A-FORCE (2009, 2013), and KORUS-AQ (2016)*”.

S3 - p1| 22: Please specify biases here refer to both urban and non-urban.

Response: We changed the sentence “*Overall, MOPITT performs reasonably well over both urban and non-urban regions, overall biases for V8J and V8T vary from -0.7% to 0.0%, and from 2.0% to 3.5%, respectively*.”

to

“*In general, MOPITT agrees reasonably well with the in-situ profiles, over both urban and non-urban regions. Version 8 multispectral product (V8J) biases vary from -0.7% to 0.0% and version 8 thermal-infrared product (TIR) biases vary from 2.0% to 3.5%.*”

S4 - p1| 22: Why is V8N disregarded in the abstract? Low DFS?

Response: Thank you for pointing this out. MOPITT V8N product has relatively lower DFS (Figure 3), and are not as widely used as the V8J and V8T. Therefore, V8N is not the main focus of this study and it is only analyzed in Table 2.

S5 - p2| 30: specify the other levels

Response: We added “(e.g., 600 hPa)” in the sentence.

S6 - p2| 48: Surely MOPITT itself is not doing the CO retrievals, but rather a team at NCAR? Suggest updating to: “Observations from the Measurements…satellite have been used for retrieving…”

Response: We updated the sentence accordingly.

S7 - p2| 51-52: Similar to last comment, I think MOPITT just makes the measurement and NCAR provides the product. Suggest update to: …products, a multispectral TIRNIR product is also produced, which…

Response: We changed “*MOPITT also provides the multispectral TIR-NIR product*” to “*the MOPITT multispectral TIR-NIR product is also provided*”.

S8 - p4| 109: I think it would help to update to “quantities in the state vector.”

Response: We updated the sentence accordingly.

S9 - p5| 142: Similar to S6 & S7 – “There are 121 profiles over four urban regions from DISCOVER-AQ.”

Response: We updated the sentence accordingly.

S10 - p5| 145: “… campaign obtained 45 profiles in total sampled over...”

Response: We revised the sentence.

S11 - p6| 161: I'm curious, why not just take the aircraft data as high as it goes and then use the model for the rest? Why include more interpolation with a Pinterp parameter than needed?

Response: Thank you. We did take the aircraft data as high as it goes, and use the reanalysis data for the rest. P_{interp} is used because the rest of the pressure levels need to be filled differently. For pressure levels below P_{interp} (lower altitude), values are linearly interpolated using highest-altitude aircraft measurement and reanalysis data at P_{interp}. For pressure levels above P_{interp} (higher altitude), reanalysis data are used directly. If we use reanalysis data to fill all the levels directly, the extended vertical profile may not be continuous at the highest-altitude the aircraft profile. In addition, the use of P_{interp} allows us to test the sensitivity to the use of model or reanalysis data as because parameter P_{interp} controls the impact of the model-based profile extension on the shape and value of in-situ profiles (see Figure S5).

S12 - p6| 169: This is a little circular, comparing MOPITT retrievals with data that assimilated another version of MOPITT retrievals. It would be helpful to let readers know here that you do not compare with these higher levels later, and that they are expected to have a minimal impact on the lower levels you use in the comparison.

Response: We thank the reviewer for pointing this out. We added the following statement in the text:

"*We note that as we do not compare with these higher levels later, the use of CAMS reanalysis is expected to have a minimal impact on the lower levels we use in the comparison (e.g., surface, 800 hPa, and 600 hPa).*"

S13 - p6| 173: Is this mass weighted? If not add in "unweighted averaging"

Response: We added the adjective "unweighted" before "averaging".

S14 – p7| 176: Please clarify, are these MOPITT profiles with the center point in the radius, or the entire footprint in the radius?

Response: Thank you. It is the center point that needs to be in the radius. We revised the sentence to:

"*MOPITT profiles are considered co-located with the aircraft profile and are selected for comparison only if their center points are within the radius of 100 km and within 12 hours of the acquisition of the aircraft profile.*"

S15 - p7| 181-183: Please clarify that these are not the profiles, but rather the state vectors. (You could also remind readers the state vectors and profiles are related by log10).

Response: As shown in the following revised text, we have re-emphasized that all CO profiles ($x_{in-situ}$, x_a, and $x_{transformed}$) appearing in Equation (2) are expressed in terms of $\log_{10}$(VMR).

"*For each pair of co-located MOPITT retrieved and in-situ profiles, we apply the MOPITT a priori profile and averaging kernel to the in-situ profile as in Eq. (1). Thus, after converting from profiles of the in-situ and a priori CO concentrations to $\log_{10}$(VMR) profiles ($x_{in-situ}$ and x_a,), we calculate*

$$x_{transformed} = x_a + A(x_{in-situ} - x_a) \qquad (2)$$

so that the $\log_{10}$(VMR)-based transformed in-situ profile ($x_{transformed}$) has the same degree of *smoothing and a priori dependence as the MOPITT retrieved log_{10}(VMR) profile (x_{rtv}).*"

S16 – p7| 197: Specify, what does "uniformly weighted" mean here? In pressure? For MOPITT, isn't the surface level an exception to "uniformly"?
Response: "uniformly weighted" is about the way that the MOPITT retrieval algorithm internally converts from 'retrieval-grid' CO profiles (10 levels/layers) to higher-resolution 'model-grid' CO profiles (35 levels/layers) that are needed for the radiative transfer model. For V5 and later products, the algorithm assumes uniform weighting when it makes this conversion. So, for example, the retrieved CO at 900 hPa represents a layer from 900 to 800 hPa with a constant VMR within that layer. Internally, this means that model levels at 900, 875 and 850 are all assigned the same VMR value. We notice "uniformly weighted" is not relevant in this sentence and hence changed it to "uniform".

S17 – p7| 198: "vertical and horizontal" here is a little confusing (at first I thought it was in km, but realized it is variation in CO). You could reword to "The standard deviation of the original aircraft CO observations in each MOPITT layer are also shown, which is due to horizontal and vertical variability in CO."
Response: We revised the sentence accordingly.

S18 - p7| 201: Numerically, what is the xtransformed and xrtv difference?
Response: We added the numerical value (12.4 ppb) in the sentence.

S19 - p8| 207: Even if it was not the focus, MOPITT has been compared against other observing systems in urban regions prior to this paper. For example, Buchholz (doi: 10.5194/amt-10-1927-2017) compared MOPITT observations with ground-based observations in urban areas including Toronto and Bremen.
Response: Thank you for pointing this out. We added a statement regarding the comparisons with ground-based spectrometric column retrievals, and citation for Buchholz et al. (2017) and Hedelius et al. (2019), in the introduction:
"*In addition, MOPITT products have also been compared with ground-based spectrometric column retrievals (e.g., Buchholz et al., 2017; Hedelius et al., 2019).*"
And for the sentence pointed out by the reviewer *"… that MOPITT has yet to be validated over urban regions*", we added specific description *"… that MOPITT has yet to be validated over urban regions with in-situ observations.*"

S20 - p9| 238: Does this range of percentages include the 900 hPa and 700 hPa layers not shown? Why did you decide not to show these layers? Presumably you already did most the work for their comparison too.
Response: The range of percentages only includes what is shown in the Table 1 (i.e., surface layer, 800 hPa, and 600 hPa). We present results on surface layer, 800 hPa, and 600 hPa in this study to represent three levels, namely the surface level, the top of the PBL (800 hPa), and the free troposphere (600 hPa). We do not show results for 900 hPa and 700 hPa as surface layer, 800 hPa, and 600 hPa are representative, as shown by the in-situ profiles (Figures 2, S1, S3, S4) and the averaging kernels at those layers (Figure 3).

S21 - p9| 239: Why wasn't V8N also included? Low DFS? Coverage over land only?

Response: Please see the reply to the comment S4.

S22 – p9| 239: Consider changing "lower" to "smaller." I initially interpreted "lower" to mean "less than" (or more negative), but I think you mean "closer to zero." Same comment for "higher" on line 241.
Response: We changed "lower" to "smaller", and "higher" to "larger".

S23 - p9| 240: Is "-0.2" supposed to be "-2.0" based on Table 2?
Response: We thank the reviewer for noticing this. We changed "*from -0.2% to -0.8%*" to "*from -0.8% to -2%*"

S24 – p9| 242: It looks like you can omit "generally" here, as it appears to be true for all (unless 700 hPa and 900 hPa are exceptions).
Response: We deleted "generally".

S25 - p9| 244-246: This line is hard to read because of the number density, and the information is already in Table 2. I suggest omitting it completely.
Response: We removed the sentence "*For example, for the V8J product, correlation coefficients over urban regions are 0.53, 0.57, and 0.53 at the surface, 800 hPa, and 600 hPa, respectively, while over non-urban regions, the corresponding correlation coefficients are 0.76, 0.73 and 0.67.*"

S26 – p9| 247: If you specified for T and J, then you could remove "generally" on this line.
Response: We changed the sentence "*We also notice that V8 products generally have higher correlation coefficients with in-situ measurements than V7 over non-urban regions, whereas over urban regions, V8 products generally have lower correlation coefficients than V7.*"
to
"*We also notice that for TIR-NIR and TIR-only products, V8 have higher correlation coefficients with in-situ measurements than V7 over non-urban regions, whereas over urban regions, V8 products have lower correlation coefficients than V7 (except for 600 hPa).*"

S27 – p9| 259: I suggest "at 600 hPa" -> "at the 600-500 hPa layer" (same for line 261). Otherwise it sounds like the comparison is at a specific level.
Response: We changed "at 600 hPa" to "at the 600-hPa layer (i.e., the 100 hPa uniform layer immediately above 600 hPa)", and changed "at the surface" to "at the surface layer (i.e., the uniform layer immediately above the surface)". In addition, we changed 600 hPa, 800 hPa, and the surface to the 600-hPa layer, the 800-hPa layer, and the surface level throughout the paper to indicate the comparisons are not for a specific pressure level.

S28 – p11| 295: Consider rewording "this validation of MOPITT at higher CO concentrations..." which sounds like it is the process does not work as well, rather than the results being further off.
Response: We changed "*this validation of MOPITT at higher CO concentrations*" to "*the agreement between MOPITT and the in-situ profiles at higher CO concentrations*".

S29 – p12| 335: Quantitatively how much larger are the "larger biases"?
Response: We changed the sentence "*The validation results using 100 hPa as Pinterp have larger biases.*" to "*The validation results against DISCOVER-AQ CA using 100 hPa or 200 hPa as*

Pinterp have larger biases at the 600-hPa layer (~25%)."

S30 – p12| 339: This is repetitive with a sentence a few lines up (line 329). Can you just separate out a paragraph for DISCOVER-AQ CA so you do not have to mention it 6 times?
Response: We deleted this sentence "*As mentioned in Section 3.2, the DISCOVER-AQ CA aircraft measurements concentrate below 600 hPa, so CO values in the in-situ profiles at 600 hPa and above are filled with and are more sensitive to CAMS reanalysis data.*" Please also see the response to the comment #G2.

S31 - p12| 344: "Previous MOPITT evaluation results," are these previous studies? Could you cite a few examples?
Response: We added Deeter et al. (2012) and Deeter et al. (2016) as examples.

S32 - p13| 376: Does this section and 4.4 use both urban and non-urban observations?
Response: Yes. We added the following statement at the beginning of Section 4:
"*In Section 3, we compared profiles over urban and non-urban regions separately to MOPITT V8T, V8N, V8J, V7T, V7N, and V7J. In this section, we compare only the MOPITT V8J product to all the in-situ profiles (both over urban and non-urban regions) described in Table 1 to test the sensitivity of results to the assumptions made during the comparisons.*"

S33 - p13| 380: Please be quantitative with "long enough lifetime" and include a reference.
Response: The typical lifetime of CO is approximately a month. We added Gamnitzer et al. (2006) as reference.

S34 - p14| 394: What is the L3 grid size?
Response: The resolution is 1°×1°, we added this in the text.

S35 – p15| 422: "MOPITT biases" -> "MOPITT mean biases"
Response: Thank you. We have revised accordingly.

S36 – p15| 422: Please provide a reference for "10% required accuracy"
Response: We added the following reference:
Drummond, J. R., & Mand, G. S. (1996). The Measurements of Pollution in the Troposphere (MOPITT) instrument: Overall performance and calibration requirements. Journal of Atmospheric and Oceanic Technology, 13(2), 314-320.

S37 – p16| 453: Do you know about how many profiles go into each grid cell for Level 3? If it's 1x1 degrees then a 100 km radius is larger. In this case the overall agreement may actually be worsened further by too few MOPITT soundings (if this is what you mean by "which is unlikely to happen when generating L3").
Response: We thank the reviewer for bringing this issue up.
As described in the MOPITT Version 8 Product User's Guide (https://www2.acom.ucar.edu/sites/default/files/mopitt/v8_users_guide_201812.pdf), MOPITT Level 2 data all feed in the specific filtering rules (both pixel filtering and signal-to-noise ratio). Data Averaging to generate MOPITT Level 3 data is performed on a one-degree latitude/longitude grid (1°×1°). The original reason for implying which is unlikely to happen when generating Level

3 is that the daily MOPITT L2 data size is large, and globally speaking there will be enormous data. However, it is true that 1°×1° pixel size is smaller than 100 km radius. And as the reviewer points out, for each individual 1°×1° grid, we should not expect to have much more data to perform the filter and averaging. So we have deleted the statement in the Section 4.4 "However, when generating Level 3 data from Level 2 data, the circumstance is different as there are usually much more data to perform the filter and averaging." We also deleted the statement "which is unlikely to happen when generating Level 3 data" in the Section 5.

However, we disagree that the overall agreement may actually be worsened further by too few MOPITT soundings in MOPITT Level 3 data. We notice that some of the discussions in the manuscript may be misleading. For example, the statement "*In some cases, applying the SNR filters degrades the validation results (e.g., DISCOVER-AQ DC at the surface, DISCOVER-AQ CA at the surface, KORUS-AQ at 600 hPa, and ARIAs at the surface, 800 hPa, and 600 hPa)"* in the section 4.4 to avoid confusion" indicate that applying SNR filter may worsen the agreement in some cases. While this statement is true, it is misleading because the readers may ignore the fact that applying SNR filter also improve the agreement in some cases (e.g., DISCOVER-AQ DC at the 600-hPa layer and the DISCOVER-AQ TX at the 600-hPa layer, and that the overall agreement does not change significantly. Therefore, we delete the aforementioned statement. In addition, we change also changed the statement "*We find that applying the SNR filters does not improve the overall agreement between MOPITT retrievals and the in-situ profiles*" in Section 4.4 to "*We find that applying the SNR filters does not significantly change the overall agreement between MOPITT retrievals and the in-situ profiles used in this study.*" We also changed the statement in the section 5 *"Applying SNR filters does not necessarily improve the overall agreement between MOPITT retrievals and in-situ profiles used in this study*" to "*Applying SNR filters does not necessarily change the overall agreement between MOPITT retrievals and in-situ profiles used in this study significantly*".

In addition, even though applying SNR filter when generating Level 3 data does not significantly change the agreement with the in-situ profiles used in this study, by excluding low-SNR observations from the Level 3 cell-averaged values raises overall mean DFS values (MOPITT Version 8 Product User's Guide, 2018). In addition, the Level 3 product typically are less affected by random retrieval errors (e.g., due to instrument noise or geophysical noise). We have added this statement at the end of section 4.4.

Note that we are not suggesting the comparisons between MOPITT Level 3 product and aircraft measurements. Because the MOPITT Level 3 product is gridded data and represent the average value in a 1°×1° grid. Comparing the grid average value to an aircraft profile within it may be subject to large representativeness errors. Here we only show the sensitivity of agreement between MOPITT Level 2 data and aircraft profiles to the application SNR filter. This statement is also added to the section 4.4 of the manuscript. More discussion on the issue of representativeness errors can be found in section 5.

S38 – p16| 463-468: These 2 sentences are very late in the paper. They should be earlier, like in Section 2.2.

Response: We moved the two sentences to section 2.3 where we discuss the sub-grid variability and representativeness error in Figure 2.

S39 – p16| 468-476: This discussion on NO2 variability from GeoTASO to try to constrain CO variability seems irrelevant and late in the paper. I think the whole thing should be omitted.

Response: We think this discussion is highly relevant to the future direction of this study and other comparisons that have issues with urban variability for satellite spatial resolution. Nevertheless, we addressed the reviewer's comment by reducing the discussion on GeoTASO substantially to one sentence:
"*One possible way is to study NO_2 data retrieved from the Geostationary Trace Gas and Aerosol Sensor Optimization (GeoTASO) at very high resolution (250 m×250 m), to provide an upper estimate on CO variability*".

S40 – p17| 480: All the references to data should be split out into a "Data availability" section. See "Manuscript Composition" here: https://publications.copernicus.org/for_authors/manuscript_preparation.html.
Response: Thank you. We added a Data availability section and moved the relevant part from acknowledgement to it.

S41 – p17| 480: Best practice is for all data to be in a public repository. If this is not possible, then please provide contact information for how the aircraft data can be obtained including ARIAs and A-FORCE.
Response: We added contact info for ARIAs and A-FORCE in the Data availability section.

S42 – p17| 480: Include a last access date with all URLs.
Response: We included last access dates for all the URLs in the data availability section.

S43 – p17| 481-482: These seem like 2 references to the same MOPITT data? Which one should readers use?
Response: MOPITT data are available at both URLs. To avoid confusion, we deleted the second one "*https://earthdata.nasa.gov/ (Last access date: Jan 14th, 2020)*."

S44 – p17| 493: An "Author Contribution" section is needed: https://publications.copernicus.org/for_authors/manuscript_preparation.html.
Response: Thank you for pointing this out. We added an author contribution section after the data availability section.

S45 – p24 Table1: Could you please include the accuracy of CO from aircraft measurements somewhere?
Response: We added "Uncertainty" that includes precision and/or accuracy for each instrument in the Table 1.

S46 – p24| Table1: For better traceability, please list the CO scale the aircraft measurements were tied to (e.g., WMO-CO-X2004, WMOCO-X2014, WMO-CO-X2014A, CSIRO...).
Response: We agree this traceability is preferred, however, we were only able to confirm this for ARIAs (WMO-CO-X2014A), KORUS-AQ (WMO-CO-X2014A), SEAC^{4}RS (WMO-CO-X2004), and DISCOVER-AQ MD, TX, and CA (WMO-CO-X2004). We added the available information to Table 1. However, the precisions/accuracies in Table 1 are based on the referenced literature for the aircraft measurements, which should be sufficient to describe the data we used.

S47 – p26: (No response requested) – Figure 1 is well done and has high information content

Response: Thank you.

S48 – p27| 776-777: I don't understand "vertical and horizontal variability" here. Does "horizontal" somehow correspond to how many km were flown? Or are these just the standard deviations of aircraft profiles. If so, then just say "are the standard deviations of the original aircraft observations."
Response: Thank you. We changed "*vertical and horizontal variability*" to "*the standard deviations of the original aircraft observations*".

S49 – p27| 777-778: I guess this is why the black and orange traces do not always match. Optionally consider plotting at layer centers (shifting up by about 50 hPa).
Response: Thank you. Ploting the $x_{rtv}, x_a, x_{in-situ}, and\ x_{transformed}$ on the surface, 900 hPa, 800 hPa etc layers is consistent with the naming in the main text as well as other figures and tables. We also have mentioned that "each MOPITT retrieval level corresponds to a uniform layer immediately above that level" in the figure caption as well as the main text.

S50 – p29| 793: Please define Delta log(VMR) here and explicitly include the base of the logarithm.
Response: Thank you. $\Delta\log_{10}$(VMR) is defined as x_{rtv} - x_a for MOPITT profiles and $x_{transformed}$-x_a for the in-situ profiles. The use of $\Delta\log_{10}$(VMR) allows us to remove the impact of the a priori in the comparisons. We added this statement in the caption. We also added the base of the log (i.e., 10) here as well as in a few places in the main text.

S51 – p30| 803: Should "Figure 2" be "Figure 4"?
Response: Yes. Thank you for noticing this. We changed "Figure 2" to "Figure 4".

S52 – p31| 810: Numerically, what are considered "outliers"? Please also add to captions of Figures 8-11. Or just reference the caption the Figure 6 so it is less repetitive.
Response: An outlier is a value that is more than 1.5 times the interquartile range away from the top or bottom of the box. We added this statement in the caption of Figure 6, and referred to it in the caption of Figures 8-11.

S53 – p33| Figure8: Are these MOPITT biases compared with aircraft still? So the "200 hPa" values are the same as yellow values in Figure 6? Please specify or optionally consider showing as a % bias compared with the baseline "200 hPa" results.
Response: The biases are calculated against all (both urban and non-urban) in-situ profiles listed in Table 1. We have added this statement in the captions of Figures 8-11. In addition, the "200 hPa" values (gray) in Figure 8 are the same as yellow values (for all data) in Figure 6; the "100 km" values (gray) in Figure 9 are the same as yellow values (for all data) in Figure 6; the "12 h" values (gray) in Figure 10 are the same as yellow values (for all data) in Figure 6; and the "without SNR filter" values in Figure 11 are the same as yellow values (for all data) in Figure 6. We added this information in the corresponding captions too.

S54 – p33| Figure8: Please clarify that you are using "all" observations (both urban and non-urban).
Response: See reply to the comment # S53.

S55 – p34 Figure9: Optionally consider comparing against 100 km.
Response: See reply to the comment # S53.

S56 – p35 Figure10: Optionally consider comparing against 12 hours.
Response: See reply to the comment # S53.

S57 – Figure S3: It would help to remind readers that "in situ" is a combination of aircraft and models since values at 1050 hPa do not make much physical sense. (As a side observation it is interesting that MOPITT is so insensitive).
Response: We added "*in-situ profiles (combination of aircraft and reanalysis data as described in Section 2.3)*". in the caption of Figure S3.

S58 – Figure S5: Could you please provide more detail in the caption? Consider marking the level of highest aircraft measurements (presumably this is why there are straight lines).
Response: We extended the caption of Figure S5 to include more details:
"***Figure S5****. Averaged in-situ profiles (combination of aircraft and reanalysis data as described in Section 2.3) under different assumptions of* P_{interp}*. For pressure levels below* P_{interp} *(lower altitude), values are linearly interpolated using the highest-altitude aircraft measurement and reanalysis data at* P_{interp}*. For pressure levels above* P_{interp} *(higher altitude), reanalysis data are used directly. For* P_{interp} *equals 100 hPa, 200 hPa, 300 hPa, 400 hPa, and 500 hPa, the corresponding averaged in-situ profiles are shown by the blue, gray, yellow, green, and red lines, respectively. Taking the* P_{interp} *equal to 100 hPa (blue line) as an example: for pressure levels below 100 hPa but above the highest-altitude aircraft measurement, the CO values are filled by linearly interpolation between CO values at the highest-altitude aircraft measurement and reanalysis data at 100 hPa; for levels above 100 hPa, CO values from reanalysis data are added to the in-situ profile directly.*"

Technical comments:

T1 - p1| 17: "The performance of the" could be omitted (and update has -> have)
Response: We have revised accordingly.

T2 - p1| 19-20: E.g., validate -> compare, using -> with
Response: We have revised accordingly.

T3 - p1| 25: suggest "performance" -> "agreement" and adding "with aircraft measurements" after V8T
Response: We have revised accordingly.

T4 - p2| 32: "allowed maximum" -> "maximum allowed" and "as criteria" -> "criterion"
Response: We have revised accordingly.

T5 - p2| 34-35: suggest "hence few MOPITT retrievals are included in the comparison."
Response: We have revised accordingly.

T6 - p2| 36: "overall smaller" -> "smaller overall"
Response: We have revised accordingly.

T7 - p2| 40: "retrievals that result for comparison." -> "retrievals for the comparison."
Response: We have revised accordingly.

T8 - p3| 58: "the most recently"
Response: We have revised accordingly.

T9 - p3| 83-84: suggest "...we compare MOPITT version...regions with aircraft profiles made over..."
Response: We have revised accordingly.

T10 - p3| 83: "version" is lowercase here but capitalized on p2| 55. Please be consistent throughout.
Response: Thank you. We changed "version" to be capitalized throughout the manuscript.

T11 - p4| 100: "retrievals" -> "observations"
Response: We have revised accordingly.

T12 - p4| 111: "The two" -> "These two"
Response: We have revised accordingly.

T13 - p5| 124: "determined" -> "considered"
Response: We have revised accordingly.

T14 - p5| 128-130: move/modify "the profiles over urban and non-urban areas are similar" to right after "We also notice..."
Response: The sentence "*We also notice for aircraft profiles sampled during KORUS-AQ, even though the averaged profile over urban regions has slightly higher CO concentration near the surface, the profiles over urban and non-urban are close.*" was revised to "*For aircraft profiles sampled during KORUS-AQ, the CO profiles over urban and non-urban regions are similar, even though the averaged profile over urban regions has slightly higher CO concentration near the surface.*"

T15 - p5| 139: omit "different instruments" (it's implied)
Response: We have revised accordingly.

T16 - p6| 148: "Only few" -> "Few"
Response: We have revised accordingly.

T17 - p6| 162: omit "below" (I initially was confused because I thought "below" meant lower pressure/higher altitude)
Response: Thank you for pointing this out. We omitted "below".

T18 - p6| 174-175: "We have investigated the..."
Response: We changed the sentence "*We have conducted further calculations to investigate the sensitivity of validation results to P_{interp} in Section 4.1.*" to "*We investigate the sensitivity of validation results to P_{interp} in Section 4.1*."
T19 - p7| 179: "have been further" -> "are"
Response: We have revised accordingly.

T20 - p7| 186: "If fewer than five MOPITT retrievals are co-located with an in-situ profile, the..."
Response: We have revised accordingly.

T21 - p7| 187-191: I think it would be clearer if you reordered the description. a) In situ profile individually applied to AK and prior of each MOPITT retrieval to get xtranformed. b) xtransform averaged as log10. c) Corresponding MOPITT profile retrievals also averaged
Response: Thank you for the suggestion. We have rephrased these sentences "*If an in-situ profile is co-located with five or more MOPITT retrievals, these co-located MOPITT profiles are averaged as log_{10}(VMR). These transformed in-situ profiles that are generated from the same in-situ profile are also averaged. Applying these corresponding different MOPITT a priori profiles and averaging kernels to the same in-situ profile results in different transformed in-situ profiles. These transformed in-situ profiles that are generated from the same in-situ profile are also averaged.*"
to
"*If an in-situ profile is co-located with five or more MOPITT retrievals (assume the number to be $N_{retrieval}$), then the following steps are used in the comparison with MOPITT: (a) the averaging kernel and a prior of each co-located MOPITT retrieval are applied to the in-situ profile (through equation 2) to obtain $N_{retrieval}$ of $x_{transformed}$. Note that applying these $N_{retrieval}$ sets of MOPITT a priori profiles and averaging kernels to the same in-situ profile results in differently transformed in-situ profiles; (b) the $N_{retrieval}$ of $x_{transformed}$ are averaged in log_{10}(VMR) space; and (c) the $N_{retrieval}$ of MOPITT retrievals x_{rtv} are also averaged.*"

T22 - p7| 200: "variability" -> "standard deviation"
Response: We have revised accordingly.

T23 - p7| 202 & 203: omit "retrieval" (the size does not depend on the retrieval algorithm, but is inherent in the MOPITT observation system)
Response: We have deleted "retrieval".

T24 - p8| 205: omit "very"
Response: We have deleted "very".

T25 - p8| 207: "validated" -> "compared with aircraft observations"
Response: We have revised accordingly.

T26 - p8| 209-210: "...show a sensitivity analysis in Section 4 to provide..."
Response: We have revised accordingly.

T27 - p8| 211: omit "validation"
Response: We have revised accordingly.

T28 - p8| 215: maybe "validation" -> "comparison with aircraft profiles"
Response: We changed the section title to "*3. MOPITT comparisons with aircraft profiles over urban and non-urban regions*"

T29 - p8| 224: "against observations" -> "against in situ observations"
Response: We have revised accordingly.

T30 - p9| 236: "Corresponding results" -> "These comparisons"
Response: We have revised accordingly.

T31 - p9| 244: "three levels" -> "three levels in Table 2"
Response: We have revised accordingly.

T32 – p9| 251: "...in terms of higher correlation coefficients and smaller biases..."
Response: Thank you. We have revised accordingly.

T33 - p9| 253: "provide" -> "evaluate", "evaluation against" -> "retrievals during", "campaigns" -> "campaigns with results"
Response: We have revised accordingly.

T34 - p9| 264: "in more favorable weather conditions" -> "during times with greater vertical mixing"
Response: We have revised accordingly.

T35 - p9| 271: "xin-situ, the" -> "xin-situ over non-urban areas, the"
Response: We have revised accordingly.

T36 - p9| 291: "concentrations all" -> "concentration for all"
Response: We have revised accordingly.

T37 - p9| 292: "For both" -> "For the higher 50% of measured mixing ratios both" and omit "if only the upper 50% of measured mixing ratios are considered"
Response: We have changed the sentence accordingly.

T38 - p12| 335-336: Suggest omitting "The validation results", and the second "validation" and changing "are" -> "does", "different for" -> change
Response: We have changed the sentence "*The validation results using 300, 400, or 500 hPa as P_{interp} are not significantly different for the validation results against DISCOVER-AQ CA.*"
to
"*Using 300, 400, or 500 hPa as P_{interp} does not significantly change the results against DISCOVER-AQ CA.*"

T39 - p12| 347: “the radius” -> “a radius”
Response: We have revised accordingly.

T40 - p12| 350: “close” -> “similar”
Response: We have revised accordingly.

T41 - p13| 356: “a smaller number of included” -> “including fewer”
Response: We have revised accordingly.

T42 - p13| 359: “a a more more” -> “a more”
Response: We have revised accordingly.

T43 - p13| 367: “especially” -> “including”
Response: We have revised accordingly.

T44 - p14| 399: “Level 3” -> “the Level 3”
Response: We have revised accordingly.

T45 - p14| 411: omit “process”
Response: We have revised accordingly.

T46 – p15| 423: suggest “overall” -> “mean”
Response: We have revised accordingly.

T47 – p15| 424: “to 3.5% for different levels”
Response: We have revised accordingly.

T48 – p15| 429: “to” -> “into”
Response: We have revised accordingly.

T49 – p15| 431: “compared with low”
Response: We have revised accordingly.

T50 – p15| 440: “as co-location criteria” -> “as a co-location criterion”
Response: We have revised accordingly.

T51 – p15| 441: “where a” -> “where only a”
Response: We have revised accordingly.

Anonymous Referee #2

This manuscript by Wenfu Tang et al presented a comparison of the latest MOPITT CO V8 retrievals with aircraft measurements from DISCOVER-AQ, SEAC4RS, ARIAs, A-FORCE, and KORUS-AQ campaigns conducted over the US or East Asia. In addition, the sensitivities of validation results to assumptions and data filters applied during the comparisons of MOPITT retrievals and in-situ profiles were also performed and analyzed. The comparison between the MOPITT CO product with various version and the coincident observations has been previously performed by many scientists in many groups around the world. This study is an extension of previous study and the strategy for comparison has been used extensively in previous MOPITT evaluation and validation studies. However, this study is one of few studies that focus on comparison over around urban regions, this is interesting. Overall, this paper is well written and fits well within the scope of AMT. I recommend for publication though I rate the novelty of this paper as moderate. Since referee # 1 has listed numerous technical comments which are mostly overlapped with my comments. Here I don't present the repeated correction request.
Response: Thank you for your time and effort in reviewing our manuscript. We have addressed the comments accordingly. Please see below for details.

Extra minor revisions or comments are:

1. The Base map and color bar in Fig. 1 can be improved. It is hard to distinguish one from another. In latitude and longitude axis, the number like 30 should be 30◦.
Response: We have changed colormap, color scale, and increased marker size. We also added the symbol for degree (°) in the latitude and longitude. See the updated Figure 1 in the manuscript for details.

2. What does the dashed line in Figs. 4 and 5 mean? The one to one lineïij§Should be stated in the caption.
Response: The dashed lines are one-to-one ratio lines. We added this information in the captions of Figures 4 and 5.

3. If you only compare the results at surface, 800 hPa, and 600 hPa. Then the expression should be the concentrations at surface, 800 hPa, and 600 hPa rather than the profiles at surface, 800 hPa, and 600 hPa.
Response: As described in the Section 2.3, we did compare the 10-level MOPITT profiles to 10-level in-situ profiles. Due to the lack of observations above 600 hPa, we only showed and discussed the results of comparisons below 600 hPa. The surface layer, 800-hPa layer, and the 600-hPa layer are selected to represent different conditions of the profiles below 600 hPa. Please also see the response to the comment # S20 from the reviewer 1. Nevertheless, we thank the reviewer for bringing this up, and changed the term "profile" to "concentration"/"value" when discussing a single layer. For example, we changed "*the overall agreements between values of MOPITT and in-situ profiles at the 800-hPa layer*" to "*the overall agreements between MOPITT and in-situ profiles at the 800-hPa layer*" in the section 4.1 to emphasize this statement is only for one layer.

4. Another confusing thing is that the MOPITT could have a very low DOFS at a given level with a limited range (Fig. 3). Thus, the retrieval should come more from a priori information rather than

the measurement. In other words, I guess, the good agreement between the MOPITT and aircraft at a given level is largely attributed to the a priori information and the smoothing effect in equation 2.

Response: The MOPITT V8N product does have a lower degree of freedom for signal compared to the MOPITT V8T and V8J products. Note that this manuscript mainly focuses on the V8T and V8J products (see the reply to the comment S4 of the reviewer 1). It is true that applying MOPITT AK and a priori (the smoothing effect in equation 2) to in-situ profile would reduce the difference between MOPITT profile and the in-situ profiles. However, this is the only correct way to perform such comparison. As stated by the MOPITT Version 8 Product User's Guide (available online at https://www2.acom.ucar.edu/sites/default/files/mopitt/v8_users_guide_201812.pdf), because of the dependence of MOPITT on the a priori information, users must transform these comparison datasets using the equation 2, so that the comparison data exhibit the same degree of smoothing and a priori dependence as the MOPITT product. We are aware of the impact of the a priori information in the retrievals. However, as described in Section 3.1, we explicitly removed the a priori information in the validation process following the method described in Deeter et al. (2017). Therefore, the good agreement between the MOPITT and aircraft at a given level is not largely attributed to the a priori information. In fact, the agreement would be much better than it is now if we did not remove the a priori information in the validation process.

Anonymous Referee #3

General comments:
The aim of this paper is to evaluate two versions of MOPITT CO (V7 and V8) by comparison with aircraft observations from diverse campaigns all over the globe. Each version has two sub versions (V7-8T, V7-8N, V7-8J for thermal, NIR and TIR+NIR, respectively). Urban and non urban areas are the focus of the evaluation. This is a paper that complete the list of publications of the evaluation of the different versions of MOPITT CO. Lots of statistics are provided and the MOPITT users community could find some interest in order to interpret MOPITT data over urban areas.
Response: Thank you for your time and effort in reviewing our manuscript.

However, I found the comparison sometimes difficult to follow because of the large number of campaigns, the number of aircraft profiles by campaign, number of aircraft profiles over urban regions, the number of MOPITT CO profiles in different circles,.. Table 1 helps but if possible it would be nice to simplify in the text. Moreover, the title does not reflect totally the subject of the paper: the validation of the MOPITT CO retrievals is also over non urban regions. I suggest to change the title in that way.
Response: Please see the responses to the comments # G2 and S1 of the reviewer 1.

1) Moreover, the distinction of urban and non urban regions for the comparison of MOPITT CO with aircraft observations could mislead the reader. What is important in this study, is it the carbone monoxide emitted from the urban region or just the urban region with surface parameters different from non urban regions? Such surface parameters that are used in the retrievals of MOPITT CO (surface temperature, emissivity). At 600 hPa, some comparisons are done but this is above the boundary layer. There is a great chance that the CO measured by both MOPITT and the aircraft is transported from other regions that are not representative of urban regions. The author should clarify this point.
Response: We thank the reviewer for bringing this question up. The urban regions often have different surface parameters (e.g., surface temperature and emissivity), and usually but not always have higher CO concentrations than non-urban regions. However, the surface parameters are unlikely to impact the ultimate quality of MOPITT retrieval products (Pan et al., 1998; Ho et al., 2005). The goal of this study is to understand if MOPITT retrievals are able to represent conditions over urban regions given sampling, and cloud cover. In addition, the relatively large spatial and temporal variability of CO concentrations over urban regions makes the validation even more complex. Because of the complexity of urban regions and their connection with non-urban regions nearby, we also provide analysis at high CO concentrations regardless of landcover type. As the reviewer pointed out, the comparisons are done for the 600-hPa layer (usually in the free troposphere). It is possible that CO concentrations at this layer are transported from other regions that are not representative of urban regions. Even so, MOPITT retrievals at the 600-hPa layer are still impacted by the CO concentrations at other layers including the surface layer (equation 1). Therefore, the comparisons at 600 hPa is necessary. We have added the discussions above to the section 2.2. See the manuscript for details.

2) Also, it would be nice to have a clear recommandation on which MOPITT CO version to use. For example, after reading the table 2 of the paper, I found difficult to conclude on which version to use for urban or a non urban study as well. The statistics are often very similar and I was wondering what is the added value of V8 vs V7 and how significant the values are? It would be nice the authors discuss this point and conclude with clear recommendation in the conclusions on the use of the different versions of MOPITT CO.

Response: We thank the reviewer for pointing this out. The main goal of this study is not to compare MOPITT V8 and V7 products, but rather to validate the performance of MOPITT products over urban regions versus non-urban regions. The finding is that in general, MOPITT agrees reasonably well with the in-situ profiles over both urban and non-urban regions. As the reviewer pointed out, the statistics are often very similar, therefore we do not have recommendation for which version to use in terms of urban versus non-urban regions.

The MOPITT TIR-only, and TIR-NIR products both have their own advantages and disadvantages. MOPITT TIR-NIR products usually have higher DFSs and have enhanced the sensitivity to near-surface CO but may have larger retrieval noise compared to the TIR-only products (Deeter et al., 2011, 2013; Worden et al., 2010). The MOPITT V8 uses a new parameterized radiance bias correction method to minimize retrieval biases, therefore in general the MOPITT V8 performs better than V7 and is recommended (Deeter et al., 2019). A detailed description of MOPITT V8 products and their comparisons to MOPITT V7 products can be found in Deeter et al. (2019). We added the discussion below to the section 5 of the manuscript:

"*The statistics are often very similar between different versions and products over urban and non-urban regions, and in general, MOPITT agrees reasonably well with the in-situ profiles in both cases. There is not, therefore, any reason to recommend the continued use of MOPITT versions earlier than V8 based on urban or non-urban region considerations. In general, MOPITT V8 is recommended (Deeter et al., 2019) as it uses a new parameterized radiance bias correction method to minimize retrieval biases, and has updated spectroscopic data for water vapor and nitrogen.*"

3) The Section 4.4 (Sensitivity to the signal-to-noise ratio (SNR) filters) is unclear to me. What are the conclusions we can draw from this section? Is level 3 useless? I didn't catch the point of this section. Maybe the authors could clarify on how to use Level 3 data over urban and non urban regions in the light of the use of such SNR filter.

Response: Please see the response to the comments # S37 of the reviewer 1.

Specific Comments:

Abstract:

In the paper, V7 and V8 of MOPITT CO are evaluated whereas only V8 is mentioned in the abstract.

Response: MOPITT V7 products is only used as a reference in the sub-Section 3.1 and is not the focus of this study. To avoid the confusion, we changed the sentence "*We focus on evaluating the recently released Version 8, as well as the Version 7, of the MOPITT TIR, NIR, and multispectral*

TIR-NIR products." in the Section 2.1 (MOPITT retrievals and products) to "*We focus on validating the recently released Version 8 of the MOPITT TIR, NIR, and multispectral TIR-NIR products. We also include comparisons with the MOPITT Version 7 TIR, NIR, and multispectral TIR-NIR products in the Section 3.1 for reference.*"

Section 3.3
L 300-301: This means MOPITT CO concentrations are highly variable in circles where true concentrations are high. In this condition, what are the retrieval errors for these MOPITT pixels?
Response: We thank the reviewer for the question. We have conducted the calculation of the retrieval uncertainties, and added the statement below to section 3.3:
"*At higher 50% CO concentrations, the averaged retrieval uncertainties for the 600-hPa, 800-hPa, and surface layers, are 28%, 28%, and 29%, respectively. This is smaller than the averaged retrieval uncertainties at lower 50% CO concentrations (28%, 29%, and 30% for the 600-hPa, 800-hPa, and surface layers, respectively). We therefore conclude that the larger apparent biases at high CO concentrations are related to greater CO variability and representativeness error of the in-situ profile within the co-location radius used for analyzing the MOPITT data, rather than indicating larger retrieval uncertainties. Theoretically, MOPITT retrievals perform better with higher CO concentrations. The larger biases at high CO concentrations in Figure 7 implies that the relatively greater CO variability may overcome the impact of high CO concentrations. Addressing representativeness error and spatial variability in the comparisons between satellite and in-situ profiles is challenging, and will be discussed further in Section 5.*"

L 334: please correct the sentence
Response: The sentence is changed to "*At the 600-hPa layer, the agreements between the values of MOPITT and in-situ profiles are affected more by P_{interp} compared to the those at the surface layer and the 800-hPa layer for comparisons with all the campaigns.*"

L 360: please correct the sentence
Response: We changed the sentence to "*We note that the usage of the largest radius (200 km) in this paper does not appear to degrade the overall results, even though representativeness errors generated from CO spatial and/or temporal variability are expected to increase. However, the use of the smallest radius (25 km) degrades the overall results by reducing the number of included MOPITT retrievals.*"

L 369: The sentence 'we note..' is unclear to me. Please clarify if necessary.
Response: We changed this sentence to
"*We note that when comparing to the ARIAs campaign, using 1h as the allowed maximum time difference decreases the biases at the surface layer, the 800-hPa layer, and the 600-hPa layer, compared to the cases using longer allowed maximum time difference (i.e., 3h, 6h, and 12h). This implies that the temporal variability is relatively large in the region.*"

Assessing MOPITT carbon monoxide retrievals over urban versus non-urban regions

Wenfu Tang[1,2], Helen M. Worden[2], Merritt N. Deeter[2], David P. Edwards[2], Louisa K. Emmons[2], Sara Martínez-Alonso[2], Benjamin Gaubert[2], Rebecca R. Buchholz[2], Glenn S. Diskin[3], Russell R. Dickerson[4], Xinrong Ren[4,5], Hao He[4], and Yutaka Kondo[6]

[1]Advanced Study Program, National Center for Atmospheric Research, Boulder, CO, USA

[2]Atmospheric Chemistry Observations and Modeling, National Center for Atmospheric Research, Boulder, CO, USA

[3]NASA Langley Research Center, Hampton, VA, USA

[4]Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD, USA

[5]Air Resources Laboratory, National Oceanic and Atmospheric Administration, College Park, MD, USA

[6]National Institute of Polar Research, Tachikawa, Japan

Correspondence: Wenfu Tang (wenfut@ucar.edu)

Abstract

The Measurements of Pollution in the Troposphere (MOPITT) retrievals over urban regions have not been validated systematically, even though MOPITT observations are widely used to study CO over urban regions. Here we compare MOPITT products over urban and non-urban regions with aircraft measurements from DISCOVER-AQ (2011-2014), SEAC^{4}RS (2013), ARIAs (2016), A-FORCE (2009; 2013), and KORUS-AQ (2016) campaigns. In general, MOPITT agrees reasonably well with the in-situ profiles, over both urban and non-urban regions. Version 8 multispectral product (V8J) biases vary from -0.7% to 0.0% and version 8 thermal-infrared product (TIR) biases vary from 2.0% to 3.5%. The evaluation statistics of MOPITT V8J and V8T over non-urban regions are better than that over urban regions with smaller biases and higher correlation coefficients. We find that the agreement of MOPITT V8J and V8T with aircraft measurements at high CO concentrations is not as good as that at low CO concentrations, although CO variability may tend to exaggerate retrieval biases in heavily-polluted scenes. We test the

Deleted: Validation of MOPITT Carbon Monoxide (CO) retrievals over urban regions

Deleted: performance of the

Deleted: has

Deleted: validate

Deleted: using

Deleted: Overall, MOPITT performs reasonably well over both urban and non-urban regions, overall biases for V8J and V8T vary from -0.7% to 0.0%, and from 2.0% to 3.5%, respectively.

Deleted: performance

sensitivities of the agreements between MOPITT and in-situ profiles to assumptions and data filters applied during the comparisons of MOPITT retrievals and in-situ profiles. The results at the surface layer are insensitive to the model-based profile extension (required due to aircraft altitude limitations) whereas the results at levels with limited aircraft observations (e.g., the 600-hPa layer) are more sensitive to the model-based profile extension. The results are insensitive to the maximum allowed time difference criterion for co-location (12 hours, 6 hours, 3 hours, and 1 hour), and are generally insensitive to the radius for co-location, except for the case where the radius is small (25 km) and hence few MOPITT retrievals are included in the comparison. Daytime MOPITT products have smaller overall biases than nighttime MOPITT products when comparing both MOPITT daytime and nighttime retrievals to the daytime aircraft observations. However, it would be premature to draw conclusions on the performance of MOPITT nighttime retrievals without nighttime aircraft observations. Applying signal-to-noise ratio (SNR) filters does not necessarily improve the overall agreement between MOPITT retrievals and in-situ profiles, likely due to the reduced number of MOPITT retrievals for comparison. Comparisons of MOPITT retrievals and in-situ profiles over complex urban or polluted regimes are inherently challenging due to spatial and temporal variabilities of CO within MOPITT retrieval pixels (i.e., footprints). We demonstrate that some of the errors are due to CO representativeness with these sensitivity tests, but further quantification of representativeness errors due to CO variability within the MOPITT footprint will require future work.

1. Introduction

Observations from the Measurements of Pollution in the Troposphere (MOPITT) instrument onboard the NASA Terra satellite have been used for retrieving total column amounts and volume mixing ratio (VMR) profiles of carbon monoxide (CO) using both thermal-infrared (TIR) and near-infrared (NIR) measurements since March 2000. Besides the TIR-only and NIR-only products, the MOPITT multispectral TIR-NIR product is also provided, which has enhanced the sensitivity to near-surface CO (Deeter et al., 2011, 2013; Worden et al., 2010). Since the start of the mission, the MOPITT CO retrieval algorithm has been improved and enhanced continuously (Worden et al., 2014). For example, the Version 6 product improvements included the reduction of both a geolocation bias and a significant latitude-dependent retrieval bias in the upper

Deleted: validation results
Deleted: validation
Deleted: allowed maximum
Deleted: as criteria
Deleted: hence the MOPITT retrievals included in the validation become very small
Deleted: smaller
Deleted: that result
Deleted: the
Deleted: that
Deleted: validation
Deleted: T
Deleted: has
Deleted: ,
Deleted: MOPITT also provides

troposphere (Deeter et al., 2014). In the Version 7 products, a new strategy for radiance-bias correction and an improved method for calibrating MOPITT's NIR radiances were included (Deeter et al., 2017). For the most recently released MOPITT Version 8 products, enhancements include a new radiance bias correction method (Deeter et al., 2019). Meanwhile, the MOPITT products have been extensively evaluated and validated with in-situ measurements, though this has been done primarily over non-urban areas (Deeter et al., 2010, 2012, 2013, 2014, 2016, 2017, 2019; Emmons et al., 2004, 2007, 2009). In addition, MOPITT products have also been compared with ground-based spectrometric column retrievals (e.g., Buchholz et al., 2017; Hedelius et al., 2019). For the past two decades, MOPITT CO products have been widely used for various applications, including understanding atmospheric composition, evaluating atmospheric chemistry models, and constraining inverse analyses of CO emissions (e.g., Arellano et al., 2004, 2006, 2007; Chen et al., 2009; Edwards et al., 2006; Emmons et al., 2010; Fortems-Cheiney et al., 2011; Gaubert et al., 2016; Heald et al., 2004; Jiang et al., 2018; Kopacz et al., 2009, 2010; Kumar et al., 2012; Lamarque et al., 2012; Tang et al., 2018; Yurganov et al., 2005).

MOPITT products are particularly useful for monitoring and analyzing air pollution over urban regions because of the enhanced retrieval sensitivity to near-surface CO and the long-term record (e.g., Clerbaux et al., 2008; Girach and Nair, 2014; Jiang et al., 2015, 2018; Kar et al., 2010; Tang et al., 2019; Worden et al., 2010; Li and Liu, 2011; He et al., 2013; Aliyu and Botai, 2018; Kanakidou et al., 2011). However, the performance of MOPITT retrievals over urban regions has not yet been validated systematically. Furthermore, in-situ observations of CO profiles over urban areas are limited, especially in Asia. Indeed, along with the non-urban validation exercises mentioned above, development and validation of the MOPITT retrieval algorithm relies heavily on in-situ measurements over remote regions, such as measurements from the HIAPER Pole-to-Pole Observations (HIPPO) and the Atmospheric Tomography Mission (ATom) campaigns (e.g., Deeter et al., 2013, 2014, 2017, 2019). Comparisons of MOPITT products to measurements with aircraft profiles during the Korea United States Air Quality (KORUS-AQ) campaign over South Korea have only recently been made in Deeter et al. (2019), but without explicitly analyzing MOPITT performance over urban regions.

In this study, we compare MOPITT Version 8 and 7 products with aircraft profiles made over urban regions (as well as non-urban regions) from campaigns including: Deriving Information

Deleted:

Deleted: validate

Deleted: version

Deleted: over urban regions by comparing

Deleted: that are over

on Surface conditions from Column and Vertically Resolved Observations Relevant to Air Quality (DISCOVER-AQ); the Studies of Emissions and Atmospheric Composition, Clouds, and Climate Coupling by Regional Surveys (SEAC[4]RS); the Air Chemistry Research In Asia (ARIAs); the Aerosol Radiative Forcing in East Asia (A-FORCE); and KORUS-AQ. These campaigns are described in Section 2, along with a brief description of the MOPITT products and the methodology used. We present the comparisons of MOPITT products to aircraft profiles, and discuss the impacts of key factors in the retrieval process on the retrieval results in Section 3. In Section 4, we discuss the sensitivities of results to the assumptions and data filters made for aircraft-satellite comparisons not only in this study, but also in previous evaluation studies of MOPITT and other satellite products. Section 5 gives the conclusions of the study.

Deleted: introduced

Deleted: introduction

Deleted: validation

Deleted: validation results

2. Data and methods

2.1 MOPITT retrievals and products

MOPITT is a nadir sounding satellite instrument flying on the NASA Terra satellite. It uses a gas filter correlation radiometer and measures radiance at both the TIR band near 4.7 μm and the NIR band near 2.3 μm. These observations have a spatial resolution of about 22 km × 22 km with satellite overpass time at approximately 10:30 and 22:30 (local time). To determine a unique CO concentration profile from the MOPITT measured radiances, an optimal estimation-based retrieval algorithm, and a fast radiative transfer model are used (Deeter et al., 2003; Edwards et al., 1999). The retrieved state vector (x_{rtv}) for optimal estimation-based retrievals can be expressed as

Deleted: retrievals

$$x_{rtv} = x_a + \boldsymbol{A}(x_{true} - x_a) + \epsilon \quad (1)$$

x_a and x_{true} are the a priori state vector and the true state vector, respectively. $\boldsymbol{A}$ (which has a size of 10×10) is the retrieval averaging kernel matrix (AK) that represents the sensitivity of retrieved profiles to actual profiles and ϵ is the random error vector. Note that CO quantities in the state vector are retrieved as $\log_{10}$(VMR).

Deleted: profiles

Deleted: quantities

We focus on validating the recently released Version 8 of the MOPITT TIR, NIR, and multispectral TIR-NIR products. We also include comparisons with the MOPITT Version 7 TIR, NIR, and multispectral TIR-NIR products in the Section 3.1 for reference. These two versions of MOPITT products were introduced in detail in Deeter et al. (2017) and Deeter et al. (2019).

Deleted: We focus on evaluating the recently released version 8, as well as the version 7, of the MOPITT TIR, NIR, and multispectral TIR-NIR products.

2.2 Aircraft measurements used for comparisons

Aircraft-sampled profiles of CO concentrations during the DISCOVER-AQ, SEAC^{4}RS, ARIAs, A-FORCE, and KORUS-AQ campaigns are used for comparisons with MOPITT-retrieved profiles. DISCOVER-AQ, and SEAC^{4}RS were conducted over the US, while ARIAs, A-FORCE, and KORUS-AQ were conducted over East Asia. Locations of the aircraft profiles from these campaigns are compared with the MODIS (Moderate Resolution Imaging Spectroradiometer) Terra+Aqua Land Cover Type Climate Modeling Grid Yearly Level 3 Version 6 0.05°×0.05° Global product (MCD12C1 v006) (Friedl and Sulla-Menashe, 2015) to determine if a profile was sampled over an urban or non-urban region. Specifically, for each aircraft profile, a 0.5°×0.5° box centered over the location of the aircraft profile (determined by averaged latitude and longitude of aircraft observations in the profile) is selected. If the urban and built-up fraction in the box is larger than 10%, the profile is considered to be an urban profile. Overall, for each campaign, the averaged aircraft profile over urban regions has higher CO concentrations compared to that over non-urban regions, especially near the surface (see Figure S1). Profiles during ARIAs, which are sampled over Hebei province in China, are exceptional, as the averaged profile over non-urban regions has higher CO concentrations especially near the surface, indicating high CO levels in the entire study region. We note that Hebei is one of the most heavily industrialized and polluted regions, and the difference in CO profiles is driven less by urban versus rural than by synoptic and mesoscale meteorology. In addition, Hebei is an arid region and subject to strong nocturnal inversions, so the surface CO can be very high. For aircraft profiles sampled during KORUS-AQ, the CO profiles over urban and non-urban regions are similar, even though the averaged profile over urban regions has slightly higher CO concentration near the surface. This is largely due to the fact that many of the non-urban aircraft profiles are sampled over the Taehwa forest site, which is impacted by CO transported from the nearby Seoul urban region. The urban regions often have different surface parameters (e.g., surface temperature and emissivity), and usually but not always have higher CO concentrations than non-urban regions. However, the surface parameters are unlikely to impact the ultimate quality of MOPITT retrieval products (Pan et al., 1998; Ho et al., 2005). The goal of this study is to understand if MOPITT retrievals are able to represent conditions over urban regions given sampling, and cloud cover. In addition, the relatively large spatial and temporal variability of CO concentrations over urban regions makes the validation even more complex. Because of the complexity of urban regions and their connection with non-urban regions nearby, we also provide

Deleted: (EA)

Deleted: version

Deleted: is

Deleted: s

Deleted: of

Deleted: determined

Deleted: the exception

Deleted: .

Deleted: We also notice f

Deleted: or aircraft profiles sampled during KORUS-AQ, even though the averaged profile over urban regions has slightly higher CO concentration near the surface, the profiles over urban and non-urban are close analysis at high CO concentrations regardless of landcover type. As the reviewer pointed out, the comparisons are done for the 600-hPa layer (usually in the free troposphere). It is possible that CO concentrations at this layer are transported from other regions that are not representative of urban regions. Even so, MOPITT retrievals at the 600-hPa layer are still impacted by the CO concentrations at other layers including the surface layer (equation 1). Therefore, the comparisons at 600 hPa is necessary.

The campaigns and profiles are summarized in the Table 1 and Figure 1. During DISCOVER-AQ, SEAC^{4}RS, and KORUS-AQ, CO concentrations were measured by the NASA Differential Absorption Carbon monOxide Measurement (DACOM), whereas during ARIAs and A-FORCE, CO concentrations were measured by Picarro G2401-m and Aero-Laser GmbH AL5002, respectively. Note that the primary goal of DISCOVER-AQ was to provide aircraft observation methodologies for satellite validation (e.g., Lamsal et al. (2014)). There are 121 profiles over four urban regions from DISCOVER-AQ, making it particularly useful for the goal of this study. Because of this, our results are heavily driven by aircraft profiles from DISCOVER-AQ. Even though there are only two profiles sampled over urban regions, the A-FORCE campaign obtained 45 profiles in total sampled over East Asia during Spring 2009, Winter 2013, and Summer 2013. The seasonal and spatial coverage of the dataset makes it representative of the region. The ARIAs campaign provides 19 profiles and three of these were sampled over Chinese urban regions. Few previous studies have validated MOPITT products over China (e.g., Hedelius et al., 2019), so aircraft profiles from ARIAs have also been included in this study.

Deleted: Urban regions do not always have higher CO concentrations than non-urban regions. Therefore, because of the complexity of urban regions and their connection with non-urban regions nearby, we also provide analysis of validation at high CO concentrations regardless of landcover type.

Deleted: different instruments, a

Deleted: I

Deleted: DISCOVER-AQ provides 121 profiles over four urban regions

Deleted: validation

Deleted: provides in total

Deleted: EA

Deleted: Only

Deleted: f

2.3 Method for comparing MOPITT profiles to aircraft measurements

Deleted: aircraft measurements and

We generally follow the method that has been used in previous MOPITT evaluation and validation studies (Deeter et al., 2010, 2012, 2013, 2014, 2016, 2017, 2019; Emmons et al., 2004, 2007, 2009). There are four main steps in aircraft versus MOPITT comparisons.

(1) Because of aircraft altitude limitations, in-situ data from field campaigns do not typically reach the highest altitudes at which MOPITT radiances are sensitive. Therefore, to obtain a complete vertical profile as required for comparison with MOPITT retrievals, each in-situ profile is extended vertically using the following steps: (i) the aircraft measurements are interpolated to the 35-level vertical grid used in MOPITT forward model calculations (0.2–1060 hPa); (ii) the levels from the surface to the lowest-altitude aircraft measurement are filled with the value of the in-situ

measurement at the lowest-altitude aircraft measurement; (iii) for levels above a certain pressure level P_{interp} (higher altitude), model or reanalysis data are used directly; (iv) for levels between the highest-altitude aircraft measurement and the altitude of P_{interp}, values are linearly interpolated. Unlike the previous MOPITT evaluation studies that used monthly model results from MOZART (Model for OZone And Related chemical Tracers) (Emmons et al., 2010) or CAM-chem (Community Atmosphere Model with chemistry) (Lamarque et al., 2012), here we use 3-hourly Copernicus Atmosphere Monitoring Service (CAMS) reanalysis of CO produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). CAMS CO reanalysis has a horizontal resolution of 80 km × 80 km, and 60 vertical grids (from surface to 0.1 hPa). Satellite retrievals of atmospheric composition including MOPITT TIR Version 6 total column CO retrievals are assimilated in the CAMS reanalysis (Inness et al., 2019; https://confluence.ecmwf.int/pages/viewpage.action?pageId=83396018). We note that as we do not compare with these higher levels later, the use of CAMS reanalysis is expected to have a minimal impact on the lower levels we use in the comparison (e.g., the surface layer, the 800-hPa layer, and the 600-hPa layer). The final CO profile at the 35-level vertical grid is then regridded onto a coarser 10-level grid (for consistency with the actual MOPITT retrieval grid) by unweighted averaging the fine-grid VMR values in the layers immediately above the corresponding levels in the retrieval grid. We investigate the sensitivity of the results to P_{interp} in Section 4.1.

(2) For a given in-situ profile, MOPITT profiles are considered co-located with the aircraft profile and are selected for comparison only if their center points are within the radius of 100 km and within 12 hours of the acquisition of the aircraft profile. Sensitivities of the results to the radius and time criteria for co-location selection are further investigated in Section 4.2.

(3) For each pair of co-located MOPITT retrieved and in-situ profiles, we apply the MOPITT a priori profile and averaging kernel to the in-situ profile as in Eq. (1). Thus, after converting from profiles of the in-situ and a priori CO concentrations to $\log_{10}$(VMR) profiles ($x_{in-situ}$ and x_a), we calculate

$$x_{transformed} = x_a + A(x_{in-situ} - x_a) \quad (2)$$

so that the $\log_{10}$(VMR)-based transformed in-situ profile ($x_{transformed}$) has the same degree of smoothing and a priori dependence as the MOPITT retrieved $\log_{10}$(VMR) profile (x_{rtv}).

Deleted: e.g., 200 hPa

Deleted: below

Deleted: have conducted further calculations to

Deleted: validation

Deleted: only

Deleted: retrieved within the radius of 100 km and within 12 hours of the acquisition of the aircraft profile

Deleted: s

Deleted: validation

Deleted: have been

Formatted: Right

Formatted: Space Before: 6 pt, After: 6 pt

Deleted: For each pair of co-located MOPITT retrieval and in-situ profiles, we apply the MOPITT a priori profile and averaging kernel to the in-situ profile, ... [1]

(4) For each in-situ profile, there are likely to be multiple MOPITT retrievals that meet the above co-location criteria. If fewer than five MOPITT retrievals are co-located with an in-situ profile, the in-situ profile is not used in the following study and analysis. If an in-situ profile is co-located with five or more MOPITT retrievals (assume the number to be $N_{retrieval}$), then the following steps are used in the comparison with MOPITT: (a) the averaging kernel and a prior of each co-located MOPITT retrieval are applied to the in-situ profile (through equation 2) to obtain $N_{retrieval}$ of $x_{transformed}$. Note that applying these $N_{retrieval}$ sets of MOPITT a priori profiles and averaging kernels to the same in-situ profile results in differently transformed in-situ profiles; (b) the $N_{retrieval}$ of $x_{transformed}$ are averaged in $\log_{10}$(VMR) space; and (c) the $N_{retrieval}$ of MOPITT retrievals x_{rtv} are also averaged.

Figure 2 shows an example of profile comparisons (the original aircraft profile, aircraft profile extended with CAMS reanalysis data and regridded to 35-level grid, $x_{in-situ}$, x_a, $x_{transformed}$, and x_{rtv}) in VMR for an aircraft profile sampled on July 22, 2011 during DISCOVER-AQ in Maryland (MD). Figure 2 also demonstrates what to expect within a MOPITT retrieval pixel and vertical level. The MOPITT retrievals have a spatial resolution of about 22 km × 22 km, and each MOPITT retrieval level corresponds to a layer immediately above that level. The standard deviation of the original aircraft CO observations in each MOPITT layer are also shown, which is due to horizontal and vertical variability in CO. Taking the 800-hPa layer as an example, the standard deviation of the original aircraft CO observations in the level is 21.4 ppb, which is larger than the difference between $x_{transformed}$ and x_{rtv} at that level (12.4 ppb). We also show the relative scale of the aircraft profile (3 km × 5 km) and a MOPITT pixel (22 km × 22 km) in Figure 2. We expect the variability of CO within a MOPITT pixel to be even larger than the CO variability within the scale of 3 km × 5 km. The variability within a satellite pixel and the representativeness error in the satellite retrieval and aircraft profile comparisons make it challenging to compare satellite retrievals to aircraft observations. This is one of the major reasons that MOPITT has yet to be compared with aircraft observations over urban regions with in-situ observations. The representativeness error has been discussed in previous studies (Fishman et al., 2011; Follette-Cook et al., 2015; Judd et al., 2019). Follette-Cook et al. (2015) quantified spatial and temporal variability of column integrated air pollutants, including CO, during DISCOVER-AQ MD from modeling perspective (using the Weather Research and Forecasting model coupled

Deleted: If an in-situ profile is co-located with fewer than five MOPITT retrievals

Formatted: Font color: Text 1

Formatted: Font:Not Italic, Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font:Not Italic, Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font:Not Italic, Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font:Not Italic, Font color: Text 1

Formatted: Font color: Text 1

Deleted: If an in-situ profile is co-located with five or more MOPITT retrievals, these co-located MOPITT profiles are averaged as $\log_{10}$(VMR). Applying these corresponding different MOPITT a priori profiles and averaging kernels to the same in-situ profile results in different transformed in-situ profiles. These transformed in-situ profiles that are generated from the same in-situ profile are also averaged.

Deleted: DC

Deleted: uniformly-weighted

Deleted: The vertical and horizontal variability of the original aircraft CO observations in each MOPITT layer (represented by standard deviation) are also shown

Deleted: level of

Deleted:

Deleted: variability

Deleted: retrieval

Deleted: retrieval

Deleted: very

Deleted: validate

Deleted: against

Deleted: validated

with Chemistry – WRF-Chem). They found that during the July 2011 DISCOVER-AQ campaign, the mean CO difference at the distance of 20-24 km is ~30 ppb (derived from the aircraft observations) and ~40 ppb (derived from co-located WRF-Chem output), based on structure function analyses. In this study, we demonstrate this challenge with an example in Figure 2. We also show a sensitivity analysis in Section 4 to provide perspectives on how the spatial and temporal representativeness may change the results. Further quantification of the variability within MOPITT pixels would be very challenging (partially due to limited coverage of the observational data), and we will elaborate more on this issue in Section 5.

Deleted: in Section 4 the

Deleted: validation

3. MOPITT comparisons with aircraft profiles over urban and non-urban regions

Deleted: validation

Deleted: regions

In this section, the results for MOPITT comparisons with aircraft profiles are provided for only daytime retrievals (i.e., solar zenith angle < 80° in the retrieval), because (1) MOPITT retrievals generally contain more CO profile information in daytime, which is reflected in AKs and Degrees of Freedom for Signal (DFS) in Figure 3, and (2) most aircraft profiles are sampled during daytime. In Section 4.3, we discuss the sensitivity to the inclusion of MOPITT nighttime retrievals in MOPITT comparisons with aircraft profiles. In addition, many aircraft profiles, especially those from DISCOVER-AQ, lack observations above 600 hPa. Even though we extended the aircraft profiles vertically with reanalysis data (as discussed in Section 2.3), this still prevents the use of these profiles for validating MOPITT retrievals at upper levels against in-situ observations. In this paper, we only focus on comparing MOPITT retrievals below the altitude of 600 hPa to aircraft profiles. Nevertheless, since the CO retrievals below 600 hPa are still weakly impacted by CO fields in the upper levels (as shown by the AKs in Figure 3), in Section 4.1 we perform sensitivity tests on how augmenting the aircraft profiles with reanalysis fields affects the comparison results.

Deleted: validation results

Deleted: the

Deleted: validation process

Deleted: validating

Deleted: validation

3.1 Overall statistics

The overall comparison results are presented in Table 2. Following Deeter et al. (2017), retrieval biases and standard deviation (SD) are calculated based on mean x_{rtv} and $x_{transformed}$ for each in-situ profile, and converted from $\log_{10}$(VMR) to percent. The correlation coefficient (r) is quantified based on $(x_{rtv} - x_a)$ and the corresponding $\left(x_{transformed} - x_a\right)$ to avoid

Deleted: validation

Deleted: log(VMR)

correlations which mainly result from the variability of the a priori. x_{rtv}, $x_{transformed}$, and x_a are in $\log_{10}$(VMR) space in order to apply the AKs, which are computed for x_{rtv} in $\log_{10}$(VMR). These comparisons for MOPITT Version 8 TIR-only (V8T) and Version 8 TIR-NIR (V8J) are shown in Figures 4 (for all profiles) and 5 (for urban profiles). Overall biases for V8J products (averaged over all campaigns in Table 1) vary from -0.7% to 0.0%, which are lower than biases for V8T (from 2.0% to 3.5%). Overall biases for V8J products are also smaller than biases for V7J (from -0.5% to -5.4%). For V8J and V7J, biases over urban regions vary from -0.8% to -2% and from -1.4% to -8.9%, respectively, which are generally larger than biases over non-urban regions (-0.3%~1.1% and -3.3%~0.1%). Correlation coefficients over non-urban regions are higher than those over urban regions for all six products (V7T, V8T, V7N, V8N, V7J, V8J) at all three levels in Table 2 (the surface layer, the 800-hPa layer, and the 600-hPa layer). We also notice that for TIR-NIR and TIR-only products, V8 have higher correlation coefficients with in-situ measurements than V7 over non-urban regions, whereas over urban regions, V8 products have lower correlation coefficients than V7 (except for the 600-hPa layer). Overall, MOPITT products (especially V8J) perform reasonably well over both urban and non-urban regions. Performance over non-urban regions is better than that over urban regions in terms of higher correlation coefficients and smaller biases for V8J and V7J.

3.2 Discussions on individual campaigns

We also evaluate MOPITT V8J retrievals during individual field campaigns with results in Figure 6. The corresponding results for MOPITT V8T are summarized in Figure S2. The patterns of biases are very similar for MOPITT V8J and V8T. Thus, in this sub-section, we focus on V8J unless stated otherwise. Overall, except comparisons with A-FORCE and ARIAs, biases over urban regions and non-urban regions do not have a significant difference. Neither do biases determined for campaigns over the US and East Asia differ significantly, either.

When compared to DISCOVER-AQ CA, MOPITT CO values are generally higher than in-situ profiles at the 600-hPa layer (i.e., the 100 hPa uniform layer immediately above 600 hPa) but not at the surface layer (i.e., the uniform layer immediately above the surface). This is likely related to the fact that the DISCOVER-AQ CA aircraft profiles are mostly below 600 hPa, and hence CO values of these in-situ profiles at 600 hPa and above are filled with CAMS reanalysis data. In addition, DISCOVER-AQ CA was conducted in the winter when boundary layer height is

Deleted: Corresponding results

Deleted: lower

Deleted: from -0.2% to -0.8%

Deleted: -8.9% to

Deleted: higher

Deleted: generally

Deleted:

Deleted:

Deleted: For example, for the V8J product, correlation coefficients over urban regions are 0.53, 0.57, and 0.53 at the surface, 800 hPa, and 600 hPa, respectively, while over non-urban regions, the corresponding correlation coefficients are 0.76, 0.73 and 0.67.

Deleted: products generally

Deleted: generally

Deleted: in terms of correlation coefficients and biases

Deleted: provide

Deleted: evaluation against

Deleted: besides

Deleted: EA

Deleted: at 600 hPa

Deleted:

at lower altitudes, which could also explain the difference, in particular since most of the other campaigns are during times with greater vertical mixing. The lack of aircraft observations at 600 hPa and above also has a smaller impact on the biases at the 800-hPa layer through applying AK (see Figure 3).

During the A-FORCE campaign, only 2 in-situ profiles out of 45 were sampled over urban regions. The locations of the two profiles are close to each other and they are both sampled on/near the coast of South Korea (Figure 1). MOPITT has large negative biases (-30%~-40%) when compared to these two profiles. The averaged $x_{in-situ}$, x_a, $x_{transformed}$, and x_{rtv} over non-urban regions during A-FORCE and the $x_{in-situ}$, x_a, $x_{transformed}$, and x_{rtv} of the two profiles over urban regions are shown in Figure S3. Compared to the averaged $x_{in-situ}$ over non-urban regions, the $x_{in-situ}$ for the two profiles over the urban regions have large enhancements near the surface and between 600~800 hPa. Even though the x_a and x_{rtv} for the two profiles have higher CO concentrations (~400 ppb at the surface layer) than the averaged x_a and x_{rtv} (~200 ppb at the surface layer), they are still lower than the $x_{transformed}$.

As for KORUS-AQ, MOPITT also has a negative bias (though smaller) when compared to the profiles over urban regions. Most of these KORUS-AQ profiles were located near the two profiles from A-FORCE but farther from the coast. The negative bias is not seen over non-urban regions during KORUS-AQ at the surface layer.

When compared to the in-situ profiles from ARIAs, MOPITT has a large positive bias, especially over urban regions (20%~30%). During ARIAs, in-situ profiles over urban regions have lower CO values (~200 ppb at the surface layer) than those in-situ profiles over non-urban regions (~ 400 ppb at the surface layer; Figure S4). We note there are only a small number of in-situ profiles over urban regions in East Asia used in this study, compared to what is provided by DISCOVER-AQ in the US. The large negative biases against A-FORCE and large positive biases against ARIAs point to the need for more in-situ observations over East Asia.

3.3 MOPITT comparisons with aircraft profiles at high CO concentrations

Urban regions are often associated with high CO concentrations. But this is not always the case (e.g., Figure S4). Here we separate the in-situ profiles at the surface layer, the 800-hPa layer, and the 600-hPa layer into lower 50% CO values and higher 50% CO values based on CO values

Deleted: in

Deleted: more favorable weather conditions

Deleted:

Deleted: ,

Deleted: EA

Deleted: EA

Deleted: Validation

Deleted: the surface, 800 hPa, and 600 hPa at each level to demonstrate the impact of CO concentrations on the MOPITT product validation (Figure 7). For V8J, MOPITT has smaller biases at higher 50% CO concentrations for all three levels, whereas for V8T, MOPITT has larger biases at the surface layer and the 600-hPa layer at higher 50% CO concentrations. For the higher 50% of measured mixing ratios both V8J and V8T have larger SDs and lower correlation coefficients at the surface layer, the 800-hPa layer, and the 600-hPa layer, suggesting that the agreement between MOPITT and the in-situ profiles at higher CO concentrations is not as good as that at lower CO concentrations. In contrast, Deeter et al. (2016) found that the retrieval biases do not visibly increase at the upper range of CO concentrations when compared to aircraft measurements over the Amazon basin. The vertical error bars in Figure 7 (caused by the multiple co-located MOPITT profiles with one in-situ profile) represent the variability (standard deviation) of the MOPITT data used to calculate each of the plotted mean values. For an in-situ profile, the variability of the MOPITT data located within its radius of 100 km and within 12 hours is larger when the in-situ profile has higher CO values, indicated by larger error bars at higher 50% CO concentrations. At higher 50% CO concentrations, the averaged retrieval uncertainties for the 600-hPa, 800-hPa, and surface layers, are 28%, 28%, and 29%, respectively. This is smaller than the averaged retrieval uncertainties at lower 50% CO concentrations (28%, 29%, and 30% for the 600-hPa, 800-hPa, and surface layers, respectively). We therefore conclude that the larger apparent biases at high CO concentrations are related to greater CO variability and representativeness error of the in-situ profile within the co-location radius used for analyzing the MOPITT data, rather than indicating larger retrieval uncertainties. Theoretically, MOPITT retrievals perform better with higher CO concentrations. The larger biases at high CO concentrations in Figure 7 implies that the relatively greater CO variability may overcome the impact of high CO concentrations. Addressing representativeness error and spatial variability in the comparisons between satellite and in-situ profiles is challenging, and will be discussed further in Section 5.

We will discuss the sensitivity of radius and time difference for the selection of co-located data in Section 4. The difference in the variability at different CO concentrations was not found in Deeter et al. (2016). It could be partially due to the fact that the aircraft profiles over the Amazon basin used in Deeter et al. (2016) were sampled under more geographically homogeneous conditions, whereas the profiles used in this study are from different campaigns, and high CO

Deleted:

Deleted: For both

Deleted: ,

Deleted: MOPITT

Deleted: s

Deleted: the surface, 800 hPa, and 600 hPa if only the upper 50% of measured CO mixing ratios are considered

Deleted: this

Deleted: validation

Deleted: of

Deleted: However, it is unclear whether the larger apparent bias at high CO concentration actually represents larger retrieval uncertainties or could be related to greater CO variability and representativeness of the in situ profile within the co-location radius used for analyzing the MOPITT data.

Deleted: in

concentrations over and near urban regions might be associated with more complex and inhomogeneous conditions.

4. Sensitivities to assumptions made for aircraft-satellite comparisons

In Section 3, we compared profiles over urban and non-urban regions separately to MOPITT V8T, V8N, V8J, V7T, V7N, and V7J. In this section, we compare only the MOPITT V8J product to all the in-situ profiles (both over urban and non-urban regions) described in Table 1 to test the sensitivity of results to the assumptions made during the comparisons.

4.1 Sensitivity to the in-situ profile extension

As discussed in Section 2.3, the in-situ profiles must be vertically extrapolated or extended to compare with MOPITT products due to aircraft altitude limits. Thus, model or reanalysis data must be merged with the in-situ data to generate a complete CO profile for comparisons with MOPITT satellite retrievals. The use of model or reanalysis data may introduce uncertainties in the comparison results as they are not measured directly. The parameter P_{interp} controls the impact of the model-based profile extension on the shape and value of in-situ profiles (see Figure S5). Here we test the sensitivity of validation results to various P_{interp} values (100 hPa, 200 hPa, 300 hPa, 400 hPa, 500 hPa) to demonstrate the potential impact of the profile extension. Note that the model-based profile extension and the value of P_{interp} impacts the validation results through changing the augmented observational profile, which is different from the other sensitivity tests in this study that change the selection of MOPITT data. The agreements between the values of MOPITT and in-situ profiles at the surface layer are insensitive to the selection of P_{interp} (Figure 8). The overall agreements between the values of MOPITT and in-situ profiles at the 800-hPa layer are also not sensitive to P_{interp}, except for the results against DISCOVER-AQ CA which have slightly larger biases when P_{interp} is 200 hPa or 100 hPa since the DISCOVER-AQ CA aircraft profiles at 600 hPa and above are mostly extended using reanalysis data. Therefore, the comparisons with DISCOVER-AQ CA are more likely to be affected by P_{interp} compared to other campaigns which typically obtained higher maximum aircraft altitudes. At the 600-hPa layer, the agreements between the values of MOPITT and in-situ profiles are affected more by P_{interp} compared to the those at the surface layer and the 800-hPa layer for comparisons with all the campaigns. The overall validation results using 100 hPa as P_{interp} have larger biases than using

Formatted: Font:Not Bold
Formatted: Font:Not Bold
Deleted: for use in
Deleted: validation
Deleted: validation
Deleted: on the validation results
Deleted: validation results
Deleted: validation results
Deleted:
Deleted: validation
Deleted: .
Deleted: As mentioned in Section 3.2,
Deleted: are mostly below 600 hPa, and hence CO values of these in-situ profiles
Deleted: validation results against
Deleted:
Deleted: validation results
Deleted: more
Deleted:
Deleted:

other values of P_{interp}. At 400-hPa layer and 200-hPa layer, the comparisons are even more sensitive to P_{interp} for all the campaigns (Figure S6). The CAMS 3-hourly reanalysis data are constrained by observations, but its usage may still introduce the uncertainties in the validation results especially at upper pressure levels (e.g., 200 hPa and 400 hPa). Previous MOPITT evaluation results may be subject to larger uncertainties by using CAM-chem monthly CO fields that are not constrained by observations (e.g., Deeter et al., 2012, 2016).

4.2 Sensitivity to the radius and allowed maximum time difference as criteria for co-location

The criteria for co-location in this study (within a radius of 100 km and within 12 hours of the acquisition of the aircraft profile) generally follow previous MOPITT validation studies (e.g., Deeter et al., 2016, 2019) and are chosen empirically. They are selected based on a trade-off between uncertainties generated from CO spatial and/or temporal variability, and the number of included MOPITT retrievals that impacts the statistical robustness. Here we test the sensitivity of the results to the two criteria for co-location. The boxplot of biases calculated with different radii (200 km, 100 km, 50 km, and 25 km) at the surface layer, the 800-hPa layer, and the 600-hPa layer are shown in Figure 9. Overall, the biases calculated with radius of 200 km, 100 km and 50 km are similar, whereas the biases calculated with the radius of 25 km are different from others. The comparisons of MOPITT to in-situ profiles results using the radius of 25 km generally have larger biases and SD, due to including fewer MOPITT retrievals. In some cases, there are no matched MOPITT retrievals within the radius of 25 km of the aircraft profile (e.g., DISCOVER-AQ CA and ARIAs). In addition, representativeness errors would be expected to go up if there are only a few retrievals over a more polluted and perhaps heterogeneous area. We note that the usage of the largest radius (200 km) in this paper does not appear to degrade the overall results, even though representativeness errors generated from CO spatial and/or temporal variability are expected to increase. However, the use of the smallest radius (25 km) degrades the overall results by reducing the number of included MOPITT retrievals.

The boxplot of biases calculated with four sets of allowed maximum time difference (12 hours, 6 hours, 3 hours, and 1 hours) are shown in Figure 10. The overall results are not sensitive to the selection of allowed maximum time difference. One exception is the comparisons to the SEAC[4]RS campaign at the 600-hPa layer, due to a smaller number of MOPITT retrievals in the shorter time window. We note that when comparing to the ARIAs campaign, using 1h as the

Deleted: The validation results using 300, 400, or 500 hPa as P_{interp} are not significantly different for the validation results against DISCOVER-AQ CA. The validation results against DISCOVER-AQ CA using 200 hPa as P_{interp} show similar results as those using 100 hPa as P_{interp}.

Deleted: T

Deleted: validation results

Deleted: to the P_{interp} at 400 hPa and 200 hPa

Deleted: with larger biases

Deleted: As mentioned in Section 3.2, the DISCOVER-AQ CA aircraft measurements concentrate below 600 hPa, so CO values in the in-situ profiles at 600 hPa and above are filled with and are more sensitive to CAMS reanalysis data.

Deleted: the

Deleted: validation

Deleted: the surface, 800 hPa, and 600 hPa

Deleted: close

Deleted: validation

Deleted: a smaller number of included

Deleted: a more

Deleted: through

Deleted: introducing

Deleted: ,

Deleted: whereas

Deleted: validation

Deleted: , especially at the surface

Deleted: validation results

Deleted: against

Deleted:

allowed maximum time difference decreases the biases at the surface layer, the 800-hPa layer, and the 600-hPa layer, compared to the cases using longer allowed maximum time difference (i.e., 3h, 6h, and 12h). This implies that the temporal variability is relatively large in the region. And the improvement observed for ARIAs for the shortest time also points to the possibility that short term emission sources might be responsible for the large biases there. On the other hand, when the allowed maximum time difference equals 1 hour, there are only 6 aircraft profiles that have matched MOPITT retrievals.

Deleted: We note that when validated against the ARIAs campaign, the biases at the surface, 800 hPa and 600 hPa are smaller with the allowed maximum time difference as 1h, indicating the temporal variability is relatively large in the region.

4.3 Sensitivity to the inclusion of MOPITT nighttime retrievals

Previous MOPITT validation studies have only included MOPITT daytime observations. Over land, MOPITT retrievals for daytime and nighttime overpasses are characterized by significantly different averaging kernels (Figure 3), and may be subject to different types of retrieval error (Deeter et al., 2007). CO has a long enough lifetime (approximately a month; Gamnitzer et al., 2006) in the free troposphere that nighttime observations could be potentially comparable, in general, to the daytime flights for remote sites. However, for urban regions where the spatiotemporal variability of the emissions and evolution of the planetary boundary layer drives large changes in the measured CO, comparisons of MOPITT nighttime observations to aircraft profiles sampled during daytime may introduce representative uncertainties, especially for areas that are subject to strong nocturnal inversions and the surface CO can be enhanced. It is difficult to disentangle the effects of the MOPITT daytime/nighttime performance and the uncertainty from the temporal representativeness, based on the comparison of the MOPITT daytime/nighttime retrievals with daytime aircraft profiles. Therefore, we only include the results in Figure S7 and briefly describe the results here without drawing any further conclusions. Overall, MOPITT nighttime retrievals have larger biases than daytime retrievals, which could be expected since most of the aircraft profiles are sampled during daytime. Flight campaigns with nighttime observations are needed to validate MOPITT nighttime retrievals.

4.4 Sensitivity to the signal-to-noise ratio (SNR) filters

The MOPITT Level 3 data are generated from Level 2 data, and are available as gridded (1°×1°) daily-mean and monthly-mean files. Pixel filtering and signal-to-noise ratio (SNR) thresholds for Channel 5 and 6 Average radiances are used when averaging Level 2 data into Level 3 data, and this increases overall mean DFS values (details can be found in the MOPITT Version

8 Product User's Guide, 2018). Taking MOPITT V8J daytime product as an example, the Level 3 data product excludes all observations from Pixel 3 (one of the four elements of MOPITT's linear detector array that has highly variable Channel 7 SNR values), or observations where both the Channel 5 Average radiances SNR < 1000 and the Channel 6 Average radiances SNR < 400. In Figure 11, we test the impact of applying the aforementioned SNR filters on the agreement between MOPITT and in-situ profiles. Note that we are not suggesting the comparisons between MOPITT Level 3 product and aircraft measurements. Because the MOPITT Level 3 product is gridded data and represent the average value in a 1°×1° grid. Comparing the grid average value to an aircraft profile within it may be subject to large representativeness errors. Here we only show the sensitivity of agreement between MOPITT Level 2 data and aircraft profiles to the application SNR filter. We find that applying the SNR filters does not significantly change the overall agreement between MOPITT retrievals and the in-situ profiles used in this study. This is mostly because applying the SNR filters reduces the number of MOPITT retrievals included in the comparisons. This effect is particularly important if there are not many MOPITT retrievals to begin with (such as our comparisons with in-situ profiles in this study). Even though applying SNR filter when generating Level 3 data does not significantly change the agreement with the in-situ profiles used in this study, by excluding low-SNR observations from the Level 3 cell-averaged values raises overall mean DFS values (MOPITT Version 8 Product User's Guide, 2018). In addition, the Level 3 product typically are less affected by random retrieval errors (e.g., due to instrument noise or geophysical noise).

Deleted: user guide; https://www2.acom.ucar.edu/sites/default/files/mopitt/v8_users_guide_201812.pdf

Deleted: validation results

Deleted: improve

Deleted: In some cases, applying the SNR filters degrades the validation results (e.g., DISCOVER-AQ DC at the surface, DISCOVER-AQ CA at the surface, KORUS-AQ at 600 hPa, and ARIAs at the surface, 800 hPa, and 600 hPa).

Deleted: However, when generating Level 3 data from Level 2 data, the circumstance is different as there are usually much more data to perform the filter and averaging process.

5. Discussion and conclusions

MOPITT products are widely used for monitoring and analyzing CO over urban regions. However, systematic validation against observations over urban regions has been lacking. In this study, we compared MOPITT products over urban regions to aircraft measurements from DISCOVER-AQ, SEAC^{4}RS, ARIAs, A-FORCE, and KORUS-AQ campaigns. The DISCOVER-AQ campaign was designed primarily with satellite validation in mind, and the campaign over MD, CA, TX, and CO together contributes 64.8% (232 out of 358) of the aircraft profiles and 91.0% (121 out of 133) of the aircraft profiles over the urban regions in this study (Table 1). Therefore, the DISCOVER-AQ campaign largely contributes to the results and the statistics in this study. We

Deleted: DC

Deleted: validation

found that MOPITT mean biases are well within the 10% required accuracy (Drummond and Mand, 1996) for both urban and non-urban regions (mean biases for V8J and V8T vary from -0.7% to 0.0%, and from 2.0% to 3.5% for different levels). The performance over non-urban regions is better than that over urban regions in terms of correlation coefficients for the 6 products in Table 2, and biases of V8J and V7J. However, the in-situ profiles over East Asia used in this study are limited, especially over urban regions (only 11 profiles). The large biases against aircraft profiles from the A-FORCE and ARIAs campaigns point to the need for more in-situ observations over East Asia. We also studied the impact of CO concentrations on the agreement between MOPITT products and in-situ profiles by dividing the aircraft profiles of CO into two groups of high CO (upper 50%) and low CO (lower 50%). We found that MOPITT retrievals at high CO concentrations have higher biases and lower correlations compared with low CO concentrations, although CO variability may tend to exaggerate retrieval biases in heavily-polluted scenes. The statistics are often very similar between different versions and products over urban and non-urban regions, and in general, MOPITT agrees reasonably well with the in-situ profiles in both cases. There is not, therefore, any reason to recommend the continued use of MOPITT versions earlier than V8 based on urban or non-urban region considerations. In general, MOPITT V8 is recommended (Deeter et al., 2019) as it uses a new parameterized radiance bias correction method to minimize retrieval biases, and has updated spectroscopic data for water vapor and nitrogen.

In addition, the assumptions and data filters made during aircraft-satellite comparisons may impact the validation results. We tested the sensitivities of the results to assumptions and data filters, including the model-based extension to the in-situ profile, radius and allowed maximum time difference as criteria for the selection of co-located data, the inclusion of nighttime MOPITT data, and the SNR filters. The agreements between the values of MOPITT and in-situ profiles at the surface layer are insensitive to the model-based profile extension, whereas the results at upper levels (e.g., 400 hPa and 200 hPa) are more sensitive to the profile extension, as there are very limited aircraft observations. The results are insensitive to the allowed maximum time difference as a co-location criteria, and are generally insensitive to the radius for co-location except for the case with a radius of 25 km, where only a small number of MOPITT retrievals are included in the comparisons. Overall, daytime MOPITT products overall have smaller biases than nighttime MOPITT products. However, conclusions regarding the performance of MOPITT daytime and nighttime retrievals cannot be drawn due to the fact that most of the aircraft profiles are sampled

Deleted: overall

Deleted: EA

Deleted: EA

Deleted: validation

Formatted: Font:Not Italic, Font color: Text 1

Formatted: Font color: Text 1

Deleted: validation

Deleted: validation results

Deleted: validation

Deleted: validation

Deleted:

Deleted: validation

during daytime. As we mentioned earlier, MOPITT daytime and nighttime retrievals may be subject to different retrieval errors. In addition, previous studies suggest pollutants themselves may have different characteristics during daytime and nighttime (e.g., Yan et al., 2018). Therefore, validation of MOPITT nighttime retrievals, with a sufficient number of nighttime airborne profiles, is needed in order to study nighttime CO characteristics and trends. Applying SNR filters does not necessarily change the overall agreement between MOPITT retrievals and in-situ profiles used in this study significantly, and this may be partially caused by the smaller number of MOPITT retrievals in the comparisons after the SNR filters. We note that comparisons to ARIAs are exceptional in a few sensitivity tests due to rather a limited number of aircraft measurements. Given the large biases against aircraft profiles from the ARIAs campaign, more in-situ observations over East Asia especially China are needed in order to validate MOPITT products in the region.

Validation and evaluation of satellite retrievals with aircraft observations are very challenging, and assumptions have to be made for the comparisons. As discussed in Section 2, the CO spatial variability within MOPITT retrieval pixels and the representativeness error of aircraft profiles when compared to MOPITT retrievals may introduce uncertainties in the validation results. This issue is difficult to address and quantify due to the limited spatial coverage of dense aircraft observations. One possible way is to study NO_2 data retrieved from the Geostationary Trace Gas and Aerosol Sensor Optimization (GeoTASO) at very high resolution (250 m × 250 m), to provide an upper estimate on CO variability. Besides, the variability of Tropospheric Monitoring Instrument (TROPOMI) CO retrievals (resolution: 7 km×7 km; Landgraf et al., 2016) might also provide information on MOPITT sub-pixel variability. Further research on trace gas spatial variability within satellite retrieval pixels, and quantification of the representativeness error incurred by comparing individual aircraft profiles to satellite products is needed, and will be the subject of a follow-up study.

Data availability

MOPITT products are available at https://www2.acom.ucar.edu/mopitt (Last access date: January 14th, 2020). MOPITT Version 8 Product User's Guide is available online at https://www2.acom.ucar.edu/sites/default/files/mopitt/v8_users_guide_201812.pdf (Last access

Deleted: improve

Deleted: validation

Deleted: process

Deleted: , which is unlikely to happen when generating Level 3 data

Deleted: validation results against

Deleted: an

Deleted: n

Deleted: EA

Deleted: Follette-Cook et al. (2015) quantified spatial and temporal variability of column integrated air pollutants, including CO, during DISCOVER-AQ DC from modeling perspective (using the Weather Research and Forecasting model coupled with Chemistry - WRF-Chem). They found that during the July 2011 DISCOVER-AQ campaign, the mean CO difference at the distance of 20-24 km is ~30 ppb (derived from the aircraft observations) and ~40 ppb (derived from co-located WRF-Chem output), based on structure function analyses. Judd et al. (2019) explored the impact of spatial resolution on tropospheric NO_2 column retrievals with NASA Geostationary Trace Gas and Aerosol Sensor Optimization (GeoTASO). We expect CO to have a smaller spatial and temporal variability than NO_2 due primarily to its relatively longer lifetime, though

Deleted: future analyses of NO_2 variability within urban regions using GeoTASO could

Deleted: In addition

Deleted: ,

Deleted: with a pixel size of

Deleted: (

Deleted: .

Deleted: within the larger MOPITT footprint

Deleted: using

Deleted: in

Deleted: validation

Deleted: comparisons

Formatted: Font:Bold

Formatted: Indent: First line: 0"

Moved (insertion) [2]

Formatted: Superscript

date: January 15th, 2020). DISCOVER-AQ data can be accessed at https://www-air.larc.nasa.gov/missions/discover-aq/discover-aq.html (Last access date: January 14th, 2020). SEAC^{4}RS data can be accessed at https://www-air.larc.nasa.gov/missions/seac4rs/ (Last access date: January 14th, 2020). KORUS-AQ and ARIAs data can be accessed at https://www-air.larc.nasa.gov/missions/korus-aq/index.html (Last access date: January 14th, 2020). A-FORCE data are available upon request (Yutaka Kondo: kondo.yutaka@nipr.ac.jp). MODIS Land Cover Type Global product (MCD12C1 v006) is available at https://earthdata.nasa.gov/ (Last access date: January 14th, 2020).

Author contribution

WT, HMW, and MND designed the study. WT analyzed the data with help from MND, SMA, and LKE. GSD provided CO measurements during DISCOVER-AQ SEAC^{4}RS, and KORUS-AQ. RRD, XR, and HH provided CO measurements during ARIAs. YK provided CO measurements during A-FORCE. HMW, MND, DPE, LKE, BG, RRB, and XR offered valuable discussions and comments in improving the study. WT prepared the manuscript with improvements from all the other coauthors.

Acknowledgements

The National Center for Atmospheric Research (NCAR) is sponsored by the National Science Foundation (NSF). W. Tang is supported by a NCAR Advanced Study Program Postdoctoral Fellowship. The NCAR MOPITT project is supported by the National Aeronautics and Space Administration (NASA) Earth Observing System (EOS) Program. The authors thank the DISCOVER-AQ, SEAC^{4}RS, ARIAs, A-FORCE, and KORUS-AQ Science Teams for the valuable in-situ observations. We thank Drs. Naga Oshima and Makoto Koike for the A-FORCE data. ARIAs was supported by NSF (grant # 1558259) and National Institute of Standards and Technology (NIST, grant #70NANB14H332). The authors thank Dr. Frank Flocke for helpful comments on the manuscript. Wenfu Tang thanks Dr. Cenlin He for helpful discussions.

References

Moved (insertion) [3]

Deleted: MOPITT data can be downloaded at https://earthdata.nasa.gov/.

Moved (insertion) [4]

Formatted: Font:Bold

Formatted: Superscript

Formatted: Font color: Auto, (Asian) Chinese (PRC)

Moved (insertion) [1]

Deleted:

Moved up [2]: MOPITT products are available at https://www2.acom.ucar.edu/mopitt. MOPITT data can be downloaded at https://earthdata.nasa.gov/.

Moved up [3]: DISCOVER-AQ data can be accessed at https://www-air.larc.nasa.gov/missions/discover-aq/discover-aq.html. SEAC^{4}RS data can be accessed at https://www-air.larc.nasa.gov/missions/seac4rs/.

Moved up [4]: KORUS-AQ data can be accessed at https://www-air.larc.nasa.gov/missions/korus-aq/index.html.

Moved up [1]: The National Center for Atmospheric Research (NCAR) is sponsored by the National Science Foundation. W. Tang is supported by a NCAR Advanced Study Program Postdoctoral Fellowship.

Al-Saadi, J., Carmichael, G., Crawford, J., Emmons, L., Song, C. K., Chang, L. S., ... & Park, R. (2015). NASA contributions to KORUS-AQ: An international cooperative air quality field study in Korea. NASA White Paper available at: https://goo. gl/VhssdX (last access: 3 May 2016).

Aliyu, Y. A., & Botai, J. O.: Appraising city-scale pollution monitoring capabilities of multi-satellite datasets using portable pollutant monitors. *Atmospheric environment*, *179*, 239-249, 2018.

Arellano, A. F., Kasibhatla, P. S., Giglio, L., van der Werf, G. R., and Randerson, J. T.: Topdown estimates of global CO sources using MOPITT measurements, Geophys. Res. Lett., 31, L01104, doi:10.1029/2003GL018609, 2004.

Arellano, A. F., Kasibhatla, P. S., Giglio, L., van der Werf, G. R., Randerson, J. T., and Collatz, G. J.:, Time-dependent inversion estimates of global biomass-burning CO emissions using Measurement of Pollution in the Troposphere (MOPITT) measurements, J. Geophys. Res., 111, D09303, doi:10.1029/2005JD006613, 2006.

Arellano A. F. Jr., Raeder, K., Anderson, J. L., Hess, P. G., Emmons, L. K., Edwards, D. P., Pfister, G. G., Campos, T. L., and Sachse, G. W.: Evaluating model performance of an ensemble-based chemical data assimilation system during INTEX-B field mission, Atmos. Chem. Phys., 7, 5695-5710, https://doi.org/10.5194/acp-7-5695-2007, 2007.

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, Atmos. Meas. Tech., 10, 1927–1956, https://doi.org/10.5194/amt-10-1927-2017, 2017.

Chen, D., Wang, Y., McElroy, M. B., He, K., Yantosca, R. M., and Le Sager, P.: Regional CO pollution and export in China simulated by the high-resolution nested-grid GEOS-Chem model, Atmos. Chem. Phys., 9, 3825-3839, https://doi.org/10.5194/acp-9-3825-2009, 2009.

Clerbaux, C., Edwards, D. P., Deeter, M., Emmons, L., Lamarque, J.-F., Tie, X. X., Massie, S. T., and Gille, J.: Carbon monoxide pollution from cities and urban areas observed by the Terra/MOPITT mission, Geophys. Res. Lett., 35, 3817, doi:10.1029/2007GL032300, 2008.

Deeter, M. N., Emmons, L. K., Francis, G. L., Edwards, D. P., Gille, J. C., Warner, J. X., Khattatov, B., Ziskin, D., Lamarque, J.-F., Ho, S.-P., Yudin, V., Attie, J.-L., Packman, D., Chen, J., Mao, D., and Drummond, J. R.: Operational carbon monoxide retrieval algorithm and selected results for the MOPITT instrument, J. Geophys. Res., 108(D14), 4399, doi:10.1029/2002JD003186, 2003.

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, J. Geophys. Res., 115, D07306, doi:10.1029/2009JD013005, 2010.

Deeter, M. N., Worden, H. M., Gille, J. C., Edwards, D. P., Mao, D., and Drummond, J. R.: MOPITT multispectral CO retrievals: Origins and effects of geophysical radiance errors, J. Geophys. Res., 116, D15303, doi:10.1029/2011JD015703, 2011.

Deeter, M. N., Worden, H. M., Edwards, D. P., Gille, J. C., and Andrews, A. E.: Evaluation of MOPITT Retrievals of Lowertropospheric Carbon Monoxide over the United States, J. Geophys. Res., 117, D13306, doi:10.1029/2012JD017553, 2012.

Deeter, M. N., Martínez-Alonso, S., Edwards, D. P., Emmons, L. K., Gille, J. C., Worden, H. M., Pittman, J. V., Daube, B. C., and Wofsy, S. C.: Validation of MOPITT Version 5 thermalinfrared, near-infrared, and multispectral carbon monoxide profile retrievals for 2000–2011, J. Geophys. Res., 118, 6710–6725, doi:10.1002/jgrd.50272, 2013.

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, Atmos. Meas. Tech., 7, 3623–3632, https://doi.org/10.5194/amt-7-3623-2014, 2014.

Deeter, M. N., Martínez-Alonso, S., Gatti, L. V., Gloor, M., Miller, J. B., Domingues, L. G., and Correia, C. S. C.: Validation and analysis of MOPITT CO observations of the Amazon Basin, Atmos. Meas. Tech., 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, Atmos. Meas. Tech., 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., Martínez-Alonso, S., Worden, H. M., Ziskin, D., and Andreae, M. O.: Radiance-based retrieval bias mitigation for the MOPITT instrument: the version 8 product, Atmos. Meas. Tech., 12, 4561–4580, https://doi.org/10.5194/amt-12-4561-2019, 2019.

Drummond, J. R., & Mand, G. S. (1996). The Measurements of Pollution in the Troposphere (MOPITT) instrument: Overall performance and calibration requirements. Journal of Atmospheric and Oceanic Technology, 13(2), 314-320.

Edwards, D., Halvorson, C., and Gille, J.: Radiative transfer modeling for the EOS Terra satellite Measurement of Pollution in the Troposphere (MOPITT) instrument, J. Geophys. Res., 104, 16755–16775, 1999.

Edwards, D. P., Petron, G., Novelli, P. C., Emmons, L. K., Gille, J. C., and Drummond, J. R.: Southern Hemisphere carbon monoxide interannual variability observed by Terra/Measurement of Pollution in the Troposphere (MOPITT), J. Geophys. Res., 111, D16303, doi:10.1029/2006JD007079, 2006.

Emmons, L. K., Deeter, M. N., Gille, J. C., et al.: Validation of Measurements of Pollution in the Troposphere (MOPITT) CO retrievals with aircraft in situ profiles, J. Geophys. Res., 109, D03309, doi:10.1029/2003JD004101, 2004.

Emmons, L. K., Pfister, G. G., Edwards, D. P., Gille, J. C., Sachse, G., Blake, D., Wofsy, S., Gerbig, C., Matross, D., and N´ed´elec, P.: Measurements of Pollution in the Troposphere (MOPITT) validation exercises during summer 2004 field campaigns over North America, J. Geophys. Res., 112, D12S02, doi:0.1029/2006JD007833, 2007.

Emmons, L. K., Edwards, D. P., Deeter, M. N., Gille, J. C., Campos, T., Nédélec, P., Novelli, P., and Sachse, G.: Measurements of Pollution In The Troposphere (MOPITT) validation through 2006, Atmos. Chem. Phys., 9, 1795-1803, https://doi.org/10.5194/acp-9-1795-2009, 2009.

Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J.-F., Pfister, G. G., Fillmore, D., Granier, C., Guenther, A., Kinnison, D., Laepple, T., Orlando, J., Tie, X., Tyndall, G., Wiedinmyer, C., Baughcum, S. L., and Kloster, S.: Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4), Geosci. Model Dev., 3, 43–67, https://doi.org/10.5194/gmd-3-43-2010, 2010.

Fishman, J., Silverman, M. L., Crawford, J. H., & Creilson, J. K.: A study of regional-scale variability of in situ and model-generated tropospheric trace gases: Insights into observational requirements for a satellite in geostationary orbit. Atmospheric environment, 45(27), 4682-4694, 2011.

Follette-Cook, M., Pickering, K., Crawford, J., Duncan, B., Loughner, C., Diskin, G., Fried, A., and Weinheimer, A.: Spatial and temporal variability of trace gas columns derived from WRF/Chem regional model output: Planning for geostationary observations of atmospheric composition, Atmos. Environ., 118, 28–44, doi:10.1016/j.atmosenv.2015.07.024, 2015.

Fortems-Cheiney, A., Chevallier, F., Pison, I., Bousquet, P., Szopa, S., Deeter, M. N., and Clerbaux, C.: Ten years of CO emissions as seen from Measurements of Pollution in the Troposphere (MOPITT), J. Geophys. Res., 116, D05304, doi:10.1029/2010JD014416, 2011.

Friedl, M., Sulla-Menashe, D. (2015). MCD12C1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 0.05Deg CMG V006 [Data set]. NASA EOSDIS Land Processes DAAC. Accessed 2019-08-12 from https://doi.org/10.5067/MODIS/MCD12C1.006.

Gamnitzer, U., Karstens, U., Kromer, B., Neubert, R. E., Meijer, H. A., Schroeder, H., & Levin, I. (2006). Carbon monoxide: A quantitative tracer for fossil fuel CO2? Journal of Geophysical Research, 111, D22302. https://doi.org/10.1029/2005JD006966

Gaubert, B., Arellano, A. F., Barré, J., Worden, H. M., Emmons, L. K., Tilmes, S., Buchholz, R. R.,Vitt, F., Raeder, K., Collins, N., Anderson, J. L., Wiedinmyer, C., Martinez Alonso, S., Edwards, D. P., Andreae, M. O., Hannigan, J. W., Petri, C., Strong, K., and Jones, N.: Toward a chemical reanalysis in a coupled chemistry-climate model: An evaluation of MOPITT CO assimilation and its impact on tropospheric composition, J. Geophys. Res.-Atmos., 121, 7310–7343, https://doi.org/10.1002/2016JD024863, 2016.

Girach, I. A. and Nair, P. R.: Carbon monoxide over Indian region as observed by MOPITT, Atmos. Environ., 99, 599–609, 2014.

He, H., Stehr, J. W., Hains, J. C., Krask, D. J., Doddridge, B. G., Vinnikov, K. Y., Canty, T. P., Hosley, K. M., Salawitch, R. J., Worden, H. M., and Dickerson, R. R.: Trends in emissions and concentrations of air pollutants in the lower troposphere in the Baltimore/Washington airshed from 1997 to 2011, Atmos. Chem. Phys., 13, 7859-7874, https://doi.org/10.5194/acp-13-7859-2013, 2013.

Heald, C. L., Jacob, D. J., Jones, D. B. A., Palmer, P. I., Logan, J. A., Streets, D. G., Sachse, G. W., Gille, J. C., Hoffman, R. N., and Nehrkorn, T.: Comparative inverse analysis of satellite (MOPITT) and aircraft (TRACE-P) observations to estimate Asian sources of carbon monoxide, J. Geophys. Res., 109, D23306, doi:10.1029/2004JD005185, 2004.

Hedelius, J. K., He, T.-L., Jones, D. B. A., Baier, B. C., Buchholz, R. R., De Mazière, M., Deutscher, N. M., Dubey, M. K., Feist, D. G., Griffith, D. W. T., Hase, F., Iraci, L. T., Jeseck, P., Kiel, M., Kivi, R., Liu, C., Morino, I., Notholt, J., Oh, Y.-S., Ohyama, H., Pollard, D. F., Rettinger, M., Roche, S., Roehl, C. M., Schneider, M., Shiomi, K., Strong, K., Sussmann, R., Sweeney, C., Té, Y., Uchino, O., Velazco, V. A., Wang, W., Warneke, T., Wennberg, P. O., Worden, H. M., and Wunch, D.: Evaluation of MOPITT Version 7 joint TIR–NIR XCO retrievals with TCCON, Atmos. Meas. Tech., 12, 5547–5572, https://doi.org/10.5194/amt-12-5547-2019, 2019.

Ho, S. P., Edwards, D. P., Gille, J. C., Chen, J., Ziskin, D., Francis, G. L., ... & Drummond, J. R. (2005). Estimates of 4.7 μm surface emissivity and their impact on the retrieval of tropospheric carbon monoxide by Measurements of Pollution in the Troposphere (MOPITT). Journal of Geophysical Research: Atmospheres, 110(D21).

Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.-H., Razinger, M., Remy, S., Schulz, M., and Suttie, M.: The CAMS reanalysis of atmospheric composition, Atmos. Chem. Phys., 19, 3515-3556, https://doi.org/10.5194/acp-19-3515-2019, 2019.

Jiang, Z., Worden, J. R., Jones, D. B. A., Lin, J.-T., Verstraeten, W. W., and Henze, D. K.: Constraints on Asian ozone using Aura TES, OMI and Terra MOPITT, Atmos. Chem. Phys., 15, 99-112, https://doi.org/10.5194/acp-15-99-2015, 2015.

Jiang, Z., McDonald, B. C., Worden, H., Worden, J. R., Miyazaki, K., Qu, Z., Henze, D. K., Jones, D. B. A., Arellano, A. F., Fischer, E. V., Zhu, L. Y., and Boersma, K. F.: Unexpected slowdown of US pollutant emission reduction in the past decade, P. Natl. Acad. Sci. USA, 115, 5099–5104, https://doi.org/10.1073/pnas.1801191115, 2018.

Judd, L. M., Al-Saadi, J. A., Janz, S. J., Kowalewski, M. G., Pierce, R. B., Szykman, J. J., Valin, L. C., Swap, R., Cede, A., Mueller, M., Tiefengraber, M., Abuhassan, N., and Williams, D.: Evaluating the impact of spatial resolution on tropospheric NO2 column comparisons within urban areas using high-resolution airborne data, Atmos. Meas. Tech., https://doi.org/10.5194/amt-2019-161, 2019.

Kanakidou, M., Mihalopoulos, N., Kindap, T., Im, U., Vrekoussis, M., Dermitzaki, E., Gerasopoulos, E. , Unal, A., Koc¸ak, M., Markakis, K., Melas, D., Youssef, A. F., and Moubasher, H.: Megacities as hot spots of air pollution in the East Mediterranean, doi:10.1016/j.atmosenv.2010.11.048, 2011.

Kar, J., Deeter, M. N., Fishman, J., Liu, Z., Omar, A., Creilson, J. K., Trepte, C. R., Vaughan, M. A., and Winker, D. M.: Wintertime pollution over the Eastern Indo-Gangetic Plains as observed from MOPITT, CALIPSO and tropospheric ozone residual data, Atmos. Chem. Phys., 10, 12273-12283, https://doi.org/10.5194/acp-10-12273-2010, 2010.

Kopacz, M., Jacob, D. J., Henze, D. K., Heald, C. L., Streets, D. G., and Zhang, Q.: Comparison of adjoint and analytical Bayesian inversion methods for constraining Asian sources of carbon monoxide using satellite (MOPITT) measurements of CO columns, J. Geophys. Res., 114, D04305, doi:10.1029/2007JD009264., 2009.

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), Atmos. Chem. Phys., 10, 855-876, https://doi.org/10.5194/acp-10-855-2010, 2010.

Kumar, R., Naja, M., Pfister, G. G., Barth, M. C., Wiedinmyer, C., and Brasseur, G. P.: Simulations over South Asia using the Weather Research and Forecasting model with Chemistry (WRF-Chem): chemistry evaluation and initial results, Geosci. Model Dev., 5, 619–648, https://doi.org/10.5194/gmd-5-619-2012, 2012.

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, Geosci. Model Dev., 5, 369–411, https://doi.org/10.5194/gmd-5-369-2012, 2012.

Lamsal, L. N., Krotkov, N. A., Celarier, E. A., Swartz, W. H., Pickering, K. E., Bucsela, E. J., Gleason, J. F., Martin, R. V., Philip, S., Irie, H., Cede, A., Herman, J., Weinheimer, A., Szykman, J. J., and Knepp, T. N.: Evaluation of OMI operational standard NO2 column retrievals using in situ and surface-based NO2 observations, Atmos. Chem. Phys., 14, 11587-11609, https://doi.org/10.5194/acp-14-11587-2014, 2014.

Landgraf, J., aan de Brugh, J., Scheepmaker, R., Borsdorff, T., Hu, H., Houweling, S., Butz, A., Aben, I., and Hasekamp, O.: Carbon monoxide total column retrievals from TROPOMI shortwave infrared measurements, Atmos. Meas. Tech., 9, 4955–4975, https://doi.org/10.5194/amt-9-4955-2016, 2016.

Li, L. and Liu, Y.: Space-borne and ground observations of the characteristics of CO pollution in Beijing, 2000–2010, Atmos. Environ., 45, 2367, doi:10.1016/j.atmosenv.2011.02.026, 2011.

Pan, L., Gille, J. C., Edwards, D. P., Bailey, P. L., & Rodgers, C. D. (1998). Retrieval of tropospheric carbon monoxide for the MOPITT experiment. Journal of Geophysical Research: Atmospheres, 103(D24), 32277-32290.

Tang, W., Arellano, A. F., DiGangi, J. P., Choi, Y., Diskin, G. S., Agustí-Panareda, A., Parrington, M., Massart, S., Gaubert, B., Lee, Y., Kim, D., Jung, J., Hong, J., Hong, J.-W., Kanaya, Y., Lee, M., Stauffer, R. M., Thompson, A. M., Flynn, J. H., and Woo, J.-H.: Evaluating high-resolution forecasts of atmospheric CO and CO2 from a global prediction system during KORUS-AQ field campaign, Atmos. Chem. Phys., 18, 11007–11030, https://doi.org/10.5194/acp-18-11007-2018, 2018.

Tang, W., Arellano, A. F., Gaubert, B., Miyazaki, K., and Worden, H. M.: Satellite data reveal a common combustion emission pathway for major cities in China, Atmos. Chem. Phys., 19, 4269–4288, https://doi.org/10.5194/acp-19-4269-2019, 2019.

Toon, O. B., Maring, H., Dibb, J., Ferrare, R., Jacob, D. J., Jensen, E. J., Luo, Z. J., Mace, G. G., Pan, L. L., Pfister, L., and Rosenlof, K. H.: Planning, implementation, and scientific goals of the Studies of Emissions and Atmospheric Composition, Clouds and Climate Coupling by Regional Surveys (SEAC4RS) field mission, J. Geophys. Res.-Atmos., 121, 4967–5009, doi:10.1002/2015JD024297, 2016.

Wang, F., Li, Z., Ren, X., Jiang, Q., He, H., Dickerson, R. R., Dong, X., and Lv, F.: Vertical distributions of aerosol optical properties during the spring 2016 ARIAs airborne campaign in the North China Plain, Atmos. Chem. Phys., 18, 8995–9010, https://doi.org/10.5194/acp-18-8995-2018, 2018.

Worden, H. M., Deeter, M. N., Edwards, D. P., Gille, J. C., Drummond, J. R., and Nédélec, P.: Observations of near-surface carbon monoxide from space using MOPITT multispectral retrievals, J. Geophys. Res., 115, D18314, https://doi.org/10.1029/2010JD014242, 2010.

Worden, H. M., Deeter, M. N., Edwards, D. P., Gille, J., Drummond, J., Emmons, L. K., Francis, G., and Martínez-Alonso, S.: 13 years of MOPITT operations: lessons from MOPITT retrieval algorithm development, Ann. Geophys., 56, 1–5, https://doi.org/10.4401/ag-6330, 2014.

Yurganov, L. N., Duchatelet, P., Dzhola, A. V., Edwards, D. P., Hase, F., Kramer, I., Mahieu, E., Mellqvist, J., Notholt, J., Novelli, P. C., Rockmann, A., Scheel, H. E., Schneider, M., Schulz, A., Strandberg, A., Sussmann, R., Tanimoto, H., Velazco, V., Drummond, J. R., and Gille, J. C.: Increased Northern Hemispheric carbon monoxide burden in the troposphere in 2002 and 2003 detected from the ground and from space, Atmos. Chem. Phys., 5, 563-573, https://doi.org/10.5194/acp-5-563-2005, 2005.

Table 1. In-situ datasets of CO used for MOPITT products validation in this study.

	Period	Region	Number of profiles	Number of profiles over urban	Technique	Uncertainty	Reference
DISCOVER-AQ MD	Jul, 2011	Baltimore-Washington, D.C., US	80	36	NASA DACOM		
DISCOVER-AQ CA	Jan-Feb, 2013	California, US	35	12	NASA DACOM	Precision < 1% or 0.1 ppbv; Accuracy 2%	https://www-air.larc.nasa.gov/missions/discover-aq/
DISCOVER-AQ TX	Sep, 2013	Texas, US	61	37	NASA DACOM		
DISCOVER-AQ CO	Jul-Aug, 2014	Colorado, US	56	36	NASA DACOM		
SEAC^{4}RS	Aug-Sep, 2013	US	15	1	NASA DACOM	Precision < 1% or 0.1 ppbv; Accuracy 2%	Toon et al. (2016)
A-FORCE	Mar-Apr, 2009; Feb-Mar, 2013; Jun-Jul, 2013	Japan, South Korea, Pacific Ocean	45	2	AL5002, Aero-Laser GmbH	Precision ~0.5%; Accuracy 2%	Oshima et al. (2012); Kondo et al. (2016)
KORUS-AQ	May-Jun, 2016	South Korea	47	6	NASA DACOM	Precision < 1% or 0.1 ppbv; Accuracy 2%	Al-Saadi et al. (2015)
ARIAs	May-Jun, 2016	Hebei, East China	19	3	Picarro G2401-m	Precision of ±4 ppbv	Wang et al. (2018)

The CO scale used for SEAC^{4}RS, and DISCOVER-AQ MD, TX, and CA is WMO-CO-X2004, while the CO scale used for ARIAs and KORUS-AQ is WMO-CO-X2014A.

Formatted Table
Deleted: DC
Formatted: Font:10 pt
Formatted: Font:10 pt, Not Bold
Formatted: Font:10 pt, Not Bold
Formatted: Font:10 pt, Not Bold
Deleted: i
Formatted: Font:10 pt, Not Bold
Formatted: Justified
Formatted: Font:10 pt
Formatted: Font:10 pt, Not Bold
Formatted: Font:10 pt
Deleted:
Formatted: Font:10 pt, Not Bold
Formatted: Font:10 pt

Table 2. Summarized validation results for V7 and V8 TIR-only (V7T and V8T), NIR-only (V7N and V8N) and TIR-NIR (V7J and V8J) products based on in-situ profiles from DISCOVER-AQ, SEAC^{4}RS, A-FORCE, KORUS-AQ, and ARIAs.

		Surface layer			800 hPa layer			600 hPa layer		
		All	Urban	Non-urban	All	Urban	Non-urban	All	Urban	Non-u
V7T	Bias (%)	0.1	-1.7	1.1	0.8	-0.6	1.7	4.0	3.9	4.0
	SD (%)	9.5	8.6	9.8	11.0	9.0	11.9	11.4	9.0	12.7
	r	0.71	0.67	0.72	0.66	0.65	0.66	0.63	0.58	0.64
V8T	Bias (%)	2.0	0.9	2.7	2.2	1.4	2.7	3.5	3.5	3.5
	SD (%)	9.3	9.6	9.0	10.7	9.7	11.2	11.7	10.0	12.6
	r	0.70	0.58	0.75	0.66	0.58	0.69	0.63	0.54	0.66
V7N	Bias (%)	-2.0	-2.8	-1.5	-1.6	-2.1	-1.1	-1.6	-1.9	-1.3
	SD (%)	6.7	6.4	6.9	5.7	5.2	6.0	4.3	4.2	4.4
	r	0.62	0.54	0.67	0.56	0.45	0.61	0.61	0.48	0.68
V8N	Bias (%)	1.4	0.4	2.2	1.6	0.9	2.1	1.2	0.8	1.5
	SD (%)	6.9	6.7	6.9	6.0	5.8	6.1	4.6	4.7	4.5
	r	0.60	0.52	0.67	0.54	0.40	0.62	0.59	0.42	0.68
V7J	Bias (%)	-5.4	-8.9	-3.3	-3.9	-6.5	-2.4	-0.5	-1.4	0.1
	SD (%)	13.5	12.1	13.9	14.2	12.4	15.0	13.6	11.0	14.8
	r	0.68	0.63	0.70	0.64	0.58	0.66	0.60	0.52	0.62
V8J	Bias (%)	0.0	-2.0	1.1	-0.7	-1.6	-0.1	-0.5	-0.8	-0.3
	SD (%)	12.7	13.7	12.0	12.9	12.5	13.1	12.8	10.9	13.8
	r	0.69	0.53	0.76	0.69	0.57	0.73	0.65	0.53	0.67

Deleted:

Deleted:

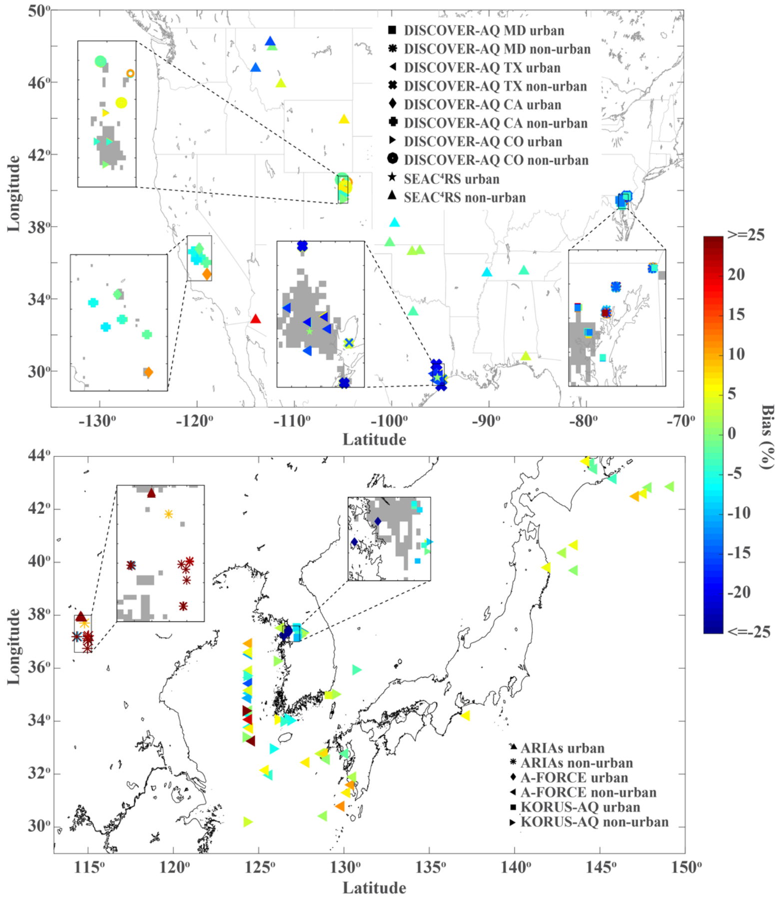
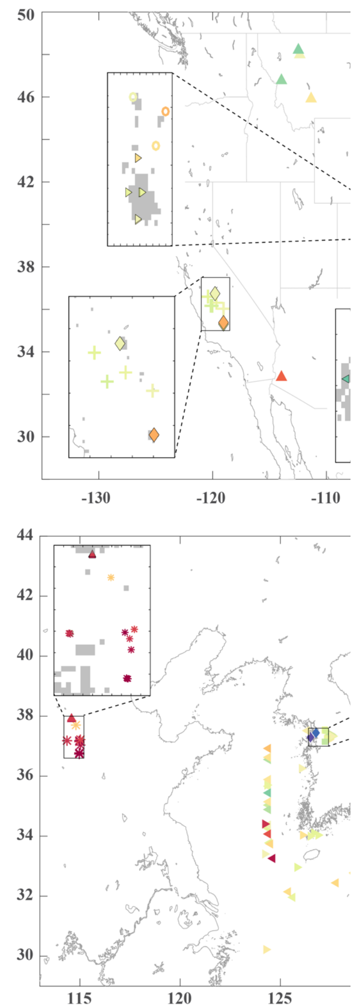


Figure 1. Spatial distributions of aircraft profiles from the DISCOVER-AQ, SEAC^{4}RS, ARIAs, A-FORCE, and KORUS-AQ campaigns. Urban and built-up land cover (from MCD12C1 v006) are shown by gray shade in the boxes. Biases of MOPITT V8J comparing to the aircraft profile at the surface layer are shown by the color of the profile.



Deleted:

Formatted: Centered

Deleted: level

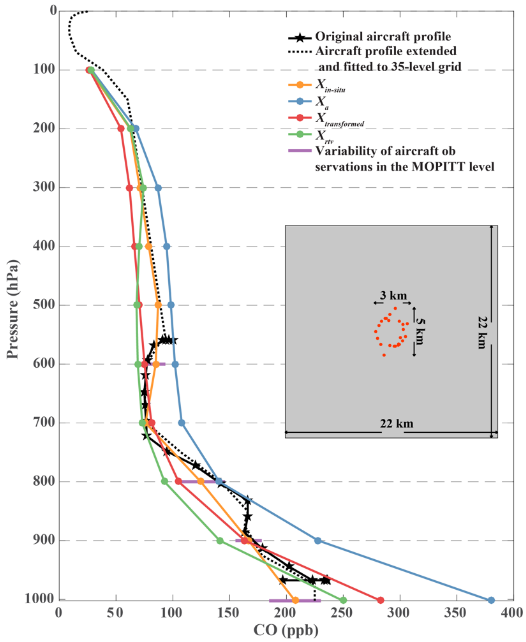


Figure 2. Example of profile comparisons for an aircraft profile sampled on July 22, 2011 during DISCOVER-AQ MD. The black solid line represents the original aircraft profile and the stars represent the original aircraft observations, the black dotted line is the aircraft profile extended with CAMS reanalysis data, and regridded to 35-level grid. The in-situ profile regridded at 10-level grid ($x_{in-situ}$), the MOPITT a priori profile (x_a), the in-situ profile transformed with the MOPITT a priori and AK ($x_{transformed}$), and the MOPITT retrieved profile (x_{rtv}) are shown in colored lines with dots. The purple bars centered at the $x_{in-situ}$ at each MOPITT retrieval level show the standard deviations of the original aircraft observations in the MOPITT layer. Note that each MOPITT retrieval level corresponds to a uniform layer immediately above that level. Superimposed gray box shows the horizontal scale of the profile (each aircraft observation is represented by a red dot) and a MOPITT pixel (gray box).

Deleted: DC

Deleted: vertical and horizontal variability

Deleted: , indicated by standard deviation

Deleted: ly-weighted

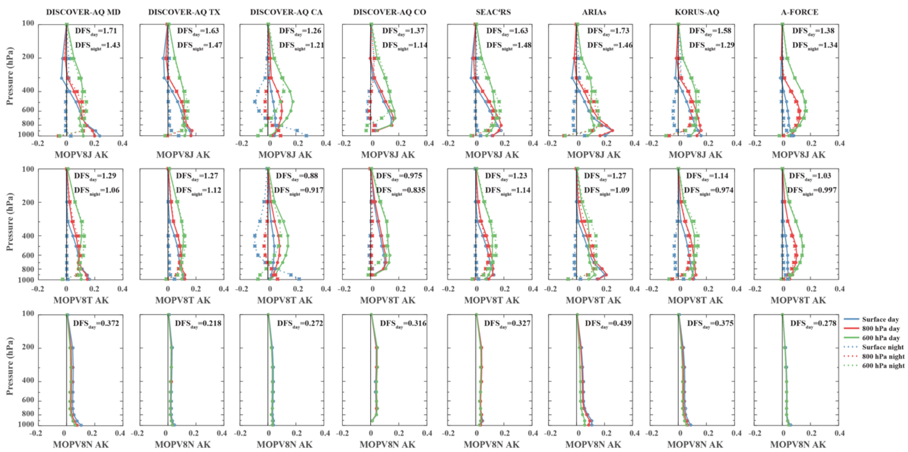


Figure 3. Mean retrieval averaging kernels for the MOPITT V8J, V8T, and V8N for the corresponding in-situ profiles from the DISCOVER-AQ, SEAC[4]RS, ARIAs, KORUS-AQ, and A-FORCE at daytime (solid lines) and nighttime (dashed lines).

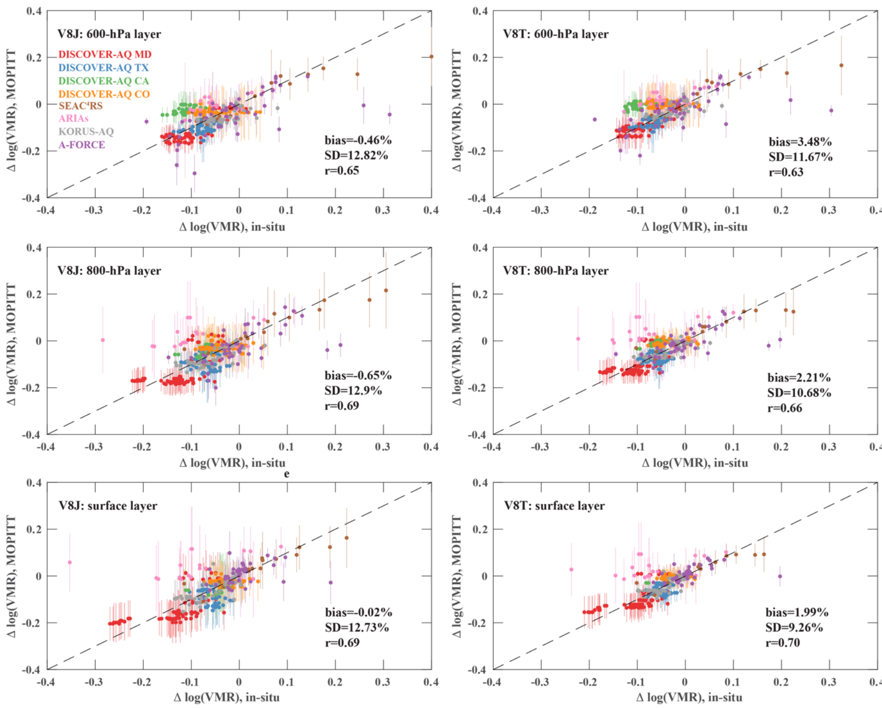


Figure 4. MOPITT V8J and V8T validation results over both urban and non-urban regions at 600-hPa layer, 800-hPa layer, and the surface layer in terms of $\Delta\log_{10}$(VMR). $\Delta\log_{10}$(VMR) is defined as $x_{rtv}-x_a$ for MOPITT profiles and $x_{transformed}-x_a$ for the in-situ profiles. The use of $\Delta\log_{10}$(VMR) allows us to remove the impact of the a priori in the comparisons. The variability of the MOPITT data used to calculate each of the plotted mean values are represented by the vertical error bars. The dashed lines are one-to-one ratio lines.

Deleted:

Deleted:

Deleted: log

Deleted:

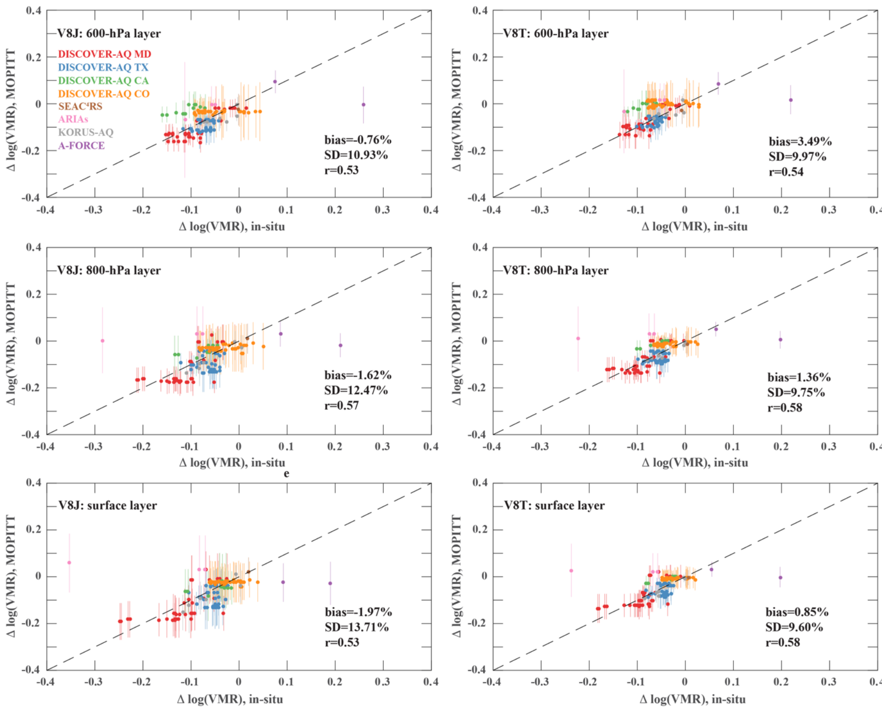


Figure 5. MOPITT V8J and V8T validation results against aircraft profiles over urban regions at the 600-hPa layer, the 800-hPa layer, and the surface layer in terms of Δlog (VMR). The dashed lines are one-to-one ratio lines. See the caption of Figure 4.

Deleted:

Deleted:

Deleted: See caption to Figure 2

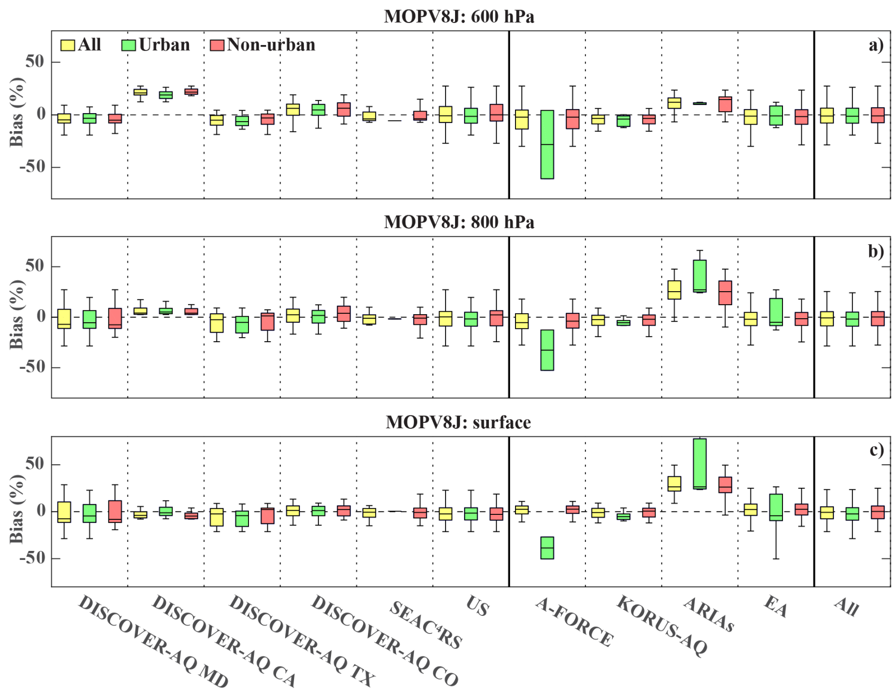


Figure 6. Boxplot (with medians represented by middle bars, interquartile ranges between 25th and 75th percentiles represented by boxes, and the most extreme data points not considered outliers represented by whiskers) for biases (%) for the profiles over both urban and non-urban regions (yellow), profiles over urban regions (green), and profiles over non-urban regions (red) at 600-hPa layer (panel a), 800-hPa layer (panel b), and the surface layer (panel c). An outlier is a value that is more than 1.5 times the interquartile range away from the top or bottom of the box.

Deleted:

Deleted:

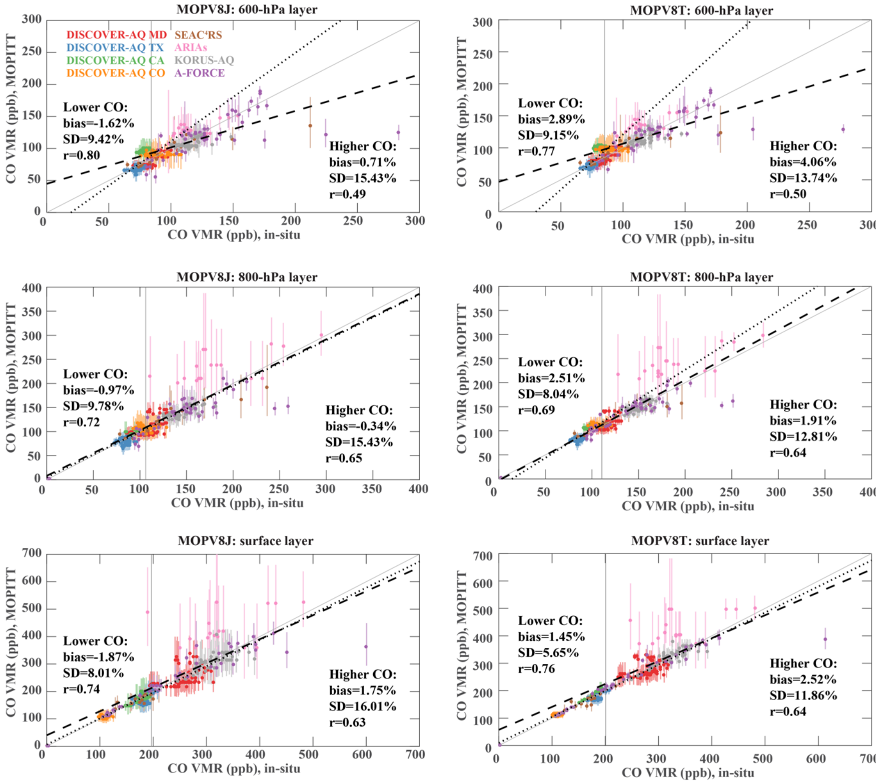


Figure 7. MOPITT V8J and V8T validation results at 600-hPa layer, 800-hPa layer, and the surface layer against the lower 50% in-situ profiles of CO and higher 50% in-situ profiles of CO. The variability of the MOPITT data used to calculate each of the plotted mean values are represented by the vertical error bars. Each panel shows the least-squares best-fit lines for the lower 50% CO concentrations (dotted line) and the higher 50% CO concentrations (dashed line).

Deleted:

Deleted:

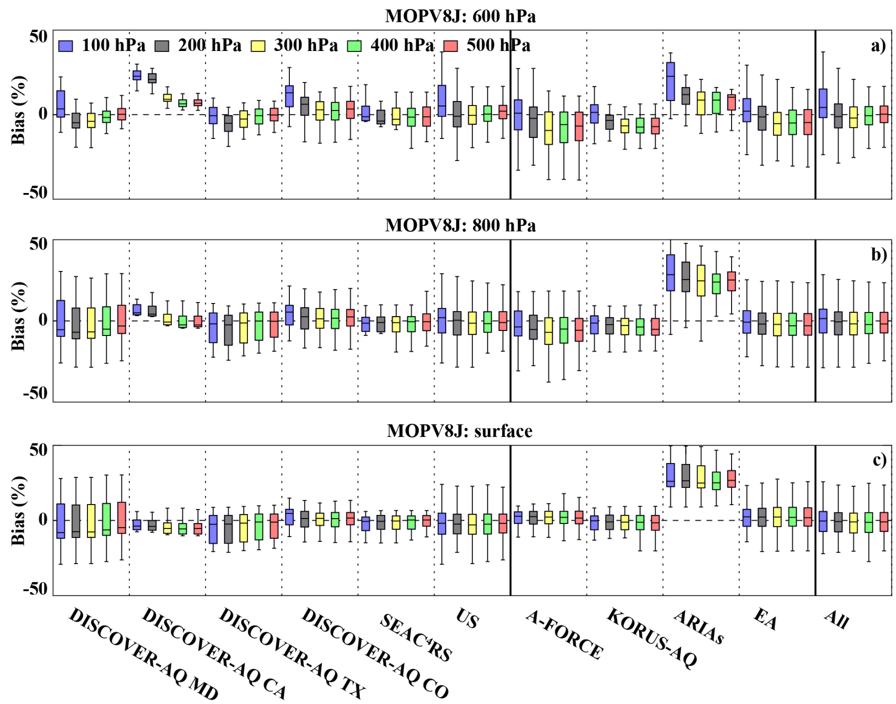


Figure 8. Sensitivity to P_{interp}. Biases (%) using 100 hPa (blue), 200 hPa (gray), 300 hPa (yellow), 400 hPa (green), and 500 hPa (red) as P_{interp} at 600-hPa layer (panel a), 800-hPa layer (panel b), and the surface layer (panel c) are shown by boxplot (with medians represented by middle bars, interquartile ranges between 25th and 75th percentiles represented by boxes, and the most extreme data points not considered outliers represented by whiskers). The biases are calculated against all (both urban and non-urban) in-situ profiles listed in Table 1. The "200 hPa" values (gray) in are the same as yellow values (for all data) in Figure 6. See the caption of Figure 6 for the definition of outliers.

Deleted:

Deleted:

Deleted:

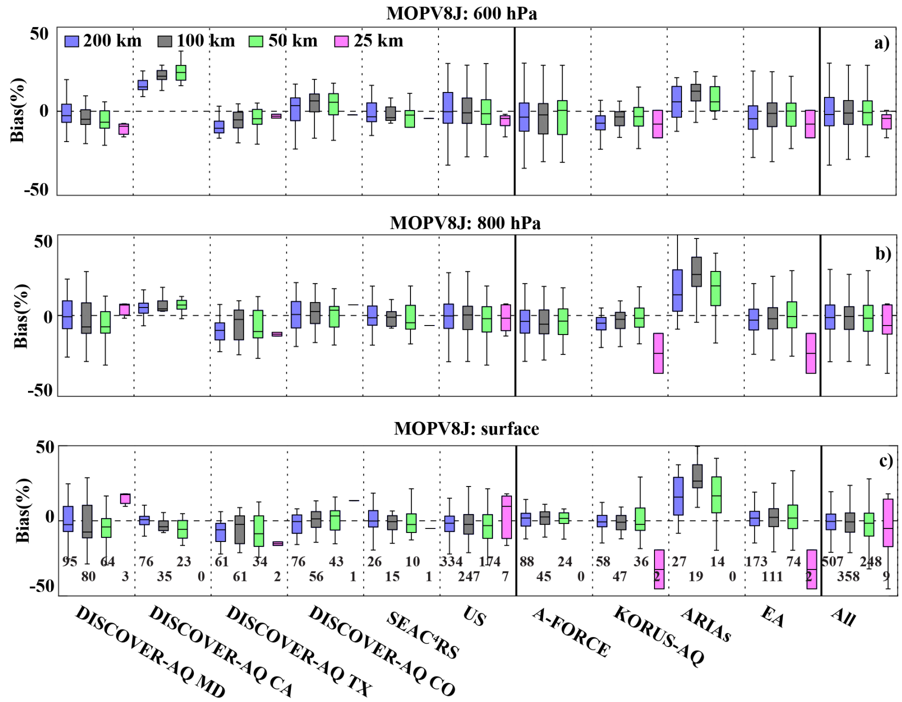
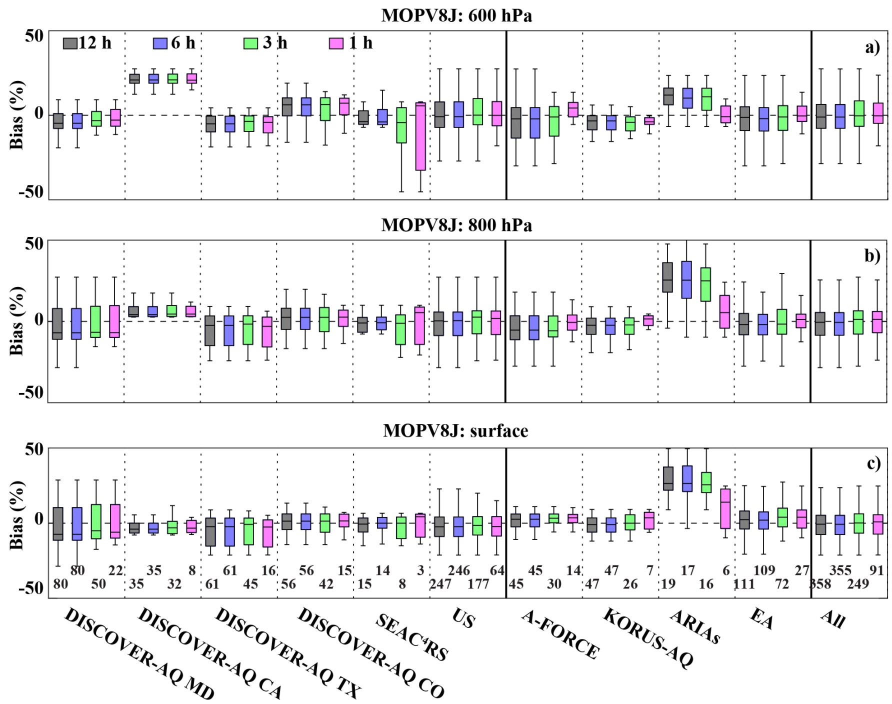
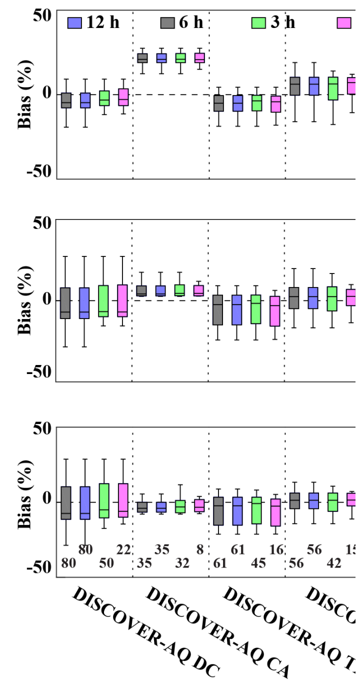


Figure 9. Sensitivity to the radius as criteria for co-location. Biases (%) using 200 km (blue), 100 km (gray), 50 km (green), and 25 km (pink) as the radius for co-location at 600-hPa layer (panel a), 800-hPa layer (panel b), and the surface layer (panel c) are shown by boxplot (with medians represented by middle bars, interquartile ranges between 25th and 75th percentiles represented by boxes, and the most extreme data points not considered outliers represented by whiskers). The numbers in panel c correspond to the number of in-situ profiles qualified for validation within the given radius. The biases are calculated against all (both urban and non-urban) in-situ profiles listed in Table 1. The "100 km" values (gray) are the same as yellow values (for all data) in Figure 6. See the caption of Figure 6 for the definition of outliers.

Deleted:

Deleted:





Deleted:

Figure 10. Sensitivity to the allowed maximum time difference as criteria for co-location. Biases (%) using 12 hour (gray), 6 hour (blue), 3 hour (green), and 1 hour (pink) as the allowed maximum time difference for co-location at 600-hPa layer (panel a), 800-hPa layer (panel b), and the surface layer (panel c) are shown by boxplot (with medians represented by middle bars, interquartile ranges between 25th and 75th percentiles represented by boxes, and the most extreme data points not considered outliers represented by whiskers). The numbers in panel c correspond to the number of in-situ profiles qualified for validation within the given allowed maximum time difference. The biases are calculated against all (both urban and non-urban) in-situ profiles listed in Table 1. the “12 h” values (gray) are the same as yellow values (for all data) in Figure 6. See the caption of Figure 6 for the definition of outliers.

Deleted: blue

Deleted: gray

Deleted:

Deleted:

Deleted: Page Break

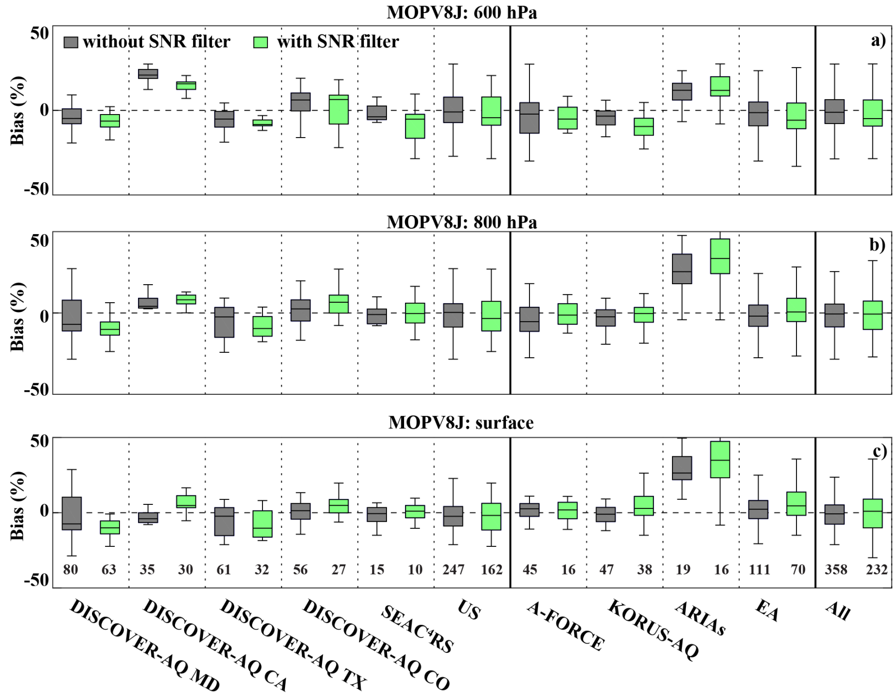


Figure 11. Sensitivity to the signal-to-noise ratio (SNR) filters. Biases (%) for MOPITT retrievals without SNR filters (gray), and MOPITT retrievals with SNR filters (green) at 600 hPa layer (panel a), 800 hPa layer (panel b), and the surface layer (panel c) are shown by boxplot (with medians represented by middle bars, interquartile ranges between 25th and 75th percentiles represented by boxes, and the most extreme data points not considered outliers represented by whiskers). The numbers in panel c correspond to the number of in-situ profiles qualified for validation without or with SNR filters. The biases are calculated against all (both urban and non-urban) in-situ profiles listed in Table 1. the "without SNR filter" values in are the same as yellow values (for all data) in Figure 6. See the caption of Figure 6 for the definition of outliers.

Page 7: [1] Deleted	Wenfu Tang	1/22/20 5:16:00 PM

For each pair of co-located MOPITT retrieval and in-situ profiles, we apply the MOPITT a priori profile and averaging kernel to the in-situ profile,

$$x_{transformed} = x_a + A(x_{in-situ} - x_a) \quad (2)$$

so that the transformed in-situ profile ($x_{transformed}$) has the same degree of smoothing and a priori dependence as the MOPITT profile.