

Review of Hedelius et al., 2019. "Evaluation of MOPITT version 7 joint TIR-NIR XCO retrievals with TCCON"

This paper gives a relatively extensive comparison of MOPITT column-average CO (XCO) retrievals with (supposedly more accurate & precise) XCO values from TCCON. The paper is careful to make averaging kernel corrections and correct for a common prior. The paper proposes additional filtering for things like SNR, and proposes a 2-tiered bias correction, an individual correction for each of the four MOPITT pixels, and an over-arching feature-based correction over land, which is related to how "wiggly" the retrieved profile is (the more wiggly, the larger the correction". The paper is chock full of equations, careful tests, and statistical results. It also includes a brief comparison of assimilating the standard MOPITT XCO product into a CO flux inversion system, vs. using the proposed filtering and bias correction.

In general, the paper is well-written. The grammar and spelling are both nearly flawless, for which this reviewer is extremely grateful. However, to some degree this paper is extremely detail-heavy and short on high-level take-aways. The paper could be enhanced by adding some additional high-level conclusions in terms of a) which areas the filters and bias correction help most in, and (b) the real impact of the proposed filters and bias correction in terms of inverse flux modelers at both local scales, as well as regional/global scales. (Ie, do they matter more in the tropics, or high latitudes; more over land vs. ocean, etc). Beyond this limited suggestion, the paper is nearly ready for publication with only minor revisions necessary.

- Sect 2.2 (TCCON), page 6 : Please state the estimated possible mean biases per site (in ppb), and variable errors (noise plus faster systematic errors). You currently only list this single 4% number, which presumably corresponds to the latter.
- Do you consider collocation errors with TCCON, or their possible size? For CO from local emission sources, it seems like collocation errors could be large, depending on your collocation criteria.
- Section 2.3: Please state something about the accuracy and precision of AirCore CO measurements.
- Phase "truth metric" seems an imperfect term. I would prefer truth estimate or truth proxy. "metric" doesn't convey the imperfection of the actual data sources you will use.
- S3, SM, and associated text: What was the source of data you used to construct these plots? Presumably some high-resolution model?
- TCCON-MOPITT trends over the 13 years?

- Figure S4 (filter plot):
 - How can 108.13% of points pass? This problem also exists in S5.

- Why do you show the MODIS snow/ice flag if it passes everything and appears to do nothing?
 - SNR, plot c, consider plotting SNR on a log scale so we can see the behavior at low SNR more clearly.
 - Plot g, why not also have an upper limit on $T_{sfc} - T_{sfc,a}$?
 - Also in Table 3. Please state what the three $tr(S)$ values are, physically, including units.
 - Sum(Reetr anom diag) – why is there no histogram?
 - Do these filter plots include any bias correction (pixel and/or feature)? Why or why not?
- Figure S5
 - Please fix/add histograms for panels b and d.
 - Given the strong bias and scatter driven by “max diff btwn adj levels”, panel j, why not more strongly filter on this variable? Based on the criteria you set in section 3.3, (2 ppb bias, 6 ppb scatter), it seems like a reasonable cut-off would be more like 150 ppb (over land), rather than 300 ppb.
- Page 14, end of section 3: please describe the “maximum difference between adjacent levels” variable physically, and why it might be correlated with bias.
 - Section 4.1 - please state if there a noticeable improvement when using the altitude adjustment in comparing to TCCON, in particular for the highest altitude sites, or sites in mountainous regions.
 - Page 19, please state at what level you can rule out a long-term trend between MOPITT & TCCON, at for example 95% confidence (2 sigma). 0.1 %/yr ? 0.01 %/yr? Etc.
 - Page 20, top: Are any of the observed seasonal biases mitigated using your 1-parameter “feature” bias correction?
 - Page 21 bottom to page 22, top: I’m not sure what the DOF and information content analysis is adding. You state numbers but not something that we qualitatively learned. Suggesting removing, or add a sentence or two to true to put the numbers into some kind of context for what we actually learn from them.
 - Page 23, top: Because you only assimilate MOPITT, it’s not clear what the difference is between assimilations 2 and 3. Are the bias corrections different between the two, i.e. one is still biased high by 6% relative to (bias-corrected) TCCON, and one is not? Please state what the purpose is to have three tests rather than just two. Ie, you are proposing a new filtering and bias correction relative to the standard MOPITT product, so the logical thing to do here would be to test those two things. You are testing three things, so please state clearly why.
 - “Assimilation 2 results are lower.” Lower in terms of what, inferred CO fluxes? Please clarify.

- Page 24: Finally, in your comparison to HIPPO, it's not clear if any assimilation is clearly "better" in terms of comparison to HIPPO to any other. Please discuss this somewhere in the paper. Currently, your readers cannot tell if your new filtering and bias correction made any real improvement to the MOPITT data in terms of inferred fluxes.
- Appendix D & Section 4.4: In terms of the need to adjust to common priors, and correct for AK differences you never state (1) the general size of these adjustments (in % or ppb), or (2) the effect on the apparent quality of the TCCON vs. MOPITT comparison. Does the comparison improve when you employ one or both of these corrections? You have gobs of equations and figures but not just a sentence or two saying something like "In general, these corrections are critically important in order to obtain meaningful results" or "In general, these corrections are small and may be in practice ignored", or something in between. Sometimes readers want to skip the equations and numbers and get to the main point!
- Table D3: what are the units? Ppb? % ?

Grammar/Technical:

Page 8: "As a second example, empirical corrections" - please add comma after the leading clause.

Page 9, "To examine the effects of averaging kernels" - please add comma after leading clause.

Page 12: "To reduce the likelihood of overfitting, " - please add comma after leading clause.

Page 13: What is the approximate size of the AK correction typically? Can you apply the AK correction for all MOPITT soundings as compared to TCCON?