Review comments on the preprint: „A comparison of carbon monoxide retrievals between the MOPITT satellite and Canadian High-Arctic ground-based NDACC and TCCON FTIR measurements" by Ali Jalali et al.

General comments:

The manuscript compares MOPITT CO v8 retrievals from 2006-2019 with TCCON and NDACC FTIR observations from the PEARL Eureka site. MOPITT observations within 110km and for 24h intervals are compared. The authors thoroughly describe the methodologies for comparing space-based total column retrievals with those ground-based networks. Beside the pixel-to-pixel bias the work examines and compares MOPITT NIR, TIR and NIR-TIR retrievals with the ground-based site and provides comprehensive of statistics. The results and improvements of v8 are thoroughly discussed and described.

- consider to mention that the fundamental and first overtone of CO are relevant in the TIR/NIR spectral region

- do I get this right: each channel has four pixles, and each pixel in a channel sounds a certain part of a single CO line? if so does that mean that the inner pixel record transmittance from the line and the outer pixel from the line wing? why not sensing mutliple CO lines?

Specific comments:

Page3, line 2:
Consider to include the MAPS mission aboard the space shuttle (e.g. Reichle 1999)

p3,18:
Does NDACC only measure absorption of solar radiation or does the network also measure emission from the atmosphere?

p4,21: Maybe add one sentence why log(vmr) is fitted?

p4,23: One or two words explaining the ‚outer' and ‚inner' pixels meaning (line centre vs line wing?). Also see general comment above.

p5,33 and Fig.1:
Just for confirmation, so the weighted average of all pixels means that each pixel's average within the 30day/100km was multiplied by 0.25 and then summed?

p5, Fig1:
Is a joint TIR-NIR retrieval possible over water? does the signal then basically only come from the TIR channel?

p9,4:
Maybe add one sentence on how well the assumption for the small-area approximation is fulfilled at the Eureka site (to justify the chosen 1° radius criteria).

p9,16 and Fig. 4:
It is mentioned (in p13,30) that DOFS represent the information content of the retrieval, so why do high DOFS lead to large delta xCO?
Does it mean that the priors are already close to the true values?
Are low DOFS retrievals trustworthy fits?
Please clarify the meaning of DOFS for the present work.

p10, Fig. 4:
Why are the SZA values ranging from 60-120°? I would expect it from 60-90°.
Also check p16,5 which designates daytime measurements (SZA < 90°).
Note that the figure caption says the RMS is represented in blue when it actually is yellow.

p12,20:
Is there a reason why NDACC is not using 4 times daily (6 hourly) NCEP data?

p13, Sec.4:
Consider to include a sentence which states that the total column averaging kernel is computed by a total column operator (converts profile concentrations to a column concentration, see Deeter 2002)

p18,3:
what is the motivation for 500 grid points per MOPITT layer? just to be dense enough?

p18,13:
What is the motivation for selecting a critical difference of 80hPa?
is it proposed somewhere (in Kerzenmacher et al. 2012?)

p18,27:
Consider to explain why MOPITT retrievals are in log space (see previous comment p4,21)

p19,,11:
If TCCON method IV from Hedelius is similar to the method used for NDACC (more consistent?), why not using it instead of method II?

p20, Eq.(10):
Why is the number of measurements N the same for the MOPITT and FTIR (because of the averaged MOPITT measurements?)

p22,5:
is the larger correlation to NDACC caused by the TIR interval?
is the Fig. 9 plot also available for MOPITT NIR?

p29, Sec. 6.4:
likely different co-location and filter criteria across studies are responsible for some of the difference (as you mention three times more V8 comparisons than V6 in Buchholz).
Consider mentioning the effect of coincide criteria w.r.t. Hedelius 2019 in paragraph p29,12.

Technical corrections:

p31, Table 3:
Is the entry for the `V7 NOAA all sites` correlation coefficient a typo?

Write text in equations upright in order to discriminate it from variables.