Comment on amt-2021-67

Corinne Vigouroux (Referee)

Referee comment on "Improved ozone monitoring by ground-based FTIR spectrometry" by
Omaira Elena García et al., Atmos. Meas. Tech. Discuss.,
https://doi.org/10.5194/amt-2021-67-RC2, 2021

General comments

The study of García et al. (2021) examines the performance of different O_3 retrieval
strategies from FTIR (Fourier Transform InfraRed) spectrometry at the subtropical Izaña
site. In particular, it studies the effect of the spectral region used for O_3 retrievals and of
the inclusion of an atmospheric temperature profile fit, which is of high interest for the
whole IRWG (Infra-Red Working Group) of NDACC (Network for the Detection of
Atmospheric Composition Change) that aims at providing the best possible O_3 product.
The quality assessment of the different FTIR O_3 products (total columns and profiles) is
carefully led, both theoretically and experimentally by comparing with Brewer and sondes
coincident measurements.

Therefore, I recommend the publication of this paper in AMT, after a few
comments/suggestions and questions (listed below) are addressed.

Specific comments:

- Section Introduction, l. 47 "However, others are still flexible and station-dependent
(e.g. the inclusion of a temperature retrieval…"

The temperature retrieval is kept as an alternative in the document IRWG (2014), but in practice, in the NDACC archive, all sites are consistent in not doing the temperature retrieval. It would be better to specify this to not let the NDACC users think that the NDACC products are not harmonized in the choice of retrieval settings. Therefore, the current NDACC homogenization is in very good shape; mainly the ILS treatment is not harmonized. And the impact of different ILS treatment is not treated in the current work. Therefore, I would correct l.47 (and followings) accordingly.

Also, in l. 50 "..much efforts should be paid…" should be replaced by "… additional efforts could …" Although this additional effort is mainly related to ILS, and not treated here, so this statement could also be put in the conclusions as a perspective to be done in the IRWG, rather than an introduction to the present work.

However, even if the retrieval settings are well harmonized, this does not mean that they cannot be improved. So I would more emphasize the present work to be a research towards a better strategy (micro-windows; temperature retrievals) to be proposed – if proven better - to the IRWG in replacement of the current harmonized one (this presentation of the study is actually well done in the conclusions). To my opinion the present work is not improving the harmonization and the network consistency, but is pushing towards an improvement of the retrieval strategy itself (which is very valuable).

For this exercise (finding a better strategy than the current IRWG one) it would have been good to include a few more stations to prove that the conclusions at Izaña are valid at other sites as well.

- Section 2.2, Brewer and ECC sondes: when you give the uncertainty for Brewer and sondes, is it the random, systematic or total one? It should be specified for the interpretation of the comparisons with FTIR (bias, standard deviation).

- Section 3.1 Ozone retrieval strategies:

- l.159: H2O treatment: did you test to simultaneously retrieve the H2O profile in a one-step approach (as done at Lauder/ Wollongong in Vigouroux et al. 2015)? The results might be equivalent to your 2 step approach, while being more simple.

- l.165: ILS treatment

In your retrievals, the ILS is fixed to the results obtained by LINEFIT using the N2O cell-measurements. Did you try to retrieve it? (starting from LINEFIT results as a priori values), in order to e.g. improve the comparisons in the 1999-2008 periods. The LINEFIT results are also obtained with some uncertainty, and averaging kernels that do not have a full sensitivity for the whole OPD. It would be interesting to see if the results of your quality assessment (by comparing with Brewer and sondes) could be improved by fitting the ILS. This would also be an interesting result for the whole IRWG and the harmonized strategy. Could you add this test for own of your set-up (e.g. 4MWS)?

- l.168: temperature profiles

NCEP provides now 6-hourly temperature, pressure, H2O profiles. I guess that if you would use these 6-hourly profiles instead of daily means, you would decrease the effect of retrieving vs fixing your temperature profiles. And you would have a temperature covariance matrix that should have reduced values, which would decrease the uncertainties due to the fixed temperature. Why not using the best NCEP available values if it is proven that for O3 the temperature is a leading source of uncertainty?

- Section 3.2.2 Uncertainty analysis, l. 233: "where the spectroscopic SY errors determine the total uncertainty budget (with values of □5%)"

To my knowledge, the uncertainty due to O3 line intensity (dominating the systematic error on the total column) has been set to 3% in the IRWG (SFIT4 new release, agreement with PROFFIT users as well, B. Langerock, F. Hase, personal communication). This is quite in agreement with your Table 2 for the best measurement periods (2008-2018): bias with Brewer below 3.4%.

I would change this 5% value here and p. 15. L. 324.

- l.240: smoothing error

If the smoothing error is getting more important when fitting the temperature, then it is important to give total error budget with smoothing included (also in Fig.5 / Table 2). To check if it's worth fitting the temperature at the end. Decision should be made using total uncertainty, smoothing included.

- Discussion p. 14 l. 296- 307:

It looks like the extreme RD values occur mainly during the 120M measurements period. So could it simply be that T retrieval are less stable with 120M (bad ILS), and therefore gives outliers in some of the retrievals?

- p.14, l.320 & p.15 Table 2:

It would be better for the discussion to include the total statistical error in Table 2 for different set-ups / period, and/or the root-square-sum of the precision of Brewer+ FTIR. Note that the smoothing error must be included in the total budget.

- p. 15, l. 335: "inconsistency in the parametrisation of the spectroscopic parameters at higher wavenumbers"

Do you mean that at 1012 cm-1 the spectroscopic parameters linked to temperature dependence are not consistent? Did you check the origin (studies) used for the parameters in hitran? Is it different studies for 1000-1005 cm-1 and 1012 cm-1?

- p.17, L. 351: "the scatter found is noticeably lower than that predicted when the temperature fit is not considered"

Indeed. This would mean that the a-priori temperature covariance matrix (SaT), constructed following Schneider et al. (2008a), is chosen with too large uncertainty parameters (-3.5K at the surface to + 4K at 30km). This is quite an important statement since the theoretical demonstration that the temperature fit is improve the retrievals (when stable instrument) is based on this SaT matrix (which presently gives large theoretical uncertainty when T is not retrieved).

This should be recalled also in the conclusions, **p. 23, l. 445: "Theoretically, the total error of O3 TCs is halved when applying a temperature fit":** probably the effect will be less if smaller values in SaT are used (as suggested by the observed scatter).

- p. 17: discussion seasonal cycle l. 355-365:

I suggest to add scatter plots Brewer vs FTIR set-ups in Fig. 8. Offset and slope will

distinguish the constant bias between Brewer and FTIR and the proportional one (which gives a seasonal effect of RD)

- p. 18; l. 381 and p.20 Table 3: comparison at the representative altitudes of 5, 18, and 29 km:

Why not using partial columns comparisons as in García et al. (2012)? It should be more stable because the wider layers are then less dependent on the smoothing error, less dependent on the DOFS (which are quite variable, especially in the 120M period, Table 1) than a single point on the profile. Overall smaller uncertainty on wider layers than on single point profile.

This "single point" comparisons of scatter vs theoretical error budget (cf. the discussion p. 21 l.407) is also then not straightforward because the uncertainty profiles (Fig.5, Sect3.2.2.) are not independent (the covariance matrices are not diagonal).

- p.23, l.437: "Quality of the FTIR O3 products improves as the retrieval strategies become more refined by including O3 absorption lines in specific narrow micro-window"

The conclusions are less clear when comparing to sondes (p. 21, l. 420), and this should also be written in the conclusions. Probably a clearer conclusion would have helped to convince the IRWG (more than 20 sites) to re-perform their retrievals, re-archive in NDACC their data, using improved settings. The very detailed and careful analysis performed in this study could have had more fast and clear impact on the IRWG if it would have been applied to at least 1 or 2 other sites. This could have helped to strengthen the findings (humid sites for the effect of narrow mws avoiding strong H2O lines present in the broad mw; site with coincident Lidar measurements to check the effect of retrieval settings at higher altitudes, where the expected ozone recovery should be detected first). Let's hope volunteer sites will try this exercise now independently, otherwise the impact of the current study on the IRWG harmonization will be more limited.

Minor or technical comments:

- Section Introduction, l.26 and following places: "O3 measurements...."

Specify that you are talking about o3 total and/or stratospheric ozone measurements when you discuss ozone decline.

- p. 9, l .221: "do depend" instead of "do dependent"?

- p. 10, Fig. 4: "for O3 TCs...": in the text you use "O3 SC". Same remark for p.11 l. 224 & 258: SC instead of TC?

- p. 11. L. 243 and l. 254: 1000T (not 100T)

- p. 14, Figure 6: explains what represents the shaded areas (1-sigma, 2-sigma of the monthly means?). Maybe enlarge / add some grids, to better see steps / better compare (b) and (c).