Review of Twenty years of ground-based NDACC FTIR spectrometry at Izaña Observatory - overview and long-term comparison to other techniques by Garcia et al.
Anonymous Referee #1

Referee comment on "Twenty years of ground-based NDACC FTIR spectrometry at Izaña Observatory – overview and long-term comparison to other techniques" by Omaira E. García et al., Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2021-307-RC1, 2021

This is an extensive and well written review of the measurements and analyses programs in place at the Izaña Observatory over the past twenty years. It is a review paper, differing from most in that the focus is on the FTIR measurements at this site rather than a broad overview of some subject matter. This is not a problem as the measurement and science record is extensive. This paper and the references therein provide a useful compendium of the research carried out at Izaña by the FTIR group and their colleagues, and it will provide a good starting point for researchers looking to understand the context of the work carried out there. It discusses some of the more significant results and provides the references to the supporting papers should a more in-depth look be of interest to the reader. The assessment of instrument performance over the 20 years is comprehensive and fairly well presented. It is difficult to judge this paper in terms of "substantial new concepts, ideas, methods, or data" as presented herein, but nevertheless I feel the paper does have merit in presenting in one place the extensive data record and the use of that data record to address many of the more prominent areas of atmospheric research.

Some general remarks:

The paper refers to the IFS120/5 HR. Is this to denote that the instrument is of the family of the IFS120HR and the later IFS125HR, or is it to denote that it entered service as an IFS120HR and was later modified to the specification of the newer IFS125HR?

With some of the figures that have many data points, the use of circles and filled circles (dots) of similar or the same colour can be problematic. It would be better to use clearly different symbols. Looking specifically at Figure 6, the use of gray circles and black circles is not a good idea.

Specific comments:
For Figure 2, could some indication of the mean number of measurements per day (binned per month) be given?

Line 120: Was the IFS 120/5HR operated in "vented mode" at all times, or just for the time period of the comparison with the IFS 120M?

Table 2: As formatted, this table is difficult to read. Is there a difference between the Target Gas and Gas columns? Perhaps the “Gas” should also be Target Gas. It would improve readability significantly if a few more spaces or a vertical bar were added to more clearly separate the left-hand grouping of gases from the right-hand grouping.

Line 223: This would be more readable if it was written: This figure depicts the rows of the averaging kernel matrix $A$ ...

Table 3: For readability rather than spacing the columns equally, I would suggest adding more space between the columns for the combined pairs, and removing a space between the M and $\sigma$, so the first line would look like:

$$\begin{array}{cccc}
C_2H_6 & 13625 & 1.48 & 0.15 & 1.81 & 0.39 & 5.44 & 0.17 \\
\end{array}$$

Figure 3: As there are many similar averaging kernels, it would be of interest to see some explanation of why the particular kernel was selected for display in the figure.

Line 236: Are “great values” to mean those values that show large statistical uncertainty?

Line 241: maximum instead of maximal

Line 243: The systematic uncertainty budget is dominated by spectroscopic errors (instead of "led")

Figure 4: The choice of lines and lines + circles in the same colour makes the plot difficult to evaluate for some constituents. Making the plot wider would help. Also, use just symbols for those constituents that have sufficient points and are fairly "straight" such that the lines are not necessary. H$_2$CO and HCl are both shown in cyan with the former being a thinner line. I can’t distinguish a difference in thickness in the lines in the plot.

Line 250: Do the authors mean to say that the predominant errors are located in one of the troposphere, upper troposphere/lower stratosphere or middle/upper stratosphere, or that the errors are in all three regions? If the latter, they do not appear to be equally concentrated in the regions and that should be discussed.

Line 257: Particularly, high error profiles ... would read better as "large"error profiles unless the authors mean to suggest that there is also an altitude dependent component which does not seem to be the case as the lines are straight in Figure 4.

Figure 6: Caption refers to ‘grey-white dots” which is, I think, a grey circle that is not filled? See earlier comment about the use of circles and dots. Also the grey circle is easily confused with a black circle. Note also that the $R_{XCO2}$ quantity is used here but not yet defined in the text.

Line 312: I don’t think that ‘considering’ is the proper word here. I would suggest ‘using’. For example: “… using a scaling retrieval with a fixed WACCM a priori VMR profile, and PROFFIT software.”
Line 366: At the resolution the data are plotted with, it is not possible to discern a seasonal cycle in $X_{\text{air}} N_2$. If the authors consider this to be a significant finding, then overlay a trace on the data that would display it. Otherwise, the statement could be left out without weakening the section.

Line 384: “… causing punctual downward and upward shift of the UTLS region,…” The use of punctual doesn’t make much sense. Punctual implies arriving on schedule. Is that what the authors mean to say?

Line 390/Figure 7: It would be a good idea to move the HF and $N_2O$ frames to be next to each other to make the anti-correlation of the two easier to detect.

Line 559: There is a set of ellipses between CCMVal initiative and Schneider et al. Are there more missing from the list? There are 8 models listed and 10 papers referenced. It might make it clearer to list the CCM followed directly by the reference.

Lines 561-562: This is an awkward and confusing sentence and the details are lacking. There is mention of “attribute of sources/sinks”, but no comment on whether or not there is agreement. Similarly for the “representation of moist processes” evaluation. As the paper is already quite long, it might be best to make a general statement about the comparison to the CCMs and referring the reader to the above references for details rather than to try to call out these areas without further discussion.

Line 693 and below: Generally, when speaking of time the word used is coincident rather than collocation.

Line 690: It is unclear whether a actual single temporal criterion is applied here as multiple gases are being discussed. Sentence should read either “Similar temporal criteria…” or “A similar temporal criterion….” as the authors see fit.

Line 699: Over what time interval are the FTIR observations averaged? Are the total columns derived from a series of sequential spectra, averaged?

Line 703: Need a comma after “correction”

Line 706: If I understand what is being said, this might read better as “… each FTIR measurement is only paired once to the reference observation that minimizes the time difference within the temporal collocation window.” And I would use coincidence rather than collocation.

Figure 14: Presumably the “TRO” quantities refer to the tropospheric portions?

Line 785: In what manner are the PWV values over IZO “slowing down”?

Line 823: Do the authors mean to say that the NDACC FTIR product is able to capture “only” a part of its tropospheric variations?

Line 842: What is meant by “pure free conditions”?

Line 845: I suggest the wording be: Despite a considerable decrease in the number of coincidences, ...

Line 849: I suggest the wording: …which makes the comparison of the remote sensing and in-situ profiles difficult.