Replies to RC1: 'Comment on egusphere-2024-30', Anonymous Referee #1, 15 Apr 2024

We thank both reviewers for their thorough reading of the paper and their appropriate suggestions, which helped us improve our manuscript.

Black: reviewer comment
Blue: authors reply
Red: manuscript correction

General comment:

The other important source of validation data are the surface remote sensing FTIR instruments (NDACC). These are mentioned briefly but I would like to read more about the comparisons with the findings from these evaluations (e.g. EUMETSAT validation report). Are the main conclusions similar? Is the underestimate reported for the TCC quantitatively in agreement with FTIR evaluations?

The main conclusions of the EUMETSAT validation report are : good agreement between FORLI-CO total columns from IASI/Metop-C and NDACC-FTIR data; average of the relative differences (IASI compared to NDACC stations) of 2.7%, average of the Pearson correlation coefficient of 0.89, (within values reported for IASI-A and -B). We have added these details to the text of the article.
Line 43: "According to a recent validation report (Langerock et al., 2021), FORLI-CO total columns from IASI/Metop-C show a very good agreement with NDACC-FTIR data with an average relative difference of 2.7% and a Pearson correlation coefficient of 0.89 . Furthermore this document shows that the distributions of IASI-A, -B and -C are highly consistent."

Abstract:

l 5: It would be good to mention that MLS CO is used for the stratospheric part.

CO is mostly a tropospheric compound and the most important are the IAGOS data. We therefore prefer not to mention MLS already in the abstract.

l 9: Does the period "2008-2020" cover 12 or 13 years?

The period is from beginning of 2008 to the end of 2019 : 12 years. We removed the only few data from 2020 at Frankfurt. This does not change the results for this airport due to the very large number of data from 2008 to 2019.

l 12: What is meant by "to capture the CO variabilities"? Variability in time, or in the profile? I suggest to replace "variabilities" by "variability".

Done

l 13: Same question for "correlation coefficients". Does this refer to temporal correlations or vertical profile correlations?

Correlations for the temporal variability. We added "for the timeseries".

l 21: The last sentence "Our validation results will provide a better characterisation of IASI-CO data to the users and help improve the retrievals for future versions" can be formulated better. I would replace "will provide" by "do provide".

What does "better characterisation of IASI-CO data to the users" mean? Please reformulate "improve the retrievals for future versions".

We have changed the statement to : "Our validation results provide an overview of the quality of IASI-CO retrievals to the users and insights for improving the retrievals in the future to the developers."

l 36: Remove the "(" before "Hurtmans".

Done

l 47: The introduction gives the impression that the study of De Wachter 2012 is the only comparison with IAGOS. It would be good to mention that George 2015 also includes comparisons with IAGOS profiles, although this comparison is somewhat limited.

De Wachter et al. (2012) is the only study to date to have performed a thorough validation with full timeseries of IAGOS profiles. We acknowledge that some comparisons were made in George et al. (2015): "George et al. (2015) also used some IAGOS profiles for comparisons with IASI-FORLI and MOPITT data."

Introduction: The CO retrievals are introduced with a good set of references. But I am missing a paragraph in the introduction on IAGOS. What is it, what can be measured, some key achievements and key references, including validation work, to provide the reader with a background and further reading.

We added a paragraph and references to better introduce IAGOS: "IAGOS uses commercial aircraft for automatic and routine in-situ measurements of atmospheric composition including reactive gases (e.g. ozone and CO), greenhouse gases, aerosols, and cloud particles along with essential thermodynamic parameters (Thouret al., 2006, Nedeec et al., 2015, Petzold et al., 2015). IAGOS provides regular observations in the upper troposphere and lower stratosphere (UTLS) during the cruise phase, and vertical profiles in the troposphere during landing and take-off, and in particular over regions that are never or poorly sampled. This long term quasi-global dataset has been used in a wide range of atmospheric studies, e.g. process studies, trend analysis, validation of climate and air quality models (Clark et al., 2021, Tsivlidou et al., 2023, Coheur et al. 2024), as well as for the calibration of space sensors and the validation of their retrievals (De Wachter et al., 2012, De laat et al. 2012)."

Sec 2.1: The key reference for SOFRID is De Wachter. But this paper is more than 10 years old. Has there been any CO retrieval development in the meantime based on SOFRID? Are the updates made to SOFRID CO documented somewhere in more detail (is there a recent ATBD)?

The product that is distributed until now is the one described in De Wachter et al. (2012) with minor changes. The major updates are described here to introduce the new version which includes simultaneous CO and N2O retrievals.

l 70: "instead of operational EUMETSAT Level 2 IASI products". Why?

The FORLI team collaborates with EUMETSAT and uses their L2 products for performing operational retrievals. The SOFRID team collaborates with colleagues from Météo-France/CERFACS (Toulouse) that are using ECMWF operational analyses in their assimilation system. To make comparisons between assimilated and retrieved O3 (Emili et al. 2019), in a coherent way it is better to use the same temperature and humidity products. The

SOFRID team therefore uses ECMWF operational analyses for SOFRID O3 and CO since a long time. As the results are satisfactory, they go on using ECMWF operational analyses.

The interesting point highlighted in the paper is that using ECMWF analyses with SOFRID allow to show that EUMETSAT Temperature discontinuities are probably responsible for temporal variations of CO columns biases in the FORLI data (see especially Figure 7, Frankfurt timeserie). Furthermore, some similar issues of continuity have been encountered in the FORLI-O3 data with a sudden drop in 2011 explained by changes in the EUMETSAT Temperature products (Boynard et al. 2016, 2018).

Nevertheless, this comment highlights that the phrasing with "instead" was misleading. We have therefore removed this statement.

l 71: Could you please motivate why "The noise of the measurement covariance matrix has been reduced from 1.4 to 1.0 · 10−8 W/(cm2 sr cm−1)" I assume that the performance of IASI has not improved over time?

The retrieval noise is not IASI radiometric noise but needs to be tuned to take other error sources (errors in ancillary data such as temperature and humidity profiles, errors in RT modeling) and needs to be consitent with the Sa matrix. Here the retrieval algorithm is modified with a new version of the RT code, a larger spectral window, simultaneous retrieval of N2O and CO profiles, and the retrieval noise has to be re-tuned to re-optimize the retrievals. In De Wachter et al. (2012), the retrieval noise was chosen very conservatively in order to avoid providing anomalous CO values. In the version presented here, the experience and the use of long time series to test the retrievals allow to lower this parameter enhancing the information content and still avoiding retrieving noise. Furthermore, the N2O variability is much lower than the CO variability and it is important to tune the retrieval noise as low as possible to retrieve information about the N2O variability. We have improved the description as follows:

"The noise of the measurement covariance matrix has been reduced from 1.4 to 1.0 · 10−8 W/(cm2 sr cm−1) in order to better capture the N2O variations. N2O spatio-temporal variations are indeed very low (less than 5%) compared to CO variations (one order of magnitude). This noise level is still very conservative and much larger than the radiometric noise of IASI-A estimated to be around 1.5 x10-9 W/(cm2 sr cm−1) in the CO spectral window (Clerbaux et al., 2009). The retrieval noise indeed takes other sources of errors into account such as errors on ancillary data (temperature and humidity profiles) or radiative transfer modeling errors. It was optimized with sensitivity tests performed on the CO IAGOS validation database."

l 80: Same question as for SOFRID. The key ref for FORLI is from 2012 as well, again more that 10 years old. I noted the ATBD is from 2014. Are there updates compared to the 2012 Hurtmans paper? Any relevant evaluations of the FORLI CO retrieval published after 2012? Does FORLI v20151001 introduce any important changes compared to Hurtmans 2012?

FORLI v20151001 is mainly a technical and spectroscopy update. We have added some details to the text of the article l87: "For this validation study FORLI-CO v20151001 was used. This version is an updated version from the one described in Hurtmans et al (2012), using look-up tables recalculated to cover a larger spectral range with a more recent version of the HITRAN spectroscopic database (HITRAN 2012) and implementing numerical corrections. It was validated with NDACC-FTIR data (Langerock et al., 2021). This version was installed..."

l 97: What is the reason that "only airports providing at least 60 days with valid data " were selected? Is a number of 60 linked to the quality of the comparison? Even just one profile can still provide a useful comparison.

We have used a threshold in order to have time series for each selected airport representative of CO temporal variations at their location. Airports with too limited number of profiles (some just provide a couple or a dozen of profiles) do not provide valuable information in our perspective: no time evolution and no seasonal or inter annual variability. Their use is rather counter productive in our general analysis aiming at characterizing the ability of the retrievals to capture the regional (latitudinal) and temporal variations of CO. We have added the following sentence at the end of the paragraph l111:

"The remaining airports provide temporally sparse profiles, which do not allow for sampling the temporal variabilities representative of their location."

l 122: The difference in DFS is striking. The statement that "the reduction of the noise of the measurement covariance matrix relative to De Wachter et al. (2012)" is partly responsible askes for some more explanation. How can the noise be reduced by such a large factor? How does this compare to the a-priori noise assumed in FORLI? Please provide more detail.

As mentioned in reply to comment l71 above, the retrieval noise has been reduced from 1.4 to 1.0 × 10−8 W (cm2 sr cm−1)−1 corresponding to a factor of 2 for the noise variance. It may appear to be important but the retrieval noise remains large compared to IASI estimated radiometric noise of 1.5x10-9 W (cm2 sr cm−1)−1. As mentioned in the manuscript, the retrieval window has also been extended from 2143–2181 cm−1 (De Wachter et al., 2012) to 2143-2218 cm−1 in order to improve N2O retrievals. The 0-1 CO absorption band is composed of its P branch below about 2140 and of the symmetrical R branch between 2140 cm-1 and 2225 cm-1. There are about as much CO absorption lines and information about CO in the 2181-2218 window than in the 2143-2181 cm-1 spectral range (Stepanov et al., 2020). The addition of the two effects leads to a significant increase of DFS documented in the paper. The DFS from FORLI and SOFRID were close to each other (between 1 and 2) in De Wachter et al. (2012) and there is indeed a "striking" difference now because of the changes in SOFRID detailed above. This was already explained in the manuscript: "The larger information content from SOFRID is due to (i) the extension of the spectral window and (ii) the reduction of the noise of the measurement covariance matrix relative to De Wachter et al. (2012). Both modifications are related to the combination of CO with N 2 O retrievals."

We have improved this explanation about the enhanced DFS in SOFRID with more details:

"In the former validation study (De Wachter, 2012), the SOFRID and FORLI DFS were close to each other, ranging between 1 and 2. The larger information content from SOFRID present version is due to two effects related to the simultaneous CO and N2O retrievals. First, the extension of the spectral window from 2143–2181 cm−1 (De Wachter et al., 2012) to 2143-2218 cm−1. The 2181-2218 cm−1 window indeed contains about half of the nu3 N2O absorption band (Barret et al., 2020). The 0-1 CO absorption band is composed of its P branch below about 2140 and of the symmetrical R branch between 2140 and 2225 cm-1 (Stepanov et al., 2020). The extension of the spectral window is therefore roughly doubling the number of CO absorption lines compared to De Wachter et al. (2012). Second, the retrieval noise variance has been reduced by a factor of 2 (see above) in order to improve the ability of the retrieval to capture N2O variations."

l 119, 133: I note that TCC is used in two ways, either as total column (retrieval of 1 quantity) or as total atmosphere (for DFS = 2.9). Maybe better to use "Total atmosphere" instead of TCC in table 1.

We have changed TCC for total atmosphere in Table 1 and in the text discussing DFS.

l 177: I find the statement "because they provide the best assessment of the real differences between the in-situ and the remote sensed data" a bit dubious. Equation 1 describes how the retrieval relates to the real profile, and provides the best way of comparing. This is also evidenced by the validation results, e.g. l 238.

Equation 1 allows to take the two largest sources of error into account for comparison: the effect of the a priori on possible biases and the smoothing of the true profile due to the limited vertical resolution. It is crucial for model validation because in that case the aim is to assess the quality of the model and not the quality of the satellite retrievals. Here, we want to assess the quality of the retrievals with reference in-situ data. The use of Equation 1 is indeed improving the comparisons but at the price of hiding the largest errors of the retrieved product and therefore prevents providing the complete evaluation of the product. A lot of papers only provide comparisons with smoothed profiles and the user cannot know the quality of the satellite product. We therefore consider the comparison with raw data as the "real" comparison because it allows a potential user to know what the product is worth but we also provide and discuss results using Equation 1.

l 297: This mentions "two major updates of EUMETSAT Level 2 data processing". Could you please provide details on these updates in Sec. 2.

These major updates of EUMETSAT Level 2 data processing improved retrieval of the vertical temperature profiles, and cloudy data flagging for the second one. We have completed with: "These updates improved the retrieval of the vertical temperature profiles, and the cloudy data flagging for the second only."

l 377: "timeseries"

Done

l 387:What does "not statistically significant" refer to? From the paper I got the message that TCC globally is not really different between SOFRID/FORLI, but that negative biases are observed.

The message is indeed close to what the reviewer reports. With significant, we meant that the RMSD was larger than the bias but this is not a sufficient condition to state that the bias is not significant so we removed "not statistically significant".