Reply to Comment on amt-2021-329 by Anonymous Referee #3

Referee comment on "Sentinel-5P TROPOMI NO2 retrieval: impact of version v2.2 improvements and comparisons with OMI and ground-based data" by Jos van Geffen et al., Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2021-329-RC2, 2021

$\Longrightarrow$ The referee report is copied below; the reply is preceeded by an arrow, like this text.

The paper "Sentinel-5P TROPOMI NO2 retrieval: impact of version v2.2 improvements and comparisons with OMI and ground-based data" by Geffen et al. presents the improvements made in the official TROPOMI NO2 product. It is well written, matches the scope of AMT, and will serve as a key reference for the TROPOMI NO2 data set which is frequently used by the community.

$\Longrightarrow$ We thank the referee for these kind words.

I recommend publication after dealing with some minor comments.

Minor comments

- Page 2 line 2: I understood that the main effect is on the tropospheric column. I would skip "and total" here.

  $\Longrightarrow$ As concluded in Sect. 5, the situation for individual measurents may be quite complex, both for tropospheric and total columns – hence this phrasing.

- Page 2 line 21: "over e.g.". I think that there are more similar products, especially over the US.

  $\Longrightarrow$ We are not aware of regional data products over the US based on TROPOMI data. There is Laughner et al, 2018 (doi:10.5194/essd-2018-66) but that uses OMI data. We are told NASA is working on a TROPOMI version, but details are not known.

- Page 3 line 2: As this often leads to confusion, it might be worth pointing out that the AMF depend on the profile *shape*, but not on the absolute concentration levels.

  $\Longrightarrow$ You are right and it is indeed good to emphasize this:

  ... pressure, the shape of the NO_2 vertical profile (not of its absolute concentration levels), and ...

- Page 3 line 5: "v1.2/v1.3": It would be helpful if the authors would add a reference to a table listing the different NO2 product versions and what is new in 1.3/1.4/2.1.

  $\Longrightarrow$ This comes in Sect. 1.1; a forward reference to that is added to the end of the preceeding paragraph:

  ... ground pixel in question. An overview of the NO_2 data versions is given below in Sect. 1.1.

- Page 3 line 20: I appreciate that the authors explicitly provide this scaling factor. In addition, I would appreciate if the unit molec/cm2 would be added at the top of all colorbars or at additional top/right axis.

  $\Longrightarrow$ The molec/cm2 unit is available on the right axis of Fig. 3. That unit has also been added at some locations (Sects. 3.1, 3.2, 3.3, 4.3, 4.4) and to the upper axis of scatter plots (Figs. 6, 11, 12, 13; addition also to the right axis is unpractical). Adding that unit to the maps (Fig. 2, 5), which are made using Panoply, in the form of a double colour bar is unfortunately not feasible.

- Page 3 line 26: "and include information on earlier versions." I don't understand this: does this mean that the linked document includes this information? (then add an "s"). Or is this a reminder to the author to include this information here?

  $\Longrightarrow$ That should have been "and includes information" – thanks for spotting this.

- Page 3 line 28: Please name the institutes directly. Pointing out the country here reads like strange kind of nationalism.

$\Longrightarrow$ The country was given because of the "NL" in the processor name. But you are right: this formulation doesn't look good – it has been corrected:

... the so-called NLL2DP ("Netherlands level-2 data processor") that provides the TROPOMI data products for which KNMI and SRON are responsible, ...

- Page 4: Here comes the information I was asking for above. I still think it would be helpful to have a table with the information in short version, even if this would add some redundancy.

  $\Longrightarrow$ The itemized list of the different versions contains all necessary information and is, we believe, easier to read than a table. Adding this also in the form of a table would be an exact doublure, as there is no clear way to summarise the already short descriptions.

- Page 4 line 16: I have seen TROPOMI results before 30 April 2018, and as far as I am aware the quality of TROPOMI spectra was already quite good since beginning of 2018. I would highly appreciate if the reprocessing would be extended back to 1 Jan 2018.

  $\Longrightarrow$ Several things are important here.
  a) 30 April 2018 was the beginning of the so-called operational phase, under the responsibility of ESA, who is also responsible for the reprocessing, and 30 April 2018 thus marks the start of the official data product release.
  b) There is indeed data before that, but during this so-called commissioning phase, especially up to early March there are large gaps in the data (when instruments tests and special measurements were done). Regular data is available only since 15 March 2018.
  c) For the NO_2 data assimilation with TM5-MP to provide reliable results, a spin-up periods of several days is required, hence data prior to mid March 2018 would not be of optimal quality if it was processed. Note that because of the spin-up, the actual reprocessing will start on 15 April 2018, with data released as of 30 April.

- Section 3.2: A similar outlier removal was introduced by Richter et al., https://doi.org/10.5194/amt-4-1147-2011. Please add a reference here.

  $\Longrightarrow$ Not quite "similar": the outlier removal we use sets limits based on the inter-quartile range and does this only once (that is: after the first round of the removal, there is no second round), while Richter et al. 2011 "itteratively scan the residual for points having a value larger than 10 times the average residual of the fit. [...] This procedure is repeated until no further outliers are identified." Since there are other ways to do outlier removal and a discussion of these approaches is beyond the scope of the paper, adding this one reference does not seem appropriate.

- Fig. 2: The figure illustrates that the outlier removal reduces the error, but the numbers are out of context. Instead (or in addition) of showing the difference I would propose to show the maps of SCD errors with and without outlier removel directly.

  $\Longrightarrow$ The main idea of the plot is to show that the SCD error decreases (negative values on the map) and where (mainly in the SAA). What the absolute SCD error values are is of less importance, but to give some idea average SCD error are shown in Fig. 3. To nevertheless accomodate the referee's request, maps of the SCD error values with and without outlier removal are added as figure in the Appendix. Note that referee # asked for a map of the SCD value difference, and this was added to Fig. 2; for completeness sake Fig. 3 was then also expanded with the SCD (in terms of the GCD).

- Table 3: Again, it is difficult to assess what a relative change of SCD error of 1% means, if the SCD error itself is not given. I would propose to compare the SCD error with and without outlier removal etc. directly.

  $\Longrightarrow$ Comparing the SCD error with and without outlier removal is done in Figs. 2 and 3 for one example orbit, with the v2.2 processor, i.e. no other differences than the outlier removal. This is dedicated processing and therefore not available for all periods. The difference between DDS and OFFL contains more than just the outlier removal, showing the total effect of all improvements is in Fig. 4 and Table 3. To accomodate the referee's remark and similar remarks of referee #1, the absolute change of the SCD and the (corrected) SCD error are now listed in Table 3, while the change in VCDstrat is moved to a new table.

  Please add a footnote why the last entry for Vstrat is missing.

$\Longrightarrow$ With the VCDstrat data moved to a new table, this is no longer necessary.

- Page 11 line 11: "strong decrease in the SCD error and RMS error": I would like to have this "strong" decrease given in absolute numbers in the presented figures and tables.

  $\Longrightarrow$ The absolute SCD error changes are now listed in Table 3 (and an example can been seen in the new Fig. A1). Absolute changes in the RMS are not listed because the absolute RMS cannot be compared to other retrieval approaches due to differences in the RMS definition. (In addition, Table 3 currently fits just in one fullpagewidth.)

- Page 14 line 15: Please explain why a dedicated TROPOMI version for FRESCO was necessary and FRESCO+ could not be used directly.

  $\Longrightarrow$ In answer to this and the following two referee comments.
  Unfortunately there is no official FRESCO-S or FRESCO-wide publication (paper or published report) as yet. The idea was here to refer to the discussion in the Eskes et al. paper that is "in preparation". Since that paper is still in preparation and in view of this referee's comments, the number of references to that paper is reduced and it was decided to adapt the beginning of Sect. 4.1:

  The FRESCO+ algorithm (Wang et al, 2008) retrieves cloud information from the O_2 A-band around 758 nm (cloud fraction and cloud pressure) as well as scene parameters assuming clear-sky (scene albedo and scene pressure) and was developed for the GOME-2 instrument. Due to the high spectral resolution of TROPOMI compared to GOME-2, the fact that TROPOMI has a spectral smile (cf. Sect. 3.1), and because of TROPOMI's row-dependent instrument spectral response function (ISRF, known also as slit function) with spectral shifts caused by inhomogenous slit illumination, the FRESCO+ algorithm needed to be re-written and the corresponding lookup tables needed to be generated once more. The resulting implementation is called FRESCO-S (short for FRESCO-Sentinel) and its cloud pressure data is used for the v1.2-v1.3 NO_2 data product.
  ...
  FRESCO+ (Wang et al., 2008) makes use of the wavelength ranges $758-759$ nm, $760-761$ nm and $765-766$ nm. For the FRESCO-S implementation the first window, representing the continuum, was shifted a little to $757-758$ nm. As a further improvement of the cloud retrieval, nicknamed FRESCO-wide, the third window is extended to $765-770$ nm in order to include more of the weaker O_2 absorption lines. This extention mainly impacts the lower clouds, generally decreasing the cloud pressure in the order of 50 hPa, and is relevant for all instruments where FRESCO has been applied. For high clouds the FRESCO versions deliver very similar cloud heights on average. Further details are given in the ATBD (van Geffen et al, 2021).
  FRESCO-wide, used as of NO_2 v1.4, provides a more realistic ...

- Page 14 line 18: Does "FRESCO" here mean (a) all FRESCO versions or (b) the original FRESCO version, different from FRESCO+ and FRESCO-S?

  $\Longrightarrow$ FRESCO here refers to the general FRESCO approach, not a specific version; the text makes this clearer:

  ... (which the FRESCO algorithm sees as an effective cloud) ...

- Page 15 line 3: Please add a reference to "FRESCO-wide", or provide more detail here.

  $\Longrightarrow$ See above.

- Figure 9: With the chosen color bar, it is impossible to discriminate between 100I propose to use a colobar which is monochromatic from 0 to 100 (e.g. dark blue to light blue) and have additinal discrete and distinctive colors for the discrete cases occuring above 100.

  $\Longrightarrow$ The point of the maps and the text in Sect. 4.2 is that the quality of the ECMWF data is much better than the NISE data in resolution and an particular in coastal areas, and that the ECMWF does not contain the problematic flags 252-254 of NISE. It is indeed difficult to see the difference between the flags 100, 101 and 103, but that distinction is not relevant for the NO_2 data: when in the snow/ice mode – which is when the snow/ice flag is 003 larger (but not 255) – the amount or type of snow/ice is not used. As additional information to the referee, the maps below zoom in on the flage 100-103. To better reflect the usage in the NO_2 processing, the following sentence is added to the end of the second paragraph of Sect. 4.2:

  ... the problematic NISE flags 252-254. The cloud fraction and cloud pressure are used for the AMF calculation for pixels flagged as ocean (255), snow-free land (000) or a percentage sea-ice flag smaller than for the ECMWF data (in case of NISE data this was smaller than 002, since NISE has the

tendency to underestimate snow/ice cover); in case of other flags the scene parameters are used.

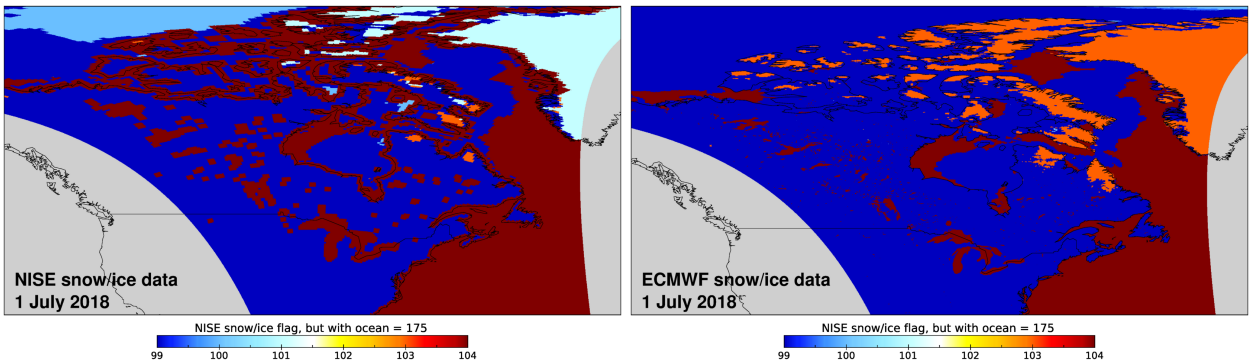

Figure 1: Maps of Fig. 9 of the manuscript focussing on the flags 100 (100% sea-ice), 101 (permanent ice) and 103 (snow) in the NISE (left) and ECMWF (right) snow/ice cover data.

- Figure 13: "version" should be "versus"?

  $\Longrightarrow$ Thanks for spotting this – it has been corrected.

- Figure 15: The meaning of IP68 has to be explained in the caption.

  $\Longrightarrow$ Addition to the figure caption, at the and of (a):

  ... (dotted lines). IP68 is the central 68 interpercentile range, the difference between the 84th and 16th percentiles, a measure for the dispersion. Δ is the difference "S5P-GB", where GB stands for ground-based.

  $\Longrightarrow$ And the text near the end of the first paragraph of Sect. 5.1 is expanded:

  ... and the dispersion (half of the central 68 interpercentile, shorthand 0.5 IP68) of the difference between ground-based ("GB") and S5P ...

- Page 29 lines 12-17: Please check if you could be more concrete by now on these future plans.

  $\Longrightarrow$ The text in Sects. 1.1, 6.x and here have been adapted to reflect that NO2 v2.4 and the subsequent mission reprocessing will contain i) a new version level-1b spectra with radiance degradation correction, and ii) the new TROPOMI surface albedo climatology.