Final response to referee comments on paper amt-2022-258

First of all, we would like to thank reviewer #1 for his/her constructive comments, which helped to improve the manuscript. Below we give answers and clarifications to all comments made by the referee (repeated in italics).

Anonymous Referee #1

Specific comments

Reviewer: *The paper made extensive comparisons between v1.5 and v1.8 of the TROPOMI/WFMD algorithm. The previous algorithm paper (Schneising et al., 2019) is focused on v1.2, and it appears that a detailed description of v1.5 has not yet been published. It would be helpful to add a table that summarizes the differences between the three versions (v1.2, v1.5, and v1.8).*

Authors: A corresponding table is added to the revised version summarising the respective differences in the algorithm setup of the different versions.

Reviewer: *Figure 1 and section 3.1: the increase of polynomials in fitting is an updated applied to all retrievals. Can the authors include some results for other areas?*

Authors: Section 3.1 is completely revised also discussing the Taymyr region in Northern Siberia and the global impact of the increase of polynomial degree quantitatively using a global map of the induced differences. For TROPOMI/WFMD the impact is typically small and is largest for the Etosha pan. That was the reason to focus on this region.

Reviewer: *Figure 3a: there appears to be a gradient along 60 N, especially over Siberia. Is there any explanation for this?*

Authors: This gradient can be attributed to the GMTED2010 data. This is a known limitation of GMTED2010, as the DEM is composed of various datasets and the main dataset used in GMTED2010, the Digital Terrain Elevation Data v2 (DTED 2), is only available for latitudes between 60°N and 56°S. The topography information at higher latitudes is based on older data with lower spatial resolution. This discussion is added to the revised version. That the high latitude difference pattern is an issue of GMTED2010 and not of GLO-90 can also be seen in a new Figure showing the methane distribution for several TROPOMI/WFMD versions over Northern Siberia demonstrating that v1.8, which is based on GLO-90, exhibits the most homogeneous methane distribution in this region.

Reviewer: *Figure 4: perhaps it would be useful to compare the same version of the algorithm with the DEM model as the only difference.*

Authors: v1.5 was the previous officially released data set before v1.8. Thus, we think that the comparison of these officially released products is most expedient. Furthermore, the other changes are negligible compared to the impact of the DEM change over Greenland, which can occasionally exceed 100 ppb. The difference pattern is reflected 1:1 in the inhomogenities of

v1.5 at coastal Greenland as can be seen in Figure 3b (Figure 6b of the revised version). For example, there are no significant changes due to the increase of the polynomial degree over Greenland as can be seen in the newly added figure demonstrating the global impact of this modification (see also above).

Reviewer: *Section 3.3.1/3.3.2, does the change in TROPOMI spatial resolution have any effects on the trained models for cloud screening and bias correction? Do the models need to be re-trained to account for the change in resolution (and also the cross-track index)?*

Authors: There is only one common quality filter and calibration for both periods before and after the change of spatial resolution and thus no re-training. The relevant time periods for training the quality filter and calibration (subsets of 2018 and 2019 covering both footprint sizes) are added in the revised manuscript. The footprint size change in August 2019 only (slightly) changes the along-track dimension of the measurements, the across-track expanse does not change. Therefore, there is no issue with the across-track index used in the calibration associated with this change.

Reviewer: *Lines 191-193: how were the parameters a, b, and c determined?*

Authors: These parameters were determined empirically to distinguish between typical values of the root mean square of the fit residual (as function of the sun-normalised radiance) and outliers for whatever reason, such as specific scenes with intense aerosol exposure exhibiting reduced fit quality relative to scenes with similar radiance. This is better explained in the revised version.

Reviewer: *Section 3.3.2, given that the cross-track index is an input to the trained model, would one expect the striping to be eliminated or reduced by the bias correction procedure?*

Authors: The inclusion of the across-track index in the calibration accounts for recurring systematics, e.g. potential viewing zenith angle issues or temporally constant striping patterns. As a result, striping is improved to some extent but complete destriping, in particular with respect to temporally variable striping patterns, cannot be entirely achieved by a shallow calibration. Therefore, the remaining vertical stripes in the satellite data after calibration are efficiently removed orbitwise in v1.8 by combined wavelet-Fourier filtering. This discussion is added to the introduction of Section 3.4.

Reviewer: *Section 3.4 – is the destriping scheme run after the bias correction? Please clarify.*

Authors: It is run after calibration. This is clarified in the revised version (see also answer to previous comment).

Reviewer: *Line 296: it would be helpful to briefly introduce Coif16 to readers who are less familiar with the method.*

Authors: We added a short explanation and reference in the main text and a figure to the appendix.

Reviewer: *Table 1: Why are the random error and especially systematic error for v1.2 different from the values given in Schneising et al. (2019)? Also it appears that the systematic error in this paper follows a different definition?*

Authors: There are 5 additional TCCON sites involved in the validation and the validation time period is extended by 1 year. Therefore, exactly the same numbers cannot be expected. The definition of systematic error has slightly changed also taking seasonal biases into account. These differences are clarified in the revised version.