Author's reply to Anonymous Referee #1

We would like to thank referee #1 for his/her comments that helped improve the quality of the paper. Below we addressed the comments one-by-one; comments from the referee are typeset in italic, our replies are in normal font, and our changes in manuscript are in blue. Line, page and figure numbers in the referee's comments refer to the original manuscript, whereas in our reply we give page and line numbers that refer to the revised manuscript.

General Comments

1) Surface albedo is strongly wavelength-dependent, for most (if not all) land surface types, however, the TROPOMI algorithm assumes a simple first-order spectral dependence across the NIR and SWIR bands. For example, green vegetation exhibits a sharp increase in reflectance near the TROPOMI NIR band (757-774 nm) and the reflectance of vegetation is likely not linear across this spectral band. Various types of minerals and soils also exhibit complex spectral dependence of surface reflectance in both the NIR and SWIR bands. Why was this effect not represented in the simulations?
Regarding the TROPOMI CH4 retrieval algorithm, the baseline is to retrieve the surface albedo up to first-order spectral dependence, however the option exist to retrieve higher-order spectral dependence. This option has been tested with GOSAT observations and it will be examined also for TROPOMI when real data is availiable. To clarify this, we have added the following sentences on page 11, line 292:
For simplicity, we assume a constant surface albedo within the NIR and SWIR band in our simulations. We expect no difficulties to fit more realistic spectral dependent surface albedo based on our experience with real GOSAT data.
and on page 6, footnote:
Note that the algorithm has the option to fit higher order spectral dependence of the surface albedo. It will be investigated on real data whether this is needed.
Further, the surface albedo used in the simulations are taken from MODIS and SCIAMACHY albedo product for the NIR and SWIR spectral bands, respectively. As far as we are aware of there exist no global database of the spectral dependence of the albedo in the NIR and SWIR window, hence the simulations assume a constant albedo within a band.

2) As described in Section 2.3, filters are used to discard those observations where the reliability of the methane retrieval is expected to be poor. Globally, approximately one half of the observations will be rejected. This will include scenes where aerosols exhibit high optical depth, large particles, and are located at high altitude. Thus, it should be expected that rejected observations will not be random, but will probably occur much more frequently in certain geographical regions (and during certain seasons) compared to others. This is extremely important information for potential users and should be specifically addressed in the revised manuscript. Where will the TROPOMI methane retrieval algorithm generally be useful and where won't it be useful? For example, it appears that there are very few successful retrievals over China and India (in Fig. 7), but it is not clear if this might just be due to the lack of ensemble 'test profiles' located there. Maps of the locations of rejected retrievals should be shown
Although, it is in general true that geographical regions with high aerosol loads have more chance to be filtered out, we prefer not to draw too strong conclusions on which regions and to what extent based on our global ensemble because it was not designed for this purpose. Rather, it was designed to test the mean retrieval performance per season based on a representative day. So, detailed information on regional scales down to the individual TROPOMI pixel size are not given by the

ensemble. First of all, the spatial resolution used in our study (~3x3 degrees) is much coarser than TROPOMI's resolution, thus where in our study whole regions might be filtered out, TROPOMI may still have useful observations. Secondly, since our simulations only represent one day per season with a mean atmospheric state, and one should therefore be careful in generalizing to observations entire seasons. To summarize, it is not that not whole regions will be filtered out but only that certain regions will have less days with success retrievals during certain seasons.

With this in mind and to illustrate which regions are more likely to be filtered out by our method, we have added the worldmap of unfiltered retrievals in Fig. 4a on page 14. and added a paragraph on page 13, line 332:
Note further that most of the rejected retrievals are found in the desert and dust regions in Northern Africa and Central Asia in April and July where our simulations capture dust storm events. For real TROPOMI measurements though, we will less likely filter out large regions because TROPOMI has higher spatial resolution than our simulations. Also, our simulations only represent one day per season with a mean atmospheric state. Therefore our results merely indicate that certain regions will have less days with successful retrievals in certain seasons, not that that entire regions will be rejected.

Furthermore, on closer investigation, we found that Fig. 7 was generated with wrong (stricter) filtering that lead to fewer points than shown in Fig. 4. We fixed this and replaced Fig. 7 which is now consistent with Fig, 4 in terms of plotted points. The situation is thus less severe as we have more successful retrievals in China and India than shown before.

Minor Revisions and Technical Corrections

p. 5, line 125. Does the entire forward model grid really change as p_sfc changes (in order to create 36 layers with equal pressure widths), or is there a fixed pressure grid above the surface?
It is as explained on page 5, line 126, so no changes are made in the manuscript. To answer the reviewer's question: the forward model pressure grid will differ per retrieval in case the surface pressure is different. During a single retrieval though, the pressure grid remains fixed as the algorithm does not retrieve the surface pressure in the baseline configuration.

p. 5, l. 139 (and following lines). Various assumptions are made in the aerosol model, including values for r_1, r_2, and the aerosol complex refractive index. How is it known that these are reasonable values?
The aerosol complex refractive index we use is an average complex refractive index weighted by the relative mass contribution of the individual aerosol chemical types in the ECHAM5-HAM model per spectal window. It has been shown by Butz et al 2010 that the results are not sensitive to the exact choice of the refractive indices.
To clarify the above, we have added on page 5, line 143:
which are averaged values derived from the ECHAM5-HAM model (Stier et al. 2005).
and on page 6, line 154:
Butz et al. 2010 found that the the exact choice of the fixed-value parameters such as the refractive indices or the width of the height distribution are not affecting the retrievals significantly.
Concerning the values for r_1 and r_2, these are from Mischenko et al 1999 as referenced in the paper on page 5, line 144.

p. 6, l. 157. Define 's' in Eq. 3.
We have added a sentence on page 6, line 164:

$F^{surf}_{s,755}$ and s represent the chlorophyll emission at 755 nm and its spectral slope, respectively.

p. 6, l. 161. What is the relationship between the 12-layer retrieval grid and the 36-layer forward model grid?

They share the same interfaces, thus there are three sublayers of the model atmoshere grid in the retrieval grid. This is now described explicitly in Section 2.1.2, page 6, line 167

The 12 retrieval layers are related to 36-layer model atmosphere by the shared interfaces, i.e. each retrieval layer is divided in three sublayers for the forward model calculations.

p. 6, l. 162. It appears that the retrieval algorithm simultaneously retrieves both methane and CO total column amounts. Is this CO total column value just a diagnostic for the methane retrieval, or is this the official TROPOMI CO product?

Our retrieved CO column is not the official TROPOMI CO product, but indeed a diagnostic of the CH_4 retrieval. The CO retrieval algorithm for TROPOMI is described in a separate publication by Landgraf et al. in this special issue. We have added a sentence in the paper to clarify this, page 7, line 178

Further, we like to mention that our retrieved CO column should be regarded as a diagnostic for the CH_4 retrieval. The official TROPOMI CO product has a dedicated retrieval algorithm described by Landgraf et al 2016.

p. 8, l. 200. Suggest rewording phrase 'have sensitivity down to the ground'

See page 8, line 215, rephrased to:

This illustrates that the retrieval of methane columns from the SWIR has a nearly ideal sensitivity to methane in the troposphere and the tropospheric boundary layer.

p. 8, l. 211. Is there a reference for this equation for total column retrieval uncertainty? Please show how this equation was derived. As shown in Rodgers' book (Section 4.3, 'Best Estimate of a Function of the State Vector'), the measurement error for total column will involve the matrix Sx as well as the linear operator which relates the retrieved profile and the total column.

We have added a reference to Rodgers (2000) on page 9, line 225. Note that Eq, (7) follows directly from Rodger's equation for the error covariance (h^T S h) in Section 4.3 after equation (4.48), where the linear operator that relates XCH_4 to the state vector is given in our paper in Eq. (4).

p. 8 l. 213. It should be noted somewhere that the simulations will not represent all types of forward model errors, such as errors in the underlying spectroscopic database.

This was already mentioned in in Section 5, but we also changed and added sentences on page 12, line 299:

While Butz et al. (2012) investigated only the scattering induced error, we increased the inconsistency between the simulation forward model and the retrieval forward model model. We have attempted to include the most important contributions to the forward model error, except for errors due to the underlying spectroscopic database which have been investigated elsewhere, see Galli et al. (2012) and Checa-Garcia et al. (2015).

p. 11, l. 251. Meaning of the phrase 'to better oppose them with the algorithm performance' is unclear to me.

We removed this phrase to avoid confusion.

p. 11, l. 256. Properties of the ensemble are very important to interpreting the simulation experiments. How are the ensemble profiles distributed geographically? What is

the range of aerosol optical depths in the ensemble?
As mentioned at the beginning of Section 3.1, the ensemble is already described in detail by Butz et al 2012, we therefore found it unnecessary to repeat exaclty the same plots. However, we included a reference to the relevant figures at the appropriate places, page 12, line 295:
We refer to Fig.2 and Fig. 3 of Butz et al. 2012 for the geographical distribution of the total optical thickness and surface albedo, respectively, used in the simulations.

p. 11, l. 265. Clarify that simulations assume constant surface albedo within NIR and SWIR bands.
We have added the sentence on page 12, line 292:
For simplicity, we assume a constant surface albedo within the NIR and SWIR band in our simulations. We expect no difficulties to fit more realistic spectral dependent surface albedo based on our experience with real GOSAT data.

p. 12, l. 286. Are these SNR values really 'design goals' or are they based on actual instrument engineering data?
They are requirements/design goals, as mentioned on page 11, line 276.

p. 13, Sec. 3.2.1, first paragraph. Is the a priori methane profile experiment really done using the 'latitudinal mean' (based on an average over all latitudes) or 'zonal mean' (based on an average over longitude)? In either case, please provide some more detail, such as the width of the latitude or longitude bins used to calculate the mean.
Indeed, we should have use 'zonal mean'. This has been corrected and clarified, e.g. on page 13, line 346:
zonal mean, where we averaged over all longitudes within a 2.79 degree latitude bin,

p. 21, Conclusion. The conclusion should include information about the filtering algorithm, and what scenarios (geographical regions and seasons) most often lead to filtered retrievals.
W have added an paragraph on page 13, see our reply to General Comment 2. where we also explain why we are careful not to draw too strong conclusions on the filtering with respect to geographical regions and seasons. Our global ensemble is not designed to study this effect. Therefore, we wish to avoid repeating this discussion in the Conclusions section as we feel it would put too much emphasize on it.