Review by Anonymous Referee #1

The authors have considered all my comments and revised their manuscript accordingly. This is a very nice and valuable study and deserves to be published. I only have some minor technical issues that remain:

Thank you for your positive feedback as well as for reading thoroughly and spotting mistakes.

- *P1, L4: Write (to be more clear) "the extension of the clear-sky to cloudy scenes"*

  We have rephrased the sentence as follows to be more clear: "Compared to the previous clear-sky-only data product, coverage is greatly enhanced by including scenes over low clouds, . . . "

- *P1, L10-11: I guess both holds for the mid and low latitudes. Thus, I would suggest to rewrite the sentence as follows: "In low and mid latitudes, the bias is small at low-altitude stations, but has a larger value at high altitude stations." Note, I exchanged here "latitude" by "altitude" since I guess you meant here "altitude" and not "latitude". There is another occasion in the manuscript where I think it was also written "latitude" instead of "altitude". Please check.*

  We do mean high latitude stations (at low altitudes). The whole sentence is about low altitude stations, which is the reason our original sentence starts with that phrase, and then distinguishes between low and mid latitudes on the one hand and high latitudes on the other hand (both at low altitudes). To improve understandability, we have rephrased it as follows: "At low-altitude stations, the bias is small at low and middle latitudes and has a larger value at high latitudes."

- *P1, L14: Plays a role for what exactly? Please add.*

  We have rephrased this sentence as follows: ". . . since the information is filled up by the prior, a realistic shape of the prior is important for realistic total column estimation in these cases."

- *P3, L74: Abbreviation TM5 not introduced.*

  We now write "the global chemistry Transport Model, version 5 (TM5)".

- *P4, L92: Add "to" so that it reads ". . . .and to stabilize. . . .".*

  Done.

- *P7, Fig 3 caption: Use capital letters, thus "RMS" instead of "rms".*

  Changed.

- *P9, L139: A closing parentheses is missing in this sentence.*

  Added after "(a. s. l.)".

- *P9, L140–142: This is still difficult to understand. On P17, L261 you give a much better explanation. I thus would suggest to write similar here as there.*

  We have rephrased it as follows: "If the altitude difference between station and satellite ground pixel is too large, both observe too different partial columns which leads to errors. That is the case for high-altitude stations that are typically located on mountains so that most co-located ground pixels have significantly lower surface height."

- *P9, L147: Instead of "i is TROPOMI and j is TCCON" I would suggest to write "i denotes (or is used) for TROPOMI and . . . ."*

  We have changed "is" to "denotes".

- *P10, L158: theses $\mapsto$ these*

  Corrected.

- *P10, L160: "are biased to each other" I think this is not correct English wording. I would suggest to rephrase this.*

  We have changed this to "do not agree".

- *P10, L164: below? Where exactly? Please add this to the sentence.*

  We have changed "below" to "in the following" as this derivation is described in the remainder of that paragraph.

- *P11, Fig 5 caption: (a) is not a histogram, but a correlation. Further, you should put the labels for the respective panels behind the text that describes what is shown, as it is done also for the other figures. Add also what the dashed line is showing.*

  To be precise, (a) shows a two-dimensional histogram of correlations. The dashed line represents the one-to-one line. The figure caption is updated accordingly. In order to unify the placement of the labels (a), (b), etc., we have moved them to the front position in the other figures since we think that makes more clear which sentence belongs to which panel.

- *P13, L218: Better to write "explained" instead of "connected"?*

  We have changed "connected to" to "explained by".

- *P14, Fig 17 caption: Also mention here in the caption what is shown by the colours and that the respective fits are shown.*

  Added.

- *P17, L240: Better to write "generally" instead of "typically"?*

  We have moved "typically" further behind in the sentence (now before "large variations"), since we mean that large variations in surface altitude and albedo within a ground pixel are typical for Garmisch, although some ground pixels may lie in the valley only and thus not have these large variations.

- *P17, L250: I would rather write "former" instead of "old", but if you have a version number you could also use that one here.*

  We have changed "old" to "former" and also inserted "former" in the first sentence of the subsection.

- *P21, L302: "....as for TCCON" $\mapsto$ "....as derived for TCCON"*

  Changed.

- *P21, L309: Add a short explanation why or refer to the respective section where you discuss this.*

  We have added a half-sentence explaining why and a reference to Section 4.1.

- *P22, Fig 13 caption: Put (a), (b) behind the text describing the respective panels as it is done for the other figures.*

  To unify the placement of the labels of the panels (a), (b), etc., we have moved them to front position in the other figures.

- *P23, Fig 14 caption: Add also here that fitted lines are shown.*

  Done similarly as in Fig. 7.

- *P25, Fig 16 caption: Same as for Fig 13 caption.*

  See answer for Fig. 13.

- *P27, L353: State more clearly that you mean the data that has been presented in this manuscript.*

  We now write "these TROPOMI data".

Review by Anonymous Referee #3

The authors have, in a broad sense, responded well to most comments but have been deflecting at times. A few comments:

Thank you for your review.

- *While showing the spectral fits, the authors state that "Nevertheless, the fit is reasonable". I don't know what that actually means. In fact, I consider the fit to be pretty mediocre and a reason for concern. The systematic residuals by far exceed the noise estimates and are systematic throughout the entire spectral range. I would be very interested if the authors actually find reduced chi2 that are anyhwere close to one in a typical case (i.e. not exceedingly dark). As the authors know, the entire posterior error analysis assumes no forward model error, which is clearly not the case. Thus, I expect systematic errors to far exceed your estimated precision errors. I would say this warrants a discussion (esp. for moving forward) instead of stating that they look "reasonable". E.g. in panel B, I see multiple lines that are saturated (H2O absorption) in the model but clearly not in the measurement. For some reason, I would have expected opposing behavior in residuals for some of the other H2O lines, as the fit would otherwise be able to adjust H2O to match the observation.*

  The radiance noise estimate is taken from the L1B files and is based on the read noise and shot noise as determined in the detector characterization, i. e. purely the statistical noise in the detector signal, but it does not include errors due to correction (offset, dark current, memory, PRNU, straylight) and conversion steps in the processor. Although these errors are mostly systematic, they should be taken into account partially. That means that the noise estimation does not include all components. Therefore, the use of this radiance noise estimation causes the residuals to exceed the noise estimate and leads to high χ^2 values. Thus, χ^2 does not correspond to the real fit quality. That happens especially for bright scenes (e. g. over the Sahara) where the sensor noise is very small so that the real noise is significantly underestimated. The spectral fit in our previous version shows the most extreme case with χ^2 very near to the filter threshold. Nevertheless, the absolute value of the residual is about 2% of the spectrum. To be more illustrative, we now show a typical case which is randomly chosen among clear-sky scenes co-located with Karlsruhe TCCON station near equinox. The reduced χ^2 value is 2.5.

- *Figure 17: I honestly don't fully understand what I am seeing. While these are single measurements, as opposed to an average as in Figure 16, I would expect them to be broadly consistent in range and distribution. However, the dynamic range in H2O is minute while delta-D covers kind of everything, including (in my mind) somewhat unphysically low delta-D values for the tropics (especially at that high H2O concentration). It seems unlikely to get column delta-D values that are more depleted that the stratosphere. The discussion of this plot is also rather speculative to be honest.*

  Thank you for bringing up this problem. Actually, a bug in the plot script caused the contour lines to be shown transposed (i. e. H_2O and δD axes swapped). This has been fixed now. In this context, the description of the plot has been changed as well. We agree that the description is to some degree speculative. Thus, we have changed "is most likely" to "may be".

- *Figure 12: I would have preferred a linear pressure scale. The scale height of H2O is really low, so this figure emphasized pressure (i.e. height) ranges that are somewhat irrelevant. A linear pressure scale would emphasize the lower height ranges (say 0–3km) more.*

  The averaging kernel plots now have a linear pressure scale.

- *"That has been corrected in the new version. However, there is still a stripe in the new version of the plot. The requested plot with the location of the low humidity data is shown in Figure 1. Currently it is unclear whether the stripe is an artefact. The issue will be further investigated."*

  Have you looked, at least in a cursory fashion, at a potential cause for this artifact? It seems important to me to figure that out, rather that just deferring to further analysis. It can't take too much time to at least dig into it a bit deeper than the authors did so far. It would be a shame to have a plot with an obvious problem in a peer reviewed publication, as a reviewer I shouldn't really allow it. A simple explanation might be that the error in delta-D blows up if H2O is low (but this can be filtered).

Before our last answer, we have looked at several quantities such as column averaging kernels at various altitudes near the surface, surface albedos, effective cloud parameters, reduced χ^2 of pre-fit and final. We show most of these plots in Fig. 1 and 2. We do not see any significant differences to other regions where δD values are considered realistic. For example, the ground pixels you consider artefacts have increased averaging kernel values as expected below clouds, but other ground pixels with realistic δD values have similar averaging kernel values.

With a filter for low H_2O the stripe could be eliminated by construction, however filter criteria should not be based on a single feature in a single overpass in a single orbit. It is important to consider effects for different types of scenes when changing filter parameters. Filtering out H_2O columns below 7×10^{21} molec cm^{-2} (which removes the stripe in question) slightly worsens the validation, although the difference is small. This does not support such a filter for the absolute value of H_2O.

The stripe consists of a relatively low number of ground pixels that were highlighted due to the logarithmic colour scale of the plot. We have changed the colour scale to linear to give a better impression.

Editor comments

Thank you for your suggestions for improvement.

Technical corrections:

- *Abstract: A key feature of the retrieval is that also measurements under cloudy conditions are considered but **only if the cloud height is low**. However, the abstract reads as if **all** clouded observations are considered. Please specify this important selection of low clouds only already in the abstract.*

  This is now mentioned in the third sentence of the abstract.

- *Lines 35/36: Consider reformulation; in the current text, the reader might refer the "upgrade" of the spatial resolution to Sentinel 5.*

  We now write "5.5 km $\times$ 7 km (7 km $\times$ 7 km before August 2019)".

- *Line 62: Please provide an argument why H218O can be fixed in the retrieval.*

  We have added the half-sentence "since the absorption is very weak".

- *TCCON: I propose to also add a map of all stations, or to add the stations to Fig. 16 (a).*

  Locations of the stations shown in Fig. 9 have been added to Fig. 16a.

- *Fig. 18: Colormaps are somehow a matter of taste, but since clouds are white I was first confused that the cloud map uses white for indicating no clouds. You might consider a different colormap e.g. from darkblue or dark gray to white.*

  The colormap for cloud fraction has been changed to blues (from blue to white).

- *The author contributions are quite vague. It is not clear who did all the analysis. Please specify.*

  We now specify the contributions as follows: "AS made the retrievals and performed the analysis with help from TB, JadB, AL and JL."

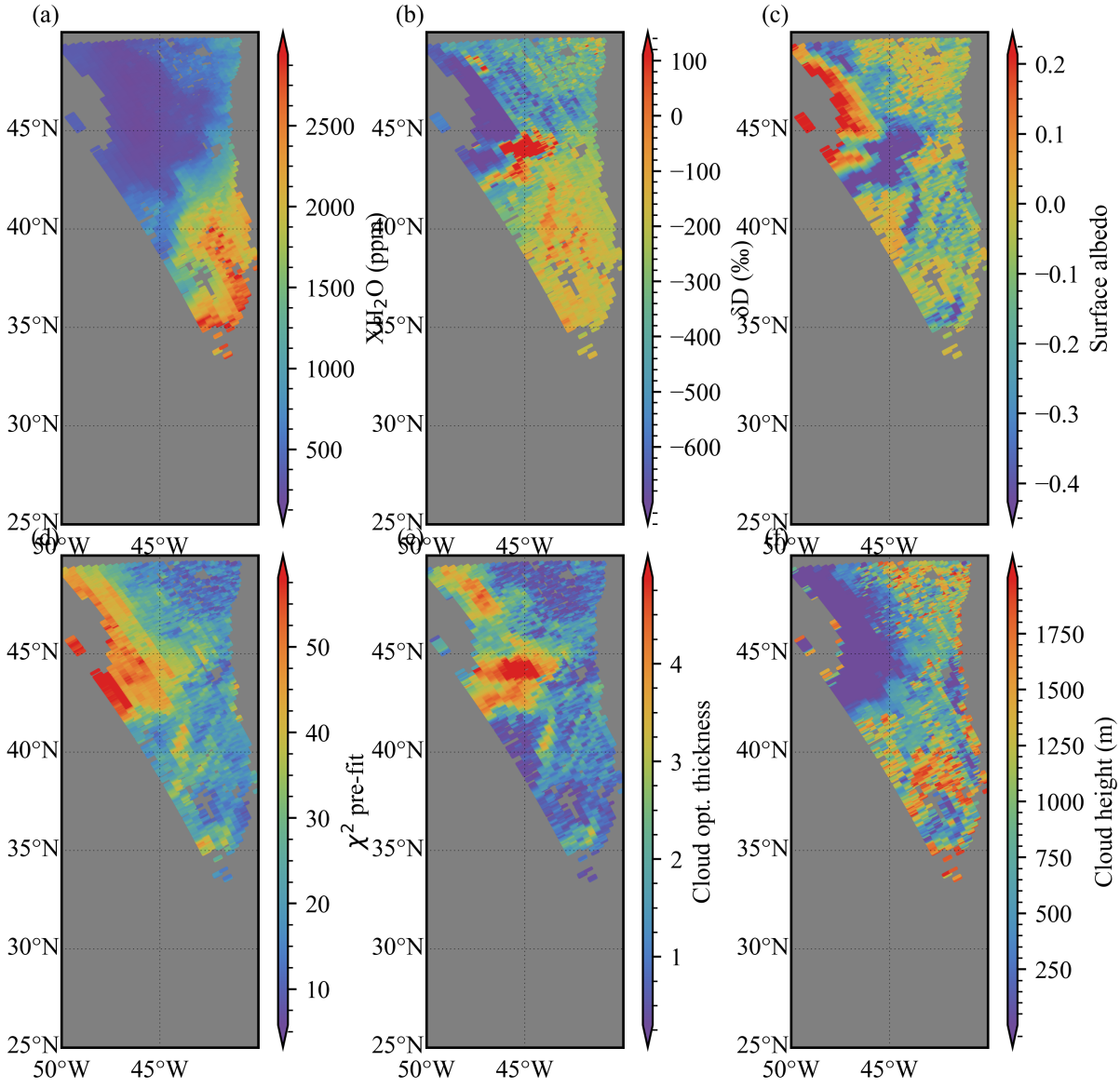

Figure 1: Single overpass results in the cold sector from orbit 11750 on 19 January 2020 showing **(a)** XH_2O, **(b)** δD, **(c)** surface albedo, **(d)** χ^2 in the pre-fit, **(e)** effective cloud optical thickness, and **(f)** effective cloud height.

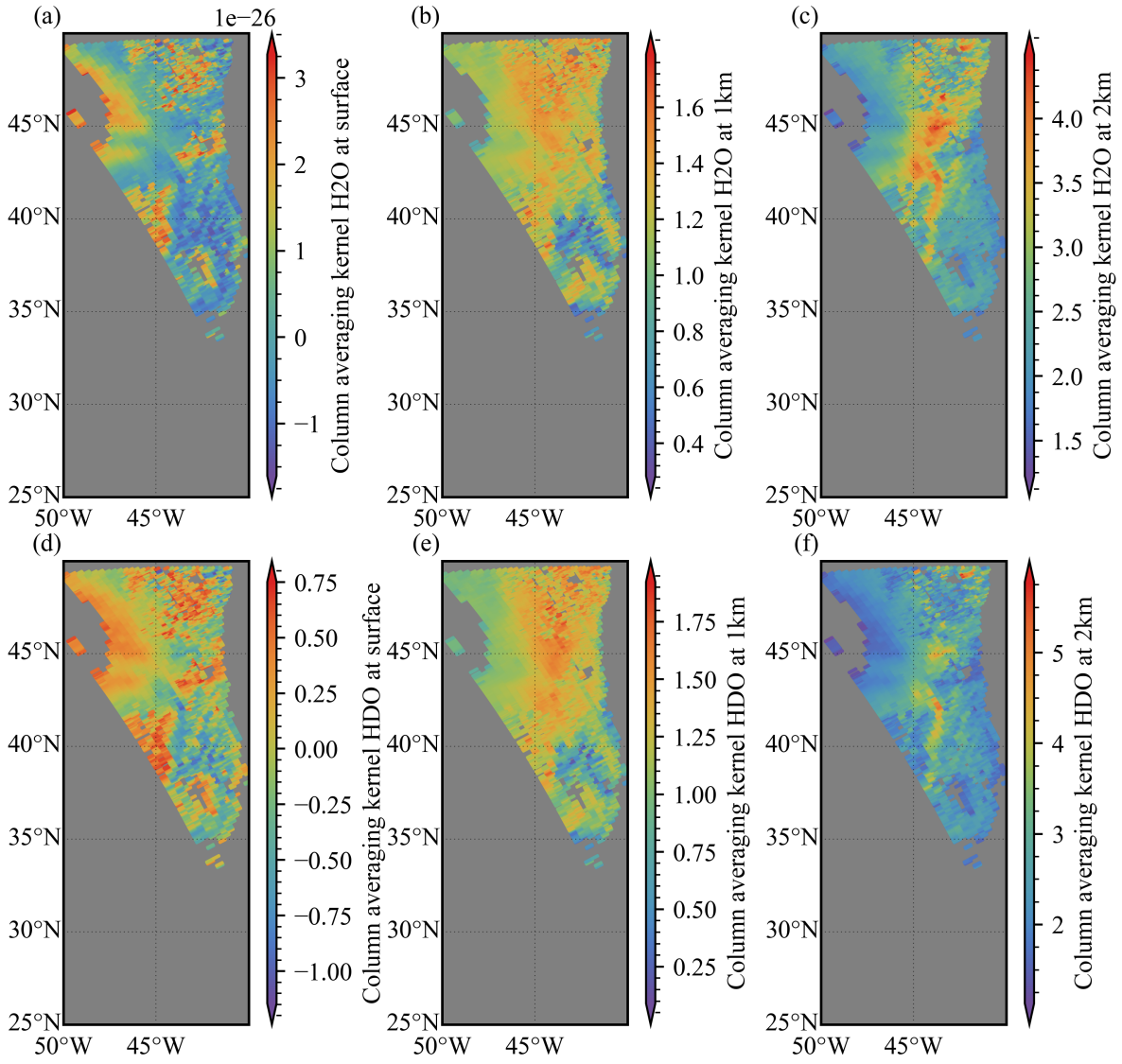

Figure 2: Single overpass results in the cold sector from orbit 11750 on 19 January 2020 showing column averaging kernels **(a)** of H_2O at the surface, **(b)** of H_2O at 1 km altitude, **(c)** of H_2O at 2 km altitude, **(d)** of HDO at the surface, **(e)** of HDO a 1 km altitude, and **(f)** of HDO at 2 km altitude.