Authors' response to the comments of Referee #2 on "MICRU background map and effective cloud fraction algorithms designed for UV/vis satellite instruments with large viewing angles" by Holger Sihler *et al.*

We would like to thank Referee #2 for the review of our submission to AMTD and for contributing helpful comments and suggestions to improve the quality and clarity of our manuscript.

For reference, the original Referee comments below are typeset in black, our responses in blue. Modifications of the original manuscript (green) are indicated in red.

Review of "MICRU background map and effective cloud fraction algorithms designed for UV/vis satellite instruments with large viewing angles" by Sihler et al.

The manuscript describes a model for accounting for anisotropic reflection of solar light from Earth's surface in an effective cloud fraction algorithm designed for UV/Vis satellite instruments. Results of the application of the algorithm to GOME-2 data are compared with other cloud fraction algorithms. Appendices provide technical details of the developed algorithm. The manuscript is clearly relevant for AMT. Even though the material is not a significant advance in remote sensing of clouds it could be published to document the GOME-2 cloud fraction algorithm in the literature. The abstract provides a concise and complete summary of the paper. The earlier work is properly credited. I recommend publication of this manuscript only after major revisions which address the following comments.

General comments

1. The authors do not clearly state what are the main improvements of the proposed algorithm as compared with the existing cloud algorithms which also accounts for surface BRDF. It would be useful to summarize those improvements in Conclusions.

To the knowledge of the authors, there are currently no cloud products for GOME-2 featuring BRDF effects for the background map. The improvements of this feature of MICRU compared to OCRA and FRESCO products are investigated in the manuscript.

MICRU features a background map computed from the measurements themselves also considering two parameters for degradation. These measures have the potential to improve the accuracy of small CF significantly.

The proposed MICRU implementation also features cloud fractions measured at different acquisition times minimizing systematic errors due to spatial aliasing, which is typical for GOME/GOME-2 instruments.

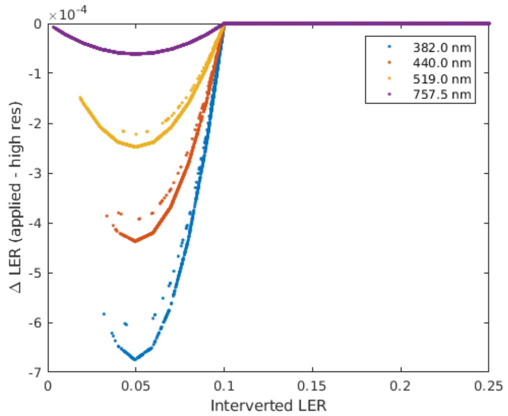
Hence, MICRU is a universal algorithm, which may be consistently applied to other spectrometers and also imaging instruments as well.

We edited the conclusions of the revised manuscript:

The unique feature of MICRU is the application of an empirical BRDF surface model accounting for viewing angle dependencies in the cloud retrieval. The paper demonstrates that MICRU CF depend significantly less on VZA compared to other available CF products for GOME-2 and, hence, are significantly more accurate. MICRU determines the lower threshold from the measurements themselves furthermore reducing biases due to calibration and degradation issues.

2. Low values of LER are of the primary interest for the construction of a minimum LER map (background map) which is the core of the developed algorithm. The existing surface reflectance data sets (see e.g. Kleipool et al., 2008) show that an overwhelming fraction of Earth's surface has reflectance lower than 0.1-0.15 in the UV/Vis spectral range with wavelengths shorter than 500 nm. This spectral range is most important for trace-gas retrievals. The background map is constructed using a look-up table that relates top-of-the-atmosphere radiance and LER. Table 4 lists the nodes of this look-up table. The step of 0.1 in LER nodes in Table 4 is quite insufficient for calculations in the low LER range. Any interpolation with so sparse nodes would lead to high errors in the low LER range thus in the minimum LER map. The authors should add more nodes of LER for its low values and provide an estimate of interpolation errors. The paper cannot be recommended for publication without addressing this comment.

We thank the Referee for highlighting this issue. We conducted some test in beforehand and we were also surprised, how small the error caused by this simplification actually is. We performed the reflectance to LER conversion for one orbit of GOME-2 MSC data for four different wavelengths using two different LUTs, one as described in the paper and one, where 9 additional nodes between 0 and 0.1 at steps of 0.01 are included. The differences of the inverted LER are:



Obviously, the error introduced by linear interpolation applying the LUTs in question is always smaller than 0.001. The error for the CF must be even smaller, as the same systematic error is performed twice (background map calculation and its later Look-up), which should cancel out almost completely due to a change in sign. Furthermore, we would like to note that interpolation errors applying basic linear interpolation for strictly monotonic functions - as the TOA radiance to LER dependence – are small. Hence, we decided against performing our calculation with larger LUTs but rather informing the reader of the interpolation error.

We add the following paragraph to Section 2.2 the revised manuscript:

It needs to be noted that the resolution of the LUTs in LER direction may appear rather coarse. However, the difference of the obtained results to preliminary RT computations featuring a 10 times higher resolution were found to be < 0.001 in the UV and even one order of magnitude less in the red spectral region.

3. In Appendix C, the authors consider the spectral dependence of BRDF model parameters. Those internal parameters are used to build the minimum LER map. It would be useful if the authors would consider the spectral dependence of the final product of the developed algorithm, namely the effective cloud fraction. There is some contradiction in interpreting the spectral dependence of the

effective cloud fraction. Formally, the effective cloud fraction is wavelength dependent because it is defined by spectral quantities (Stammes et al., JGR, 2008). However the radiative transfer simulations show that the cloud fraction is nearly invariant with wavelength over a wide spectral range (Gupta et al., AMT, 2016).

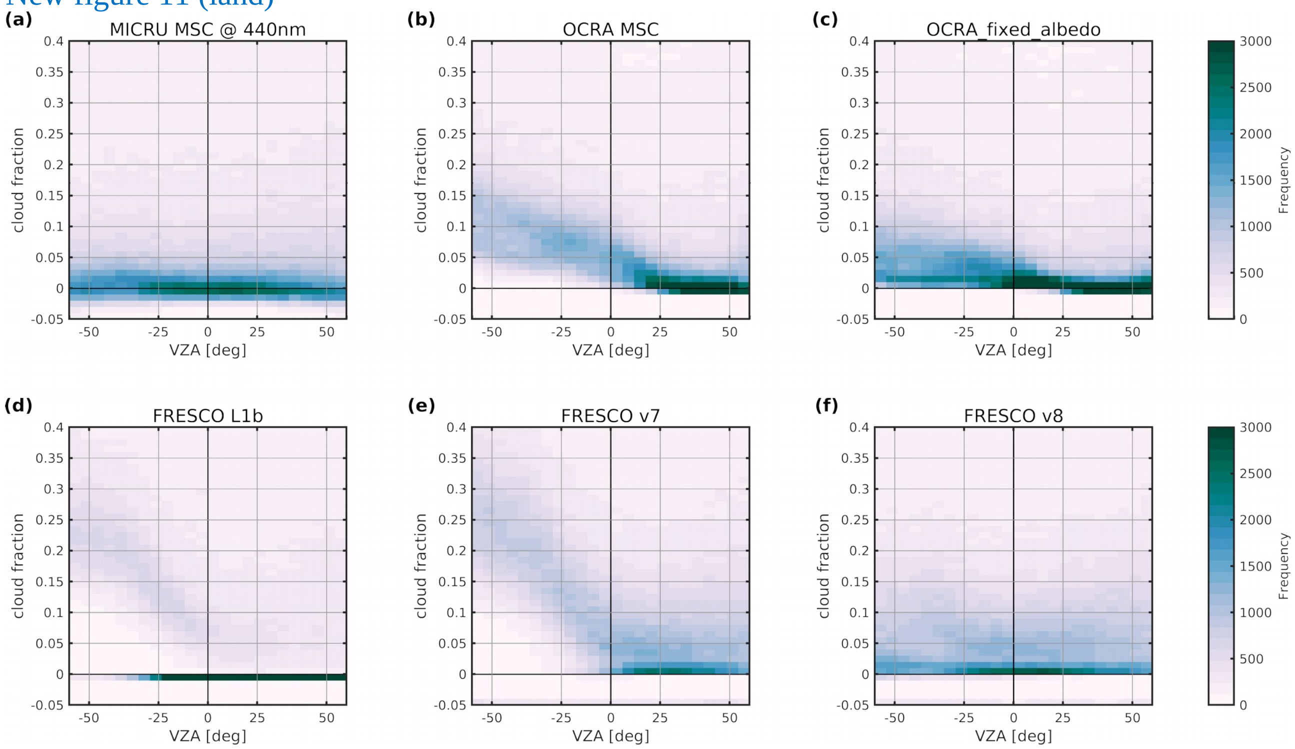
We thank the Referee for this interesting question. We, however, doubt that our results are suitable to address this issue for a number of reasons:

- MICRU is optimised for retrieving small CFs at high accuracy over the entire swath of GOME-2. Larger CF have larger errors.
- The applied GOME-2 level 1b suffers from degradation and its absolute accuracy is not ideal for such comparisons.
- The lower threshold maps are not absolute LER as they also account for degradation effects, which differ between MICRU channels.
- GOME-2 is a scanning instrument resulting in different measurement PSFs at different wavelengths, that is spatial aliasing.

4. I strongly recommend to show and analyze the cross-track dependence, i.e. dependence on VZA, of the cloud fraction. Accounting for BRDF effects on the cloud fraction would flatten the cloud fraction cross-track dependence reducing possible biases related to not accounting for anisotropic reflection of solar light from Earth's surface. Particularly, it is important for the ocean where the sun glitter can significantly affects cloud pressure retrievals. The authors are encouraged to compare their results with those in the following paper:
Fasnacht et al., A geometry-dependent surface Lambertian-equivalent reflectivity product for UV-Vis retrievals – Part 2: Evaluation over open ocean, Atmos. Meas. Tech., 12, 6749–6769, 2019.

We thank the reviewer for this valuable suggestion. We performed an analysis of the statistics depending on viewing direction and included the following two figures to the revised manuscript.

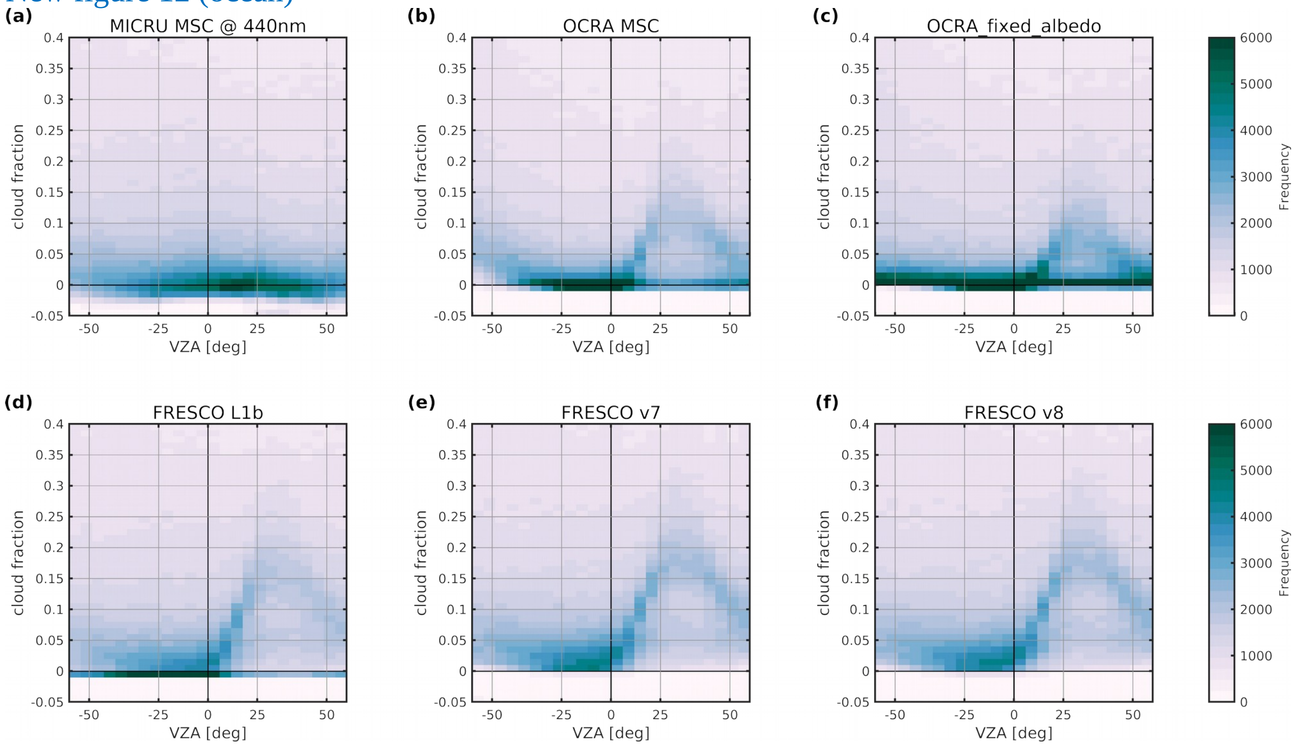
New figure 11 (land)



Caption: Comparison of viewing angle dependence of small CFs over land between the 55° parallels: (a) MICRU MSC at 440 nm, (b) OCRA MSC, (c) OCRA_fixed_albedo, (d) FRESCO

L1b, (e) FRESCO v7, and (f) FRESCO v8. Statistics based on April 2010 data.

New figure 12 (ocean)



Caption: Same as Fig. 11 but over ocean. The influence of sun glitter is evident for positive VZA for most products, even though measurements flagged sunglint risk (Table 7) are filtered.

Accordingly, we added to Section 3.1:

Another estimate of the residual VZA-dependence may be assessed by analysing the cross-track dependence of the lower CF accumulation point displayed in Figs. 11(a) and 12(a) for land and ocean surfaces, respectively. Over land, the small CF accumulate between -0.02 and 0.03 almost evenly over the entire swath. Over ocean, small CF are slightly more scattered. The distribution dilutes significantly and reveals a slight positive bias towards west (negative VZA in Fig. 12).

and to Section 3.4.1:

An additional view on the VZA-dependence of MICRU and OCRA CF is provided by Figs. 11(a)–(c) and 12(a)–(c) for land and ocean surfaces, respectively. Over land, the accumulation points of small OCRA MSC are significantly biased high for all negative VZA. The albedo correction for OCRA_fixed_albedo (Fig. 11(c)) improves the situation significantly confining the accumulation point to CF < 0.1, which is still significantly larger than the MICRU CF in Fig. 11(a). Over ocean, both OCRA investigated versions reveal a significant bias from sun glitter for positive VZA in Figs. 11(b) and (c). Towards the west (negative VZA), the lower accumulation point of OCRA_fixed_albedo is more populated than MICRU.

and to Section 3.4.2:

Figures 11(d)–(f) and 12(d)–(f) detail the VZA-dependence of the three FRESCO versions over land and ocean, respectively. Over land, the accumulation points of small CF are significantly biased high for negative VZA, especially for FRESCO L1b and v7. The distribution of FRESCO v8 CF (Fig. 11(f)), however, is almost independent of VZA. Over ocean, the differences between the

FRESCO versions are small and the bias from sun glitter is again significant, which is consistent with other results.

While we acknowledge the results by Fasnacht et al., they may not easily be compared to our results for two reasons. Firstly, MICRU retrieves CF using a lower threshold, which may not be readily compared to the GLER retrieved by the cited work. Secondly, it is not straightforward to compare OMI and GOME-2 due to their different orbital parameters and measurement times. Unfortunately, the authors of Fasnacht et al. responded to our request of GOME-2 data that they did not yet apply their algorithm to GOME-2.

However, we include a reference to Fasnacht et al. to our short review in the introduction (p.4, l. 10) to acknowledge their results.

5. In my opinion, the manuscript is too long and somewhat overloaded with technical details. More technical details could be moved in Appendices. For instance, Section 2.3.2 can be either cut down or moved to an appendix. Section 4 returns to Fig. 8-20 which were already discussed in the previous section. I would recommend to combine Sect. 3 and Sect. 4 to avoid possible duplication.

We understand the Referees concerns regarding the structure and length of the manuscript. We agree that it is not brief, but we believe its length is justified because we compare our results to several other products. This careful comparison is at least partially motivated by our activities in the verification of the operational TROPOMI/S5P algorithms. Our feedback from the TROPOMI/S5P and S5 community is quite positive and, therefore, we would like to not shorten our result section further.

Furthermore, we refrained from combining Sects. 3 and 4 because we see significant scientific advantage in separating the result from their discussion because we want to base our thorough discussion on all results, which, logically, need to be presented before. Short comments in Section are merely signposts to guide the reader.

As suggested by the Referee, the iterative surface fitting section (Sect. 2.3.2) is moved to the Appendix in the revised manuscript. Citations to this section are edited from Sect. to Appendix, respectively.

Specific comments

The title does not clearly reflect the contents of the paper. MICRU is not a common acronym. It is not clear what "background map" means. The title does not reflect that the paper is dealing with accounting for anisotropic reflection (BRDF) of solar light from Earth's surface in cloud algorithms.

We opted for a brief title and omitted the specific feature of MICRU background maps:
MICRU: a sophisticated effective cloud fraction algorithm designed for UV/vis satellite instruments with large viewing angles

P.6, L4 and elsewhere. The letter T is commonly used to denote the transmittance in radiative transfer. To avoid confusion it is desirable to select a different symbol for LER.

We agree with the Referee, that the choice of the symbol T for LER was not ideal and may also lead to confusion with the lower threshold T_{min}. Instead of replacing T we decided to remove this symbol from the manuscript altogether and replacing it by the term (the) LER as often applied in AMT articles.

P.10, Fig.4. T_min in the figure capture is not defined yet.

The last sentence of the caption is rephrased in the revised manuscript (also see comment by Referee #1 on same caption):

The highest resolutions for MSC and PMD lower threshold maps are 0.1°×0.05° and 0.0125°×0.05°, respectively, and denoted binning #1 as in Table 6.

P.11, L.1. It is not clear how the land sea mask is applied to a nominal GOME-2 pixel? Is a land/sea fraction within a pixel known? Please clarify.

The algorithm to determine the land/sea fraction in each pixel is described in the second and third paragraph in Sect. 2.1.2 (P. 10). The second paragraph of Section 2.1.2 of the revised manuscript is changed to:

MICRU features a separate T_{min} parametrisation for measurements over land and ocean, respectively. An accurate description of the land and water transition is therefore crucial for the accurate interpolation of T_{min} at coasts. An algorithm specifically developed for MICRU derives the fraction of water and land in each satellite pixel at high resolution. As input, the land sea mask (LSM) compiled from revision 679 of the GSHHG coast line database (Wessel and Smith, 1996; NOAA, 2018) is applied. The polygon data from intermediate GSHHG resolution neglecting polygons smaller in area than one GOME-2 pixel is first sampled at 0.1°×0.05° and 0.0125°×0.05° for MSC and PMD, respectively, and then convolved with the corresponding PSF (cf. Figure 4). The convolution yields a global map of fractional land cover ranging between 0 and 1 representing complete water and land coverage, respectively. Hence, T_{min} values for land and ocean may be interpolated for each satellite pixel based on the convolved land cover map. The interpolated fractional land cover values are later also used for flagging (Sect. 2.5.3).

where the sentence

The LSM is processed at eight times higher longitudinal resolution for PMD compared to MSC taking advantage of the smaller PMD pixel size.

is erased.

P.11, L.7. Please provide a reference to GTOPO30.

Citation to https://doi.org/10.5066/F7DF6PQS added.

P.11, L.26. Please specify the wavelengths at which the absorbing aerosol index is defined. Its threshold value used for filtering the data depends on the wavelengths.

We agree with the Referee that the wavelength should be specified in the manuscript. The revised manuscript now includes:

The reflectances used for the determination of the AAI at MSC resolution are centred at 340 and 380 nm. For the AAI at PMD resolution, PMD-PP bands 4 and 6 at 338 and 382 nm are applied, respectively.

P.12, L.3. While doing RT computations in a spherical atmosphere the authors do not account for the atmospheric refraction. Please provide a justification for neglecting the refraction effect?

Actually, SCIATRAN is able to do calculations in spherical geometry either with or without refraction. For our LUT calculations, refraction was turned on. Furthermore, we would like to note that the influence of refraction on our calculation, which do not include the limb geometry, is minimal.

In order to clarify this setting, we appended
and accounting for atmospheric refraction
after h=0m (p.12, l.4).

P.12, L.5. The use of a single value of 250 DU for total ozone column may not be sufficient for wavelengths within the Chappuis absorption bands in case of high solar zenith angles (Table 4 lists the angles up to 87 deg.).

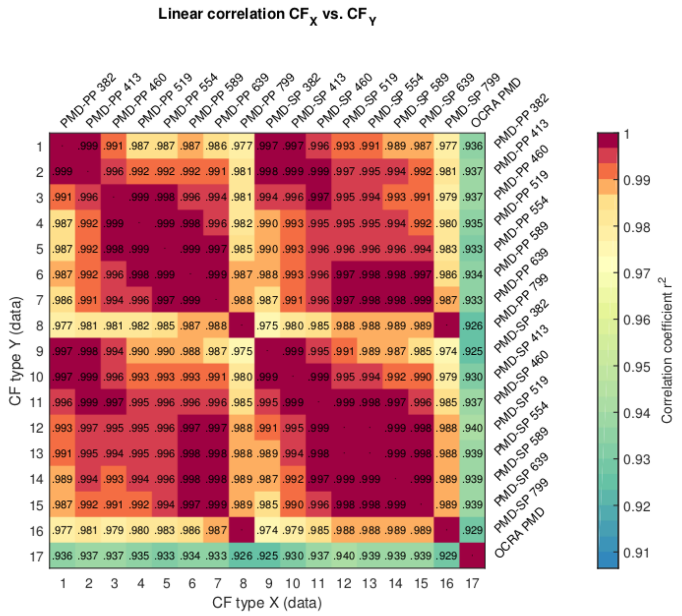
We thank the Referee for addressing our simplification of the RT. Our motivation is to decrease the amount of input information on MICRU and, hence, to actually provide an independent product/piece of information. In order to estimate the influence of variations of the ozone column on the CF retrieved within the Chappuis band, we want to perform a worst case estimation:

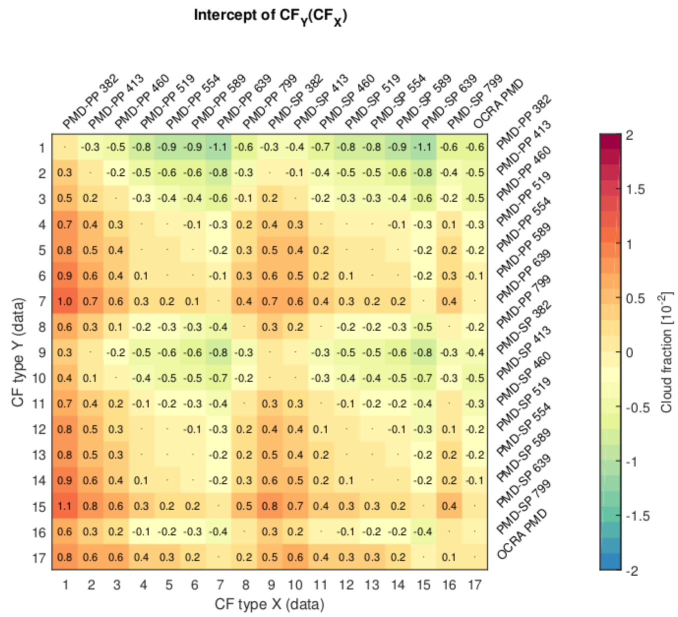
- The MICRU channels most affected by the Chappuis band absorption are PMD channels 6 and 14 measuring between 568-613 nm, where the O3 cross section \sigma=4.8e-21 cm^2/molec. For MSC channels 11 and 12, the cross section is less than 2e-21.
- The worst case air mass factor up to 55° latitude computes from an SZA=86 and VZA=54 using AMF = 1/cos(SZA) + 1/cos(VZA) to 16
- RT calculations are performed using 250 DU=6.75e18 molec/cm^2. We will compare the calculated TOA reflectances to those at 500DU=6.75e18. Multiplied by the AMF, this yields slant columns of S1=1.1e+20 and S2=2.2e+20 molec/cm^2, respectively.
- The Transmission at 250DU then computes to T(250DU)=exp(-S1*\sigma)=0.59 and T(500DU) = 0.35, respectively. The difference corresponds to a reduction to approximately 60% TOA reflectance compared to a measurement affected by 250DU.
- For MICRU, the goal is to achieve 4% accuracy on the lower threshold, which is the minimum TOA reflectance. At 600nm, the minimum TOA reflectance is below 0.1 for 80% of the measurements (determined from an April example orbit, partially over Africa).
- With an average upper threshold of 0.73, this can yield a CF error of 0.1*0.6/0.63 = 10% for the lower threshold. This exceeds the 4% error goal of MICRU for the accuracy of small cloud fractions.

It needs to be noted, that this 10% error is a conservative estimate of the error caused by Ozone column variations. The issue is compensated by the following factors:

- The AMF of 16 is quite an extreme case, which only seldomly occurs for tropospheric measurements at latitudes lower than 55°.
- The column variation of Ozone at the same latitude are usually much smaller than 250DU. Exception: Ozone hole conditions.
- The empirical approach of MICRU may also compensate systematic influences of the RT. For example, the offset may compensate an errorenous average ozone column and the fitted parabola may account for its influences for larger viewing angles. Furthermore, the manuscript discusses that systematic instrumental effects have, for GOME-2, a larger effect on the CF accuracy compared to RT errors.

In total, we only find a negligible effect in the MICRU results. The correlation matrices for all PMD channels corresponding to Figures D2(a) and (b) are





Here, channels 6 and 14 do not show a significant variations compared to their respective neighbors. Reduced linear correlation with increased channel difference may be explained by spatial aliasing. The CF intercepts are mostly less than 1% and all well below the 4% limit.

In order to convey our considerations to the reader, the revised manuscript is edited to

The O_3 column is fixed to 250 Dobson Units (DU) in order to reduce the number of required input parameters. This simplification may affect MICRU retrievals within the Ozone Chappuis band, most notably PMD channels 6 and 14, and, to a lesser extent, MSC channels 11 and 12. Preliminary results, however, showed that errors are on average negligible as the empirical approach of MICRU reduces the influence of systematic errors.

P.14, L.6. Why is the glitter reflectance, r_g, defined as an independent variable? It depends on the sun-view geometry, e.g. on the viewing zenith angle which is specified as an independent variable. Please clarify.

MICRU applies an empirical surface reflectance model, where the parameter specifying the contribution by sun glitter is an independent parameter. The sun glitter itself is a function of viewing geometry and wind speed, and, hence, linear independent from the other three independent variables: time, VZA, and scattering angle.

P.19, L.23-24. The authors say " Longer time-series increase the probability of including measurements not contaminated by clouds." Please provide actual numbers that characterize the duration of timeseries.

This issue is thoroughly discussed by Krijger et al., 2007. We include this citation in the revised manuscript. The paragraph at the specified location now reads in the revised manuscript:

Temporally, larger subsets should be favoured over smaller ones unless there are significant changes of surface properties or the instrument response degrades much differently than considered in the model (Sect. 2.3.1). For example, for MSC binning 1 there are 400 equatorial bins over land with 15 or less measurements considered cloud-free by the T_{min}-retrieval despite applying a study period of 77 months. Longer time-series increase the probability of including measurements not contaminated by clouds (Krijger et al., 2007).

P.22, L.4. Please give a reference to FRESCO v8.

Actually, there is no peer reviewed paper about FRESCO v8 for GOME-2. The algorithm, however, is described in a EUMETSAT report about FRESCO+ v2 using DLER, which is the same as FRESCO v8 at KNMI and in our work. The report is available from the TEMIS website

Wang, P., Tuinder, O., and (KNMI), P. S.: FRESCO+ version 2 for GOME-2 Metop-C processing, Internet, http://www.temis.nl/fresco/frescopv2_metopc_WP1_report_20181026.pdf, last access: 16 November 2020, 2018.

and we appended the reference (Wang et al., 2018) to line 4 on page 22 of the revised manuscript.

P.24, l.21-22. Fig. 8(f) shows the minimum LER residuals, T_min (stated in Line 14). However, the authors say that "... average deviations much smaller than 0.04, which is the targeted accuracy of MICRU CF". Please clarify how the LER residual of 0.04 is related to the targeted accuracy of cloud fraction.

We thank the Referee for this helpful comment. We agree that this statement may be confusing for the reader. We decided that the subclause

, which is the targeted accuracy of MICRU CF

is actually redundant and removed it from the revised manuscript.

P.25, L.2. In the discussion of Fig. 9, CF is mentioned. Please clarify what parameter (LER or CF) is shown in Fig. 9.

We thank the Referee for pointing this out. We replaced CF by LER in the revised manuscript.

Section 3.2 compares cloud fractions from different algorithms. Please specify the wavelengths at which the cloud fractions are retrieved. Can the observed differences between MICRU and FRESCO/OCRA CF retrievals be due to the wavelength difference?

We thank the Referee for this thoughtful consideration. However, we draw a different conclusion from our results, because CF from all MICRU channels are consistent (Figures 15, E1, and E2). The intercomparison of MICRU results furthermore illustrates that noise increases towards larger wavelengths due to larger reflectivity and increased uncertainty. FRESCO and OCRA both apply measurements from the red spectral region, which may partially explain the differences. Furthermore, as presented in Section 3.2, MICRU performs much more reliable over cloud-free scenes compared to FRESCO and OCRA because it applies a more elaborate parameterisation of the lower threshold, which is an effect independent from wavelength.

In order to clarify this issue, the applied wavelengths are added to Section 3.2 of the revised manuscript as requested by the Referee.

MICRU MSC and PMD results are specifically obtained at 440 and 460nm, respectively. OCRA results are based on PMD measurements between 321 and 804nm. FRESCO applies the O2A-band at 757.5nm.

and added to the third paragraph of Section 4.1 (Discussion) of the revised manuscript:

This effect is almost independent from wavelength and consistent for all MICRU channels (cf. Fig.

15 and Appendix E).

Section 3.3.2. Please explain why there are CF differences retrieved at different wavelengths for high values of CF. For high values of CF, possible surface effects could be neglected. Given the cloud backscatter spectrally independent, the CF values at different wavelengths seem to be same.

We agree with the Referee that this issue should be addressed in the manuscript. We are certain that this effect is dominated by data and instrumental deficiencies. We added to Section 3.3.2 of the revised manuscript:

Figure 15 furthermore indicates, that the CF slope differs between MICRU channels, which is discussed in Sect. 4.1.

We furthermore modified the respective paragraph in Section 4.1:

Another aspect of the MICRU MSC channel intercomparison are differencies at different wavelengths for high values of CF and, hence, slopes deviating from unity as, for example, shown in Fig. 15(a). CF at 382 nm are biased high with respect to those retrieved at 440 nm while the intercept at zero CF is negligible. Hence, the definition of T_{max} apparently deviates between MICRU channels, which should be independent from surface effects. Figure E2(c) comprehensively compares the slopes of all MICRU channels. There is a significantly biased slope for MSC channels 1–4 retrieved at 389.7 and below. This step between MSC channels 4 and 5 may be attributed to the application of different GOME-2 bands, specifically bands 2B and 3, from which the MICRU channels are extracted (cf. Table 1). Hence, we conclude that differences between MICRU channels at high CF values are dominated by instrumental effects and calibration deficiencies of the input data. We would like to note that we observed also the CF accuracy degrading near GOME-2 band edges when fine-tuning the MSC channel definitions (Table 1). The degradation depends only weakly on kernel width leading to the conclusion that this is a broadband effect. Furthermore, interferences with molecular absorption and atmospheric scattering above the clouds resulting in a wavelength dependent R may also cause a systematic slope bias. It needs to be noted, however, that the influence of the slope on the accuracy on small cloud fractions is minor.

Section 3.4.1. What is a conclusion of the comparison of MICRU and OCRA? Do the authors attribute the differences between the algorithms to the different treatment of surface BRDF?

The comparison between MICRU and OCRA is discussed in Section 4.2. The revised manuscript now states:

- The accuracy of singular OCRA measurements, however, is significantly and consistently lower compared to MICRU as revealed by the larger scatter of OCRA CF for very small MICRU CF than vice versa
- it can be concluded that OCRA's empirical correction algorithm is a bit too optimistic
- OCRA seems to properly account for this effect on PMD resolution
- This indicates that BRDF effects have a stronger influence on OCRA results for observation geometries opposing the sun and that the empirical VZA correction performed by OCRA is not sufficient.

P.31, L.6. "... different definition of the upper threshold." What do you mean? What is the upper threshold for OCRA?

This issue is now discussed in more detail following the comments of Referee #1. Changes to manuscript are detailed in our answers to his/her comments.

Section 3.4.2. Please formulate a purpose of comparing MICRU with three versions of FRESCO. Why do not select just the latest version of FRESCO?

All three FRESCO versions apply different strategies for supplying the background map leading to significantly different results as presented in the manuscript. The latest version actually not outperforms its predecessors in all aspects as already discussed in Section 3.4.2

We added the following motivation to Section 2.6 (Comparison data):

in order to study the particular differences with respect to background map generation and residual VZA dependence