Dear Reviewer 3,

Thank you for reviewing our manuscript. You provided your full review already during the quick access phase. That is why we already sent a full reply to the editor after the quick access reviews. To provide an improved manuscript to the other reviewers, we already did consider the changes to the manuscript text you requested before the discussion phase. That means, the other reviewers already see a manuscript version which includes the changes discussed below.

We considered many of your smaller points and do agree, but we do not agree with your main point of criticism. Please see below for our point-by-point reply.

Best regards,
Tobias Zinner (and co-authors)

Reply to reviewer 3:

Reviewer comments are highlighted in gray.

I include my complete review. It can be used when/if the paper is posted in AMTD.

The paper presents an algorithm for determination of cloud shape/geometry using cross-track scanning spectrometer making measurements in O2 A-band. The authors made an extensive retrieval error estimation effort and honestly discuss challenges facing the proposed approach. These challenges appear to be so serious that the whole retrieval algorithm needs substantial modification (see my comments below for details). I hope that the authors will be able to improve their technique during the discussion and reply to reviews period, so the paper could be then accepted after major changes.

General comments

The main issue affecting the proposed retrieval technique is multiple scattering of light in essentially 3D cloud geometry. This creates a kind of a circle: cloud shape is determined point-by-point using LUT based on 3D radiation transfer model, which itself has to imply a specific shape of cloud as a whole. Simplified assumption of cloud shape ("cloud wall") leads to substantial bias (3.8 km) in the derived cloud position compared to an independent stereo dataset (which is assumed to be more robust).

The authors should also note that the same stereo dataset cannot be used for both correction of A-band retrievals and for validation of them. The 3.8-km offset applied to A-band data (making it to agree better with stereo results) is discussed in the paper, however, it is not even mentioned in the abstract when the accuracy of the proposed approach is described. This misleads the reader.

The paper in its present form describes a technique that currently does not work, and its results are artificially forced to agree with stereo dataset for "validation". This is certainly is not worth publishing. In my opinion the only way to save the paper is to follow the authors' own suggestion (last paragraph of Conclusions Sec.) and incorporate stereo measurements into A-band algorithm as an occasional "calibration" source. (Note that stereo measurements may not be used for validation then.) Only after this is successfully done (which is a "major change") paper can be accepted for publication.

English in the paper is generally acceptable except for few sentences that need clarification. However, the paper (Conclusions in particular) would certainly benefit from proof-reading by a native English speaker.

We do agree with the reviewer that we should have made limitations of our method even more clear already in the abstract. We extended abstract and description accordingly. However we do not think that a fundamental extension of our method with additional "calibration" stereo data is necessary or would change the message of this paper. Already in the original manuscript we do suggest the combination of both data sources for our purpose and discuss the limitations.

Our main goal is a determination of cloud surface orientation and cloud points' vertical height for our cloud side view for a specific campaign data set (ACRIDICON-CHUVA). For a plausibility check we compare our data to stereo-points. We find an offset (3.8 km) which is most likely caused by 3D effects. Apart from this deviation, errors are small and within the expected and described ranges. As 3D effects are the most likely explanation for sign and size of the offset, we suggest to correct the offset to match the mean stereo distance.

On purpose, we did not use the word "validation" in the manuscript for this comparison of stereo and oxygen-A derived distance, because we do not consider the stereo distances to be free of uncertainty and even systematic effects (e.g. a window distortion effect).

For one of our main goals, the determination of cloud surface orientation for the correction of 3D effects of other retrievals, the mentioned offset does not introduce any error. For our second goal, the determination of a cloud height for specMACS derived products, we would add an additional (and problematic) bias, if our attribution of the offset to 3D effects is not correct. The second point has been clarified in the revised manuscript.

Nonetheless, we consider our proposed method and its detailed analysis to be careful and sufficiently complete to be instructive for the growing community using spectral solar remote sensing methods and by our group for application in successive studies. Any following paper on use of oxygen-A band distance derivation in complex cloud situations, whether our own group's application or any other author's, could be compared to our open and detailed analysis in this manuscript.

Apart from these argumentation, it has to be emphasized that all passive solar cloud remote sensing techniques suffer from systematic 3D effects due to real clouds' shape, well known by the community from the very beginning, but only incrementally quantified over time. Nonetheless a vast number of important and significant publications used these techniques, the good ones keeping in mind the limitations introduced even without proper quantification.

Point-by-point reply

Line comments
----

p.1, l.1:
Is ACRIDICON-CHUVA an acronym? If yes, spell it out (may be in the paper text rather than in the abstract).

Changed in the text.

p.1, ll.12-14:
The last sentence of abstract is misleading (see general comment above), since it does not mention the 3.8-km offset subtracted from A-band retrievals. This operation makes A-band retrievals dependent on stereo dataset, which then cannot be used as direct validation source.

It is mentioned in the abstract now.

p.1, ll.13-14:
Distance accuracy should be compared with typical intrinsic property of cloud field, such as horizontal scale (what was typical cloud size?), not with typical distance between cloud and aircraft.

The typical cloud type was added.

p.3, l.14:
Making measurements through aircraft window (rather than through instrument's own window) may present a problem. Any window can be dirty or scratched, and this should be factored into instrument's calibration. Unlike instrument's own window, aircraft's window cannot be taken into calibration lab. How specMACS was calibrated? Can additional absorption by the window potentially bias A-band measurements?

The specMACS calibration in Ewald et al 2016 includes the window.

p.3. l.18:
Add publication year to "Ewald et al."

Added.

p.6., ll.16-19:
This paragraph needs clarification: has the spherical cloud been moved up and down relative to the sensor height in order to achieve different cloud surface angles?

Correct. The text has been clarified there.

p.7. Fig.7m:
"sphere angle" is not defined or used in text.

Changed to "surface angle" as used and defined in the text.

p.7, caption; p.8, l.1; p.9, ll.11-12; p.10, l.6; etc.:
Change "zenith" to "zenith angle".

I did a "Search and replace" throughout the document.

p.7, caption:
Use subscript 's' for sensor altitude 'z' (as it is used later in text).

Done.

p.7, caption:
Caption is longer than the page, only part of it can be seen.

Corrected. Seems to be a problem of the discussion format.

p.8, ll.7-19:
Parts of this list are repeated in Table 1 and can be omitted.

That is correct. The list has been removed.

p.8, l.8:
The whole line is unclear. As viewing direction is not necessarily horizontal, what is "horizontal extent in viewing direction". What does "infinite, periodic perpendicular to it" mean? Please, clarify.

That is the the horizontal extent in cross-track direction using the aircraft coordinate system (away from an aircraft). It is not the along-track (flight direction) extent. In the model setup this second dimension is periodic. A 2 km thick cloud wall is used. I have added a sentence to explain the geometry.

p.8, l.19:
I suggest to include 'z_s' on this line, since it is used below it without definition.

No this is not "z_s". z_s is sensor height while the derived height is the "z_s – z_o2a". It now reads "In order to check the validity of distance and height measurements derived ...".

p.8, ll.23-26:
"For a given sensor height the cloud wall is visible under certain observation zenith angles only up to a certain distance." This implies certain finite cloud top and bottom heights, which are not defined or discussed. What are they in the model? This may also contradict l.8 (see above) stating that cloud is somehow "infinite". Please, clarify.

See answer to l.8 above.

p.9, l.14:
The phrase "filtering cloud edge areas for the remaining cloud situations" is completely unclear. And what is "cloud edge" (also used in other places in the paper) compared to "cloud surface"?

I tried to clarify. It became "cloud tops, which are always more or less horizontal". Other appearance of "edge" was changed accordingly.

p.10, l.4:
I suggest to define "R = R_O2A" here as it is used without subscript or with different subscripts below this line.

Added your suggestion.

p.10, eqs.2&3:
I am confused. Dependences of R = R_O2A on all parameters presented in Fig. 3 seem to be monotonic. Then, R_min < R_LUT < R_max or R_max < R_LUT < R_min (depending on definition of R_min and R_max) for any parameter (pressure, etc.). In both cases R_LUT cancels from the numerator of Eq. (2), thus,
Delta R = 2 sigma_R.
Why this relation is not satisfied in Table 1? Why both Delta R and sigma_R are needed?

You are right. Your question led to confusion on my side too. Originally this was meant to provide some complicated but didactically nice derivation. Due to your question I had to reconstruct why the numbers in table 1 for DeltaR and sigmaR did not show a perfect factor 2. Obviously numbers of a preliminary analysis using DeltaR and a final analysis using sigmaR both made it into the table. I should have noticed that before. I have removed the DeltaR definition and part of the explanations from the text and the column with the incomplete DeltaR values from the table.

p.10, l.11:
sigma_R defined by Eq. (3) should be called "relative dispersion" (ratio of the standard deviation to the mean) rather than "standard deviation".

I would prefer to stick to "standard deviation from the mean".

p.11, Fig.5:
Here and in the rest of the paper it is not clear what the authors mean by "distance": horizontal distance or distance along the viewing direction. Please, clarify.

Tried to clarify in two places of the retrieval chapter.

p.11, Fig.5 and ll.3-17:
Important: the absolute accuracies (in meters, not %) of distance measurement must be presented instead of relative ones. Measured distance is only an intermediate step towards derivation of cloud shape and geographic location, which are independent of the sensor position. Thus, only the absolute values of uncertainties in derived distance have physical meaning. The same, say, 600 m accuracy of cloud location determination (the only thing that matters) can be 1% if the distance between cloud and sensor is 60 km or 10% if this distance is 6 km, thus, speculation that 1% is better than 10% does not make any sense.

I would like to leave this point open for the public discussion. Use of relative values shows differences more nicely and are more intuitive to understand. Using absolute values would basically only show an increasing stddev with increasing distance. In comparison literature usually both are provided.

p.11, ll.3-5:
The sentence sounds a bit awkward. I suggest:
"At low altitudes most cloud sides can be observed looking upward, while for high altitudes they are observed looking downward."

I used your suggestion. Thank you.

p.12, ll.15-16:
"uncertainty of oxygen-A retrievals is likely to be larger than that of the stereo derivation": How much larger? Comparison is hard since the stereo accuracy is given in meters (200-300 m), while that of A-band measurements is given in Section 2.2.3 in % of the distance between cloud and sensor (see comment to Fig. 5 above).

I have tried to clarify that and to add a few numbers from the earlier publications. They usually provide only a few numbers, and usually absolute values together with typical cloud distance as well as relative numbers.

p.12, l.23:

How relative azimuth is defined in the paper? If here ~180 deg azimuth means that the sun is behind the sensor, does it mean that the azimuth of 0 deg in Table 1 corresponds to the sun in front of the sensor (which would not make much sense)?

Thanks for pointing out this inconsistency. I have corrected the values in table 1.

p.12, l.34:
I suggest to replace "shown as well" by "also shown in Fig. 6".

Done.

p.13, Figs.6(c)&7(a):
Are these distances horizontal or along sensor's line of view to the point?

See above.

p.13, Fig.6(c,d):
It is difficult to see comparison between A-band and stereo distances/heights since the former are shown by color, while the latter - by numbers. I suggest to include (in addition to this figure) several single-scan plots showing 2D cross-section of cloud surface derived from A-band measurements with stereo points plotted over. This could also show how/if stereo points can be used for "calibration" of A-band retrievals on single scan or single cloud scene bases.

You mean a 2D cross section in x-z plane? x being the horizontal direction along line-of-sight, across-track. That would mean to show one x-z cross-section through the oxygen-A derived surface as a line for each stereo value (sometimes two), because the surface changes quickly in y (flight) direction. We are not sure, if another figure improves this comparison much. A third chance to directly compare them is given in Fig 7 where the turquoise points show the values again.

p.14, Fig.7:
Make both plots square (since x and y have the same physical meaning and scale) and set them side by side. Provide means and standard deviations of (y-x).

Will be done.

p.14, Fig.7:
The meaning of points' colors is not explained neither in text nor in caption. Do they correspond to the mentioned 27 cloud scenes?

Yes. Included in text and caption.

p.14, ll.3-4:
Replace "200 x 200 m^2" by "200 m x 200 m".

Done.

p.15, ll.9-10; p.16, ll.12-13:
My impression from looking at Fig. 7(a) is that if different cloud scenes (assuming that they are identified by point colors) are considered separately the fitting offsets in them would significantly vary (and may even have different signs). I do not see convincing justification that all these scenes belong to the same "type" so the same 3.8-km offset is characteristic for all of them.

That there is variation is certainly true and to be expected as not two clouds will display the exact same geometry and 3D effects. On the other hand not each single stereo result is to be trusted 100%. For all larger groups the offset lies between 0 and 6 km. Only a clear minority shows the opposite sign and most of them show large uncertainty bars labeling them as difficult situations. The use of a single offset for the whole campaign data set is certainly a compromise limiting accuracy. On the other hand, we are confident that our overall uncertainty values are correct, as they are nicely confirmed by this comparison (after offset correction). Use of one offset of 3.8 km might of course leave individual cloud scenes with systematic remaining biases of mostly 1-2 km. We mention this in the text now.

p.15, ll.18-19; p.16, ll.17-18:
"In general, such a "calibration" of the method could also be reached using a limited number of synthetic cloud model based simulations." Here the authors try to downplay the complexity of 3D RT simulations contradicting their own words in the first paragraph of Sec. 2.2.2. I do not believe in "universal" 3D-RT LUT based on a few of cloud scenes.

We tried to clarify.

"In general, such a ``calibration`` of the method could also be reached using a small set of selected 3D simulations. A small number a few hundred test results (as in our comparison to stereo points) at a low Monte Carlo accuracy - a high noise level - would be sufficient to characterise a single typical offset value. It could be provided at a small computation effort (several orders of magnitude below the effort to provide non-vertical slopes). "

For explanation: A use of a limited number of cloud resolving model produced scenes can be very effectively probed by a limited number of randomly selected single line-of-sight simulations, say 300 as in our stereo-comparison. Monte Carlo accuracy can be quite low for these 300 as we only try to characterise one number – the offset. Compared to such a set of simulations, even the "cloud wall" LUT simulations of the manuscript (4 SZA x 29 VZA x 17 distances x 10 altitudes = 19720) are still about 500 to 1000 more costly.

p.16, l.9:
Delete "towards the horizontal", change "shortcuts" to "decreases".

Done.

p.16, l.10:
Replace "towards" by "for".

I have replaced "towards" by "approaching".

p.16. l.12:
What is "simulated geometry path"?

Bad wording. Changed to "simulated absorption path".

p.16, ll.
"Eventually a combination of stereo and absorption methods might be a beneficial approach, as advantages of both methods could be combined". This is what needs to be done in the next revision of this paper (not eventually), potentially creating a working retrieval method. Applying a campaign-wide offset to all A-band retrievals can be considered as a first approximation to a

working synergistic approach in which stereo-based corrections should be made at each individual cloud scene.

Basically this is what we already do in this paper!? We use an oxygen-A retrieval to provide cloud distances for all cloud surfaces with uncharacterised 3D cloud effect. We use a very limited number of stereo derived points to characterise the campaign-typical 3D effect. We suggest to use the offset corrected oxygen-A results as we are confident that we understood all limitations after careful analysis. We do not see how the message or validity of results of this manuscript would change, if we would use only half of the stereo points for "calibration" and use the other half for "validation".