We thank the referee for taking the time to read through the manuscript and for the constructive suggestions and criticisms. Below we reproduce the referee's comments (printed in italic) and address them individually:

1. *Methodology*

   There are A LOT of assumptions, retrieval variables and tuning in the proposed retrieval method. Observations are supposed to provide evidence of cloud and drizzle profiles to allow us to explore new features, or to test if current assumptions and parameterizations are appropriate in models. If possible, we should let observations speak themselves, rather than forcing all kind of assumptions in the retrieval process. Although the authors mention that these assumptions are based on some other independent observations, it would be good to keep in mind that these assumptions are based on very limited observations, and may not work everywhere. As the authors may have already realised, adjustment of these assumptions is needed when this algorithm is applied to different cloud regimes. It would be good to know where the assumption fails, and how this failure affects the overall retrieval. Any limitation does not undermine the value of the proposed work/method.

   These assumptions also intrinsically introduce many variables to be retrieved. We need to keep in mind that we only have lidar backscatter, radar reflectivity, and microwave temperature measurements. These are limited observations after all, so we should ask ourselves if these observations really contain sufficient information content to retrieve all the variables proposed in the manuscript. The answer is clearly, a No, and thats why the authors use tuning so often in the manuscript. In the end, it will be quite hard to track/ensure that there is no compensating error in the retrieval process. Could the authors comment on this and perhaps have a way to prevent the compensating errors?

   Given the available data, it is necessary to make assumptions to reduce the number of variables that are needed to profile the cloud and drizzle properties. In making these assumptions, we use past observational and theoretical studies as a guide to retrieve realistic cloud and drizzle profiles.

   The information content is indeed limited, which can lead to compensating errors. In this respect, the retrieved cloud droplet number concentration (N), effective radius (r_e) and the shape parameter are quite vulnerable, since they rely largely on radar reflectivity (Z) alone. Z is proportional to both N and to the 6th power of droplet size, which makes N a lot more sensitive to changes in Z than r_e. This results in N fluctuating rather strongly from profile to profile. As shown in the LES exercise (section 3.1), apriori knowledge of shape parameter can help stabilise N. Finally, a good statistics of N over an extended time period is needed for a more reliable interpretation.

   The algorithm looks a bit unnecessarily complicated to me. For example, I dont quite understand why is needed to go through all the trouble to tune cloud base height.

   The tuning and smoothing are improvements made to the way the cloud model is used in the retrieval to account for observational features or limitations that are not part of the cloud model. Below we summarize why we think that it is a necessary step (a more lengthy description is provided in the redrafted section 2.2.1).

   There are at least 4 parameters to determine when specifying the cloud LWC according to the model in eq. (8) and (9), e.g. the parameters W and H in eq. (9) and the cloud top and cloud base absolute heights. One can estimate the locations of cloud boundaries from lidar and radar observations, but using these estimates as they are does not do justice to the model. The radar-lidar estimates are limited by the range resolution, whereas the model is not. The 'tuning' is therefore applied so the model fitting can account for this spatial limitation when looking for the best-fit solution.

   A further adjustment of cloud LWC model near the cloud base is ad-hoc but justified by what is observed by the lidar (as shown in Fig. 2). The cloud model has a well defined cloud base and shows a considerable increase of extinction coefficient immediately above the cloud base, resulting in a fast and pronounced rise in the lidar backscatter signal. However, what is often observed is a slower increase in the lidar backscatter signal around the cloud base because real cloud boundaries are not sharp (due to e.g. turbulence, changing cloud base heights within measurement temporal resolution) and are not distinct to model precision. We incorporate this aspect in the model via the convolution or smoothing.

As shown in Figure 2, the authors apply an ad-hoc smoothing in order to get reasonable cloud base height that matches with lidar measurements. Why not using the observed cloud base height instead, and then discuss/understand how sensitive the retrieval will be to the accuracy of the observed cloud base height?

We feel that it is important to make sure that the observations are reproduced as well as possible in the retrieval. We have tried before to use the "observed" cloud base height in the retrieval (using the synthetic signals in the LES exercise in section 3.1) and we found that the model often failed to produce a good fit to the lidar backscatter at the cloud base. As a result, the true LWC and other propertes were not well recovered. For this reason, we let the algorithm decide where the model cloud base height should be located, guided by the cloud base height estimate from the lidar signal.

We realized that our description of the retrieval algorithm in the old manuscript caused some confusion and misunderstandings. We rewrote a large part of sections 2.2.1-2.2.3 to try to explain the algorithm better and with clarity. In the revised section 2.2.1, we describe the cloud base height determination differently in the light of reviewer's comments on this subject. Here, we avoid the word 'tuning' altogether since its use seems to have been misleading.

There are also a lot of ad-hoc smoothing bits and thresholds in the proposed method. Rigorous scientific justifications about their choices are needed. For example, why choosing only 1 or 2 radar range gates to classify non-drizzling case. When should we use 1, and when to use 2 gates? Does this really perform better than threshold-based approaches, or they actually agree to a large extent?

The eventual decision in the algorithm to classify a column profile as drizzling or non drizzling stems from the use of a droplet size (r_e) threshold (the scientific justification of using this threshold is given in the introduction and repeated again in section 2.3.2 for clarity). Our approach is thus, threshold-based.

This r_e threshold (13 microns) largely determines the proportion of cloud and drizzle signal. The cloud is deemed to be non drizzling when, after applying the r_e threshold, there is no or very little drizzle reflectivity signal left. We define 'very little' as having drizzle signal at less than 3 range gates (hence 1 or 2 radar range gates). In this case, we decide to write off drizzle because it makes little sense to construct a profile from less than 3 data points, apart from the fact that noise is likely to prevail in such weak/few detections.

The text describing this in the manuscript has been edited to provide more clariy.

Page 5: The justification of constant cloud droplet number concentration (N) with height is a bit misleading, and I feel that the authors are stretching this a bit too far. There is really no sufficient information to infer the vertical profile of N from radar/lidar/microwave measurements. Yes, some could probably use a stronger priori, but the result will not be mainly determined by observations. Saying that a constant N is adopting the homogeneous mixing case is just not quite right.

It is true that radar, lidar and MWR measurements alone do not provide sufficient information to infer the vertical profile of N. In such a case, the simplest approach is to then assume that N is constant with height. We reworded the last part of section 2.1.2 to avoid confusion.

2. *Evaluation*

   Synthetic data set:

   The key point of the manuscript is about cloud/drizzle properties. I am surprised that the authors chose a non-drizzling case in the synthetic data test. Without the presence of drizzle particles, I am less convinced about the performance of the proposed method. I think demonstrating a drizzling case is necessary.

   The retrieval for the drizzling case using synthetic data is now included in the revised manuscript (section 3.2).

   For the current case, I am not sure I understand the results. In section 2, the authors keep emphasizing that they apply a droplet size threshold to separate the cloud and drizzle regime. As a result, they use

13 microns as the separation threshold, meaning that at any altitude, cloud effective radius has to be smaller than 13 microns and drizzle effective radius cannot be less than 13 microns (page 11). If thats the case, how come cloud effective radius in Figure 4 clearly exceeds 13 microns for most altitudes? I also dont understand why the majority of radar reflectivity is greater than 20 dBZ for a non-drizzling cloud (and interestingly, it is opposite for the case from the ACCEPT campaign; see next).

We recognize the reviewer's concern that the magnitude of the cloud properties in the LES exercise is at odds with the droplet size threshold or with the ACCEPT data/retrieval. However, this does not interfere with the purpose of the LES exercise, that is to verify the forward models and the cloud LWC model regardless of the magnitude of N or r_e or Z. This LES scene was set up to test the cloud-only retrieval. In the set up, the cloud droplet number concentration was arbitrarily set to a small number, resulting in large effective radius and radar reflectivity. The absence of drizzle in this synthetic scene means that the droplet size threshold of 13 microns is irrelevant.

Also, the narrow range of cloud droplet number concentration may not be the best case for testing whether the retrieval method is robust.

In the newly added section 3.2, the cloud droplet number concentration is set to a more realistic number (about 6 times larger than in section 3.1). From the retrievals in both sections, we did not find any indications that different values of N could affect the retrieval performance.

The ACCEPT campaign:
Could the authors please modify the range of colour bar of radar reflectivity in Figure 6?

Done. The colour bar range of Z in Fig. A1 was also modified accordingly.

It is unclear if radar reflectivity is much higher than -30 dBZ or not. If not, it is surprising to see such low radar reflectivity corresponds to drizzle effective radius up to 60 microns.

The radar reflectivity is typically not higher than -28 dBZ, except for several profiles during the intense drizzle periods around 3.8hr and 5.6hr UTC where Z can go above -20 dBZ (the maximum is -12 dBZ).

Also, this time series does not include many precipitating profiles. It would be much better to choose another time period that includes a wide range of precipitating conditions.

The profiles in our selected 4-hour period cover all the cloud cases that we consider in this retrieval method (non drizzling and drizzling). The majority of those profiles show radar detection below the cloud base, varying in spatial extent and intensity. Profiles that show radar detection down to 200 m or lower are not retrieved because there is no meaningful lidar information below 200 m (incomplete overlap region). We also do not perform retrieval when precipitation is detected on the ground, due to the compromised accuracy of the microwave brightness temperature measurements in such events.

Could the authors include any independent datasets for evaluations? For example, compare optical depth as shown in Figure 5?

Section 5 presents the evaluations of the retrieval results using independent retrieval methods and independent datasets. The independent datasets include lidar depolarization signals (section 5.1), radar doppler spectra (section 5.2) and radar doppler moments and ceilometer backscatter signals (section 5.3). Independent optical depth retrieval is unfortunately not available, to our knowledge. The closest we can get to this is the extinction coefficient comparison in Fig. 9 (which is now Fig. 10 in the revised manuscript).

3. *Presentation*
Re-organisation of the section 2: I would suggest starting the section with 2.3.2, and making Figure 1 more understandable and stand-alone. The authors need to refer to Figure 1 in a bit more detail to guide readers to understand the overall structure/flow of the retrieval method. It would be nice

to construct Figure 1 into a number of main components, provide an overall flow and linkage of all components in the first paragraph, and then synthesize the details in each component.

We have simplified Fig. 1 and reorganized section 2 based on the reviewer's suggestion. We refer to Fig. 1 when describing the flow of the retrieval at the start of the section 2 and synthesize the details of the main steps in the following subsections.

Some examples that need better connections and wording:
Page 9, Line 1821: The sentence was talking about z_cb and z_peak, and then the equation below uses z_max and z_min. After reading the line below equation(17), it is unclear how the equation links to z_cb and z_peak.

We clarify z_max and z_min in the revised manuscript (page 8).

Also, many methods for determining cloud base height from lidar measurements have been proposed and compared; what has presented in this paragraph is a result of the uncertainty in cloud base height determination. Why not mentioning this to justify what has been done here, instead of presenting them as the actual cloud base and the model cloud base?

We rephrased the text that explains the cloud base height determination (see the redrafted section 2.2.1).

More importantly, what is the implication of the need to find the optimal cloud base height? It is not very good news if retrieval needs such precise determination of cloud base height.

The need to find the optimal cloud base height implies the need to incorporate observational aspects when using the cloud model in a retrieval scheme that simulates observables for comparison with real observations.

Page 9, Line 29: It is unclear what as the maximum number of consecutive range gates around z_cb means. Do you mean, if z_cb is at range gate #10, for example, then the maximum number of consecutive range gates is 10?

No, that is not what we mean. We provide a more quantitative and detailed description of the smoothing in the revised section 2.2.1 (just under eq. 17) to remove the ambiguity in the previous version of the text.

What is the physical justification for this smoothing? Softening certain behaviour to get rid of something does not sound very scientific to me. It would be much more appropriate and convincing if the authors could link this behaviour to some sources of uncertainty/noise for justification.

The physical justification of the smoothing is provided above (page 1 of this document) and also in section 2.2.1. The smoothing (or softening) is not aimed to get rid of anything. It is aimed to reproduce what is observed in the lidar signal and not taken into account in the cloud LWC model.

Page 9, Line 29: Is p_cb the pressure at z_cb? It may be obvious for readers, but all variables should be denoted clearly.

No, p_cb is not the pressure at z_cb. p_cb is a variable in the state vector that is used in the cloud base smoothing. This variable is better explained in the revised section 2.2.1.

There are also quite a few repetitions. Could the authors please read the manuscript carefully and clean things up?

The manuscript has been revised.

4. *Finally, I feel the manuscript could use a bit more positive attitude/tone - we dont need to play down other peoples work to justify our work.*

   We have attempted to be fair and objective when citing others' work. We certainly did not intend to play down the work of others and indeed we are somewhat puzzled by this remark. We can not identify any specific passage where we feel we are being too negative. However, we would be happy to receive specific suggestions from the reviewer as to how we can improve this aspect of the paper. Furthermore, in an attempt to conclude the paper on a high note we have added appropriate text to the Summary (see last paragraph of Page 24).

The manuscript has been revised based on the input from two anonymous reviewers. The notable changes are:

1. The flowchart in Figure 1 has been simplified. In the flowchart, we include references to the sections where more details can be found.

2. Section 2 has been reconstructed and reorganized to provide a clearer and a more coherent description of the retrieval method. No changes were made in the conceptual design or the implementation of the method.

3. Section 3 has been expanded to include the retrieval of drizzling clouds using synthetic data. Along with this, an additional figure was produced (shown as Fig. 6 in the revised manuscript).

Changes or new additions in the text are marked in red in the revised manuscript.