Author Responses:

Anonymous Referee #1

Received and published: 4 October 2017

General remarks:

The manuscripts provides a detailed analyses of two well established cloud retrieval methods using passive satellite measurements of solar radiation. Bispectral and polarimetric retrieval of cloud optical thickness, droplet effective radius, and effective variance are compared for cases of liquid maritime clouds. The study is not based on measurements of real clouds. To analyze the limits of the physics behind the retrieval approaches, cloud fields provided by LES model runs are used to generate synthetic radiation measurements by radiative transfer simulation. This approach has the advantage of being independent on different uncertainties introduced by real observations and that the retrieved quantities can be compared to the truth given by the LES model. It is concluded that the bispectral retrieval shows a higher uncertainty for the retrieval of cloud droplet size compared to polarimetric retrieval while cloud optical thickness agrees between both approaches.

The results presented by this study are of high value for current and future satellite remote sensing. Retrieval uncertainties, which originate from the general limitations of the retrieval algorithms are clearly quantified and may help to improve the interpretation of satellite cloud products. In this regard, the manuscript provides an important contribution to current and future research and is worth to be published.

However, in my opinion the manuscript lacks of some major issues which have to be reassessed in detail before publishing the manuscript. By neglecting measurement uncertainties, the retrieval comparison might be only of academic value because it does not reflect the real uncertainties of both retrieval approaches when real satellite observations are considered. Furthermore, new developments of the bispectral retrieval are not considered in the study and limit the conclusions for future satellite employments. The bias of the bispectral retrieval is surprisingly high considering the ideal setup of the study. A more accurate treatment of the vertical weighting function of the bispectral retrievals needs to be applied in order to guaranty the comparability with the LES and the polarimetric retrieval. Below, I compiled a list of comments which have to be considered in a revised version of the paper. There might be some contradictory statements resulting from my misinterpretation of the text when first reading. I am sure the authors will know how to weight in such cases and how to improve the text to avoid misinterpretations by other readers.

Major comments:

1. **Neglecting measurement uncertainties**
   I understand the approach of the authors to use synthetic measurements generated from LES cloud fields and radiative transfer simulations. This approach leaves only a limited number of causes which can explain the difference between both retrieval approaches, such as the complexity of the cloud representation in the radiative transfer model (vertical profile). In this regard, the study provides good insight into the physics of the retrieval approaches. This is worth to be published but might be only of academic value. However, the conclusions on the performance of the retrieval approaches might change when measurement errors are considered. Uncertainties of the spectral radiance measured by the satellite sensors, e.g. radiometric calibration, might propagate differently in both retrieval approaches. An uncertainty of spectral radiance might have larger consequences on retrieved cloud properties compared to uncertainties in the polarimetric measurements. To judge, which retrieval algorithm provides the more accurate cloud properties when applied to real satellite measurements such a propagation of the measurement uncertainty has to be considered and analyzed. This should not replace the current results of the study. Please keep these results. I rather suggest to add an additional exercise with focus on the propagation of measurement uncertainties. On basis of the available data set, this should be easy to realize. The simulated radiances which are the exact synthetic measurement are available. By generating synthetic measurements including a measurement uncertainty and propagating through the retrieval algorithm should already give an estimate of these retrieval uncertainties. Your motivation to use a LES cloud field and IPA simulations would still hold for such a study, as 3D-radiative effect, etc. still can be ruled out. Only the propagation of pure sensor uncertainties will be analyzed.

   a. I can add a section and a figure focusing on the impact of measurement uncertainty, but the behavior of uncertainty for these two retrievals is highly algorithm and instrument dependent. It could narrow the scope of applicability of the paper – at the cost of making it more applicable to a single pair of instruments.

2. **Vertical weighting functions**
   As discussed by the authors, the vertical weighting function is essential to compare retrieved cloud properties with the LES model clouds. Therefore, I am wondering why a relative crude assumption for the weighting function of the bispectral retrieval is assumed. The two-way transmittance function is valid for single-scattering only which holds for the polarimetric retrieval where single scattering features are extracted from the measurements. But the bispectral retrieval certainly are effected by multiple scattering. Platnick (2000) clearly shows that the vertical weighting functions significantly extend into the lower cloud layers. Even for 3.7 µm cloud layers at optical thickness larger than $\tau > 2$ contribute to the weighting function while the 2WT weighting already becomes zero for $\tau > 2$.

   First, I was surprised by the relative large differences between bispectral retrieval and LES-truth because the setup of the study was chosen well and should not allow large differences. But the treatment of the vertical weighting functions may explain these differences. Considering the idealized setup using the LES clouds and the independent pixel approximation to generate the synthetic measurements, I do not see many sources of error than the vertical distribution of cloud particles and how these are represented in the radiative transfer model. I assume, that the calculation of the synthetic measurements and the calculation of the LUTs use the same radiative transfer code. For the synthetic measurements, the vertical cloud profile is considered, but not for the LUTs of the retrieval. So the radiative transfer code itself is no issue.

   The inaccurate treatment of the vertical weighting fits also to the results shown in Figure 3a. The slight shift of the bispectral retrieval to smaller particle sizes compared to the 2WT weighing might result from different vertical weighting function. While the 2WT weighting only considered the larger particles at cloud top, the bi-spectral retrieval is also influenced by smaller particles at lower cloud levels. This could already lead to C4 the observed differences. Therefore, I suggest to use a more realistic vertical weighting function for comparing the bispectral retrieval with the LES model. The weighting function considering multiple scattering can be easily calculated by the method presented by Platnick (2000). As an approximation the weighting functions can be calculated assuming vertically homogeneous clouds as for the retrieval LUTs. With this assumption they can be easily extracted from the LUTs as the slope of reflectance with increasing optical thickness. So all required simulations should be available.

   a. I am not sure that a vertical weighting based on homogeneous vertical profile assumptions would be appropriate for this analysis. The LES clouds are not vertically homogeneous and there can be significant extinction cross-section variability within cloud vertical profiles. The intent of the applied vertical weighting techniques is to account for that vertical inhomogeneity directly.

b. The authors agree that the single scattering (2WT) vertical weighting may not sufficiently describe the behavior of scattering in the 2.13 µm spectral band. One of the reasons for implementing a single vertical weighting definition throughout this paper was to ensure that we were not comparing retrievals to a "moving target." Additionally, from our previous work in Miller et al. (2016) we knew that r_e(2WT) matched the r_e(3.75 µm) spectral retrieval reasonably well.

c. Despite the large variability in the comparison of vertical weighting and bispectral retrievals it is important to note that the mean biases are quite small. The logarithmic histogram in the joint PDF can possibly be emphasizing features that are occurring for clouds that have significant vertically inhomogeneous cloud tops.

d. To test the impact of using a more accurate vertical weighting function we implemented the approach described in equation 4 of Zhang et al. 2017, which includes an additional factor that accounts for multiple scattering contributions:

$$W(\tau) = c\tau^b \exp\left[-\tau\left(\frac{1}{\mu} + \frac{1}{\mu_0}\right)\right].$$

Where the τ^b factor is introduced to account for multiple scattering, and the rest is the same as previously defined in the paper. For b=0 we get back the original 2WT vertical weighting used previously. Below is a figure displaying the scene-averaged vertical weighting functions from the DYCOMS-II case, note that similar results (throughout this discussion) were found in other LES cases. Each of the curves corresponds to a different value of b and each is also color coded to indicate the resulting average vertically weighted cloud microphysical property r_e(VW) and v_e(VW). Increasing the value of b causes the weighting function to extend deeper into the cloud, changing the mean value of r_e(VW) and v_e(VW). The change in the mean value of r_e(vw) for these different weightings does not exceed 0.5 µm and the mean value of v_e(VW) changes less than 0.05. The small variability in r_e has a lot to do with the actual profile of $r_e(\tau)$, which has little variability near cloud top. In contrast the large variability in v_e is related to the rapid change in $v_e(\tau)$ just below cloud top. It should be noted that the b=0 vertical weighting has a similar r_e(VW) result as the b=9 case, which offers an explanation as to why r_e(2WT) just happened to be a decent point of comparison.

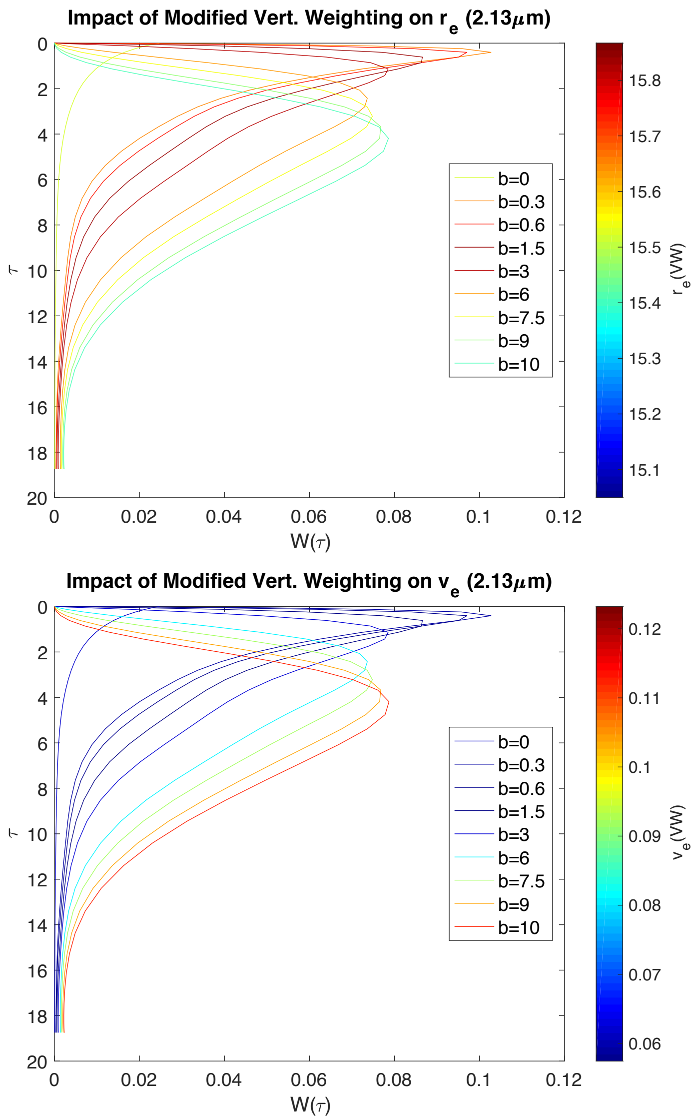

Figure 1: Scene average vertical weighting functions for the DYCOMS-II case with varying values of the parameter *b*. Lines are colored by the corresponding vertically weighted properties (r_e(VW,2.13) or v_e(VW,2.13))

Using this parameterization, we can tune the "quality of vertical weighting" by comparing the shape of the distribution of the bias between r_e(2.13) or r_e (3.75) with respect to their respective vertical weightings r_e(VW,2.13) and r_e(VW,3.75). By searching this histogram with different values of *b* we can find a parameterization that minimizes the mean bias, as well as the variability of the distribution – indicating a more appropriate vertical weighting definition for the bispectral retrieval in question. Several animated gif's of the behavior of this histogram as a function of b can be found here:

r_e(2.13)-r_e(VW,2.13) Histogram varying *b*
r_e(2.13)-r_e(VW,2.13) Histogram varying *b* (FOR τ>5 ONLY!)
r_e(3.75)-r_e(VW,3.75) histogram varying *b*

r_e(3.75,coupled to v_e(pol))-r_e(VW,3.75) Histogram varying b

Where for the 3.75μm spectral band we found that a b coefficient of 0.3 minimized mean bias and variability (when coupled with v_e information), whereas for 2.13μm a value of b=10 was found to be optimal. It is also important to note that the appropriate vertical weighting also depends on total optical thickness. As a consequence of the deep penetration of the 2.13 vertical weighting function some extremely high biased values of r_e(VW) exist for $\tau_{tot}<5$. This threshold was selected because $\tau=5$ is roughly the location of the peak of this vertical weighting function. For the ~5% of profiles with $\tau_{tot}<5$ we simply define b=0 in order to avoid applying an inaccurate multiple scattering vertical weighting.

The regression histograms of these new results are shown in Figure 2. The primary finding of this comparison is that the new flexible vertical weighting function produces results for the 2.13 μm band that have a far tighter regression. There is also a new bump in the comparison for 2.13 with $\tau_{tot}>3$ only, but this is associated with the population of high-biased r_e(VW) that remain below $\tau_{tot}<5$.

I also found a possible mistake in my plotting routines, I was only plotting nadir viewing r_e(2WT) and v_e(2WT) values for comparisons to the polarimetric retrieval. I have modified these results to include the other viewing geometries and those are shown in Figure 3.

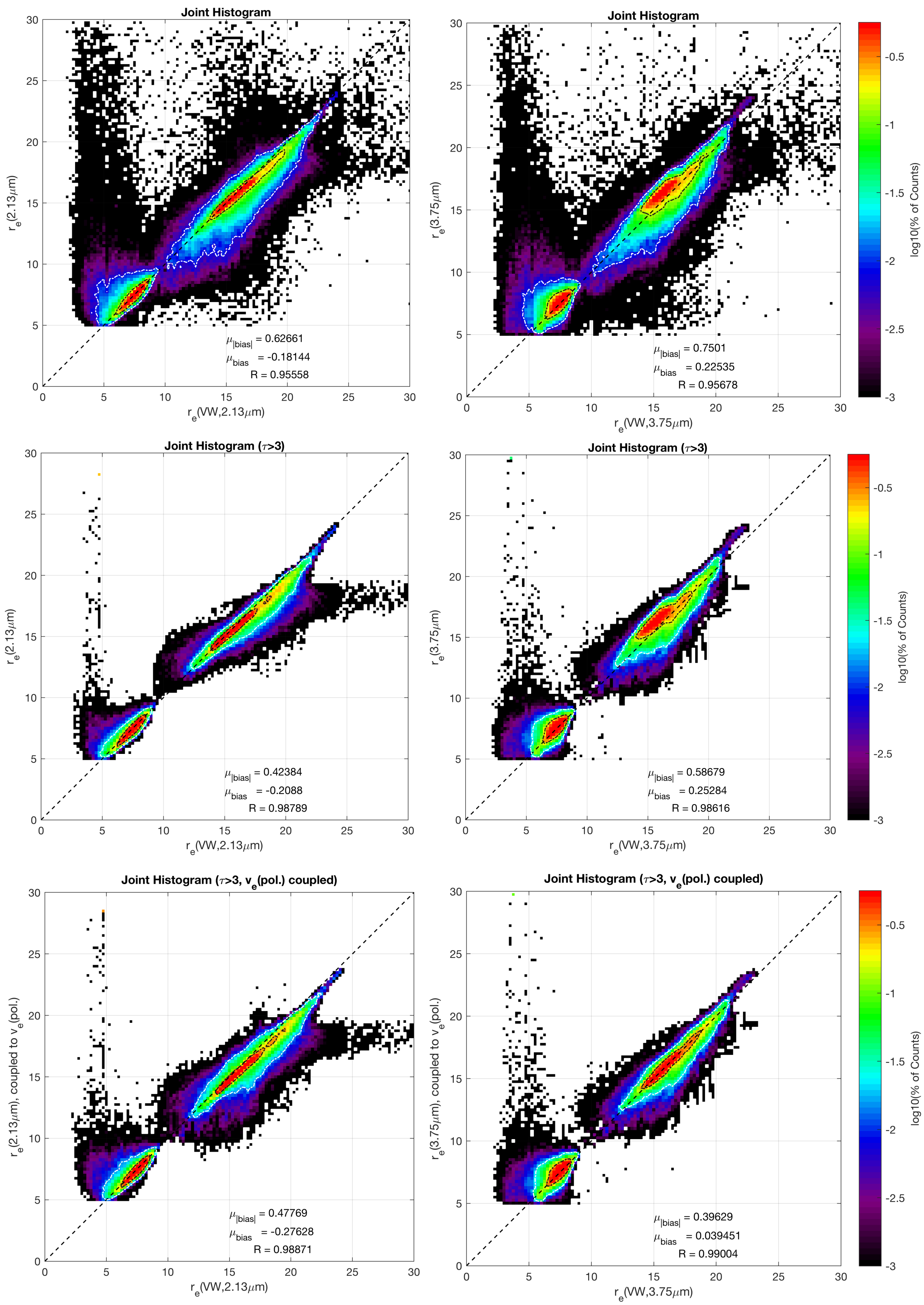

Figure 2: Refer to Figure 3 of the original manuscript for more figure information. This new version compares to a more flexible vertical weighting definition (denoted VW) than the single-scattering 2WT scheme.

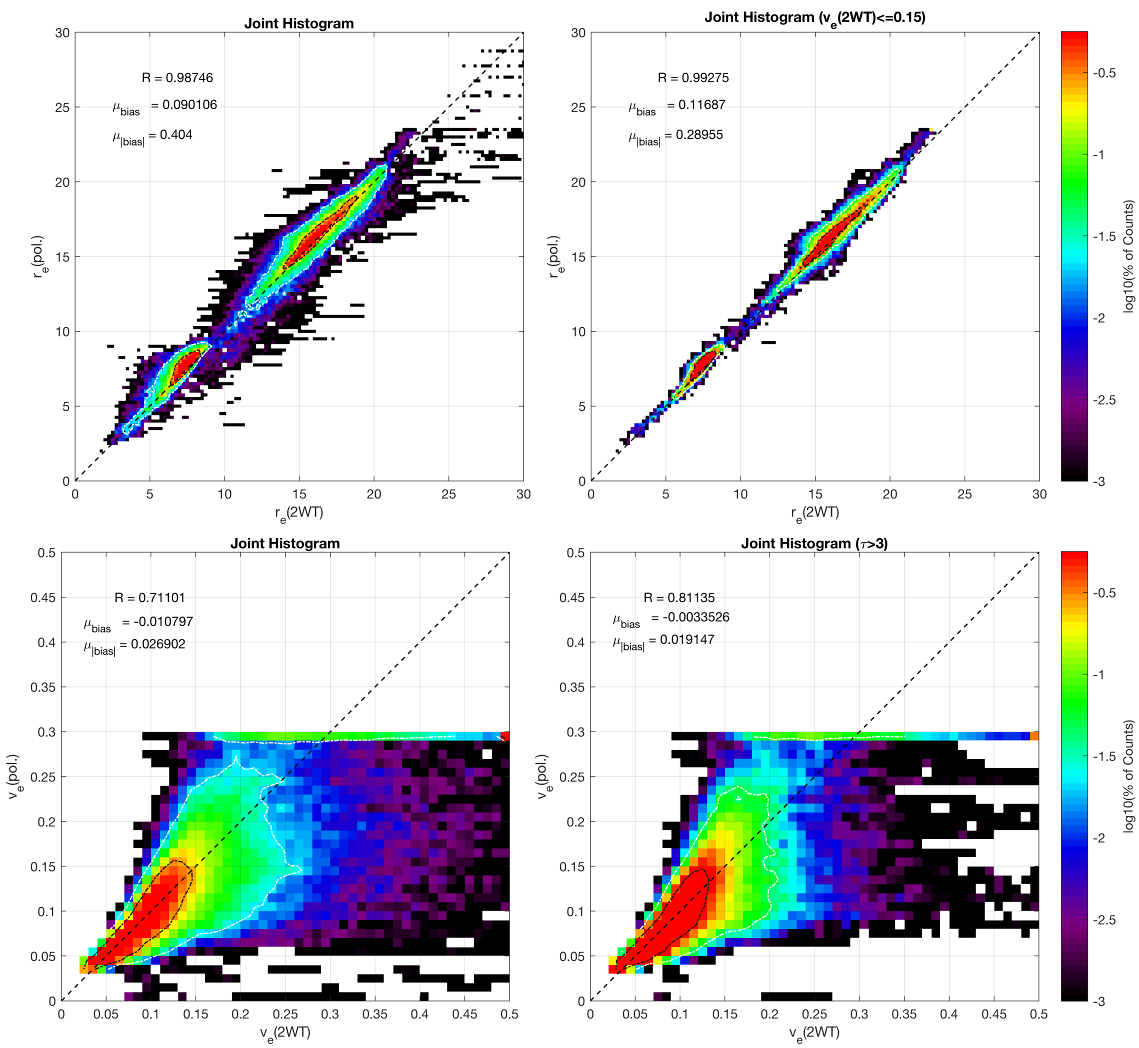

Figure 3: Refer to Figure 4 of the original manuscript for more figure information. These polarimetric retrieval comparison plots have been updated to include data from all the same viewing geometries used in the bispectral retrieval.

Zhang, Z., X. Dong, B. Xi, H. Song, P. L. Ma, S. J. Ghan, S. Platnick, and P. Minnis (2017), Intercomparisons of marine boundary layer cloud properties from the ARM CAP-MBL campaign and two MODIS cloud products, *J. Geophys. Res.*, *122*(4), 2351–2365, doi:10.1002/2016JD025763.

3. Radiance Ratio Retrieval

The manuscript motivates the study by future satellite missions providing multi-angular polarimetric observations. However, also the classic bispectral observation will profit from continuous improvement of the retrieval algorithms. In recent studies, the radiance ratio retrieval approach has been proposed (actually by one of the co-authors) to reduce some limitations of the bispectral retrieval (Werner el al. 2013, Ehrlich et al. 2017, LeBlanc et al. 2015, *Brückner et al. 2014). Using ratios of spectral radiance instead of absolute radiance improves the orthogonality of the LUTs and the impact of measurement uncertainties. Therefore, some limitations discussed for the bispectral retrieval might be improved. E.g. LUTs of radiance ratios are more spread for small cloud optical thickness, the PPH-bias is likely reduced having more orthogonal LUTs (similar to the differences between 2.1 µm and 3.7 µm). This improved retrieval approach should be considered in the manuscript as a third retrieval approach if it aims to be relevant for future satellite observations.*

Werner, F., Siebert, H., Pilewskie, P., Schmeissner, T., Shaw, R. A., and Wendisch, M.: New airborne retrieval approach for trade wind cumulus properties under overlying cirrus, J. Geophys. Res.-Atmos., 118, 3634–3649, https://doi.org/10.1002/jgrd.50334, 2013.

Ehrlich, A., Bierwirth, E., Istomina, L., and Wendisch, M.: Combined retrieval of Arctic C5 liquid water cloud and surface snow properties using airborne spectral solar remote sensing, Atmos. Meas. Tech., 10, 3215-3230, https://doi.org/10.5194/amt-10-3215- 2017, 2017.

LeBlanc, S. E., Pilewskie, P., Schmidt, K. S., and Coddington, O.: A spectral method for discriminating thermodynamic phase and retrieving cloud optical thickness and effective radius using transmitted solar radiance spectra, Atmos. Meas. Tech., 8, 1361– 1383, https://doi.org/10.5194/amt-8-1361-2015, 2015.

Brückner, M., Pospichal, B., Macke, A., and Wendisch, M.: A new multispectral cloud retrieval method for ship-based solar transmissivity measurements, J. Geophys. Res., 119, 11338–11354, https://doi.org/10.1002/2014JD021775, 2014.

a. We did a sensitivity study comparing the standard MODIS approach to a ratio retrieval built using the available bands in the MODIS LUT reflectance library. We found that for plane-parallel homogeneous clouds the ratio retrieval resulted in a reduction of uncertainty << 1%. Both retrieval approaches have significant issues for low optical thicknesses. The reflectance ratio method discussed in Werner et al. 2013 makes use of hyperspectral observations with 2 nm resolution. Our current thinking is that this approach becomes less fruitful with wider spectral bands. The MODIS spectral bands are significantly wider than this (on the order of 100nm), as are many imaging radiometers as a result of instrument design trade offs between spectral resolution and spatial coverage/resolution. Because the LUT used in this study is based on MODIS it is unfortunately not straightforward to reimplement a ratio retrieval for a hypothetical hyperspectral instrument using the same datasets.

b. One of the primary reasons for applying the ratio retrieval is to minimize the PPH bias associated with the impact of the curvature of the NJK LUT for broken/inhomogeneous clouds (at coarse resolutions). For the standard Nakajima-King approach, our group has been working on a technique that can provide a simple correction or estimate of this bias discussed in Zhang et al. (2016) and others. As a consequence, we believe that the PPH bias will be less severe with the development of this technique -- removing one of the motives for implementing a ratio retrieval.

c. We will introduce some further discussion of the ratio retrieval approach, and why it isn't applicable in this study, but may be interesting to explore for hyperspectral instruments that stretch far enough into the SWIR.

Zhang, Z., F. Werner, H. M. Cho, and G. Wind (2016), A framework based on 2-D Taylor expansion for quantifying the impacts of subpixel reflectance variance and covariance on cloud optical thickness and effective radius retrievals based on the bispectral method, *Journal of Geophysical Research: Atmospheres*, doi:10.1063/1.4975502.

Werner, F., Siebert, H., Pilewskie, P., Schmeissner, T., Shaw, R. A., and Wendisch, M.: New airborne retrieval approach for trade wind cumulus properties under overlying cirrus, J. Geophys. Res.-Atmos., 118, 3634–3649, https://doi.org/10.1002/jgrd.50334, 2013.

4. **Polarimetric Retrieval**
How meaningful are the results of the study on effects of the horizontal resolution for the polarimetric retrieval? In the motivation it was mentioned, that for POLDER a footprint of 150 km has to be used to obtain measurements of the cloudbow? This is far from the scales analyzed here with the LES clouds. Is the spatial resolution of future spaceborne polarization sensors comparable to the scales analyzed in this study? The results presented in the manuscript suggest, that in the scales analyzed here, polarimetric measurements are not strongly effected by cloud inhomogeneities. Can this conclusion also be transferred to larger spatial scales? These issues should be discussed somehow in the manuscript.

a. You are correct to point out a "vastness of scale" problem between LES (50 m) and the POLDER satellite retrievals (~50-150 km depending on what paper you look at). The intent of this work is not to compare to the POLDER instrument, but to the newer actively developing spaceborne polarimeters that are expected to launch in the next decade. Perhaps we can make that argument more clearly in the introduction. The POLDER retrieval and its resolution are highlighted for historical context on spaceborne polarimetric retrievals because it's still the only spaceborne polarimeter that has looked at cloud microphysics.

b. The latest polarimetric instruments intended for spaceborne applications such as HARP (see link below) are currently aiming for a retrieval spatial resolution of ~2.5 km for cloud retrievals. That means that the LES retrievals shown here at up to 800 m resolutions are within an order of magnitude of the expected results. At 800 m we have 64 pixels of observations, and to avoid the poor statistics for coarse resolutions we stopped there. It should also be noted that the airborne versions of the polarimeters currently being developed often make observations at nadir resolutions within the ranges discussed here.
https://userpages.umbc.edu/~martins/laco/harp.htm

Minor comments

1. *P1 L1: Title: The study is limited to three very specific cases of liquid low level cloud over the ocean (trade wind cumulus, stratocumulus). At least "liquid clouds" has to be added in the title. "marine" or similar indicating, that only clouds over water have been analyzed should also be considered. Retrieval of ice clouds will certainly differ from the study presented here. Also results for clouds over land can differ due to surface albedo and the different cloud dynamics over land (vertical profile).*

    a. Response: This is a good point. The title will be amended as follows to indicate the focus on marine boundary layer clouds:
    "Comparisons of bispectral and polarimetric retrievals of marine boundary layer cloud microphysics: Case studies using a LES-satellite retrieval simulator"

2. *P5 L10: The reflectance at SWIR and VNIR bands both depend on optical thickness and effective radius. It is simply wrong to indicate that the sensitivities are decoupled. The lookup table shown in the manuscript clearly reveal the non-orthogonality especially for small optical thickness. This coupling has different implication on the bispectral retrieval (PPH-bias) which partly are already used to discus the retrieval biases.*

    a. This was not the intention, I will parse that sentence more carefully. Would the following be more appropriate?

"The VNIR band, dominated by multiple scattering, and the SWIR band, where liquid water droplets are moderately absorptive, contain two independent pieces of information that can be used to remotely infer τ and r_e."

3. *<u>P8 L12:</u> The polarized phase function and the modeled polarized reflectance are two different quantities as far as I understood. How these can be fitted to each other? The degree of linear polarization calculated from polarized reflectance would be comparable to P12.*

4. *<u>P9 L27: Eq. 4:</u> Wouldn't it be better to use/write the size of the coarser resolution pixel into brackets of the mean value. Instead R(0.865 µm, 50 m) better R(0.865 µm, 800 m)? The mean value is calculated for the coarse resolution pixel and independent on the fine resolution of 50 m.*
    a. No, the mean value at 800m is calculated using the 50m data. The values inside the brackets are simply the highest resolution data I have. For example, an 800m pixel has 16x16 pixels at 50m within it that are used in this calculation to get the mean and standard deviation. Perhaps I can come up with a clearer syntax…

5. *<u>P10 L11:</u> Is the comparison only done for a specific solar zenith angle or are all simulations mixed? In Sect. 4.2, Fig. 5 it was explicitly mentioned that all cases and geometries are included. Should be done here as well. P11 L3: Footnote: Why this was written as footnote? The explanation given in the footnote should be presented directly in the main text because it is needed to understand the systematic bias. Putting such parts into a footnote only disturbs the flow of reading.*
    a. All geometries and simulations are mixed, I will make an effort to update that. And the footnote will be worked in to flow with the test.

6. *<u>P14 L5: Figure 8:</u> This comparison has to be done with respect to the LES-truth (see also comment to Figure 5). Only then you can judge which retrieval has a bias and which not. Comparing both retrieval to each other merges effects and does not tell which retrieval is closer to reality. In P14 L9 the differences between bispectral and polarimetric retrieval are rated by assuming the polarimetric retrieval to be the truth. This should be avoided as also the polarimetric retrieval may have caused these differences. You should always refer to the truth solution which is given by the LES cloud fields.*
    a. Refer to the response to the second reviewer regarding figure 8.

7. *<u>P14 L10: typo:</u> "less" and "lower"*

8. *<u>Figure 1:</u> Panel a): Something is wrong because the color codes do not fit to spectral bands! Likely the labeling of x-y axis is switched.*
    a. Indeed, this was an accidental figure editing error that got fixed already but didn't make it into the submitted copy.

9. *Figure 2: Indicate horizontal scale!*
    a. I can include axes labels for these, but they will get even smaller… Would it be sufficient to indicate resolution and scale in the figure caption?

10. *Figure 3: Typo in caption: "or" should be "of"*

11. *Figure 5: I do not see a need for these plots. Comparing both approaches separately to the LES-truth already tells where the uncertainties of the individual approaches are. Comparing both to each other makes interpretation only very difficult but does not give any new conclusions. Both retrieval have to be compared to the LES-truth. The comparison in figure 5 also results in some incorrect conclusions (at least when these are only followed from Fig. 5 alone). The polarimetric retrieval has been found to be better compared to the bispectral retrieval. But this conclusion can not come from Fig. 5 because Fig 5 does not compare to the truth values. Therefore, I suggest to remove Fig. 5 or exchange by similar comparisons with the LES-truth. Also the corresponding discussion (P 12 L 20-30) should use the LES-truths as the reference.*
    a. These plots are necessary because, as indicated in the introduction, this is the sort of plot used in any instrument intercomparison. Put another way, for observational data sets, there is no such thing as an LES truth. Both retrieval approaches have already been compared to the LES truth in Figure 3 and 4. I guess I don't understand how to better address this concern.

12. *Figure 6: Some data does not fit into the LUT. Is this necessary? The range of optical thickness can be extended in your simulations? You should be able to calculate the maximum optical thickness from the LES field in advance. Or is there any other reason why these data does not fit?*
    a. The LUT has a maximum optical thickness below the maximum optical thickness of the LES. Unfortunately, I did not have this version of the LES when we first made the LUT several years ago and so here we are today. At this point it would be computationally expensive to extend the LUT for all geometries and combinations of r_e, v_e, τ – and it wouldn't alter any conclusions.

13. *Figure 6: What is the range of optical thickness? Can be labeled similar to the particle size.*
    a. The range of the optical thickness was stated in the background section (τ=[0.1:100] with 101 logarithmically spaced grid points), but a few labels could be added to this figure.

14. *Figure 7: This figure is also not needed. Both results have already been compared to the LES-truth. Figure 8: Very hard do distinguish the color code*

and circle size. Especially the size of the circles is not visible in the center of the data cloud. Only outliers are visible

a. As stated previously, the comparison figures are part of the primary objective of the paper. Folks comparing these two techniques to one another once they are in orbit don't have the advantage of comparison to LES, so this type of plot is useful in their context. This figure is intended to emphasize that the two retrievals of τ perform approximately the same when they are compared to one another.
b. Regarding figure 8 we have made substantial changes to the way we are attempting to visually display this conclusion. Refer to our response to the second reviewer for more information.