Authors' Reply

The authors would like to thank the referees for their constructive feedback, that helped in improving the manuscript. In the following, all revision points are addressed and the resulting text edits are included in the following way:
The comments are repeated and the responses are given below. Changes made in the manuscript are indicated in blue. Figure numbers with "R" correspond to figures in this reply not included in the manuscript.

Reply to Anonymous Referee #2

Comment: I think a bit more discussion of the science possible with the integration of the radar and Lidar datasets, and of the radar data itself would enhance the scientific impact of the manuscript, but these constitute minor comments, incorporated into those provided below.
Response: We added an outlook to combine the results of the present study with radar and lidar datasets to study the presence and condensate loads of different shallow trade wind cumulus types. We added to Sec. 7: "With respect to trade wind cumuli, the products of the present study in combination with cloud boundary estimations from the radar and backscatter lidar will be used to evaluate the condensate loads of different shallow trade wind cumulus types in large eddy simulations. For example, radar and lidar both detect shallow convection or shallow outflow anvils as depicted in Fig. 10. In addition, the lidar also allows detecting boundary layer driven clouds, which have tops around 1 km and are below the radar sensitivity."

Comment: Also, while I am not sure of the Copernicus standards, I would recommend that DOIs be generated for the datasets and included within the manuscript.
Response: The DOI assignment was in preparation and is completed now. DOIs are added to the references.

Comment: p.2 line 14: also mention the clear-sky contribution to the field of view (it is mentioned later but the sentence suggest precip is the major error source).
Response: We added two sentences to that paragraph to address this point. "Furthermore, the observed LWP per se is an average over the sensors field of view, which is affected by cloud and rain inhomogeneity, and clear sky contribution. Therefore, the spatial resolution is a key information to interpret LWP statistics."

Comment: p. 2 Line 27: what do Greenwald et al and other conclude for the tropical Atlantic region you are interested in?
Response: It is difficult to give any quantitative estimate from the coarse figures provided by Greenwald et al. (2018) for the North Atlantic tropical region. However, our study region in the tropics behaves close to the global average conditions. Nevertheless, from Elsaesser et al. (2017) we added two values. "Elsaesser et al. (2017) additionally estimate the contribution RWP to the total LWP by a simple parametrization and recommend to only use those values with a ratio RWP:LWP of less than 0.2. The average MAC RWP:LWP ratio in our area of interest is 0.23 and 0.30 in December 2013 and August 2016, respectively. Therefore, a more detailed assessment of the rain cloud partitioning is important to better interpret satellite measurements in our study area."

Comment: P. 3. some where I think the adiabatic constraint on LWP is worth mentioning. Is it possible to construct an adiabatic estimate from the Lidar cloud top height and dropsonde RH-derived cloud base do you think? This is an earnest question - I am not sure how well this would work. But it would provide an additional constraint on the retrieval that might be more physical than the imposed 1000 g/m^2 (and its relaxation), and could also provide some additional physical insights. For example, in clean marine stratocumulus regions, the adiabatic constraint on LWP seems to hold well until about 200 g/m^2, at which point precipitation begins to deplete LWP (Zuidema et all, 2005, fig. 8 and 9). I think during RICO the adiabaticity deviated more quickly from the theoretical maximum because of mixing with environment air (Rauber et al., 2007). Related to this I do not see any discussion on the radar or cloud top height at which precipitation becomes discernible later on in the manuscript - perhaps I missed it.
Response: Thanks for these interesting thoughts. However, two aspects limit the applicability of the adiabatic theory in our

case. First, the adiabatic assumption requires, that the cloud develops through vertical transport, i.e. is buoyancy driven, and without horizontal exchange. This is more realistic in stratocumulus situations as addressed by Zuidema et al. (2005) than in the trade wind cumuli cases as in our study. Quite often we see clouds in the radar, that are not buoyancy driven which is also the case in the example of Fig. 10. The radar echo between 17:40:00 and 17:41:40 looks more like a shallow outflow of the nearby precipitating core. This shallow outflow anvil is not buoyancy driven, as its radar echo shows no link to the lifting condensation level (lcl), which is roughly at 700 m. Second, the vertical cloud extend Δz is very important as the adiabatic water content is proportional to Δz^2, but the exact estimation of Δz is difficult. The estimation of the cloud base hight using the lcl derived from near surface dropsonde data has an accuracy of 50 to 100 m. The cloud top height estimation of shallow cumulus adds additional uncertainty as the lidar sees the cloud top typically 250 m higher in altitude than the radar, in cases in which the radar sees a cloud at all. Thus we think, that a comparison to a somehow estimated adiabatic LWP raises more methodological questions than it would help to constrain the LWP.

Comment: P. 3 lines 11-19: what is approximately the spatial footprint of the HAMP instrumentation? It would be nice to see this number in relation to the satellite spatial footprints. On p. 6 you mention that the different footprints and sensitivities of the instruments are covered in Stevens et al 2019, but a brief summary here would be useful.
Response: The reference to Schnitt et al. (2017) reads now: "Their study uses the 1 km resolution HAMP data to show the sub-footprint variability of spaceborne CLWP estimation of about 30 km resolution. Further they illustrate how MODIS products at 1 km resolution likely underestimate CLWP of thick clouds due to MODIS' sensitivity towards the upper part of the cloud."

Comment: P. 7 line 23: how is scattering off of the ocean surface dealt with?
Response: We understand the "scattering off of the ocean" as surface reflection. Surface reflectivity is also calculated using FASTEM5. The related sentence in Sec. 2.3 reads now: "The emissivity and reflectivity of the sea ocean surface is calculated by the FAST microwave Emissivity Model version 5 (FASTEM5; Liu et al. (2011)), which is a modification of the Fresnel coefficients including corrections for ocean surface roughness and foam building as a function of wind speed."

Comment: p.9 line 24: where is the ocean emissivity represented? It would be nice to see a bit more discussion of the ocean surface microwave radiation characteristics in general. A figure of the emission/scattering as a function of SST and wind speed, for the 2 frequencies would be nice, for example. How much does error in the surface characterization contribute to the overall error?
Response: Emission by the surface is only implicitly included in the retrieval. When generating the database, ocean emissivity is calculated by FASTEM5 using sea surface temperature (sst) and 10 m wind speed as input. Based on a comment by referee #1 we investigated the uncertainty due to sst and 10 m wind speed. (See Fig. R1 and R2 and discussion in our answer to referee #1.)

Comment: p.13 line 4; Is there any cloud fraction within a model column? At a grid spacing of 0.5 degree, clouds will not necessarily fill the full grid box.
Response: This is probably a misunderstanding. The ICON simulations were run at 1.25 km resolution. Afterwards, the data used for the retrieval database was coarse grained to 0.5° as explained in the first paragraph of Sec. 2.3.

Comment: p. 17 : it would be nice to see the retrieved LWP/RWP as a function of the vertically integrated reflectivity from both campaigns as part of fig. 11. They should look the same, if not, that may tell you something about the cloud droplet number concentration variation between the two seasons.
Response: The RWP retrieval already includes the vertically integrated reflectivity Z_{int}. Therefore, the empirical relation between RWP and Z_{int} would be established from two dependent variables. Each of our retrievals was trained for each campaign individually by using simulations for the respective period. Thus, differences in the RWP-Z_{int} relation would also represent the different training datasets.
Technically, we have to exclude scenes with no radar echo above noise level because $Z_{\text{int}} = 0$ can not be represented on a decibel scale from such comparison. A logarithmic scale is required for displaying in analogy to dBZ. This means that such a figure excludes clouds, that were to thin to be detectable by the radar but were detected by the lidar. Nevertheless, we prepared Fig. R1 showing the relation between LWP and Z_{int} for scenes, where $10\log_{10}(Z_{\text{int}}) > -30$. The scatter plot shows, that there

are less scenes with LWP > $300\,\mathrm{g\,m^{-2}}$ during NARVAL1 than NARVAL2 as it can also be seen on Fig. 11. For NARVAL1, there is a pronounced maximum of combinations for Z_{int} from 15 dB at $100\,\mathrm{g\,m^{-2}}$ increasing to 40 dB at LWP > $400\,\mathrm{g\,m^{-2}}$. A similar relation can be seen also for NARVAL2. In addition there is an second mode for LWP < $200\,\mathrm{g\,m^{-2}}$ with Z_{int} being about 0 dB. Scenes with lower Z_{int} most likely consist of smaller droplets for the same LWP as $Z_{\mathrm{int}} \propto D^6\,\Delta z$ and LWP $\propto D^3\,\Delta z$. So, probably clouds with smaller droplets were slightly more prominent during NARVAL1 than NARVAL2.
Figure R1 is an interesting starting point for a microphysical study. However, to present this topic in an appropriate manner more work has to be done, which will be included on a follow-up study.

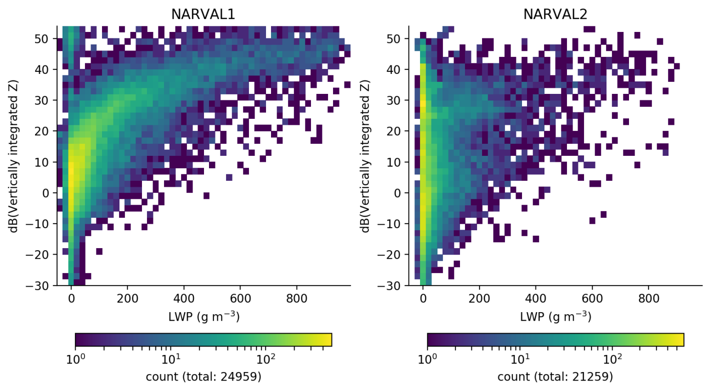

Figure R1. Decibel of vertically integrated reflectivity ($10\log_{10}(Z_{\mathrm{int}})$) versus LWP during (left) NARVAL1 and (right) NARVAL2.

Comment: How does RWP spatial heterogeneity affect the retrieval do you think?
Response: The spatial heterogeneity of rain affects the airborne HAMP measurements less than microwave satellites as the HAMPs spatial resolution is at least an order of magnitude better than satellite resolution. To illustrate the scale on that HAMP resolves precipitation, we added a km scale to Fig. 10. Also we added to the fourth paragraph of Sec. 5: "The figure shows how HAMP is able to resolve spatial features of showering cells, which were observed with a cross section of several HAMP footprints."

Comment: P. 18: how does WVP vary in this example?
Response: See Fig. R2. We added the summary "the IWV varies around $31.5 \pm 1.5\,\mathrm{kg\,m^{-2}}$ in this scene" to the Fig. 10 caption in the manuscript, as the main aspect of the figure is the liquid phase and IWV variation is only of secondary interest in that example.

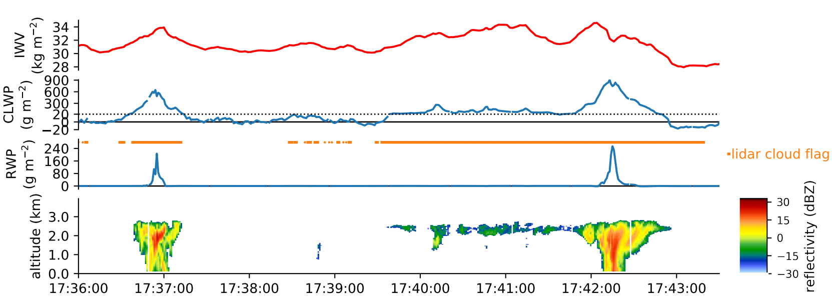

Figure R2. As Fig. 10 but with additional time series of IWV.

Comment: p. 19, lines 17-18: Some discussion of the sampling of the diurnal cycle (I presume HALO only flew during the day, were cumuli more prevalent in the afternoon?), and how that might alias into the results from the 2 seasons would be nice. I presume the BCO LWP measurements mentioned are diurnal averages
Response: Yes you are right, we added that "flights were scheduled during local daytime" to the second paragraph of Sec. 2.1. In the conclusions we added that "sound conclusions on the diurnal cycle can not be drawn from the data presented here, as the spatial variability of the clouds on the observed mesoscale was higher than an expected effect of the diurnal cycle." Radar time-height-plots of the NARVAL1 flights that were flown from Barbados to the East and back are rather symmetrical to the return point due to large-scale patterns. This means, a potentially diurnal cycle during the flights is overlaid by the changes in the larger scale cloud field. Additionally, we added in the outlook the following with respect to EUREC4A: "Also, more locally targeted flights, distributed over the daytime are planed to study the diurnal cycle."

Comment: P.11: an easy additional plot would be how LWP and RWP vary with lidar-derived cloud top height. This would be of scientific interest. How would that compare to, e.g., Byers and Hall, 1955?
Response: Indeed, liquid condensate load versus cloud extend is an interesting comparison. Figure R3 depicts the relations during the two campaigns. The overall impression of increasing rain amount with increasing cloud top height during NARVAL1 agrees with the findings by Byers and Hall (1955). The dry winter season during the NARVAL1 campaign compares best to their pioneer study according to their description. Differences exist in details and but are also partially due to the analysis approach. Byers and Hall subjectively identified cloud objects while we analyze the data profile-wise. For example, Byers and Hall (1955) found the lowest cloud top of a precipitating cloud near 1.8 km, whereas we already observed RWP > 10 $g\,m^{-2}$ for cloud top heights near 1.0 km. Stevens et al. (2019, in press) present in Fig. 7 a cloud object oriented analysis of NARVAL, that directly uses radar reflectivity instead of the RWP retrieval presented in our manuscript. A more detailed analysis of cloud dimensions in relation to their LWP and RWP will follow in a subsequent study, which is in preparation at the moment.

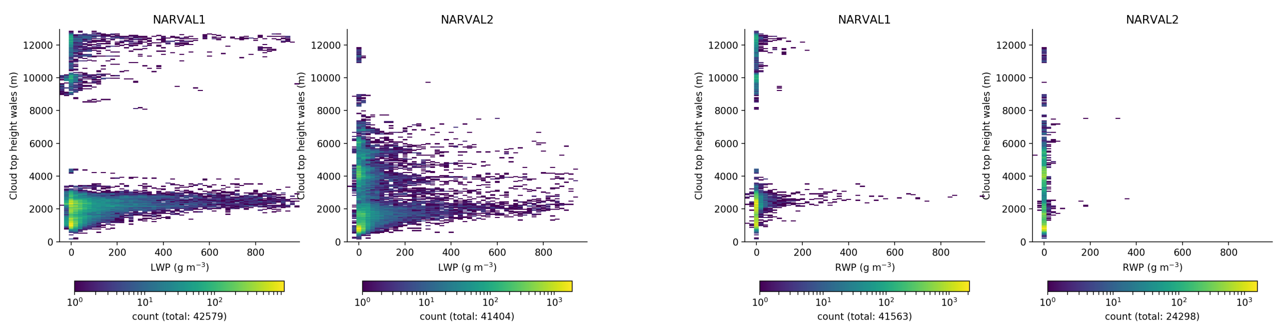

Figure R3. Lidar (WALES) derived cloud top height in relation to LWP (left) and RWP (right) for both campaigns.

Comment: P. 23, data availability: do the datasets have dois? They should.
Response: The DOI assignment was in preparation and is completed now. DOIs are added in the references.

Comment: The writing overall is fine, but there are small awkward uses of the English language sprinkled throughout that reflect English as a second language. If it is possible to find a native English speaker to read it that would polish the manuscript.
Response: Referee #1 pointed out problematic sentences and typos that are remedied now.

Comment: In particular the abstract and its first sentence needs a revisit (you could consider just removing the first sentence).
Other comments on the abstract: mention the frequencies you use. You don't mention the linear regression approach, is that intentional? Mention clear-sky frequency and LWP statistics, as opposed to focusing on IWV - the title only mentions LWP after all. Overall the abstract seems to have been written in a hurry.

Response: We removed the first sentence from the abstract as suggested and rearranged most parts of the abstract for improved comprehensibility. We incorporated the radiometer frequencies into the abstract. We don't mention the linear regression approach on purpose, as it is mainly used as reference to classical retrievals. Averages for LWP, RWP and cloudiness were added and the sentence order was rearranged such that it has the dry and wet season in a consistent order. Further, we reformulated the closing two sentences of the abstract after the discussion of the flight patterns question by referee #1 and reconsidering that issue. We conclude that our former formulation was to negative. The revised abstract reads now:

"Liquid water path (LWP) is an important quantity to characterize clouds. Passive microwave satellite sensors provide the most direct estimate on global scale, but suffer from high uncertainties due to large footprints and the superposition of cloud and precipitation signals. Here, we use high spatial resolution airborne microwave radiometer (MWR) measurements together with cloud radar and lidar observations to better understand the LWP of warm clouds over the tropical North Atlantic. The nadir measurements were taken by the German High Altitude and Long range research aircraft (HALO) in December 2013 (dry season) and August 2016 (wet season) during two Next generation Advanced Remote sensing for VALidation (NARVAL) campaigns.

Microwave retrievals of integrated water vapor (IWV), LWP and rain water path (RWP) are developed using artificial neural network techniques. A retrieval database is created using unique cloud-resolving simulations with 1.25 km grid spacing. The IWV and LWP retrievals share the same eight MWR frequency channels in the range from 22 GHz to 31 GHz and at 90 GHz as their sole input. The RWP retrieval combines active and passive microwave observations and is able to detect drizzle and light precipitation. The comparison of retrieved IWV with coincident dropsondes and water vapor lidar measurements shows root-mean-square deviations below $1.4\,\mathrm{kg\,m^{-2}}$ over the range from $20\,\mathrm{kg\,m^{-2}}$ to $60\,\mathrm{kg\,m^{-2}}$. This comparison raises the confidence in LWP retrievals which can only be assessed theoretically. The theoretical analysis shows that the LWP error is constant with $20\,\mathrm{g\,m^{-2}}$ for LWP below $100\,\mathrm{g\,m^{-2}}$. While the absolute LWP error increases with increasing LWP, the relative one decreases from 20 % at $100\,\mathrm{g\,m^{-2}}$ to 10 % at $500\,\mathrm{g\,m^{-2}}$. The identification of clear sky scenes by ancillary measurements, here backscatter lidar, is crucial for thin clouds (LWP < $12\,\mathrm{g\,m^{-2}}$) as the microwave retrieved LWP uncertainty is higher than 100 %.

The analysis of both campaigns reveals that clouds were more frequent (47 % vs. 30 % of the time) in the dry than in the wet season. Their average LWP ($63\,\mathrm{g\,m^{-2}}$ vs. $40\,\mathrm{g\,m^{-2}}$) and RWP ($6.7\,\mathrm{g\,m^{-2}}$ vs. $2.7\,\mathrm{g\,m^{-2}}$) were higher as well. Microwave scattering of ice, however, was observed less frequently in the dry season (0.5 % vs. 1.6 % of the time). We hypothesize that higher degree of cloud organization on larger scales in the wet season reduces the overall cloud cover and observed LWP. As to be expected, the observed IWV clearly shows that the dry season is on average less humid than the wet season ($28\,\mathrm{kg\,m^{-2}}$ vs. $41\,\mathrm{kg\,m^{-2}}$). The results reveal that the observed frequency distributions of IWV are substantially affected by the choice of the flight pattern. This should be kept in mind when using the airborne observations to carefully mediate between long-term ground-based and spaceborne measurements to draw statistically sound conclusions. "