Response to the comments from Anonymous Referee 1 for the submitted AMT paper: ´Dorff, H. et al. 2021: Horizontal geometry of trade-wind cumuli – aircraft observations from shortwave infrared imager versus radar profiler

We thank the AMT associating editor, Maximilian Maahn, as well as the Anonymous Referee #1, for this enlightening review. Please find below our response (in standard font) to the remarks from the Anonymous Referee #1 (in *italics*).

This paper contrasts 1D and 2D observations of cloud (size) during the NARVAL campaign. In general, I think the paper is great, and should be published quickly. Many of the questions I had while reading it were actually answered a section later.

Response: First, we want to thank you expressly for this motivating and enlightening overall feedback of our manuscript.

1. *Is AMT the best venue for this paper? Sure, it is nominally about comparing different observational techniques, but the results are much more broadly applicable to the cloud physics community. I feel ACP would be the more appropriate journal in the EGU stable. I therefore would suggest* ***moving to a different journal,*** *but otherwise* ***minor revisions***

Response: Indeed, our manuscript deals with a variety of characteristics and metrics relevant for better understanding of cloud physics. Since we link our findings to boundary layer conditions (primarily the wind field), we can understand your suggestion for another journal specialized on cloud physics and strongly thank you for this advice. However, we intend to approach the overall topic of trade-wind cumuli geometries from the observation perspective and in particular, from how the measurement methods can deteriorate our understanding of prevailing cloud geometries. We propose future work considering our technical methods to investigate the interactions of cloud geometries under the impact of the trade-wind boundary layer in more detail. A follow-up study using EUREC4A data, which extends the cloud sample significantly together with a well-defined characterisation of the boundary layer by dropsondes might be a helpful contribution to, e.g., ACP.

2. *In Section 4.1, I would like to see a few more statistics about the compatibility between radar and imager. For instance, what is the % agreement between the two on clear/cloudy pixels (false positive/negative rate, if you will). Is this a function of certain parameters and choices of thresholds?*

Response: We determined a rated matrix according to your suggestion including true positive rate (TPR), false negative rate (FNR), false positive rate (FPR), true negative rate (TNR), where we refer radar to regridded imager. We use this convention for clarity without assuming that the imager reflects necessarily the truth. We compile one matrix for the entire period (Table 1) and one (Table 2) for the flights RF03 and RF06 only, which were dominated by shallow convection.

Table 1: Rate matrix for clouds seen in radar and imager for all collocated flight periods.

%	radar cloud	no radar cloud
imager cloud	77.94 (TPR)	22.06 (FNR)
no imager cloud	7.41 (FPR)	92.59 (TNR)

Table 2: Rate matrix for cloud pixels seen in radar and imager for shallow convection dominated flights (RF03, RF06).

%	radar cloud	no radar cloud
imager cloud	32.83 (TPR)	67.16 (FNR)
No imager cloud	3.30 (FPR)	96.70 (TNR)

In fact, the radar is clearly less sensitive and detects only 78% (33%) of the all (shallow convective) clouds. At the same time, there are hardly any clouds, which are observed by the radar only. Just the morphological closing applied to the radar cloud mask may create faulty clouds. This effect results in a small false-positive rate of 7%(3%) for all (just shallow convective) clouds.

Apart from predominant cloud types, the results depend on the thresholds for assigning coarse gridded imager pixels as cloudy or clear (set to 50% in Sec. 3.3). Figure 1 illustrates how the device differing results (FPR and FNR) depend on this threshold for the specific flights.

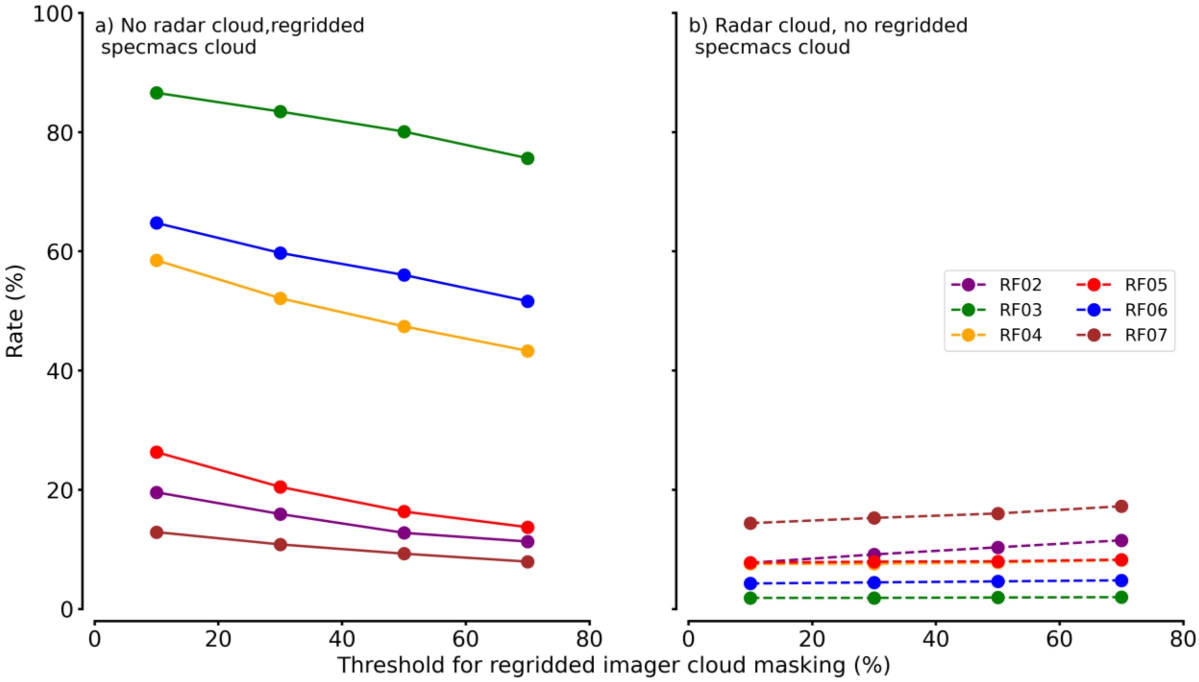

Figure 1:a) Ratio of no radar cloud pixels to cloud pixels in regridded imager (false negative rate) and b) ratio of radar cloud pixels to no regridded imager cloud pixels (false positive rate) for given threshold of imager pixels being cloud in high resolution for one single pixel in coarse-gridded resolution.

Increasing the threshold for the regridded cloud masking, the false negative rate decreases for all flights (Fig. 1a) showing that lower sensitivity in coarse-gridding creates more cloud gaps as seen by the radar. However, in particular for RF03 and RF06, the false negative rate remains significantly above 50 % as many clouds are not only fragmented but also completely undetected by the radar. Between the flights, the false negative rates show large spreads, while the false positive rate varies much less between flights and thresholds (Fig. 1b). For flights in shallow convection (RF03 and RF06), false positive cloud pixels from radar exist for less than 5% for all masking thresholds.

In conclusion, we see a higher dependency of radar based false negative rate on cloud types and on regridding thresholds as for the false positive rate. At this point, we want to remind the reader that we are using an imager cloud mask, which is also dependent on various thresholds. Altogether, this supplement analysis underline the complexity in how to determine the amount of clouds at a given region even when using collocating measurements. Since the here presented plots focus more on impacts on cloud fraction without considering individual cloud objects, this supplementary study is slightly above the scope of this manuscript where we want to focus on the representation of coherent cloud objects and their geometries. Nonetheless, we really appreciate your suggestion so that we included the following sentences in our manuscript at the beginning of Section 4 before coming to the coherent cloud objects:

“Comparing cloud fraction from regridded imager and radar after the FOV adjustment (Sec. 3.1), the radar is clearly less sensitive and detects only 78% (33%) of the all (shallow convective) clouds. At the same time, there are hardly any clouds, which are observed by the radar only. Just the morphological closing applied to the radar cloud mask may create faulty cloud pixels. This results in a small false-positive rate of 7% (3%) for all (just shallow convective) clouds. The rates also depend on our set regridding threshold (Sec. 3.3) but to a lesser extent. “

b) Do the 1D CSDs from both instruments pass a KS test for certain sets of parameters?

Response: Indeed, the KS-test represents a very useful statistical tool to compare the two samples (regridded imager and radar cloud size) and their distribution. We used different viewing curtains of the imager (light lines in Fig. 6) to indicate how the distribution varies for different viewing angles. Their envelope shows the robust difference between the radar distribution compared to the coarse-gridded imager for the entirety of imager viewing angles. This difference originates especially from the unresolved shallow low-level clouds leading to a higher relative contribution of larger clouds in the radar-based distribution. This effect partially decreases by increasing the cloud mask threshold when regridding. However, we cannot detect cases where the KS test will reveal an equal probability distribution.

3. *I noticed some choices of words where I am not sure I would have given those words the same meaning. Some suggestions for alternatives are below whenever I found them – not exhaustive, and maybe not always what you intended to say. It would be good to go through the paper with a non-native reader in mind with a somewhat limited English vocabulary, and when in doubt just use the simplest words possible. Otherwise the paper is written in a very clear language.*

Thank you for your advice regarding wording and phrasing. In the following, we specify our adjustments in the revised manuscript according to your suggestions.

Minor comments and word suggestions:

L 14: While-> Since

Response: the word has been deleted but its sentence was also rephrased (see next response).

L15: Do clouds become invisible, or simply gridpoint? The lower end of your CSDs is not much discussed, other than by the scale break. I can see several different mechanisms at play here

Response: We modified the sentence to the following: The radar encounters difficulties to represent clouds shorter than 200 m as they are either completely unresolved or considered like single grid points. Very shallow clouds can also remain unresolved due to a too low radar sensitivity. Both facts deteriorate the cloud size distribution significantly at this scale.

L42: Why is that a challenge? If anything, perhaps “2D imagers are better equipped to address the challenge…” or so

Response: We changed the sentence accordingly.

L50 barely -> rarely.

Response: we replaced the word.

L56 LES has been able to do this for a while, but now also for large domains (>100km)

Response: We added this for clarification at this place.

L73: How collocated are the instruments? How many meters away in spanwise and streamwise direction? I doubt that at least the streamwise direction is going to matter much (after the correction you talk about later), but good to mention here either way.

Response: Yes, this is a crucial aspect. The devices have fixed viewing directions on the aircraft frame and hence also between each other. As shown in Fig. 2,4,7,9, the orange line, indicating the radar FOV, is not centered in the imager FOV, but slightly shifted. In specific, the imager looks 2.6° ahead of the radar. The central imager across-track pixel is located by 0.55° shifted to the left in flight direction. For the modification of the manuscript, we have the impression that giving all these details in the introduction can be a bit distracting for the reader. Therefore, before going into detail in Section 3, we briefly referred to it in line 71 as follows: “Section 3 encompasses our applied synergy of both cloud masks comprising FOV adaptation between both devices and coordinate transformation in order to […].”

L99: How much is this FOV in practice in meters? and what is the typical resolution in meters? I’m not a fan of pixel# as a unit. Perhaps the spatial equivalents of Time and Pixel units can be put on secondary x/y axes in figs2,4,7,9?

Response: Typical resolutions are given in Sec 2.3, but still we fully understand your impression and suggestion regarding the plot axes. However, adding secondary distance-based axes leads to issues for the shown illustration of the imager because the aircraft has a changing air attitude (pitch, yaw, roll angle). Moreover, the distance between signal source (cloud or ocean) and aircraft changes at every point. Requiring the coordinate transformation for each cloud (Sec 3.3) as done in later analysis, the squared panel in the mentioned figures become snappy at the borders when showing the transferred cloud mask. At this place, we consider this as inappropriate for the readability. In order to still taking into account your helpful comment, we included some short explanations of the respective viewing distance in the first captions (Fig.2, Fig.4) such as repeating the FOV of each pixel that Sec. 2.3 specifies in more detail.

L102: pronounce -> result

Response: wording changed

L105 non-zero reflectivity

Response: *declaration adapted*

L143: Is CTH the correct metric? Since you're integrating over the entire depth of the cloud, mid-cloud level would be more precise, I guess. Again, shouldn't matter much in practice for these shallow clouds.

Response: We confirm that our method possesses weaknesses. They result from the restricted abilities we have. Principally, the imager sees clouds from the top, and the horizontally projected cloud top reveals the cloud shapes in the 2D mask. Accordingly, CTH is the correct metric, although at the edges of the FOV clouds are slightly more captured from the side as described in Sec. 3.2.

However, indeed, a certain cloud penetration path exists for the imager and radar. Comparing with simultaneous LIDAR measurements (e.g. Gutleben et al, 2019), we have found tendencies of the radar to underestimate cloud top height slightly. In particular, for the shallow low-level clouds, radar-based CTHs (as a function of time) are supposed to be too low due to the radar sensitivity, so that the signal actually originates from slightly deeper inside the cloud a little bit towards mid-level height.

Therefore, we conducted various visual inspections of the collocated cloud masks such as we illustrate in Fig. 4. They show us an overall reasonable time shift being adapted to the radar for several cloud cases. Furthermore, when considering the uncertainties of CTH (Sec. 3.2), our results remain robust.

L143: "lower and further to the aircraft" not sure what that means exactly.

Response: We changed that to: "The lower and more distant the clouds are to the aircraft, […]"

L149: Emphasize 2D connectivity

Response: the emphasis on two dimensions is now given.

L174: dammed -> limited

Response: We replaced the word.

L183: This does introduce the bias that cloud size is artificially limited by the FOV size.

Response: Yes, it truly does. This explains why we also compare all along-track cloud sizes at least to lengths up to 10 km when remaining in the aircraft-following coordinate system without neglecting incomplete clouds in Sec 4.2 and 4.3. Both approaches have their pros and cons. To make your point clearer, we rephrased the bullet point as follows: "Clouds extending out of the FOV are neglected as their 2D geometries exterior of the FOV are not obtainable. In addition, we exclude clouds having an along-track cloud size bigger than the imager across-track FOV, as otherwise large clouds orientated along the flight path would be considered preferentially. Since cloud size then becomes artificially limited by the FOV size, cloud size statistics remain biased for undersampled larger clouds approaching the typical image scale.

L240: This may be cloud misrepresentation, but it is the fair comparison between the two instruments. This is an important part, because it validates the radar for use in the (extremely common) situation that no imager is available.

Response: We agree that this is a fair comparison, which is why we considered the overlapping 1D path in the following sections from both devices. To make our statement on the effects from 2D to 1D more understandable, we rephrased the second last sentence of the paragraph as follows: "[…] fraction of ID labelled clouds from the 2D field reaches into radar curtain (Fig. 5). Several cloud fragments therefrom can appear distinctively within the 1D curtain although they actually belong to one cloud.

L281: I would be interested in a bit more discussion of the scale break, as it is located much sooner than often reported for shallow Cu (1km+). Are the authors sure that this is not an artifact of the observational strategies/instrument resolution?

Response: Thank you for mentioning the complexity in the interpretation of the scale-break. Although there exist studies showing higher scale-break values, we found several studies considering observations of comparable resolution that suggest plausibility of our results. In our literature review we conducted before, we found various sources, that resolve clouds below hectometer scale, locating scale breaks between 0.5-1.0 km (such as Zhao and Di Girolamo, 2007; Dawe and Austin, 2012; Heus and Seifert, 2013). Another source we want to highlight in this regard is Mieslinger et al. (2019). Using spaceborne observations from ASTER having comparable resolution to the airborne imager, Mieslinger et al, (2019) so far comprise the largest dataset of marine shallow cumulus clouds covering also our region of interest. Since the dataset of Mieslinger et al (2019) includes decameter resolution for domain sizes in the order of hundred kilometers, we put a lot of trust in their findings.

Higher values of scale-breaks with 1km+ mostly occur for dataset constrained to coarser resolutions (e.g. Wood et al, 2011) not capturing the ubiquity of very small-scale clouds, which may then be merged to single hectometer scale cloud objects. Yet, we agree on your remark highlighting the complexity of scale break location and its origin that might be atmospherically driven or method-specific, why literature keeps on actively debating.

In our manuscript, we gave some respective information in Sec. 3.4 beforehand and afterwards in Sec. 4.3 when considering the entire 2D FOV with a larger cloud sample. We have the impression that it is more appropriate to add some final discussion of the scale-break at the end of 4.3. Therefore, we added to line 281 (of the preprint): “We discuss its location further in the following section when considering the entire 2D FOV with increasing cloud sample size.”

In Sec. 4.3, we included more literature to compare scale break values. In accordance with the remarks of *RC2*, we added the following after line 321 of the preprint:

“Thus, observations from the imager specMACS during NARVAL-II well reproduce trade-wind cloud size distributions found in comparable studies.

Nonetheless, we highlight the ongoing debate in literature about location and artificial or boundary-layer driven origin of scale breaks (Mieslinger et al., 2019). We see that resolution affects the location of the scale-break in a way that it is missing for airborne observations in hectometer resolution. Although Wood and Field (2011) locate scale breaks above 1 km using hectometer scale spaceborne data, we cannot identify this from the radar curtain samples. On the other hand, larger clouds may be misinterpreted from the imager if only their edges reach into the imager FOV and artificially enhance the scale break through length underestimation. If we completely neglect clouds reaching out of the FOV, we do also produce biases with increasing cloud size (Sec. 3.2). Due to the complexity of scale break origin, some studies, e.g. van Laar et al. (2019), suggest to apply exponential power law fits (Ding et al, 2014) to prevent the scale break by a modified single distribution.

L298: Could be interest to compare the overlap corrections from Sulak et al (JGR, 2020).

Response: We agree that investigating the cloud overlap ratio from the radar is applicable and of certain interest. Unfortunately, we are not certain where to place this source at L298 as we completely remain in the horizontal projection and do not consider vertical overlap, here. Yet, as your feedback is inspiring, we included an average horizontally inverse length factor defined as the average ratio of maximum along-track length divided by effective length in the radar-equivalent imager curtain (red distance divided by blue distance in Fig.07). This is added to the manuscript after Preprint line 302 as follows: “[…] For all imager clouds reaching into the radar curtain, we calculate a mean inverse length factor, defined as the average ratio of maximum along-track length divided by the effective cloud length in the radar curtain (blue and red distances in Fig. 7). For these clouds, this horizontal analogue to the vertical overlap ratio in Sulak et al. (2020) reveals that, on average, maximum along-track cloud length differs to curtain length by a factor of 2.63. Since this factor has strong cloud length dependency, its value strongly varies with cloud size and is presumably underestimated due to the FOV limits for larger clouds.”

We further included some suggestions for complementary studies following Sulak et al. 2020 in the outlook, as given: “With this 3D representation based on merged imager and profiler, inverse cloud overlap studies such as from Sulak et al. (2020) can be complemented for vertical and horizontal perspective under the influence of BL characteristics derived from dropsondes.”