Reconciling Chord Length Distributions and Area Distributions for Fields of Fractal Cumulus Clouds

Nicholas R. Barron [1,†] [ID], **Shawn D. Ryan** [2] [ID] **and Thijs Heus** [1,*] [ID]

[1]   Department of Physics, Cleveland State University, Cleveland, OH 44115, USA; nrb171@psu.edu
[2]   Department of Mathematics and Statistics, Cleveland State University, Cleveland, OH 44115, USA; s.d.ryan@csuohio.edu
[*]   Correspondence: t.heus@csuohio.edu
[†]   Current address: Department of Meteorology and Atmospheric Science, Penn State University, State College, PA 16801, USA.

check for updates

Abstract: While the total cover of broken cloud fields can in principle be obtained from one-dimensional measurements, the cloud size distribution normally differs between two-dimensional (area) and one-dimensional retrieval (chord length) methods. In this study, we use output from high-resolution Large Eddy Simulations to generate a transfer function between the two. We retrieve chord lengths and areas for many clouds, and plot the one as a function of the other, and vice versa. We find that the cloud area distribution conditional on the chord length behaves like a gamma distribution with well-behaved parameters, with a mean $\mu = 1.1L$ and a shape parameter $\beta = L^{-0.645}$. Using this information, we are able to generate a transfer function that can adjust the chord length distribution so that it comes much closer to the cloud area distribution. Our transfer function improves the error in predicting the mean cloud size, and is performs without strong biases for smaller sample sizes. However, we find that the method is still has difficulties in accurately predicting the frequency of occurrence of the largest cloud sizes.

Keywords: cumulus clouds; cloud size distributions; Large Eddy Simulations; fractals

1. Introduction

Clouds are a challenging component of the atmosphere to model [1]. This is particularly true as we enter the gray zone of convection [2,3], where some convection is resolved but smaller clouds still need to be represented in the subgrid parameterization. One approach to resolve this problem is to formulate the convection parameterization as a function of cloud size [4–6]. In order to do so, a necessary first step is to observe and understand the behavior of the cloud size distribution. Many studies have retrieved cloud size distributions, usually either from a large observational dataset with a wide variety of synoptic conditions [7–9] or from a numerical simulation [10,11]. For a reliable, high-resolution cloud size distribution of shallow convection based on observations, vertically pointing instruments (e.g., radar, lidar, and ceilometer) are often used [12]. However, this generates only one-dimensional transects through the clouds. In other words, a vertically pointing measurement would result in a chord length distribution, while a cloud area distribution is arguably more relevant one for parameterizations of convective mass flux and of cloud radiative transfer [13]. A good understanding of the chord length distribution also helps in comparing the cloud base properties between observations and simulations [14]. In theory, we could keep using a high-resolution numerical model (like a Large Eddy Simulation (LES)) as an intermediary between observations and parameterizations, but it would clearly be beneficial to have a transfer function to generate a cloud area distribution directly from observations.

There are several biases in a chord length distribution based on a straight line through a cloud field, and a naïve interpretation of a chord length distribution could lead to a drastically different

interpretation of the cloud field, as can be seen from the difference between Figure 1a,b. Some of the more prominent biases are

1. a higher chance of missing small clouds all together biases towards large chord lengths;
2. a high likelihood of hitting clouds off-center biases towards small chord lengths; and
3. irregularities in cloud shapes, such as gaps and the fractal dimension of the cloud edge [15], bias toward small chord lengths.

According to previous studies [7], the net effect is that the chord length distribution skews small, although it is unclear by how much and how this is a function of cloud distribution and spacing.
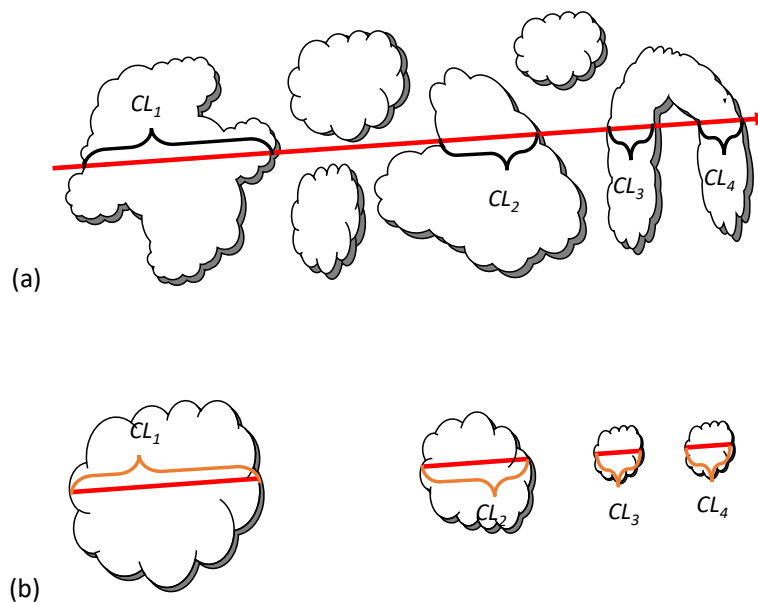

Figure 1. Illustration of potential biases in the chord length distribution. (**a**) A linear observation through a cloud field. (**b**) A naïve interpretation of the cloud field based on those chord lengths.

Previous research has been performed on the comparison between chord length and cloud area distributions [7,16,17]. In particular, Romps and Vogelmann (2017, hereafter RV17) use a mainly theoretical approach to create a conversion by assuming simple geometric shapes for the clouds, such as circles or squares. This allows for an analytical relationship between chord length and cloud area distribution, and for addressing the first two biases listed above. In other words, RV17 assumed that additional cloud transects due to the irregular shape of the clouds can be safely neglected.

It is important to first properly define cloud *size*. A chord length L is defined as the length of a contiguous line segment within a cloud. The area A is simply the projected area of the cloud. To easily compare the two, it is useful to define a linear cloud size D based on the area, but with dimensions of length instead of length squared.

RV17 uses the maximum width of the cloud as its definition for the linear cloud size D, which works well for regularly shaped clouds. In what they call the "simple method", RV17 assumes properly aligned square clouds, so that the maximum width of the cloud is equal to the square-root of the cloud area, $D = \sqrt{A}$. This is also consistent with the cloud size definition used in other works [7,18], so that is what we will use as well. In the remainder of this manuscript, we will use this linear cloud size D as our proxy for cloud area. We like to stress that while we use a linear cloud size D, we do not assume any particular cloud shape here. D is merely a proxy for A, which happens to be equal to the side of a cloud for a square shape, or equal to $\sqrt{\pi}R$ for a circular cloud, with R the radius.

Given a number density $n(D)$ of clouds with widths D, the probability density of sampling a cloud of width D is equal to $P(D) = \frac{n(D)}{n_0}$, with n_0 the total number of clouds in the population. The area weighted average cloud size is then

$$\langle A \rangle = \int_0^\infty P(D)A(D)dD = \int_0^\infty P(D)D^2 dD. \tag{1}$$

We can now define the probability of sampling a cloud chord length $p(L|D)dL$ between L and $L + dL$, conditional on a cloud width D. That means that the probability density of sampling a chord length L is equal to

$$P(L) = \int_0^\infty n(D)DP(L|D)dD. \tag{2}$$

The overarching goal of this paper, and of RV17, is to calculate $n(D)$ from $P(L)$, which means finding an expression for $p(L|D)$, or rather its inverse: $p(D|L)$. The major difference between the current paper and RV17 is that RV17 started from theoretical and Euclidean clouds (circles and squares), while we will take an empirical approach with LES fields of cumulus clouds, which includes the complex shape of these clouds.

For square clouds, aligned in parallel with the direction of measurement, RV17 derives a trivial relationship:

$$n(D) = \frac{P(L)}{L} = \frac{P(D)}{D}, \tag{3}$$

which means that the probability density function of the chord length is identical to that of the square root of the area, which is the linear cloud size used in [7,18]. For circular clouds, the RV17 result is an inverse Abel transform, which is a more involved operation:

$$n(D) = -\frac{2D}{\pi} \int_D^\infty \frac{\frac{d}{dL}\left[\frac{P(L)}{L}\right]}{\sqrt{L^2 - D^2}} dL. \tag{4}$$

For Landsat imagery, RV17 finds little difference for the area-average cloud size between either method, although the Abel transformation performs better in approximating the true distribution.

In this paper, we present an empirical statistical relationship between chord length and linear cloud size for fields of shallow cumulus clouds, which includes the impact of the non-Euclidean shape of these clouds. We derive this relationship using high-resolution (25 m), large domain (25 km) Large Eddy Simulations (LESs). To lead to this goal, the outline of this paper is as follows. After a brief overview of the methodology, we will first study the chord length distribution given a certain cloud, which we then can generalize to all clouds of a specific area. Then, we will invert the problem to look at the linear cloud size distribution conditional on an observed chord length. As it turns out, this cloud area distribution can be well described as a gamma distribution, so we will then model the parameters of this gamma distribution as a function of chord length. This yields a model to convert chord length to linear cloud size distribution. Finally, we apply this model to a variety of simulated cloud fields to establish its general validity, and compare it to the RV17 model.

2. Methods

We are basing our analysis on cloud fields generated with MicroHH [19]. This modern, fast Large Eddy Simulation model has been validated against a wide range of standard cases, including all the intercomparison cases used in this study: BOMEX [20], which are non-precipitating marine clouds; ARM-SGP [21] a diurnal cycle over land; and RICO [22], somewhat deeper and precipitating marine clouds.

We simulated all cases using the forcing and settings as described in the respective case description papers. Each simulation was run on a 25 m resolution in both the horizontal and vertical directions, with a horizontal domain size of 25 km^2. For BOMEX, we simulated 10 h, and discarded the first 3

as spin up. As BOMEX is a steady-state case, the final 7 h are aggregated in our analysis. For RICO, we simulated 60 h, resulting in a deeper and more organized cloud field [23]. We used a 2-moment bulk microphysics scheme [24] to include precipitation in RICO. Note that RICO is not a steady-state case, and it therefore allows us to explore a larger range of cloud size distributions. Similarly, as ARM-SGP is a diurnal cycle over land, the cloud size distribution varies significantly between scenes, with a dominance of small forced clouds especially in the earlier stages.

As a validation of our method, we applied our algorithm to one case from the Large Eddy Simulation (LES) ARM Symbiotic Simulation and Observation (LASSO) [25] database. These are realistic and routine simulations of cumulus fields over the ARM-Southern Great Planes observatory in Oklahoma. From this LASSO database, we used 9 June 2015 in the current study and selected the configurations with the best match to the observations in cloud cover and liquid water path. As the simulations in the LASSO database were done on a relatively coarse resolution of 100 m, we re-ran all cases with MicroHH on a 25 m higher resolution.

In Section 3.1, we use a simple retrieval method for the chord length: Line segments along the x-direction and the y-direction of the LES grid are used to get an estimate of the chord length distribution. In Section 3.2 and onwards, we use a more detailed method to focus on the effect of small scale irregularities: For every LES grid point inside each cloud, we draw a number of lines under different angles. Each of these lines generates one or more chord lengths through this particular cloud; the chord length distribution is then retrieved by taking the probability density of all of these chord lengths. Sensitivity tests show a similar outcome for different numbers of points, or different numbers of lines through those points. Our results were also robust for slightly different methods, such as picking lines between every point on the perimeter of each cloud.

To generate aggregated distributions over the entire cloud field, we then collect chord length distributions of all the clouds of a particular linear size. As we know the actual linear cloud size distribution, we can then invert this to find the linear cloud size distribution conditional on chord length.

3. Results and Discussion

3.1. Sensitivity and Convergence

We define clouds based on their projected area, so a cloud object is defined as a contiguous region with a non-zero liquid water path (LWP). A larger threshold on the LWP would result in smaller cloud sizes, but our results are independent thereof: As we argue that the irregularities are important in the discrepancies between the chord length distribution and the linear cloud size distribution, we show the perimeter vs. linear cloud size in Figure 2. The best fit is a power law with a slope of 1.32, independent of the LWP threshold. This is in agreement with previously report fractal dimensions [15]. Given the same fractal dimension, there is no reason to expect sensitivity to the LWP thresholds for our results (and indeed, we did not found any).
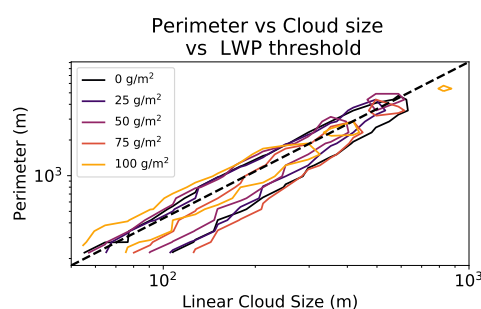

Figure 2. Linear cloud size vs. perimeter for different liquid water path thresholds. The slope of the best fit line signifies a fractal dimension of 1.32.

It is well known that the sample size has significant influence on the variability in the observables, such as the cloud number distribution [26,27]. In Figure 3a,b, we show a box–whisker plot of the synthetic observations of the cloud cover as a function of sample size, using the relative coverage of a random subdomain of size $\mathcal{L}^2$ in Figure 3a, and the relative in-cloud portion of a linear observation of length $\mathcal{L}$ in Figure 3b. For the 2D observations, we divide our LES domain into subdomains, and then calculate the cloud cover for that particular subdomain. For the 1D synthetic observations, we use every possible line segment along the x- and y-direction in our domain, and determine the cloud cover for it. As a direct consequence, the distribution is near-binary for the smallest sample sizes (either totally cloudy or totally clear), and a cloud cover of zero only ceases to be a likely possibility once the sample size is larger than the typical nearest neighbor size, in the order of a few km.
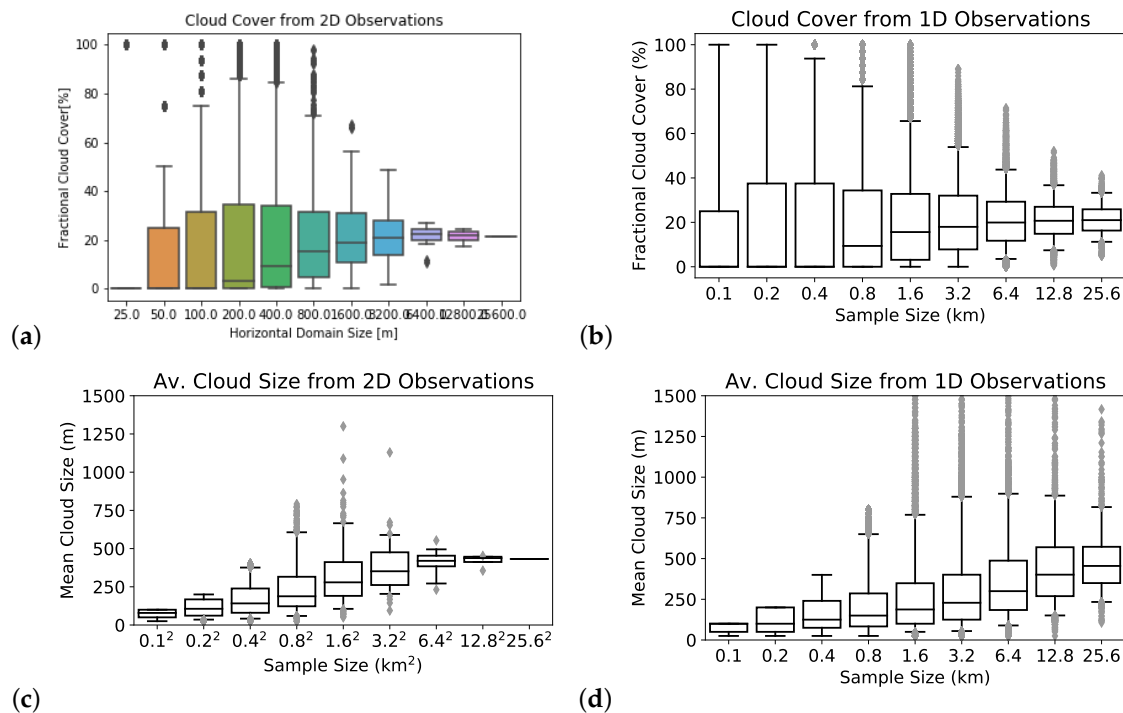

(**a**)   (**b**)

(**c**)   (**d**)

Figure 3. Box–whisker plot of the LES-retrieved fractional cloud cover and cloud size as a function of sample size: (**a,b**) cloud cover and (**c,d**) weighted average cloud size; (**a,c**) for area measurements and (**b,d**) based on chord length observations. Based on the 8th hour of RICO. Fifty percent of the measurements falls within the box, and 95% of the measurements within the whiskers. The line within the box gives the median value.

Based on this figure, it is clear that both area measurements (such as satellite pictures or Total Sky Imagery (TSI)) and linear measurements (such as ceilometers or airplanes) of cloud cover converge to the same value, as expected. However, it takes a much longer linear sample size to reduce the variance in observed cloud cover, especially in comparison with the area measurement. While linear measurements of 3 km yield a correct cloud cover on average, the variance is still significant, even after 25 km, similar to what was found by other studies [27–29]. In Figure 3c,d, we do the same exercise, but now for the area-averaged cloud size, and we observe a similar picture, although we now see that even the median cloud size has not converged for one-dimensional observations.

No convergence is expected between chord length size and linear cloud area size; for sufficiently long sample sizes, either should converge to its own value, which clearly has not happened yet for the one dimensional observations. As a back of the envelope calculation, a 25 km line observation with 10 m/s wind speed means a measurement time of ~40 min, in which a cloud field would have changed drastically during a diurnal cycle. Clearly, to retrieve a reliable estimate for cloud size or even for cloud

cover, a single 1-dimensional observation does not suffice. To overcome this in observations, one will either have to rely on 2D observations such as TSI, or use multiple vertically pointing instruments.

Of course, in LES the sample size does not need to be a problem; in the remainder of this paper, we will use a large number of crossings per cloud to ensure that our results are converged. To illustrate this method, it is most instructive to start by looking at the possible chord lengths retrieved from a single cloud first, which we will do in the next section.

3.2. Chord Length Distribution Conditional on Linear Size

As an example, we first explore the chord length distribution that is retrieved from a single cloud (Figure 4a), with a number of different crossings through it for illustration. This particular cloud is retrieved from the RICO simulation, although qualitatively identical results can be retrieved from any other simulation.

For the chord length distribution of a single cloud, we take transects through several locations inside each cloud, under different angles. The number of locations and angles are chosen randomly, and to a sufficient amount to result in a converging chord length distribution. If a particular transect leaves and then enters the cloud multiple times, all chords are individually included in the distribution.

Even from the crossings that were drawn, it is suggestive that it is likely to cross the cloud through its edge, resulting in relatively small chord lengths. Figure 4b quantifies this suggestion in the chord length distribution for this particular cloud. The distribution is characterized by a strong peak that represents chord lengths through the cloud edge, and a broad area of chord lengths that are retrieved from crossings through the bulk of the cloud, resulting in a chord length close to the linear cloud size of this cloud.
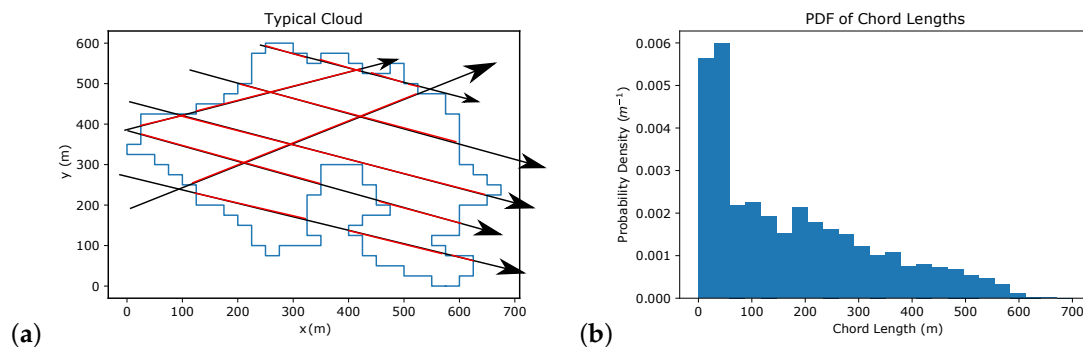

Figure 4. Chord length distribution through a single cloud from the 8th hour of the RICO simulation. (**a**) The projected footprint of the cloud, with transects for illustrative purposes. (**b**) The resulting chord length distribution.

By aggregating all clouds of a single size, a chord length probability density conditional on the cloud size $P(L|D)$ can be retrieved. Examples thereof are given in Figure 5a, for several different cloud sizes, all taken from the 8th hour of RICO. For all but the smallest cloud sizes, we observe a first peak that represents transects through the edges, and a second broad peak slightly below linear cloud size. A small portion of the chord lengths are larger than the linear cloud size, thanks to a crossing through the cloud that is longer.

The first peak is clearly the signal of the cloud being a fractal. Given its peak below 100 m and close to the 25 m grid spacing of our simulations, it is difficult to study the details of this distribution, without a massive decrease of the grid spacing of the simulation. It is also a signal that can be easily missed in satellite observations, and even in situ measurements or ground-based remote sensing may often ignore a signal that has a duration of less than 10 s [30]. If we would be able to ignore this first peak, a distribution emerges that is closely aligned with the RV17 results: The typical chord length slightly underestimates the linear size of the cloud. We should also note that the width of the second peak is significant, with a standard deviation in the order of 100 m, meaning that a single chord length

yields little information on the actual cloud area, and that a large sample size is needed to reconstruct the linear cloud size distribution.

In Figure 5b, the location of the local maximum in the distribution is plotted as a function of linear cloud size, which is equal to the modal chord length—but ignoring a maximum at zero, which is fueled by edge effects. For each linear cloud size, the modal chord length 5 to 10% lower than the linear cloud size, due to the prevalence of off-center transects.
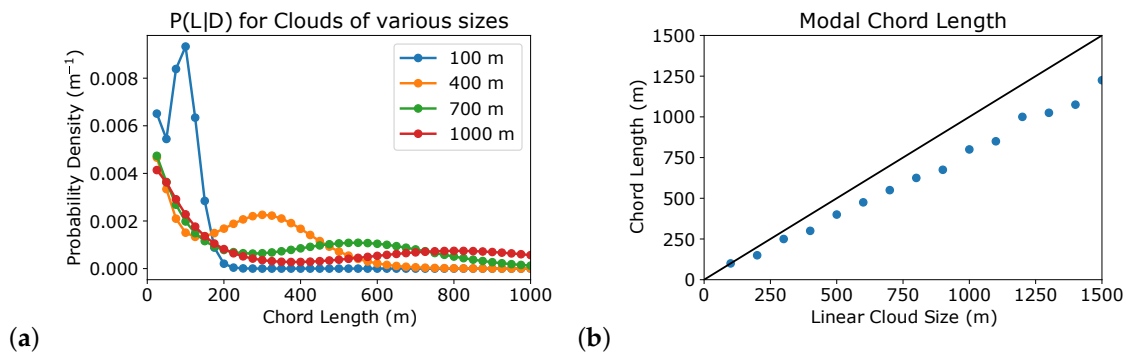

Figure 5. (**a**) Chord length distribution through all clouds with the same linear size in the 8th hour of the RICO simulation. (**b**) Modal chord length as a function of linear cloud size.

3.3. Linear Cloud Size Distribution Conditional on Chord Length

To retrieve the linear cloud size distribution conditional on the chord length $P(D|L)dD$, we now invert the exercise of the previous section: For each chord length L, we generate the linear cloud size distribution of the clouds that were transected. Figure 6 shows the results for several chord lengths.

For chord lengths beyond 200 m, the distribution can be well approximated by a gamma distribution:

$$P_\Gamma(x)dx = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x} \tag{5}$$

with $\Gamma(\alpha)$ being the Gamma function. However, for smaller chord lengths a secondary peak shows, which is related to chord lengths through the edges of larger clouds. For illustrative purposes, we have modeled the secondary peak as a gamma distribution as well, although there is considerable variability in the fit parameters of that peak, and we did not see much consistent behavior for this secondary peak from simulation to simulation, as it is likely strongly dependent on the overall cloud size distribution.

A gamma distribution can be characterized by several sets of parameters, for instance, its mean $\mu = \frac{\alpha}{\beta}$ and the normalized shape parameter $\beta = \frac{\sqrt{\sigma^2}}{\mu}$. We retrieved these parameters as a function of chord length for each of the snapshots from each of our training simulations (ARM, BOMEX, and RICO), and then took the mean value of them to create a general transfer function that is applicable to a wider range of cloud fields.

The results are plotted in Figure 7. In these two panels, the dots are the average values for these parameters taken over all simulations, with the lines denoting the standard deviation. Several features are clear from these graphs. As expected based on Figure 6, for the very smallest chord lengths (<100 m), the average associated cloud size μ is much larger than the chord length, which is because a single gamma distribution is an ill fit for these bimodal distributions, see also Figure 6. A peak linear cloud size close to the chord length is still there, but a secondary peak at much larger cloud sizes contributes as well. For those same small chord lengths, the shape parameter tapers of slightly, but only because the μ in the denominator is increasing. For much of the range of chord lengths, the mean cloud size goes linearly, and the shape parameter falls off with a power law. For the largest clouds a linear function tends to overpredict the mean cloud size, as the scale break in the cloud size distribution limits the number of clouds with a large area.
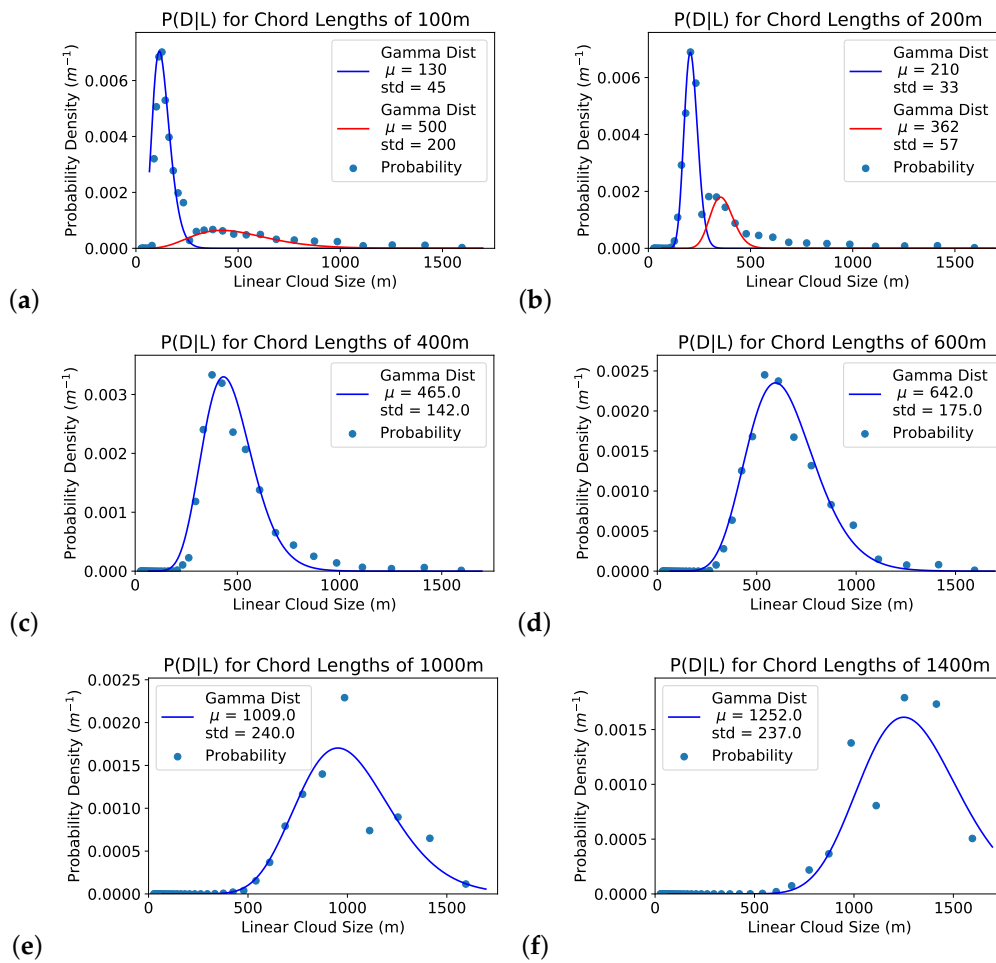
Figure 6. Cloud size distribution conditional on the chord length for the 8th hour of the RICO simulation. (**a**) $L = 100$ m. (**b**) $L = 200$ m. (**c**) $L = 400$ m. (**d**) $L = 600$ m. (**e**) $L = 1000$ m. (**f**) $L = 1400$ m.
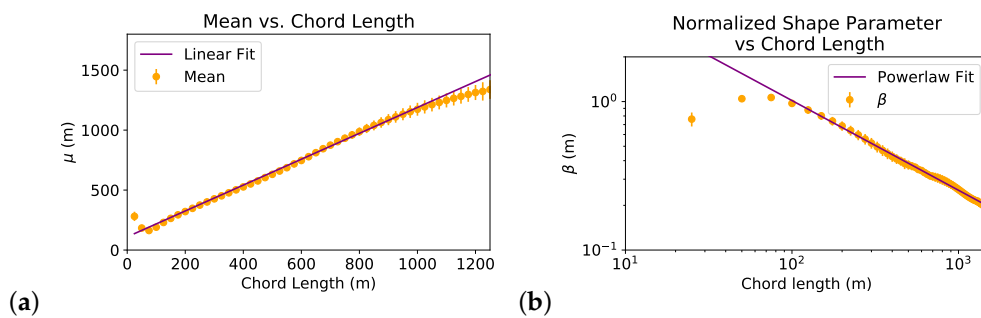

Figure 7. Fitting parameters of the gamma distribution, as a function of chord length. The vertical lines depict the standard deviation between different cases. (**a**) The mean μ. (**b**) The shape parameter β.

We can now model the gamma parameters as a function of chord length. First, for the mean we use a simple linear fit:

$$\mu = 1.1L + 110, \tag{6}$$

which fits well for all but the largest chord lengths, and makes intuitive sense in that the chord length tends to underestimate the linear cloud size.

For the shape parameter, we use a power law fit:

$$\beta = 21L^{-0.645}, \tag{7}$$

This fit ignores the chord lengths below 100 m, but otherwise fits well until $L = 2$ km. We emphasize that this is an empirical fit; while it makes sense that the mean μ is linear with a slope of slightly above 1, we do not have a good physical interpretation for the shape parameter, other than that we expect a strongly declining curve. We hypothesize that this may be related to the cloud size distribution, as larger clouds are much more rare than smaller clouds, given distributions that are often characterized by a D^b power law, with $b \approx -2$.

Using our empirical fit for the gamma distribution $P_\Gamma(L; \mu, \beta)$, we are now able to reconstruct the linear cloud size distribution $P(D)dD$ from the chord length distribution $P(L)dL$, using

$$P(D)dL = \int_0^\infty P(L)P_\Gamma(L; \mu(L), \beta(L))dL. \tag{8}$$

Figure 8 shows the chord length, adjusted chord length, and linear cloud size distribution for different cloud fields. Remembering that the RV17 "simple method" assumed well-aligned squares for its clouds, the unadjusted chord length distribution (in red) should be equivalent to the linear cloud size distribution (in black). As reported by RV17, this approach tends to underestimate the real linear cloud size, and Figure 8 confirms that. However, after adjustment by our algorithm the chord length distribution (blue) is much closer to the linear cloud size.
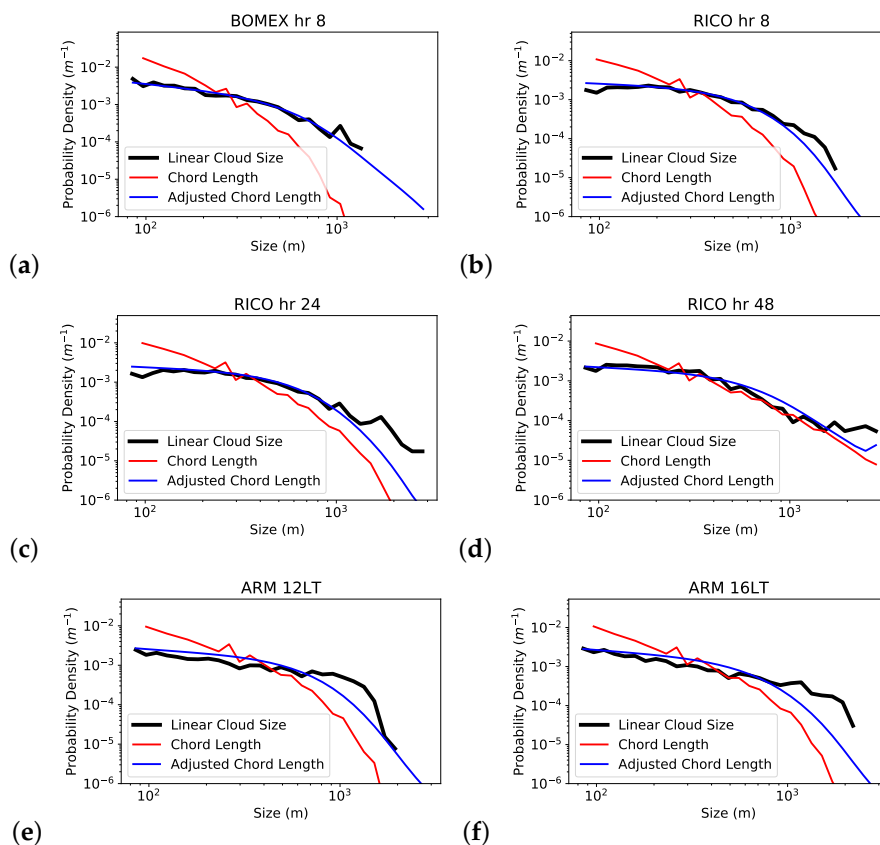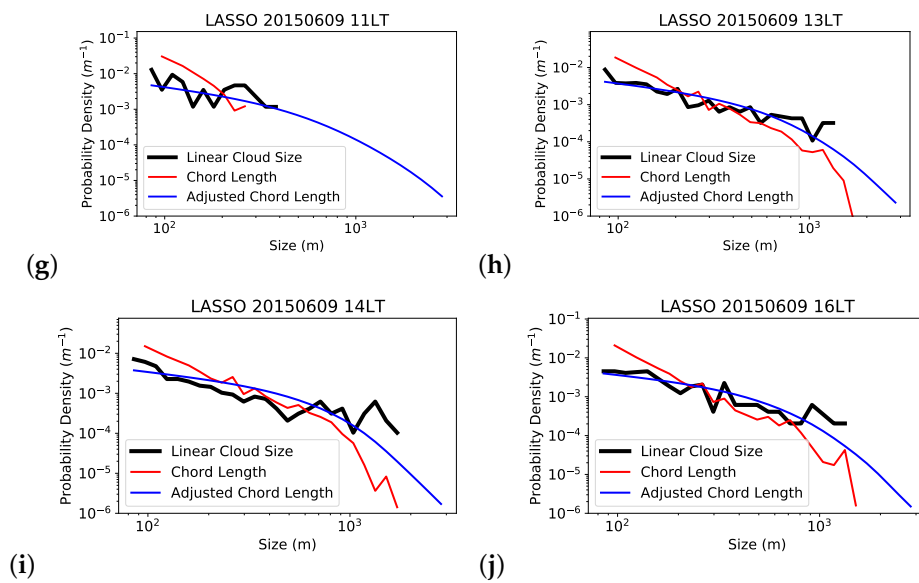

Figure 8. *Cont.*

Figure 8. Reconstructed linear cloud size distributions for different cases, with the linear cloud size (black), the chord length distribution (red), and the adjusted chord length distribution (blue). (**a**) BOMEX 8th hour. (**b**–**d**) RICO hr 8, 24, and 48. (**e**–**f**) ARM 12LT and 16LT. (**g**–**j**) LASSO 11, 13, 14, and 16 LT.

Figure 8g–j shows the cloud size distribution for the LASSO 2015/06/09 case, which was not used in training the algorithm. In this independent data set, the adjusted chord length distribution is also close to the linear cloud size distribution. While our method does well in reproducing most of the cloud size distribution, it does not have good knowledge in predicting the maximum cloud size. As a result, the adjusted chord length tends to predict large clouds that are not present early in the day for the diurnal cycle cases (see panel (g)), and has a tendency to under predict the largest clouds in mature and organized fields, resulting in an overprediction of medium cloud sizes (in the order of 200–600 m; see panels (c–f)). In general, however, the adjusted chord length distribution follows the shape of the linear cloud size distribution well.

Finally, Table 1 shows the area-average cloud size, as retrieved from the true cloud size distribution, the chord length distribution, and the adjusted chord length distribution using Equation (1). The adjusted chord length improves the predicted area-average cloud size for every case, although it still tends to underpredict the cloud size by 14% on average, instead of 46% without correction. It is also clear that the correction tends to underpredict more significantly for larger clouds (see Figure 8). This is likely due to the limited amount of training data; a larger set with a wider range of cloud sizes should improve our results. Another avenue of improvement would be to have a better estimate for the scale break and the maximum cloud size. If the maximum cloud size is underpredicted, the impact of these largest clouds on the typical cloud size can still be significant.

Table 1. The area-weighted average cloud size for different instances, calculated using the actual cloud area, the chord length, and the adjusted chord length, together with the relative error for the latter two methods. For RICO and BOMEX, the simulation time is shown. For ARM and LASSO, the local time is shown.

	Linear	Chord Length	pct Change	Adjusted Chord Length	pct Change
RICO h8	545	292	−47	478	−12
RICO h24	741	354	−52	524	−29
RICO h40	716	506	−29	643	−10
RICO h48	790	554	−30	697	−12
RICO h56	656	512	−22	649	−1
BOMEX h4	435	210	−52	470	8
BOMEX h5	467	227	−51	474	1
BOMEX h6	446	216	−52	470	5
BOMEX h7	445	214	−52	469	5
BOMEX h8	452	227	−50	472	4
ARM h11	383	249	−35	466	22
ARM h12	713	343	−52	518	−27
ARM h13	905	376	−58	540	−40
ARM h14	909	367	−60	533	−41
ARM h15	888	359	−60	526	−41
ARM h16	826	358	−57	525	−37
ARM h17	823	354	−57	521	−37
ARM h18	744	326	−56	502	−32
LASSO h12.0	392	219	−44	511	30
LASSO h13.0	560	330	−41	517	−8
LASSO h14.0	743	345	−54	515	−31
LASSO h15.0	639	324	−49	504	−21
LASSO h16.0	504	306	−39	494	−2

3.4. Robustness for Small Sample Sizes

As was already suggested in Section 3.1, the perceived cloud statistics are a strong function of sample size, especially for 1-dimensional observations. Meanwhile, we have developed our transfer function based on an infinite number of chord lengths. Therefore, how will the transfer function hold up if only a few chord lengths are observed? To test this, we take a random sample of 20 chords from the full set of chord lengths, which is the equivalent of a 25 km transect through the LES domain, with similar mean and variance in cloud cover and size as the last column in Figure 3b,d. We then apply our transfer function on these 20 chord lengths, and compare it with the actual linear cloud size of Figure 8b. To get an idea of the robustness of our transfer function, we take 20 different transects of chord lengths, and do the same exercise for different equivalent transect lengths ranging from 6.25 km to 200 km, or from five chord lengths per transect to 160 chord lengths. The results are shown in Figure 9. From this figure we see that while there is a clear deviation from the linear cloud size for the smallest sample sizes (cloud sizes under 200 m), the mean of the 20 transects (after the application of the transfer function) is close to the linear cloud size of the overall field. Of course, one cannot expect a 6.25 km long individual sample to be close to the correct value, but the variability between samples becomes less for transect lengths of 25–50 km, first for the smaller sizes, and later for the largest ones.
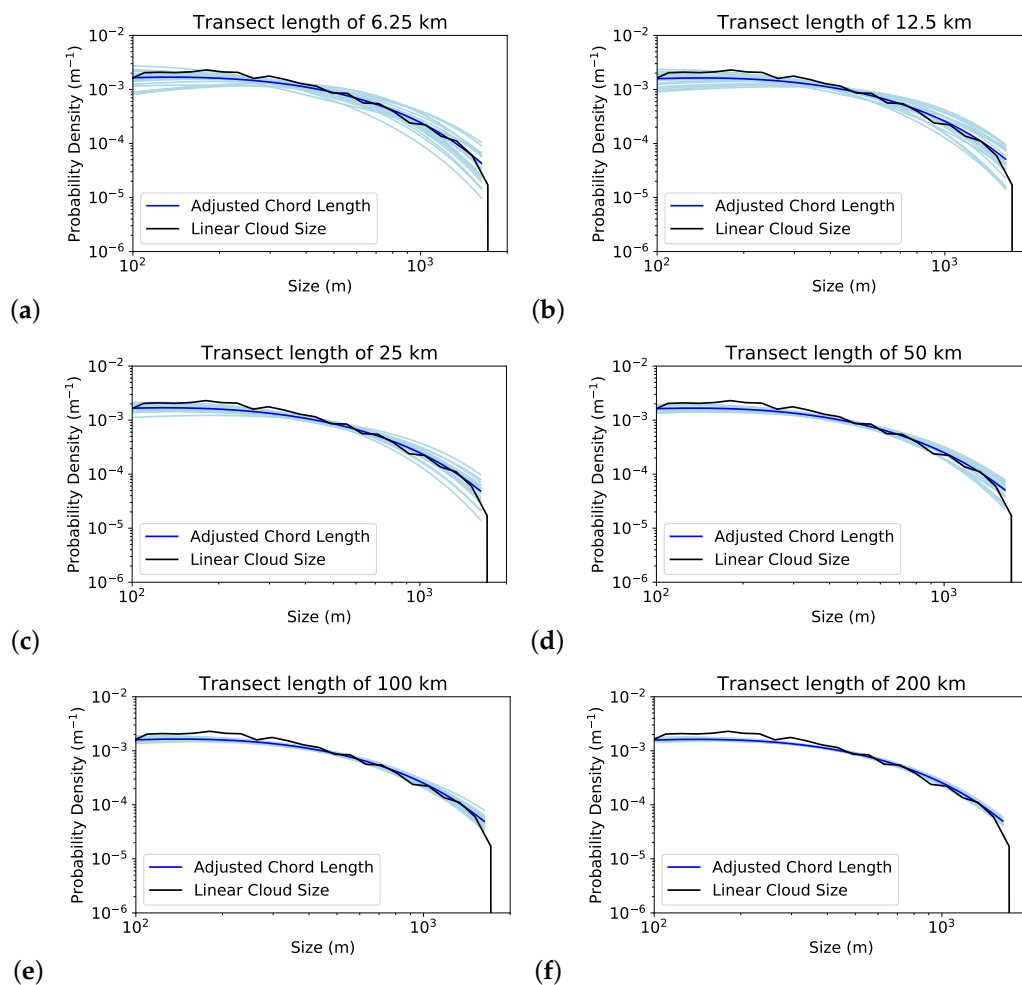
Figure 9. Reconstructed linear cloud size distributions for different transect lengths, with the linear cloud size (black), the mean adjusted chord length distribution (blue), and the adjusted chord length distribution of individual transects (light blue) for transects of (**a**) 6.25 km, (**b**) 12.5 km, (**c**) 25 km, (**d**) 50 km, (**e**) 100 km, and (**f**) 200 km.

4. Conclusions

In this paper, we developed a transfer function to convert the chord length distribution to the cloud linear size distribution. Based on empirical LES data, we see that the cloud size distribution, conditional on the chord length, is a gamma distribution, which is usually closely related to the actual cloud size. This means that the RV17 method of retrieving the typical cloud size by using a calibrated version of the typical chord length is usually sound, with two caveats: (1) chord lengths of less than 100 m can come from any cloud, and yield little to no information about the cloud size distribution, and (2) to retrieve the entire distribution, the spread of clouds that can produce a particular chord length has to be taken into account.

Of course, many practical applications would already remove the smallest chord lengths from their sample [30]; a 100 m chord length is approximately equal to a 10 s measurement window for radar, or less than 1 s for an airplane. This paper gives some theoretical foundation for that practice. However, to retrieve a cloud cover that is converging to the correct value, let alone a area-average cloud size, one needs a significant amount of data. In a swiftly changing field of cumulus such as a simple diurnal cycle, a single vertically pointing instrument will not be sufficient for this, and a network of instruments is needed instead. Alternatively, spatial approaches like satellites, scanning radar, stereo cameras [31], and Total Sky Imagery [32] seem more suitable alternatives for retrieving the typical cloud size.

Our current study still has some limitations: It is highly likely that we are underestimating the number of small chord lengths, as our simulations were limited by their grid spacing of 25 m. However, even in our simulations, the bimodal distribution speaks to the multifractal nature of these clouds. Finally, our model is not properly aware of the scale break and the maximum size of the clouds. This is a topic of ongoing research for the authors.

References

1. Bony, S.; Stevens, B.; Frierson, D.M.W.; Jakob, C.; Kageyama, M.; Pincus, R.; Shepherd, T.G.; Sherwood, S.C.; Siebesma, A.P.; Sobel, A.H.; et al. Clouds, circulation and climate sensitivity. *Nat. Geosci.* **2015**, *8*, 261–268. [CrossRef]

2. Wyngaard, J. Toward numerical modeling in the "Terra Incognita". *J. Atmos. Sci.* **2004**, *61*, 1816–1826. [CrossRef]

3. Honnert, R.; Masson, V.; Couvreux, F. A Diagnostic for Evaluating the Representation of Turbulence in Atmospheric Models at the Kilometric Scale. *J. Atmos. Sci.* **2011**, *68*, 3112–3131. [CrossRef]

4. Neggers, R.A.J. Exploring bin-macrophysics models for moist convective transport and clouds. *J. Adv. Model. Earth Syst.* **2015**, *7*, 2079–2104. [CrossRef]

5. Sakradzija, M.; Seifert, A.; Heus, T. Fluctuations in a quasi-stationary shallow cumulus cloud ensemble. *Nonlinear Process. Geophys.* **2015**, *22*, 65–85. [CrossRef]

6. Plant, R.S.; Craig, G.C. A Stochastic Parameterization for Deep Convection Based on Equilibrium Statistics. *J. Atmos. Sci.* **2008**, *65*, 87–105. [CrossRef]

7. Rodts, S.M.A.; Duynkerke, P.G.; Jonker, H.J.J. Size distributions and dynamical properties of shallow cumulus clouds from aircraft observations and satellite data. *J. Atmos. Sci.* **2003**, *60*, 1895–1912. [CrossRef]

8. Wood, R.; Field, P.R. The Distribution of Cloud Horizontal Sizes. *J. Clim.* **2011**, *24*, 4800–4816. [CrossRef]

9. Trivej, P.; Stevens, B. The Echo Size Distribution of Precipitating Shallow Cumuli. *J. Atmos. Sci.* **2010**, *67*, 788–804. [CrossRef]

10. Neggers, R.A.J.; Jonker, H.J.J.; Siebesma, A.P. Size statistics of cumulus cloud populations in large-eddy simulations. *J. Atmos. Sci.* **2003**, *60*, 1060–1074. [CrossRef]

11. Van Laar, T.W.; Schemann, V.; Neggers, R.A.J. Investigating the Diurnal Evolution of the Cloud Size Distribution of Continental Cumulus Convection Using Multiday LES. *J. Atmos. Sci.* **2019**, *76*, 729–747. [CrossRef]

12. Chandra, A.S.; Kollias, P.; Albrecht, B.A. Multiyear Summertime Observations of Daytime Fair-Weather Cumuli at the ARM Southern Great Plains Facility. *J. Clim.* **2013**, *26*, 10031–10050. [CrossRef]

13. Sulak, A.; Calabrase, W.; Ryan, S.D.; Heus, T. The Contributions of Shear and Turbulence to Cloud Overlap for Cumulus Clouds. *J. Geophys. Res. Atmos.* **2020**, *125*, e2019JD032017. [CrossRef]

14. Griewank, P.J.; Heus, T.; Lareau, N.P.; Neggers, R.A.J. Size-dependence in chord characteristics from simulated and observed continental shallow cumulus. *Atmos. Chem. Phys. Discuss.* **2020**, *2020*, 1–32. [CrossRef]

15. Siebesma, A.P.; Jonker, H.J.J. Anomalous scaling of cumulus cloud boundaries. *Phys. Rev. Lett.* **2000**, *85*, 214–217. [CrossRef] [PubMed]

16. Romps, D.M.; Vogelmann, A.M. Methods for Estimating 2D Cloud Size Distributions from 1D Observations. *J. Atmos. Sci.* **2017**, *74*, 3405–3417. [CrossRef]

17. Yau, M.K.; Rogers, R.R. An Inversion Problem on Inferring the Size Distribution of Precipitation Areas from Raingage Measurements. *J. Atmos. Sci.* **1984**, *41*, 439–448. [CrossRef]

18. Neggers, R.A.J.; Duynkerke, P.G.; Rodts, S.M.A. Shallow cumulus convection: A validation of large-eddy simulation against aircraft and Landsat observations. *Q. J. R. Meteorol. Soc.* **2003**, *129*, 2671–2696. [CrossRef]

19. Van Heerwaarden, C.C.; van Stratum, B.J.H.; Heus, T.; Gibbs, J.A.; Fedorovich, E.; Mellado, J.P. MicroHH 1.0: A computational fluid dynamics code for direct numerical simulation and large-eddy simulation of atmospheric boundary layer flows. *Geosci. Model Dev.* **2017**, *10*, 3145–3165. [CrossRef]

20. Siebesma, A.P.; Bretherton, C.S.; Brown, A.; Chlond, A.; Cuxart, J.; Duynkerke, P.G.; Jiang, H.L.; Khairoutdinov, M.; Lewellen, D.; Moeng, C.H.; et al. A large eddy simulation intercomparison study of shallow cumulus convection. *J. Atmos. Sci.* **2003**, *60*, 1201–1219. [CrossRef]

21. Brown, A.R.; Cederwall, R.T.; Chlond, A.; Duynkerke, P.G.; Golaz, J.C.; Khairoutdinov, M.; Lewellen, D.C.; Lock, A.P.; MacVean, M.K.; Moeng, C.H.; et al. Large-eddy simulation of the diurnal cycle of shallow cumulus convection over land. *Q. J. R. Meteorol. Soc.* **2002**, *128*, 1075–1093. [CrossRef]

22. VanZanten, M.C.; Stevens, B.; Nuijens, L.; Siebesma, A.P.; Ackerman, A.S.; Burnet, F.; Cheng, A.; Couvreux, F.; Jiang, H.; Khairoutdinov, M.; et al. Controls on precipitation and cloudiness in simulations of trade-wind cumulus as observed during RICO. *J. Adv. Model. Earth Syst.* **2011**, *3*, M06001. [CrossRef]

23. Seifert, A.; Heus, T. Large-eddy simulation of organized precipitating trade wind cumulus clouds. *Atmos. Chem. Phys.* **2013**, *13*, 5631–5645. [CrossRef]

24. Seifert, A. On the shape-slope relation of drop size distributions in convective rain. *J. Appl. Meteorol.* **2005**, *44*, 1146–1151. [CrossRef]

25. Gustafson, W.I., Jr.; Vogelmann, A.M.; Li, Z.; Cheng, X.; Dumas, K.K.; Endo, S.; Johnson, K.L.; Krishna, B.; Fairless, T.; Xiao, H. The Large-Eddy Simulation (LES) Atmospheric Radiation Measurement (ARM) Symbiotic Simulation and Observation (LASSO) Activity for Continental Shallow Convection. *Bull. Am. Meteorol. Soc.* **2020**, *101*, E462–E479. [CrossRef]

26. Neggers, R.A.J.; Griewank, P.J.; Heus, T. Power-Law Scaling in the Internal Variability of Cumulus Cloud Size Distributions due to Subsampling and Spatial Organization. *J. Atmos. Sci.* **2019**, *76*, 1489–1503. [CrossRef]

27. Oue, M.; Tatarevic, A.; Kollias, P.; Wang, D.; Yu, K.; Vogelmann, A.M. The Cloud-resolving model Radar SIMulator (CR-SIM) Version 3.3: Description and applications of a virtual observatory. *Geosci. Model Dev.* **2020**, *13*, 1975–1998. [CrossRef]

28. Oue, M.; Kollias, P.; North, K.W.; Tatarevic, A.; Endo, S.; Vogelmann, A.M.; Gustafson, W.I., Jr. Estimation of cloud fraction profile in shallow convection using a scanning cloud radar. *Geophys. Res. Lett.* **2016**, *43*, 10,998–11,006. [CrossRef]

29. Lamer, K.; Puigdomènech Treserras, B.; Zhu, Z.; Isom, B.; Bharadwaj, N.; Kollias, P. Characterization of shallow oceanic precipitation using profiling and scanning radar observations at the Eastern North Atlantic ARM observatory. *Atmos. Meas. Tech.* **2019**, *12*, 4931–4947. [CrossRef]

30. Heus, T.; Pols, C.F.J.; Jonker, H.J.J.; Van den Akker, H.E.A.; Lenschow, D.H. Observational validation of the compensating mass flux through the shell around cumulus clouds. *Q. J. R. Meteorol. Soc.* **2009**, *135*, 101–112. [CrossRef]

31. Romps, D.M.; Öktem, R. Observing Clouds in 4D with Multiview Stereophotogrammetry. *Bull. Am. Meteorol. Soc.* **2019**, *99*, 2575–2586, [CrossRef]

32. Kleiss, J.M.; Riley, E.A.; Long, C.N.; Riihimaki, L.D.; Berg, L.K.; Morris, V.R.; Kassianov, E. Cloud Area Distributions of Shallow Cumuli: A New Method for Ground-Based Images. *Atmosphere* **2018**, *9*, 258. [CrossRef]