Investigating the Diurnal Evolution of the Cloud Size Distribution of Continental Cumulus Convection Using Multiday LES

THIRZA W. VAN LAAR, VERA SCHEMANN, AND ROEL A. J. NEGGERS

University of Cologne, Cologne, Germany

(Manuscript received 14 March 2018, in final form 11 December 2018)

ABSTRACT

The diurnal dependence of cumulus cloud size distributions over land is investigated by means of an ensemble of large-eddy simulations (LESs). A total of 146 days of transient continental shallow cumulus are selected and simulated, reflecting a low midday maximum of total cloud cover, weak synoptic forcing, and the absence of strong surface precipitation. The LESs are semi-idealized, forced by large-scale model output but using an interactive surface. This multitude of cases covers a large parameter space of environmental conditions, which is necessary for identifying any diurnal dependencies in cloud size distributions. A power-law exponential function is found to describe the shape of the cloud size distributions for these days well, with the exponential component capturing the departure from power-law scaling at the larger cloud sizes. To assess what controls the largest cloud size in the distribution, the correlation coefficients between the maximum cloud size and various candidate variables reflecting the boundary layer state are computed. The strongest correlation is found between total cloud cover and maximum cloud size. Studying the size density of the cloud area revealed that larger clouds contribute most to a larger total cloud cover, and not the smaller ones. Besides cloud cover, cloud-base and cloud-top height are also found to weakly correlate with the maximum cloud size, suggesting that the classic idea of deeper boundary layers accommodating larger convective thermals still holds for shallow cumulus. Sensitivity tests reveal that the results are only minimally affected by the representation of microphysics and the output resolution.

1. Introduction

Shallow cumulus clouds play an important role in Earth's weather and climate system. Both the mean amplitude and the variability of cloud cover have a significant impact on Earth's radiation budget (Cahalan et al. 1994). This requires a correct representation of the spatial structure of a shallow cumulus cloud field in climate and weather prediction models, in order to adequately simulate the radiative fluxes.

A complicating factor is that shallow cumulus clouds need to be parameterized because of their small and highly variable temporal and spatial scales. Various recent studies have identified such cumulus parameterizations to be at the heart of problems in both numerical weather prediction and climate simulations (Bony and Dufresne 2005; Neggers and Siebesma 2013; Vial et al. 2016). An active area of research that addresses a part of this problem focuses on the feedback of shallow cumulus

clouds to climate perturbations (Zhang et al. 2013; Brient et al. 2015; Dal Gesso et al. 2015). Another problem for cumulus parameterization is the increasing resolution of the climate and NWP models and thereby the approach of the gray zone (Wyngaard 2004). This means that parameterization schemes have to become sensitive to the resolution used in the model and be scale adaptive. Moreover, it also means that our understanding of the spatial structure and diurnal cycles of cumulus cloud fields has to be improved. This area of interest is therefore actively being researched at the moment (Arakawa and Wu 2013; Dorrestijn et al. 2013; Kwon and Hong 2017; Honnert 2016).

In essence, making a cumulus scheme scale aware means that size information somehow has to be included. This has recently motivated researchers to revisit the approach of formulating models in terms of cloud size distributions (CSDs), following Arakawa and Schubert (1974). Recent studies with CSD-based schemes by Wagner and Graf (2010), Park (2014), Neggers (2015), and Brast et al. (2018) report promising skill in reproducing scale adaptivity; however, closure of the CSD

Corresponding author: Thirza W. van Laar, vanlaar@meteo.uni-koeln.de

is still needed. This closure is still an open research question, and needs to be informed by reliable statistics on cloud sizes and their dependence on meteorological conditions.

Cloud size distributions for shallow convection have been investigated in numerous previous studies, using both observational and model data. Some older studies, like [Plank \(1969\)](#) and [Wielicki and Welch \(1986\)](#), report an exponential distribution, whereas later work commonly describes the functional form of the cloud size distribution with a power law:

$$N(l) = al^b, \quad (1)$$

with l the cloud size and a and b fitting constants. A power law for cloud size distributions has been applied in observational studies like [Benner and Curry \(1998\)](#) and [Zhao and Di Girolamo \(2007\)](#) as well as LES studies like [Neggers et al. \(2003\)](#) and [Heus and Seifert \(2013\)](#). [Jiang et al. \(2008\)](#) used a power law as well to describe cloud size properties and they showed a good comparison between the observations and model data. Recently, [Feingold et al. \(2017\)](#) used a power-law distribution for their study on the relation between albedo and cloud cover. The power-law exponent b from Eq. (1) describes the slope of the distribution. Typical values that have been found are between -1.7 and -2.5 ([Rieck et al. 2014](#)).

Many observational and modeling studies report the existence of a scale break in the power law [recently, e.g., [Trivej and Stevens \(2010\)](#) and [Heus and Seifert \(2013\)](#)], meaning that the pure power-law fit is only valid for the small cloud sizes. For clouds larger than the scale break the slope of the power law needs to be adjusted to fit the data, resulting in a double power law. How to determine the location of this scale break in a methodical way, as well as the underlying cause for this break in scaling, is still actively debated in the literature. [Benner and Curry \(1998\)](#) mention a break in the fractal dimension at a location similar to the scale break in their CSD for which the underlying reason could be a maximum size of individual convective cells ([Joseph and Cahalan 1990](#)), while [Wood and Field \(2011\)](#) hypothesize that the Rossby radius controls the characteristic cloud sizes and thereby the scale break. A third possible cause for the scale break is insufficient statistics; the sampling size is too small to capture the full distribution, in particular at the largest sizes, which occur least frequently. Also the domain size and resolution might influence the location of the scale break in a simulated cloud field.

Since it is hard to estimate the position of the scale break in a standardized way and its background is unclear, we follow [Windmiller \(2017\)](#) and [Peters et al. \(2009\)](#) in

their approach of applying a power-law exponential function to the CSD. The power-law exponential fit is flexible enough to capture many shapes, from a pure power law to a more curved fit. By applying this fit, we avoid having to explicitly deal with the scale break. The power-law exponential function is based on percolation theory and is defined as ([Ding et al. 2014](#))

$$N(l) = al^b \exp(cl), \quad (2)$$

with c representing another fitting constant. Both b and c are expected to be smaller than or equal to 0 and together they reflect the shape of the distribution. With decreasing c the distribution follows less of a power law and is increasingly dominated by the exponential, especially at the large cloud sizes.

While the shape is one defining aspect of the CSD, its range in terms of size is another. An early study by [Joseph and Cahalan \(1990\)](#) suggests that the maximum cloud size scales with the depth of the boundary layer. More recently, [Rieck et al. \(2014\)](#) confirmed this dependence, but also reported that surface heterogeneity plays a role. [Dawe and Austin \(2012\)](#) found a strong relationship between the cloud area and the eventual height reached by the cloud. [Sakradzija and Hohenegger \(2017\)](#) showed that the shape of the distribution of cloud-base mass flux, which is closely related to the cloud size distribution, is determined by the Bowen ratio at the surface. While these new insights are encouraging, many of these studies were limited to single cases, or even single snapshots. Such snapshots do not contain any information on the diurnal evolution of the maximum cumulus cloud size. In addition, the use of single cases limits the statistical significance of the obtained results. To assess the robustness of the results and to obtain information on the diurnal signal of the maximum cloud size, one would need a database of many different cases covering a broad parameter space of large-scale conditions.

For this study, a library with 146 individual cumulus days is created to investigate the behavior of the size distribution during diurnal cycles of cumulus over land. Continental shallow cumulus cases are chosen because of their transient nature, with a strong temporal evolution in the amount of clouds, their elevation and depth, and the size distribution. Using many different days over the course of 5 years enables us to do statistical studies on cloud size controlling factors on a diurnal time scale. For all the selected days a 24-h large-eddy simulation (LES) is performed, and a clustering algorithm is applied to compute the cloud size distribution and study both its shape and the maximum cloud size. Correlations between the maximum cloud size and a set of candidate variables are calculated to establish the role of these

variables in controlling the upper limit of the distribution of cloud sizes. Sensitivity tests are performed with a subset of cases on the output resolution and the microphysics representation.

The configuration of the LES experiments and the case selection is described in [section 2](#). In [section 3](#) the clustering algorithm and the derivation of the cloud size statistics are explained, followed by the results in [section 4](#) and [5](#). Finally, discussion and conclusions can be found in [sections 6](#) and [7](#).

2. Multiday LES

The derivation of statistically significant size distributions requires 3D cloud fields that contain a sufficient number of clouds and are available at a frequency high enough to resolve any diurnal signal. While the latest scanning radar strategies show promising capability in detecting CSDs in nature ([Borque et al. 2014](#)), this technique is still in its infancy and needs to be fully explored. We therefore still rely on LESs to complement observational data with 3D fields at high frequencies ([Neggers et al. 2012](#); [Schalkwijk et al. 2015](#); [Zhang et al. 2017](#)). For this study, a daily LES as applied at the Jülich Observatory for Cloud Evolution (JOYCE) meteorological supersite in Germany will be used.

a. JOYCE

The observational JOYCE supersite ([Löhnert et al. 2015](#)) is a continental midlatitude site, well suited to studying diurnal cycles. It is equipped with a suite of state-of-the-art cloud detection instrumentation that is operational on the long term. Midlatitude cumulus clouds are regularly present at JOYCE during the summer, since the site is close enough to the sea to ensure that low-level humidity is frequently high enough to allow daytime cumulus formation. The abundance of long-term observations allows for confronting LES with relevant observational data; this is, however, not the focus of this study and is considered a future research topic. The measurements taken at JOYCE will only be used for the selection of shallow cumulus days and a short comparison of the measured and modeled cloud cover.

b. Cumulus day selection

Since the focus of this research is on boundary layer cumulus clouds, days that show this cloud regime have been selected to be simulated. Several specific criteria are applied to visualized reflectivity of lidar and radar observations taken at JOYCE. Selected days do not show any precipitation and no significant synoptic-scale activity during the daytime. They do have a significant

period of cumulus convection around noon and an increase in the LCL during the day. Cloud cover is small and the cumulus events are isolated, which means that they are not influenced by synoptic events at the beginning and ending of the day. These criteria are applied during the summer months of the years 2012–16, and during the winter months the conditions are usually too cold for shallow cumulus to form. The result is a selection of 146 days in total. How these days are distributed over the months and years can be seen in [Fig. 1](#). The resulting library of cases provides us with a parameter space of considerable width. By simulating all 146 days individually, and not simulating a single composite like [Zhang et al. \(2017\)](#), the internal variability of the ensemble is resolved. This provides additional information on the robustness of any diurnal signal in the CSD properties, as well as its spread for the environmental conditions covered by the library of cases. Since we are interested in the diurnal cycle, only the daytime hours between 9 and 17 h are used for further analysis.

c. DALES

The Dutch Atmospheric Large-Eddy Simulation (DALES; [Heus et al. \(2010\)](#)) model is used for the simulation of the selected days. DALES has been employed in many recent intercomparison studies on shallow cumulus convection (e.g., [van Zanten et al. 2011](#); [van der Dussen et al. 2013](#)), which document the model's skill in simulating this cloud regime. The land surface parameterization scheme is described in [Heus et al. \(2010\)](#) and it is similar to the European Centre for Medium-Range Weather Forecasts (ECMWF) surface scheme. In the simulations, the state of the land surface is evolving with time-varying temperature and soil moisture data. As a result, the surface energy fluxes are also time varying, calculated interactively through the surface energy budget. For the calculation of the vertical transfer of radiative energy Monte Carlo spectral integration is used, as described in detail by [Heus et al. \(2010\)](#). The resolved turbulent domain reaches up to 5 km. Above this height, the vertical profile of the thermodynamic and cloudy state of the IFS analysis is used to calculate the downward transfer of radiative energy into the turbulent domain, by which the impact of clouds overhead are accounted for. In addition, a climatological vertical profile of ozone is provided to the radiation scheme. The domain size of the simulations is $12.8 \text{ km} \times 12.8 \text{ km} \times 5 \text{ km}$, with a resolution of 50 m in the horizontal and 40 m in the vertical. Simulations start at 0000 LT (LT leads UTC by 2 h during the summer in Germany) and cover one full day (24 h). Every 15 min, every fourth point of the three-dimensional field of all required model variables is saved for further analysis.

FIG. 1. Histogram showing how the 146 days are spread over the months and years.

This subsampling is done in order to keep the amount of data manageable. The impact of the reduced data output on the CSD will be assessed (as reported in the [appendix](#)), showing only marginal dependence, which supports taking this approach. [Zhang et al. \(2017\)](#) report a dependence of the total cloud cover on the featured microphysics scheme. In the simulations the two-moment parameterization scheme of [Seifert and Beheng \(2004\)](#) is used, but its dependence on the results is tested. A sensitivity study is performed by using some additional simulations with a simple nonprecipitating all-or-nothing cloud scheme by [Sommeria \(1976\)](#) that are done for the years 2013 and 2014.

d. Initialization, boundary conditions, and large-scale forcing

The 24-h-long simulations are initialized and driven by time-varying boundary conditions and large-scale forcings derived from the Integrated Forecasting System (IFS) of ECMWF. The method as applied here to drive the LES model was first described by [Neggers et al. \(2012\)](#), and was recently used in slightly modified form in the study by [Gesso and Neggers \(2018\)](#). All fields provided to the LES as input are based on two analysis fields per day, at 0000 and 1200 UTC, supplemented by short-range forecast fields to cover the intermediate time points at 3, 6, and 9 h. This effectively yields a forcing dataset covering 24 h at 3-h time resolution, which is assumed sufficient to resolve diurnal and synoptic signals in the large-scale forcing.

The initialization of the height-dependent atmospheric state variables concerns the wind components u and v ; the liquid water potential temperature θ_l ; the total water specific humidity q_t ; the cloud state variables liquid and ice condensate q_l and q_i , respectively; and ozone (for the radiation scheme). Also initialized are the soil temperature and moisture, with the latter scaled to preserve the fraction between the field capacity and wilting point ([Neggers et al. 2012](#)). After initialization, all these variables are allowed to evolve freely, apart from ozone, which is not changing with time.

The prescribed boundary conditions at the surface include the surface geopotential, the albedo, the roughness lengths for heat and momentum, the leaf area index (LAI), and the near-surface air pressure. For both the LAI and the roughness lengths the IFS values at this location are extracted from the ECMWF's meteorological archive (MARS; [ECMWF 2017](#)). Only the near-surface air pressure changes with time, determining the geometric height of the pressure levels in the profile above. The soil temperature and moisture can evolve freely but interact with the atmosphere, through the surface energy and humidity budgets. In this respect the LES experiments in this study differ from those in other recent LES studies at supersites, which typically use a fully prescribed soil state ([Heinze et al. 2016](#); [Gustafson et al. 2017](#)).

The impact of large-scale weather systems, which can be significant at midlatitude locations, is represented by means of prescribed advective forcing tendencies in the budget equations of the four atmospheric state variables $\{q_t, \theta_l, u, v\}$. This forcing is both height and time dependent, and is also continuous, being interpolated linearly between the 3-hourly IFS time points. IFS data at pressure levels are used to this purpose, in order to avoid problems with advective calculations in areas of steep orography that arise when terrain-following coordinates are used ([Simmons and Burridge 1981](#); [Simmons 1986](#)). Orography is not included in the simulation, which is justified by the fact that the direct vicinity of the site is relatively flat. The horizontal tendencies are calculated from the wind components and the local gradients on the $0.1^\circ \times 0.1^\circ$ operational IFS grid, averaged over a $0.5^\circ \times 0.5^\circ$ horizontal area around the site of interest to obtain smoothly varying forcing fields. The vertical component of large-scale advection is calculated interactively from a prescribed subsidence profile that acts on the vertical gradients in the model state variables. Momentum is forced using a prescribed large-scale horizontal pressure gradient, appearing in the resolved Navier–Stokes equations as a prescribed geostrophic wind, and by accounting for the Coriolis force. The momentum equations take the form as discussed in any

FIG. 2. A 5-yr climatology of large-scale forcing at 500-m height at the JOYCE site for the months of April–September during the period of 2012–16, as derived from IFS data. (a) The probability density function of wind speed, and time averages of the (b) pressure velocity Ω , (c) temperature advection, and (d) humidity advection.

standard academic textbook on atmospheric dynamics, for example, [Holton and Hakim \(2012\)](#).

Finally, following [Neggers et al. \(2012\)](#) the profiles of $\{q_t, \theta_t, u, v\}$ are continuously nudged to the IFS-derived state at a synoptic time scale of 6 h. Relaxation at this synoptic time scale has been proven tight enough to effectively limit excessive model drift, but still loose enough to allow the turbulence to act freely.

Figure 2 gives some insight into the typical summertime large-scale forcing at the JOYCE site, derived from the IFS fields for the period 2012–16. The focus is on a height of 500 m that is normally inside the daytime boundary layer. Some features are visible that can be considered typical for midlatitude near-coastal sites in northwestern Europe, making it different from, for example, the ARM SGP site ([Zhang et al. 2017](#)). First, stronger wind speeds are predominantly from the south-southwest, a clear fingerprint of the polar jet stream and its quasigeostrophic forcing in this area during summer.

Second, there is a rough east–west divide in the sign of the temperature and humidity advection (Figs. 2c,d), with westerlies associated with moistening and cooling and easterlies with warming and drying. This is explained by the proximity of the ocean in the west, which tends to cool and moisten the coastal regions during summertime. In general, the amplitude of the advection increases with wind speed, as can be expected from its definition. Interestingly, the large-scale subsidence roughly correlates with the advective tendencies, with strong westerlies being associated with stronger subsidence, and easterlies with lifting. The latter could partially reflect the cases of convective instability in the summertime associated with low surface pressure and strong convergence.

3. Methods

Three-dimensional instantaneous snapshots of liquid water and buoyancy from the LES serve as input for the

derivation of the CSDs. Alternatively, one can track individual clouds through their life cycle and thereby calculate their size (Heus and Seifert 2013). The choice to use snapshots instead is motivated by two reasons. First, the typical lifetime of a single cumulus cloud is much shorter than the diurnal signal. Second, using snapshots significantly simplifies the analysis, which is essential for processing the 146 daily simulations with multidimensional high-frequency output.

a. Cloud definitions

The starting point for defining the clouds is to define a single LES grid box as either cloudy or noncloudy. Two different criteria for cloudiness are used: 1) the liquid water content is greater than zero or 2) the liquid water content is greater than zero and the buoyancy is positive as well. The combined criteria of condensate presence and positive buoyancy is often referred to as the *cloud core* (Siebesma and Cuijpers 1995). This option is added because it is suggested in the literature that clouds with and without positive buoyancy have different dynamics and a different vertical extent and should therefore be considered separately (Zhang and Klein 2013). Unless otherwise indicated, in this study, only the liquid water content of a grid box is used as a criterion for cloud presence.

b. Clustering algorithm and CSD calculation

Details of the computation of the cloud sizes and their distribution can be found in Neggers et al. (2003), but the basics will be described here. All horizontally and vertically neighboring cloudy grid boxes are clustered into clouds and their size is calculated. Cloud size is defined as the square root of the covered horizontal area, where this area is the number of cloudy grid cells times the area per grid cell. There are two ways to determine the horizontal area covered by the cloud, either as the projected area, or as the horizontal area averaged over the depth of the cloud [for mathematical definitions, see Neggers et al. (2003)]. By using the projected area, it is assumed that cloud sizes do not vary much with height and that horizontal wind does not influence the aspect ratio. Cloud statistics based on the projected area are more suitable for comparison with satellite observations. Therefore, the projected cloud cover will be the default definition in this study. All sizes of the population in a single snapshot are sorted into a histogram, which is a discretized form of a CSD. The maximum cloud size $l_{\max}$, which is used later, is defined as the biggest cloud of the field.

Because of the subsampling of the original LES output, the resolution of the model output fields is 4 times smaller than the simulation resolution of 50 m.

The derived cloud sizes are therefore organized in bins of 200 m. The amount of clouds per bin is divided by the total number of clouds in the field, resulting in the normalized CSD, denoted as $N^*(l)$. The computed CSDs with a frequency of 15 min are averaged per hour. This yields a time frequency at which the diurnal cycle can be considered well resolved, while at the same time the statistical reliability is increased.

4. Results: Exploring the variability

The 146 LESs produce a large amount of data with considerable variability. First, some average properties and individual days will be studied. The findings based on a subset of the data will guide a more thorough analysis using the complete dataset in the next section.

a. Averaged mean state

In the average cloud fraction of all simulated days (Fig. 3a) some typical features in the vertical structure for diurnal cycles of shallow cumulus over land can be recognized and it shows many similarities with, for example, Zhang et al. (2017). First, during the day the well-defined cloud base rises in a quasi-linear way. Second, the layer with high-cloud fractions deepens gradually, reaching a depth of about 500–1000 m. Third, the averaged cloud cover does not exceed 10%, which is typical of fair-weather cumulus (Nuijens et al. 2015). These results suggest that the ensemble of selected cases matches the prototype behavior of continental shallow cumulus as reported in previous studies (Brown et al. 2002; Zhang et al. 2017).

However, there are some features visible in Fig. 3 that are different from previously mentioned studies, but typical for our specific case of midlatitude continental shallow cumulus at a site in relative proximity to the coast. Due to morning fog there is a low cloud cover close to the ground before sunrise. With time this fog breaks up, resulting in low lifting condensation levels during the early morning. There is also significant cloud cover during the night; this probably points to some synoptic activity.

Similar features as for cloud fraction can be seen for the averaged relative humidity (Fig. 3b). There is a relatively high humidity in the cloud layer, starting low at the surface in the early morning hours because of fog. The cloud layer gradually grows deeper during the day. The lowest levels of humidity at the surface are found at midday. During these hours the moisture is effectively transported upward, a prerequisite for cloud formation at greater heights. Besides this, drier air from the free troposphere is transported to the surface by effective entrainment at the inversion height (Betts 2009).

FIG. 3. The simulated (top) cloud cover and (bottom) relative humidity, averaged over all the simulated days.

b. Cloud cover comparison

Cloud cover is an important variable in this study. Therefore, despite a comparison of model and observational data not being a primary goal of this study, the modeled cloud cover will be shortly evaluated using ceilometer data. For all of the 146 selected cumulus days the cloud cover up to 3 km is determined from the ceilometer observations and from the LES output. The median and upper and lower quartiles of both the ceilometer data and the simulated cloud cover as a function of time are shown in Fig. 4. The spread in the ceilometer data is considerably larger than for the model data and the median cloud cover during the day (except for the last hour) is 5%–10% smaller for the model data. Both observations and simulations do however show the typical diurnal cycle for shallow cumulus of an increased cloud cover in the morning and decreased cloud cover in the afternoon.

c. Individual cumulus days

There is much variability among the days that establish the averaged behavior that is shown in Fig. 3. Five random individual days are selected to showcase this variability, each with a different cloud cover around noon (Fig. 5). The 2 days with the lowest cloud covers (28 May 2012 and 23 June 2016) both show a thin cloud layer around noon. The main difference between these 2 days is the development of the depth of the cloud layer. For 28 May 2012 this happens gradually, whereas for 23 June 2016 there is already a peak in cloud depth and cloud cover during the early morning. On 29 July 2012

FIG. 4. The diurnal evolution of the median (solid lines) and top and bottom quartiles (shading) of cloud cover as observed by a ceilometer (gray) and cloud cover as modeled by DALES (dark red).

the cloud layer stays shallow during the first few hours of the day, but around noon it suddenly deepens. On 9 September 2015 there is a low LCL in the morning; this might be because of the synoptic activity during the night. There are also high cloud fractions at the top of the thick cloud layer. This suggests the presence of a layer just below an inversion with some more stratiform, capping outflow. On 18 April 2015 there is an increasing cloud base during the day and a quite shallow cloud layer, but nevertheless high-cloud fractions between 40% and 50%.

d. Cloud size distributions

The considerable variation in the cloud structure visible in Fig. 5 is also reflected in the shape and range of the associated CSDs taken at 1300 LT (Fig. 6). The dots in the figure represent the discretized distribution and the lines in the corresponding colors the fit, following the power-law exponential function from Eq. (2). All details, including the constants of the fit and the maximum cloud size of the distributions are given in Table 1. Judging by eye, the power-law exponential function seems to fit the data well, since for most days shown in the figure it captures the power law part, as well as the tail of the distribution. Three features stand out in these CSDs. First, an increasing projected cloud cover largely coincides with an increasing simulated maximum cloud size $l_{\max}$. Second, the power-law scaling is visible for the smaller cloud sizes; for greater cloud cover, it is also visible for the larger cloud sizes. The small values for c in Table 1 indicate that for greater cloud cover, the power law dominates across all scales. Third, the data show more noise at larger cloud sizes, especially for greater cloud cover (shown in red). We speculate that the reason behind this is the undersampling of the least frequently occurring larger cloud sizes. This suggests that statistics

FIG. 5. The cloud fraction for five simulated days, all with a different maximum cloud cover at 1300 LT. The dashed line represents the highest cloud top, and the solid line the lowest cloud base.

need to be improved to obtain convergence. In this study the undersampling at larger sizes is addressed by considering many more days, since due to computational feasibility considerations the domain size and sampling frequency are limited.

As an example of the diurnal evolution, the CSDs for 29 July 2012 (Fig. 5c) are shown in Fig. 7. The values for b , c , cloud cover (cc), and $l_{\max}$ are shown in Table 2. The diurnal cycle of $l_{\max}$ shows a peak at midday, and

FIG. 6. The CSDs at 1300 LT for five different days (see also Table 1). Each color represents a different day with a different cloud cover.

a similar pattern is present for the cloud cover. The fitting constants b and c have a somewhat different pattern. Over the course of the day, b decreases while c increases.

5. Results: Shape of the cloud size distribution

The CSDs for the individual days suggest a relation between cloud cover and $l_{\max}$, and between cloud cover and the shape of the distribution. Now the complete dataset with all its variability will be used to investigate these relations in more detail and increase the statistical reliability.

a. Evolution of the fitting constants

As a first step in a more detailed analysis, the power-law exponential fit is determined for all days and times available. This results in a set of fitting constants a , b , and c . Since the constants b and c reflect the shape of the fit, they are plotted against cc in Fig. 8. The data are colored by $l_{\max}$, revealing that $l_{\max}$ correlates well with cloud cover. The fitting constants b and c are decreasing and increasing with cloud cover, respectively, with c becoming less negative and closer to zero. For greater cloud cover b becomes more negative, which means that the slope of the power-law part of the distribution is steeper. Identifying which processes influence the slope of the power law is still a matter of debate. Sakradzija and Hohenegger (2017) proposed a mechanism based on the Bowen ratio, whereas Feingold et al. (2017)

TABLE 1. Days shown in Fig. 5 with their cloud cover, fitting constants, maximum cloud size, and color as used in Fig. 6.

Panel in Fig. 5	Date	cc (%)	a	b	$c (\times 1000)$	$l_{\max}$	Color in Fig. 6
Fig. 5a	28 May 2012	9	0.02	0.0	−7.4	1200	Green
Fig. 5b	23 Jun 2016	14	12.10	−1.45	−2.2	1800	Blue
Fig. 5c	29 Jul 2012	27	347.79	−2.15	−0.5	3000	Purple
Fig. 5d	9 Sep 2015	35	2809.85	−2.55	0.0	3800	Orange
Fig. 5e	18 Apr 2015	46	536.88	−2.25	0.0	3400	Red

argued that the oscillation of the cloud field between a relative abundance of small or large clouds could explain the steepness of the slope.

The diurnal evolution of the b and c constants with the median (solid line) and upper and lower quartiles (shaded area) results from the fit of Eq. (2) to the hourly CSDs of all days (Fig. 9). In the early morning hours when the boundary layer is rapidly deepening and clouds start to form, the power law becomes more dominant (decrease of b , less negative c). During midday b is low and c is close to zero, whereas at the end of the day c becomes more negative again and the exponential part of the fit becomes stronger again. Apparently, for this library of cases the shape of the CSD changes significantly during the day, with a shift from exponential at cloud onset to more power law at midday and back to exponential in the evening. Interpreting this result purely from the perspective of percolation theory would mean that a fit closer to a pure single power law implies an enhanced aggregation of cumulus clouds (Windmiller 2017). However, this should be applied with caution; it does not have to mean that individual cumulus clouds actually merge during their ascent. All it reflects is that the cloud objects are perhaps bigger and are situated closer together, so that the system is closer to critical percolation.

b. Interpreting the power-law fits

As discussed earlier, power-law functions [Eq. (1)] have been commonly used in the past to fit to CSDs, in contrast to power-law exponentials. A major drawback of a single pure power law is that it only holds for the smaller cloud sizes and does not describe the distribution for the larger cloud sizes. The slope of the CSD can change considerably across the size range, sometimes featuring abrupt changes. To gain more insight into the dependence of the CSD slope on cloud size and cloud cover, we take advantage of the number of simulations at our disposal.

The following “goodness of fit” method is applied on all CSDs at 1300 LT. First, for each day a single power-law fit is applied to the full width of the CSD, covering all size bins. This yields an exponent b and a standard error (standard deviation) of this b (SE). The fit is then

applied on a reduced number of bins, leaving out the first bin at the lower end of the spectrum. The result is a new exponent and SE. This is repeated until the minimum amount of three bins to perform a fit is reached. The slope of the fit will vary with the amount of bins taken into account. This method highlights the variation of the CSD slopes across the size range, giving the opportunity to assess its dependence on the cloud sizes that are covered.

The results are visualized in Fig. 10. Box-and-whisker plots for every cloud cover bin show the median values and the spread in the data. The x axes denote the cloud size of the starting bin for the fitting range. Going from left to right means using fewer and fewer cloud sizes for applying the power-law fit, since the starting cloud size becomes bigger. For both the exponent (top row) and the SE (bottom row) the divergence increases toward a smaller fitting range.

The values of the exponent b at 200-m cloud size (smallest bin) compare well to results found in the literature (Rieck et al. 2014). The increasing divergence

FIG. 7. The CSDs for all available hours of 29 Jul 2012 (see also Table 2).

TABLE 2. Fitting constants, cloud cover, and $l_{\max}$ for Fig. 7 of the hourly CSDs of 29 Jul 2012.

Time of day (LT)	b	c ($\times 1000$)	cc (%)	$l_{\max}$
1000	0	-6.1	10.3	1200
1100	-1.25	-2.3	15.0	1800
1200	-2.4	-0.1	29.2	3400
1300	-2.15	-0.5	26.7	3000
1400	-1.75	-1.2	23.5	2200
1500	-1.9	-1.2	24.6	2200
1600	-2.6	-0.3	23.0	2400
1700	-2.55	-0.2	6.8	1200

from this value depends on the cloud cover. This divergence can also be seen for SE, and can be explained in two ways. At least partially because of the statistics being less robust for a shorter fitting range, the uncertainty and therefore the SE becomes higher. However, it also shows that especially for smaller cloud covers the power law is not sufficient to properly describe the CSD. Despite the statistical noise, the stratification as a function of cloud cover shows that for higher cloud covers the variations in b and SE are smaller than for lower cloud cover. This confirms the results from the previous section, that at larger cloud cover the power law reflects the shape of the CSD on a broader range of sizes. Since the exponential part is not taken into account in this analysis, the fit for low cloud cover is not as good as for large cloud cover and there is more variation in the slope.

c. Maximum cloud size

Apart from its shape, the CSD derived from the simulations is also defined by the maximum cloud size $l_{\max}$ that it covers. Several physical processes are candidates for controlling this $l_{\max}$. The boundary layer height, for example, limits the vertical size of the largest convective element that can be accommodated in the layer, an argument often used in boundary layer scaling. This in turn could also constrain the horizontal size of the element, although it is already shown by [de Roode et al. \(2004\)](#) that this is not the case for stratocumulus. The surface heat fluxes control the boundary layer height, and thereby they also might control the maximum cloud size. Recently, [Sakradzija and Hohenegger \(2017\)](#) showed that the Bowen ratio controls the shape of the mass flux distribution through the efficiency of the moist heat cycle in the subcloud layer. The shape of the mass flux and the cloud size distribution are similar, so therefore the surface flux partitioning could also have an influence on the cloud size. Finally, humidity of the environmental air at the top of the boundary layer might have an impact on the maximum cloud size. When the environmental air is more humid, clouds become less

FIG. 8. Cloud cover against the fitting constants (top) b and (bottom) c associated with the power-law exponential fit from Eq. (2). Each dot indicates the properties of a single hourly CSD, with the color indicating the maximum cloud size.

diluted at the edges and can grow bigger. Dry surroundings may limit the horizontal growth of the cloud by making cloud evaporation more efficient at the cloud edges. A priori it is not known which processes will control $l_{\max}$ since this has not been studied before. However, the considerable variation in cloud structure and amount, covered by the library of cases in this study, might provide a large-enough parameter space to gain more insight.

As a first step, the ensemble-mean time evolution of several candidate variables expressing the state of the boundary layer are compared to that of $l_{\max}$ (Fig. 11). The solid line in each panel is the median over all cases, while the shaded areas represent the bottom and top quartiles. To study the relations between the variables and $l_{\max}$, the Pearson correlation coefficient is calculated between all individual data points as

$$R_{x_1, x_2} = \frac{\text{COV}(x_1, x_2)}{\sigma_{x_1} \sigma_{x_2}}, \quad (3)$$

and is given in the upper-right corner of the panels in Fig. 11. COV is the covariance and σ is the standard deviation. The order of the panels reflects the degree of correlation with $l_{\max}$. The time series of both the convective velocity scale w_* (Fig. 11a) and the buoyancy flux (Fig. 11d) show the typical evolution of a daytime convective boundary layer, with turbulence increasing

FIG. 9. Temporal evolution of fitting constants (top) b and (bottom) c associated with the power-law exponential fit [Eq. (2)] as applied to the hourly cloud size distribution.

during the morning and decaying again toward sunset. The Bowen ratio (Fig. 11b) shows on average a decrease during the whole day, explaining the poor correlation with $l_{\max}$. The relative humidity (RH; Fig. 11c) averaged over the subcloud layer shows a decrease as the boundary layer develops and becomes deeper. Minimum cloud base (zb; Fig. 11e) shows the ballistic growth that has been described earlier by Stevens (2007) for this type of boundary layer. Boundary layer depth (zi, based on the Richardson number; Fig. 11f) increases strongly during the early morning hours and after that continues to deepen at a lower rate. The convective inhibition (CIN; Fig. 11g), maximum cloud-layer-top height

(zt; Fig. 11h), and the cloud cover up to 3 km (cc; Fig. 11i) all show a similar evolution. There is an increase in the beginning of the day, and a decrease at the end of the afternoon. This behavior better resembles the evolution of $l_{\max}$. The cloud cover up to 3 km (cc; Fig. 11g) shows the strongest diurnal cycle of all candidate variables. The fact that CIN, zt, and cc have well-defined maxima during the afternoon explains their good correlation with $l_{\max}$ (Fig. 11j), which shares this feature. The evolution of $l_{\max}$ also compares well with the observed cloud chord length as shown in Zhang et al. (2017), even though the observations are taken at a different location.

d. The influence of cloud cover

To further disentangle the possibly simultaneous impact of multiple variables on $l_{\max}$, the results are stratified on the strongest correlating variable, which is cloud cover. Stratifying the data on the dominant dependence can act to highlight any secondary controls on $l_{\max}$. To this end, the dataset is divided into bins of 10% cloud cover. With this bin size we maintain sufficient statistical quality (more than 400 points for a cloud cover up to 10% and around 200 points for the other cloud cover bins), while narrowing the bin width enough to notice possible patterns. Resulting correlation coefficients are shown in Table 3, with scatterplots for the boldface numbers in Figs. 12a–c. The correlation between $l_{\max}$ and each of zi, zt, zb, and CIN is stronger for a cloud percentage of 0%–10% than for all data together. For the higher cloud cover the correlation coefficients drop for zt, whereas they stay relatively constant for zi and zb. The correlation coefficients for CIN decrease most and show no strong correlation between CIN and $l_{\max}$ for the larger cloud cover. For low cloud cover the correlation coefficient is highest for zt. Figure 12c illustrates this high correlation coefficient; there are few outliers and

FIG. 10. Box-and-whisker plots of the exponent of the (top) power law and (bottom) standard error of the exponent SE. The x axes denote the starting size of the fitting range, so going from left to right means fewer and fewer cloud sizes are taken into account. Every column presents the data for a different cloud cover bin.

FIG. 11. Temporal evolution of a selection of variables. The correlation coefficient (upper-right corner in each panel) shows the correlation between (a)–(i) the depicted variable and (j) the maximum cloud size.

no offset at the y axis. Remarkably, the correlation coefficient for z_i is lower than for z_t and z_b . However, it should be noted that even the 146 days of data might be insufficient for reliable correlation coefficients. The computed correlation coefficient can still be influenced by some outliers, as can be seen, for example, in Fig. 12b. Nevertheless, the results suggest that after cloud cover, the depth of the cloud layer, which is represented by cloud base and cloud top, plays a secondary but still

important role in establishing the maximum cloud size in the population.

The four variables showing the best correlation with $l_{\max}$ are now investigated more closely. The variables considered are boundary layer height, cloud top, cloud base, and cloud cover. To assess the robustness of these high correlations, the calculation is repeated using slightly different definitions for a cloud and its size as described in section 3a. In Table 4 the correlation

TABLE 3. The correlation coefficient for projected maximum cloud size and boundary layer length scales, organized by different cloud covers. The boldface values correspond to the R values in Figs. 12a–c.

cc (%)	zi	zt	zb	CIN
0–10	0.32	0.66	0.47	0.38
10–20	0.29	0.59	0.5	0.26
20–30	0.32	0.48	0.5	0.03
30–40	0.42	0.35	0.45	0.22

coefficients are given between all five variables and $l_{\max}$, and for both definitions of a cloud and its size. For all four definitions a strong correlation exists between $l_{\max}$ and cloud cover. When taking only the cloud core into account, the correlation for all variables decreases slightly or stays the same. The difference between the projected and height-averaged cloud sizes is also small. There is a slightly better correlation for zb, but for the

other variables there is a small drop in the correlation coefficient. The high correlation coefficient for the projected $l_{\max}$ and cc is shown in Fig. 12d.

The strong correlation between $l_{\max}$ and cloud cover is robust for all tested cloud and cloud size definitions, as seen in the last row of Table 4. This could mean that fields with larger cloud cover feature larger clouds, instead of a larger number of smaller clouds. To test this hypothesis, we look at the cloud cover decomposition $\alpha(l)$:

$$\alpha(l) = \frac{l^2 N(l)}{L_x L_y}, \quad (4)$$

where L_x and L_y are the horizontal dimensions of the domain (Neggers et al. 2003). The cloud cover decomposition gives the contribution to the total cloud cover as a function of cloud size. To answer the question

FIG. 12. Scatterplots of $l_{\max}$ and (a) boundary layer height, (b) cloud base, (c) cloud top, and (d) cloud cover. For (a)–(c) only the days with cloud cover up to 10% are taken into account. The correlation coefficients R can also be found in boldface in Tables 3 and 4.

TABLE 4. The correlation coefficient for the variables that show the strongest correlation with $l_{\max}$ for different cloud and cloud size definitions. The boldface value corresponds to the R value in Fig. 12d.

	Projected	Height averaged	Buoyant	
			Projected	Height averaged
zi	0.24	0.18	0.35	0.29
zt	0.58	0.5	0.53	0.42
zb	0.17	0.19	0.07	0.09
cc	0.7	0.7	0.72	0.65

of the relative contribution of larger clouds to high cloud cover, we again divide the database of cloud cover into bins of 10%. For every cloud cover bin the cloud cover decomposition $\alpha(l)$ is shown in Fig. 13.

All cloud covers show a decreasing contribution to total cloud cover with increasing cloud size. This decrease is stronger for smaller cloud covers whereas the slope is less steep for larger cloud covers. The small cloud sizes contribute less to the total cloud cover for fields with cloud covers up to 10% as compared with larger cloud cover. The differences for larger cloud cover are minimal in this range. This is different for larger cloud sizes; here, an increase in contribution is visible when going from small to larger cloud cover. These findings confirm that an increase in cloud cover goes together with an increasing contribution of large clouds instead of a larger number of small clouds.

e. Impact of microphysics

The way condensation and rain formation are described by a microphysics scheme in a LES could in principle affect the cloud sizes and their distribution (Zhang et al. 2017). This is investigated by using an extra set of simulations of the days in 2013 and 2014 with the microphysics scheme of Sommeria (1976), from here on referred to as the Sommeria–Deardorff scheme. This is a scheme often operated by early LES models (e.g., Siebesma and Cuijpers 1995). The simulations in this paper used for analysis are done with the Seifert–Beheng scheme (Seifert and Beheng 2004). Independent of the microphysical scheme the all-or-nothing approach for cloud cover is used. This defines a grid box as cloudy once it is completely saturated. The main difference between the schemes is that in the Seifert–Beheng scheme rain is included and that it features number as well as mass densities. The formation and evaporation of rain has an impact on the heat and moisture budgets of the boundary layer. This could in turn influence cloud formation processes, and thereby cloud sizes. This short sensitivity test focuses on the first-order impact of the microphysics scheme on cloud sizes. Individual components of the schemes are not addressed since this is beyond the scope of this study.

FIG. 13. Cloud size against the cloud cover decomposition [Eq. (4)]. The dataset is divided into bins with similar cloud covers.

For both microphysics representations the diurnal evolution of the shape of the CSD as expressed by the hourly averaged fitting constants b and c is shown in Fig. 14. The diurnal evolution for both constants is similar for the two microphysical schemes, in the morning b becomes smaller and c bigger. There is a small difference between the two schemes, with b reaching the lower and c the higher values for the Seifert–Beheng scheme as compared to the Sommeria–Deardorff scheme. This combination of fitting constants is associated with a higher cloud cover, which can indeed be observed for the diurnal evolution of cloud cover shown in Fig. 14c. Besides these small differences, the signal in the diurnal evolution for the shape of CSD is the same for both microphysical schemes.

Second, we test the strong correlation we found between $l_{\max}$ and cloud cover for both microphysical parameterizations. Again, using the data from 2013 and 2014, the correlation coefficients are computed between the hourly averaged $l_{\max}$ and cc. For the Sommeria–Deardorff scheme this results in an R of 0.785, while for Seifert–Beheng the R is 0.782. In addition, the scatterplots with regression lines show only small differences (Fig. 15). Both regression lines show a similar slope, although the Seifert–Beheng line is a bit steeper. Probably due to the more advanced microphysics $l_{\max}$ is slightly bigger for these simulations. To summarize, the correlation found between $l_{\max}$ and cc agrees with these results not being dependent on whether the Sommeria–Deardorff or the Seifert–Beheng microphysical parameterization scheme is used.

FIG. 14. The diurnal evolution of the fitting constants (a) b and (b) c according to the power-law exponential function for two different microphysical schemes. (c) The accompanying cloud cover.

6. Discussion

The main goal of this study is to quantify the diurnal evolution of the CSD for continental shallow cumulus with a focus on its shape and range.

Concerning the range of the CSD, we find high correlations between $l_{\max}$ and cloud cover, cloud base and cloud top, and less so with the boundary layer height. The latter is surprising, because it is classically used as a scale that controls many aspects of the boundary layer. It is often used to normalize height and is seen as a measure for the vertical scale of the largest thermals. It would thereby control the maximum horizontal cloud size as well. However, here evidence is found that other controlling factors exist that codetermine the size of the cloud. The fact that correlation coefficients for cloud cover, cloud base, and cloud top are higher than for boundary layer depth suggests that there might be another factor that eventually effectively controls the

FIG. 15. Scatterplots of $l_{\max}$ and cloud cover for two different microphysical schemes. The blue circles represent the Sommeria-Deardorff scheme, and the orange triangles the Seifert-Beheng scheme. The correlation coefficients R for both sets of data are depicted in the upper-left corner.

cloud size. This also means that it might be necessary to reconsider the strong assumption that is used in some cumulus parameterizations.

A possible physical explanation for the high correlation between cloud cover and $l_{\max}$ could be the dryness of the cloud environment. The small cloud cover (up to 10%) can be seen as forced cumuli, that reach the inversion and just condense, but their environment is very dry, preventing them from growing further, both in the vertical and the horizontal. When the environmental air that surrounds the rising cumuli at the top of the boundary layer is sufficiently humid or close to saturation, the stronger cumuli manage to condense a much larger part of their area, not just their core. In addition, clouds dilute less quickly in more humid air, prolonging their lifetime and promoting the formation of outflow or “anvil” clouds. Further research is necessary to investigate this relation between the moisture content of the environmental air and the cloud cover/ $l_{\max}$. Some first steps were taken on this, but addressing the humidity in the direct surroundings of the cloud is not so straightforward since this really depends on the height and structure of the cloud. It is therefore considered outside the scope of the present study and is a topic for further research.

The strong correlation between $l_{\max}$ and cloud cover reveals something important about the response of a cumulus cloud population to a diurnally varying external forcing. We find that an increase in cloud cover is not

generated by a larger number of small clouds, but is instead carried by larger clouds. This behavior is relevant for radiation schemes in GCMs. The effective vertical transfer of radiation depends on the sizes and horizontal distribution of clouds (Barker and Räisänen 2005). Organizing the data by cloud cover showed a secondary role for cloud-base height and cloud-top height in controlling $l_{\max}$. Although strong correlations do not reveal anything about causality, this does indicate that besides cloud cover, the heights of the cloud base and cloud top play a secondary but still important role in establishing the maximum cloud size in the population.

The study by Sakradzija and Hohenegger (2017) suggested a strong link between the surface Bowen ratio and the mass flux distribution. The low correlation between $l_{\max}$ and the Bowen ratio reported in this study does not necessarily contradict that finding. It rather suggests that the vertical velocity component of the mass flux could carry the strongest dependence on the Bowen ratio (Lamer and Kollias 2015), and not the area of the largest thermal. Proving this hypothesis requires further research.

7. Conclusions

To improve our understanding of what controls the shape and range of cumulus cloud size distributions, we performed an ensemble of LESs of 146 cases of diurnal cycles of summertime shallow convection over land at a midlatitude meteorological site in northwestern Germany. To this purpose, the relation between the defining aspects of the CSD, including the slope and the maximum size, with various candidate variables expressing the boundary layer state was investigated.

The power-law exponential function, which was used to fit the cloud size distributions, shows the typical slope that has been reported in previous studies. This slope covers a larger range of cloud sizes as cloud cover increases. The shape of the CSD changes significantly during the day, with a shift from exponential at cloud onset to power law at midday and back to exponential in the evening. What this means in terms of the spatial structure of the cumulus cloud field, including the effects of the merging of clouds and neighbor spacing, is not explained by the results obtained in this study. Future research will therefore focus on the interaction between clouds in order to better understand the intercloud dynamics. A possible way to do this is to determine the cloud spacing in relation to cloud size and its relation with cloud cover.

A dominating correlation between cloud cover and maximum cloud size is found, although a weaker secondary relation with cloud-base height and cloud-top height is also identified. The correlation between cloud

cover and maximum size implies that it is not a larger number of small clouds, but rather larger clouds that are responsible for an increase in cloud cover. The possible role of environmental humidity in the strong link between cloud cover and maximum size needs to be investigated further.

With a subset of the data, sensitivity tests are performed on the output resolution and the microphysics scheme used in the LES. Both subsampling of the original resolution and the use of a less involved microphysics parameterization have no influence on the conclusions of this study.

Results of this study showed that the shape of the CSD is related to the cloud cover. This could explain the different CSDs found previously in the literature. Insights like this are beneficial for the development of scale-aware parameterization schemes based on CSDs, as they need to be informed about the shape of the CSD for closure and calibration (Wagner and Graf 2010; Neggers 2015; Brast et al. 2018).

Acknowledgments. This research is supported by the SFB-TR32 project “Patterns in soil-vegetation-atmosphere systems: Monitoring, modelling and data assimilation,” funded by the German Science Foundation (DFG). The authors are thankful for being given the opportunity to run simulations on CHEOPS, the cluster of the Regionales Rechenzentrum of the University of Cologne as well as on the JURECA supercomputer at the Jülich Supercomputing Centre (JSC). Observations used in this study originate from the German Research Foundation (DFG) funded Core Facility (JOYCE-CF) under DFG Research Grant LO 901/7-1. The observational data used in this study are available from the Institute for Geophysics and Meteorology of the University of Cologne. Finally, the authors want to thank Steve Klein and two anonymous reviewers for their useful comments and their help in improving this paper.

APPENDIX

Comparison of the Original- and Coarse-Resolution DALES Output

For computational reasons only every fourth point of the original LES output is used in this study, resulting in an effective horizontal resolution of 200 m. However, the simulations are done with a resolution of 50 m. In this section we will evaluate the consequences this subsampling of the original data might have.

We only take into account the 52 simulated days of the years 2013 and 2014. As a first step to seeing the influence of resolution on the CSD, the normalized distributions

FIG. A1. Cloud size densities for the original and coarse resolution (dots), including fits according to the power-law exponential function (lines). Densities are at 1300 LT 4 Apr 2013. The gray dotted line is the reflection of the fit through the subsampled data, only plotted in such a way that the densities match at the smallest bin of the coarse resolution.

are computed for a single moment, at 1300 LT 4 July 2013 (Fig. A1). Through the data (dots), the power-law exponential is fitted (lines). Green denotes the sub-sampled data (as used in this study) and shows a power law for the smaller sizes with an exponential for the larger sizes. The original resolution (red) displays a power law across all sizes. It seems that this distribution is heavily influenced by the smaller cloud sizes. For a fairer comparison the fit is therefore also determined by using the same cloud size as the starting point (orange line). This adjusted fit is based on the data for the original resolution, but ignores cloud sizes smaller than the smallest bin of the subsampled data. Now again the exponential part is seen for the bigger cloud sizes. This is supported by the values of the fitting constants, as shown in Table A1. Since for the original-resolution clouds smaller than 200 m are included as well, the total number of clouds is larger. This leads to a lower normalized CSD, as is also visible in Fig. A1. For an easier comparison of the slope and shape

TABLE A1. Fitting constants for CSD for 4 Apr 2013 for different output resolutions. The CSDs are shown in Fig. A1.

	a	b	$c (\times 1000)$
Original resolution	108.97	−2.3	0.0
Original, adjusted fit	0.46	−1.35	−1.0
Coarse resolution	52.78	−1.8	−0.8

FIG. A2. Evolution of the fitting constants (a) b and (b) c following the power-law exponential function, and (c) $l_{\max}$ from DALES. They are averaged over the CSDs at each time point for the original and coarse resolution.

of the different fits, the fit through the subsampled data is plotted again in dotted gray. This is done in such a way that the two different densities match at the smallest bin of the coarse resolution.

To study the temporal evolution of the shape of the fit for both resolutions, all days during 2013 and 2014 are used. The averaged b and c constants of the fit are plotted against time (Figs. A2a,b), with colors corresponding to the ones in Fig. A1. It can be seen that no matter the resolution, the temporal evolution of the fitting constants is similar. The value of b decreases during the day and c increases, and this process reverses at the end of the day. The largest differences between the resolutions can be seen in the early morning hours. For b , the adjusted fit for the original resolution is closest to the coarse resolution during the afternoon. For c , this is the case for the un-adjusted original resolution. The maximum cloud size is an important variable in this study and is therefore also

tested for its influence of subsampling. During the whole day, the averaged $l_{\max}$ as simulated by DALES (Fig. A2c) is very similar to the original resolution and the subsampled one, with an increase in the morning, a peak during midafternoon, and a small decrease afterward.

The magnitude and evolution of the b and c fitting constants of the power-law exponential fit are similar for both resolutions, and $l_{\max}$ is not influenced by subsampling. Therefore, the use of coarse resolution for the analysis in this paper is justified.

REFERENCES

- Arakawa, A., and W. H. Schubert, 1974: Interaction of a cumulus cloud ensemble with the large-scale environment, part I. *J. Atmos. Sci.*, **31**, 674–701, [https://doi.org/10.1175/1520-0469\(1974\)031<0674:IOACCE>2.0.CO;2](https://doi.org/10.1175/1520-0469(1974)031<0674:IOACCE>2.0.CO;2).
- , and C.-M. Wu, 2013: A unified representation of deep moist convection in numerical modeling of the atmosphere. Part I. *J. Atmos. Sci.*, **70**, 1977–1992, <https://doi.org/10.1175/JAS-D-12-0330.1>.
- Barker, H. W., and P. Räisänen, 2005: Radiative sensitivities for cloud structural properties that are unresolved by conventional GCMs. *Quart. J. Roy. Meteor. Soc.*, **131**, 3103–3122, <https://doi.org/10.1256/qj.04.174>.
- Benner, T. C., and J. A. Curry, 1998: Characteristics of small tropical cumulus clouds and their impact on the environment. *J. Geophys. Res.*, **103**, 28 753–28 767, <https://doi.org/10.1029/98JD02579>.
- Betts, A. K., 2009: Land-surface-atmosphere coupling in observations and models. *J. Adv. Model. Earth Syst.*, **1**, 4, <https://doi.org/10.3894/james.2009.1.4>.
- Bony, S., and J.-L. Dufresne, 2005: Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophys. Res. Lett.*, **32**, L20806, <https://doi.org/10.1029/2005GL023851>.
- Borque, P., P. Kollias, and S. Giangrande, 2014: First observations of tracking clouds using scanning ARM cloud radars. *J. Appl. Meteor. Climatol.*, **53**, 2732–2746, <https://doi.org/10.1175/JAMC-D-13-0182.1>.
- Brast, M., V. Schemann, and R. A. J. Neggers, 2018: Investigating the scale adaptivity of a size-filtered mass flux parameterization in the gray zone of shallow cumulus convection. *J. Atmos. Sci.*, **75**, 1195–1214, <https://doi.org/10.1175/JAS-D-17-0231.1>.
- Brient, F., T. Schneider, Z. Tan, S. Bony, X. Qu, and A. Hall, 2015: Shallowness of tropical low clouds as a predictor of climate models' response to warming. *Climate Dyn.*, **47**, 433–449, <https://doi.org/10.1007/s00382-015-2846-0>.
- Brown, A. R., and Coauthors, 2002: Large-eddy simulation of the diurnal cycle of shallow cumulus convection over land. *Quart. J. Roy. Meteor. Soc.*, **128**, 1075–1093, <https://doi.org/10.1256/003590002320373210>.
- Cahalan, R. F., W. Ridgway, W. J. Wiscombe, T. L. Bell, and J. B. Snider, 1994: The albedo of fractal stratocumulus clouds. *J. Atmos. Sci.*, **51**, 2434–2455, [https://doi.org/10.1175/1520-0469\(1994\)051<2434:TAOFSC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1994)051<2434:TAOFSC>2.0.CO;2).
- Dal Gesso, S., J. J. van der Dussen, A. P. Siebesma, S. R. de Roode, I. A. Boutle, Y. Kamae, R. Roehrig, and J. Vial, 2015: A single-column model intercomparison on the stratocumulus representation in present-day and future climate. *J. Adv. Model. Earth Syst.*, **7**, 617–647, <https://doi.org/10.1002/2014MS000377>.
- Dawe, J. T., and P. H. Austin, 2012: Statistical analysis of an LES shallow cumulus cloud ensemble using a cloud tracking algorithm. *Atmos. Chem. Phys.*, **12**, 1101–1119, <https://doi.org/10.5194/acp-12-1101-2012>.
- de Roode, S. R., P. G. Duynkerke, and H. J. J. Jonker, 2004: Large-eddy simulation: How large is large enough? *J. Atmos. Sci.*, **61**, 403–421, [https://doi.org/10.1175/1520-0469\(2004\)061<0403:LSHLIL>2.0.CO;2](https://doi.org/10.1175/1520-0469(2004)061<0403:LSHLIL>2.0.CO;2).
- Ding, B., C. Li, M. Zhang, G. Lu, and F. Ji, 2014: Numerical analysis of percolation cluster size distribution in two-dimensional and three-dimensional lattices. *Eur. Phys. J.*, **87B**, 179, <https://doi.org/10.1140/epjb/e2014-40996-4>.
- Dorrestijn, J., D. T. Crommelin, A. P. Siebesma, and H. J. J. Jonker, 2013: Stochastic parameterization of shallow cumulus convection estimated from high-resolution model data. *Theor. Comput. Fluid Dyn.*, **27**, 133–148, <https://doi.org/10.1007/s00162-012-0281-y>.
- ECMWF, 2017: Part IV: Physical processes. IFS Documentation Cy40r1, ECMWF, 190 pp., https://www.ecmwf.int/sites/default/files/IFS_CY40R1_Part4.pdf.
- Feingold, G., J. Balsells, F. Glassmeier, T. Yamaguchi, J. Kazil, and A. McComiskey, 2017: Analysis of albedo versus cloud fraction relationships in liquid water clouds using heuristic models and large eddy simulation. *J. Geophys. Res. Atmos.*, **122**, 7086–7102, <https://doi.org/10.1002/2017JD026467>.
- Gesso, S. D., and R. A. J. Neggers, 2018: Can we use single-column models for understanding the boundary layer cloud-climate feedback? *J. Adv. Model. Earth Syst.*, **10**, 245–261, <https://doi.org/10.1002/2017MS001113>.
- Gustafson, W. I., A. M. Vogelmann, X. Cheng, S. Endo, B. Krishna, Z. Li, T. Toto, and H. Xiao, 2017: Description of the LASSO alpha 1 release. DOE Tech. Rep. DOE/SC-ARM-TR-194, 43 pp. + appendixes, <https://doi.org/10.2172/1373564>.
- Heinze, R., C. Moseley, L. N. Böske, S. Muppa, V. Maurer, S. Raasch, and B. Stevens, 2016: Evaluation of large-eddy simulations forced with mesoscale model output for a multi-week period during a measurement campaign. *Atmos. Chem. Phys.*, **17**, 7083–7109, <https://doi.org/10.5194/acp-2016-498>.
- Heus, T., and A. Seifert, 2013: Automated tracking of shallow cumulus clouds in large domain, long duration large eddy simulations. *Geosci. Model Dev.*, **6**, 1261–1273, <https://doi.org/10.5194/gmd-6-1261-2013>.
- , and Coauthors, 2010: Formulation of the Dutch Atmospheric Large-Eddy Simulation (DALES) and overview of its applications. *Geosci. Model Dev.*, **3**, 415–444, <https://doi.org/10.5194/gmd-3-415-2010>.
- Holton, J. R., and G. J. Hakim, 2012: *An Introduction to Dynamic Meteorology*. 5th ed. International Geophysics Series, Vol. 88, Academic Press, 552 pp.
- Honnert, R., 2016: Representation of the grey zone of turbulence in the atmospheric boundary layer. *Adv. Sci. Res.*, **13**, 63–67, <https://doi.org/10.5194/asr-13-63-2016>.
- Jiang, H., G. Feingold, H. H. Jonsson, M.-L. Lu, P. Y. Chuang, R. C. Flagan, and J. H. Seinfeld, 2008: Statistical comparison of properties of simulated and observed cumulus clouds in the vicinity of Houston during the Gulf of Mexico Atmospheric Composition and Climate Study (GoMACCS). *J. Geophys. Res.*, **113**, D13205, <https://doi.org/10.1029/2007JD009304>.
- Joseph, J. H., and R. F. Cahalan, 1990: Nearest neighbor spacing of fair weather cumulus clouds. *J. Appl. Meteor.*, **29**, 793–805, [https://doi.org/10.1175/1520-0450\(1990\)029<0793:NNSOFW>2.0.CO;2](https://doi.org/10.1175/1520-0450(1990)029<0793:NNSOFW>2.0.CO;2).
- Kwon, Y. C., and S.-Y. Hong, 2017: A mass-flux cumulus parameterization scheme across gray-zone resolutions. *Mon. Wea. Rev.*, **145**, 583–598, <https://doi.org/10.1175/MWR-D-16-0034.1>.

- Lamer, K., and P. Kollias, 2015: Observations of fair-weather cumuli over land: Dynamical factors controlling cloud size and cover. *Geophys. Res. Lett.*, **42**, 8693–8701, <https://doi.org/10.1002/2015GL064534>.
- Löhnert, U., and Coauthors, 2015: JOYCE: Jülich Observatory for Cloud Evolution. *Bull. Amer. Meteor. Soc.*, **96**, 1157–1174, <https://doi.org/10.1175/BAMS-D-14-00105.1>.
- Negggers, R. A. J., 2015: Exploring bin-macrophysics models for moist convective transport and clouds. *J. Adv. Model. Earth Syst.*, **7**, 2079–2104, <https://doi.org/10.1002/2015MS000502>.
- , and A. P. Siebesma, 2013: Constraining a system of interacting parameterizations through multiple-parameter evaluation: Tracing a compensating error between cloud vertical structure and cloud overlap. *J. Climate*, **26**, 6698–6715, <https://doi.org/10.1175/JCLI-D-12-00779.1>.
- , H. J. J. Jonker, and A. P. Siebesma, 2003: Size statistics of cumulus cloud populations in large-eddy simulations. *J. Atmos. Sci.*, **60**, 1060–1074, [https://doi.org/10.1175/1520-0469\(2003\)60<1060:SSOCCP>2.0.CO;2](https://doi.org/10.1175/1520-0469(2003)60<1060:SSOCCP>2.0.CO;2).
- , A. P. Siebesma, and T. Heus, 2012: Continuous single-column model evaluation at a permanent meteorological supersite. *Bull. Amer. Meteor. Soc.*, **93**, 1389–1400, <https://doi.org/10.1175/BAMS-D-11-00162.1>.
- Nuijens, L., B. Medeiros, I. Sandu, and M. Ahlgrimm, 2015: Observed and modeled patterns of covariability between low-level cloudiness and the structure of the trade-wind layer. *J. Adv. Model. Earth Syst.*, **7**, 1741–1764, <https://doi.org/10.1002/2015MS000483>.
- Park, S., 2014: A unified convection scheme (UNICON). Part I: Formulation. *J. Atmos. Sci.*, **71**, 3902–3930, <https://doi.org/10.1175/JAS-D-13-0233.1>.
- Peters, O., J. D. Neelin, and S. W. Nesbitt, 2009: Mesoscale convective systems and critical clusters. *J. Atmos. Sci.*, **66**, 2913–2924, <https://doi.org/10.1175/2008JAS2761.1>.
- Plank, V. G., 1969: The size distribution of cumulus clouds in representative Florida populations. *J. Appl. Meteor.*, **8**, 46–67, [https://doi.org/10.1175/1520-0450\(1969\)008<0046:TSDOCC>2.0.CO;2](https://doi.org/10.1175/1520-0450(1969)008<0046:TSDOCC>2.0.CO;2).
- Rieck, M., C. Hohenegger, and C. C. van Heerwaarden, 2014: The influence of land surface heterogeneities on cloud size development. *Mon. Wea. Rev.*, **142**, 3830–3846, <https://doi.org/10.1175/MWR-D-13-00354.1>.
- Sakradzija, M., and C. Hohenegger, 2017: What determines the distribution of shallow convective mass flux through a cloud base? *J. Atmos. Sci.*, **74**, 2615–2632, <https://doi.org/10.1175/JAS-D-16-0326.1>.
- Schalkwijk, J., H. J. J. Jonker, A. P. Siebesma, and F. C. Bosveld, 2015: A year-long large-eddy simulation of the weather over Cabauw: An overview. *Mon. Wea. Rev.*, **143**, 828–844, <https://doi.org/10.1175/MWR-D-14-00293.1>.
- Seifert, A., and K. D. Beheng, 2004: A two-moment cloud microphysics parameterization for mixed-phase clouds. Part 1: Model description. *Meteor. Atmos. Phys.*, **92**, 45–66, <https://doi.org/10.1007/s00703-005-0112-4>.
- Siebesma, A. P., and J. W. M. Cuijpers, 1995: Evaluation of parametric assumptions for shallow cumulus convection. *J. Atmos. Sci.*, **52**, 650–666, [https://doi.org/10.1175/1520-0469\(1995\)052<0650:EOPAFS>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<0650:EOPAFS>2.0.CO;2).
- Simmons, A. J., 1986: Orography and the development of the ECMWF model. *Workshop on Observation, Theory and Modelling of Orographic Effects*, Vol. 2, Shinfield Park, Reading, United Kingdom, ECMWF, 129–164, <http://www.ecmwf.int/en/learning/workshops-and-seminars/past-workshops/1986-annual-seminar-and-workshop>.
- , and D. M. Burridge, 1981: An energy and angular-momentum conserving vertical finite-difference scheme and hybrid vertical coordinates. *Mon. Wea. Rev.*, **109**, 758–766, [https://doi.org/10.1175/1520-0493\(1981\)109<0758:AEAAMC>2.0.CO;2](https://doi.org/10.1175/1520-0493(1981)109<0758:AEAAMC>2.0.CO;2).
- Sommeria, G., 1976: Three-dimensional simulation of turbulent processes in an undisturbed trade wind boundary layer. *J. Atmos. Sci.*, **33**, 216–241, [https://doi.org/10.1175/1520-0469\(1976\)033<0216:TDSOTP>2.0.CO;2](https://doi.org/10.1175/1520-0469(1976)033<0216:TDSOTP>2.0.CO;2).
- Stevens, B., 2007: On the growth of layers of nonprecipitating cumulus convection. *J. Atmos. Sci.*, **64**, 2916–2931, <https://doi.org/10.1175/JAS3983.1>.
- Trivej, P., and B. Stevens, 2010: The echo size distribution of precipitating shallow cumuli. *J. Atmos. Sci.*, **67**, 788–804, <https://doi.org/10.1175/2009JAS3178.1>.
- van der Dussen, J. J., and Coauthors, 2013: The GASS/EUCLIPSE model intercomparison of the stratocumulus transition as observed during ASTEX: LES results. *J. Adv. Model. Earth Syst.*, **5**, 483–499, <https://doi.org/10.1002/jame.20033>.
- van Zanten, M. C., and Coauthors, 2011: Controls on precipitation and cloudiness in simulations of trade-wind cumulus as observed during RICO. *J. Adv. Model. Earth Syst.*, **3**, M06001, <https://doi.org/10.1029/2011MS000056>.
- Vial, J., S. Bony, J.-L. Dufresne, and R. Roehrig, 2016: Coupling between lower-tropospheric convective mixing and low-level clouds: Physical mechanisms and dependence on convection scheme. *J. Adv. Model. Earth Syst.*, **8**, 1892–1911, <https://doi.org/10.1002/2016MS000740>.
- Wagner, T. M., and H.-F. Graf, 2010: An ensemble cumulus convection parameterization with explicit cloud treatment. *J. Atmos. Sci.*, **67**, 3854–3869, <https://doi.org/10.1175/2010JAS3485.1>.
- Wielicki, B. A., and R. M. Welch, 1986: Cumulus cloud properties derived using Landsat satellite data. *J. Climate Appl. Meteor.*, **25**, 261–276, [https://doi.org/10.1175/1520-0450\(1986\)025<0261:CCPDUL>2.0.CO;2](https://doi.org/10.1175/1520-0450(1986)025<0261:CCPDUL>2.0.CO;2).
- Windmiller, J., 2017: Organization of tropical convection. Ph.D. thesis, Ludwig-Maximilian University, Munich, Germany, 124 pp., <https://edoc.ub.uni-muenchen.de/21245/>.
- Wood, R., and P. R. Field, 2011: The distribution of cloud horizontal sizes. *J. Climate*, **24**, 4800–4816, <https://doi.org/10.1175/2011JCLI4056.1>.
- Wyngaard, J. C., 2004: Toward numerical modeling in the “terra incognita.” *J. Atmos. Sci.*, **61**, 1816–1826, [https://doi.org/10.1175/1520-0469\(2004\)061<1816:TNNMIT>2.0.CO;2](https://doi.org/10.1175/1520-0469(2004)061<1816:TNNMIT>2.0.CO;2).
- Zhang, M., and Coauthors, 2013: CGILS: Results from the first phase of an international project to understand the physical mechanisms of low cloud feedbacks in single column models. *J. Adv. Model. Earth Syst.*, **5**, 826–842, <https://doi.org/10.1002/2013MS000246>.
- Zhang, Y., and S. A. Klein, 2013: Factors controlling the vertical extent of fair-weather shallow cumulus clouds over land: Investigation of diurnal-cycle observations collected at the ARM Southern Great Plains site. *J. Atmos. Sci.*, **70**, 1297–1315, <https://doi.org/10.1175/JAS-D-12-0131.1>.
- , —, J. Fan, A. S. Chandra, P. Kollias, S. Xie, and S. Tang, 2017: Large-eddy simulation of shallow cumulus over land: A composite case based on ARM long-term observations at its Southern Great Plains site. *J. Atmos. Sci.*, **74**, 3229–3251, <https://doi.org/10.1175/JAS-D-16-0317.1>.
- Zhao, G., and L. Di Girolamo, 2007: Statistics on the macrophysical properties of trade wind cumuli over the tropical western Atlantic. *J. Geophys. Res.*, **112**, D10204, <https://doi.org/10.1029/2006JD007371>.