Cleveland State University
EngagedScholarship@CSU

Physics Faculty Publications | Physics Department

6-2019

Power-Law Scaling in the Internal Variability of Cumulus Cloud Size Distributions due to Subsampling and Spatial Organization

R. A. J. Neggers
University of Cologne, neggers@meteo.uni-koeln.de

P. J. Griewank
University of Cologne

Thijs Heus
Cleveland State University, t.heus@csuohio.edu

Follow this and additional works at: https://engagedscholarship.csuohio.edu/sciphysics_facpub

 Part of the Atmospheric Sciences Commons, and the Physics Commons

How does access to this work benefit you? Let us know!

Repository Citation

Neggers, R. A. J.; Griewank, P. J.; and Heus, Thijs, "Power-Law Scaling in the Internal Variability of Cumulus Cloud Size Distributions due to Subsampling and Spatial Organization" (2019). *Physics Faculty Publications*. 387.
https://engagedscholarship.csuohio.edu/sciphysics_facpub/387

This Article is brought to you for free and open access by the Physics Department at EngagedScholarship@CSU. It has been accepted for inclusion in Physics Faculty Publications by an authorized administrator of EngagedScholarship@CSU. For more information, please contact library.es@csuohio.edu.

Power-Law Scaling in the Internal Variability of Cumulus Cloud Size Distributions due to Subsampling and Spatial Organization

R. A. J. NEGGERS AND P. J. GRIEWANK

University of Cologne, Cologne, Germany

T. HEUS

Cleveland State University, Cleveland, Ohio

(Manuscript received 8 July 2018, in final form 27 November 2018)

ABSTRACT

In this study, the spatial structure of cumulus cloud populations is investigated using three-dimensional snapshots from large-domain LES experiments. The aim is to understand and quantify the internal variability in cloud size distributions due to subsampling effects and spatial organization. A set of idealized shallow cumulus cases is selected with varying degrees of spatial organization, including a slowly organizing marine precipitating case and five more quickly organizing diurnal cases over land. A subdomain analysis is applied, yielding cloud number distributions at sample sizes ranging from severely undersampled to nearly complete. A strong power-law scaling is found in the relation between cloud number variability and subdomain size, reflecting an inverse linear relation. Scaling subdomain size by cloud size yields a data collapse across time points and cases, highlighting the role played by cloud spacing in controlling the stochastic variability. Spatial organization acts on top of this baseline model by increasing the maximum cloud size and by enhancing the variability in the number of smallest clouds. This reflects that the smaller clouds start to live on top of larger-scale thermodynamic structures, such as cold pools, which favor or inhibit their formation. Compositing all continental cumulus cases suggests the existence of a prototype diurnal time dependence in the spatial organization. A simple stochastic expression for cloud number variability is proposed that is formulated in terms of two dimensionless groups, which allows objective estimation of the degree of spatial organization in simulated and observed cumulus cloud populations.

1. Introduction

Recent advances in supercomputing have introduced a "gray zone" in the representation of cumulus convection in general circulation models (GCMs), in which this physical process is getting partially resolved (Wyngaard 2004). Recent studies have defined the gray zone as the range of resolutions within which the subgrid and resolved contribution to turbulence and transport are of the same order of magnitude (e.g., Dorrestijn et al. 2013; Honnert et al. 2011). Existing parameterization schemes, built on the assumption that cumulus populations are fully sampled in the GCM grid box, often lack the scale awareness and scale adaptivity to conceptually and practically deal with this situation. How to best achieve this is still an open research question.

Research into the spatial structure of cumulus cloud populations goes back decades (e.g., Cahalan and Joseph 1989; Sengupta et al. 1990; Nair et al. 1998) but has recently intensified because of the arrival of the gray zone problem. A natural way of characterizing the scale dependence within a cumulus population is the *size distribution* of a cloud field, which has been scientifically established for many cloud regimes using a large variety of instrumentation (e.g., Plank 1969; Raga et al. 1990; Benner and Curry 1998; Zhao and Di Girolamo 2007; Yuan 2011) and finescale cloud-resolving simulation (e.g., Neggers et al. 2003; Rieck et al. 2014; Senf et al. 2018). A characteristic feature of the size density of cumulus cloud number [a cloud size density (CSD)] is its shape, which has been described by lognormal, exponential, and power-law functions. These shapes

Denotes content that is immediately available upon publication as open access.

Corresponding author: R. A. J. Neggers, neggers@meteo.uni-koeln.de

DOI: 10.1175/JAS-D-18-0194.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

are argued to reflect the underlying physical–dynamical processes of cloud formation, interaction, and subcritical percolation (Newman 2005; Cohen and Craig 2006; Ding et al. 2014; Garrett et al. 2018). The size range covered by the distribution has been found to differ between marine and continental cloud populations (Sakradzija and Hohenegger 2017). Over land, the maximum size of the distribution exhibits a distinct diurnal dependence. With many studies relying on finescale simulations of cumulus cloud populations, the confrontation of simulated and observed cumulus populations has recently become an important focus point. Both area-covering measurements (Kassianov et al. 2005) and vertically pointing measurements (Lareau et al. 2018) have started to provide a wealth of new information about cumulus cloud populations at high resolutions and frequencies.

Building on these new results and insights, a new class of convection schemes has recently emerged in which the transport is explicitly formulated in terms of discretized CSDs (Wagner and Graf 2010; Park 2014; Neggers 2015). In essence, these schemes adopt the spectral approach in convective modeling as first proposed by Arakawa and Schubert (1974). Compared to bulk convection schemes, the CSD is a new variable that requires closure, concerning both its functional form and its range. Differences exist among schemes in how the CSD is reconstructed. But a general benefit is that the CSD is in general well observable, allowing such models to be constrained by both existing and new datasets. Furthermore, scale awareness is by definition present at the foundation of the scheme (Neggers 2015). The reconstructed CSD can be size filtered, which introduces scale adaptivity in the parameterized convective transport and clouds (Brast et al. 2018).

A complicating factor is that the CSD is not a global constant and exhibits both external and internal variability. External variability arises from various sources such as large-scale synoptic conditions or surface properties (Rieck et al. 2014). In contrast, internal variability arises because of spatial sampling issues. Given a well-defined CSD reflecting a cloud population covering an infinitely large area, a limited spatial sampling leads to stochastic variability (e.g., Nair et al. 1998; Cohen and Craig 2006). This situation applies to convective modeling in the gray zone, in which the grid box has become too small to include the full population. In CSDs diagnosed in LES realizations of a relatively small domain, this stochastic signal is sometimes present at the largest cloud sizes, which are often poorly sampled (Neggers et al. 2003). The spatial organization in a cloud field can add to this stochastic variability by introducing perturbations in the cloud populations on scales much larger than the boundary layer depth (e.g., Seifert and Heus 2013). Physical–dynamical processes that drive spatial organization include cold pool formation (e.g., Schlemmer and Hohenegger 2014) and oscillations (Sakradzija et al. 2015; Feingold et al. 2017).

In recent years, super-large-domain LES has become computationally feasible (Khairoutdinov and Randall 2006; Satoh et al. 2008; Heinze et al. 2017) and has offered new ways to study the CSD and its variability. The availability of high spatial resolution in combination with a large domain size yields cloud size distributions that can freely form and evolve across a range of scales that is less and less artificially constrained at both ends. The large domain allows more complete sampling of the cumulus population, as well as better coverage of the rarely occurring largest clouds. With the cumulus clouds still resolved, the ever-larger domain of the simulation also allows mesoscale fluctuations to form naturally and start to affect the CSD. The opportunities created by super-large LES to investigate the behavior of cumulus CSDs and their interaction with mesoscale organization have only recently begun to be explored (e.g., Senf et al. 2018).

The goal of this study is to use large-domain simulations to investigate and quantify the internal variability in a cumulus CSD. We ask how this variability depends on (i) the domain size of the analysis and (ii) the degree of spatial organization within the cloud population. The aim is to work with domain sizes large enough to approach full sample size when diagnosing the CSDs. While this study is exclusively limited to shallow cumulus convection, two distinctive convective regimes are investigated. The first is a slowly organizing marine cumulus case based on the Rain in Cumulus over the Ocean (RICO) field campaign (Rauber et al. 2007). The second regime reflects continental summertime conditions at the Southern Great Plains (SGP) site of the Atmospheric Radiation Measurement (ARM; Stokes and Schwartz 1994) program. A subdomain analysis is performed for all fields to quantify the impact of subsampling on the CSD. The difference in organization speed between the two regimes then provides insight on how this process affects the internal variability of shallow cumulus CSDs. The implications of the obtained results for our understanding of cumulus populations in general and for stochastic cumulus parameterization in the gray zone in particular will be discussed.

2. Method

a. Large-eddy simulations

This study makes use of LES results for two well-defined shallow cumulus regimes, in which the pace of the spatial organization differs considerably. The first, a slowly organizing cloud regime, reflects marine subtropical

subsidence conditions as observed during the RICO field campaign (Rauber et al. 2007). The spatial organization in this case is related to warm precipitation processes, producing cold pools with low cloud mass that are separated by convergence lines where clouds tend to congregate. The spatial organization in this case is a slow process, taking multiple days. The second, faster organizing cloud regime reflects diurnal cycles of continental shallow cumulus at the ARM SGP site. For this regime, five day-long cases are chosen, which are all part of the archive of simulations for the ARM SGP site in 2016 generated by the LES ARM Symbiotic Simulation and Observation Workflow (LASSO; Gustafson et al. 2017b). The cases include 18 May, 19 and 25 June, 16 July, and 18 August 2016 (hereinafter referred to as the 20160518, 20160619, 20160625, 20160716, and 20160818 cases). These days are selected because they best reflect the prototype view of diurnal cycles of shallow convection, with most daytime cloudiness associated with surface-driven convective clouds and with more or less clear-sky conditions before and afterward.

The LES experiments for these six cases have already been described in detail in the literature, and only the details relevant for this study will be summarized here. The RICO composite shallow cumulus case (vanZanten et al. 2011) is simulated using the University of California, Los Angeles, LES (UCLA-LES) code, as described in detail by Seifert and Heus (2013); their simulations are also used in this study. The simulated domain size is $D_x \times D_y \times D_z = 51.2\,\text{km} \times 51.2\,\text{km} \times 4.0\,\text{km}$. A period of 48 h is simulated, considered long enough for the spatial organization to take place. Full three-dimensional fields are stored at 8-h intervals. The LASSO cases have been resimulated with the MicroHH code (van Heerwaarden et al. 2017), using the WRF-based prescribed large-scale forcings and boundary conditions that are part of the LASSO Alpha 2 dataset (Gustafson et al. 2017a). Compared to the standard LASSO simulations, the horizontal domain size is extended to 25.6 km in the resimulations in order to maximize the domain range available for the subdomain analysis of CSDs.

Although two different LES codes are used, the experiment setups share some important aspects. A spatial resolution of 25 m × 25 m × 25 m is always applied, as is adaptive time stepping. Both the UCLA-LES and the MicroHH codes apply the Smagorinsky scheme for subgrid transport of momentum, energy, and heat, and both make use of the warm-cloud double-moment microphysics scheme of Seifert and Beheng (2001). Horizontally periodic boundary conditions are applied, as well as a sponge layer in the top third of the domain to dampen any gravity waves. A prescribed surface temperature is used in both the RICO and LASSO cases, the only difference being the presence of a diurnal cycle in the latter.

The evolution of the simulated cloud population for the RICO case is illustrated in Fig. 1, showing the vertically projected cloud mask at each storage time point. The cloud field at the first two time points still looks more or less homogeneously distributed, but increasingly, the cloud field becomes more organized, with large cloud structures appearing after about one day of simulation associated with cold pool development (Seifert and Heus 2013). Figure 2 shows the same for the LASSO 20160518 case, revealing the distinct diurnal evolution of the cloud population. The spacing between the largest clouds increases during the day, likely reflecting the continuous ongoing deepening of the boundary layer that is typical of such diurnal cycles (Brown et al. 2002; Zhang et al. 2017). Organization in the cloud field becomes apparent during the final period, with small clouds increasingly clustering around the larger ones, being separated by large areas without any cloud mass. This may be related to the decay of the largest clouds at the end of the day, when the surface forcing of the turbulence weakens.

To provide further confidence in the realism of the simulations, the LES cloud cover for the five LASSO cases is compared to two ARM observational products. The Active Remote Sensing of Clouds (ARSCL) value-added product (Clothiaux et al. 2000, 2001) combines lidar and radar measurements, while the total sky imager (TSI; Kassianov et al. 2005) retrieves cloud cover from wide-angle imagery. The comparison for the five cases is shown in Fig. 3, only showing daytime hours. While some differences exist between the two measurements, they agree reasonably well on the general evolution and amplitude of the cloud cover. The model does reproduce these trends; however, the amplitude seems to be a bit underestimated. This slight bias has been revealed in other recent LES evaluation studies at supersites (Schalkwijk et al. 2015; Zhang et al. 2017). In addition, on 18 May and 18 August, the LES seems to underestimate cloud cover in the morning, which we speculate is due to either (i) high-altitude cloudiness or (ii) spinup effects. Despite these minor shortcomings, the results do suggest that the experiment setup captures the prototype diurnal variation in cloudiness present in the observations to a reasonable degree, as well as its modulation due to large-scale forcing. This agreement suggests the simulations are representative of nature and justifies their use for subsequent cloud studies.

b. *Deriving cloud size distributions*

Cloud size distributions can be derived in many ways, involving choices concerning what defines a cloud and

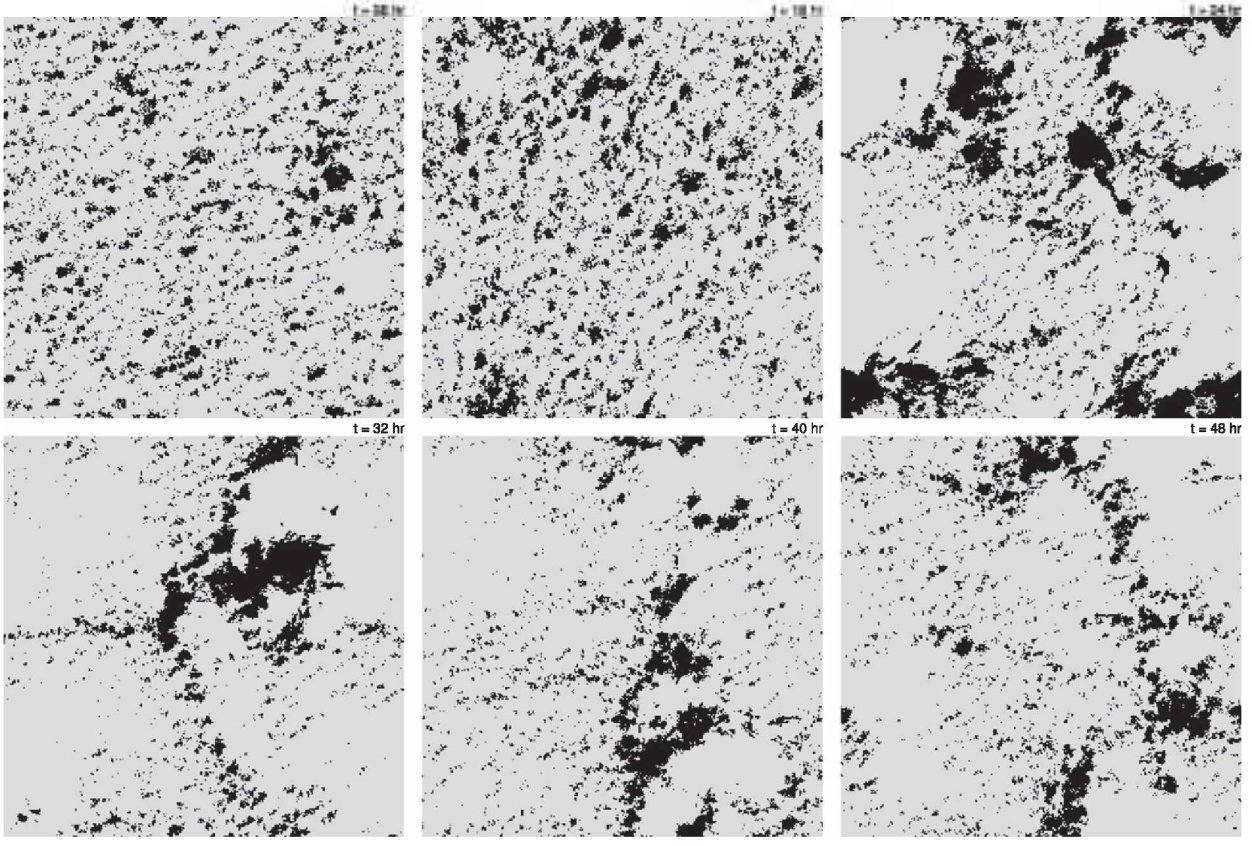

FIG. 1. Vertically projected cloud masks for the six time points during the LES of the RICO case that are considered in this study. The domain size shown is 51.2 km by 51.2 km. A grid point is masked (black) when any condensate exists in the vertical column.

its size. In this study, a simple definition is adopted, to be applicable to both observed and simulated cloud populations, and that allows the use of past data records as much as possible. The first simplification is to disregard cloud life cycles by only analyzing standalone instantaneous three-dimensional fields. Often, the time frequency of observations of cloud populations is simply too low to allow detection of cloud life cycle properties. Also, we hypothesize that the CSD internal variability should already be detectable in instantaneous snapshots. The second choice is to use the presence of any cloud condensate mass to label a grid box as cloudy or noncloudy. This allows the analysis to be used with any LES, no matter how complex its representation of cloud microphysics. The third choice is to define cloud size as the square root of the projected area fraction of the object defined on the 3D grid. This choice is made to allow comparison to previous CSD studies, which often relied on this definition.

In practice, the "cusize" algorithm is applied as described by Neggers et al. (2003) to investigate diurnal cycles in the CSD. Given a 3D field of cloud mask, first, all objects are identified that consist of neighboring cloudy grid boxes and that are separated by clear air. The size of each object is then established as the square root of its projected cloud cover. These are then sorted

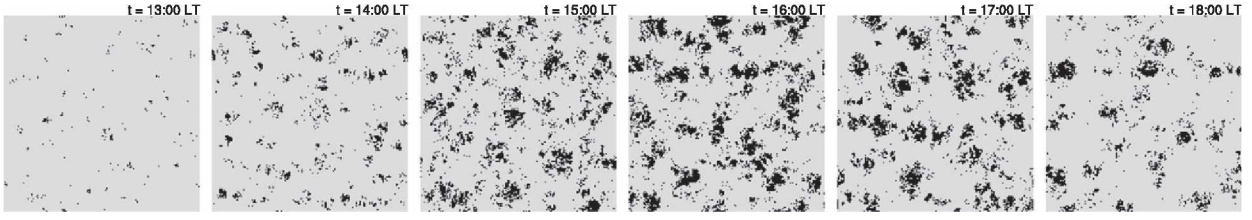

FIG. 2. As in Fig. 1, but for the cloud mask for six time points during the LASSO 20160518 case. The domain size is 25 km by 25 km.

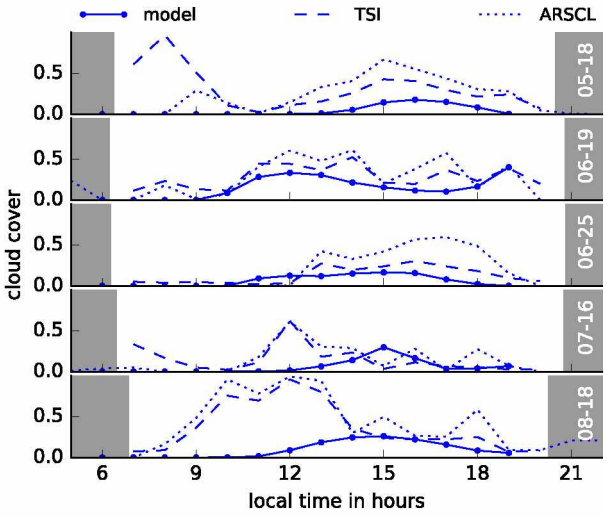

FIG. 3. Simulated and observed total cloud cover for the five LASSO cases investigated in this study. TSI is an estimate of the total sky cover and is only available during daytime, while ARSCL represents cloud cover as diagnosed over only the lowest 5 km. The gray bars denote the period before sunrise and after sunset.

into a histogram of cloud number $N(l)$, yielding the CSD $\mathscr{N}(l)$ defined as

$$\mathscr{N}(l) = N(l)\, dl^{-1}, \tag{1}$$

with dl the width of the size bins. As illustration, the resulting size densities of cloud number in log–log space for the RICO case and the LASSO 20160518 case are shown in Fig. 4. Each line represents a single instantaneous 3D field. The typically observed behavior that cumulus cloud occurrence sharply decreases with size is also evident in these cases, as expected. The RICO case shows power-law scaling in the size range $l < 300$ m, above which the CSD decreases more rapidly with size. The first cloudy hours in the LASSO case also have this shape, but as time progresses, the slope of the CSD becomes more constant across the covered size range. Note that a higher vertical position in the frame corresponds to a larger number of clouds, because of the use of nonnormalized densities. In the LASSO case, cloud number strongly increases across the spectrum after cloud onset and decreases again toward the end of the day. What both cases share is the increasing spread on the vertical axis (number density) toward the larger sizes, in the right tail. This spread, referred to as *CSD internal variability* or *stochastic variability*, is the main focus of this study.

c. *Subdomain analysis*

The subdomain analysis method used by Dorrestijn et al. (2013) to investigate the scale dependence of turbulent transport is applied here to cloud number N. To this purpose, the full horizontal LES domain $D_x \times D_y$ is first subdivided into small square subdomains of horizontal size $L_0 = 1.6$ km (L is used to indicate subdomain size, referring to the length of the edge of the square object). This basic size was chosen to still contain a sufficient number of grid points on a horizontal slice (64^2) and to be large enough to contain the largest cloud but, on the other hand, to still be significantly smaller than the typical neighbor spacing of large cumulus clouds (Joseph and Cahalan 1990). Also, with this value

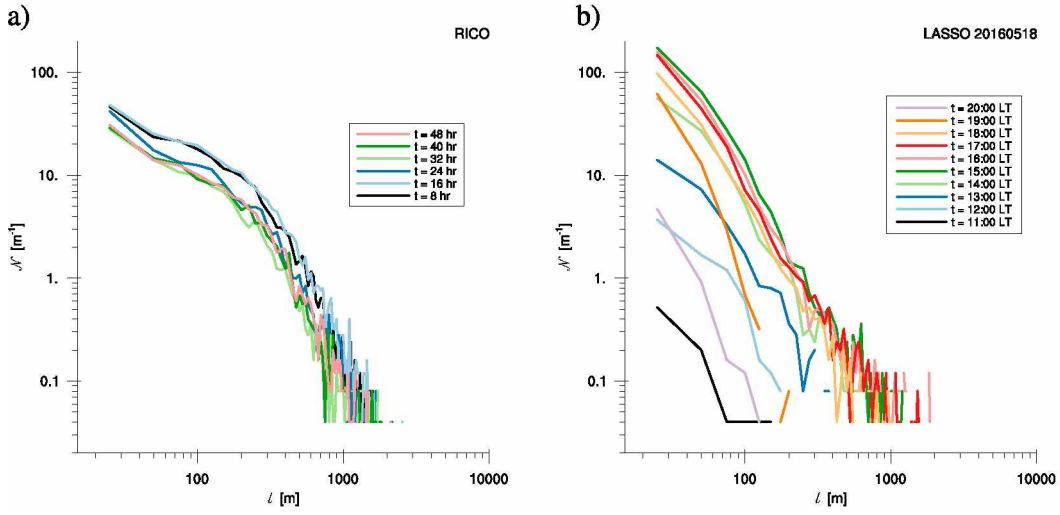

FIG. 4. Cloud number densities $\mathscr{N}$ for (a) the RICO case and (b) the LASSO 20160518 case, as derived from the full-domain LES fields. Each line represents a single instantaneous 3D cloud field, with t its time point in hours. The clouds were sorted in histograms using a constant bin width of 25 m, matching the LES resolution.

of L_0, the full domain size is exactly covered for both the RICO and LASSO cases.

The next step is to define a set of bigger square subdomains of size L that consist of multiple adjacent smallest subdomains L_0:

$$L = nL_0, \tag{2}$$

with $n \in \{1, 2, 4, 8, 16\}$ for LASSO and $n \in \{1, 2, 4, 8, 16, 32\}$ for RICO. This yields a number of K subdomains of size L, with

$$K = \left(\frac{D_x}{L}\right)^2 = \left(\frac{D_y}{L}\right)^2. \tag{3}$$

In each of these K subdomains, indexed by subscript k, the number of clouds N of size l are then counted, indicated as $N_k(L, l)$. In this process, no cloud is counted twice; in other words, should a cloud object cross subdomain boundaries, then the cloud is included in the count for the subdomain in which it is identified first. As a result, the sum of N_k over all subdomains always exactly equals the total cloud number counted for the full domain.

With N_k known, this finally allows the calculation of the mean cloud number per subdomain $\overline{N}(L, l)$, as well as the associated standard deviation $\sigma(L, l)$:

$$\overline{N}(L,l) = \frac{1}{K}\sum_{k=1}^{K} N_k(L,l), \tag{4}$$

$$\sigma(L,l) = \left\{\frac{1}{K}\sum_{k=1}^{K}[N_k(L,l) - \overline{N}(L,l)]^2\right\}^{1/2}. \tag{5}$$

Equation (5) expresses how variable the number of clouds of a certain size l is when considering subdomains of size L. The variable σ is thus a measure of the CSD internal variability, visible in Fig. 4 as the vertical spread in the right tail of the CSD. In practice, it is only interpreted for $K \geq 4$ to ensure statistical significance. It is important to note that this definition of σ accounts for the impact of no clouds occurring in a subdomain (i.e., $N_k = 0$). This has the potential to significantly increase σ in situations when the subdomain size is of the same order of magnitude as the cloud spacing, as will be illustrated later.

3. Results

a. Number densities in subdomains

Figure 5 shows a subdomain analysis of the size density of the number of clouds (the number density) of the first 3D snapshot from the RICO case. By visual comparison in Fig. 1 this cloud field shows the lowest degree of spatial organization, so that the CSD internal variability will be predominantly due to subsampling. Unlike the CSDs as shown in Fig. 4 the data are now plotted not as histograms but as two-dimensional PDFs in $(l, \mathcal{N}_*)$ space. The normalized size density for subdomains $\mathcal{N}_*(L, l)$ is obtained by dividing $N(L, l)$ by the average total number of clouds in a subdomain of size L:

$$\mathcal{N}_*(L,l) = \frac{N(L,l)}{\sum_l \overline{N}(L,l)}\, dl^{-1}, \tag{6}$$

Normalized number densities are commonly used in studies of cumulus cloud population, an advantage being that, per definition, the integral of $\mathcal{N}_*$ with l always adds up to 1. As a result, in $(l, \mathcal{N}_*)$ space, the vertical position of the density is preserved for subdomain size, which facilitates their comparison. The two-dimensional PDF is created by sorting the number densities $\mathcal{N}_*(L, l)$ for all subdomains of size L at a given l onto a 100-bin histogram on the vertical axis, covering the range shown in Fig. 5. For a given cloud size l, the shading thus reflects the relative probability of occurrence of a value of $\mathcal{N}_*$ among the subdomains, with red marking the value that occurs most. Two lines are added for reference. The curved line reflects a power-law exponential fit on the full domain CSD as proposed by Ding et al. (2014). The horizontal line indicates the value of a single cloud occurrence in the average subdomain, which is constant for l because of the use of a constant linear bin width dl in the calculation of the CSD. As a result of the use of the normalized number density on the y axis, the single cloud value increases with smaller L because of the decreasing number of clouds present in the subdomain.

Figure 5a shows the analysis for $n = 32$ so that the subdomain equals the full domain. As a result, the plotted data are equivalent to that shown in Fig. 4a. A few aspects catch the eye. First, the power-law exponential fit reproduces the shape of the CSD in this case to a reasonable degree. Second, the horizontal spacing in the left tail of the PDF reflects the horizontal discretization of the LES, allowing only a few specific cloud sizes. Third, the vertical spread in the data increases toward the largest sizes, which is an expression of an increase in the CSD internal variability. Also, the data tend to organize along horizontal lines, with the lowest line of points (with lowest $\mathcal{N}_*$ value) coinciding with the single cloud line. This implies that only one cloud object of this size is present in the domain.

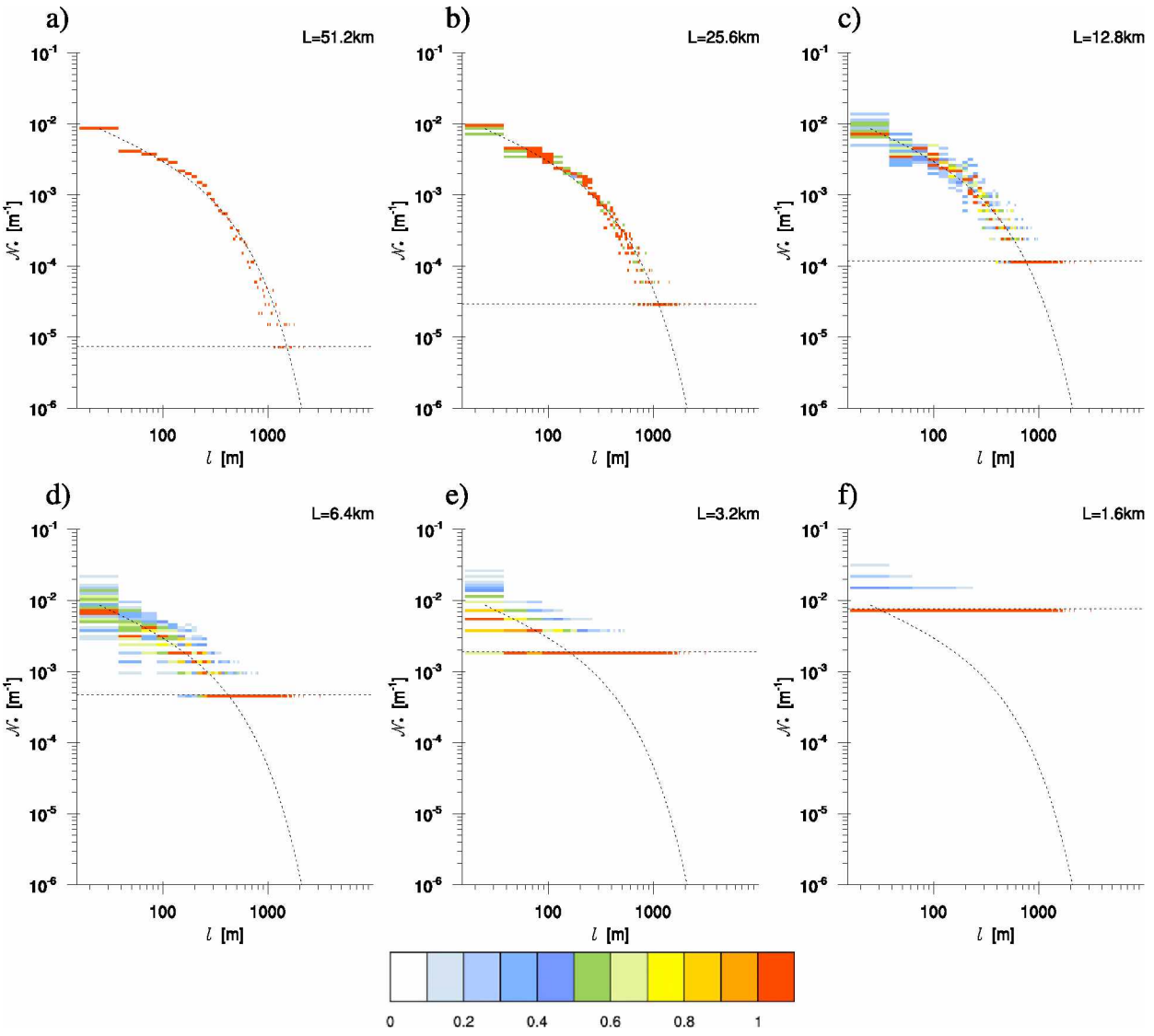

FIG. 5. The normalized cloud number density $\mathcal{N}_*$ for the first RICO snapshot (t = 8 h) calculated for various subdomain sizes. The densities are plotted as two-dimensional PDFs in $(l, \mathcal{N}_*)$ space. The shading indicates the probability of occurrence among the subdomains, normalized such that the most frequently occurring value of $\mathcal{N}_*$ at a given value of l has a value of 1 (red). The curved dotted line is the least squares fit of a power-law exponential function to the full domain CSD, while the horizontal dotted line indicates the value of a single cloud in the average subdomain, calculated as $\left[\sum_l \overline{N}(L,l)dl\right]^{-1}$.

The following panels show the analysis for decreasing subdomain size L, for $n \in \{16, 8, 4, 2, 1\}$. With decreasing L, the single cloud line starts to intersect the full domain fit at ever smaller l. To the right of the intersection size, the points in the PDF are situated at the single cloud line, which thus again acts as a minimum possible value of $\mathcal{N}_*$. For the smallest domain size L = 1.6 km, only N = 2, N = 3, and N = 4 occur. This behavior can be explained by the subdomain becoming comparable or even smaller than the cloud spacing so that sometimes no clouds are present in it. This binary-like occurrence of clouds can significantly boost the variance σ^2 at these sizes. More insight is provided by Fig. 6, showing the probability to encounter a cloud-free subdomain. In general, this probability is highest for the largest clouds and lowest for the smallest, a result that supports this hypothesis. This transition shifts with decreasing subdomain size L, with no-cloud occurrences becoming more frequent at smaller l. At small-to-intermediate L, the transition is gradual, while for large L, the transition is abrupt, almost binary. From these results, one concludes that the no-cloud occurrence and its impact on the internal variability of the CSD is in principle a stochastic behavior, reflecting the subsampling of populations.

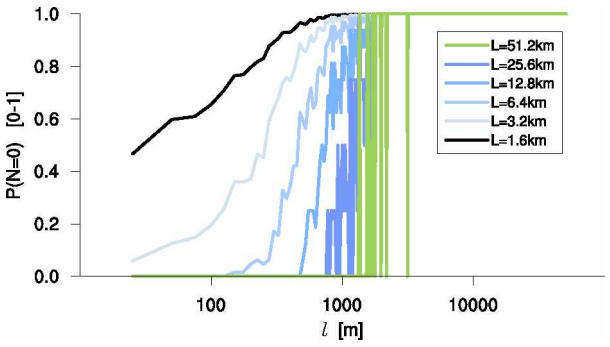

FIG. 6. The probability P that no clouds are present in a subdomain, as a function of cloud size l, for a range of subdomain sizes L for the first snapshot of the RICO case.

b. *Stochastic variability*

The next step is to quantify the internal variability of the CSD due to subsampling as analyzed in the previous section. This is achieved by calculating the variance in cloud number σ^2 for all cloud- and subdomain sizes, as defined by (5). To make sure that the results will only reflect subsampling, again the first RICO time point ($t = 8$ h) is used that shows the least signs of spatial organization. Figure 7a shows σ normalized by the average cloud number per subdomain $\overline{N}$ as a function of subdomain size L. Data points are equidistant in the horizontal because of the set of n values that was chosen. For all cloud sizes, a well-defined monotone relation exists between these variables in log–log space, indicating a strong power-law scaling. Variability in cloud number also increases with cloud size, shifting the relation vertically in a fairly equidistant way, leaving the slope unaffected. In general, the relation is best defined at the smaller domain sizes but shows more noise at the larger subdomain sizes. This is caused by the number of subdomains K as used in the variance calculation decreasing quadratically with domain size (e.g., $K = 4$ at $L = 0.5D$).

The fact that the power-law scaling is similar for all cloud sizes, and that the lines of constant cloud size are more or less equidistant in the vertical, motivates reformulation in terms of two dimensionless groups. Figure 7b shows the same log–log plot but now with subdomain size L divided by cloud size l. This transformation yields a data collapse. A first-order estimation of the slope suggests a power-law exponent $b = -1$, a fit of which indeed captures the dependence to a high degree (dotted line). This dependence reflects an inverse linear relation between the dimensionless groups $\sigma/\overline{N}$ and L/l.

c. *The impact of subsampling*

What could explain the dependence of the CSD variability on both the subdomain size L and the cloud size l? Let us first consider the inverse linear dependence on L. Inverse power-law forms have been found for many natural phenomena, including cloud fields (Newman 2005). The starting point for understanding its appearance in this problem is that the variability σ in the cloud number N present in a subdomain can be expected to be significant when L is of the same order of magnitude as the spacing between the clouds. Consider to this purpose

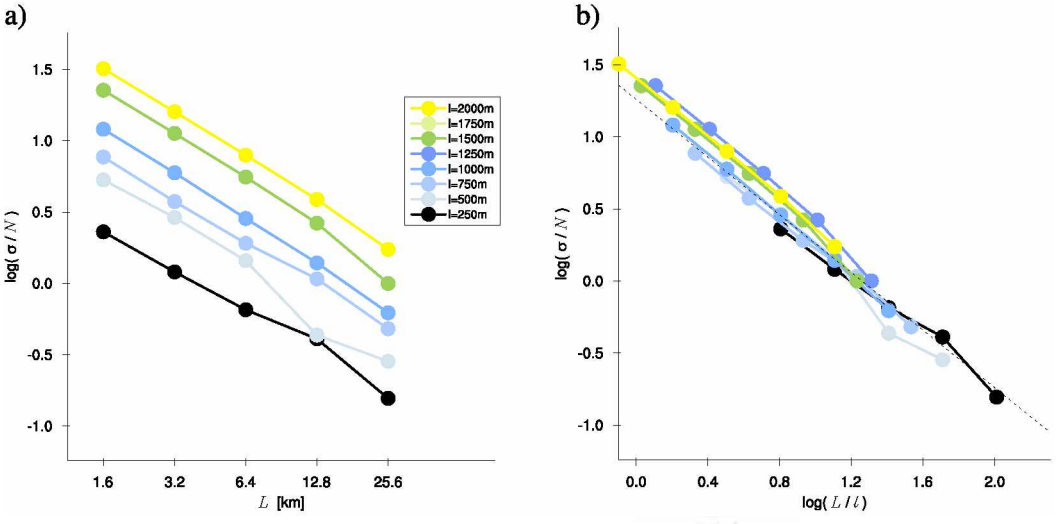

FIG. 7. Log–log plot of the normalized standard deviation in cloud number $\sigma/\overline{N}$ for the first snapshot ($t = 8$ h) of the RICO case, plotted as a function of (a) subdomain size L and (b) subdomain size divided by cloud size L/l. A selection of sizes across the spectrum is shown, as indicated by the colors. The dotted line in (b) represents a fit of the baseline model for the impact of subsampling on cloud number variability with $b = -1$, as discussed in section 3c.

the hypothetical case of a population of clouds of a single size l that is regularly distributed horizontally, with each cloud and the cloud-free area around it occupying a "unit area" of spacing size l_s. This situation is illustrated in Fig. 8. A subdomain is considered that covers a number of cloud-containing units, some of which are only partially covered (illustrated here as one row and one column at the edge). In this situation, the number of clouds in subdomain N_k can be written as

$$N_k = \left(\frac{L}{l_s}\right)^2 \pm O\left(\frac{L}{l_s}\right). \tag{7}$$

The first term on the right-hand side is a *quadratic* function of L and represents the number of clouds included in the subdomain that one expects statistically after many subdomains are considered and is conceptually equal to $\overline{N}$. In contrast, the second term is a *linear* function of L and represents the number of clouds near the subdomain's edges that may or may not be situated inside its area. For the schematic cloud scene shown in Fig. 8, with $L = 3.5l_s$ this gives $N_k = 12.25 \pm 3.5$, implying that the number of clouds in a single arbitrary snapshot can be anywhere between 9 and 16. The larger the L/l_s is, the smaller the remainder term relative to the quadratic term, and the smaller the variation in the estimate of N_k. This corresponds to a less subsampled cloud population. What this implies for the cloud number variance σ^2 can then simply be understood by substituting (7) in (5), which gives

$$\sigma = O\left(\frac{L}{l_s}\right), \tag{8}$$

where the quadratic term has disappeared because of the appearance of $\overline{N}$ in the definition of variance so that only the linear "boundary effect" term remains. Finally, normalizing by the average cloud number $\overline{N}$ yields

$$\frac{\sigma}{\overline{N}} \propto \left(\frac{L}{l_s}\right)\left(\frac{L}{l_s}\right)^{-2} = \left(\frac{L}{l_s}\right)^{-1}. \tag{9}$$

This proportionality explains the inverse linear dependence of the normalized standard deviation in cloud number on the subdomain size that is purely due to subsampling. The power-law exponent of -1 can thus be considered a baseline model for this process.

What is also clear is that the effect of subsampling is first experienced by larger clouds. We speculate that this is explained by the observation that larger clouds typically have larger spacing l_s (Joseph and Cahalan 1990). The relation between l and l_s could carry some dependence on the cumulus regime but also critically depends on how the cloud size l and the neighbor spacing l_s are exactly defined. More research for more cumulus cases is needed to gain insight.

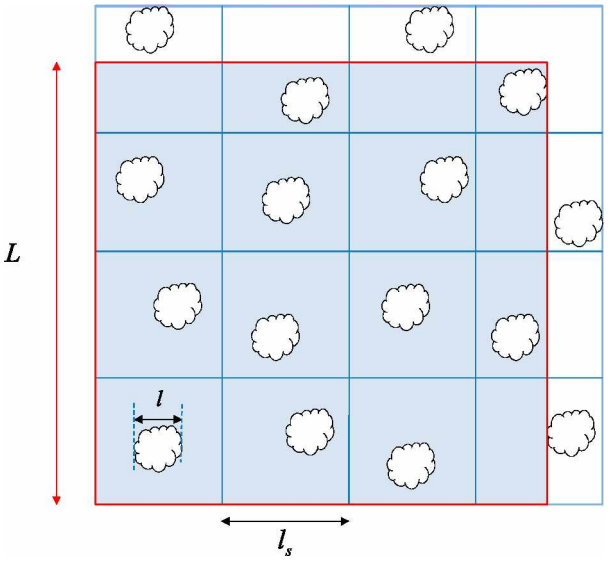

FIG. 8. Schematic illustration of the subsampling of an equally spaced single-size cloud population. The red box indicates the subdomain of size L, while the blue grid boxes represent the spatial unit occupied by a single cloud and its spacing l_s. Variables are explained in the text.

d. *The impact of spatial organization*

The analysis is now repeated for all six RICO fields, which become increasingly organized as time progresses, as shown in Fig. 1. The time evolution of the relation between $\sigma/\overline{N}$ and L is shown in Fig. 9a, first for a single cloud size (l = 25 m). In general, the variability increases with time, already suggesting that some process is affecting the occurrence of the smallest clouds on the grid. Most importantly, the increase in variability is not uniform, but its amplitude increases with subdomain size L. Power-law scaling is still evident at all time points but with an increasingly reduced exponent. This means that, especially at larger L, the variance in the number of small clouds is larger than can be expected from pure subsampling effects, as expressed by the $b = -1$ baseline model.

More insight is provided by Fig. 9b, showing the power-law exponent for all cloud sizes l as a function of time. Two features stand out. First, the maximum cloud size in the domain increases with time, approaching 10 km at the t = {24, 32}-h time points. This reflects the emergence of large cloud structures as visible in Fig. 1, although their number is still very small. Second, the impact of the spatial organization in the population on the power-law exponent appears limited to the smaller

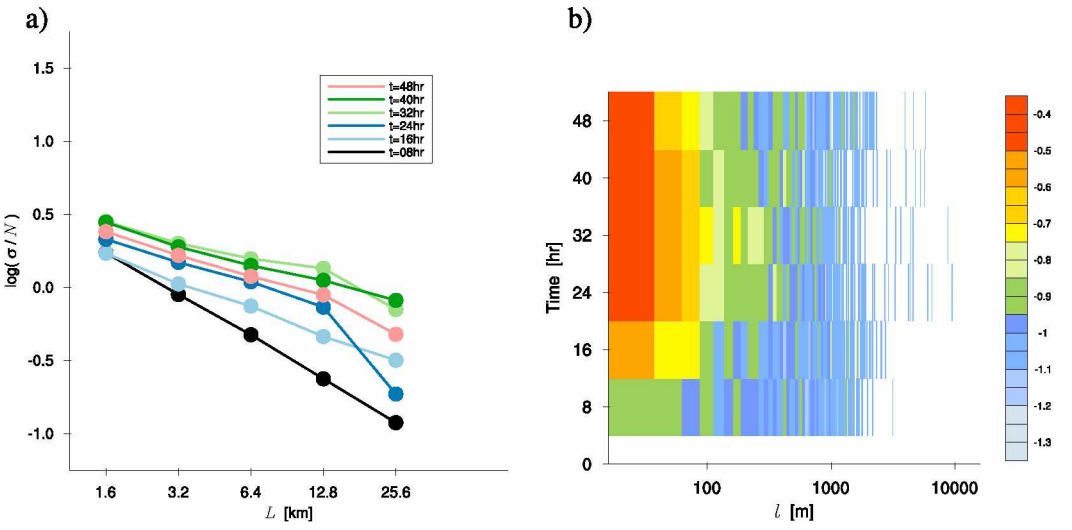

FIG. 9. (a) As in Fig. 7a, but for only a single cloud size $l = 25$ m as a function of time in the RICO case. (b) Power-law exponent b as a function of cloud size and time in the RICO case, resulting from linear fits in $\log(\sigma/\overline{N})$–$\log(L/l)$ space.

end of the distribution and gradually disappears above about $l = 250$ m. These two findings suggest that spatial organization apparently affects both ends of the CSD. On the one end, the size of the largest cloud increases, while on the other, the variability in the number of small clouds is enhanced. This reflects that small clouds then "live" on top of larger-scale thermodynamic structures in the boundary layer, such as cold pools or convergence lines, which favor or inhibit their formation.

To investigate the robustness of the impact of spatial organization on the CSD variability, the analysis is now repeated for all LASSO cases, as shown in Figs. 10a–e. The time evolution of the CSD is similar in all LASSO cases, with a gradually increasing maximum cloud size in the period after cloud onset. A reduction in the power-law exponent at the smaller cloud sizes is also present; however, it is only visible toward the end of the day. A clear difference with the RICO case is the much shorter time span during which this happens. This suggests that the physical process of spatial organization causing this behavior is of a slightly different nature. After inspecting the cloud mask fields in Fig. 2, one notices that, toward the end of the day, small clouds start to surround the bigger clouds in the population. One speculates that this is caused by big clouds falling apart toward sunset, when turbulence dies. In that sense, one expects that this behavior is a typical and prototype feature of diurnal cycles of shallow cumulus over land. This motivates compositing the five cases to enhance the statistical significance, as shown in Fig. 10f. The compositing indeed makes the contoured field smoother. In addition, the compositing emphasizes that the reduction of the power-law exponent in the last hours of convective cloud existence is a robust feature.

4. Discussion

The results obtained so far for the RICO and LASSO cases suggest that the impact of spatial organization on the variability in cloud number is superimposed on the impact of subsampling alone. Two different mechanisms of spatial organization have been encountered in this study, including cold pool dynamics in the RICO case and cloud decay in the LASSO cases. But apparently, the impact of subsampling always takes place in the same way. This is further illustrated by Fig. 11, showing the scaled variances of all time points from all cases in one frame. All cloud sizes $l > 250$ m are included, thus excluding the size range that is most affected by organization in the cases considered. Combining all fields from all cases still yields a data collapse for this size range. A power-law fit is included in Fig. 11 (dashed line), yielding $b = -0.92$, which is very close to the baseline model of $b = -1$ as formulated in section 3c (dotted line). This result suggests that the baseline model for the impact of subsampling on the variability in the cumulus CSD is indeed generally applicable.

This motivates expressing the CSD internal variability in terms of two dimensionless groups:

$$\left(\frac{\sigma}{\overline{N}}\right) = a\left(\frac{L}{l}\right)^{b}, \tag{10}$$

with exponent b written as

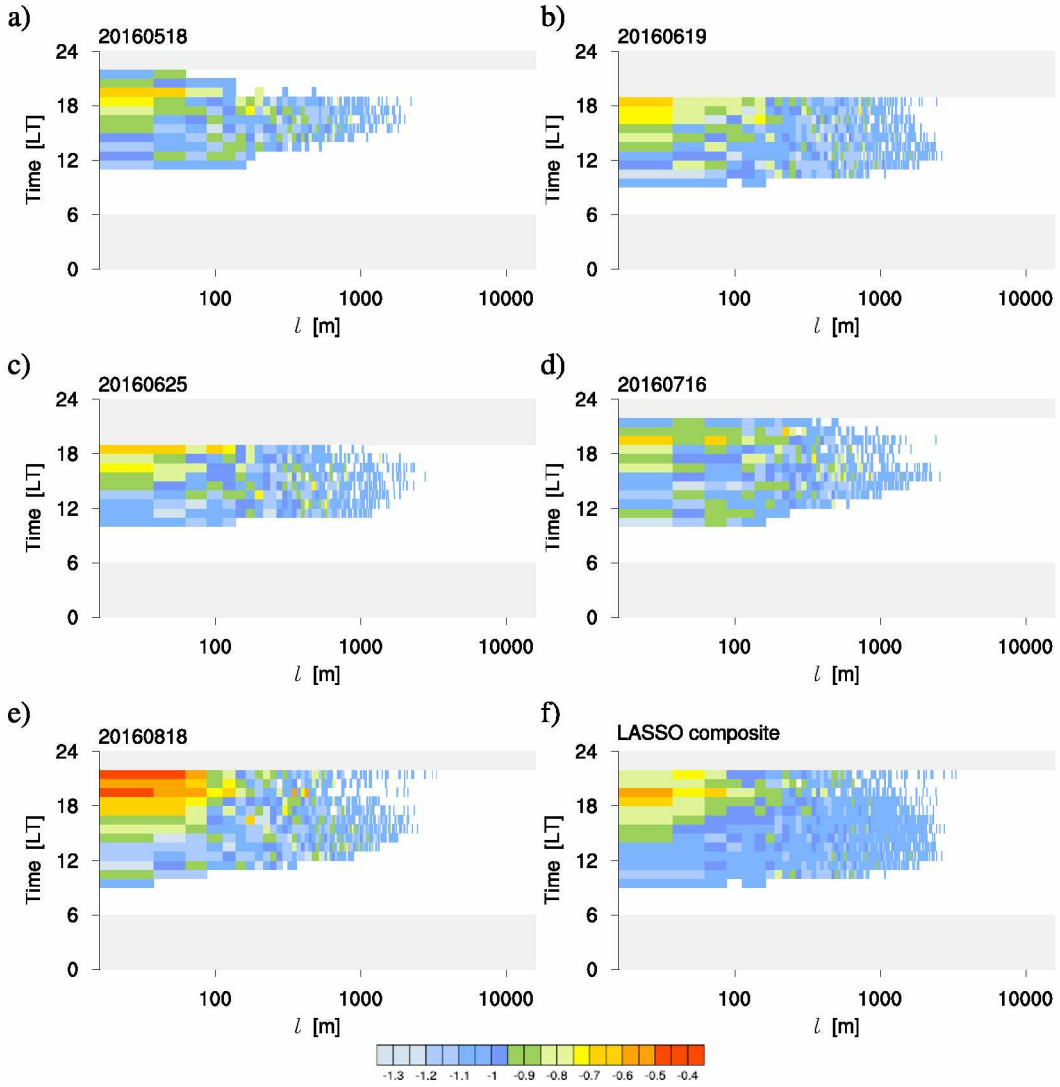

FIG. 10. As in Fig. 9b, but for the LASSO cases (a) 20160518, (b) 20160619, (c) 20160625, (d) 20160716, and (e) 20160818 and (f) a composite average of (a)–(e). Nonsimulated time periods are shaded light gray.

$$b = -1 + \mathcal{F}(\mathrm{Org}). \tag{11}$$

The baseline value of $b = -1$ exclusively reflects the impact of subsampling due to a too-small subdomain size. On top of this baseline exponent, the impact of the degree of spatial organization in the population is then superimposed through an extra term, depending on the degree of organization. While the latter dependence still needs to be established, one option could be the Org parameter as proposed by Mapes and Neale (2011). The parameter a, used here as constant of proportionality to fit the scaling relation (10) to the LES data, effectively translates the relation vertically in log–log space. A shift in this parameter thus reflects an overall change in the variability for all subdomain sizes. Constant a also depends on the bin width of the discretized CSD, as explained in appendix B. Whether a carries an additional dependence on Org is unknown and requires more research.

The scaling relation (10) has various potential uses. For example, it can inform the development of parameterizations in the gray zone of moist convection by acting as a benchmark relation that stochastic parameterizations of cumulus cloud populations have to reproduce. This is in particularly applicable to schemes that are formulated in terms of reconstructed CSDs. Complex population dynamics models might let the number density grow from interaction between sizes, while simpler ones using assumed functional forms for the CSD could have the impact of subsampling and

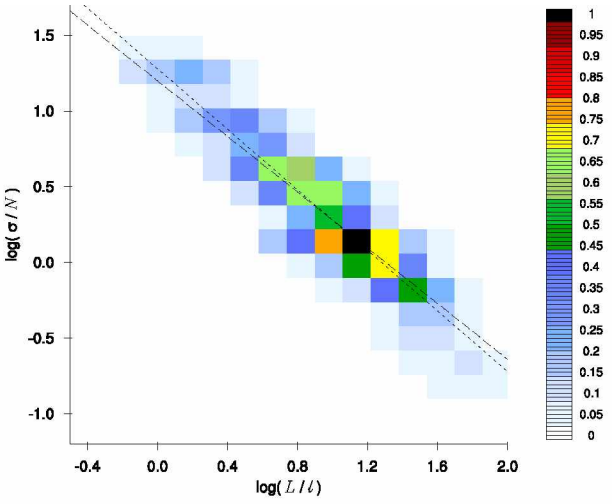

FIG. 11. As in Fig. 7b, but for all snapshots from all six cases. All combinations of $\sigma/\overline{N}$ and L/l are sorted on a 2D histogram covering the parameter space shown above. The shading reflects the probability of occurrence, normalized such that the maximum probability is 1. All clouds with sizes $l > 250$ m are included. The dashed line is a least squares fit of the function $F = F_0 + b\log(L/l)$ to the composite data, yielding $b = -0.92$ at $F_0 = 1.2$. For reference, the fit of the baseline model with prescribed $b = -1$, as already shown in Fig. 7b, is also included (dotted line).

spatial organization as established in this study superimposed. Regardless of the design of the stochastic parameterization, they can be called successful when they reproduce the $b = -1$ scaling for unorganized cloud populations in the gray zone of convection.

The results obtained in this study highlight the important role of cloud spacing in controlling the stochastic variability. Normalizing subdomain size by cloud size yields a data collapse in the scaled variance, argued here to reflect that cloud size is proportional to the cloud spacing. The data collapse suggests that the relation between cloud size and cloud spacing could be pretty robust. More research is needed to confirm and quantify this proportionality. Previous observational studies have indeed hinted at its existence in nature (e.g., Joseph and Cahalan 1990; Sengupta et al. 1990), but typically, these relations still carried considerable scatter. In addition, the definition of cloud spacing used in most previous studies is not equivalent to what is used in this study, which is the spacing between clouds *of equal size*. Exploring both observed and simulated datasets to investigate various definitions of cloud spacing is a future research topic.

This study also sheds more light on the nature of spatial organization in shallow cumulus cloud fields. Two types were encountered in the cases covered, including cold pool formation in the marine case and large cloud decay in the diurnal cycle cases. While this study does not explain how the organization takes place, what the results do reveal is that both types of spatial organization are associated with a similar shift in the power-law exponent in the relation between $\sigma/\overline{N}$ and L/l. Perhaps this shared behavior is something typical for spatial organization. More cases need to be investigated before being able to claim general applicability of this behavior, for example, for deeper precipitating convection.

An interesting idea inspired by these results is to use relation (11) in reverse and objectively estimate the degree of organization $B_{\rm org}$ from the power-law fit in (10) as diagnosed from simulated or observed cumulus cloud populations:

$$B_{\rm org} = b + 1. \tag{12}$$

This would be one possible way to define the degree of spatial organization $B_{\rm org}$, in effect expressing it as the deviation from the inverse linear power law. Clouds of a certain size are unorganized if $b = -1$ (perfectly inverse linear) and become organized as soon as $b > -1$. Figures 9b and 10 could thus be interpreted as representing maps of $B_{\rm org}$ in size–time space. Recently, Tompkins and Semie (2017) proposed a metric for spatial organization $I_{\rm org}$, which is similar to $B_{\rm org}$ in its dependence on the nearest-neighbor spacing between cloudy objects. However, a key difference lies in the exact definition of the spacing; while $I_{\rm org}$ works with the distance to the nearest cloud of *any* size, $B_{\rm org}$ implicitly depends on the distance between clouds of *equal* size through the normalization of L by l. This makes $B_{\rm org}$ particularly applicable for representing stochastic effects in convection schemes predicting the number of cumulus clouds of a certain size.

Some aspects of our analysis could affect the scaling. The first is numerical and concerns the LES discretization, which could artificially affect the variability of the smallest clouds. This is investigated in more detail in appendix A. We find that while the power-law scaling in the cloud number variability (slope) is unaffected, the amplitude increases at cloud sizes smaller than $4\Delta x$. A close proximity to discretization thus slightly enhances the variability in cloud number. Note that these numerically affected clouds were not included in the fitting exercise shown in Fig. 11. Nevertheless, because the spatial organization affects the smallest clouds most, additional simulations at even higher resolution are still advisable to investigate at which point convergence takes place. A second important aspect concerns the definition of cloud size, for which more complex alternatives can be used. For example, cloud objects could be tracked through time, which might change how their variability behaves. Perhaps this study, based on a simple definition of size, can serve as a starting point for such analyses.

5. Summary, conclusions, and outlook

The main results of this study can be summarized as follows:

- Multiple three-dimensional snapshots from various large-domain LESs of marine and continental shallow cumulus cases are used to investigate the behavior of variability in cumulus cloud number due to subsampling and spatial organization.
- Strong power-law scaling is found in the relation between cloud number variability and subdomain size, reflecting an inverse linear relation.
- Cloud spacing as a function of cloud size is a crucial parameter, explaining the data collapse across cases and time points.
- The impact of spatial organization on the variability in the CSD is found to act on top of this baseline model by enhancing its amplitude, in particular at the smaller cloud sizes. The power-law scaling is still preserved but with a reduced exponent. This impact reflects that the small clouds start to live on top of larger-scale thermodynamic structures, favoring or inhibiting their formation.
- A simple expression for the CSD internal variability is proposed that is formulated in terms of two dimensionless groups and that captures the impact of both subsampling and spatial organization.
- This scaling relation provides a new way to objectively estimate the degree of organization $B_{\rm org}$ in simulated or observed cumulus cloud populations as the deviation from a theoretical power-law scaling.
- Compositing all continental cumulus cases suggests the existence of a prototype diurnal time dependence in the spatial organization $B_{\rm org}$, partially reflecting the decay of large convective cumulus clouds toward sunset.

The subdomain analysis adopted in this study needs to be repeated for many more cumulus cloud scenes in order to broaden the parameter space of environmental large-scale conditions and thus enhance the statistical significance of the obtained results. This is required to better calibrate constants of proportionality and also to better understand the nature of spatial organization in cumulus cloud populations. Of particular relevance is the behavior of nearest-neighbor spacing between clouds of a certain size, as this dependence is at the foundation of the scaling relation found in this study. In an ongoing effort by the authors, data from super-large-domain LES experiments with the new Icosahedral Nonhydrostatic (ICON; Zangl et al. 2015) model in the Caribbean dry season are used for this purpose. In addition, existing observational datasets of shallow cumulus cloud populations are revisited to seek observational support for the scaling found in the simulations. These data consist of satellite imagery as well as data from cloud-detecting ground-based imaging and remote sensing instrumentation at the ARM SGP site.

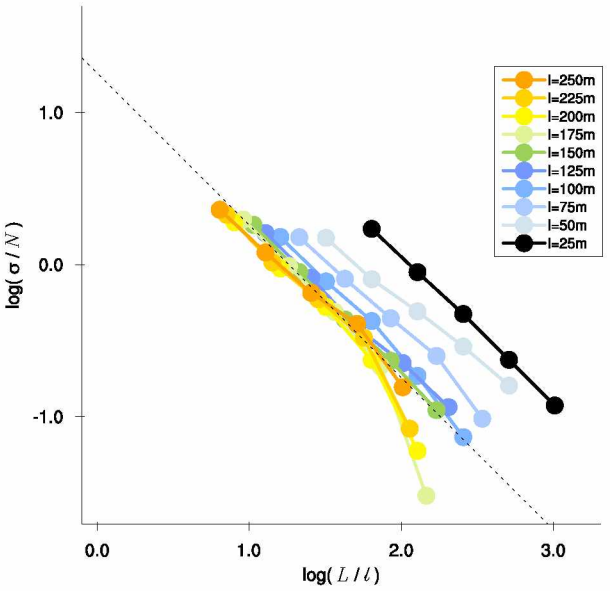

FIG. A1. As in Fig. 7b, but for clouds with $l < 250$ m.

Acknowledgments. This research was supported in part by the U.S. Department of Energy's Atmospheric System Research, an Office of Science, Office of Biological and Environmental Research program, under Grant DE-SC0017999. Data from the DOE as part of the Atmospheric Radiation Measurement (ARM) Climate Research Facility Southern Great Plains site were used to support the simulations. We thank Andy Vogelmann and the LASSO team for providing the forcing data defining the five LES cases at the ARM SGP site and for kindly providing the ARSCL data. The resimulations of the LASSO cases with the MicroHH code were performed using the supercomputing facilities of the Ohio Supercomputer Center (2016). We finally thank three anonymous reviewers for their constructive comments, which greatly helped to improve this manuscript.

APPENDIX A

Impacts of Discretization

In Fig. A1, the σ/N scaling is investigated for cloud sizes smaller than 10 times the LES discretization. For the clouds smaller than about $4\Delta x$, the scaling relation starts to be situated above the baseline model (dotted line) but remains more or less parallel. This implies that, at sizes near the discretization scale, the CSD variability

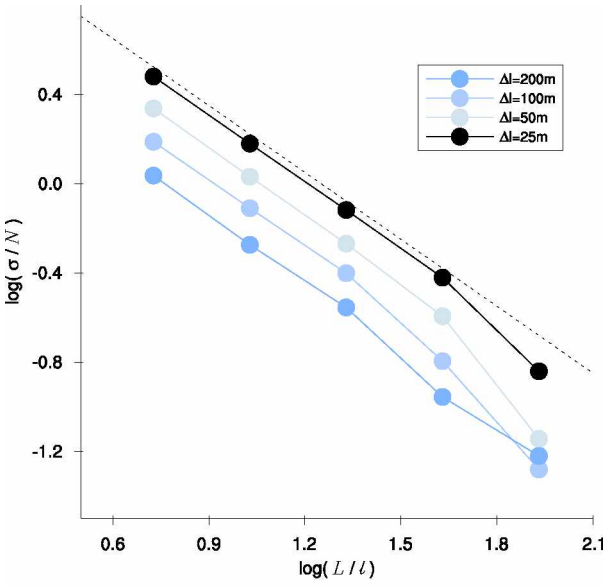

FIG. B1. As in Fig. 7b, but for the sensitivity to CSD bin width Δl. Only the size range $200 < l < 400$ m is considered.

is somewhat larger than can be expected from subsampling alone. We speculate that this numerical artifact reflects on–off cloud behavior in single grid boxes that artificially boosts the variability in their number. However, most importantly, the power-law exponent is more or less unaffected by the proximity to discretization, meaning that the scaling is maintained even for barely resolved clouds.

APPENDIX B

Sensitivity to CSD Bin Width

The bin width that was chosen to discretize the CSD might affect the scaling relation in (10). To examine this possible impact in more detail, the scaling is now recalculated for a specific size range of the CSD in the first snapshot of the RICO case for various bin widths Δl. The size range $200 < l < 400$ m is chosen to this purpose, as it is far enough removed from discretization (see appendix A) but covers sizes still small enough to be well sampled. For $\Delta l \in \{25, 50, 100, 200\}$ m, this yields a number of {8, 4, 2, 1} estimates of relation (10), respectively. The averages over these estimates are shown in Fig. B1. Increasing the bin width shifts the relation downward. This means that while cloud number N increases with larger bin width, the variability in cloud number σ also increases but to a lesser extent. This probably reflects the smaller chance of zero cloud occurrence at a larger cloud number per bin. However, most importantly, the power-law exponent is always preserved and is independent of the bin width.

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.

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.

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.

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**, 1075–1094, https://doi.org/10.1256/003590002320373210.

Cahalan, R. F., and J. H. Joseph, 1989: Fractal statistics of cloud fields. *Mon. Wea. Rev.*, **117**, 261–272, https://doi.org/10.1175/1520-0493(1989)117<0261:FSOCF>2.0.CO;2.

Clothiaux, E. E., T. P. Ackerman, G. G. Mace, K. P. Moran, R. T. Marchand, M. A. Miller, and B. E. Martner, 2000: Objective determination of cloud heights and radar reflectivities using a combination of active remote sensors at the ARM CART sites. *J. Appl. Meteor.*, **39**, 645–665, https://doi.org/10.1175/1520-0450(2000)039<0645:ODOCHA>2.0.CO;2.

——, and Coauthors, 2001: The ARM millimeter wave cloud radars (MMCRs) and the active remote sensing of clouds (ARSCL) value added product (VAP). DOE Tech. Memo. ARM VAP-002.1, 38 pp.

Cohen, B. G., and G. C. Craig, 2006: Fluctuations in an equilibrium convective ensemble. Part II: Numerical experiments. *J. Atmos. Sci.*, **63**, 2005–2015, https://doi.org/10.1175/JAS3710.1.

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. Crommelin, A. Siebesma, and H. 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.

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.

Garrett, T. J., I. B. Glenn, and S. K. Krueger, 2018: Thermodynamic constraints on the size distributions of tropical clouds. *J. Geophys. Res. Atmos.*, **123**, 8832–8849, https://doi.org/10.1029/2018JD028803.

Gustafson, W. I., A. M. Vogelmann, X. Cheng, S. Endo, B. Krishna, Z. Li, T. Toto, and H. Xiao, 2017a: LASSO Alpha 2 data bundles—36°36′18.0″N, 97°29′6.0″W. ARM Research Facility, accessed 23 April 2018, https://doi.org/10.5439/1342961.

——, ——, ——, ——, ——, ——, ——, and ——, 2017b: Description of the LASSO Alpha 1 release. R. Stafford, Ed., ARM Research Facility Rep. DOE/SC-ARM-TR-194, 57 pp., https://doi.org/10.2172/1373564.

Heinze, R., and Coauthors, 2017: Large-eddy simulations over Germany using ICON: A comprehensive evaluation. *Quart. J. Roy. Meteor. Soc.*, **143**, 69–100, https://doi.org/10.1002/qj.2947.

Honnert, R., V. Masson, and F. Couvreux, 2011: A diagnostic for evaluating the representation of turbulence in atmospheric models at the kilometric scale. *J. Atmos. Sci.*, **68**, 3112–3131, https://doi.org/10.1175/JAS-D-11-061.1.

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.

Kassianov, E., C. N. Long, and M. Ovtchinnikov, 2005: Cloud sky cover versus cloud fraction: Whole-sky simulations and observations. *J. Appl. Meteor.*, **44**, 86–98, https://doi.org/10.1175/JAM-2184.1.

Khairoutdinov, M. F., and D. A. Randall, 2006: High-resolution simulation of shallow-to-deep convection transition over land. *J. Atmos. Sci.*, **63**, 3421–3436, https://doi.org/10.1175/JAS3810.1.

Lareau, N. P., Y. Zhang, and S. A. Klein, 2018: Observed boundary layer controls on shallow cumulus at the ARM Southern Great Plains site. *J. Atmos. Sci.*, **75**, 2235–2255, https://doi.org/10.1175/JAS-D-17-0244.1.

Mapes, B., and R. Neale, 2011: Parameterizing convective organization to escape the entrainment dilemma. *J. Adv. Model. Earth Syst.*, **3**, M06004, https://doi.org/10.1029/2011MS000042.

Nair, U. S., R. C. Weger, K. S. Kuo, and R. M. Welch, 1998: Clustering, randomness, and regularity in cloud fields: 5. The nature of regular cumulus cloud fields. *J. Geophys. Res.*, **103**, 11 363–11 380, https://doi.org/10.1029/98JD00088.

Neggers, 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.

——, 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.

Newman, M. E. J., 2005: Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.*, **46**, 323–351, https://doi.org/10.1080/00107510500052444.

Ohio Supercomputer Center, 2016: Owens supercomputer. Ohio Supercomputer Center, http://osc.edu/ark:/19495/hpc6h5b1.

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.

Plank, V. G., 1969: The size distributions 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.

Raga, G., J. Jensen, and M. Baker, 1990: Characteristics of cumulus band clouds off the coast of Hawaii. *J. Atmos. Sci.*, **47**, 338–355, https://doi.org/10.1175/1520-0469(1990)047<0338:COCBCO>2.0.CO;2.

Rauber, R. M., and Coauthors, 2007: Rain in shallow cumulus over the ocean: The RICO campaign. *Bull. Amer. Meteor. Soc.*, **88**, 1912–1928, https://doi.org/10.1175/BAMS-88-12-1912.

Rieck, M., C. Hohenegger, and 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.

——, A. Seifert, and T. Heus, 2015: Fluctuations in a quasi-stationary shallow cumulus cloud ensemble. *Nonlinear Processes Geophys.*, **22**, 65–85, https://doi.org/10.5194/npgd-1-1223-2014.

Satoh, M., T. Matsuno, H. Tomita, H. Miura, T. Nasuno, and S. Iga, 2008: Nonhydrostatic Icosahedral Atmospheric Model (NICAM) for global cloud resolving simulations. *J. Comput. Phys.*, **227**, 3486–3514, https://doi.org/10.1016/j.jcp.2007.02.006.

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.

Schlemmer, L., and C. Hohenegger, 2014: The formation of wider and deeper clouds as a result of cold-pool dynamics. *J. Atmos. Sci.*, **71**, 2842–2858, https://doi.org/10.1175/JAS-D-13-0170.1.

Seifert, A., and K. D. Beheng, 2001: A double-moment parameterization for simulating autoconversion, accretion and selfcollection. *Atmos. Res.*, **59–60**, 265–281, https://doi.org/10.1016/S0169-8095(01)00126-0.

——, and T. Heus, 2013: Large-eddy simulation of organized precipitating trade wind cumulus clouds. *Atmos. Chem. Phys.*, **13**, 5631–5645, https://doi.org/10.5194/acp-13-5631-2013.

Senf, F., D. Klocke, and M. Brueck, 2018: Size-resolved evaluation of simulated deep tropical convection. *Mon. Wea. Rev.*, **146**, 2161–2182, https://doi.org/10.1175/MWR-D-17-0378.1.

Sengupta, S. K., R. M. Welch, M. S. Navar, T. A. Berendes, and D. W. Chen, 1990: Cumulus cloud field morphology and spatial patterns derived from high spatial resolution Landsat imagery. *J. Appl. Meteor.*, **29**, 1245–1267, https://doi.org/10.1175/1520-0450(1990)029<1245:CCFMAS>2.0.CO;2.

Stokes, G. M., and S. E. Schwartz, 1994: The Atmospheric Radiation Measurement (ARM) program: Programmatic background and design of the cloud and radiation test bed. *Bull. Amer. Meteor. Soc.*, **75**, 1201–1222, https://doi.org/10.1175/1520-0477(1994)075<1201:TARMPP>2.0.CO;2.

Tompkins, A. M., and A. G. Semie, 2017: Organization of tropical convection in low vertical wind shears: Role of updraft entrainment. *J. Adv. Model. Earth Syst.*, **9**, 1046–1068, https://doi.org/10.1002/2016MS000802.

van Heerwaarden, C., B. van Stratum, T. Heus, J. Gibbs, E. Fedorovich, and J. P. Mellado, 2017: MicroHH 1.0: A computational fluid dynamics code for direct numerical simulation and large-eddy simulation of atmospheric boundary layer flows. *Geosci. Model Dev.*, **10**, 3145–3165, https://doi.org/10.5194/gmd-10-3145-2017.

vanZanten, 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.

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.

Wyngaard, J., 2004: Toward numerical modeling in the "terra incognita." *J. Atmos. Sci.*, **61**, 1816–1826, https://doi.org/10.1175/1520-0469(2004)061<1816:TNMITT>2.0.CO;2.

Yuan, T., 2011: Cloud macroscopic organization: Order emerging from randomness. *Atmos. Chem. Phys.*, **11**, 7483–7490, https://doi.org/10.5194/acp-11-7483-2011.

Zängl, G., D. Reinert, M. P. Ripodas, and M. Baldauf, 2015: The ICON (Icosahedral Nonhydrostatic) modelling framework of DWD and MPI-M: Description of the nonhydrostatic dynamical core. *Quart. J. Roy. Meteor. Soc.*, **141**, 563–579, https://doi.org/10.1002/qj.2378.

Zhang, Y., S. A. Klein, 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.