

## What Determines the Distribution of Shallow Convective Mass Flux through a Cloud Base?

MIRJANA SAKRADZIJA AND CATHY HOHENEGGER

*Max Planck Institute for Meteorology, Hamburg, Germany*

(Manuscript received 14 November 2016, in final form 8 March 2017)

### ABSTRACT

The distribution of cloud-base mass flux is studied using large-eddy simulations (LESs) of two reference cases: one representing conditions over the tropical ocean and another one representing midlatitude conditions over land. To examine what sets the difference between the two distributions, nine additional LES cases are set up as variations of the two reference cases. It is found that the total surface heat flux and its changes over the diurnal cycle do not influence the distribution shape. The latter is also not determined by the level of organization in the cloud field. It is instead determined by the ratio of the surface sensible heat flux to the latent heat flux, that is, the Bowen ratio  $B$ . This ratio sets the thermodynamic efficiency of the moist convective heat cycle, which determines the portion of the total surface heat flux that can be transformed into mechanical work of convection against mechanical dissipation. The thermodynamic moist heat cycle sets the average mass flux per cloud  $\langle m \rangle$ , and through  $\langle m \rangle$  it also controls the shape of the distribution. An expression for  $\langle m \rangle$  is derived based on the moist convective heat cycle and is evaluated against LES. This expression can be used in shallow cumulus parameterizations as a physical constraint on the mass flux distribution. The similarity between the mass flux and the cloud area distributions indicates that  $B$  also has a role in shaping the cloud area distribution, which could explain its different shapes and slopes observed in previous studies.

### 1. Introduction

Since the seminal work on parameterization of cumulus clouds by [Arakawa and Schubert \(1974, hereafter AS-74\)](#), the understanding of the spectral distribution of cloud properties and how it is controlled by the large-scale environment remains an obstacle for the formulation of convection parameterizations. In their paper, [AS-74](#) (p. 687) wrote, “Our final problem is to find the mass flux distribution function. The real conceptual difficulty in parameterizing cumulus convection starts from this point. We must determine how the large-scale processes control the spectral distribution of clouds, in terms of the mass flux distribution function, if they indeed do so. This is the essence of the parameterization problem.” With this in mind, it is the goal of our paper to determine how the mass flux distribution of shallow cumulus clouds  $p(m)$  is controlled by the underlying physical processes and large-scale conditions.

In the formulation of the [AS-74](#) parameterization, the mass flux distribution function refers to the spectral

distribution of cloud subensembles. The subensembles encompass clouds of different types based on their sizes and cloud-top heights. This distribution is estimated in [AS-74](#) by numerical solution of the Fredholm integral equation assuming convective quasi equilibrium (QE). Here, we instead regard the mass flux distribution as an asymptotic distribution of the spectral subensembles that are reduced to single clouds, which then can be classified as a cloud population distribution. In this way, we approach the problem from another point of view: instead of assuming convective QE and solving for the spectral distribution of mass fluxes numerically, we focus on the underlying physical principles that determine the shape of  $p(m)$  and its parameters.

The decision to examine the population distribution  $p(m)$  instead of the spectral distribution based on cloud types comes from the need to formulate a scale-aware parameterization. As the model resolution increases to the kilometer scale, the separation of the cloud ensemble into spectral bins that represent clouds of different types loses statistical significance. Instead, a cloud sample within a grid box can be viewed as a random sample of clouds drawn from the cloud population. The clouds are grouped by the gridbox boundaries regardless

---

*Corresponding author:* Mirjana Sakradzija, mirjana.sakradzija@mpimet.mpg.de

of the cloud types. The total mass flux in a grid box  $M$  is then a sum over the sampled clouds,  $M = \sum_{i=1}^n m_i$ , and its distribution  $p(M)$  is characterized by a spectrum of shapes starting from a normal-like distribution on the coarse grids, toward a long-tailed distribution on the kilometer-scale grids (Craig and Cohen 2006; Sakradzija et al. 2015). The distribution of the total mass flux within model boxes  $p(M)$  has been parameterized based on the principles of statistical mechanics and has been applied to deep convection by Plant and Craig (2008) and further developed to a parameterization of shallow convection by Sakradzija et al. (2015, 2016). In the context of such a parameterization, it is important to understand the physical constraints on  $p(m)$  because fluctuations of the subgrid-scale convective tendencies influence convective regimes and their organization, as well as the energetics of the explicitly modeled atmospheric flows (Sakradzija et al. 2016).

The evidence about  $p(m)$  based on observations is not extensive. A few observational studies that examined  $p(m)$  among other cloud statistics were focused on cumulonimbus clouds, for which  $p(m)$  was fitted to a log-normal distribution function (LeMone and Zipser 1980; Jorgensen and LeMone 1989). More evidence about  $p(m)$  has been provided by modeling studies using cloud-resolving models (CRM) or large-eddy simulations (LESs). In a CRM study of an equilibrium deep-convective ensemble under homogeneous large-scale forcing  $p(m)$  was fitted to an exponential function (Cohen and Craig 2006). This fit was supported by theoretical derivation using the formalism of the Gibbs canonical ensemble from statistical mechanics (Craig and Cohen 2006). As more computing power allowed performing simulations with resolutions on the order of 100 m, it was revealed that the shape of this distribution is dependent on the horizontal resolution. With kilometer-scale resolution, where the deep cumulus clouds are not fully resolved,  $p(m)$  takes an exponential-like shape, while the shape changes toward a power-law distribution when using higher resolution (Scheufele 2014). Scheufele (2014) further demonstrated that the power-law-like shape emerges as a result of self-organization of the individual cloud updrafts.

For shallow cumulus clouds over the ocean, Sakradzija et al. (2015) found that the overall shape of the mass flux distribution results from the superposition of two distribution modes: one corresponding to the active buoyant clouds and the other one to nonbuoyant clouds. The two modes of the cumulus cloud distribution deviate from an exponential shape as a result of correlation between cloud mass fluxes and cloud lifetimes. Each mode can be described using a Weibull distribution with two parameters, shape  $k$  and scale  $\lambda$  [see Eq. (13)

and also Sakradzija et al. (2015)]. In the case of shallow cumulus clouds, the shape parameter of the Weibull distribution is less than one,  $k < 1$ , which signifies that it is a heavy-tailed distribution. The combination of at least two Weibull distribution modes results in a distribution of the shallow cumulus mass flux that takes an overall power-law-like shape (see section 6). Hence, it appears that different mechanisms can lead to power-law distributions (see, e.g., Mitzenmacher 2003; Newman 2005). Moreover, either a power-law or a lognormal distribution can be generated by the same underlying mechanism under slightly different conditions (e.g., Mitzenmacher 2003), and it is often difficult to rule out one or the other functional form.

It might be possible to gain more insight into the mass flux distribution  $p(m)$  by making a parallel to the distribution of cloud sizes. Based on the findings of modeling and observational studies, there is no consensus on the functional form that best describes the cloud size distribution. The suggested functions span from exponential (Plank 1969; Hozumi et al. 1982; Astin and Latter 1998), to log-normal (López 1977; LeMone and Zipser 1980; Jorgensen and LeMone 1989), to power-law functions with single (Lovejoy 1982; Zhao and Di Girolamo 2007; Wood and Field 2011; Dawe and Austin 2012) or double slopes (Cahalan and Joseph 1989; Sengupta et al. 1990; Nair et al. 1998; Benner and Curry 1998; Neggers et al. 2003; Trivej and Stevens 2010; Heus and Seifert 2013). Most studies, in particular more recent ones, suggest power laws, with or without a break in the power-law scaling at the intermediate cloud sizes. This scale break manifests itself as a change in the slope of a power-law distribution or as an exponential cutoff near the distribution tail. However, no explanation supported by evidence has been provided for the observed differences in the distribution shapes and slopes, and some of these differences may just reflect different meteorological conditions.

Given that the characteristics of cloud updrafts are substantially different between tropical oceanic and midlatitude continental cumulus convection (Xu and Randall 2001), the dependency of  $p(m)$  on meteorological conditions is not surprising. We nevertheless suspect that there are some dominant macroscopic parameters or processes that determine the characteristic cloud size and the mass flux that cause the variations in  $p(m)$  between different cases and locations. Instead of assuming a distribution functional form and estimating the distribution parameters by statistical fitting of modeled or observed clouds, we set out to identify the physical mechanisms that might lead to a specific distribution functional form and a characteristic scale. We use LES of shallow cumulus convection based on two

measurement campaigns, Rain in Cumulus over the Ocean (RICO) to represent conditions over the ocean and measurements in an Atmospheric Radiative Measurement (ARM) site to represent conditions over land (section 2). We aim to reveal what makes the difference in  $p(m)$  between these two reference cases and to derive a parameterization for the distribution parameters that applies to oceanic and land conditions.

In nine additional simulations, the two reference cases are modified (see section 2) to test the impacts of the large-scale forcing and surface conditions on  $p(m)$ . Cloud life cycles are studied using the method of cloud tracking, also described in section 2. This method provides the lifetime averaged cloud mass flux distribution defined in section 3. Several reasons for the difference in  $p(m)$  between the two reference cases are hypothesized and tested in section 4. In section 5 we describe the physical principle that explains the difference between the two characteristic distribution shapes. The distribution is then fitted to the mixed Weibull function to estimate the remaining unknown parameters (section 6). Conclusions are given in section 7.

## 2. LES case studies

Simulations were performed using the University of California, Los Angeles, large-eddy simulation (UCLA-LES) model (Stevens et al. 1999; Stevens 2010). A detailed description of the UCLA-LES model and the specification of the parameters and constants used in our study are provided in Stevens (2010). The UCLA-LES model solves the Ogura–Phillips anelastic equations, discretized over the doubly periodic uniform Arakawa C grid. The prognostic variables include the wind components  $u$ ,  $v$ , and  $w$ ; liquid water potential temperature  $\theta_l$ ; total water mixing ratio  $q_t$ ; and, in the precipitating cases (see the next paragraph), rain mass mixing ratio  $q_r$  and rain number mixing ratio  $N_r$ . In the precipitating cases, the double-moment warm-rain scheme of Seifert and Beheng (2001) is used to compute the cloud microphysics. The subgrid turbulent fluxes are computed using the Smagorinsky–Lilly scheme [as described in Stevens et al. (1999) and Stevens (2010)]. A third-order Runge–Kutta method is used for numerical time integration, a directionally split monotone upwind scheme is used for the advection of scalars, and a directionally split fourth-order centered scheme is used for the momentum advection (see Stevens 2010). The effects of radiation are prescribed as net forcing tendencies.

As a first reference case (R-base), an LES case study of shallow convection based on the RICO measurement campaign (Raubert et al. 2007) is used to represent conditions over the tropical ocean. The field

measurements were taken during the winter season 2004/05 in the trade wind region of the western Atlantic upwind of the islands of Antigua and Barbuda (Raubert et al. 2007). The initial profiles of potential temperature  $\theta$ , specific humidity  $q_v$ , and the horizontal winds  $u$  and  $v$  are constructed as piecewise linear fits of the averaged profiles from the radiosonde measurements taken over Barbuda during a period with no disturbance due to mesoscale convective systems [Fig. 2 and Table 2 in van Zanten et al. (2011)]. Vertical time-invariant profiles of the subsidence rate and of horizontal advection of moisture and temperature are prescribed and act on the thermodynamic quantities at each time step [Table 2 in van Zanten et al. (2011)]. The radiative and advective cooling rates are prescribed jointly as a large-scale vertically homogeneous cooling rate profile of  $2.5 \text{ K day}^{-1}$ . The sea surface temperature is set to  $299.8 \text{ K}$ , while the surface fluxes are computed interactively using a surface-layer bulk aerodynamic parameterization (see van Zanten et al. 2011). The geostrophic wind profiles are prescribed as time invariant and equal to the initial wind profiles, and the background wind is set to  $u = -5 \text{ m s}^{-1}$  and  $v = -4 \text{ m s}^{-1}$ . Duration of the R-base simulation is 60 h.

To represent conditions over land, a second reference case (A-base) is set up based on the ARM campaign, as in Brown et al. (2002). This case is forced by the averaged observed conditions at the Southern Great Plains (SGP) site on 21 June 1997. The start of the simulation is set to 1130 UTC (0630 local time), a time before convection initiates, and is integrated over a single diurnal cycle until 0200 UTC the next day (2100 local time). The initial vertical profiles of the thermodynamics quantities are computed based on the averaged soundings from that day (Fig. 1 in Brown et al. 2002). The wind direction did not change significantly during that day, so the initial wind profile is set to a constant wind of  $u = 10 \text{ m s}^{-1}$  and  $v = 0 \text{ m s}^{-1}$  at all levels. The geostrophic wind is also set to these values, while the background wind is set to  $u = 0 \text{ m s}^{-1}$  and  $v = 7 \text{ m s}^{-1}$ . At the surface, the turbulent heat fluxes are prescribed following Brown et al. (2002; see their Fig. 3) and exhibit a strong diurnal cycle. Weak large-scale forcing tendencies due to horizontal advection of moisture and temperature as well as radiative cooling rates are prescribed following the diurnal cycle; however, they have only a minor impact on the simulation.

The two reference LES cases, R-base and A-base, are further modified to test the effects of surface conditions, diurnal cycle, and large-scale forcing on the cloud statistics (Table 1). For all LES cases, the simulations are performed over a domain of  $51.2 \text{ km} \times 51.2 \text{ km}$ , with a horizontal grid spacing of 25 m and a vertical resolution

TABLE 1. List of the LES cases with the abbreviations used in the text, the case on which the simulations are based, the maximum Bowen ratio  $B_{\max}$ , the total surface turbulent heat flux averaged over the simulation period  $\overline{F}_{\text{in}} = \overline{F}_{\text{sh}} + \overline{F}_{\text{lh}}$ , and the type and duration of the large-scale forcing  $\text{LS}_{\text{forc}}$ .

Abbreviation	Reference case	$B_{\max}$	$F_{\text{in}}$ ( $\text{W m}^{-2}$ )	$\text{LS}_{\text{forc}}$
R-base	RICO	0.06	171	Const
R-0.24	RICO	0.24	152	Const
R-0.33	RICO	0.33	152	Const
A-base	ARM	0.36	343	14 h 30 min
A-0.5	ARM	0.50	340	14 h 30 min
A-0.1	ARM	0.11	347	14 h 30 min
A-0.06	ARM	0.06	348	14 h 30 min
A-0.03	ARM	0.03	349	14 h 30 min
A-lowflx	ARM	0.36	274	14 h 30 min
A-short	ARM	0.36	341	10 h
A-long	ARM	0.36	344	19 h

of 25 m up to a height of 5 km (domain top). Five vertical grid levels are used as damping layers at the top of the domain.

In the first group of simulations (R-base, R-0.24, R-0.33, A-base, A-0.5, A-0.1, A-0.06, and A-0.03; Fig. 1), we have prescribed a range of values of the ratio of the sensible to latent heat fluxes at the surface, the Bowen

ratio, to both cases starting from  $B = 0.03$  to  $B = 0.5$ . This range of values is selected because it encompasses the typical values of  $B$  characteristic for the regions of the tropical oceans to midlatitude continental conditions. The purpose of these simulations is to investigate the hypothesis that the differences between the two reference cases come from different Bowen ratios. The average Bowen ratio in R-base is around 0.03 and is approximately constant, while in A-base the starting value of  $B$  is around 0.3, and it decreases slightly over the diurnal cycle (Fig. 1a). The two simulations based on RICO, R-0.24 and R-0.33, are set up by fixing the surface heat fluxes instead of the fixed SST. The total heat flux magnitude is kept equal to the reference RICO case, but the ratio of sensible to latent heat flux is changed to result in the wanted  $B$  value, 0.24 in the first and 0.33 in the second case. In the ARM-based cases (A-0.5, A-0.1, A-0.06, and A-0.03), the total surface heat fluxes are kept the same, but the ratio of sensible to latent heat flux is changed to result in the targeted  $B$  values of 0.5, 0.1, 0.06, and 0.03. These new  $B$  values are set at the beginning of the diurnal cycle and are decreasing over the cycle at the same rate as in A-base (see Fig. 1a). Note that the total surface heat flux in the RICO-based cases is,

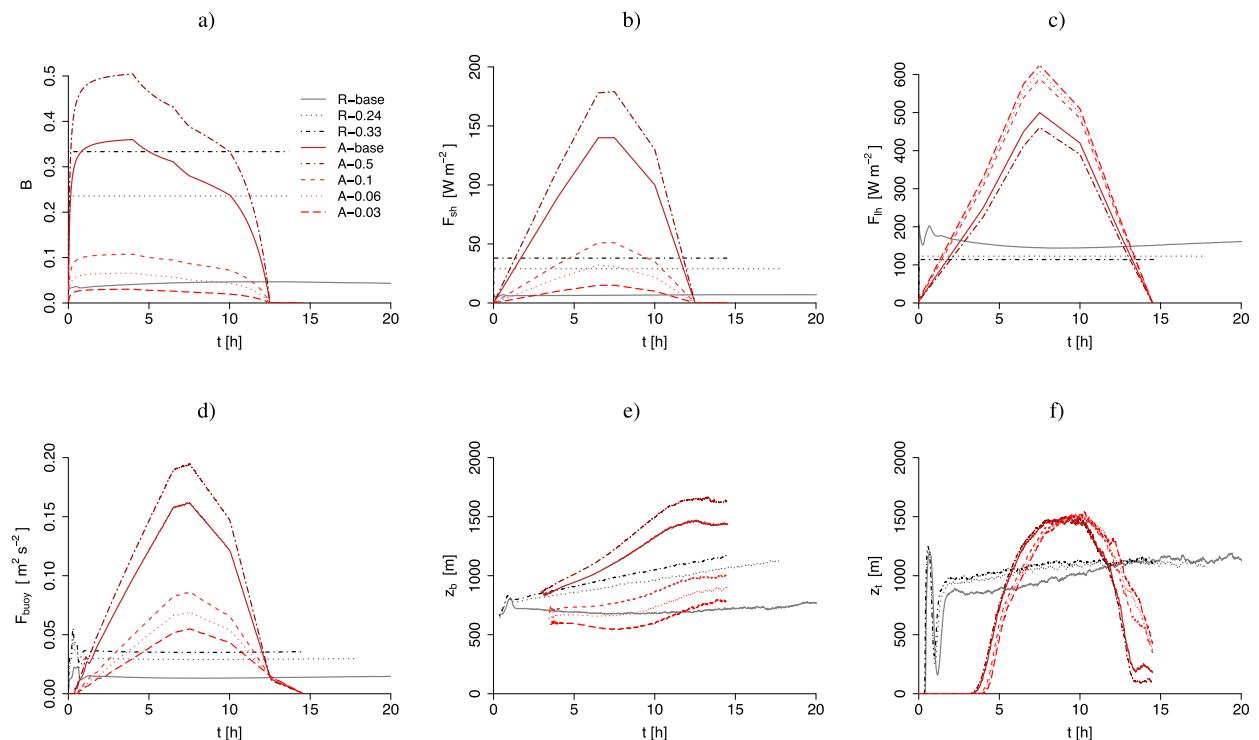


FIG. 1. Time series of the surface forcing in the first group of eight LES cases from Table 1: (a) Bowen ratio  $B$ , (b) surface sensible heat flux  $F_{\text{sh}}$ , (c) surface latent heat flux  $F_{\text{lh}}$ , (d) surface buoyancy flux  $F_{\text{buoy}}$ , and the resulting (e) cloud-base  $z_b$  and (f) cloud-top heights  $z_t$ . The difference between these simulations is set through the Bowen ratio, which is indicated in the case abbreviations and line colors: R-base (black solid), R-0.24 (black dots), R-0.33 (black dot-dash), A-base (red solid), A-05 (red dot-dash), A-0.1 (red short dash), A-0.06 (red dot), and A-0.03 (red long dash). Time from the start of the simulation is shown on the x axis.

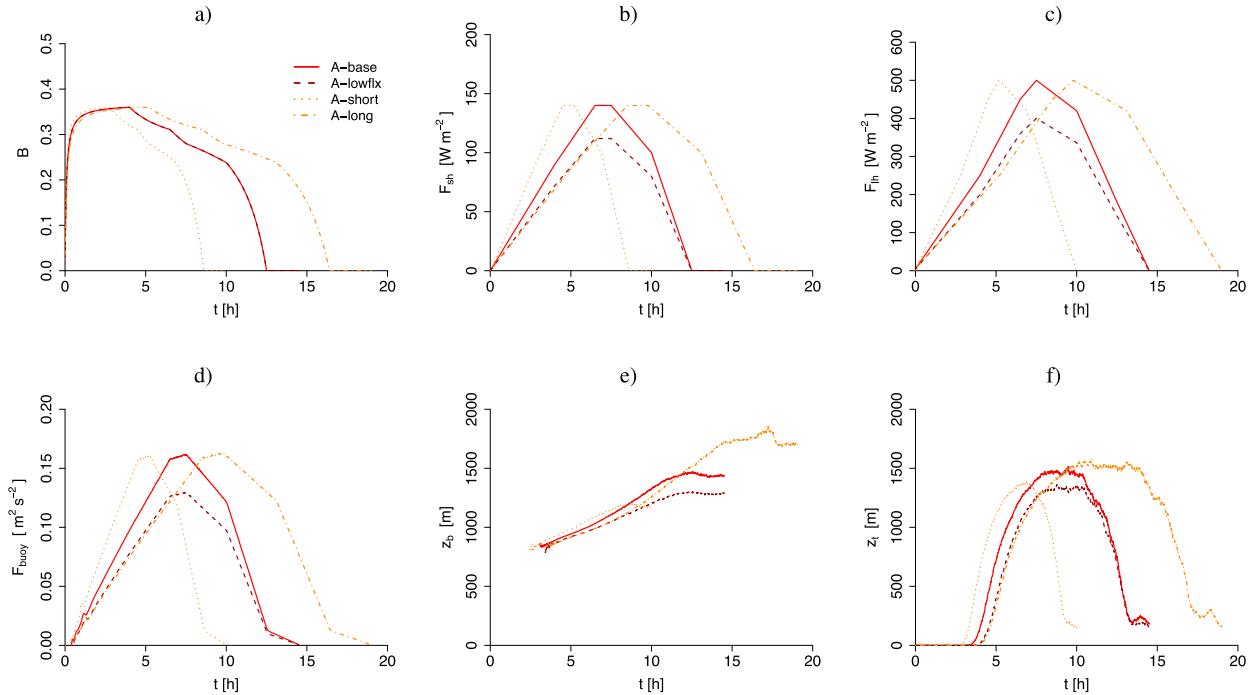


FIG. 2. As in Fig. 1, but for forcing in the second group of the LES cases based on ARM: A-base (red solid), A-lowflx (dark red dashes), A-short (orange dots), and A-long (orange dot-dash). The difference between these simulations is set through the period of the large-scale forcing (A-short and A-long) and through the total surface heat flux magnitude (A-lowflx).

in average, more than twice lower than the total surface heat flux in the ARM-based cases (Table 1). By comparing the maximum values of the total surface heat flux or of the buoyancy flux near the peak of the diurnal cycle (Figs. 1b–d), the difference between the two reference cases is even up to 4 times.

As expected, the mean thermodynamic state of the subcloud layer is affected by the changes in the Bowen ratio. Increase of the Bowen ratio from 0.03 to 0.33 in the RICO-based cases causes an increase of the liquid water potential temperature by 1 K and a decrease in the total water mixing ratio by  $1 \text{ g kg}^{-1}$ , as averaged over a 500-m-thick layer starting from the surface. In the ARM case, a decrease of the Bowen ratio from 0.33 in A-base to 0.03 in A-0.03 causes a decrease of the liquid water potential temperature by 2 K, and an increase of the total water mixing ratio by  $2 \text{ g kg}^{-1}$ , averaged over a 500-m-thick layer at the surface. Clearly, all these test cases have a different thermodynamic state in the boundary layer, even though the Bowen ratios might have the same values.

The depth of the subcloud layer is controlled by the surface buoyancy flux  $F_{\text{buoy}}$  (Stevens 2007) with the higher cloud-base heights in the simulations with higher surface buoyancy fluxes (Figs. 1d,e). The rate of growth of the subcloud layer is also influenced by  $B$ , and it is higher in the cases with higher  $B$  [Fig. 1e; see

also Schrieber et al. (1996)]. Convective clouds are initiated sooner for the higher values of  $B$  (Fig. 1f). Except for the R-base case where the surface fluxes are not fixed, the top of the cloud layer does not seem to be significantly influenced by the changes in  $B$  or  $F_{\text{buoy}}$  (Fig. 1f). This indicates that the processes in the cloud layer are to some extent detached from the surface forcing.

In the second group of simulations (A-lowflx, A-short, and A-long; Fig. 2), we have kept the Bowen ratio to its assigned values, but changed other key aspects of the forcing that are distinct between the two reference cases. The effect of the diurnal cycle in ARM is tested by shortening it by 1/3 (A-short), or by prolonging it by 1/3 (A-long), by applying these changes to the cycle period of the surface fluxes (see Figs. 2b,c) and the large-scale forcing tendencies. The effect of the value of the total surface heat flux is tested by reducing it by 20% in ARM (A-lowflx). As can be seen in Fig. 2e, the rate of the growth of the cloud-base height is not affected by these changes. However, if there is more time for the cloudy boundary layer to develop, as in A-long, a higher cloud-base height is reached. The cloud layer deepens further either with an increase in the forcing period or with stronger total surface heat fluxes, although the differences are only around 100 m (Fig. 2f).

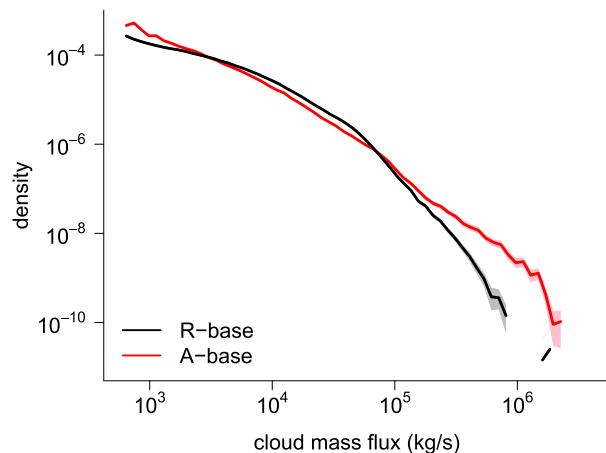


FIG. 3. The probability density distribution of the lifetime average cloud-base mass fluxes. Clouds are sampled from hour 6 to hour 22 from the simulation start in the R-base case; and from hour 9 to hour 12 after the simulation start in the A-base case. Clouds with mass flux values  $< 600 \text{ kg s}^{-1}$  are discarded from the plot to remove possible numerical noise, since those are mostly the clouds that cover only a single grid cell. The 95% confidence bands are plotted as shaded areas.

### Cloud tracking

The cloud tracking algorithm developed by Heus and Seifert (2013) is applied to the simulated cloud fields in postprocessing of the LES. In the tracking algorithm, clouds are identified as the adjacent grid points that hold the liquid water path exceeding a threshold value of  $5 \text{ g m}^{-2}$ . In that way, the identified cloud area is a projection of a cloud from all vertical levels that can be tracked through space and time. Using the temporal resolution of 1 min, cloud areas, vertical velocities, and cloud lifetimes are recorded for each cloud in the simulation. A cloud splitting algorithm is then used to separate and track the individual cloud elements that form the multicore clouds or the merged cloud clusters. These cloud elements are defined as holding a buoyant core with the maximum in-cloud virtual potential temperature  $\theta_v$  excess larger than a chosen threshold of 0.5 K. More details and validation of the tracking method are provided in Heus and Seifert (2013).

To develop a cloud parameterization based on the mass flux approach, the cloud mass flux has to be estimated near the cloud-base level. For this reason, we have developed a secondary tracking routine, as in Sakradzija et al. (2015), in which we record the area that every cloud occupies at the level that lies 100 m above the lifting condensation level (LCL). We define this area as the area that contains all the points with liquid water content greater than zero.

### 3. Cumulus cloud population statistics

The upward flux of mass through cloud base of the  $i$ th cloud is defined as

$$m_i = \rho a_i w_i, \quad (1)$$

where  $a_i$  is the area ( $\text{m}^2$ ) occupied by points holding liquid water at a level 100 m above LCL and  $w_i$  ( $\text{m s}^{-1}$ ) is the vertical velocity averaged over the area  $a_i$ . To compute the distribution of the cloud mass flux  $p(m)$  we average  $m_i$  over the lifetime of each cloud. Similar results can nevertheless be obtained by looking at the instantaneous values. The choice of computing the lifetime averaged mass fluxes comes from the possibility to reconstruct cloud life cycles for the purpose of a parameterization, as in Sakradzija et al. (2015).

The distribution of cloud-base mass fluxes is calculated for the two reference cases, RICO and ARM (Fig. 3). The probability density distribution is computed using the generic R function hist (R Core Team 2015). The width of the bins used to compute the probability density of mass fluxes is logarithmically increasing with higher mass flux values. The sampling period in RICO is from hour 6 to hour 22 of the simulation, while in the ARM case clouds are sampled from hour 6 (1730 UTC) to hour 12 (2330 UTC) of the simulation. Only those clouds that were initialized during the sampling period are included in the calculation. Clouds that lasted longer than this sampling time period are followed beyond the time limit to finalize their life cycles. The sample size of the lifetime average cloud-base mass flux distribution is 317 014 clouds in the RICO case and 120 292 clouds in the ARM case.

The two reference LES cases exhibit distinct horizontal and vertical extents of the clouds, number of clouds, and their spacing because of different initial conditions, surface, and large-scale forcing. The mass flux distributions corresponding to these two reference cases have different shapes, and they cover different ranges of the mass flux values (Fig. 3). The distribution of the cloud-base mass flux in the ARM case shows a straight-line shape on a log-log plot, similar to a power-law distribution over a range of three orders of magnitude. In contrast, the distribution in the RICO case shows a more concave shape. In previous literature on the cloud size distribution, such a type of a concave shape has often been identified as a double power-law distribution with two distinct slopes and a scale-break point at the intermediate cloud size (Cahalan and Joseph 1989; Sengupta et al. 1990; Nair et al. 1998; Benner and Curry 1998; Neggers et al. 2003; Trivej and Stevens

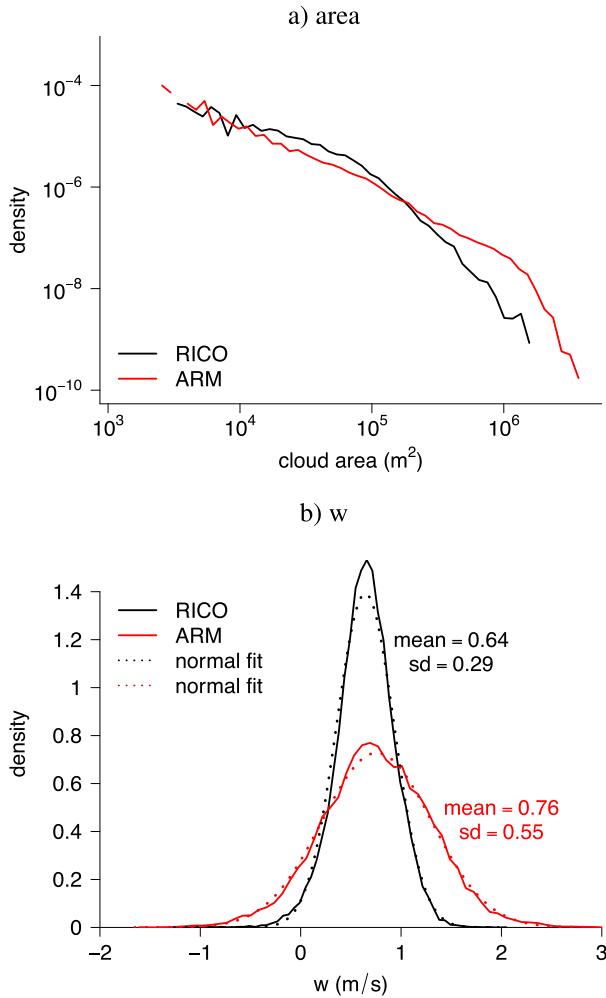


FIG. 4. The probability density distribution of (a) lifetime averaged cloud-base areas and (b) vertical velocity ( $w$ ) through cloud base for RICO (black) and ARM (red). As in the previous figure, clouds with mass flux values  $< 600 \text{ kg s}^{-1}$  are discarded from the plot to remove possible numerical noise.

2010; Heus and Seifert 2013). To make a parallel to these studies, we identify the scale break in the mass flux distribution of the R-base case at a value of the cloud-base mass flux close to  $1 \times 10^5 \text{ kg s}^{-1}$  (Fig. 3). Based on the qualitative comparison of the mass flux distributions of the R-base and A-base case, we conclude that there is no universality in the distribution slopes on a log-log plot (Fig. 3). As we will show in section 4c, the slope of the mass flux distribution changes with the change of a control parameter of the simulations.

The sampling variability of the mass flux distributions is very low in both reference cases except near the end of the right tails of the distributions (Fig. 3), which is a sign of a limited sample size of the largest possible cloud mass flux values. This portion of the distribution tail has higher

sampling variability based on the 95% confidence intervals computed for each distribution bin (shaded areas in Fig. 3). The confidence intervals were calculated using a bootstrap method with replacement using 1000 random samples.

As a key contributor to the cloud-base mass flux, the cloud area  $a_c$  is distributed qualitatively similarly to the distribution of the mass flux (Fig. 4a). The difference between the two reference LES cases shows similar characteristics as for the two mass flux distributions. So, the knowledge about the physical mechanism that shapes  $p(m)$  might also be sufficient to describe  $p(a_c)$ . The cloud area distribution of the A-base case shows a power-law-like shape with a scale break around the value of  $10^6 \text{ m}^2$ . The scale break in the ARM-base case is located at a scale an order of magnitude larger than the one of the R-base case. These two cloud area distributions are actually very similar to the two typical cloud size distributions observed over land and over ocean as derived from the Landsat images in Sengupta et al. (1990), their Fig. 4. A similar change in the distribution behavior for the largest cloud areas is observed in the radar echo areas distribution in Trivej and Stevens (2010). Different statistics of the large echoes compared to a power-law behavior of the small echoes may be controlled by the meteorological environment. In particular, the existence of an inversion layer topping the cloud layer limits the growth of clouds beyond a certain size, which can be connected to the observed break in the scaling (Trivej and Stevens 2010). Strong subsidence inversions over the tropical oceans might explain the position of the scale break at the lower values than what is observed at midlatitudes (see also Wood and Field 2011).

The distribution of vertical velocity of individual clouds is approximately symmetric and can be well fitted using a normal distribution, as illustrated in Fig. 4b. The average vertical velocity per cloud is  $\langle w \rangle = 0.64 \text{ m s}^{-1}$  in the RICO case and  $\langle w \rangle = 0.76 \text{ m s}^{-1}$  in the ARM case. Compared to the RICO case, in the ARM case the variance of  $w$  is significantly higher, and some clouds can gain velocities larger than  $2 \text{ m s}^{-1}$ . This result is in line with the findings of Xu and Randall (2001), albeit for deep convection, where the most significant differences in the updraft intensities between tropical oceanic and midlatitude continental convection were found in the strongest 10% of the updrafts, not in the median values. The correlation between vertical velocity  $w_i$  of individual clouds and their mass fluxes  $m_i$  is very low (not shown here). This is the reason for the similarity between  $p(m)$  and  $p(a_c)$ , while  $p(w)$  belongs to a different family of distributions.

Why are the two reference population distributions different? Is the distribution shape changing under the

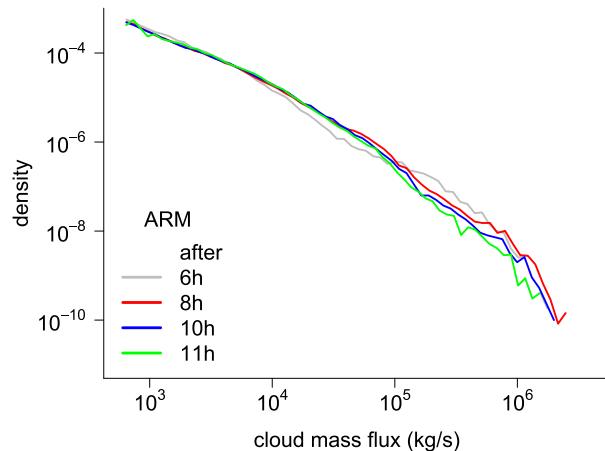


FIG. 5. The probability density distribution of the lifetime average cloud-base mass flux ( $\text{kg s}^{-1}$ ) sampled over time frames of 1-h duration over the diurnal cycle of the A-base case, starting after 6 (gray), 8 (red), 10 (blue), and 11 h (green) of simulation (at 1730, 1930, 2130, and 2230 UTC, respectively).

influence of the large-scale forcing or of the surface conditions? We address these questions in the following section.

#### 4. The three hypotheses

The main differences between the two reference LES cases are in the existence of a strong diurnal cycle over land, strong self-organization of clouds over ocean, and in the magnitude and partitioning of the surface turbulent heat fluxes (Table 1). Other aspects of the large-scale forcing are also different between the two reference cases. However, we rule out those differences as a cause of the different distribution shapes because it was hypothesized and shown in previous studies that the intensity of the convective updrafts was insensitive to changes in the large-scale forcing (e.g., Robe and Emanuel 1996; Cohen and Craig 2006; Plant and Craig 2008). Based on these facts, we propose the three hypotheses that might explain the divergence of the mass flux distribution between the two reference LES cases:

- (i) diurnal cycle of convection determines the distribution  $p(m)$ ,
- (ii) convective self-organization determines the distribution  $p(m)$ , and
- (iii) surface fluxes determine the distribution  $p(m)$ .

In the following, we test the three hypotheses by analyzing all 11 LES cases (Table 1).

##### a. The first hypothesis: Diurnal cycle of convection

Here we test if changes in the forcing associated with the convective diurnal cycle might be responsible

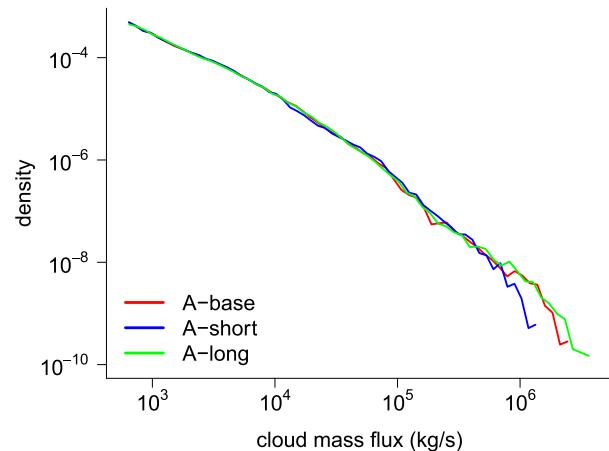


FIG. 6. The probability density distribution of the lifetime average cloud-base mass flux over the cases with different diurnal cycle periods, based on the ARM case: A-base (red), A-short (blue), and A-long (green).

for the different shapes of  $p(m)$  in the two reference cases. We sample the clouds that emerge in the ARM case during four time frames of 1-h duration, taken at different stages of the diurnal cycle, starting at 1730 UTC. The distribution of cloud-base mass flux in all four time frames is shown in Fig. 5. It is clear that there is no significant change in  $p(m)$  over the diurnal cycle of the ARM case [i.e., the distribution  $p(m)$  is stationary].

Another property of the diurnal cycle that might influence  $p(m)$  is the period of the diurnal cycle. Shorter or longer diurnal cycles imply faster or slower temporal changes in the forcing. With faster changes, clouds might have less time to develop undisturbed, so their sizes and mass fluxes might be lower. Or, with slower changes in the forcing, larger clouds might result. To test this, we investigate the results of the simulations A-short and A-long. A time frame of 1-h duration is taken around the peak of the diurnal cycle, after 9, 7, and 11 h from simulation start in A-base, A-short, and A-long, respectively, and  $p(m)$  is examined (Fig. 6). There is again no significant difference among the simulations, except near the right tail of the distribution, where the A-short case shows a faster drop-off than the other two cases. This means that the largest possible clouds cannot develop in the ARM case if the period of the forcing is too short. Overall, there is nevertheless no change in the distribution shape, and the slope of the line stays similar across the three cases. The results of these experiments demonstrate that changes of the forcing over a diurnal cycle do not shape the distribution of the cloud-base mass flux.

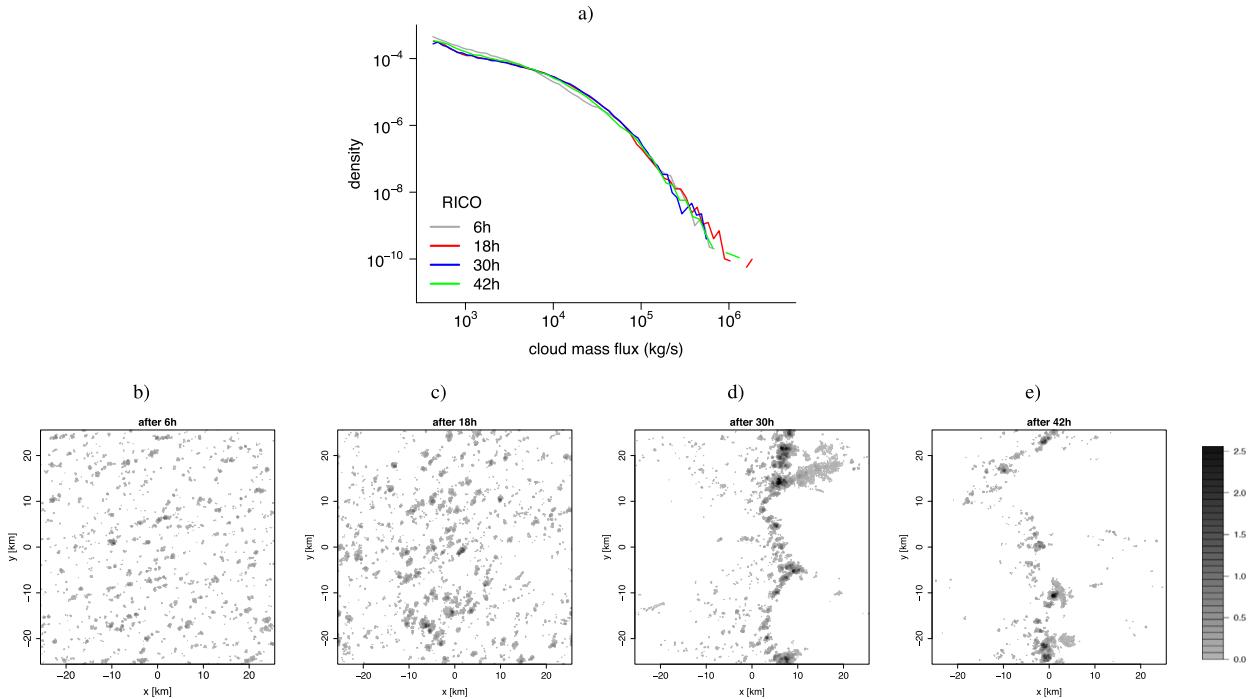


FIG. 7. (a) The probability density distribution of the lifetime average cloud-base mass flux in the R-base case, over different stages of cloud organization: 6 (gray), 18 (red), 30 (blue), and 42 h (green). (b)–(e) Visualization of the corresponding horizontal spatial distribution of the cloud field in the liquid water path ( $\text{g m}^{-2}$ ): after (left to right) 6, 18, 30, and 42 h, respectively. The gray scale is from 0.0 to 2.5 in increments of 0.5.

### b. The second hypothesis: Convective self-organization

In this section, we test how the spatial correlations during the organized phase of the RICO case influence the cloud-base mass flux distribution. Organization of convective clouds into clusters, lines, or arcs could influence  $p(m)$  by affecting the size and intensity of individual cloud elements. Here, it is important to note that the cloud tracking routine identifies the cloud entities that form the cloud clusters and performs splitting so that every element can be followed separately even when two cloud elements have merged. In contrast, past studies have investigated the distributions of merged cloud clusters and suggested self-organization as a mechanism for creating power laws (Scheufele 2014).

We choose the R-base case to test the effects of cloud organization on  $p(m)$  because this convective case is strongly organized after one day of simulation (Fig. 7). Starting from a randomly distributed field of clouds and looking into the time frames with different stages and forms of organization, we plot  $p(m)$  in Fig. 7. We find no evidence that self-organization of clouds has an effect on  $p(m)$  because the overall distribution shape stays the same in spite of organization. Hence, the different degrees of organization between RICO and ARM cannot explain the differences in  $p(m)$ .

Even though self-organization is not responsible for the final shape of the distribution, it is a process that can produce longer tails in the cloud distributions, if the cloud splitting is not performed and cloud clusters are sampled to compute  $p(m)$  (Scheufele 2014). Figure 7 indicates that this dependency vanishes if individual cloud elements are considered.

### c. The third hypothesis: Surface heat fluxes

The two reference cases have very different surface conditions: one is set over the ocean, while the other one is set over land, so the magnitudes of the surface heat fluxes differ by up to a factor of 4 between the cases (see Fig. 1). We investigate here the dependency of the distribution shape on the surface turbulent heat fluxes, which drive the boundary layer convective updrafts that ultimately form cumulus clouds at the top of the sub-cloud layer. We test the magnitude of the fluxes and their partitioning at the surface.

#### 1) THE MAGNITUDE OF THE SURFACE HEAT FLUXES

We have already concluded in the previous section for the ARM case that  $p(m)$  does not change over a single diurnal cycle (Fig. 5). From this conclusion, it also follows that  $p(m)$  is not sensitive to the surface flux

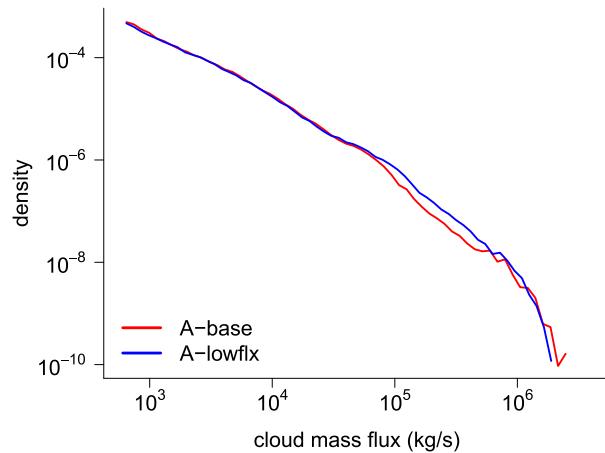


FIG. 8. The probability density distribution of the lifetime average cloud-base mass flux in the A-lowflx case (blue) compared to the distribution in the A-base case (red).

magnitude. To further prove this, we perform one additional test (A-lowflx), in which the total surface turbulent heat flux is lowered by 20% (Fig. 8). There is no significant difference between the two distributions. The A-lowflx case can simply be considered as another realization of the same shallow cloud ensemble of the A-base case.

## 2) THE RATIO OF THE SURFACE HEAT FLUXES

The ratio of the sensible and latent heat fluxes at the surface, the Bowen ratio  $B$ , is the main parameter that characterizes the two surface types, ocean and land. Though the total surface flux magnitude has no effect on  $p(m)$ , the partitioning of this flux into sensible and latent heating might have an effect. Note that the Bowen ratio does not change much over the diurnal cycle in ARM. We thus turn our attention to the sensitivity experiments using different Bowen ratios (Fig. 9).

By changing only the ratio of the surface fluxes and leaving their total summed magnitudes unchanged, the shape of the mass flux distribution can be altered. More importantly, by setting the RICO Bowen ratio in the ARM setup (A-0.03), the mass flux distribution of the RICO case is recovered (Fig. 9a). Likewise, by setting the ARM Bowen ratio in the RICO setup (R-0.33), the mass flux distribution of the ARM case is recovered (Fig. 9b). Thus, it is evident that the ratio of the surface fluxes and not their magnitudes shapes the mass flux distribution.

## 3) THE TWO MODES OF THE CLOUD DISTRIBUTION

The final shape of  $p(m)$  is a result of the superposition of the distribution modes associated with cloud groups of different subtypes: active, forced, and passive

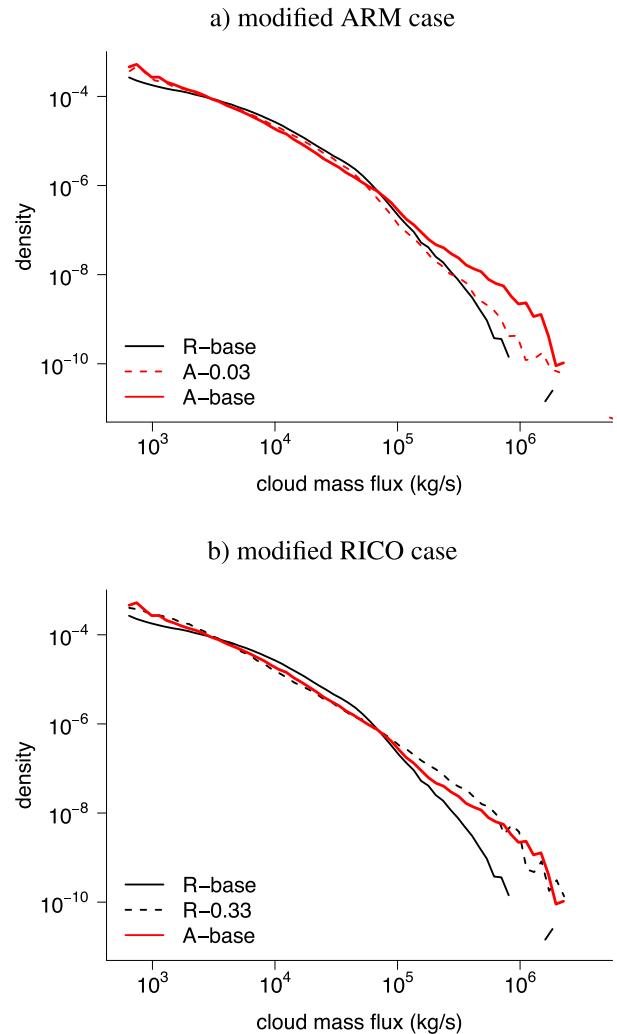


FIG. 9. Reproduction of the distribution shape of the lifetime average cloud-base mass flux of (a) the R-base case (solid black) by altering the Bowen ratio of the ARM case to  $B = 0.03$  (dashed red) and (b) the A-base case (solid red) by altering the RICO case Bowen ratio to  $B = 0.33$ .

clouds (see the classification of Stull 1985). We examine the dependency of these modes on the Bowen ratio separately. Here we simplify the classification of clouds into buoyant (active) and nonbuoyant (passive) clouds, as in Sakradzija et al. (2015). Forced clouds fall into the “passive” nonbuoyant cloud group owing to this simplification. Clouds are classified as active buoyant clouds if the excess of the vertically integrated virtual potential temperature within clouds,  $\theta_{v,up} - \theta_v$ , is larger than a threshold. The threshold is set to 0.5 K, except in a case where this leads to a too-small statistical sample, as in R-0.33. In the latter case, the threshold is set to 0.4 K.

In the RICO-base case (Fig. 10a), the cloud distribution shows shorter tails in both modes and lower mass

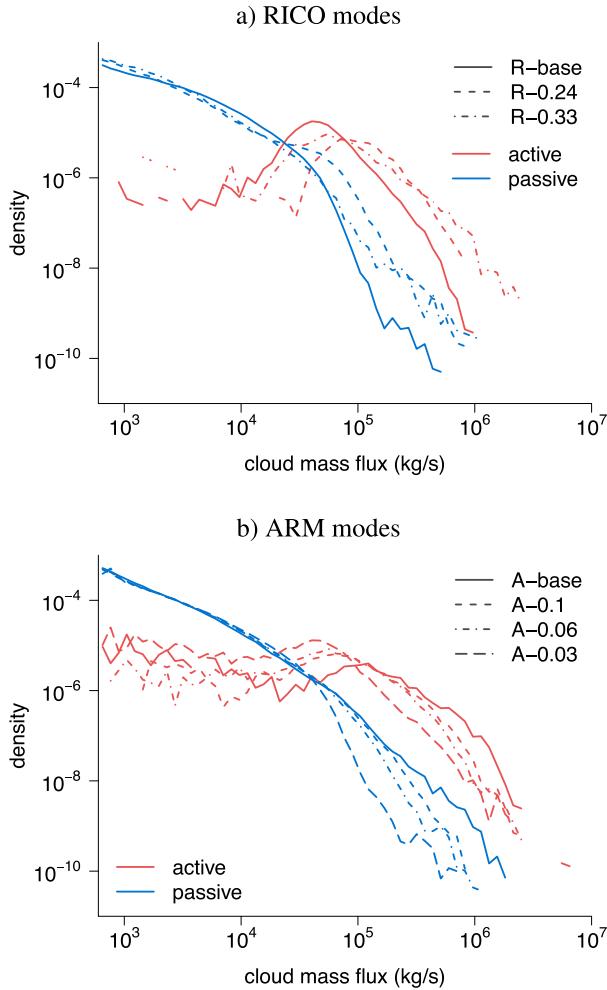


FIG. 10. The lifetime average cloud-base mass flux distribution of active (red) and passive (blue) cloud modes for different Bowen ratios: (a) RICO-based cases and (b) ARM-based cases. Full lines correspond to the reference cases, while dashed lines correspond to the cases with changed  $B$ .

fluxes in average compared to the A-base case (full lines in Fig. 10b). With increasing Bowen ratio, the active cloud modes shift toward higher mass flux values, while the slopes of the two modes become less steep (Fig. 10). Through the control on the range of mass flux values that individual modes of  $p(m)$  can take, and by setting the slope of the modes, the Bowen ratio ends up determining the average mass flux per cloud  $\langle m \rangle$  in both distribution modes.

This fact might explain why different power-law slopes are documented in different observational studies of cloud population (see Table 1 in Zhao and Di Girolamo 2007). The slopes of the observed cloud size distributions in the midlatitude regions have lower values than the slopes in the tropics (see Wood and Field 2011). These characteristics of the observed cloud size

distribution correspond to the control that  $B$  imposes on the slopes that we observe in the RICO and the ARM cloud-base mass flux distributions (Fig. 10). Higher values of  $B$  in midlatitudes produce lower slopes compared to the higher slopes that are produced as a result of low  $B$  in the tropics.

## 5. The Bowen ratio indirectly sets the average mass flux per cloud

To understand the link between  $p(m)$  and  $B$ , we aim at deriving in this section the constraints on the mass flux  $\langle m \rangle$  that an average cloud can transport based on the boundary layer energetics. As will be shown in section 6,  $\langle m \rangle$  is the key parameter through which the difference between the mass flux distributions of the two reference cases is set.

We start from the concept of atmospheric convection as a natural heat engine (Rennó and Ingersoll 1996). During a heat cycle of an average convective cloud, the heat  $Q_{\text{in}}$  (J) is input near the surface in the form of the turbulent surface heat flux  $F_{\text{in}}$  ( $\text{W m}^{-2}$ ) (sum of latent and sensible heat fluxes). This heat is partly converted into mechanical work  $W_{\text{mech}}$  of the convective overturning in the subcloud layer, and the rest is added into the cloud layer and redistributed further. Here, we define the heat cycle for the subcloud layer that lies between the surface layer over the warm ocean or land surface and the colder cloud layer above.

The efficiency of the heat cycle is defined as the ratio of mechanical work and the heat input at the surface:

$$\eta = \frac{W_{\text{mech}}}{Q_{\text{in}}}. \quad (2)$$

The theoretical maximum efficiency of the heat cycle in the subcloud layer is the Carnot efficiency, which can be defined as

$$\eta_{\text{max}} = \frac{T_{\text{sfc}} - T_{\text{lcl}}}{T_{\text{sfc}}}, \quad (3)$$

where  $T_{\text{sfc}}$  is the surface temperature and  $T_{\text{lcl}}$  is the temperature at the lifting condensation level. If the heat input at the surface would happen solely in the form of the sensible heat flux and if no heat was spent to transport water vapor out of the subcloud layer, the efficiency of the convective heat cycle would approach the Carnot efficiency. However, the thermodynamic cycle of convection in the boundary layer is a mixed moist heat cycle with an efficiency that is lower than the maximum theoretical Carnot efficiency,  $\eta < \eta_{\text{max}}$ . As shown in Shutts and Gray (1999), the efficiency of the moist heat cycle can be expressed as [see their Eq. (19)]

$$\eta = \frac{B}{1+B} \left( 1 + \frac{\varepsilon c_p T_{\text{sfc}}}{L_v B} \right) \frac{gH}{2c_p T_{\text{sfc}}}, \quad (4)$$

where  $c_p$  is the specific heat capacity of the dry air at constant pressure,  $L_v$  is the latent heat of vaporization,  $g$  is the gravitational acceleration,  $\varepsilon = 1 - R_v/R_d$ ,  $R_v$  is the gas constant for water vapor,  $R_d$  is the gas constant for dry air, and  $H$  is the subcloud layer depth. Equation (4) is derived under the assumption that the effective heat input at the surface,  $\eta F_{\text{in}}$ , is used to maintain convection against mechanical dissipation in a convective system in statistical equilibrium. The efficiency of a moist heat cycle  $\eta$  could be further explained using the entropy budget analysis, as in [Pauluis and Held \(2002\)](#). They found that convection acts both as a heat cycle and as an atmospheric dehumidifier, and the irreversible entropy production by the two processes is in competition. The more the atmosphere acts as a dehumidifier, the less effective it is to generate kinetic energy of convective circulations ([Pauluis and Held 2002](#)).

From Eq. (4), it follows that the Bowen ratio highly influences the fraction of the heat input that can be transformed into mechanical work to maintain convective circulations. The Bowen ratio  $B$  appears explicitly in Eq. (4) but also implicitly through its control on the depth of the subcloud layer  $H$  (see [Schrieber et al. 1996](#); [Stevens 2007](#)).

Equation (2) does not explicitly relate  $\langle m \rangle$  to the moist heat cycle. To do so, we proceed as follows. The average cloud-base mass flux per cloud  $\langle m \rangle$  is related to the turbulent flux of the moist static energy at the cloud-base level  $\overline{\rho w' h'}$  through the mass flux approximation as defined in [AS-74](#):

$$\overline{\rho w' h'} \approx \sum_i m_i (h_i - \bar{h}), \quad (5)$$

where  $i = 1, \dots, N$  is the index of individual clouds,  $h_i - \bar{h}$  is the excess of the moist static energy within the updrafts that form clouds with respect to the environment, and an overline denotes averaging over the domain.

As a first simple hypothesis, we assume that the turbulent flux of the moist static energy at cloud base,  $\overline{\rho w' h'}$ , is proportional to the effective surface forcing of the cloud ensemble,  $N\eta F_{\text{in}}$ , and by using Eq. (5), we write

$$\sum_i m_i (h_i - \bar{h}) \approx C_1 N \eta F_{\text{in}}, \quad (6)$$

where  $C_1$  is a proportionality constant, which can be seen as a factor of correction for further heat losses not taken into account, and  $N$  is the number of clouds in the cloud ensemble. Because the surface forcing is homogeneous,  $F_{\text{in}}$  is equal for all individual cloud heat cycles.

The efficiency is controlled by the homogeneous surface properties and the subcloud layer depth and is approximately equal among the clouds [see Eq. (4)], so  $\eta$  is treated as a constant in a single convective case. Now we apply the mass-flux-weighted averaging as defined in [Yanai et al. \(1973\)](#) to Eq. (6):

$$\frac{\sum_i m_i h_i}{\sum_i m_i} \approx \tilde{h}, \quad (7)$$

where  $\tilde{h}$  is the mass-flux-weighted average of  $h$ , which is approximately equal to the average of the moist static energy per cloud,  $\langle h \rangle$ , where the brackets  $\langle \cdot \rangle$  denote averaging over the cloud ensemble. The relative difference between the values of  $\tilde{h}$  and  $\langle h \rangle$  is lower than 0.5% as estimated from LES. So, we can rewrite the left-hand side of Eq. (6) as

$$\begin{aligned} \sum_i m_i (h_i - \bar{h}) &= \sum_i m_i h_i - \sum_i m_i \bar{h} = \frac{\sum_i m_i \sum_i m_i h_i}{\sum_i m_i} \\ &\quad - \sum_i m_i \bar{h} \approx \sum_i m_i (\langle h \rangle - \bar{h}); \end{aligned} \quad (8)$$

$$\sum_i m_i (\langle h \rangle - \bar{h}) \approx C_1 N \eta F_{\text{in}}. \quad (9)$$

An average moist heat cycle per cloud can then be expressed as

$$\langle m \rangle (\langle h \rangle - \bar{h}) \approx C_1 \eta F_{\text{in}} \quad (10)$$

and the average mass flux per cloud is then approximately equal to

$$\langle m \rangle \approx C_1 \frac{\eta F_{\text{in}}}{\langle h \rangle - \bar{h}}, \quad (11)$$

with  $\eta$  given by Eq. (4).

We look into the LES to find evidence to support Eq. (11). We base our analysis on the active cloud group, and we plot the average mass flux per active cloud  $\langle m \rangle$  versus the right-hand side of the Eq. (11) ([Fig. 11a](#)). It turns out that Eq. (11) holds remarkably well for the eight tested LES cases of [Table 1](#), which suggests that the average mass flux per cloud is determined by the moist heat cycle of the subcloud layer. The coefficient of determination of a linear regression model is  $r^2 = 0.95$ . The slope is estimated to be equal to  $C_1 = 0.13$ . The intercept parameter is nevertheless not equal to zero and results in an additional mass flux, which we will denote by  $m_0$ :

$$\langle m \rangle = m_0 + C_1 \frac{\eta F_{\text{in}}}{\langle h \rangle - \bar{h}}. \quad (12)$$

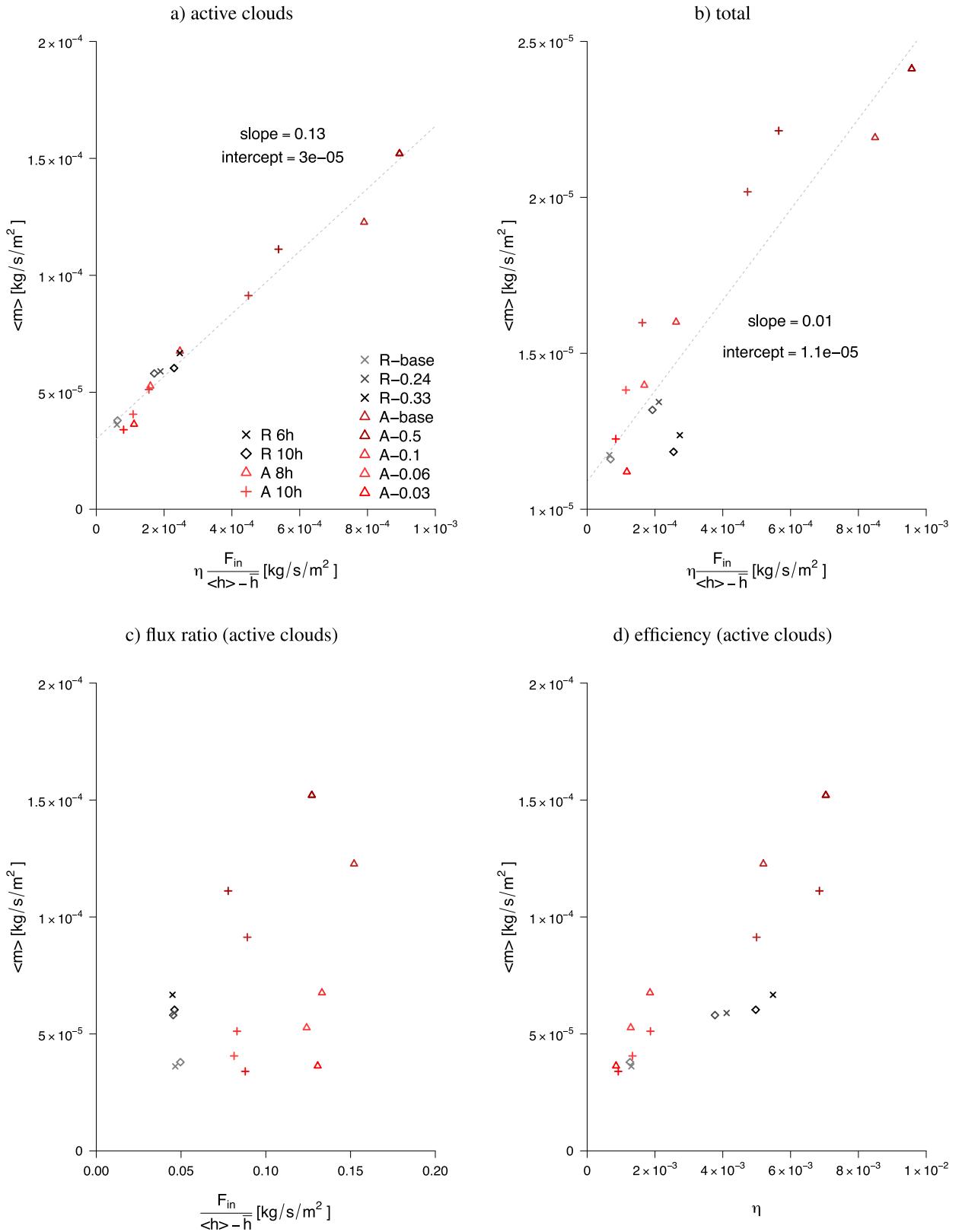


FIG. 11. Scaling of the average cloud-base mass flux ( $\langle m \rangle$ ) based on the moist heat cycle [Eq. (11)] for eight LES cases showing (a) the average mass flux of active cumulus clouds, (b) the average mass flux of the total cloud ensembles, and the right-hand side of the Eq. (11) decomposed into (c)  $F_{in}/(\langle h \rangle - \bar{h})$  and (d)  $\eta$ . Two time frames are used for these figures: the frames starting after 6 and 10 h of the simulation in the RICO-based cases and after 8 and 10 h in the ARM-based cases.

The estimated value in this study is  $m_0 = 3 \times 10^{-5} \text{ kg s}^{-1} \text{ m}^{-2}$ . Depending on the test case and the Bowen ratio value,  $\langle m \rangle$  can be 1.5–6.9 times larger than  $m_0$  (Fig. 11a).

The scaling Eq. (11) is evaluated in Fig. 11a only for the active clouds, while we do not show the scaling for the passive cloud group. This is because the buoyancy threshold used to separate the clouds into the two groups misinterprets some active clouds as passive. We can, however, show the scaling for the total cloud ensemble in Fig. 11b, which still holds.

Equation (11) is decomposed into two parts to test the dependency of  $\langle m \rangle$  of the active clouds on  $F_{\text{in}}/(\langle h \rangle - \bar{h})$  and  $\eta$  separately (Figs. 11c,d). It is clear from Fig. 11c that  $\langle m \rangle$  does not scale with  $F_{\text{in}}/(\langle h \rangle - \bar{h})$ . The points are aligned vertically in three different groups associated with the three main values of the ratio  $F_{\text{in}}/(\langle h \rangle - \bar{h})$ : that is, 0.05, 0.08, and 0.13. The increase in  $\langle m \rangle$  in each of these three groups of points is due to changing values of  $B$ . Thus,  $\langle m \rangle$  is controlled by  $B$ , while the different mean states of the subcloud layer can still result in the same value of  $\langle m \rangle$ . It is not shown here explicitly, but  $\langle m \rangle$  also does not scale uniquely with the total surface heat flux  $F_{\text{in}}$ .

The average mass flux per cloud  $\langle m \rangle$  is also not uniquely determined by  $\eta$ . The quantity  $\eta$  sets the slope of the three lines corresponding to the three different magnitudes of the ratio  $F_{\text{in}}/(\langle h \rangle - \bar{h})$ . Furthermore, if the ratio  $F_{\text{in}}/(\langle h \rangle - \bar{h})$  in a given group of points is higher, the efficiency  $\eta$  in the same group is lower compared to the other groups. As a result,  $\langle m \rangle$  is uniquely determined by the product of the two factors [Eq. (11)], with  $\eta$  playing the key role in setting the dependence on  $B$ .

The fact that  $B$  sets the efficiency of the moist convective heat cycle, and thus also controls the expected value of  $p(m)$ , directly explains why the magnitude of the surface forcing does not influence the distribution shape. From this it also follows that changes of the surface forcing over the diurnal cycle cannot alter the distribution shape, as long as  $B$  does not change significantly over the diurnal cycle. The moist heat cycle formalism might also explain why self-organization is not a powerful driver for the distribution  $p(m)$ . The convecting system is forced by the same amount of heat input, and the efficiency of the moist heat cycle is the same at all stages of cloud organization. So, for the shape of  $p(m)$ , the spatial distribution of clouds does not play any significant role.

An important point to notice here is that the heat cycle formalism applies to the average convective cycle per cloud, and thus it determines  $\langle m \rangle$ , not the bulk contribution  $M$  of all clouds in the shallow cloud ensemble. The quantity  $M$  is not fully constrained by the heat cycle of the subcloud layer. This has an implication for the bulk closure assumptions in parameterization of

convection. To retrieve the closure of a bulk parameterization, an average mass flux per cloud  $\langle m \rangle$  has to be multiplied by the total number of clouds  $N$  to result in the bulk mass flux  $M$ . Therefore, the controls on  $M$  may be decomposed into two contributions: the surface conditions control  $\langle m \rangle$ , while, in addition to the surface conditions, the large-scale forcing acts to set the number of clouds in the ensemble  $N$ .

## 6. Parameters of the mass flux distribution

For the application to parameterizations based on the spectral cloud ensembles (AS-74) or stochastic cloud ensembles (as in Plant and Craig 2008; Sakradzija et al. 2015), a functional form for  $p(m)$ , has to be defined, and the corresponding distribution parameters have to be estimated. In the following, we adopt the mixed Weibull distribution as a functional form for  $p(m)$ , as in Sakradzija et al. (2015):

$$p(m) = (1-f) \frac{k_p}{\lambda_p} \left( \frac{m}{\lambda_p} \right)^{k_p-1} e^{-(m/\lambda_p)^{k_p}} + f \frac{k_a}{\lambda_a} \left( \frac{m}{\lambda_a} \right)^{k_a-1} e^{-(m/\lambda_a)^{k_a}}, \quad (13)$$

where  $f$  is the fraction of active cumulus clouds,  $k$  is the shape,  $\lambda$  is the scale parameter of the Weibull distribution, and subscripts  $p$  and  $a$  denote passive and active distribution modes.

From the results of the previous section, we know that  $\langle m \rangle$  varies with the surface conditions. The question nevertheless remains whether any of the remaining distribution parameters, namely  $k_{p,a}$  and  $\lambda_{p,a}$ , are universal constants. In the study of Sakradzija et al. (2015), these parameters were estimated only for the RICO case for the time period of 6 h, starting after 6 h of simulation. The estimated shape parameter was  $k_p = k_a = 0.7$  for the given cloud sample. In the following, we extend the analysis over the longer time period of the RICO case and over land conditions in the ARM case. We focus on the estimation of the shape parameters,  $k_{p,a}$ , while the scale parameters of the Weibull distribution modes,  $\lambda_{p,a}$ , can be calculated from the expected value of the distribution,  $\langle m \rangle_{p,a} = \lambda_{p,a} \Gamma[1 + (1/k_{p,a})]$ .

In shallow cumulus cloud ensembles, the shape parameter that is less than one,  $k < 1$ , indicates that the memory of cloud life cycles has an effect on the distribution shape (Sakradzija et al. 2015). This effect takes place through correlation between the cloud lifetimes  $\tau_i$  and the cloud-base mass fluxes  $m_i$ , which is already demonstrated for the RICO case in Sakradzija et al.

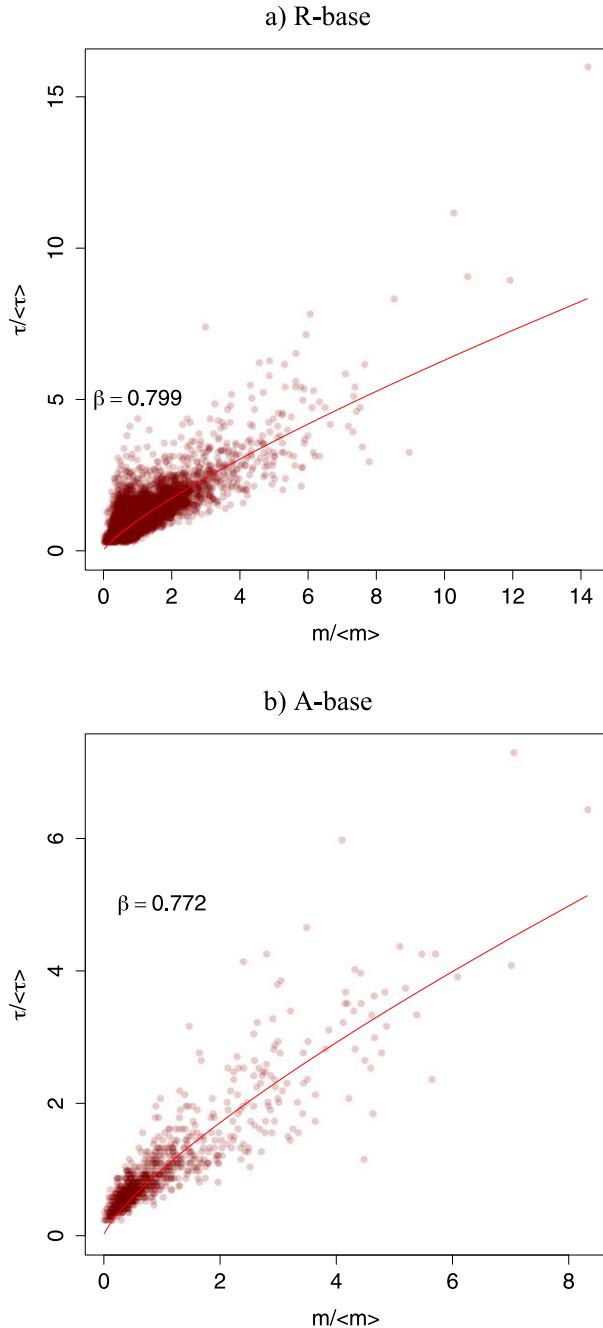


FIG. 12. Scatterplot of individual active cloud lifetimes and cloud-base mass fluxes normalized by the ensemble average values: (a) the R-base simulation and (b) the A-base simulation. Cloud samples are collected during 1 h starting from simulation hour 10. The red line is a fit of the function  $\tau/\langle \tau \rangle = (m/\langle m \rangle)^\beta$  using the nonlinear least squares.

(2015). We confirm this finding for the RICO case, and we also show that it holds in the ARM case (Fig. 12). This correlation is high, with the correlation coefficient equal to  $r_p = 0.8$  in RICO and  $r_p = 0.9$  in ARM, estimated for the active cumulus clouds. We assume that

this correlation can be described by a power-law relation  $\tau/\langle \tau \rangle = (m/\langle m \rangle)^\beta$ , where  $\beta$  is the power exponent and  $\langle \tau \rangle$  is the average cloud lifetime, as in Sakradzija et al. (2015).

In the theory of extreme events, it is known that long-term correlations with a power-law decay of the autocorrelation function lead to Weibull distributions of return intervals between rare events (e.g., Bunde et al. 2003; Blender et al. 2015). In that case, the power-law exponent of the autocorrelation function  $t^{-\beta}$  can be assumed equal to the shape parameter of the Weibull distribution  $k$  (e.g., Blender et al. 2015). Following this reasoning, the normalized lifetime expression  $\tau/\langle \tau \rangle = (m/\langle m \rangle)^\beta$  also leads to a Weibull distribution for the cloud-base mass flux distribution [Eq. (13)]. The power-law exponent can then be related to the shape parameter of the active mode of the Weibull distribution as  $k_a \approx \beta$ . The nonlinear least squares fit in Fig. 12 gives the values for the exponent  $\beta = 0.8$  in RICO and  $\beta = 0.77$  in ARM. Hence, it appears that  $\beta$  is independent of the case setup. The passive cloud group is more dispersive (not shown here), and the statistical fit is thus more uncertain; however, we will assume that  $k_p \approx k_a = 0.8$ .

Combination of the two Weibull modes of the same shape parameter  $k_p = k_a = k$ , but different  $\langle m_p \rangle$  and  $\langle m_a \rangle$ , and hence different  $\lambda_{a,p}$ , can explain the difference between the two cases (Fig. 13). To construct Fig. 13 and with the purpose of highlighting the uncertainties in  $p(m)$  as a result of the chosen value of  $k$  only, we here calculate the values of  $\langle m_p \rangle$  and  $\langle m_a \rangle$  directly from the LES output rather than using the formalism of a thermodynamic cycle. The chosen value of  $k_p = k_a = 0.8$  provides a good fit to both distributions (Fig. 13). On the same plot, we also test a broad range of values for  $k$ , which demonstrates that  $k$  is of secondary importance in determining the final shape of  $p(m)$ . It is evident that  $k$  can still take a wide range of values, [0.8, 1] for RICO and [0.5, 0.8] for ARM, for the correct reproduction of the distribution  $p(m)$ . Therefore, we conclude that the main parameter that sets the difference in  $p(m)$  among the shallow cumulus cases is  $\langle m \rangle$ .

The parameter  $f$ , which is the proportion of active clouds in the cloud ensemble, is about 4%–5% of the total cloud population both in ARM and RICO. This is valid for the distribution of lifetime average mass fluxes during time frames of 1-h duration and including only those clouds that are initiated during the time frames. We choose here to set the value of  $f$  to 0.05.

## 7. Conclusions

The probability distribution of cloud-base mass flux  $p(m)$  differs among shallow cumulus cases. These

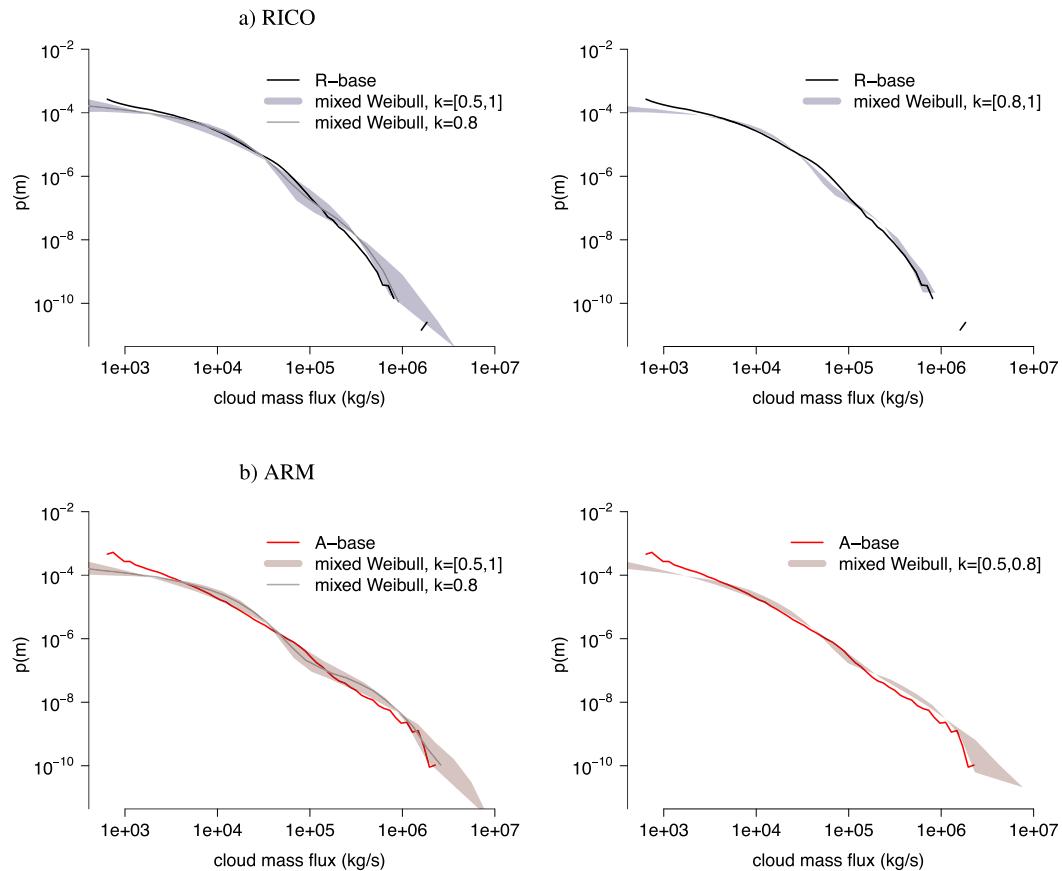


FIG. 13. The mass flux distribution approximately fitted using a bimodal Weibull function: (a) the distribution fit of RICO and (b) the distribution fit of ARM. The range of the shape parameter  $k$  is quite wide to show low sensitivity of the distribution overall shape to this parameter, while the fraction of active clouds in the ensemble is  $f = 5\%$ .

differences manifest themselves through various shapes, slopes, and scales of the distribution. Based on the examination of one typical LES case over the ocean (RICO) and one typical LES case over land (ARM), and nine variations of these two cases, we propose an explanation for the differences in  $p(m)$  among shallow cumulus cases.

The setup of the two reference LES cases differs in the strength and partitioning of the surface turbulent heat fluxes, as well as in the prescribed large-scale forcing tendencies. The ARM case has a strong diurnal cycle that is typical for land conditions, while there is no diurnal cycle in the simulation over the ocean (RICO). In addition, the cloud field in the RICO case is strongly organized, with manifestation of cold pools and arc structures. We have investigated which of these differences in the LES setup is responsible for the distinct shapes of the distribution  $p(m)$ .

Analysis demonstrates that the partitioning of the surface turbulent fluxes into sensible and latent heating,

that is the Bowen ratio  $B$ , is the only parameter that controls the shape of the distribution  $p(m)$ . This control appears to be governed by the second law of thermodynamics and can be explained by interpreting moist convection in the boundary layer as a combination of moisture and heat cycles (as in Shutts and Gray 1999; Pauluis and Held 2002). The efficiency of the moist heat cycle  $\eta$  is less than the Carnot cycle efficiency, because it is directly set by the surface Bowen ratio and the depth of the convecting layer (Shutts and Gray 1999). Through  $\eta$ , the Bowen ratio controls the average mass flux per cloud  $\langle m \rangle$ .

Using the formalism of a moist heat cycle, a scaling law for  $\langle m \rangle$  is derived [see Eq. (11)]. By this scaling, the average vertical mass flux through cloud base  $\langle m \rangle$  is proportional to the ratio of the effective surface heat flux  $\eta F_{\text{sfc}}$  and the excess in the moist static energy at the cloud base with respect to the environment  $\langle h \rangle - \bar{h}$ . This scaling holds remarkably well for the active buoyant clouds in the eight considered convective cases and thus

suggests a universal law across a wide range of the control parameter  $B$ . Passive and forced clouds are not investigated here because of their uncertain separation from the active clouds, but we show that the scaling still holds considering all cloud types.

As such,  $B$  controls the shape of the distribution  $p(m)$  through its control on  $\langle m \rangle$ . We have demonstrated that different shapes of the distribution  $p(m)$  can be well captured by a two-mode Weibull distribution function. The shape parameter of each distribution mode is  $k < 1$ , and it is of secondary importance for determining the final shape of  $p(m)$ . The reason for this robustness comes from similarity of the power-law exponent  $\beta$  in the relation between cloud lifetime and cloud-base mass flux across the LES cases. This power-law exponent sets the unique value of the shape parameter across the LES cases.

The Bowen ratios tested in this study covered the range of values between 0.03 and 0.5. This range corresponds to the span of conditions covering ocean surfaces to temperate forests and grasslands. To make the conclusions of this study more general, it would be advantageous to expand the study to dry land surfaces and to extend the analysis to cloud observations. In addition, the mechanical forcing in the two reference cases was of similar magnitude. A question left for further investigation is how stronger winds and higher wind shears might influence the convective mass flux and population statistics.

One of the key outcomes of this study is that the concept of a moist heat cycle applies to an average convective cloud cycle. To retrieve the total mass flux in a cloud ensemble  $M$ , it is necessary to set the constraints on the number of clouds  $N$  in every given case, since  $M = N\langle m \rangle$ . The number  $N$  does not appear to be constrained by the moist heat cycle. One may hypothesize that  $M$  is governed by the large-scale forcing through control on the number of clouds  $N$ , in addition to the surface conditions that impose a constraint on  $\langle m \rangle$ .

The results of this study also have implications for the cloud size distribution, which has a very similar shape to the distribution  $p(m)$ . Various shapes and slopes of the cloud size distribution that are observed and have been documented in literature may just reflect changes in Bowen ratios encountered across various studies. The various proposed distribution shapes could be encompassed by a single functional form given by the mixed Weibull distribution function. Such a multimodal distribution function already encompasses all the observed shapes, starting from an exponential shape to power laws, depending on the value of the distribution parameters. Based on this study, the expected value of the cloud size distribution might impose the only relevant

control on the distribution shape, which then could be constrained by the underlying physical processes in the boundary layer.

*Acknowledgments.* This research was carried out as part of the Hans-Ertel Centre for Weather Research, under the funding code DWD2014P1A. This research network of universities, research institutes, and the Deutscher Wetterdienst is funded by the Federal Ministry of Transport and Digital Infrastructure (BMVI). Primary data and scripts used in the analysis and other supplementary information that may be useful in reproducing the author's work are archived by the Max Planck Institute for Meteorology and can be obtained by contacting publications@mpimet.mpg.de. Computation is performed using Deutsches Klimarechenzentrum (DKRZ) resources. Plots are made using the R statistical software. The authors thank Juan-Pedro Mellado and Alberto de Lozar for fruitful discussions and comments on an early version of this manuscript. We are also grateful to two anonymous reviewers for their constructive comments and suggestions that helped improve the manuscript.

#### 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, doi:10.1175/1520-0469(1974)031<0674:IOACCE>2.0.CO;2.
- Astin, I., and B. G. Latter, 1998: A case for exponential cloud fields? *J. Appl. Meteor.*, **37**, 1375–1382, doi:10.1175/1520-0450(1998)037<1375:ACFECF>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, doi:10.1029/98JD02579.
- Blender, R., C. C. Raible, and F. Lunkeit, 2015: Non-exponential return time distributions for vorticity extremes explained by fractional Poisson processes. *Quart. J. Roy. Meteor. Soc.*, **141**, 249–257, doi:10.1002/qj.2354.
- 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, doi:10.1256/003590002320373210.
- Bunde, A., J. F. Eichner, S. Havlin, and J. W. Kantelhardt, 2003: The effect of long-term correlations on the return periods of rare events. *Physica A*, **330**, 1–7, doi:10.1016/j.physa.2003.08.004.
- Cahalan, R. F., and J. H. Joseph, 1989: Fractal statistics of cloud fields. *Mon. Wea. Rev.*, **117**, 261–272, doi:10.1175/1520-0493(1989)117<0261:FSOCF>2.0.CO;2.
- Cohen, B. G., and G. C. Craig, 2006: Fluctuations in an equilibrium convective ensemble. Part II: Numerical experiments. *J. Atmos. Sci.*, **63**, 2005–2015, doi:10.1175/JAS3710.1.
- Craig, G. C., and B. G. Cohen, 2006: Fluctuations in an equilibrium convective ensemble. Part I: Theoretical formulation. *J. Atmos. Sci.*, **63**, 1996–2004, doi:10.1175/JAS3709.1.
- 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, doi:10.5194/acp-12-1101-2012.
- 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, doi:10.5194/gmd-6-1261-2013.
- Hozumi, K., T. Harimaya, and C. Magono, 1982: The size distribution of cumulus clouds as a function of cloud amount. *J. Meteor. Soc. Japan*, **60**, 691–699, doi:10.2151/jmsj1965.60.2\_691.
- Jorgensen, D. P., and M. A. LeMone, 1989: Vertical velocity characteristics of oceanic convection. *J. Atmos. Sci.*, **46**, 621–640, doi:10.1175/1520-0469(1989)046<0621:VVCOOC>2.0.CO;2.
- LeMone, M. A., and E. J. Zipser, 1980: Cumulonimbus vertical velocity events in gate. Part I: Diameter, intensity and mass flux. *J. Atmos. Sci.*, **37**, 2444–2457, doi:10.1175/1520-0469(1980)037<2444:CVVEIG>2.0.CO;2.
- López, R. E., 1977: The lognormal distribution and cumulus cloud populations. *Mon. Wea. Rev.*, **105**, 865–872, doi:10.1175/1520-0493(1977)105<0865:TLDACC>2.0.CO;2.
- Lovejoy, S., 1982: Area-perimeter relation for rain and cloud areas. *Science*, **216**, 185–187, doi:10.1126/science.216.4542.185.
- Mitzenmacher, M., 2003: A brief history of generative models for power law and lognormal distributions. *Internet Math.*, **1**, 226–251, doi:10.1080/15427951.2004.10129088.
- 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, doi:10.1029/98JD00088.
- Neggers, R. A. J., 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, doi:10.1175/1520-0469(2003)60<1060:SSOCCP>2.0.CO;2.
- Newman, M., 2005: Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.*, **46**, 323–351, doi:10.1080/00107510500052444.
- Pauluis, O., and I. M. Held, 2002: Entropy budget of an atmosphere in radiative–convective equilibrium. Part I: Maximum work and frictional dissipation. *J. Atmos. Sci.*, **59**, 125–139, doi:10.1175/1520-0469(2002)059<0125:EBOAAI>2.0.CO;2.
- Plank, V. G., 1969: The size distribution of cumulus clouds in representative Florida populations. *J. Appl. Meteor.*, **8**, 46–67, doi:10.1175/1520-0450(1969)008<0046:TSDOCC>2.0.CO;2.
- Plant, R. S., and G. C. Craig, 2008: A stochastic parameterization for deep convection based on equilibrium statistics. *J. Atmos. Sci.*, **65**, 87–105, doi:10.1175/2007JAS2263.1.
- Rauber, R. M., and Coauthors, 2007: Rain in (shallow) cumulus over the ocean: The RICO campaign. *Bull. Amer. Meteor. Soc.*, **88**, 1912–1928, doi:10.1175/BAMS-88-12-1912.
- R Core Team, 2015: R: A language and environment for statistical computing. R Foundation for Statistical Computing. [Available online at <http://www.R-project.org/>.]
- Rennó, N. O., and A. P. Ingersoll, 1996: Natural convection as a heat engine: A theory for cape. *J. Atmos. Sci.*, **53**, 572–585, doi:10.1175/1520-0469(1996)053<0572:NCAAHE>2.0.CO;2.
- Robe, F. R., and K. A. Emanuel, 1996: Moist convective scaling: Some inferences from three-dimensional cloud ensemble simulations. *J. Atmos. Sci.*, **53**, 3265–3275, doi:10.1175/1520-0469(1996)053<3265:MCSSIF>2.0.CO;2.
- Sakradzija, M., A. Seifert, and T. Heus, 2015: Fluctuations in a quasi-stationary shallow cumulus cloud ensemble. *Nonlinear Processes Geophys.*, **22**, 65–85, doi:10.5194/npg-22-65-2015.
- , —, and A. Dipankar, 2016: A stochastic scale-aware parameterization of shallow cumulus convection across the convective gray zone. *J. Adv. Model. Earth Syst.*, **8**, 786–812, doi:10.1002/2016MS000634.
- Scheufele, K., 2014: Resolution dependence of cumulus statistics in radiative–convective equilibrium. Ph.D. thesis, Faculty of Physics, Ludwig–Maximilians–Universität München, 121 pp.
- Schrieber, K., R. Stull, and Q. Zhang, 1996: Distributions of surface-layer buoyancy versus lifting condensation level over a heterogeneous land surface. *J. Atmos. Sci.*, **53**, 1086–1107, doi:10.1175/1520-0469(1996)053<1086:DOSLBV>2.0.CO;2.
- Seifert, A., and K. D. Beheng, 2001: A double-moment parameterization for simulating autoconversion, accretion and self-collection. *Atmos. Res.*, **59–60**, 265–281, doi:10.1016/S0169-8095(01)00126-0.
- 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, doi:10.1175/1520-0450(1990)029<1245:CCFMAS>2.0.CO;2.
- Shutts, G. J., and M. E. B. Gray, 1999: Numerical simulations of convective equilibrium under prescribed forcing. *Quart. J. Roy. Meteor. Soc.*, **125**, 2767–2787, doi:10.1002/qj.49712555921.
- Stevens, B., 2007: On the growth of layers of nonprecipitating cumulus convection. *J. Atmos. Sci.*, **64**, 2916–2931, doi:10.1175/JAS3983.1.
- , 2010: Introduction to UCLA-LES, version 3.2.1. Accessed 3 July 2017, 23 pp. [Available online at [https://github.com/uclales/uclales/raw/master/doc/les\\_doc.pdf](https://github.com/uclales/uclales/raw/master/doc/les_doc.pdf).]
- , C.-H. Moeng, and P. P. Sullivan, 1999: Large-eddy simulations of radiatively driven convection: Sensitivities to the representation of small scales. *J. Atmos. Sci.*, **56**, 3963–3984, doi:10.1175/1520-0469(1999)056<3963:LESORD>2.0.CO;2.
- Stull, R. B., 1985: A fair-weather cumulus cloud classification scheme for mixed-layer studies. *J. Climate Appl. Meteor.*, **24**, 49–56, doi:10.1175/1520-0450(1985)024<0049:AFWCCC>2.0.CO;2.
- Trivej, P., and B. Stevens, 2010: The echo size distribution of precipitating shallow cumuli. *J. Atmos. Sci.*, **67**, 788–804, doi:10.1175/2009JAS3178.1.
- 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, doi:10.1029/2011MS000056.
- Wood, R., and P. R. Field, 2011: The distribution of cloud horizontal sizes. *J. Climate*, **24**, 4800–4816, doi:10.1175/2011JCLI4056.1.
- Xu, K.-M., and D. A. Randall, 2001: Updraft and downdraft statistics of simulated tropical and midlatitude cumulus convection. *J. Atmos. Sci.*, **58**, 1630–1649, doi:10.1175/1520-0469(2001)058<1630:UADSOS>2.0.CO;2.
- Yanai, M., S. Esbensen, and J.-H. Chu, 1973: Determination of bulk properties of tropical cloud clusters from large-scale heat and moisture budgets. *J. Atmos. Sci.*, **30**, 611–627, doi:10.1175/1520-0469(1973)030<0611:DOBPOT>2.0.CO;2.
- 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, doi:10.1029/2006JD007371.