

# A new subgrid-scale representation of hydrometeor fields using a multivariate PDF

Brian M. Griffin and Vincent E. Larson

University of Wisconsin – Milwaukee, Department of Mathematical Sciences, Milwaukee, WI, USA

*Correspondence to:* Brian M. Griffin (bmg2@uwm.edu)

**Abstract.** The subgrid-scale representation of hydrometeor fields is important for calculating microphysical process rates. In order to represent subgrid-scale variability, the Cloud Layers Unified By Binormals (CLUBB) parameterization uses a multivariate Probability Density Function (PDF). In addition to vertical velocity, temperature, and moisture fields, the PDF includes hydrometeor fields.

5 Previously, each hydrometeor field was assumed to follow a multivariate single lognormal distribution. Now, in order to better represent the distribution of hydrometeors, two new multivariate PDFs are formulated and introduced.

The new PDFs represent hydrometeors using either a delta-lognormal or a delta-double-lognormal shape. The two new PDF distributions, plus the previous single lognormal shape, are compared to  
10 histograms of data taken from Large-Eddy Simulations (LES) of a precipitating cumulus case, a drizzling stratocumulus case, and a deep convective case. Finally, the warm microphysical process rates produced by the different hydrometeor PDFs are compared to the same process rates produced by the LES.

## 1 Introduction

15 The atmospheric portion of the hydrological cycle depends on the formation and dissipation of precipitation. In a numerical model, precipitation processes are represented by the microphysics process rates. These process rates are highly dependent on the values of hydrometeor fields at any place and time. Hydrometeors (such as rain water mixing ratio) can vary significantly on spatial scales smaller than the size of a numerical model grid box (Boutle et al., 2014; Lebsock et al., 2013). This means  
20 that a good representation of subgrid-scale variability is important for the parameterization of microphysical process rates.

Subgrid-scale variability (but not spatial organization) can be accounted for through use of a Probability Density Function (PDF). PDFs have been used in atmospheric modeling to account for subgrid variability in moisture and temperature (e.g., Mellor, 1977; Sommeria and Deardorff,  
25 1977; Tompkins, 2002; Naumann et al., 2013) in order to calculate such fields as cloud fraction and mean (liquid) cloud mixing ratio, and have been extended to vertical velocity in order to calculate

fields such as liquid water flux (Lewellen and Yoh, 1993; Lappen and Randall, 2001; Larson et al., 2002; Bogenschutz et al., 2010; Firl and Randall, 2015). PDFs have been used in microphysics to account for subgrid variability in cloud water (Zhang et al., 2002; Morrison and Gettelman, 2008) and in warm hydrometeor fields (Larson and Griffin, 2006, 2013; Cheng and Xu, 2009; Kogan and Mechem, 2014, 2015) in order to calculate warm microphysics process rates. They also have been used to represent cloud ice (Kärcher and Burkhardt, 2008).

Regarding the PDF's functional form, generality is highly desired. For instance, we would like the PDF to be capable of representing interactions among species, such as accretion (collection) of cloud droplets by rain drops. In addition, the PDF should be able to represent a variety of cloud types, such as cumulus and stratocumulus. Generality in the PDF's functional form is important because it facilitates the formulation of unified cloud parameterizations (e.g., Lappen and Randall, 2001; Neggers et al., 2009; Sušelj et al., 2013; Bogenschutz and Krueger, 2013; Guo et al., 2015; Cheng and Xu, 2015; Thayer-Calder et al., 2015).

Cloud Layers Unified By Binormals (CLUBB) is a single-column model that uses a multivariate PDF to account for the subgrid-scale variability of model fields (Golaz et al., 2002a, b; Larson and Golaz, 2005). The original PDF used by CLUBB consisted of only vertical velocity,  $w$ , total water mixing ratio (vapor + liquid cloud),  $r_t$ , and liquid water potential temperature,  $\theta_l$ . The PDF is a weighted mixture, or sum, of two multivariate normal functions. Each one of these multivariate normal functions is known as a PDF component. Although a normal distribution is unskewed, the two-component shape makes it possible to include skewness in model fields.

Larson and Griffin (2013) extended CLUBB's PDF to account for subgrid variability in rain water mixing ratio,  $r_r$ , and rain drop concentration (per unit mass),  $N_r$ . Each of these hydrometeor species was assumed to follow a single lognormal (SL) distribution on the subgrid domain. This treatment worked well for calculating microphysics process rates in a drizzling stratocumulus case (Griffin and Larson, 2013). Subsequently, CLUBB's PDF was extended to other hydrometeor species involving ice, snow, and graupel.

However, the single lognormal treatment of hydrometeors is less successful when it is applied to a partly cloudy, precipitating case. The problem is that the single lognormal assumes that a hydrometeor is found (that is, has a value greater than 0) at every point on the subgrid domain. This is not realistic in a partly cloudy regime, such as precipitating shallow cumulus, which has non-zero precipitation over only a small fraction of the domain.

Consider an example in which rain covers 10% of the grid level. Then the in-precipitation mean of  $r_r$  is ten times greater than the grid-mean value. This can cause problems when microphysics process rates are calculated using the SL. The accretion rate of  $r_r$  is proportional to the value of  $r_r$  inside cloud. In this example, the SL, which distributes the lognormal around the grid mean, would underpredict accretion rate because it causes  $r_r$  to be too small in cloud. Likewise, evaporation rate

is proportional to the value of  $r_r$  outside cloud. The SL would overpredict evaporation rate because it spreads  $r_r$  throughout the domain, including the clear portion.

65 The solution to this problem is to account for the non-precipitating region of the subgrid domain. This is done by representing the non-precipitating region of the domain with a delta function at a value of the hydrometeor of 0. The in-precipitation portion of the subgrid domain can still be handled by using a single lognormal distribution to represent subgrid variability in the hydrometeor species. The resulting distribution is called a delta-lognormal (DL). In the above example with 10%  
70 rain fraction, the (in-precipitation) lognormal from the DL PDF would be distributed around the in-precipitation mean, as desired, rather than around the grid mean, which is a factor of 10 smaller.

Further improvements in accuracy can be achieved with relatively minor modifications to the PDF. As previously mentioned, CLUBB's PDF contains two components. Each of these components can be easily subdivided into an in-precipitation sub-component and an outside-precipitation sub-  
75 component. The result is a delta-lognormal representation of the hydrometeor field *in each PDF component*. Both delta functions are at 0 and represent the region outside of precipitation, but the in-precipitation hydrometeor values are distributed as two lognormals that may have different means and/or variances. When the two lognormals differ in some way, the resulting distribution is called a delta-double-lognormal (DDL). Figure 1 illustrates the SL, DL, and DDL hydrometeor PDF shapes.

80 The main purpose of this paper is to present the formulation of an updated multivariate PDF that extends CLUBB's traditional PDF to include the DL and DDL hydrometeor PDF shapes. Additionally, a new method is derived to divide the *grid-box* mean and variance of a hydrometeor species into *PDF component* means and standard deviations. A secondary purpose of this paper is to present a preliminary comparison of the new PDF shapes with PDFs output by large-eddy simulations (LESs).  
85 The SL, DL, and DDL hydrometeor PDF shapes are compared to histograms of hydrometeor data taken from precipitating LES. Additionally, microphysics process rates are calculated using each of the idealized PDF shapes and compared to microphysics process rates taken from the LES.

The remainder of the paper is organized as follows. Section 2 gives a detailed description of the new PDF. Section 3 discusses the PDF parameters and includes the derivation of a new method to  
90 divide the grid-box mean and variance into PDF component means and standard deviations for a hydrometeor species. Section 4 describes the LES setup and the test cases, as well as the driving of CLUBB's PDF for the tests. Section 5 presents a comparison of hydrometeors between the LES and the SL, DL, and DDL PDF shapes. The comparison includes plots of PDFs, Kolmogorov-Smirnov and Cramer-von Mises scores, and microphysics process rates. Section 6 contains all conclusions.

## 95 **2 Description of the multivariate PDF**

We now describe how the multivariate PDF used by CLUBB is modified to improve the representation of hydrometeors. Perhaps the most important modification is the introduction of precipitation

fraction,  $f_p$ , to the PDF. Precipitation fraction is defined as the fraction of the subgrid domain that contains any kind of precipitation (where any hydrometeor species has a positive value). In order to  
100 account for any precipitation-less region in the subgrid domain, the PDF is modified to add a delta function at a value of 0 for all hydrometeor species. Each PDF component contains its own precipitation fraction. Expressed generally for a PDF of  $n$  components, the overall precipitation fraction is related to the component precipitation fractions by

$$f_p = \sum_{i=1}^n \xi_{(i)} f_{p(i)}, \quad (1)$$

105 where  $f_{p(i)}$  denotes precipitation fraction in the  $i$ th PDF component, and where  $0 \leq f_{p(i)} \leq 1$  for all  $f_{p(i)}$ . Additionally

$$\sum_{i=1}^n \xi_{(i)} = 1, \quad (2)$$

where  $\xi_{(i)}$  is the relative weight, or mixture fraction, of the  $i$ th PDF component, and where  $0 < \xi_{(i)} < 1$  for all  $\xi_{(i)}$ . A PDF with more than one component requires that each PDF component have  
110 a mixture fraction.

Before writing the form of the multi-component PDF, we digress to discuss a special case, the cloud droplet concentration (per unit mass),  $N_c$ . In Larson and Griffin (2013),  $N_c$  was introduced to the PDF and was assumed to follow a single lognormal distribution. This assumption for  $N_c$  means that when any cloud is found at a grid level,  $N_c > 0$  at every point on the subgrid domain. This is  
115 unphysical in a partly cloudy situation, for cloud droplets would be found at points where cloud water is not found. Additionally, the single lognormal treatment of  $N_c$  can cause problems with the microphysics. The grid-level mean of  $N_c$ , denoted  $\overline{N_c}$  (for the remainder of this paper, an overbar denotes a grid-level mean and a prime denotes a turbulent value), is handed to the PDF by the model, and this mean value includes clear air in a partly cloudy situation. This results in a value of  $\overline{N_c}$  that is  
120 much smaller than the in-cloud values of  $N_c$ . Since the single lognormal in  $N_c$  is distributed around  $\overline{N_c}$ ,  $N_c$  is much too small in cloud for cases with small cloud fraction, leading to an excessive autoconversion (raindrop formation) rate.

In order to distribute  $N_c$  where (and only where) cloud water mixing ratio,  $r_c$ , is found on the subgrid domain, it cannot use the same method as the other hydrometeors. Hydrometeors such as  
125  $r_r$  can be found outside cloud where  $r_c$  is not found, or alternatively hydrometeors might be absent inside cloud where  $r_c$  is found. Instead the PDF is modified so that a new variable,  $N_{cn}$ , replaces  $N_c$  in the PDF. The variable  $N_{cn}$  is a mathematical construct that can be viewed as an extended cloud droplet concentration or even as a simplified, conservative cloud condensation nuclei concentration. It is distributed as a single lognormal over the subgrid domain. At points where cloud water is found,  
130  $N_c$  is set equal to  $N_{cn}$ . Otherwise,  $N_c$  is set to 0 at points where no cloud water is found (see Eq. (4) below). The value of  $\overline{N_{cn}}$  is approximately the in-cloud mean of  $N_c$ , and in special cases, is exactly the in-cloud mean of  $N_c$ . Please see Appendix B for a more detailed explanation.

The PDF includes all the hydrometeor species found in the chosen microphysics scheme with the exception of  $r_c$ , which is calculated from other variables in the PDF through a saturation adjustment, and  $N_c$ , which is described above. In addition to  $r_r$  and  $N_r$ , a microphysics scheme may include hydrometeor species such as ice mixing ratio,  $r_i$ , ice crystal concentration (per unit mass),  $N_i$ , snow mixing ratio,  $r_s$ , snowflake concentration (per unit mass),  $N_s$ , graupel mixing ratio,  $r_g$ , and graupel concentration (per unit mass),  $N_g$ . The vector containing all the hydrometeor species included in the PDF will be denoted  $\mathbf{h}$ . The full PDF can be written as  $P(w, r_t, \theta_l, N_{cn}, \mathbf{h})$ .

In order to calculate quantities that depend on saturation, such as  $\bar{r}_c$  and cloud fraction, a PDF transformation is required. The PDF transformation is a change of coordinates. The multivariate PDF undergoes translation, stretching, and rotation of the axes (Larson et al., 2005; Mellor, 1977). Within each PDF component, a separate PDF transformation takes place. The  $i$ th component PDF,  $P_{(i)}(w, r_t, \theta_l, N_{cn}, \mathbf{h})$ , is transformed to  $P_{(i)}(w, \chi, \eta, N_{cn}, \mathbf{h})$ , where  $\chi$  is an ‘‘extended’’ liquid water mixing ratio that, when the air is supersaturated, has a positive value and furthermore is equal to  $r_c$ . When the air is subsaturated,  $\chi$  has a negative value. The variable  $\eta$  is orthogonal to  $\chi$ . The variables  $r_c$  and  $N_c$  can now be written as

$$r_c = \chi H(\chi) \text{ and} \quad (3)$$

$$N_c = N_{cn} H(\chi), \quad (4)$$

where  $H(x)$  is the Heaviside step function.

The general form of a PDF with  $n$  components and  $D$  variables (whether  $D$  includes all the variables in the PDF or any subset of those variables in a multivariate marginal PDF) can be written as

$$P(x_1, x_2, \dots, x_D) = \sum_{i=1}^n \xi_{(i)} P_{(i)}(x_1, x_2, \dots, x_D). \quad (5)$$

Of the  $D$  variables listed, the first  $J$  variables are normally distributed in each PDF component (i.e.  $w, r_t$ , and  $\theta_l$ , or  $w, \chi$  and  $\eta$ ), the next  $K$  variables are lognormally distributed (i.e.  $N_{cn}$ ), and the last  $\Omega$  variables are the hydrometeor species, such that  $D = J + K + \Omega$ . The  $i$ th component of the PDF,  $P_{(i)}(x_1, x_2, \dots, x_D)$ , accounts for both the precipitating and precipitation-less regions, and is given by

$$P_{(i)}(x_1, x_2, \dots, x_D) = f_{p(i)} P_{(J, K + \Omega)(i)}(x_1, x_2, \dots, x_D) + (1 - f_{p(i)}) P_{(J, K)(i)}(x_1, x_2, \dots, x_{J+K}) \left( \prod_{\epsilon=J+K+1}^D \delta(x_\epsilon) \right). \quad (6)$$

The subscripts in the  $i$ th component,  $P_{(J, K)(i)}$  or  $P_{(J, K + \Omega)(i)}$ , denote the number of normal variates,  $J$ , and the number of lognormal variates,  $K$  or  $K + \Omega$ , used in Eq. (7).

Each original PDF component is split into precipitating and precipitation-less sub-components. The component means, variances, and correlations for variables  $x_1 \dots x_{J+K}$  do not differ between

165 the precipitating and precipitation-less parts of Eq. (6). This greatly simplifies the procedure for  
parameterizing the component means and variances, given the grid-level means and variances. Ad-  
ditionally, keeping the component means and variances the same between the in-precipitation and  
outside-precipitation parts of Eq. (6) allows the PDF to be reduced back to prior versions. For in-  
stance, the multivariate PDF in Eqs. (5) and (6) reduces to the version given in Larson and Griffin  
170 (2013) when all  $f_{p(i)} = 1$  and various PDF parameters are chosen appropriately. Furthermore, when  
microphysics is not used in a simulation, hydrometeors are not found in the PDF. In this scenario,  
the PDF reduces to the original version found in Golaz et al. (2002a).

The PDF does not contain a fraction for each hydrometeor species or type, but rather one precipi-  
tation fraction. Each PDF component is split into two sub-components (in-precipitation and outside-  
175 precipitation). Including a fraction for each hydrometeor type (rain, snow, etc.) would cause the  
number of sub-components to grow exponentially with the number of fractions. Using  $n_f$  hydrome-  
teor fractions increases the number of sub-components to  $2^{n_f}$  in each PDF component. This would  
make setting the PDF parameters associated with each sub-component increasingly difficult.

The multivariate PDF can be adjusted to account for a situation when a variable has a constant  
180 value in a PDF (sub-)component. In that situation, the variable can be reduced to a delta function at  
the (sub-)component mean value. A good example of this would be setting  $N_{cn}$  to a constant value  
in order to use a constant in-cloud value of cloud droplet concentration. This is also especially useful  
when dealing with more than one hydrometeor. If one hydrometeor species is found at a grid level,  
but another hydrometeor species is not found at that level, the hydrometeor that is not found can  
185 reduce to a delta function at 0 in the precipitating sub-component of Eq. (6).

The general form of the  $m$ -variate hybrid normal/lognormal distribution in the  $i$ th PDF compo-  
nent,  $P_{(j,k)(i)}(x_1, x_2, \dots, x_m)$ , which is found in each sub-component of Eq. (6), consists of  $j$  nor-  
mal variates and  $k$  lognormal variates, where  $m = j + k$ . The first  $j$  variables are normally distributed  
and the remaining  $k$  variables are lognormally distributed. The multivariate normal/lognormal PDF  
190 is given by (Fletcher and Zupanski, 2006)

$$P_{(j,k)(i)}(x_1, x_2, \dots, x_m) = \frac{1}{(2\pi)^{\frac{m}{2}} |\Sigma_{(i)}|^{\frac{1}{2}}} \left( \prod_{\tau=j+1}^m \frac{1}{x_\tau} \right) \times \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_{(i)})^T \Sigma_{(i)}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{(i)}) \right\}. \quad (7)$$

Both  $\mathbf{x}$  and  $\boldsymbol{\mu}_{(i)}$  are  $m \times 1$  vectors, where  $\mathbf{x}$  is a vector of the variables (in normal-space) in the PDF  
and  $\boldsymbol{\mu}_{(i)}$  is a vector of the (normal-space) PDF sub-component means. The notation  $\mathbf{T}$  denotes the  
transpose of the vector. The  $m \times m$  (normal-space) covariance matrix is denoted  $\Sigma_{(i)}$  and its determi-  
195 nant is denoted  $|\Sigma_{(i)}|$  (Fletcher and Zupanski, 2006). The advantage of a *single* multivariate PDF,  
as opposed to a collection of individual marginal PDFs, is that the multivariate PDF accounts for  
correlations among the variables in the PDF. This is advantageous when calculating such quantities  
as rain water accretion rate and rain water evaporation rate.

Although the multivariate PDF allows for the calculation or specification of the (horizontal) correlation between any two variables at the same grid level, the PDF does not contain information about vertical correlations. Vertical correlations can arise in calculations of radiative transfer, diagnosed hydrometeor sedimentation, or other processes that involve the correlation of a variable with itself at different vertical levels. Such processes are excluded from this study, and hence information about vertical correlations is not needed here. For one possible method to parameterize vertical correlations, see Larson and Schanen (2013).

When variables are integrated out of the full multivariate PDF, the result is a multivariate marginal PDF consisting of fewer variables. When all variables but one are integrated out of the PDF, the result is a univariate marginal or individual marginal PDF. For any hydrometeor species,  $h$ , found in the full multivariate PDF in Eq. (5), the univariate marginal distribution is

$$P(h) = \sum_{i=1}^n \xi_{(i)} (f_{p(i)} P_{L(i)}(h) + (1 - f_{p(i)}) \delta(h)), \quad (8)$$

where  $P_{L(i)}(h)$  is a lognormal distribution in the  $i$ th PDF component, which is given by

$$P_{L(i)}(h) = \frac{1}{(2\pi)^{\frac{1}{2}} \tilde{\sigma}_{h(i)} h} \exp \left\{ -\frac{(\ln h - \tilde{\mu}_{h(i)})^2}{2 \tilde{\sigma}_{h(i)}^2} \right\}. \quad (9)$$

The in-precipitation mean of  $h$  in the  $i$ th PDF component is  $\mu_{h(i)}$ . This is the mean of the  $i$ th lognormal of  $h$ . However,  $\tilde{\mu}_{h(i)}$ , as in Eq. (9), is the normal-space component mean of  $h$ . It is the in-precipitation mean of  $\ln h$  in the  $i$ th PDF component and is given by

$$\tilde{\mu}_{h(i)} = \ln \left( \mu_{h(i)} \left( 1 + \frac{\sigma_{h(i)}^2}{\mu_{h(i)}^2} \right)^{-\frac{1}{2}} \right), \quad (10)$$

where  $\sigma_{h(i)}$  is the in-precipitation standard deviation of  $h$  in the  $i$ th PDF component. The quantity  $\sigma_{h(i)}$  is the standard deviation of the  $i$ th lognormal of  $h$ . The normal-space component standard deviation of  $h$  is  $\tilde{\sigma}_{h(i)}$ , as found in Eq. (9). It is the in-precipitation standard deviation of  $\ln h$  in the  $i$ th PDF component and is given by

$$\tilde{\sigma}_{h(i)} = \sqrt{\ln \left( 1 + \frac{\sigma_{h(i)}^2}{\mu_{h(i)}^2} \right)}. \quad (11)$$

The variables that are distributed marginally as binormals use similar notation. For example,  $\mu_{w(i)}$  is the mean of  $w$  in the  $i$ th PDF component, or the mean of the  $i$ th normal. Likewise,  $\sigma_{w(i)}$  is the standard deviation of  $w$  in the  $i$ th PDF component, or the standard deviation of the  $i$ th normal.

### 225 3 PDF parameters

This paper will use the phrase ‘‘PDF parameters’’ to refer to the PDF *component* means, standard deviations, and correlations involving variables in the PDF, as well as the mixture fractions and the

PDF component precipitation fractions. The PDF parameters are calculated from various grid-mean input variables. In this paper, the component means, standard deviations, and correlations involving  $w$ ,  $r_t$ , and  $\theta_l$ , and the mixture fractions,  $\xi_{(1)}$  and  $\xi_{(2)}$ , are calculated according to the Analytic Double Gaussian 2 (ADG2) PDF, as described in Appendix (e) of Larson et al. (2002). ADG2 requires the following quantities as input: the overall (grid-box) mean, variance, and third-order central moment of  $w$  ( $\overline{w}$ ,  $\overline{w'^2}$ , and  $\overline{w'^3}$ , respectively), the overall mean and variance of  $r_t$  ( $\overline{r_t}$  and  $\overline{r_t'^2}$ , respectively), and the overall mean and variance of  $\theta_l$  ( $\overline{\theta_l}$  and  $\overline{\theta_l'^2}$ , respectively). ADG2 preserves the values of these input variables, meaning that the PDF parameters can be used to successfully reconstruct the values of the input variables. Additionally, ADG2 requires and preserves the overall covariance of  $w$  and  $r_t$  ( $\overline{w'r_t'}$ ), the overall covariance of  $w$  and  $\theta_l$  ( $\overline{w'\theta_l'}$ ), and the overall covariance of  $r_t$  and  $\theta_l$  ( $\overline{r_t'\theta_l'}$ ). All of the aforementioned quantities are prognosed or diagnosed in CLUBB and are not the subject of this paper.

The individual marginal distribution for  $N_{cn}$  is specified to be a single lognormal over the entire subgrid domain. This requires that both PDF component means equal the overall (grid-box) mean ( $\mu_{N_{cn}(1)} = \mu_{N_{cn}(2)} = \overline{N_{cn}}$ ). Likewise, this requires that both PDF component standard deviations equal the overall standard deviation ( $\sigma_{N_{cn}(1)} = \sigma_{N_{cn}(2)} = \overline{N_{cn}'}^{1/2}$ ).

When no hydrometeor species are found at a grid level ( $\mathbf{h} = 0$ ),  $f_p = f_{p(1)} = f_{p(2)} = 0$ . Otherwise, if any hydrometeor species in  $\mathbf{h}$  is found at a grid level (has a value greater than 0),  $f_p|_{\text{tol}} \leq f_p \leq 1$ , where  $f_p|_{\text{tol}}$  is the minimum value allowed for precipitation fraction when hydrometeors are present. We now describe how CLUBB parameterizes  $f_{p(1)}$  and  $f_{p(2)}$ , given  $f_p$ . First, we note that

$$f_p = \xi_{(1)}f_{p(1)} + \xi_{(2)}f_{p(2)}. \quad (12)$$

A tunable parameter,  $v_*$  (where the  $*$  subscript denotes a tunable or adjustable parameter), is introduced and is defined as the ratio of  $\xi_{(1)}f_{p(1)}$  to  $f_p$ , where  $0 \leq v_* \leq 1$ . The precipitation fraction of PDF component 1 is solved by

$$f_{p(1)} = \min\left(\frac{v_*f_p}{\xi_{(1)}}, 1\right). \quad (13)$$

The PDF component 2 precipitation fraction can now be solved by

$$f_{p(2)} = \min\left(\frac{f_p - \xi_{(1)}f_{p(1)}}{\xi_{(2)}}, 1\right). \quad (14)$$

When  $f_{p(1)}$  calculated by Eq. (13) is small enough to force  $f_{p(2)}$  calculated by Eq. (14) to be limited at 1, the value of  $f_{p(1)}$  is recalculated (with  $f_{p(2)} = 1$ ) and is increased enough to satisfy Eq. (12).

### 3.1 Hydrometeor PDF parameters

A mean-and-variance-preserving method is used to calculate the in-precipitation means of the hydrometeor field in the two PDF components,  $\mu_{h(1)}$  and  $\mu_{h(2)}$ , and the in-precipitation standard deviations of the hydrometeor field in the two PDF components,  $\sigma_{h(1)}$  and  $\sigma_{h(2)}$ . The fields that need



to be provided as inputs are the overall (grid-box) mean of the hydrometeor,  $\bar{h}$ , the overall variance of the hydrometeor,  $\overline{h'^2}$ , the mixture fraction in each PDF component,  $\xi_{(1)}$  and  $\xi_{(2)}$ , the overall precipitation fraction,  $f_p$ , and the precipitation fraction in each PDF component,  $f_{p(1)}$  and  $f_{p(2)}$ . Given  
 265 these inputs, the in-precipitation mean of the hydrometeor,  $\overline{h|_{ip}}$ , can be calculated by

$$\overline{h|_{ip}} = \frac{\bar{h}}{f_p}, \quad (15)$$

and the in-precipitation variance of the hydrometeor,  $\overline{h|_{ip}'^2}$ , can be calculated by

$$\overline{h|_{ip}'^2} = \frac{\overline{h'^2} + \bar{h}^2 - f_p \overline{h|_{ip}}^2}{f_p}. \quad (16)$$

The grid-level mean value of any function that is written in terms of variables involved in the PDF  
 270 can be found by integrating over the product of that function and the PDF. For example,

$$\bar{h} = \int_0^{\infty} h P(h) dh \quad \text{and} \quad \overline{h'^2} = \int_0^{\infty} (h - \bar{h})^2 P(h) dh. \quad (17)$$

After integrating, the equation for  $\bar{h}$  expressed in terms of PDF parameters is

$$\bar{h} = \xi_{(1)} f_{p(1)} \mu_{h(1)} + \xi_{(2)} f_{p(2)} \mu_{h(2)}. \quad (18)$$

Likewise, the equation for  $\overline{h'^2}$  expressed in terms of PDF parameters is

$$275 \quad \overline{h'^2} = \xi_{(1)} f_{p(1)} \left( \mu_{h(1)}^2 + \sigma_{h(1)}^2 \right) + \xi_{(2)} f_{p(2)} \left( \mu_{h(2)}^2 + \sigma_{h(2)}^2 \right) - \bar{h}^2. \quad (19)$$

When the hydrometeor is not found at a grid level,  $\bar{h} = \overline{h'^2} = 0$  and the *component* means and standard deviations of the hydrometeor also have a value of 0. When the hydrometeor is found at a grid level,  $\bar{h} > 0$ . Precipitation may be found in only PDF component 1, only PDF component 2, or in both PDF components. When precipitation is found in only PDF component 1,  $\mu_{h(2)} = \sigma_{h(2)} = 0$   
 280 and  $\mu_{h(1)}$  and  $\sigma_{h(1)}$  can easily be solved by Eq. (18) and Eq. (19). Likewise, when precipitation is found in only PDF component 2,  $\mu_{h(1)} = \sigma_{h(1)} = 0$  and  $\mu_{h(2)}$  and  $\sigma_{h(2)}$  can easily be solved by the same equation set.

When there is precipitation found in both PDF components, further information is required to solve for the two component means and the two component standard deviations. The variable  $R$  is  
 285 introduced such that

$$R \equiv \frac{\sigma_{h(2)}^2}{\mu_{h(2)}^2}. \quad (20)$$

In order to allow the ratio of  $\sigma_{h(1)}^2$  to  $\mu_{h(1)}^2$  to vary, the parameter  $\zeta_*$  is introduced, such that

$$R(1 + \zeta_*) = \frac{\sigma_{h(1)}^2}{\mu_{h(1)}^2}, \quad (21)$$

where  $\zeta_* > -1$ . When  $\zeta_* > 0$ , then  $\sigma_{h(1)}^2/\mu_{h(1)}^2$  increases at the expense of  $\sigma_{h(2)}^2/\mu_{h(2)}^2$ , which  
 290 decreases in this variance-preserving equation set. When  $\zeta_* = 0$ , then  $\sigma_{h(1)}^2/\mu_{h(1)}^2 = \sigma_{h(2)}^2/\mu_{h(2)}^2$ .  
 When  $-1 < \zeta_* < 0$ , then  $\sigma_{h(2)}^2/\mu_{h(2)}^2$  increases at the expense of  $\sigma_{h(1)}^2/\mu_{h(1)}^2$ , which decreases.  
 Combining Eq. (19), Eq. (20), and Eq. (21), the equation for  $\overline{h'^2}$  can be rewritten as

$$\overline{h'^2} = \xi_{(1)}f_{p(1)}(1 + R(1 + \zeta_*))\mu_{h(1)}^2 + \xi_{(2)}f_{p(2)}(1 + R)\mu_{h(2)}^2 - \overline{h}^2. \quad (22)$$

Both the variance of each PDF component and the spread between the means of each PDF com-  
 295 ponent contribute to the in-precipitation variance of the hydrometeor ( $\overline{h|_{ip}'^2}$ ). At one extreme, the  
 standard deviation of each component could be set to 0 and the in-precipitation variance could be  
 accounted for by spreading the PDF component (in-precipitation) means far apart. The value of  $R$   
 in this scenario would be its minimum possible value, which is 0. At the other extreme, the means  
 of each component could be set equal to each other and the in-precipitation variance could be ac-  
 300 counted for entirely by the PDF component (in-precipitation) standard deviations. The value of  $R$  in  
 this scenario would be its maximum possible value, which is  $R_{\max}$ .

In order to calculate the value of  $R_{\max}$ , set  $\mu_{h(1)} = \mu_{h(2)} = \overline{h|_{ip}}$  and  $R = R_{\max}$ . Eq. (22) becomes  

$$\overline{h'^2} + \overline{h}^2 = \overline{h|_{ip}}^2 (\xi_{(1)}f_{p(1)}(1 + R_{\max}(1 + \zeta_*)) + \xi_{(2)}f_{p(2)}(1 + R_{\max})). \quad (23)$$

When Eq. (16) is substituted into Eq. (23),  $R_{\max}$  is solved for and the equation is

$$305 \quad R_{\max} = \left( \frac{f_p}{\xi_{(1)}f_{p(1)}(1 + \zeta_*) + \xi_{(2)}f_{p(2)}} \right) \frac{\overline{h|_{ip}'^2}}{\overline{h|_{ip}}^2}. \quad (24)$$

In the scenario that  $\zeta_* = 0$  the equation for  $R_{\max}$  reduces to the ratio of  $\overline{h|_{ip}'^2}$  to  $\overline{h|_{ip}}^2$ .

In order to calculate the value of  $R$ , a parameter is used to prescribe the ratio of  $R$  to its maximum  
 value,  $R_{\max}$ . The prescribed parameter is denoted  $o_*$ , where

$$R = o_* R_{\max}, \quad (25)$$

310 and where  $0 \leq o_* \leq 1$ . Both  $R$  and  $R_{\max}$  are known functions of the inputs and tunable parameters.  
 When  $o_* = 0$ , the standard deviation of each PDF component is 0, and  $\mu_{h(1)}$  is spread far from  $\mu_{h(2)}$ .  
 When  $o_* = 1$ , then  $\mu_{h(1)} = \mu_{h(2)}$ , and the standard deviations of the PDF components account for all  
 of the in-precipitation variance. At intermediate values of  $o_*$ , the means of each PDF component are  
 somewhat spread apart and each PDF component has some width. The new equation for hydrometeor  
 315 variance becomes

$$\overline{h'^2} = \xi_{(1)}f_{p(1)}(1 + o_*R_{\max}(1 + \zeta_*))\mu_{h(1)}^2 + \xi_{(2)}f_{p(2)}(1 + o_*R_{\max})\mu_{h(2)}^2 - \overline{h}^2. \quad (26)$$

The two remaining unknowns,  $\mu_{h(1)}$  and  $\mu_{h(2)}$ , can be solved by a set of two equations, Eq. (18)  
 for  $\overline{h}$  and Eq. (26) for  $\overline{h'^2}$ . All other quantities in the equation set are known quantities. To find the  
 solution, Eq. (18) is rewritten to isolate  $\mu_{h(2)}$  such that

$$320 \quad \mu_{h(2)} = \frac{\overline{h} - \xi_{(1)}f_{p(1)}\mu_{h(1)}}{\xi_{(2)}f_{p(2)}}. \quad (27)$$

The above equation is substituted into Eq. (26). The resulting equation is rewritten in the form

$$Q_a \mu_{h(1)}^2 + Q_b \mu_{h(1)} + Q_c = 0, \quad (28)$$

so the solution to the quadratic equation for  $\mu_{h(1)}$  is

$$\mu_{h(1)} = \frac{-Q_b \pm \sqrt{Q_b^2 - 4Q_a Q_c}}{2Q_a}, \quad (29)$$

325 where:

$$\begin{aligned} Q_a &= \xi_{(1)} f_{p(1)} (1 + o_* R_{\max} (1 + \zeta_*)) + \frac{\xi_{(1)}^2 f_{p(1)}^2}{\xi_{(2)} f_{p(2)}} (1 + o_* R_{\max}), \\ Q_b &= -2 \frac{\xi_{(1)} f_{p(1)}}{\xi_{(2)} f_{p(2)}} (1 + o_* R_{\max}) \bar{h}, \text{ and} \\ Q_c &= - \left( \bar{h}'^2 + \left( 1 - \frac{1 + o_* R_{\max}}{\xi_{(2)} f_{p(2)}} \right) \bar{h}^2 \right). \end{aligned} \quad (30)$$

The value of  $Q_a$  is always positive and the value of  $Q_b$  is always negative. The value of  $Q_c$  can be positive, negative, or zero. Since  $(1 - (1 + o_* R_{\max}) / (\xi_{(2)} f_{p(2)})) \bar{h}^2$  is always negative and  $\bar{h}'^2$  is always positive, the sign of  $Q_c$  depends on which term is greater in magnitude.

330 When  $\bar{h}'^2$  is greater, the sign of  $Q_c$  is negative. This means that  $-4Q_a Q_c$  is positive, which in turn means that  $\sqrt{Q_b^2 - 4Q_a Q_c}$  is greater in magnitude than  $-Q_b$ . If the subtraction option of the  $\pm$  were to be chosen, the value of  $\mu_{h(1)}$  would be negative in this scenario. At first glance, it might appear natural to always choose the addition option. However, this set of equations was derived with the condition that  $\mu_{h(1)}$  equals  $\mu_{h(2)}$  when  $o_* = 1$ . When  $\zeta_* \geq 0$ , this happens when the addition  
335 option is chosen, but not when the subtraction option is chosen. However, when  $\zeta_* < 0$ , this happens when the subtraction option is chosen, but not when the addition option is chosen. So, the equation for  $\mu_{h(1)}$  becomes

$$\mu_{h(1)} = \begin{cases} \frac{-Q_b + \sqrt{Q_b^2 - 4Q_a Q_c}}{2Q_a}, & \text{when } \zeta_* \geq 0; \text{ and} \\ \frac{-Q_b - \sqrt{Q_b^2 - 4Q_a Q_c}}{2Q_a}, & \text{when } \zeta_* < 0. \end{cases} \quad (31)$$

The value of  $\mu_{h(2)}$  can now be found using Eq. (27). After  $\mu_{h(1)}$  and  $\mu_{h(2)}$  have been solved,  $\sigma_{h(1)}$   
340 and  $\sigma_{h(2)}$  can be solved by plugging Eq. (25) back into Eq. (21) and Eq. (20), respectively.

As the value of  $\overline{h}_{\text{ip}}'^2 / \overline{h}_{\text{ip}}^2$  increases and as the value of  $o_*$  decreases (narrowing the in-precipitation standard deviations and increasing the spread between the in-precipitation means), one of the component means may become negative. This happens because there is a limit to the amount of in-precipitation variance that can be represented by this kind of distribution. In order to prevent  
345 out-of-bounds values of  $\mu_{h(1)}$  or  $\mu_{h(2)}$ , a lower limit is declared, called  $\mu_h|_{\min}$ , where  $\mu_h|_{\min}$  is a small, positive value that is typically set to be two orders of magnitude smaller than  $\overline{h}_{\text{ip}}$ . The value of  $\mu_{h(1)}$  or  $\mu_{h(2)}$  will be limited from becoming any smaller (or negative) at this value. From there,

the value of the other hydrometeor in-precipitation component mean is easy to calculate. Then, both values will be entered into the calculation of hydrometeor variance in Eq. (22), which will be rewritten to solve for  $R$ . Then, both the hydrometeor mean and hydrometeor variance will be preserved with a valid distribution.

When the value of  $\zeta_* \geq 0$ , the value of  $\mu_{h(1)}$  tends to be larger than the value of  $\mu_{h(2)}$ . Likewise when the value of  $\zeta_* < 0$ , the value of  $\mu_{h(2)}$  tends to be larger than the value of  $\mu_{h(1)}$ . Since most cloud water and cloud fraction tends to be found in PDF component 1, it is appropriate and advantageous to have the larger in-precipitation component mean of the hydrometeor also found in PDF component 1. The recommended value of  $\zeta_*$  is a value greater than or equal to 0.

This method of closing the hydrometeor PDF parameter equation set produces a DDL hydrometeor PDF shape when  $0 < o_* < 1$  or when  $\zeta_* \neq 0$ . The DL hydrometeor PDF shape is produced simply by setting  $o_* = 1$  and  $\zeta_* = 0$ . These settings force  $\mu_{h(1)} = \mu_{h(2)}$  and  $\sigma_{h(1)} = \sigma_{h(2)}$ , which result in a single lognormal within the precipitating portion of the subgrid domain. Furthermore, if, in addition to setting  $o_* = 1$  and  $\zeta_* = 0$ , one simply sets  $f_{p(1)} = f_{p(2)} = 1$ , then precipitation is found everywhere within the subgrid domain, producing the SL hydrometeor PDF shape. Hence it is very easy to change between DDL, DL, and SL hydrometeor PDF shapes. Additionally, it should be noted that there is only one  $o_*$  and only one  $\zeta_*$  applied to all the hydrometeor species in  $\mathbf{h}$ .

In limited testing, the value of the tunable parameter  $\zeta_*$  did not affect the results much for CLUBB's DDL PDF shape. The value of  $\zeta_*$  has been left at 0, effectively eliminating a tunable or adjustable parameter from the scheme. When  $\zeta_* = 0$ , the DDL shape approaches the DL shape as  $o_*$  approaches 1. As  $o_*$  approaches 0, the DDL shape approaches a double-delta in precipitation (in addition to the delta at 0). Additionally, when  $0 < o_* < 1$ , the in-precipitation skewness of the hydrometeor field is influenced by  $v_*$ . As  $v_*$  approaches 0, the in-precipitation distribution becomes more highly (positively) skewed. In Gaussian space (see Section 5), the in-precipitation distribution is positively skewed. As  $v_*$  approaches 1, the in-precipitation distribution is less (positively) skewed. In Gaussian space, the in-precipitation distribution is negatively skewed. For the results presented in this paper for the DDL hydrometeor PDF shape, the remaining two tunable parameters have been set to the values  $o_* = 0.5$  and  $v_* = 0.55$ .

#### 4 Model setup and testing

There is insufficient data from observations to calculate all the fields that need to be input into CLUBB's PDF. However, this data can be supplied easily and plentifully by a LES. In this paper, LES output of precipitating cases is simulated by the System for Atmospheric Modeling (SAM) (Khairoutdinov and Randall, 2003). SAM uses an anelastic equation set that predicts liquid water static energy, total water mixing ratio, vertical velocity, and both the south-north and west-east components of horizontal velocity. Additionally, it predicts hydrometeor fields as directed by the chosen

microphysics scheme. A predictive 1.5-order subgrid-scale turbulent kinetic energy closure is used to compute the subgrid-scale fluxes (Deardorff, 1980). SAM uses a fixed, Cartesian spatial grid and a third-order Adams-Bashforth time-stepping scheme to advance the predictive equations of motion. It uses periodic boundary conditions and a rigid lid at the top of the domain. The second-order MP-DATA (multidimensional positive definite advection transport algorithm) scheme is used to advect the predictive variables (Smolarkiewicz and Grabowski, 1990).

In order to assess the generality of the different hydrometeor PDF shapes for different cloud regimes, SAM was used to run three idealized test cases — a precipitating shallow cumulus case, a drizzling stratocumulus case, and a deep convective case. The use of cases from differing cloud regimes help avoid overfitting the parameterizations of PDF shape. The setup for the precipitating shallow cumulus test case was based on the Rain in Cumulus over the Ocean (RICO) LES inter-comparison (van Zanten et al., 2011). The horizontal resolution was 100 m, and 256 grid boxes were used in each horizontal direction. The vertical resolution was a constant 40 m and 100 grid boxes were used in the vertical. The model top was located at 4000 m in altitude. The model time step was 1 s and the duration of the simulation was 72 hours. A vertical profile of level-averaged statistics was output every minute and a three-dimensional snapshot of hydrometeor fields was output every hour.

The RICO simulation was run with SAM’s implementation of the Khairoutdinov and Kogan (2000, hereafter KK) warm microphysics scheme. KK microphysics predicts both  $r_r$  and  $N_r$ . SAM’s implementation of KK microphysics uses a saturation adjustment to diagnose  $r_c$ , and cloud droplet concentration is set to a constant value (which is  $70 \text{ cm}^{-3}$  for RICO).

The setup for the drizzling stratocumulus test case was taken from the LES intercomparison based on research flight two (RF02) of the second Dynamics and Chemistry of Marine Stratocumulus (DYCOMS-II) field experiment (Ackerman et al., 2009). The horizontal resolution was 50 m and 128 grid boxes were used in each horizontal direction. An unevenly-spaced vertical grid was used containing 96 grid boxes and covering a domain of depth 1459.3 m. The model time step was 0.5 s and the duration of the simulation was six hours. A vertical profile of level-averaged statistics was output every minute and a three-dimensional snapshot of hydrometeor fields was output every 30 minutes. The DYCOMS-II RF02 simulation was also run with SAM’s implementation of KK microphysics and used a constant cloud droplet concentration of  $55 \text{ cm}^{-3}$ .

The setup for the deep convective test case was taken from the LES intercomparison based on the Large-Scale Biosphere-Atmosphere (LBA) experiment (Grabowski et al., 2006). The horizontal resolution was 1000 m, and 128 grid boxes were used in each horizontal direction. An unevenly-spaced vertical grid was used, containing 128 grid boxes and covering a domain of depth 27500 m. The model time step was 6 s and the duration of the simulation was six hours. A vertical profile of level-averaged statistics was output every minute and a three-dimensional snapshot of hydrometeor fields was output every 15 minutes for the final 3.5 hours of the simulation.

The LBA case requires a microphysics scheme that can account for ice-phase hydrometeor species. The LBA simulation was run with Morrison et al. (2005) microphysics, which predicts the mixing ratio and number concentration (per unit mass) of rain, cloud ice, snow, and graupel. SAM’s implementation of Morrison microphysics diagnoses  $r_c$  using a saturation adjustment right before the microphysics is called and then allows microphysics to update the value of  $r_c$ , which in turn is used to update the value  $r_t$ . Cloud droplet concentration was set to a constant value of  $100 \text{ cm}^{-3}$ .

CLUBB’s hydrometeor PDF shapes will be compared to histograms of hydrometeors produced by SAM LES data. Our goal is to isolate errors in the PDF shape itself. In order to eliminate sources of error outside of the PDF shape and provide an “apples-to-apples” comparison of CLUBB’s PDF shapes to SAM data, we drive CLUBB’s PDF using SAM LES fields, rather than perform interactive CLUBB simulations. The following fields are taken from SAM’s statistical profiles and are used as inputs to CLUBB’s PDF:  $\overline{r_t}$ ,  $\overline{\theta_l}$ ,  $\overline{w'^2}$ ,  $\overline{r_t'^2}$ ,  $\overline{\theta_l'^2}$ ,  $\overline{w'r_t'}$ ,  $\overline{w'\theta_l'}$ ,  $\overline{r_t'\theta_l'}$ ,  $\overline{w'^3}$ ,  $f_p$ ,  $\overline{r_r}$ ,  $\overline{r_r'^2}$ ,  $\overline{N_r}$ , and  $\overline{N_r'^2}$ . For the LBA case, we add  $\overline{r_i}$ ,  $\overline{r_i'^2}$ ,  $\overline{N_i}$ ,  $\overline{N_i'^2}$ ,  $\overline{r_s}$ ,  $\overline{r_s'^2}$ ,  $\overline{N_s}$ ,  $\overline{N_s'^2}$ ,  $\overline{r_g}$ ,  $\overline{r_g'^2}$ ,  $\overline{N_g}$ , and  $\overline{N_g'^2}$ . Another input to CLUBB’s PDF is  $\overline{w}$ . The value of  $\overline{w}$  from large-scale forcing is set according to case specifications in both SAM and CLUBB. CLUBB’s PDF is generated at every SAM vertical level and at every output time of SAM level-averaged statistical profiles.

Additionally, covariances that involve at least one hydrometeor are added to the above list and are used to calculate the PDF component correlations of the same two variables. These covariances are  $\overline{r_t'r_r'}$ ,  $\overline{\theta_l'r_r'}$ ,  $\overline{r_t'N_r'}$ ,  $\overline{\theta_l'N_r'}$ , and  $\overline{r_r'N_r'}$ . Please see Appendix A for more details on the calculation of PDF component correlations. The values of the component correlations do not affect the individual marginal PDFs of the hydrometeors. They are included for the calculation of microphysics process rates (see Section 5.2).

Owing to differences between the KK and Morrison microphysics schemes in SAM,  $f_p$  used by CLUBB’s PDF is computed slightly differently depending on which microphysics scheme is used by SAM. The differences are due to the number of hydrometeor species involved in the microphysics, the thresholding found internally in the microphysics codes, and the variables that are output to statistics by SAM. KK microphysics contains only rain, and SAM’s implementation of KK microphysics clips any value of  $r_r$  (and with it  $N_r$ ) below a threshold value in clear air. Therefore, it is simple to set  $f_p$  to the fraction of the domain occupied by non-zero values of  $r_r$  and  $N_r$ . Morrison microphysics predicts rain, ice, snow, and graupel. For each of these species, SAM outputs a fraction. To provide an apples-to-apples comparison with CLUBB,  $f_p$  is approximated as the greatest of these four fractions at any particular grid level.

Although  $f_p$  is provided by the LES for this study, it can be diagnosed based on the cloud fraction using a method such as that of Morrison and Gettelman (2008). If the cloud fraction, in turn, is diagnosed based on the omnipresent prediction of means, variances, and other moments — as in higher-order moment parameterizations such as CLUBB — then the onset of partial cloudiness is well defined and indeterminacy about the time of cloud initiation is avoided. In contrast, parameteri-

zations that diagnose cloud fraction based on, e.g., cloud water mixing ratio, lack crucial information in cloudless grid boxes, as discussed in Tompkins (2002). The well-defined onset of CLUBB’s cloud fraction is inherited by the precipitation fraction.

## 5 Results

460 We first evaluate the shape of the idealized PDFs directly against LES. Histograms of SAM LES data are generated from the three-dimensional snapshots of hydrometeor fields. One histogram is generated at every vertical level for each hydrometeor field. A histogram of a SAM hydrometeor field is compared to the CLUBB marginal PDF of that hydrometeor field at the same vertical level and output time. The comparison is done with each of the SL, DL, and DDL PDF shapes.

465 Figure 2 compares marginal PDFs involving  $r_r$  and  $N_r$  for the RICO case at an altitude of 380 m and a time of 4200 min. For the plot of the PDF of  $r_r$  in Fig. 2a, the delta function at  $r_r = 0$  has been omitted. The SAM data is divided into 100 bins, equally-sized in  $r_r$ , that range from the largest value of  $r_r$  to the smallest positive value of  $r_r$ . (In what follows, all histograms use 100 equal-size bins, arranged from smallest to largest value.) The SL hydrometeor PDF shape significantly overpredicts  
470 the PDF at small values of  $r_r$  and significantly underpredicts it at large values of  $r_r$ . These errors are an expected consequence of the single lognormal’s attempt to fit the precipitation-less area. The DL and DDL PDF shapes provide a much closer match qualitatively to the SAM data. A quantitative assessment of the quality of the fit will follow in Section 5.1.

Each of the CLUBB hydrometeor PDF shapes has a lognormal distribution within precipitation  
475 in each PDF component. Taking the natural logarithm of every point of a lognormal distribution produces a normal distribution, and so the plot of the PDF of  $\ln r_r$  in Fig. 2b is a normal distribution in each PDF component for each of the DDL, DL, and SL PDF shapes. The plot of the PDF of  $\ln r_r$  (hereafter referred to as the PDF of  $r_r$  in Gaussian space) complements the aforementioned plot of the PDF of  $r_r$  (Fig. 2a). The plot of the PDF of  $r_r$  is log-scaled on the y-axis, accentuating the small  
480 values of  $P(r_r)$  that are found at large values of  $r_r$ . The plot of the PDF of  $\ln r_r$  accentuates the PDF at small values of  $r_r$ .

The plot of the PDF of  $\ln r_r$  is a plot of only the in-precipitation portion of the distribution, omitting all zero-values. The in-precipitation portion of the PDF is divided by  $f_p$ , which allows the area under the curve to integrate to 1. The PDF shown in Fig. 2b is the Gaussianized form of Eq. (32).

485 Figure 2b shows that the SL hydrometeor PDF shape significantly misses the mark, for its peak is located too far to the left of the bulk of the SAM LES data. This shift of the peak to excessively small values is to be expected of a continuous PDF shape that tries to include a delta function at zero. The DL PDF shape is far too peaked in comparison to the SAM LES data, which is spread out broadly in Gaussian space. The DDL PDF shape is able to achieve a spread-out shape because it has

490 two different means within precipitation. This allows it to better fit the more platykurtic shape of the SAM LES data in Gaussian space.

The plot of the PDF of RICO  $N_r$  is found in Fig. 2c and the Gaussian-space plot of  $N_r$  is found in Fig. 2d. Similar to  $r_r$ , the SL shape overpredicts the PDF at small values of  $N_r$  and underpredicts it at large values of  $N_r$ . In Gaussian space, it is easy to see that SL's peak is located too far to the left. The DDL shape provides a better fit than the DL shape to SAM LES data in Fig. 2c. Again, the DL shape is too peaked in Fig. 2d, whereas the bimodal DDL is able to spread out, which provides a better match to SAM LES data.

Figure 3 contains scatterplots that show the bivariate PDF of  $r_r$  and  $N_r$  for both SAM LES and CLUBB's PDF in RICO at the same altitude and time as Fig. 2. The CLUBB PDF scatterpoints were generated by sampling the DDL PDF using an unweighted Monte Carlo sampling scheme. This demonstrates the advantages of the multivariate nature of CLUBB's PDF. The hydrometeor fields are correlated the same way in CLUBB's PDF as they are in SAM LES.

Figure 4 compares marginal PDFs involving  $r_r$  and  $N_r$  for the DYCOMS-II RF02 case at an altitude of 400 m and a time of 330 min. All three hydrometeor PDF shapes provide a decent match to the SAM LES data. In Fig. 4a and Fig. 4c, the SL and DL PDF shapes dip a little below the SAM LES line in the middle of the data range for  $r_r$  and  $N_r$ , respectively. The DDL PDF shape stays closer to the SAM LES line in this region. Additionally, the SL PDF shape overestimates the SAM LES line close to the y-axis. In Fig. 4b and Fig. 4d, the Gaussian-space plots show that the two components of the DDL shape superimpose more than they did for the RICO case, owing to the reduced in-precipitation variance in the drizzling stratocumulus case.

In order to assess how well the PDF shapes are able to capture ice PDFs as well as liquid PDFs, we turn to the LBA case. In LBA, liquid and ice appear at different altitudes and times. Figure 5 compares marginal PDFs involving  $r_r$  and  $N_r$  for the LBA case at an altitude of 2424 m and a time of 330 min. Compared to SAM's PDF, the DDL hydrometeor PDF shape is too bimodal, but it still provides the best visual match of the three hydrometeor PDF shapes to SAM data. The fit will be quantified in Section 5.1.

To indicate whether the three PDF shapes work for ice-phase hydrometeors, we compare marginal PDFs involving  $r_i$  and  $N_i$  for the LBA case at an altitude of 10500 m and a time of 360 min (Figure 6). Similar to the  $r_r$  and  $N_r$  plots for RICO and LBA, Fig. 6a and Fig. 6c show that the SL PDF shape overpredicts the PDF at small values of  $r_i$  and  $N_i$  and underpredicts it at large values of  $r_i$  and  $N_i$ . The DL shape provides a better fit than the SL, and the DDL has a slightly better fit than the DL. The Gaussian-space plots in Fig. 6b and Fig. 6d show that the SAM LES distribution of  $\ln r_i$  and  $\ln N_i$  is again platykurtic. The SL PDF shape has a peak that is shifted to the left. The DDL hydrometeor PDF shape is able to spread out the most to cover the platykurtic shape of the LES in Gaussian space.



Why does the DDL PDF shape match LES output better than the DL shape in the aforementioned figures? The PDFs (in Gaussian space) for the LES of RICO and LBA show a broad, flat distribution of hydrometeor values from the LES. The DL shape is too peaked in comparison to the LES data. The DDL PDF is able to spread out the component means and thereby represent the platykurtic shape more accurately. However, even the DDL PDF fails to capture the far left-hand tail of the LES PDF. In the RICO, DYCOMS-II RF02, and LBA cases, between about 5% and 20% of the LES PDF is found to the left of the DDL PDF (see Figures 2b, 4b, 5b, and 6b). However, these values of hydrometeor mixing ratios are small. They are roughly a factor of 20 or more smaller than the median value. By combining these factors, we see that the percentage contribution of hydrometeor mixing ratios that are omitted on the left-hand tail is only about 1%.

Why does SAM LES data have a platykurtic shape in Gaussian space in these cases? One possible cause is the partly cloudy (and partly rainy) nature of these cases. In these partly rainy cases, a relatively high percentage of the precipitation occurs in “edge regions” near the non-precipitating region. These regions usually correspond to the edge of cloud or outside of cloud. Evaporation (or less accretion) occurs in these regions, increasing the area occupied by smaller amounts of rain. Yet, there is also an area of more intense precipitation near the center of the precipitating region, which produces larger amounts of rain. Collectively, the areas of small and large rain amount produce the large spread in the hydrometeor spectrum.

The DYCOMS-II RF02 PDFs from the LES tend not to share the platykurtic shape seen in the other cases. The RF02 case is overcast, so there are not as many “edge” regions of precipitation as found in partly rainy cases. There is much less in-precipitation variance in the RF02 case. The simpler PDF shape is easier to fit by all the PDF shapes (SL, DL, and DDL). To further illuminate the physics underlying the PDF shapes produced by LES, further study would be needed.

## 5.1 Quality of fit: general scores

While a lot can be learned by looking at plots of the hydrometeor PDFs, they are anecdotal and cannot tell us how well the idealized PDF shapes work generally. To obtain an *overall* quantification of the quality of the fit, we calculate the Kolmogorov-Smirnov (K-S) and the Cramer-von Mises (C-vM) scores.

Both the K-S and C-vM tests compare the cumulative distribution function (CDF) of the idealized distribution to the CDF of the empirical data (in this case, SAM LES data). Both tests require that the CDFs be continuous. Therefore, the scores are calculated using only the in-precipitation portion of the hydrometeor PDF in Eq. (8). The DDL, DL, and SAM LES data all have the same precipitation fraction. The in-precipitation portion of the PDF is normalized by dividing by precipitation fraction so that it integrates to 1. The equation for the in-precipitation portion of the marginal PDF,  $P(h)|_{ip}$ ,

560 is

$$P(h)|_{\text{ip}} = \xi_{(1)} \frac{f_{p(1)}}{f_p} P_{L(1)}(h) + \xi_{(2)} \frac{f_{p(2)}}{f_p} P_{L(2)}(h), \quad (32)$$

where  $P_{L(i)}$  is given by Eq. (9).

The K-S score is the greatest difference between the empirical in-precipitation CDF,  $C_e(h)|_{\text{ip}}$ , and the idealized in-precipitation CDF,  $C(h)|_{\text{ip}}$ , at any point in  $h > 0$ . In order to run the tests, the SAM LES data from the requested level and time was sorted in the order of increasing value. This was done only for points where the requested hydrometeor was found. The K-S score is given by (Stephens, 1970)

$$\begin{aligned} \text{KS} &= \max_h \left| C_e(h)|_{\text{ip}} - C(h)|_{\text{ip}} \right| = \max(\text{KS}^+, \text{KS}^-), \quad \text{where} \\ \text{KS}^+ &= \max_{1 \leq \kappa \leq n_p} \left( \frac{\kappa}{n_p} - C(h_\kappa)|_{\text{ip}} \right) \quad \text{and} \quad \text{KS}^- = \max_{1 \leq \kappa \leq n_p} \left( C(h_\kappa)|_{\text{ip}} - \frac{\kappa-1}{n_p} \right). \end{aligned} \quad (33)$$

The number of data points in SAM LES where the hydrometeor is found is denoted  $n_p$ , and  $h_\kappa$  is the value of the hydrometeor at SAM LES ordered data point  $\kappa$ .

Unlike the K-S test, which only considers the greatest difference between the CDFs, the C-vM test is based on an integral that includes the differences between the CDFs over the entire distribution. The integral is (Anderson, 1962)

$$\omega^2 = \int \left( C_e(h)|_{\text{ip}} - C(h)|_{\text{ip}} \right)^2 d C(h)|_{\text{ip}}. \quad (34)$$

575 The C-vM score is calculated by (Anderson, 1962; Stephens, 1970)

$$\text{CVM} = \omega^2 n_p = \frac{1}{12n_p} + \sum_{\kappa=1}^{n_p} \left( \frac{2\kappa-1}{2n_p} - C(h_\kappa)|_{\text{ip}} \right)^2. \quad (35)$$

The K-S and C-vM test scores are produced at every LES vertical level and three-dimensional statistical output time for every hydrometeor species. This results in a large number of scores. We desire that each hydrometeor species have a single K-S score and a single C-vM score in order to more easily compare the DDL, DL, and SL hydrometeor shapes. We calculate this score by averaging the individual level scores over multiple levels and multiple output times. For K-S this is simple, and the result is  $\langle \text{KS} \rangle$  (where angle brackets denote an average over multiple levels and times). The C-vM test score in Eq. (35) is dependent on the number of precipitating grid points. This number changes between vertical levels and output times, so the C-vM scores cannot simply be averaged. Rather, they are normalized first by dividing CVM by  $n_p$  to produce  $\omega^2$  at every level and time. Those results are averaged to calculate  $\langle \omega^2 \rangle$ .

After inspecting profiles of SAM LES results for mean mixing ratios in height and time, regions were identified in height and time where the mean mixing ratio of a species was always at least  $5.0 \times 10^{-6} \text{ kg kg}^{-1}$ . Averaging of the scores was restricted to these regions in order to eliminate from consideration levels that do not contain the hydrometeor or contain only small amounts of the

hydrometeor with a small number of samples. RICO test scores for  $r_r$  and  $N_r$  were averaged from the surface through 2780 m and from 4200 min through 4320 min. DYCOMS-II RF02 test scores for  $r_r$  and  $N_r$  were averaged from 277 m through 808 m and from 300 min to 360 min.

The LBA case contains both liquid and frozen-phase hydrometeor species that evolve as the cloud system transitions from shallow to deep convection. The various hydrometeor species develop and maximize at different altitudes and times, so different periods and altitude ranges are chosen for averaging test scores for each species. LBA test scores for  $r_r$  and  $N_r$  were averaged from the surface through 6000 m and from 285 min through 360 min. The test scores for  $r_g$  and  $N_g$  were averaged from 4132 m through 9750 m and from 315 min through 360 min. The test scores for  $r_s$  and  $N_s$  were averaged from 5026 m through 9000 m and from 345 min through 360 min. Finally, the test scores for  $r_i$  and  $N_i$  were averaged from 10250 m through 11750 m at 360 min. For the LBA case, the value of  $f_p$  used by CLUBB's PDF was based on the greatest value of SAM output variables for rain fraction, ice fraction, snow fraction, and graupel fraction. Each of these statistics is the fraction of the SAM domain occupied by values of the relevant mixing ratio of at least  $1.0 \times 10^{-6} \text{ kg kg}^{-1}$ . In order to keep the comparison of the PDF shapes to SAM data consistent, values lower than this threshold were omitted from the calculations of the individual level-and-time scores for K-S and C-vM.

The results of  $\langle \text{KS} \rangle$  are listed in Table 1 for every hydrometeor species in every case. The DDL PDF shape has the lowest average score for every case and hydrometeor species except for one. The DL PDF shape edges out the DDL in the DYCOMS-II RF02  $N_r$  comparison. The SL PDF shape has the highest average score for every case and hydrometeor species, except for the LBA  $r_r$  comparison, where it has the second-lowest score and the DL has the highest score. The results of  $\langle \omega^2 \rangle$  are listed in Table 2. The DDL PDF shape has the lowest average score for every case and hydrometeor species, the DL shape has the second-lowest average score, and the SL shape has the highest average score.

We note the important caveat that, as compared to DL, DDL has more adjustable parameters. A parameterization with more free parameters would be expected to provide a better fit to a training data set. Therefore, although DDL matches the LES output more closely than does DL, we cannot be certain, based on the analysis presented here, that DDL will outperform DL on a different validation dataset. For a deeper analysis, one could use a model selection method that penalizes parameterizations with more parameters. We leave such an analysis for future work.

## 5.2 Microphysical process rates

A primary reason to improve the accuracy of hydrometeor PDFs is to improve the accuracy of the calculation of microphysical process rates. In this section, we compare the accuracy of calculations of microphysical process rates based on the SL, DL, and DDL PDF shapes.

In the simulations of RICO and DYCOMS-II RF02, both SAM LES and CLUBB use KK micro-  
physics. The process rates output are the mean evaporation rate of  $r_r$ , the mean accretion rate of  $r_r$ ,  
and the mean autoconversion rate of  $r_r$ . Also recorded is rain drop mean volume radius, which is  
important for sedimentation velocity of rain. In order to account for subgrid variability in the micro-  
physics, the KK microphysics process rate equations have been upscaled (to grid-box scale) using  
analytic integration over the PDF (Larson and Griffin, 2013; Griffin and Larson, 2013). The updates  
to the multivariate PDF (see Section 2) require updates to the upscaled process rate equations. The  
updated forms of these equations are listed in the Supplement.

Figure 7 shows profiles of RICO mean microphysics process rates. The mean evaporation rate  
profile in Fig. 7a shows that all three shapes over-evaporate at higher altitudes, but that SL and DL  
over-evaporate more than DDL. It should be noted that the reason for the over-evaporation at higher  
altitudes in the RICO case is the marginal PDF of  $\chi$  produced by ADG2. While it provides a good  
match between CLUBB and SAM LES in the fields of cloud fraction and  $\bar{r}_c$ , the value of  $\sigma_{\chi(1)}$  is  
far too large. When  $\chi$  and  $r_r$  (or  $N_r$ ) are distributed jointly, this results in too many large values of  
 $r_r$  (or  $N_r$ ) being placed in air that is far too dry. RICO mean evaporation rate could benefit from an  
improved ADG2 in order to produce a better marginal distribution of  $\chi$ , but that is beyond the scope  
of this paper.

Figure 7b shows that both the DL and DDL PDF shapes match the LES mean accretion rate  
profile much better than does the SL shape. The mean autoconversion rate depends on  $\chi$  and  $N_{cn}$   
but not hydrometeor variables, and so the autoconversion rate is the same for all three PDF shapes  
(not shown). The overall mean microphysics rate — i.e., the sum of the evaporation, accretion, and  
autoconversion rates — is fit best by the DDL shape and worst by the SL shape. Both DDL and DL  
are a much better match to the SAM profile of rain drop mean volume radius than SL (Fig. 7d).

Figure 8 shows that all three hydrometeor PDF shapes provide a good match to SAM LES for  
DYCOMS-II RF02. In Fig. 8d, the SL PDF shape deviates more strongly from SAM LES than does  
DL or DDL near the bottom of the profile of rain drop mean volume radius.

In the simulation of LBA, Morrison microphysics was used in both the SAM LES and CLUBB.  
In order to account for subgrid variability in the microphysics, sample points from the PDF are  
produced at every grid level using the Subgrid Importance Latin Hypercube Sampler (SILHS)  
(Raut and Larson, 2016; Larson and Schanen, 2013; Larson et al., 2005). For the LBA case, 128  
sample points were drawn. Morrison microphysics is then called using each set of sample points,  
and the results are averaged to calculate the mean microphysics process rates.

Figure 9 shows the same mean microphysics process rates as in previous figures, but here for LBA.  
The profile of mean evaporation rate in Fig. 9a shows that DDL is the best match to SAM LES. The  
profile of mean accretion rate in Fig. 9b shows that DDL is the best match to SAM, followed by DL  
and then SL. The overall (autoconversion + accretion + evaporation) warm microphysics process

rate profile is best matched by the DDL hydrometeor PDF shape, followed by the DL shape, which in turn is followed by the SL shape (Fig. 9c).

## 6 Conclusions

665 The multivariate PDF used by CLUBB has been updated to improve the subgrid representation of hydrometeor species. The most important update is the introduction of precipitation fraction to the PDF. The precipitating fraction contains any non-zero values of any hydrometeor species included in the microphysics scheme. The remainder of the subgrid domain is precipitation-less and is represented by a delta function where every hydrometeor species has a value of zero. When a hydrometeor is  
670 found at a grid level, its representation in the precipitating portion of the subgrid domain is a lognormal or double lognormal distribution. The introduction of precipitation fraction increases accretion and decreases evaporation in cumulus cases, allowing more precipitation to reach the ground.

Additionally, a new method has been developed to calculate the in-precipitation mean and standard deviation of a hydrometeor species in each component of CLUBB's two-component PDF. This  
675 method preserves the grid-box mean and variance of the hydrometeor species. By simply changing the values of tunable parameters, CLUBB's marginal PDF for a hydrometeor can be changed from a delta-double-lognormal (DDL) to a delta-lognormal (DL) or to a single-lognormal (SL) shape.

In order to compare the effectiveness of the three hydrometeor PDF shapes, three simulations – a precipitating shallow cumulus case (RICO), a drizzling stratocumulus case (DYCOMS-II RF02),  
680 and a deep convective case (LBA) – were run using SAM LES. Statistical output values from the LES for the grid-level mean and turbulent fields were used to drive the PDF for each hydrometeor PDF shape. The idealized PDF shapes were compared to the SAM LES results. The DDL PDF shape produced the lowest average K-S and average normalized C-vM scores when compared to SAM LES results, followed by the DL PDF shape. Both produced lower scores than the original SL PDF shape.  
685 However, for DYCOMS-II RF02, all three PDF shapes were in almost equal agreement with SAM LES results.

The DL and DDL PDFs possess three important properties: 1) they are multivariate, and hence can represent interactions among multiple hydrometeor species; 2) they admit a precipitation-less region, which is necessary to permit realistic process rates in cumulus cloud layers; and 3) they have  
690 realistic tails, as evidenced by the comparisons with LES shown here. Because of these three properties, the DL and DDL PDFs may be general enough and accurate enough to adequately represent hydrometeor variability over a range of important cloud types, including shallow cumulus, deep cumulus, and stratocumulus clouds. This generality, in turn, may help enable parameterization of these clouds types in a more unified way. Indeed, an early version of the DDL PDF has already been used  
695 to represent hydrometeor subgrid variability in some interactive simulations with a unified cloud parameterization. Namely, the DDL PDF was used in the interactive single-column simulations of

these cloud types by Storer et al. (2015) and in the global simulations by Thayer-Calder et al. (2015). Further testing would be required, however, to better understand the limits of the DL and DDL PDFs. Better understanding is particularly desirable in, for instance, mixed-phase and glaciated clouds. This  
700 has been left for future work.

## Appendix A: Back-solving PDF component correlations

In Section 5, mean microphysics process rates were calculated either by using the analytical integration of a local microphysics scheme or by using SILHS to sample the PDF in order to drive a local microphysics scheme. Both methods require information on the PDF component correlations.  
705 These correlations can be back-solved when given the overall (grid-box) covariance of the necessary variables.

### A1 PDF component correlation of a binormal variate and a hydrometeor

The PDF *component* correlation of a binormal variate (using  $r_t$  as an example) and a hydrometeor can be back-solved when their covariance,  $\overline{r'_t h'}$ , is provided. Their covariance can be written in terms  
710 of PDF parameters by integrating over the PDF, such that

$$\overline{r'_t h'} = \int_{-\infty}^{\infty} \int_0^{\infty} (r_t - \bar{r}_t) (h - \bar{h}) P(r_t, h) dh dr_t, \quad (\text{A1})$$

where  $P(r_t, h)$  is the bivariate marginal PDF of  $r_t$  and  $h$ . This equation can be rewritten as

$$\begin{aligned} \overline{r'_t h'} = \sum_{i=1}^n \xi_{(i)} \int_{-\infty}^{\infty} \int_0^{\infty} (r_t - \bar{r}_t) (h - \bar{h}) (f_{p(i)} P_{NL(i)}(r_t, h) \\ + (1 - f_{p(i)}) P_{N(i)}(r_t) \delta(h)) dh dr_t, \end{aligned} \quad (\text{A2})$$

where  $P_{NL(i)}(r_t, h)$  is the  $i$ th component bivariate PDF involving one normal variate and one log-normal variate, and where  $P_{N(i)}(r_t)$  is a normal distribution in the  $i$ th component. This equation is  
715 integrated and reduced, resulting in

$$\overline{r'_t h'} = \sum_{i=1}^n \xi_{(i)} f_{p(i)} (\mu_{r_t(i)} - \bar{r}_t + \tilde{\rho}_{r_t, h(i)} \sigma_{r_t(i)} \tilde{\sigma}_{h(i)}) \mu_{h(i)}, \quad (\text{A3})$$

where  $\mu_{r_t(i)}$  and  $\sigma_{r_t(i)}$  are the mean and standard deviation, respectively, of  $r_t$  in the  $i$ th PDF component.

720 The variable that needs to be solved for is  $\tilde{\rho}_{r_t, h(i)}$ , which is the in-precipitation correlation of  $r_t$  and  $\ln h$  in the  $i$ th PDF component. This is the normal-space correlation that is required for use in the microphysics. It is related to the  $i$ th component in-precipitation correlation of  $r_t$  and  $h$ ,  $\rho_{r_t, h(i)}$ , by

$$\rho_{r_t, h(i)} = \tilde{\rho}_{r_t, h(i)} \tilde{\sigma}_{h(i)} \frac{\mu_{h(i)}}{\sigma_{h(i)}}. \quad (\text{A4})$$

725 The covariance  $\overline{r'_t h'}$  given by Eq. (A3) can be written in terms of CLUBB's two-component PDF ( $n = 2$ ) as

$$\begin{aligned} \overline{r'_t h'} &= \xi_{(1)} f_{p(1)} \left( \mu_{r_t(1)} - \bar{r}_t + \tilde{\rho}_{r_t, h(1)} \sigma_{r_t(1)} \tilde{\sigma}_{h(1)} \right) \mu_{h(1)} \\ &\quad + \xi_{(2)} f_{p(2)} \left( \mu_{r_t(2)} - \bar{r}_t + \tilde{\rho}_{r_t, h(2)} \sigma_{r_t(2)} \tilde{\sigma}_{h(2)} \right) \mu_{h(2)}. \end{aligned} \quad (\text{A5})$$

The overall covariance is provided, so the component correlation can be back-solved by setting  $\tilde{\rho}_{r_t, h(1)} = \tilde{\rho}_{r_t, h(2)} (= \tilde{\rho}_{r_t, h})$ . The result is

$$730 \quad \tilde{\rho}_{r_t, h} = \frac{\overline{r'_t h'} - \xi_{(1)} f_{p(1)} \left( \mu_{r_t(1)} - \bar{r}_t \right) \mu_{h(1)} - \xi_{(2)} f_{p(2)} \left( \mu_{r_t(2)} - \bar{r}_t \right) \mu_{h(2)}}{\xi_{(1)} f_{p(1)} \sigma_{r_t(1)} \tilde{\sigma}_{h(1)} \mu_{h(1)} + \xi_{(2)} f_{p(2)} \sigma_{r_t(2)} \tilde{\sigma}_{h(2)} \mu_{h(2)}}, \quad (\text{A6})$$

where  $-1 \leq \tilde{\rho}_{r_t, h} \leq 1$ .

The equation for  $\overline{r'_t h'}$  given in Eq. (A5) is for a fully-varying PDF in both components ( $\sigma_{r_t(i)} > 0$  and  $\sigma_{h(i)} > 0$ ). A variable may have a constant value in a PDF sub-component. When this happens, the PDF of the constant variable is a delta function at the  $i$ th sub-component mean. When

735  $\sigma_{r_t(i)} > 0$  and  $\sigma_{h(i)} = 0$ ,  $r_t$  varies in  $i$ th component but  $h$  is constant within precipitation. The PDF  $P_{NL(i)}(r_t, h)$  becomes  $P_{N(i)}(r_t) \delta(h - \mu_{h(i)})$ . There also may be situations where  $\sigma_{r_t(i)} = 0$  but  $\sigma_{h(i)} > 0$ , or even where  $\sigma_{r_t(i)} = 0$  and  $\sigma_{h(i)} = 0$ .

When  $\sigma_{r_t(1)} \sigma_{h(1)} > 0$  but  $\sigma_{r_t(2)} \sigma_{h(2)} = 0$ , the equation for  $\overline{r'_t h'}$  is written as

$$\overline{r'_t h'} = \xi_{(1)} f_{p(1)} \left( \mu_{r_t(1)} - \bar{r}_t + \tilde{\rho}_{r_t, h(1)} \sigma_{r_t(1)} \tilde{\sigma}_{h(1)} \right) \mu_{h(1)} + \xi_{(2)} f_{p(2)} \left( \mu_{r_t(2)} - \bar{r}_t \right) \mu_{h(2)}. \quad (\text{A7})$$

740 The above equation can be rewritten to solve for  $\tilde{\rho}_{r_t, h(1)}$ , such that

$$\tilde{\rho}_{r_t, h(1)} = \frac{\overline{r'_t h'} - \xi_{(1)} f_{p(1)} \left( \mu_{r_t(1)} - \bar{r}_t \right) \mu_{h(1)} - \xi_{(2)} f_{p(2)} \left( \mu_{r_t(2)} - \bar{r}_t \right) \mu_{h(2)}}{\xi_{(1)} f_{p(1)} \sigma_{r_t(1)} \tilde{\sigma}_{h(1)} \mu_{h(1)}}, \quad (\text{A8})$$

while  $\tilde{\rho}_{r_t, h(2)}$  is undefined and irrelevant to the microphysics. When  $\sigma_{r_t(1)} \sigma_{h(1)} = 0$  but  $\sigma_{r_t(2)} \sigma_{h(2)} > 0$ , the equation for  $\overline{r'_t h'}$  is analogous to Eq. (A7). An equation analogous to Eq. (A8) solves for  $\tilde{\rho}_{r_t, h(2)}$ , while  $\tilde{\rho}_{r_t, h(1)}$  is undefined. In a scenario where  $\sigma_{r_t(1)} \sigma_{h(1)} = 0$  and  $\sigma_{r_t(2)} \sigma_{h(2)} =$

745 0, the equation for  $\overline{r'_t h'}$  is

$$\overline{r'_t h'} = \xi_{(1)} f_{p(1)} \left( \mu_{r_t(1)} - \bar{r}_t \right) \mu_{h(1)} + \xi_{(2)} f_{p(2)} \left( \mu_{r_t(2)} - \bar{r}_t \right) \mu_{h(2)}. \quad (\text{A9})$$

When this is the case, both  $\tilde{\rho}_{r_t, h(1)}$  and  $\tilde{\rho}_{r_t, h(2)}$  are undefined.

This method of back-solving for the component correlations was used to calculate the PDF component correlations of  $r_t$  and  $r_r$ ,  $r_t$  and  $N_r$ ,  $\theta_l$  and  $r_r$ , and  $\theta_l$  and  $N_r$ . These were the only correlations

750 of this type that were necessary to produce the microphysics process rates used in the comparison.

## A2 PDF component correlation of two hydrometeors

The PDF *component* correlation of two hydrometeors,  $h_x$  and  $h_y$ , can be back-solved when their covariance,  $\overline{h'_x h'_y}$ , is provided. Their covariance can be written in terms of PDF parameters by inte-

grating over the PDF, such that

$$755 \quad \overline{h'_x h'_y} = \int_0^\infty \int_0^\infty (h_x - \overline{h_x}) (h_y - \overline{h_y}) P(h_x, h_y) dh_y dh_x, \quad (\text{A10})$$

where  $P(h_x, h_y)$  is the bivariate marginal PDF of  $h_x$  and  $h_y$ . This equation can be rewritten as

$$\begin{aligned} \overline{h'_x h'_y} = \sum_{i=1}^n \xi_{(i)} \int_0^\infty \int_0^\infty (h_x - \overline{h_x}) (h_x - \overline{h_y}) (f_{p(i)} P_{LL(i)}(h_x, h_y) \\ + (1 - f_{p(i)}) \delta(h_x) \delta(h_y)) dh_y dh_x, \end{aligned} \quad (\text{A11})$$

where  $P_{LL(i)}(h_x, h_y)$  is the  $i$ th component bivariate PDF involving two lognormal variates.

This equation is integrated and reduced, resulting in

$$760 \quad \overline{h'_x h'_y} = -\overline{h_x} \overline{h_y} + \sum_{i=1}^n \xi_{(i)} f_{p(i)} (\mu_{h_x(i)} \mu_{h_y(i)} + \rho_{h_x, h_y(i)} \sigma_{h_x(i)} \sigma_{h_y(i)}), \quad (\text{A12})$$

where  $\rho_{h_x, h_y(i)}$  is the in-precipitation correlation of  $h_x$  and  $h_y$  in the  $i$ th PDF component. When the PDF is fully-varying in both components ( $\sigma_{h_x(i)} > 0$  and  $\sigma_{h_y(i)} > 0$ ), the covariance  $\overline{h'_x h'_y}$  given by Eq. (A12) can be written in terms of CLUBB's two-component PDF as

$$\begin{aligned} \overline{h'_x h'_y} = \xi_{(1)} f_{p(1)} (\mu_{h_x(1)} \mu_{h_y(1)} + \rho_{h_x, h_y(1)} \sigma_{h_x(1)} \sigma_{h_y(1)}) \\ + \xi_{(2)} f_{p(2)} (\mu_{h_x(2)} \mu_{h_y(2)} + \rho_{h_x, h_y(2)} \sigma_{h_x(2)} \sigma_{h_y(2)}) - \overline{h_x} \overline{h_y}. \end{aligned} \quad (\text{A13})$$

765 The overall covariance is provided, so the component correlation is solved by setting  $\rho_{h_x, h_y(1)} = \rho_{h_x, h_y(2)}$  ( $= \rho_{h_x, h_y}$ ). The result is

$$\rho_{h_x, h_y} = \frac{\overline{h'_x h'_y} + \overline{h_x} \overline{h_y} - \xi_{(1)} f_{p(1)} \mu_{h_x(1)} \mu_{h_y(1)} - \xi_{(2)} f_{p(2)} \mu_{h_x(2)} \mu_{h_y(2)}}{\xi_{(1)} f_{p(1)} \sigma_{h_x(1)} \sigma_{h_y(1)} + \xi_{(2)} f_{p(2)} \sigma_{h_x(2)} \sigma_{h_y(2)}}. \quad (\text{A14})$$

When  $\sigma_{h_x(1)} \sigma_{h_y(1)} > 0$  but  $\sigma_{h_x(2)} \sigma_{h_y(2)} = 0$ , the equation for  $\overline{h'_x h'_y}$  is written as

$$\overline{h'_x h'_y} = \xi_{(1)} f_{p(1)} (\mu_{h_x(1)} \mu_{h_y(1)} + \rho_{h_x, h_y(1)} \sigma_{h_x(1)} \sigma_{h_y(1)}) + \xi_{(2)} f_{p(2)} \mu_{h_x(2)} \mu_{h_y(2)} - \overline{h_x} \overline{h_y}. \quad (\text{A15})$$

770 The above equation can be rewritten to solve for  $\rho_{h_x, h_y(1)}$ , such that

$$\rho_{h_x, h_y(1)} = \frac{\overline{h'_x h'_y} + \overline{h_x} \overline{h_y} - \xi_{(1)} f_{p(1)} \mu_{h_x(1)} \mu_{h_y(1)} - \xi_{(2)} f_{p(2)} \mu_{h_x(2)} \mu_{h_y(2)}}{\xi_{(1)} f_{p(1)} \sigma_{h_x(1)} \sigma_{h_y(1)}}, \quad (\text{A16})$$

while  $\rho_{h_x, h_y(2)}$  is undefined and irrelevant to the microphysics. When  $\sigma_{h_x(1)} \sigma_{h_y(1)} = 0$  but  $\sigma_{h_x(2)} \sigma_{h_y(2)} > 0$ , the equation for  $\overline{h'_x h'_y}$  is analogous to Eq. (A15). An equation analogous to Eq. (A16) solves for  $\rho_{h_x, h_y(2)}$ , while  $\rho_{h_x, h_y(1)}$  is undefined. In a scenario where  $\sigma_{h_x(1)} \sigma_{h_y(1)} = 0$

775 and  $\sigma_{h_x(2)} \sigma_{h_y(2)} = 0$ , the equation for  $\overline{h'_x h'_y}$  is

$$\overline{h'_x h'_y} = \xi_{(1)} f_{p(1)} \mu_{h_x(1)} \mu_{h_y(1)} + \xi_{(2)} f_{p(2)} \mu_{h_x(2)} \mu_{h_y(2)} - \overline{h_x} \overline{h_y}. \quad (\text{A17})$$



When this is the case, both  $\rho_{h_x, h_y(1)}$  and  $\rho_{h_x, h_y(2)}$  are undefined.

The variable that needs to be solved for is  $\tilde{\rho}_{h_x, h_y(i)}$ , which is the in-precipitation correlation of  $\ln h_x$  and  $\ln h_y$  in the  $i$ th PDF component. This is the normal-space correlation that is required for use in the microphysics, and it is given by

$$\tilde{\rho}_{h_x, h_y(i)} = \frac{\ln \left( 1 + \rho_{h_x, h_y(i)} \frac{\sigma_{h_x(i)} \sigma_{h_y(i)}}{\mu_{h_x(i)} \mu_{h_y(i)}} \right)}{\tilde{\sigma}_{h_x(i)} \tilde{\sigma}_{h_y(i)}}, \quad (\text{A18})$$

where  $-1 \leq \tilde{\rho}_{h_x, h_y(i)} \leq 1$ .

This method of back-solving for the component correlations was used to calculate the PDF component correlation of  $r_r$  and  $N_r$ . This was the only correlation of this type that was necessary to produce the microphysics process rates used in the comparison.

### Appendix B: The relationship between $\overline{N_c}$ and $\overline{N_{cn}}$

The relationship between  $N_c$  and  $N_{cn}$  is given in Eq. (4) and described in Section 2. The overall (grid-box) mean value of cloud droplet concentration,  $\overline{N_c}$ , is calculated by integrating over the product of Eq. (4) and the PDF of  $\chi$  and  $N_{cn}$ , such that

$$\overline{N_c} = \int_{-\infty}^{\infty} \int_0^{\infty} N_{cn} H(\chi) \sum_{i=1}^n \xi_{(i)} P_{(i)}(\chi, N_{cn}) dN_{cn} d\chi, \quad (\text{B1})$$

where  $P_{(i)}(\chi, N_{cn})$  is the bivariate marginal PDF of  $\chi$  and  $N_{cn}$  in the  $i$ th PDF component. This equation can be rewritten as

$$\begin{aligned} \overline{N_c} &= \sum_{i=1}^n \xi_{(i)} \int_0^{\infty} \int_0^{\infty} N_{cn} P_{NL(i)}(\chi, N_{cn}) dN_{cn} d\chi \\ &= \sum_{i=1}^n \xi_{(i)} \text{NC}_{(i)}, \end{aligned} \quad (\text{B2})$$

where  $P_{NL(i)}(\chi, N_{cn})$  is the  $i$ th component bivariate PDF involving one normal variate and one lognormal variate. This equation is integrated and reduced, resulting in

$$\text{NC}_{(i)} = \frac{1}{2} \mu_{N_{cn}(i)} \text{erfc} \left( -\frac{1}{\sqrt{2}} \left( \frac{\mu_{\chi(i)}}{\sigma_{\chi(i)}} + \tilde{\rho}_{\chi, N_{cn}(i)} \tilde{\sigma}_{N_{cn}(i)} \right) \right), \quad (\text{B3})$$

where  $\text{erfc}(x)$  is the complimentary error function, and where  $\mu_{\chi(i)}$  and  $\sigma_{\chi(i)}$  are the mean and standard deviation, respectively, of  $\chi$  in the  $i$ th PDF component. Additionally,  $\tilde{\sigma}_{N_{cn}(i)}$  is the standard deviation of  $\ln N_{cn}$  in the  $i$ th PDF component and  $\tilde{\rho}_{\chi, N_{cn}(i)}$  is the correlation of  $\chi$  and  $\ln N_{cn}$  in the  $i$ th PDF component.

The evaluated integral for  $\text{NC}_{(i)}$  given in Eq. (B3) is for a fully-varying PDF in the  $i$ th component ( $\sigma_{\chi(i)} > 0$  and  $\sigma_{N_{cn}(i)} > 0$ ). There are times when a variable may have a constant value in a PDF component. When this happens, the PDF of the constant variable is a delta function at the  $i$ th

PDF component mean. When  $\sigma_{\chi(i)} > 0$  and  $\sigma_{N_{cn}(i)} = 0$ ,  $\chi$  varies in  $i$ th component but  $N_{cn}$  is constant. The PDF  $P_{NL(i)}(\chi, r_r)$  becomes  $P_{N(i)}(\chi) \delta(N_{cn} - \mu_{N_{cn}(i)})$ . The integral is solved and the equation for  $\text{NC}_{(i)}$  becomes

$$\text{NC}_{(i)} = \mu_{N_{cn}(i)} \frac{1}{2} \operatorname{erfc} \left( -\frac{\mu_{\chi(i)}}{\sqrt{2}\sigma_{\chi(i)}} \right). \quad (\text{B4})$$

Likewise, when  $\sigma_{\chi(i)} = 0$  and  $\sigma_{N_{cn}(i)} > 0$ ,  $\chi$  varies in  $i$ th component but  $N_{cn}$  is constant, and when both  $\sigma_{\chi(i)} = 0$  and  $\sigma_{N_{cn}(i)} = 0$ , both  $\chi$  and  $N_{cn}$  are constant in the  $i$ th PDF component. In either scenario, the equation becomes

$$\text{NC}_{(i)} = \begin{cases} \mu_{N_{cn}(i)}, & \text{when } \mu_{\chi(i)} > 0; \text{ and} \\ 0, & \text{when } \mu_{\chi(i)} \leq 0. \end{cases} \quad (\text{B5})$$

It is important to be able to back-solve  $\overline{N_{cn}}$  from  $\overline{N_c}$  because numerical models and microphysics schemes usually either predict  $\overline{N_c}$ , predict the mean in-cloud value of cloud droplet concentration,  $\overline{N_c}|_{\text{ic}}$ , or specify the in-cloud value of cloud droplet concentration. The value of  $\overline{N_{cn}}$  needs to be calculated from this information for use in the PDF. The relationship between grid-box mean  $N_c$  and the mean in-cloud value of cloud droplet concentration is  $\overline{N_c}|_{\text{ic}} = \overline{N_c}/f_c$ , where  $f_c$  is cloud fraction. CLUBB calculates cloud fraction by integrating over the PDF of  $\chi$ , such that

$$f_c = \int_{-\infty}^{\infty} H(\chi) \sum_{i=1}^n \xi_{(i)} P_{(i)}(\chi) d\chi = \sum_{i=1}^n \xi_{(i)} f_{c(i)}, \quad (\text{B6})$$

where  $f_{c(i)}$  is the cloud fraction in the  $i$ th PDF component, which is calculated by

$$f_{c(i)} = \begin{cases} \frac{1}{2} \operatorname{erfc} \left( -\frac{\mu_{\chi(i)}}{\sqrt{2}\sigma_{\chi(i)}} \right), & \text{when } \sigma_{\chi(i)} > 0; \\ 1, & \text{when } \sigma_{\chi(i)} = 0 \text{ and } \mu_{\chi(i)} > 0; \text{ and} \\ 0, & \text{when } \sigma_{\chi(i)} = 0 \text{ and } \mu_{\chi(i)} \leq 0. \end{cases} \quad (\text{B7})$$

The value of  $\overline{N_{cn}}$  from can be back-solved  $\overline{N_c}$ . As described in Section 3,  $\mu_{N_{cn}(i)} = \overline{N_{cn}}$  and  $\sigma_{N_{cn}(i)} = \sqrt{\overline{N_{cn}'^2}}$ . Additionally,  $\tilde{\rho}_{\chi, N_{cn}(i)}$  is related to  $\rho_{\chi, N_{cn}(i)}$  in the manner described by Eq. (A4). These substitutions are made into Eq. (B3), Eq. (B4), and Eq. (B5). However, further simplification is necessary to solve for  $\overline{N_{cn}}$ . The value of  $\overline{N_{cn}'^2}/\overline{N_{cn}}^2$  is set to a prescribed constant value, denoted  $C_{N_{cn}^*}$ . Additionally, the correlation  $\rho_{\chi, N_{cn}(i)}$  is also set to a prescribed value,  $\rho_{\chi, N_{cn}^*}$ . The value of  $N_{cn}$  can now be solved by

$$\overline{N_{cn}} = \frac{\overline{N_c}}{\sum_{i=1}^n \xi_{(i)} \text{FCTR}_{(i)}}, \quad (\text{B8})$$

where  $\text{FCCTR}_{(i)}$  is the factor given by

$$\text{FCCTR}_{(i)} = \begin{cases} \frac{1}{2} \operatorname{erfc} \left( -\frac{1}{\sqrt{2}} \left( \frac{\mu_{\chi(i)}}{\sigma_{\chi(i)}} + \rho_{\chi, N_{cn}^*} \sqrt{C_{N_{cn}^*}} \right) \right), & \text{when } \sigma_{\chi(i)} > 0 \text{ and } C_{N_{cn}^*} > 0; \\ \frac{1}{2} \operatorname{erfc} \left( -\frac{\mu_{\chi(i)}}{\sqrt{2} \sigma_{\chi(i)}} \right), & \text{when } \sigma_{\chi(i)} > 0 \text{ and } C_{N_{cn}^*} = 0; \\ 1, & \text{when } \sigma_{\chi(i)} = 0 \text{ and } \mu_{\chi(i)} > 0; \text{ and} \\ 0, & \text{when } \sigma_{\chi(i)} = 0 \text{ and } \mu_{\chi(i)} \leq 0. \end{cases} \quad (\text{B9})$$

830 When the last scenario is found in all components, the denominator of Eq. (B8) is 0. However, cloud fraction is 0, meaning  $\overline{N_c}$  is also 0, leaving  $\overline{N_{cn}}$  undefined. For numerical purposes,  $\overline{N_{cn}}$  can be set to a reasonable value in this scenario.

The relationship between  $\overline{N_{cn}}$  and  $\overline{N_{c|ic}}$  is found by comparing Eq. (B8) and Eq. (B9) with Eq. (B6) and Eq. (B7). Three scenarios emerge in the comparison. When  $N_{cn}$  is constant (used for  
835 prescribing a constant value of in-cloud  $N_c$ , as was done for the RICO, DYCOMS-II RF02, and LBA cases in this paper),  $\overline{N_{cn}} = \overline{N_{c|ic}}$ . When  $N_{cn}$  varies, but the correlation of  $\chi$  and  $N_{cn}$  is 0 (or undefined when  $\chi$  is constant), it still holds that  $\overline{N_{cn}} = \overline{N_{c|ic}}$ . However, when  $N_{cn}$  varies and the correlation of  $\chi$  and  $N_{cn}$  is a value other than 0,  $\overline{N_{cn}}$  differs from  $\overline{N_{c|ic}}$ .

In a situation where CLUBB is using SILHS with a microphysics scheme that predicts  $\overline{N_c}$ , a  
840 model time step works as follows. The time step begins with the most recent value of  $\overline{N_c}$ . The value of  $\overline{N_{cn}}$  is calculated for use in the PDF. Then, the PDF parameters are calculated, including those involving  $N_{cn}$ . SILHS uses the PDF parameters to generate sample points of all variables involved in the PDF, including  $\chi$  and  $N_{cn}$ . The value of  $N_c$  at these sample points is calculated according to Eq. (4). The sample points of  $N_c$  are then fed to the microphysics scheme, which requires  $N_c$  as  
845 input. The microphysics tendency of  $N_c$  is output from SILHS, and is used to advance the predictive equation for  $\overline{N_c}$  for the next model time step.

### Code availability

The CLUBB code is freely available for non-commercial use after registering for an account on the website <http://clubb.larson-group.com>. The specific ver-  
850 sion of CLUBB used in this paper is available in the SVN repository located at [http://carson.math.uwm.edu/repos/clubb\\_repos/tags/Hydromet\\_PDF\\_shapes](http://carson.math.uwm.edu/repos/clubb_repos/tags/Hydromet_PDF_shapes). In the repository is a file README\_Hydromet\_PDF\_shapes which gives instructions for generating the results found in this paper.

*Acknowledgements.* The authors gratefully acknowledge financial support from the National Science Founda-  
855 tion under Grant No. AGS-0968640 and the Office of Science (BER), U. S. Department of Energy under Grant

No. DE-SC0008323 (Scientific Discoveries through Advanced Computing, SciDAC). Some simulations presented here were performed on the Avi high-performance computer cluster at the University of Wisconsin – Milwaukee.

## References

- 860 Ackerman, A. S., van Zanten, M. C., Stevens, B., Savic-Jovicic, V., Bretherton, C. S., Chlond, A., Golaz, J.-C., Jiang, H., Khairoutdinov, M., Krueger, S. K., Lewellen, D. C., Lock, A., Moeng, C.-H., Nakamura, K., Petters, M. D., Snider, J. R., Weinbrecht, S., and Zulauf, M.: Large-Eddy Simulations of a Drizzling, Stratocumulus-Topped Marine Boundary Layer, *Mon. Wea. Rev.*, 137, 1083–1110, doi:10.1175/2008MWR2582.1, 2009.
- 865 Anderson, T. W.: On the Distribution of the Two-Sample Cramer-von Mises Criterion, *Ann. Math. Statist.*, 33, 1148–1159, 1962.
- Bogenschutz, P. A. and Krueger, S. K.: A simplified PDF parameterization of subgrid-scale clouds and turbulence for cloud-resolving models, *J. Adv. Model. Earth Syst.*, 5, doi:10.1002/jame.20018, 2013.
- Bogenschutz, P. A., Krueger, S. K., and Khairoutdinov, M.: Assumed Probability Density Functions for Shallow and Deep Convection, *J. Adv. Model. Earth Syst.*, 2, doi:10.3894/JAMES.2010.2.10, 2010.
- 870 Boutle, I., Abel, S., Hill, P., and Morcrette, C.: Spatial variability of liquid cloud and rain: Observations and microphysical effects, *Quarterly Journal of the Royal Meteorological Society*, 140, 583–594, 2014.
- Cheng, A. and Xu, K.-M.: A PDF-Based Microphysics Parameterization for Simulation of Drizzling Boundary Layer Clouds, *J. Atmos. Sci.*, 66, 2317–2334, 2009.
- 875 Cheng, A. and Xu, K.-M.: Improved Low-cloud Simulation from the Community Atmosphere Model with an Advanced Third-order Turbulence Closure, *Journal of Climate*, 28, doi:http://dx.doi.org/10.1175/JCLI-D-14-00776.1, 2015.
- Deardorff, J. W.: Stratocumulus-capped mixed layers derived from a three-dimensional model, *Bound.-Layer Meteor.*, 18, 495–527, 1980.
- 880 Firl, G. J. and Randall, D. A.: Fitting and Analyzing LES Using Multiple Trivariate Gaussians, *J. Atmos. Sci.*, 72, 1094–1116, doi:10.1175/JAS-D-14-0192.1, 2015.
- Fletcher, S. J. and Zupanski, M.: A hybrid multivariate Normal and lognormal distribution for data assimilation, *Atmos. Sci. Lett.*, 7, 43–46, doi:10.1002/asl.128, 2006.
- Golaz, J.-C., Larson, V. E., and Cotton, W. R.: A PDF-based model for boundary layer clouds. Part I: Method and model description, *J. Atmos. Sci.*, 59, 3540–3551, 2002a.
- 885 Golaz, J.-C., Larson, V. E., and Cotton, W. R.: A PDF-based model for boundary layer clouds. Part II: Model results, *J. Atmos. Sci.*, 59, 3552–3571, 2002b.
- Grabowski, W. W., Bechtold, P., Cheng, A., Forbes, R., Halliwell, C., Khairoutdinov, M., Lang, S., Nasuno, T., Petch, J., Tao, W. K., Wong, R., Wu, X., and Xu, K. M.: Daytime convective development over land: A model intercomparison based on LBA observations, *Quart. J. Roy. Meteor. Soc.*, 132, 317–344, 2006.
- 890 Griffin, B. M. and Larson, V. E.: Analytic upscaling of local microphysics parameterizations, Part II: Simulations, *Quart. J. Royal Met. Soc.*, 139, 58–69, 2013.
- Guo, H., Golaz, J.-C., Donner, L., Wyman, B., Zhao, M., and Ginoux, P.: CLUBB as a unified cloud parameterization: opportunities and challenges, *Geophys. Res. Lett.*, 42, 4540–4547, doi:10.1002/2015GL063672, 2015.
- 895 Kärcher, B. and Burkhardt, U.: A cirrus cloud scheme for general circulation models, *Quart. J. Royal Met. Soc.*, 134, 1439–1461, doi:10.1002/qj.301, 2008.

- Khairoutdinov, M. and Kogan, Y.: A new cloud physics parameterization in a large-eddy simulation model of marine stratocumulus, *Mon. Wea. Rev.*, 128, 229–243, 2000.
- 900 Khairoutdinov, M. and Randall, D. A.: Cloud Resolving Modeling of the ARM Summer 1997 IOP: Model Formulation, Results, Uncertainties, and Sensitivities, *J. Atmos. Sci.*, 60, 607–624, 2003.
- Kogan, Y. L. and Mechem, D. B.: A PDF-Based Microphysics Parameterization for Shallow Cumulus Clouds, *J. Atmos. Sci.*, 71, 1070–1089, doi:10.1175/JAS-D-13-0193.1, 2014.
- Kogan, Y. L. and Mechem, D. B.: A PDF-based formulation of microphysical variability in cumulus congestus  
905 clouds, *J. Atmos. Sci.*, doi:10.1175/JAS-D-15-0129.1, 2015.
- Lappen, C.-L. and Randall, D. A.: Towards a unified parameterization of the boundary layer and moist convection. Part I: A new type of mass-flux model, *J. Atmos. Sci.*, 58, 2021–2036, 2001.
- Larson, V. E. and Golaz, J.-C.: Using probability density functions to derive consistent closure relationships among higher-order moments, *Mon. Wea. Rev.*, 133, 1023–1042, 2005.
- 910 Larson, V. E. and Griffin, B. M.: Coupling microphysics parameterizations to cloud parameterizations, in: Preprints, 12th Conference on Cloud Physics, Madison, WI, American Meteorological Society, 2006.
- Larson, V. E. and Griffin, B. M.: Analytic upscaling of local microphysics parameterizations, Part I: Derivation, *Quart. J. Royal Met. Soc.*, 139, 46–57, 2013.
- Larson, V. E. and Schanen, D. P.: The Subgrid Importance Latin Hypercube Sampler (SILHS): a multivariate  
915 subcolumn generator, *Geosci. Model Dev.*, 6, 1813–1829, doi:10.5194/gmdd-6-1813-2013, 2013.
- Larson, V. E., Golaz, J.-C., and Cotton, W. R.: Small-scale and mesoscale variability in cloudy boundary layers: Joint probability density functions, *J. Atmos. Sci.*, 59, 3519–3539, 2002.
- Larson, V. E., Golaz, J.-C., Jiang, H., and Cotton, W. R.: Supplying Local Microphysics Parameterizations with Information about Subgrid Variability: Latin Hypercube Sampling, *J. Atmos. Sci.*, 62, 4010–4026, 2005.
- 920 Lebsock, M., Morrison, H., and Gettelman, A.: Microphysical implications of cloud-precipitation covariance derived from satellite remote sensing, *Journal of Geophysical Research: Atmospheres*, 118, 6521–6533, 2013.
- Lewellen, W. S. and Yoh, S.: Binormal model of ensemble partial cloudiness, *J. Atmos. Sci.*, 50, 1228–1237, 1993.
- 925 Mellor, G. L.: The Gaussian cloud model relations, *J. Atmos. Sci.*, 34, 356–358, 1977.
- Morrison, H. and Gettelman, A.: A new two-moment bulk stratiform cloud microphysics scheme in the Community Atmosphere Model, Version 3 (CAM3). Part I: Description and numerical tests, *J. Climate.*, 21, 3642–3659, 2008.
- Morrison, H., Curry, J. A., and Khvorostyanov, V. I.: A new double-moment microphysics parameterization for  
930 application in cloud and climate models. Part I: Description, *J. Atmos. Sci.*, 62, 1665–1677, 2005.
- Naumann, A. K., Seifert, A., and Mellado, J. P.: A refined statistical cloud closure using double-Gaussian probability density functions, *Geosci. Model Dev.*, 6, 1641–1657, doi:10.5194/gmd-6-1641-2013, 2013.
- Neggers, R. A. J., Köhler, M., and Beljaars, A. C. M.: A Dual Mass Flux Framework for Boundary Layer Convection. Part I: Transport, *J. Atmos. Sci.*, 66, 1465–1488, 2009.
- 935 Raut, E. K. and Larson, V. E.: A Flexible Importance Sampling Method for Integrating Subgrid Processes, *Geosci. Model Dev.*, 9, 413–429, doi:10.5194/gmd-9-413-2016, 2016.

- Smolarkiewicz, P. K. and Grabowski, W. W.: The multidimensional positive definite advection transport algorithm: nonoscillatory option, *J. Comput. Phys.*, 86, 355–375, doi:10.1016/0021-9991(90)90105-A, 1990.
- Sommeria, G. and Deardorff, J. W.: Subgrid-scale condensation in models of nonprecipitating clouds, *J. Atmos. Sci.*, 34, 344–355, 1977.
- Stephens, M. A.: Use of the Kolmogorov-Smirnov, Cramer-Von Mises and Related Statistics Without Extensive Tables, *J. Roy. Statist. Soc. Ser. B*, 32, 115–122, 1970.
- Storer, R. L., Griffin, B. M., Höft, J., Weber, J. K., Raut, E., Larson, V. E., Wang, M., and Rasch, P. J.: Parameterizing deep convection using the assumed probability density function method, *Geosci. Model Dev.*, 8, 1–19, doi:10.5194/gmd-8-1-2015, 2015.
- Sušelj, K., Teixeira, J., and Chung, D.: A unified model for moist convective boundary layers based on a stochastic eddy-diffusivity/mass-flux parameterization, *J. Atmos. Sci.*, 70, 1929–1953, 2013.
- Thayer-Calder, K., Gettelman, A., Craig, C., Goldhaber, S., Bogenschutz, P. A., Chen, C.-C., Morrison, H., Höft, J., Raut, E., Griffin, B. M., Weber, J. K., Larson, V. E., Wyant, M. C., Wang, M., Guo, Z., and Ghan, S. J.: A unified parameterization of clouds and turbulence using CLUBB and subcolumns in the Community Atmosphere Model, *Geosci. Model Dev.*, 8, 3801–3821, doi:10.5194/gmd-8-3801-2015, <http://www.geosci-model-dev.net/8/3801/2015/>, 2015.
- Tompkins, A. M.: A prognostic parameterization for the subgrid-scale variability of water vapor and clouds in large-scale models and its use to diagnose cloud cover, *J. Atmos. Sci.*, 59, 1917–1942, 2002.
- van Zanten, M., Stevens, B., Nuijens, L., Siebesma, A., Ackerman, A., Burnet, F., Cheng, A., Couvreux, F., Jiang, H., Khairoutdinov, M., Kogan, Y., Lewellen, D., Mechem, D., Nakamura, K., Noda, A., Shipway, B., Slawinska, J., Wang, S., and Wyszogrodzki, A.: 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, 2011.
- Zhang, J., Lohmann, U., and Lin, B.: A new statistically based autoconversion rate parameterization for use in large-scale models, *J. Geophys. Res.*, 107, Article No. 4750, doi:10.1029/2001JD001484, 2002.

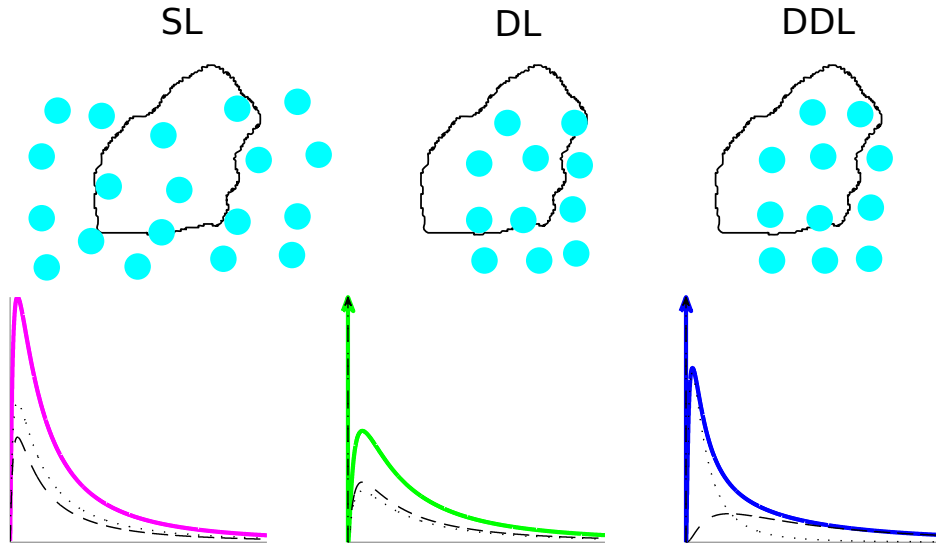
**Table 1.** Kolmogorov-Smirnov statistic averaged over multiple grid levels and statistical output timesteps comparing each of DDL, DL, and SL hydrometeor PDF shapes to SAM LES results. The best (lowest) average score for each case and hydrometeor species is listed in bold. The DDL has the lowest average score most often, and the DL has the second-lowest average score most often.

Average Kolmogorov-Smirnov Statistic			
Case-Species	$\langle \text{KS} \rangle$ DDL	$\langle \text{KS} \rangle$ DL	$\langle \text{KS} \rangle$ SL
RICO $r_r$	<b>0.223</b>	0.373	0.496
RICO $N_r$	<b>0.182</b>	0.263	0.634
RF02 $r_r$	<b>0.131</b>	0.133	0.148
RF02 $N_r$	0.152	<b>0.150</b>	0.170
LBA $r_r$	<b>0.152</b>	0.240	0.201
LBA $N_r$	<b>0.142</b>	0.187	0.295
LBA $r_g$	<b>0.197</b>	0.307	0.429
LBA $N_g$	<b>0.165</b>	0.222	0.566
LBA $r_s$	<b>0.177</b>	0.267	0.432
LBA $N_s$	<b>0.173</b>	0.238	0.492
LBA $r_i$	<b>0.212</b>	0.282	0.614
LBA $N_i$	<b>0.122</b>	0.210	0.647

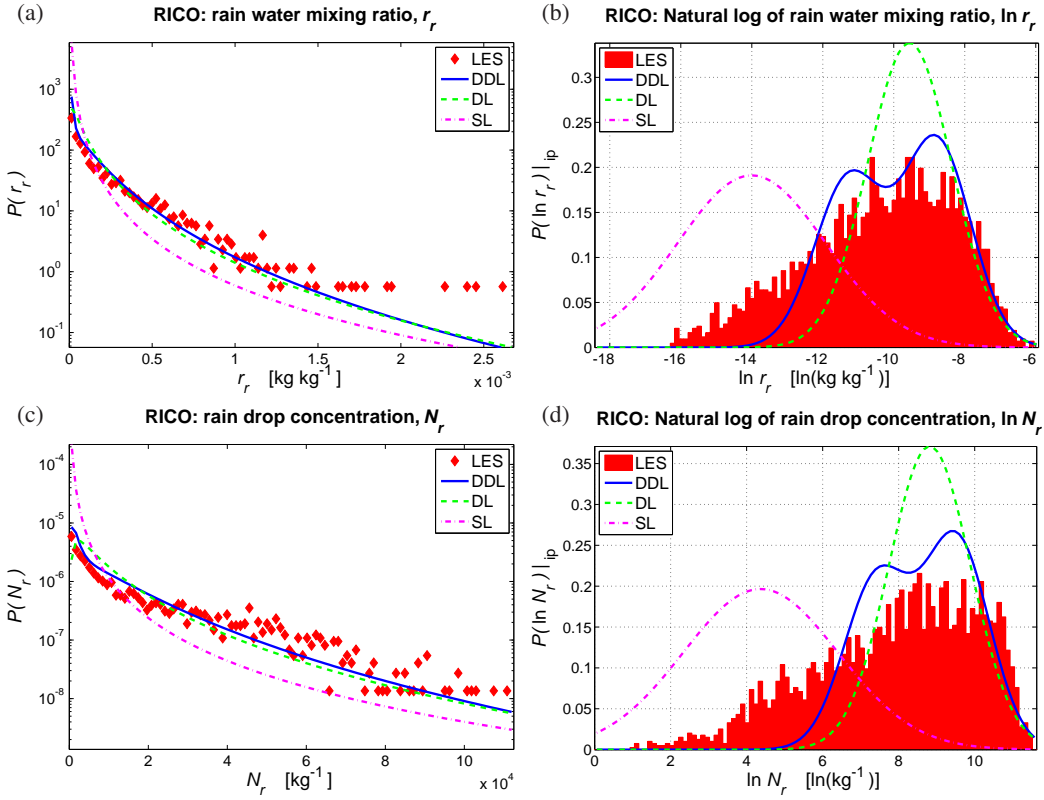
**Table 2.** Normalized Cramer-von Mises statistic averaged over multiple grid levels and statistical output timesteps comparing each of DDL, DL, and SL hydrometeor PDF shapes to SAM LES results. The best (lowest) average score for each case and hydrometeor species is listed in bold. The DDL has the lowest average score every time, and the DL has the second-lowest average score every time.

Average Normalized Cramer-von Mises Statistic			
Case-Species	$\langle \omega^2 \rangle$ DDL	$\langle \omega^2 \rangle$ DL	$\langle \omega^2 \rangle$ SL
RICO $r_r$	<b>0.0187</b>	0.0508	0.1255
RICO $N_r$	<b>0.0100</b>	0.0238	0.1872
RF02 $r_r$	<b>0.0041</b>	0.0049	0.0094
RF02 $N_r$	<b>0.0064</b>	0.0070	0.0136
LBA $r_r$	<b>0.0078</b>	0.0231	0.0282
LBA $N_r$	<b>0.0081</b>	0.0145	0.0537
LBA $r_g$	<b>0.0159</b>	0.0351	0.1092
LBA $N_g$	<b>0.0129</b>	0.0194	0.1576
LBA $r_s$	<b>0.0107</b>	0.0240	0.1072
LBA $N_s$	<b>0.0089</b>	0.0174	0.1261
LBA $r_i$	<b>0.0126</b>	0.0246	0.1968
LBA $N_i$	<b>0.0046</b>	0.0134	0.2046

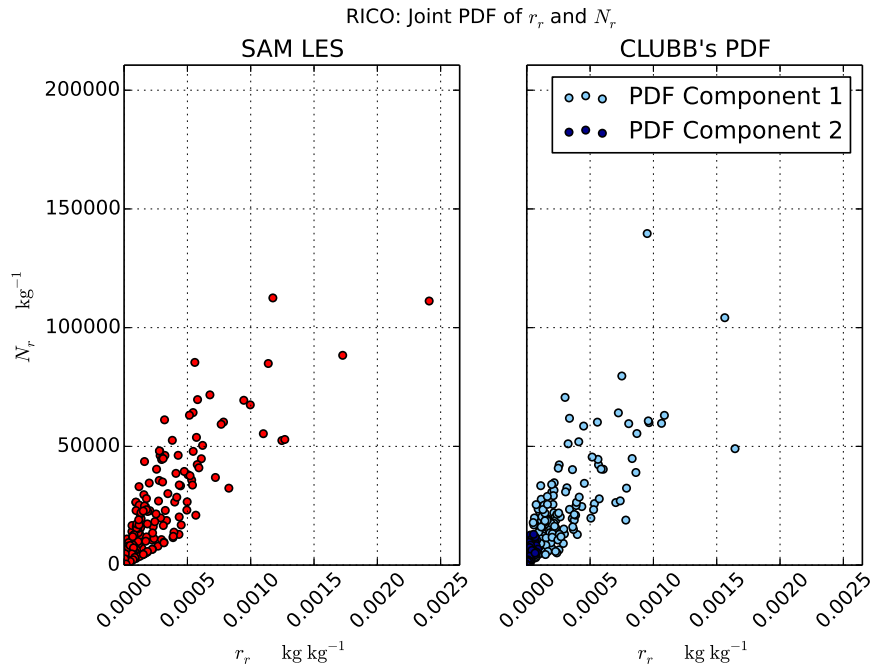




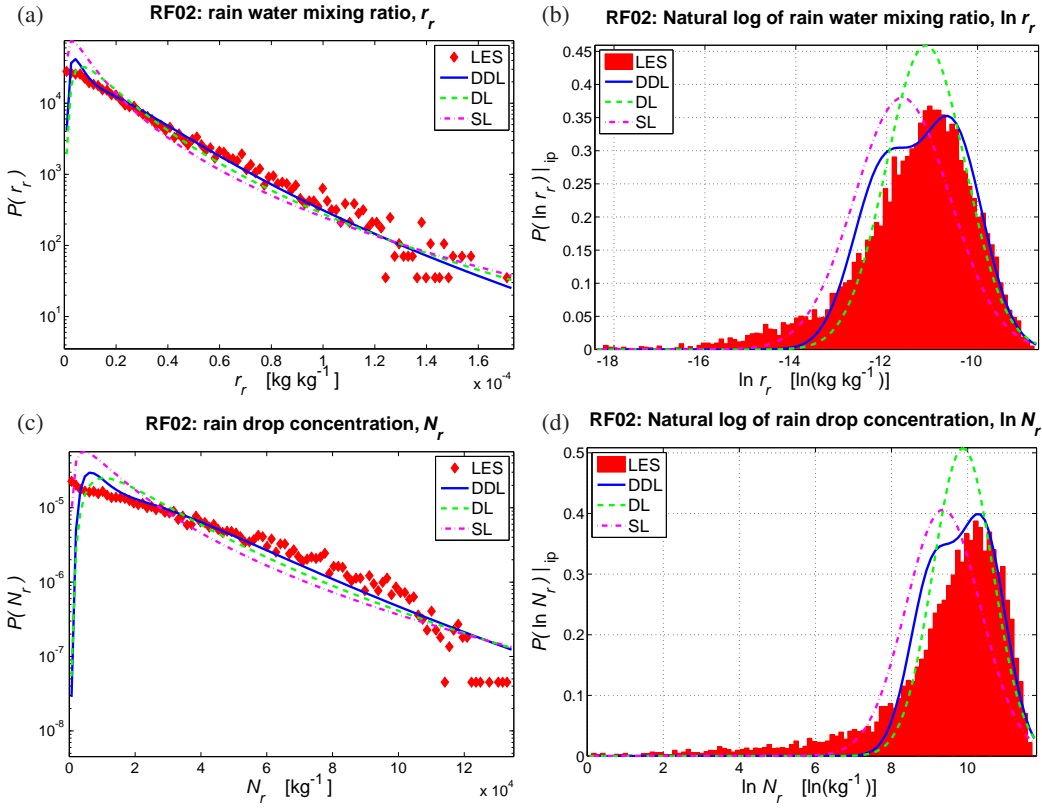
**Figure 1.** A schematic of the single lognormal (SL), delta-lognormal (DL), and delta-double-lognormal (DDL) hydrometeor PDF shapes. The SL PDF shape is precipitating over the entire subgrid domain, whereas the DL and DDL shapes are not. In all three plots of the PDFs (where each PDF is a function of a hydrometeor species, such as  $r_r$ ), the weighted PDF from each PDF component is shown (black dashes and black dots). The sum of the two are the SL (solid magenta), the DL (solid green), and the DDL (solid blue). The SL does not contain a delta at 0, and the mean and variance of each PDF component are the same. Each component of the DL has a delta at 0 (upward pointing black arrows on the y-axis). The sum of the two component deltas forms the DL's delta at 0 (upward pointing green arrow). The mean and variance of each DL PDF component are the same within precipitation. Each component of the DDL also has a delta at 0 (upward pointing black arrows). The sum of the two component deltas forms the DDL's delta at 0 (upward pointing blue arrow). The mean and/or variance differ between DDL PDF components within precipitation.



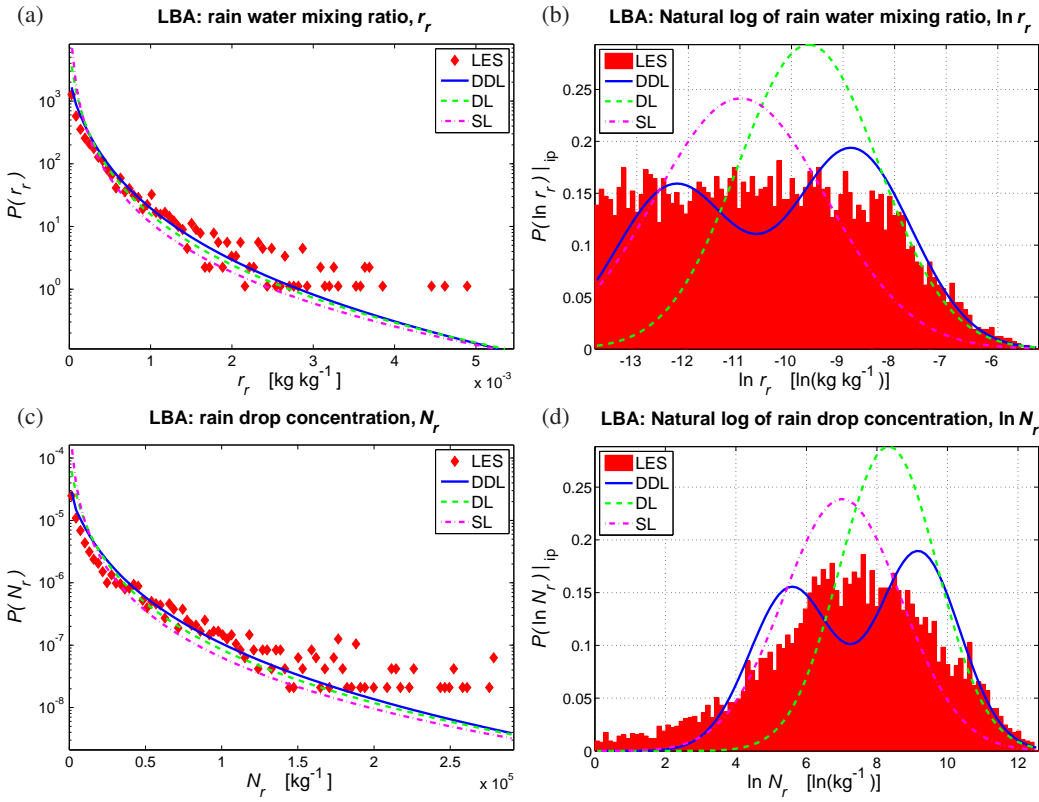
**Figure 2.** PDFs of rain in the RICO precipitating shallow cumulus case at an altitude of 380 m and a time of 4200 min. The SAM LES results are in red, the DDL results are blue solid lines, the DL results are green dashed lines, and the SL results are magenta dashed-dotted lines. (a) The marginal distribution of  $r_r$  with the delta at  $r_r = 0$  omitted. (b) The marginal distribution of  $\ln r_r$  using the “in-precipitation PDF.” This is the in-precipitation marginal PDF in Gaussian space. (c) The marginal distribution of  $N_r$  with the delta at  $N_r = 0$  omitted. (d) The marginal distribution of  $\ln N_r$  using the “in-precipitation PDF.” Again, this is the in-precipitation marginal PDF in Gaussian space. The DDL provides a better fit to SAM LES than the DL, which in turn provides a better fit than the SL.



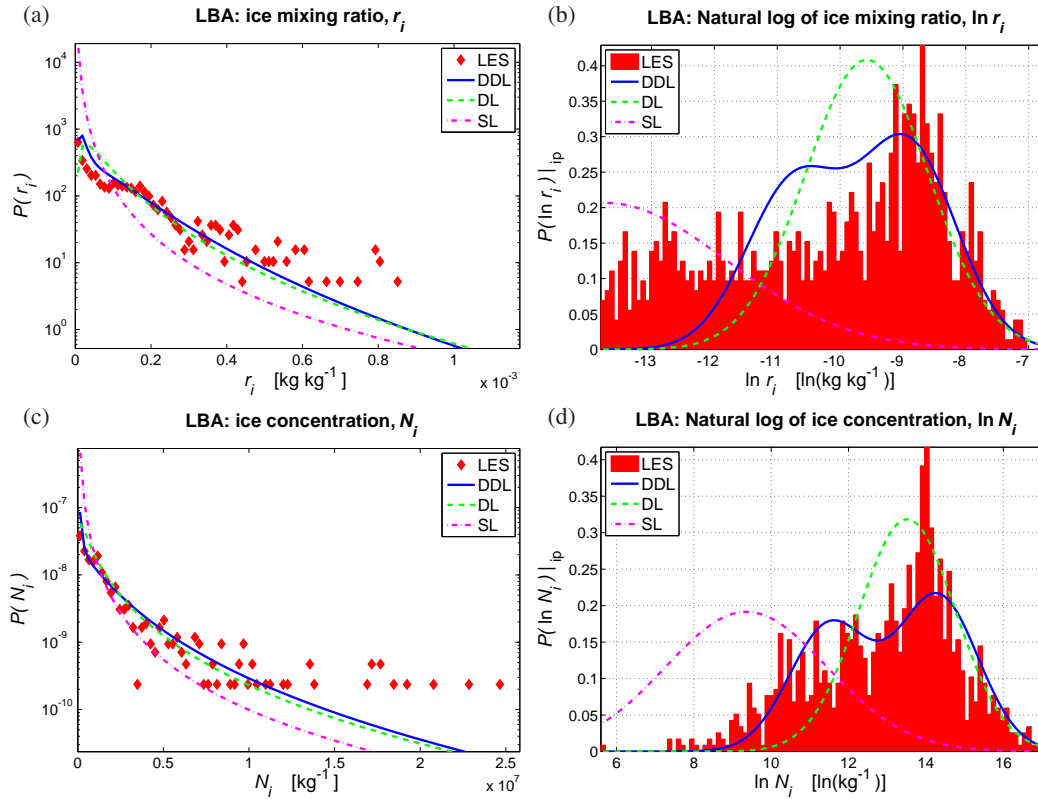
**Figure 3.** Joint PDF of  $r_r$  and  $N_r$  in the RICO precipitating shallow cumulus case at an altitude of 380 m and a time of 4200 min. SAM LES results are the red scatterpoints. CLUBB PDF scatterpoints were generated by sampling the DDL PDF using an unweighted Monte Carlo scheme. The SAM LES domain is  $256 \times 256$  grid points, so to provide for the best comparison of LES points to CLUBB PDF sample points, 65536 CLUBB PDF sample points were used. The light blue scatterpoints are from PDF component 1 and the dark blue scatterpoints are from PDF component 2. Every 10th point was plotted from both SAM LES and CLUBB's PDF. The joint nature of the PDF allows  $r_r$  and  $N_r$  to correlate the same way in CLUBB as they do in SAM.



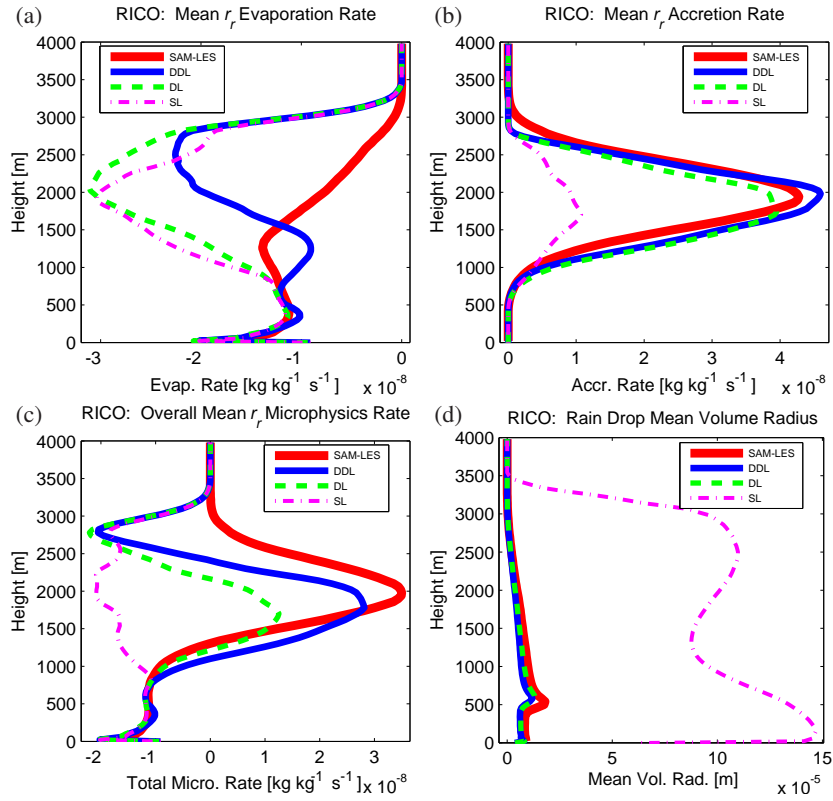
**Figure 4.** PDFs of rain in the DYCOMS-II RF02 drizzling stratocumulus case at an altitude of 400 m and a time of 330 min. The SAM LES results are in red, the DDL results are blue solid lines, the DL results are green dashed lines, and the SL results are magenta dashed-dotted lines. (a) The marginal distribution of  $r_r$  with the delta at  $r_r = 0$  omitted. (b) The marginal distribution of  $\ln r_r$  using the “in-precipitation PDF.” This is the in-precipitation marginal PDF in Gaussian space. (c) The marginal distribution of  $N_r$  with the delta at  $N_r = 0$  omitted. (d) The marginal distribution of  $\ln N_r$  using the “in-precipitation PDF”, which is the in-precipitation marginal PDF in Gaussian space. Owing to relatively low within-precipitating variance, the three hydrometeor PDF shapes are all a close match to SAM LES.



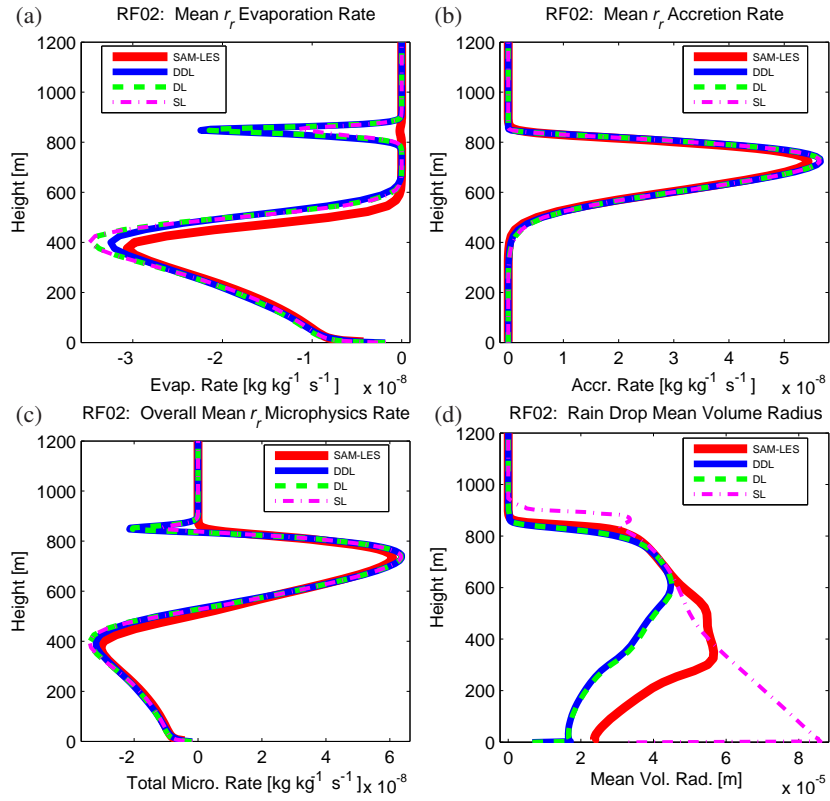
**Figure 5.** PDFs of rain in the LBA deep convective case at an altitude of 2424 m and a time of 330 min. The SAM LES results are in red, the DDL results are blue solid lines, the DL results are green dashed lines, and the SL results are magenta dashed-dotted lines. (a) The marginal distribution of  $r_r$  with the delta at  $r_r = 0$  omitted. (b) The marginal distribution of  $\ln r_r$  using the “in-precipitation PDF.” This is the in-precipitation marginal PDF in Gaussian space. (c) The marginal distribution of  $N_r$  with the delta at  $N_r = 0$  omitted. (d) The marginal distribution of  $\ln N_r$  using the “in-precipitation PDF”, which is the in-precipitation marginal PDF in Gaussian space. Again, the DDL provides the best fit to SAM LES.



**Figure 6.** PDFs of ice in the LBA deep convective case at an altitude of 10500 m and a time of 360 min. The SAM LES results are in red, the DDL results are blue solid lines, the DL results are green dashed lines, and the SL results are magenta dashed-dotted lines. (a) The marginal distribution of  $r_i$  with the delta at  $r_i = 0$  omitted. (b) The marginal distribution of  $\ln r_i$  using the “in-precipitation PDF.” This is the in-precipitation marginal PDF in Gaussian space. (c) The marginal distribution of  $N_i$  with the delta at  $N_i = 0$  omitted. (d) The marginal distribution of  $\ln N_i$  using the “in-precipitation PDF.” Again, this is the in-precipitation marginal PDF in Gaussian space. The method works for frozen hydrometeor species as well, as the DDL provides a better fit to SAM LES than the DL, which in turn provides a better fit than the SL.

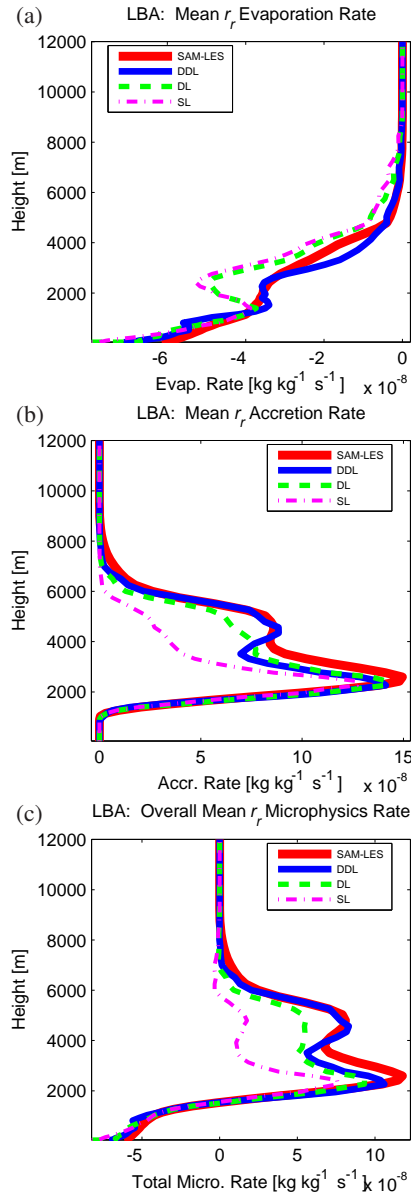


**Figure 7.** Profiles of mean microphysics process rates in the RICO precipitating shallow cumulus case time-averaged over the last two hours of the simulation (minutes 4200 through 4320). The SAM LES results are red solid lines, the DDL results are blue solid lines, the DL results are green dashed lines, and the SL results are magenta dashed-dotted lines. (a) The mean evaporation rate of  $r_r$ . (b) The mean accretion rate of  $r_r$ . (c) The overall mean microphysics tendency for  $r_r$ . (d) The mean volume radius of rain drops. Overall, the DDL provides a better fit to SAM LES than the DL, which in turn provides a better fit than the SL.



**Figure 8.** Profiles of mean microphysics process rates in the DYCOMS-II RF02 drizzling stratocumulus case time-averaged over the last hour of the simulation (minutes 300 through 360). The SAM LES results are red solid lines, the DDL results are blue solid lines, the DL results are green dashed lines, and the SL results are magenta dashed-dotted lines. (a) The mean evaporation rate of  $r_r$ . (b) The mean accretion rate of  $r_r$ . (c) The overall mean microphysics tendency for  $r_r$ . (d) The mean volume radius of rain drops. All hydrometeor PDF shapes provide a good fit to SAM LES.





**Figure 9.** Profiles of mean warm microphysics process rates in the LBA deep convective case time-averaged over the last hour of the simulation (minutes 300 through 360). The SAM LES results are red solid lines, the DDL results are blue solid lines, the DL results are green dashed lines, and the SL results are magenta dashed-dotted lines. (a) The mean evaporation rate of  $r_r$ . (b) The mean accretion rate of  $r_r$ . (c) The overall mean microphysics tendency for  $r_r$ . Again, the DDL provides a better fit to SAM LES than the DL, which in turn provides a better fit than the SL.