Comment by topical editor

It would be nice if you would make one addition. I did not find what you mean with "mixture components" (Eq. 3). Please provide a definition, somewhere close to Eq. 3. Thank you.

Authors' response

Thank you for pointing out the lack of explaination of mixture components. The revised manuscript states:

"The PDF is split into N_{comp} components; ... Each component m has a weight $\xi_{(m)}$, where $\sum_{m=1}^{N_{\text{comp}}} \xi_{(m)} = 1$, and for each m, $\xi_{(m)} \geq 0$. Each component has different means and variances for the variates, and the PDF is a mixture (i.e., weighted sum) of the components."

A flexible importance sampling method for integrating subgrid processes

E. K. Raut and V. E. Larson

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

Correspondence to: E. K. Raut (raut@uwm.edu)

Abstract

Numerical models of weather and climate need to compute grid-box-averaged rates of physical processes such as microphysics. These averages are computed by integrating subgrid variability over a grid box. For this reason, an important aspect of atmospheric modeling is
5 spatial integration over subgrid scales.

The needed integrals can be estimated by Monte Carlo integration. Monte Carlo integration is simple and general but requires many evaluations of the physical process rate. To reduce the number of function evaluations, this paper describes a new, flexible method of importance sampling. It divides the domain of integration into eight categories, such as the
10 portion that contains both precipitation and cloud, or the portion that contains precipitation but no cloud. It then allows the modeler to prescribe the density of sample points within each of the eight categories.

The new method is incorporated into the Subgrid Importance Latin Hypercube Sampler (SILHS). The resulting method is tested on drizzling cumulus and stratocumulus cases. In
15 the cumulus case, the sampling error can be considerably reduced by drawing more sample points from the region of rain evaporation.

1 Introduction

Coarse-resolution atmospheric models of weather and climate do not solve differential equations; they solve *integro*-differential equations, that is, equations containing both
20 derivatives and integrals. Although a derivation of an atmospheric model starts with differential equations, such as the Navier–Stokes or advection-diffusion equations, those equations are coarse-grained or filtered before being discretized (e.g., Leonard, 1974; Pope, 2000). Typically, a spatial running-mean filter is used, producing equations similar to Reynolds-averaged equations (e.g., Germano, 1992). Each term in the filtered equations is spatially
25 averaged over a grid box. For instance, in a prognostic equation for grid-averaged rain mixing ratio, the grid-averaged rain is updated by *grid-averaged* microphysical process rates.

2

Schematically, we may give an example of such a filtered equation:

$$\frac{\partial \bar{r}}{\partial t} = \bar{h} + \ldots, \tag{1}$$

where $\bar{r}$ denotes grid-averaged rain mixing ratio, t denotes time, and $\bar{h}$ denotes the grid-averaged microphysical time tendency of rain mass mixing ratio. Because a grid-box av-
5   erage is an integral (divided by the grid-box volume), the resulting filtered equations are integro-differential equations. Therefore, a central problem in atmospheric modeling is (subgrid-scale, spatial) integration.

Mathematically, the problem is to evaluate integrals of the form

$$\bar{h} \equiv \int h(\boldsymbol{x}) P(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \tag{2}$$

10   where $\boldsymbol{x}$ is a vector containing the relevant model fields, $h(\boldsymbol{x})$ is the time tendency of a phys-
ical process, such as autoconversion of cloud droplets to form rain drops, and $P(\boldsymbol{x})$ is the model-predicted subgrid-scale probability density function (PDF) (i.e. "filtered density func-
tion", Colucci et al., 1998) of the variables. Here, $h(\boldsymbol{x})$ could be a simple analytic function or a complex numerical subroutine. An integral such as Eq. (2) ought to be computed for each
15   of the many nonlinear process rates in the model. (However, when the grid box is assumed to be uniform, then the integral is not performed.) The integrals also need to be computed for each grid column in the horizontal and each grid level in the vertical.

To carry out this integration (i.e. "quadrature"), researchers have proposed several meth-
ods. First, the integral (Eq. 2) may be evaluated analytically (e.g., Zhang et al., 2002; Larson
20   and Griffin, 2006; Morrison and Gettelman, 2008; Cheng and Xu, 2009; Griffin and Larson, 2013; Larson and Griffin, 2013; Lebsock et al., 2013; Boutle et al., 2014). Analytic inte-
gration has the advantage of accuracy, but it can be carried out only if both the process rate $h(\boldsymbol{x})$ and the subgrid-scale PDF $P(\boldsymbol{x})$ are sufficiently simple. Furthermore, analytic integration is carried out grid level by grid level, and does not compute the vertical overlap
25   of cloud properties. Vertical overlap is related to the correlation between quantities at two

3

points in space, one located directly above the other. The degree of vertical overlap has a strong influence on, e.g., radiative transfer. Second, the integrals may be computed by deterministic quadrature (Xiu, 2009; Golaz et al., 2011; Chowdhary et al., 2015). Deterministic quadrature solves an integral by computing a weighted sum of integrand values evaluated at specially chosen quadrature points. Deterministic quadrature has a couple of advantages: unlike analytic integration, deterministic quadrature is applicable to a broad range of processes, and like analytic integration, deterministic quadrature is still accurate. Deterministic quadrature also has a disadvantage: it does not compute vertical overlap. Third, the integrals can be evaluated by Monte Carlo integration (e.g., Gentle, 2003; Kalos and Whitlock, 2008). In Monte Carlo integration, random samples are drawn from the subgrid PDF $P(x)$, the integrand is evaluated at each sample point, and the resulting values are suitably averaged. Monte Carlo integration is broadly applicable and can be configured to model vertical overlap (Barker et al., 2002, 2008; Pincus et al., 2003, 2006; Räisänen et al., 2004, 2005, 2007, 2008; Räisänen and Barker, 2004; Larson et al., 2005; Larson, 2007; Hill et al., 2011; Larson and Schanen, 2013; Tonttila et al., 2013, 2015). However, Monte Carlo integration converges slowly. Obtaining an accurate integration requires many costly evaluations of a microphysics parameterization.

To improve the convergence of Monte Carlo integration, many methods have been proposed. Two broad strategies are stratified sampling and importance sampling (Press et al., 2007; Lemieux, 2009). Stratified sampling spreads out the sample points in sample space in order to avoid clumping, which leads to poor sampling. One popular stratified sampling method is Latin hypercube sampling, which stratifies along each dimension of the integral (e.g., McKay et al., 1979; Owen, 2003). Another strategy, importance sampling, preferentially places sample points in important regions of the integration domain (Press et al., 2007; Lemieux, 2009). For instance, extra sample points may be placed within cloud because that is where important processes occur, such as the formation and growth of cloud droplets.

Some sampling methods combine stratified and importance sampling. For instance, a prior version of the Subgrid Importance Latin Hypercube Sampler (SILHS) placed sample points preferentially in cloud, and also stratified the within-cloud sample points using

4

Latin hypercube sampling (Larson et al., 2005; Larson and Schanen, 2013; Storer et al., 2015; Thayer-Calder et al., 2015). SILHS similarly stratified the points out of cloud. Although SILHS' importance sampling improved the integration of within-cloud microphysical processes, the *importance* sampling did not improve the integration of out-of-cloud pro-
5   cesses, such as evaporation of rain. This is a drawback in cases where evaporation is an important process. What is needed is a more flexible importance sampling method, one that allows the modeler to sample important processes in a more targeted way.

This paper proposes a new importance sampling method that is highly flexible. It divides the domain of integration into N_{cat} non-overlapping "categories", such as the region that
10 contains precipitation and cloud, or the region that contains precipitation but no cloud. This "nCat" method allows the modeler to prescribe the density of sample points within each category. This flexibility allows a modeler to allocate more sample points to a particular process, such as evaporation of rain drops, if evaporation is especially important to the problem of interest. Furthermore, two or more categories can be combined into a single
15 "cluster" if none of the categories in the cluster should be treated preferentially over the others.

This paper will introduce nCat sampling and evaluate it in an idealized, single-column setting. Section 2 specifies the subgrid probability density function (PDF) that our method will sample. Section 3 describes how SILHS sampled the subgrid PDF before the nCat
20 method was introduced. Section 4 details the new nCat method that has been introduced into SILHS. Section 5 explains the criteria and methodology used to evaluate the new SILHS sampling scheme, including configuration of the model. Section 6 shows tests using a precipitating shallow cumulus case and a precipitating stratocumulus case. Section 7 concludes the paper.

25 **2   The functional form of the PDF from which SILHS draws sample points**

SILHS does not *generate* sample points according to a stochastic rule; rather, SILHS merely *draws* sample points from a pre-existing subgrid PDF. In this paper, the PDF is calculated

by the Cloud Layers Unified By Binormals (CLUBB) parameterization (Larson et al., 2002; Golaz et al., 2002; Larson and Golaz, 2005; Larson et al., 2012). At each time step in a simulation, CLUBB diagnoses the subgrid PDF by means of the assumed PDF method. That is, CLUBB prognoses various subgrid higher-order moments, assumes a functional form for the PDF, and diagnoses a particular PDF for each time step and grid level that is consistent with both the moments and the functional form.

CLUBB's PDF is multivariate. It includes several variates (i.e., variables) that are useful inputs to thermodynamical and microphysical calculations. One of the PDF's variates is the extended cloud (liquid) water mass mixing ratio (χ) (Mellor, 1977; Larson et al., 2005). When $\chi > 0$, then χ equals the cloud water mass mixing ratio; when $\chi < 0$, then cloud water is assumed to be zero, and χ represents the deviation from saturation. The variate χ does not include ice. Another of the PDF's variates, related to χ, is the corresponding orthogonal variable (η) (Mellor, 1977; Larson et al., 2005). Together, χ and η are a rotation and rescaling of temperature and total water variables. CLUBB's PDF also includes the vertical velocity w, an extended cloud droplet number mixing ratio (N_{cn}), and the precipitating hydrometeor mass mixing ratios and number mixing ratios (**hm**). The cloud droplet number equals the extended cloud droplet number mixing ratio when cloud is present (that is, when $\chi > 0$); when no cloud is present, the cloud droplet number is assumed to be zero.

The functional form of CLUBB's PDF is a compromise between realism and mathematical simplicity. CLUBB's PDF may be written as

$$P(\boldsymbol{x}) = \sum_{m=1}^{N_{comp}} \xi_{(m)} \left[f_{p(m)} P_{(m)}(\chi, \eta, w, N_{cn}, \mathbf{hm}) + (1 - f_{p(m)}) \delta(\mathbf{hm}) P_{(m)}(\chi, \eta, w, N_{cn}) \right]. \quad (3)$$

The PDF ~~has~~ is split into N_{comp} ~~mixture~~ components; currently, in CLUBB, $N_{comp} = 2$. Each component m has a weight $\xi_{(m)}$, where $\sum_{m=1}^{N_{comp}} \xi_{(m)} = 1$, and for each m, $\xi_{(m)} \geq 0$. Each component has different means and variances for the variates, and the PDF is a mixture (i.e., weighted sum) of the components. The vector **hm** contains precipitating hydrometeor species that are prognosed by the microphysics scheme. The exact type and number of hydrometeors depends on the microphysics scheme used. In this paper, the micro-

6

physics scheme used is that of Khairoutdinov and Kogan (2000), in which the prognosed hydrometeor species are rain mass mixing ratio and rain number mixing ratio. The fraction $f_{\mathsf{p}(m)}$ represents the portion of mixture component m that contains at least one precipitating hydrometeor species, where $0 \leq f_{\mathsf{p}(m)} \leq 1$. The fraction $(1 - f_{\mathsf{p}(m)})$ represents portions of mixture component m that are precipitation-free (and are denoted by $\delta(\mathbf{hm})$) but may or may not contain cloud water.[1] In the portions of the PDF that contain precipitation, $P_{(m)}(\chi, \eta, w, N_{\mathsf{cn}}, \mathbf{hm})$ is a joint normal-lognormal distribution, where χ, η, and w are normally distributed, and N_{cn} and all the variables in $\mathbf{hm}$ are lognormally distributed (Larson and Griffin, 2013). (See also B. M. Griffin and V. E. Larson (2016), "A new subgrid-scale representation of hydrometeor fields using a multivariate PDF," in preparation.) In the parts of the PDF that do not contain precipitation, $P_{(m)}(\chi, \eta, w, N_{\mathsf{cn}})$ is a joint normal-lognormal distribution, as before, but all the precipitating hydrometeors are zero, rather than lognormally distributed. A simplifying feature of the functional form is that we insist that

$$P_{(m)}(\chi, \eta, w, N_{\mathsf{cn}}) = \int P_{(m)}(\chi, \eta, w, N_{\mathsf{cn}}, \mathbf{hm})\, \mathrm{d}\mathbf{hm}. \tag{4}$$

That is, the marginal distribution of cloud and turbulence within a mixture component, m, is the same both within and outside of the precipitating region. Therefore, integrating over precipitation in (3) collapses the two terms per component m into one term.

3 Prior formulation of SILHS

In both the new and prior versions of SILHS, sample points are drawn from CLUBB's PDF and fed into subroutines that compute microphysical process rates. To reduce the noise associated with the random sampling of processes, both versions of SILHS incorporate stratified sampling (specifically, Latin hypercube sampling) and importance sampling.

---

[1] In the notation used above, $\delta(\mathbf{hm})$ is a Dirac delta function and is short for $\delta(\mathsf{hm}_1)\delta(\mathsf{hm}_2)\cdots\delta(\mathsf{hm}_n)$.

The Latin hypercube algorithm is described in many sources (e.g., McKay et al., 1979; Owen, 2003; Larson et al., 2005). Intuitively, the algorithm stratifies the sample points across each variate such that for each variate, exactly one sample point falls into each of N_s intervals of equal size, where N_s is the number of sample points. For instance, if $N_s = 3$, the Latin hypercube sampling chooses low, medium, and high values of, e.g., rain mass mixing ratio.

Importance sampling is useful when a process rate is particularly large and variable within a small portion of the sample space. For instance, autoconversion of cloud droplets occurs only within cloud, which in cumulus cases often occupies a small fraction of the domain. Without importance sampling, the density of sample points in the sample space is given by the PDF $P(x)$. For example, if, according to the PDF, 10 % of the domain is occupied by cloud, then on average only 10 % of sample points will be placed within cloud. Importance sampling is used to change the sampling density so that areas of interest are sampled more frequently than less important regions, regardless of the densities given by the PDF.

The prior version of SILHS (Larson et al., 2005; Larson and Schanen, 2013) used a simple importance sampling scheme that placed half the sample points in cloud and the other half out of cloud. This importance sampling was only performed when, according to the PDF, the amount of the grid box occupied by cloud (the "cloud fraction") was between 0.5 % and 50 %. When cloud fraction was in this range, SILHS preferentially sampled cloud, thereby improving the representation of cloud processes such as autoconversion. In doing so, of course, SILHS' importance sampling degraded the representation of processes that occur out of cloud, such as evaporation of rain.

In both the new and prior versions of SILHS, a sample is first drawn from a starting grid level. This grid level is the only grid level where SILHS explicitly performs importance sampling; it is called the "importance sampling level" in this paper. The importance sampling level is chosen at each timestep to be the height level with the maximum within-cloud cloud water mass mixing ratio. To represent vertical overlap, sample points at other height levels are drawn such that they are correlated with adjacent levels according to a correlation coefficient that decreases exponentially with increasing height (see Larson and Schanen,

8

2013). This process continues to the top and bottom of the domain. The resulting vertical profile of sample points is a "subcolumn" that statistically represents a fraction of the grid column and models vertical overlap. Thus, SILHS does sample at all grid levels, but it *explicitly* performs *importance* sampling only at the importance sampling level.

4 The nCat importance sampling method

The nCat flexible importance sampling method is a generalization of the original SILHS importance sampling method described above. It is designed to give the modeler finer control over which parts of the subgrid PDF are preferentially sampled, that is, regarded as "important".

First, the domain of the PDF is split into a set of disjoint categories, C_j, that span the entire PDF domain. Here, $j = 1 \ldots N_{\text{cat}}$, where N_{cat} is the number of categories.

In this paper, eight categories are used. The definitions of the categories are based on the following three criteria: in/out of cloud, in mixture component 1/2, and in/out of precipitation. A sample point lies in cloud if and only if $\chi > 0$; if $\chi \leq 0$, the sample point lies outside cloud. To determine whether a sample point lies within mixture component 1, we generate a random number, u_{d+1}, that is uniformly distributed between $(0, 1)$. If $u_{d+1} < \xi_{(1)}$, then the sample is in mixture component 1; otherwise, the sample is in mixture component 2. To determine whether a sample point lies within precipitation, we generate another uniformly distributed random number, u_{d+2}, and check whether $u_{d+2} < f_{\mathsf{p}(m)}$, where m is the mixture component number. The eight possible combinations of cloud, mixture component, and precipitation form the eight categories used for importance sampling. The categories are shown in Table 1.

Each category C_j is associated with a certain amount of PDF mass, called the category's "original probability" and denoted as:

$$p_j = \int\limits_{C_j} P(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}. \tag{5}$$

9

Since the categories C_j span the entire PDF, we have:

$$\sum_{j=1}^{N_{\text{cat}}} p_j = 1, \tag{6}$$

where N_{cat} is the number of categories (currently $N_{\text{cat}} = 8$). Each category has $p_j \geq 0$; but naturally, categories with $p_j = 0$ need not be included in the corresponding integral (Eq. 2).

In general, the p_j values must be found by performing an integral over the PDF. For example, the amount of PDF mass in the first category (the category with cloud, precipitation, and in component 1) may be found by integrating the PDF in Eq. (3) over this portion of the PDF:

$$
\begin{aligned}
p_1 &= \int\limits_{C_1} P(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \\
&= \xi_{(1)} \int\limits_{\chi > 0} \int\limits_{\textbf{hm} > 0} \left[f_{\mathsf{p}(1)} P_{(1)}(\chi, \eta, w, N_{\mathsf{cn}}, \textbf{hm}) + (1 - f_{\mathsf{p}(1)}) \delta(\textbf{hm}) P_{(1)}(\chi, \eta, w, N_{\mathsf{cn}}) \right] \mathrm{d}\textbf{hm}\,\mathrm{d}\chi \\
&= \xi_{(1)} f_{\mathsf{p}(1)} \int\limits_{\chi > 0} P_{(1)}(\chi, \eta, w, N_{\mathsf{cn}}, \textbf{hm})\,\mathrm{d}\chi \\
&= \xi_{(1)} f_{\mathsf{p}(1)} f_{\mathsf{c}(1)}.
\end{aligned} \tag{7}
$$

In CLUBB's PDF (see Eq. 3), because cloud and precipitation are independent within a component (that is, the marginal distribution of χ, which determines cloud, is the same both in and out of precipitation), the integrals to find the category original probabilities p_j involve only simple quantities, such as $f_{\mathsf{c}(1)}$ (cloud fraction in mixture component 1), that are already computed elsewhere in CLUBB. In general, computing these constants may involve evaluating complicated numerical functions, such as the error function, which involves computational expense. However, the constants need to be computed only at the importance sampling level (not at each height level) and only once per timestep (not for each sample point), and so the additional expense is tolerable.

10

The notation introduced so far in this section relates to the PDF itself, rather than importance sampling per se. In order to implement importance sampling, we sample what we regard as the "important" categories preferentially. To do so, we introduce for each category, C_j, a user-defined probability, S_j, called the category's "modified probability". The modified

5 probability S_j of a given category is the desired probability that any sample will fall in that category. In other words, it is the expected fraction of sample points in the category when importance sampling is used. Therefore, intuitively, it is advantageous to set the modified probabilities such that the categories that are important for a process of interest are sampled more often than the unimportant categories. These modified probabilities must be set

10 such that

$$\sum_{j=1}^{N_{\text{cat}}} S_j = 1. \tag{8}$$

The sampling process is modified such that each category C_j is sampled with probability S_j rather than p_j. In order to give a mathematical form for the new PDF that points are drawn from, we introduce some notation. We define a new function, $L(\boldsymbol{x})$, called the "likelihood

15 ratio":

$$L(\boldsymbol{x}) \equiv \sum_{j=1}^{N_{\text{cat}}} \left(\frac{p_j}{S_j} \right) \cdot 1_j(\boldsymbol{x}) \equiv \sum_{j=1}^{N_{\text{cat}}} \omega_j 1_j(\boldsymbol{x}). \tag{9}$$

Here, $1_j(\boldsymbol{x})$ is the indicator function of category C_j, defined as

$$1_j(\boldsymbol{x}) = \begin{cases} 1 & \boldsymbol{x} \in C_j \\ 0 & \boldsymbol{x} \notin C_j \end{cases}, \tag{10}$$

and

20 $$\omega_j = \frac{p_j}{S_j} \tag{11}$$

11

is the weight of each sample point in category C_j. Then, the new sampling PDF, denoted $Q(\boldsymbol{x})$, is defined as

$$Q(\boldsymbol{x}) = \frac{P(\boldsymbol{x})}{L(\boldsymbol{x})}. \tag{12}$$

The new PDF, $Q(\boldsymbol{x})$, is normalized because $\sum_{j=1}^{N_{\text{cat}}} S_j = 1$. The integral in Eq. (2) is written as

$$\int h(\boldsymbol{x})P(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \int h(\boldsymbol{x})L(\boldsymbol{x})\left(\frac{P(\boldsymbol{x})}{L(\boldsymbol{x})}\right)\mathrm{d}\boldsymbol{x} = \int h(\boldsymbol{x})L(\boldsymbol{x})Q(\boldsymbol{x})\mathrm{d}\boldsymbol{x}. \tag{13}$$

Then, the new integral in Eq. (13) is approximated by drawing N_{s} sample points from the $Q(\boldsymbol{x})$ distribution and evaluating

$$\int h(\boldsymbol{x})L(\boldsymbol{x})Q(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \approx \frac{1}{N_{\text{s}}} \sum_{i=1}^{N_{\text{s}}} h(\boldsymbol{x}_i)L(\boldsymbol{x}_i), \tag{14}$$

where $\boldsymbol{x}_i$ is the i sample point drawn from the $Q(\boldsymbol{x})$ distribution. For a sample point $\boldsymbol{x}_i$ in category C_j, $L(\boldsymbol{x}_i) = \frac{p_j}{S_j} = \omega_j$.

To draw sample points from the $Q(\boldsymbol{x})$ distribution, a uniform variate, $0 < u_c < 1$, is picked for each sample point. The value of this uniform variate determines a sample point's category. For example, if $0 < u_c < S_1$, then the sample point will be associated with category C_1. If $S_1 \leq u_c < S_1 + S_2$, the sample will be associated with category C_2, and so on. Once the category has been determined, the sample point is drawn from the portion of the marginal distribution of $P(\boldsymbol{x})$ that is within the category. For example, a sample point that is to be in cloud is drawn from the distribution $P(\boldsymbol{x}|\chi > 0)$.

4.1 The weight limiter

Importance sampling allows the modeler to concentrate sample points in areas of the sample space that are considered important. But sample points given to important areas are

12

taken from unimportant areas. Therefore, if importance sampling is applied overzealously, the less important processes can become excessively noisy.

In SILHS, we wish to employ an importance sampling scheme that improves results for important processes (e.g., certain microphysical processes) while still producing reasonably accurate estimates of other "less important" or perhaps less variable processes. One reason that we wish to avoid overdoing the importance sampling is that a favorable sampling distribution at one grid level (altitude) may be unfavorable at another.

The change in accuracy for a given category due to importance sampling can be assessed by noting the weight of sample points in that category. The inverse weight of a sample point in category C_j is given by

$$\frac{1}{\omega_j} = \frac{S_j}{p_j}, \tag{15}$$

where p_j is the category's original probability and S_j is the category's modified probability due to importance sampling. The weight, ω_j, is closely related to $L(x)$ (see Eq. 9). The inverse weight, $1/\omega_j$, may be interpreted as the density of sample points per unit probability mass. When $1/\omega_j < 1$, the category is sampled less often with importance sampling, and when $1/\omega_j > 1$, the category is sampled more often with importance sampling. We are particularly concerned about large values of ω_j, because large weights are associated with undersampling and hence degradations in accuracy.

In order to mitigate the negative impact of importance sampling, we now propose a simple method to impose a maximum weight, $\omega_{\max}$, in each category. Intuitively, the algorithm works as follows. For each category C_j, we compute the minimum modified sampling probability:

$$S_{j,\min} = \frac{p_j}{\omega_{\max}}. \tag{16}$$

To ensure that the weight of a category, ω_j, does not exceed $\omega_{\max}$, the category must be sampled at least as often as $S_{j,\min}$; that is, $S_j \geq S_{j,\min}$. If any categories are undersampled,

then S_j must be increased in those categories, and probability mass must be taken from other categories (where $S_j > S_{j,\min}$) in order to ensure that the S_j probabilities sum to 1. The algorithm takes probability mass from another category in proportion to how much "extra" probability mass the category has.

Formally, the algorithm is constructed as follows. We compute the difference between the category's modified probability, S_j, and its minimum modified probability, $S_{j,\min}$:

$$S_{j,\text{diff}} = S_j - S_{j,\min}. \tag{17}$$

Let N be the set of all categories where $S_{j,\text{diff}} < 0$ (the categories where S_j needs to be increased) and M be the set of all categories where $S_{j,\text{diff}} \geq 0$ (the categories where S_j can potentially be reduced). If N is the empty set (that is, if no categories have $S_{\text{diff}} < 0$), then all categories already satisfy the weight limit, and nothing needs to be done. Otherwise, if S'_j is the new distribution of modified probabilities, then for all $C_j \in N$ we set

$$S'_j = S_{j,\min}, \tag{18}$$

and for all $C_j \in M$ we set

$$S'_j = S_j - \left[S_{j,\text{diff}} \cdot \left(\frac{\sum\limits_{C_i \in N} \left| S_{i,\text{diff}} \right|}{\sum\limits_{C_i \in M} S_{i,\text{diff}}} \right) \right]. \tag{19}$$

This method will take sampling probability away from the categories with extra probability proportionally to how much extra probability they have (i.e., how large $S_{j,\text{diff}}$ is). It can readily be shown that the fractions S'_j are bound to the range $[0, 1]$ and sum to 1.

In SILHS, we currently set $\omega_{\max} = 2$, which means that on average, a category will be sampled no less than half as often with importance sampling than without importance sampling. Consequently, the variance of the estimate of a quantity is increased (degraded) by no more than a factor of two due to importance sampling. (The standard deviation is increased by no more than a factor of $\sqrt{2} \approx 1.4$. For this estimate of variance, we assume the usual Monte Carlo convergence rate.)

14

4.2 Optimal allocation of sample points

The success of the nCat method depends on knowing how to allocate sample points to the categories. In some cases, it is easy to see how to allocate points. For example, if it is known that the process(es) of interest are active in only one of the N_{cat} categories, then we can simply put all sample points in that one category, and use the weight limiter to ensure that other categories are still adequately sampled. However, in the case that processes of interest are active in two or more categories, one needs to know how to distribute (i.e. "allocate") sample points among these categories.

For a given process rate, $h(\boldsymbol{x})$, and a given estimator of the variance, it is possible to derive the optimal allocation of sample points (Lemieux, 2009). The optimal allocation provides guidance on how to determine the S_j values. In Appendix A, it is shown that the optimal modified probabilities are:

$$S_j = \frac{p_j\sqrt{v_j}}{\sum_{i=1}^{N_{\text{cat}}} p_i\sqrt{v_i}}. \tag{20}$$

This expression shows that the optimal fraction of samples in category C_j, i.e. S_j, depends on both the original probability p_j of category C_j, and on the category-averaged standard deviation, $\sqrt{v_j}$.

4.3 A simple method of allocating sample points

One could prescribe the modified probabilities S_j directly. However, a key problem with directly prescribing the S_j is that prescribed values cannot scale with the original probabilities p_j. For instance, consider the case in which, for some category C_j, we have prescribed $S_j > 0$, but it turns out that for a particular cloud case, $p_j = 0$. Then some sample points will be placed in category C_j even though they contribute nothing to the overall sum. (These sample points have a weight of zero.) This is a needless computational expense. Instead, the sample points should be placed in other categories with non-zero p_j.

More generally, prescribing each S_j directly is akin to assuming that the contribution of each category to the total sum is constant regardless of what fraction of the PDF is occupied by each category. Instead, a more realistic assumption is that each category contributes to the total sum a constant amount *per unit original probability*, p_j. These prescribed amounts are then scaled by the original probabilities p_j to obtain the modified probabilities S_j.

Specifically, we prescribe the following normalized standard deviation of the process rate for each category C_j:

$$\gamma_j = \frac{\sqrt{v_j}}{\sum_{i=1}^{N_{\text{cat}}} \sqrt{v_i}}. \tag{21}$$

To make the γ_j easier to interpret and prescribe, we insist that $0 \leq \gamma_j \leq 1$; then the denominator is simply the sum of the numerator in all categories, so that each γ_j is a fraction with $\sum_{j=1}^{N_{\text{cat}}} \gamma_j = 1$. Specifically, γ_j is the fraction of $\sqrt{v_j}$ in category C_j. Prescribing γ_j is accurate and general when each γ_j varies little in space or time, or from case to case. Note that the numerator of Eq. (21) is the same as the numerator of Eq. (20), but without the p_j term.

Given the γ_j fractions, it is easy to determine the S_j values by dividing the numerator and denominator of Eq. (20) by $\sum_{i=1}^{N_{\text{cat}}} \sqrt{v_i}$. This yields

$$S_j = \frac{p_j \gamma_j}{\sum_{i=1}^{N_{\text{cat}}} p_i \gamma_i}. \tag{22}$$

It is clear from this equation that the S_j values are still in the range $[0, 1]$ and still sum to 1.

The prescription (Eq. 22) leads to more robust importance sampling than does prescribing S_j values as constants. With Eq. (22), the optimal γ_j in a particular category (say, in cloud and precipitation and mixture component 1) is relatively insensitive to the area occupied by that category (e.g., to the cloud fraction or precipitation fraction). The reason is that in Eq. (22), each S_j value weights γ_j by the original probability p_j of category C_j. This means that, for instance, when C_j occupies a small fraction of the domain, and γ_j is moderate, then the total fraction of sample points S_j in category C_j scales naturally to small values. Prescribing the γ_j values is akin to prescribing the inverse weights $1/\omega_j$, rather than

the sample fractions S_j. To see resemblance of γ_j and $1/\omega_j$, note from Eqs. (15) and (22) that

$$\frac{1}{\omega_j} = \frac{S_j}{p_j} = \frac{\gamma_j}{\sum_{i=1}^{N_{\text{cat}}} p_i \gamma_i}. \tag{23}$$

Both $1/\omega_j$ and γ_j are related to the *density* of sample points per unit probability mass, p_j.

4.4 Advantages of the new method

As compared to the previous version of SILHS, the chief advantage of the new nCat method is its flexibility. In particular, the user can individually prescribe the sampling density per unit probability (γ_j) in each of eight categories, C_j. This flexibility is useful when important processes, such as evaporation of rain, occur in particular categories, such as the region of rain but no cloud.

This flexibility is made possible in part by the fact that the nCat method imposes no restriction on the number of sample points used per timestep. The previous version of SILHS required an even number of sample points per timestep, because one point was placed in cloud and the other was placed outside cloud. Generalizing this method to eight categories would have required a multiple of at least eight sample points per timestep, and would not allow much flexibility in prescribing the relative importance of categories. Instead, the nCat method uses a probabilistic approach to picking a category for each sample point. This allows any number of sample points to be used at each timestep, including the use of fewer than N_{cat} samples, without causing a biased result.

4.5 Summary of steps to implement method

In summary, to implement the new importance sampling method, the following steps should be taken:

1. Pick a set of categories, C_j, that span the PDF domain. We have proposed eight categories for use, as given in Table 1.

17

2. Pick a set of sampling fractions, γ_j. A good set of values to use can be obtained by using a simulation to estimate the optimal values, given by the right-hand side of Eq. (21), as we do in Sect. 6.1 below.

3. Compute, from the fractions γ_j, the modified probabilities S_j using Eq. (22). Pick sample points from the $Q(x)$ distribution, defined in Eq. (12).

4. Compute the weight in each category, ω_j, using Eq. (11). Sample points are given a weight corresponding to the category the sample point is in. Limit the weights according to the algorithm in Sect. 4.1, if so desired.

5. Feed the (unweighted) sample points one by one into a physical parameterization (e.g., a microphysics scheme).

6. Compute a weighted average of the function of interest using Eq. (14).

5 Methodology of evaluation of the sampling methods

In order to evaluate how well the new importance sampling scheme simulates multiple cloud types, we have simulated two cloud cases. The first is a drizzling shallow cumulus case: Rain in shallow Cumulus over the Ocean (RICO), configured as in the intercomparison of vanZanten et al. (2011). RICO was a drizzling trade-wind cumulus case observed off the Caribbean islands of Antigua and Barbuda (Rauber et al., 2007). The second is a drizzling stratocumulus case: Research Flight 2 (RF02) of the DYnamics and Chemistry Of Marine Stratocumulus (DYCOMS-II), configured as in Wyant et al. (2007). DYCOMS-II RF02 was a nocturnal drizzling stratocumulus layer observed off the coast of California (Stevens et al., 2003). A key difference in the sampling of these two cases is that the stratocumulus case has a much larger cloud fraction (> 0.95) than the cumulus case (< 0.05). Therefore, without importance sampling, nearly all sample points fall in cloud in the stratocumulus case, while almost none fall in cloud in the cumulus case. Finding a single, effective sampling strategy for both the stratocumulus and cumulus cases is challenging.

18

The following four configurations of SILHS were used for comparison:

1. "LH-only": This configuration uses only Latin hypercube sampling. No importance sampling is performed. The nCat method is not used.

2. "2Cat-Cld": This configuration is functionally equivalent to the old version of SILHS that placed one point in cloud and one point out of cloud. This configuration uses two categories: in cloud, and out of cloud. The categories (c,p,1), (c,p,2), (c,np,1), and (c,np,2) are all lumped together into the "cloud" category, and the other four categories are analogously lumped into the "clear" category. That is, a point that is in cloud belongs to the cloud category regardless of whether it is in precipitation or which mixture component it is in, and similarly for points in clear air. When cloud fraction is between 0.5% and 50%, it places 50% of sample points in each of the two categories. (That is, $S_1 = S_2 = 0.5$.) Otherwise, no importance sampling is performed.

3. "2Cat-CldPcp": This configuration also uses two categories. The first consists of points that are either in cloud or in precipitation, and the second consists of the complement, namely, points that are neither in cloud nor in precipitation. That is, (nc,np,1) and (nc,np,2) are lumped into the no-cloud-or-precipitation category, and the others are lumped into the cloud-or-precipitation category. Since no microphysical processes act in the area of the domain outside of cloud and precipitation, the sample points are initially prescribed such that all points fall in the cloud-or-precipitation category (i.e. the first category). (That is, $\gamma_1 = S_1 = 1, \gamma_2 = S_2 = 0$). After the initial prescription, the weight limiter ensures that $S_2 = \frac{p_2}{\omega_{\max}} = \frac{p_2}{2}$.

4. "8Cat": This configuration uses all eight categories listed in Table 1. To determine the sampling fractions γ_j to use, a simulation was run in which SILHS was used to estimate the quantity in Eq. (21) at each timestep. One set of sampling fractions, γ_j, was used for both RICO and DYCOMS-II RF02. This is discussed in Sect. 6.1.

The simulations were non-interactive, so that errors in the SILHS simulations did not feed back into the simulated fields. This made it possible to evaluate multiple SILHS simulations

against a common analytic solution. Some notable aspects of the simulation configurations are shown in Table 2.

The microphysics scheme used in the simulations is that of Khairoutdinov and Kogan (2000). As a reference solution, an analytically upscaled version of the Khairoutdinov-Kogan microphysics scheme was used, as described in Larson and Griffin (2013). Comparison with the analytic solution indicates whether SILHS draws sample points from the correct PDF at each grid level. However, the comparison with the analytic solution does not test whether the PDFs at each level are overlapped accurately. Although the overlap assumptions do not affect these test cases, overlap does influence processes such as radiative transfer. Testing the PDF overlap assumptions is left for future research. Nevertheless, the ability to test convergence at each grid level is an advantage. For instance, early convergence tests revealed several bugs in SILHS. Many microphysics schemes in operational use do not permit analytic solution. For these microphysics schemes, a non-analytic integration method, such as SILHS, is necessary.

Each SILHS configuration was evaluated on its ability to estimate the following three microphysical processes:

1. Autoconversion: the conversion of cloud water to rain water. This process occurs within cloud, both inside and outside of precipitation (rain).

2. Accretion: the growth of rain droplets by collection of cloud water. This process occurs when both cloud and precipitation are present.

3. Evaporation: the conversion of rain water to water vapor. This process occurs in areas outside cloud but within precipitation.

6 Simulations of drizzling cumulus and stratocumulus clouds

In this section, we present results obtained using the new importance sampling method.

20

6.1 Estimation of optimal sampling fractions

Prescribing the γ_j is a useful general approach only if the γ_j vary relatively little from case to case. We test this by estimating the optimal sampling fractions, γ_j, for both the RICO and DYCOMS-II RF02 cases. The optimal γ_j values are calculated by estimating the right-hand side of Eq. (21) at each timestep at the importance sampling level. The process used, $h(x)$, is the sum of the autoconversion, accretion, and evaporation tendencies from the microphysics scheme. The importance sampling level is chosen at each timestep to be the level with the maximum within-cloud cloud water mass mixing ratio. The integral defining v_j, given in Eq. (A4), was estimated using 256 SILHS sample points. The γ_j values, averaged over all timesteps (864 total timesteps for RICO and 360 for DYCOMS-II RF02), are shown in Table 3.

We see that in both cases, the optimal γ_j values are largest in category 1 (in cloud, in precipitation, and in mixture component 1) and in category 3 (out of cloud, in precipitation, and in mixture component 1). As expected, the optimal sampling fractions for the last two categories are zero, since microphysical processes do not act in the region where neither cloud nor rain exists. The other categories show differences, which may or may not be important. To test this, the optimal fractions for DYCOMS-II RF02 shown in Table 3 are used for both RICO and DYCOMS-II RF02 simulations to be presented. Thereby, the RICO case is used to test the robustness of the DYCOMS-II RF02 sampling fractions.

6.2 Results for RICO case

Figure 1 shows a plot of the root-mean-square error (RMSE) of the SILHS RICO simulations as a function of the number of sample points used. The 8Cat method has the smallest RMSE of all three methods when estimating the sum of autoconversion, accretion, and evaporation.

The largest improvement of the 8Cat and 2Cat-CldPcp methods over the (old) 2Cat-Cld method is in sampling evaporation. In fact, even the LH-only method (no importance sampling at all) results in a better estimate of evaporation than the 2Cat-Cld method. The

reason that evaporation is so poorly sampled in the 2Cat-Cld method is that the 2Cat-Cld method performs importance sampling only within cloud. Indeed, for in-cloud processes, such as autoconversion and accretion, the 2Cat-Cld method equals or improves upon both the 2Cat-CldPcp method and the 8Cat method in the RICO cumulus case. However, the 2Cat-Cld method reduces the number of sample points outside of cloud, degrading the simulation of rain evaporation. In contrast, both the 2Cat-CldPcp and 8Cat methods preferentially sample within the region of the sample space containing evaporation (out of cloud but within precipitation), leading to large improvements.

Table 4 compares how each of the four sampling methods allocates sample points. The table shows the percentage of sample points allocated to each category, averaged over the simulation. Comparing the allocation between the methods can give insight into the strengths and weaknesses of each method. For example, evaporation is best sampled by the 8Cat and 2Cat-CldPcp methods because they are the only methods that place a sizable number of points in the two categories that are in precipitation and outside cloud. The 2Cat-Cld and 8Cat methods give the best estimate of accretion because they place the largest number of points in the categories that are within cloud and precipitation.

Figure 2 shows, for the RICO case, timeseries plots of the four tendencies at the importance sampling level. Again, the largest improvement can be seen in the sampling of evaporation. Looking at the 2Cat-Cld timeseries for evaporation, it can be seen that at many timesteps, no points are found in the evaporating region of the sample space (out of cloud and within precipitation), and the estimated evaporation tendency is zero. At the timesteps where one or more sample points are found in the evaporating region, the tendency estimate is very large because the evaporation rate within the evaporating region is much larger than the overall mean evaporation rate. In the 2Cat-CldPcp and 8Cat simulations, the evaporating region of the sample space is well sampled, and sample points in this region have small weights, leading to an estimate that is much more comparable to the overall mean.

To assess the performance of the sampling methods at levels away from the importance sampling level, profile plots (over height levels) were generated for simulations with 32 sample points. To reduce the role of a "lucky" random seed in the comparison and thereby

better distinguish the methods, an ensemble of 12 simulations was used. Figure 3 shows profile plots of the four tendencies over height levels. These plots are averaged over all 864 timesteps and over the 12 ensemble members, and serve to indicate that SILHS converges to the analytic solution at all height levels and not only the importance sampling level. Figure 4 shows the RMSE of the SILHS solutions at each height level compared to the analytic solution, for all timesteps and ensemble members. It can be seen that the 8Cat and 2Cat-CldPcp methods show improved results compared to the 2Cat-Cld method at height levels between 1000 m and 2500 m. These height levels are where the improvement in the evaporation term is strongest. At levels below 1000 m (which are far below the importance sampling level of about 2000 m), all methods start to show considerable noise. It is interesting that this noise remains even after time- and ensemble-averaging. This highlights the large degree of variability in cumulus clouds and the need for careful parameterization of this variability.

We note that, in this paper, only the profile plots display an ensemble average. The time series plots display a single simulation so that individual sample values can be seen. The plots displaying RMSE vs. the number of sample points are not strongly influenced by the choice of random seed.

6.3 Results for DYCOMS-II RF02 case

The other simulated case is DYCOMS-II RF02, a drizzling stratocumulus case. Figure 5 shows a plot of the RMS error of the SILHS simulations as a function of the number of sample points. The LH-only, 2Cat-Cld, and 2Cat-CldPcp simulations all show approximately the same amount of noise. However, the 8Cat method reduces noise in autoconversion and accretion, thereby also decreasing noise in the sum of the three tendencies.

The similarity between the LH-only, 2Cat-Cld, and 2Cat-CldPcp methods is expected. The 2Cat-Cld method, like the previous version of SILHS, includes a condition that reverts to straight Latin hypercube sampling in the event that cloud fraction exceeds 50 %. The DYCOMS-II RF02 case is a stratocumulus case, and the cloud fraction is close to 100 % for much of the simulation. Therefore, the 2Cat-Cld method behaves identically to the LH-only

method for much of the simulation. The 2Cat-CldPcp reduces the number of sample points in the second category (the category without cloud or precipitation) by only a factor of two. Since the second category is so small to begin with (that is, p_2 is very small), this reduction hardly improves the result at all.

5   The reason for the improvement using the 8Cat method can be inferred from Table 5. The table shows the percentage of sample points allocated to each category at the importance sampling level, averaged over the entire simulation. The 2Cat-CldPcp, 2Cat-Cld, and LH-only allocations are all similar, but the 8Cat allocation places more points in (c,p,1) than (c,p,2). That is, unlike the other three methods, the 8Cat method is able to preferentially
10   sample from mixture component 1. Component 1, in turn, contains larger cloud (liquid) water mixing ratios. The other three methods necessarily place more points in component 2 than component 1, because component 2 occupies more of the (original) PDF. However, it was shown in Table 3 that optimally, component 1 has a much higher *per-probability* sampling density than does component 2. This increased sampling of component 1 is the
15   source of the improvement of the 8Cat method.

    Figure 6 shows timeseries plots of the four tendencies at the importance sampling level. The estimate of evaporation is noisy in the three plotted configurations, because evaporation occurs outside of cloud, and the region of the sample space outside cloud is poorly sampled by all three methods. However, evaporation contributes little to the overall sum
20   because the original probability p_j outside of cloud is so small.

    Figure 7 shows the mean profile plots of the four tendencies. Once again, an ensemble of 12 simulations was used, and each sampling method overplotted is averaged over all timesteps and 12 ensemble members. All of the lines look similar, which indicates that all three methods do a good job of sampling the three processes with 32 sample points per
25   timestep, perhaps because the DYCOMS-II RF02 stratocumulus case is not highly variable. Figure 8 shows profile RMSE plots averaged over all timesteps and ensemble members. All sampling methods show the largest RMSE at around 800 m. At this level, the 8Cat method shows smaller error in autoconversion and accretion, but all methods show about the same error in evaporation.

24

Table 6 shows a quantitative comparison of the four configurations for both the RICO and DYCOMS-II RF02 cases. For each sampling method, Table 6 lists the approximate number of sample points needed to obtain the given time-averaged RMSE at the importance sampling level. These values are estimated visually from Figs. 1 and 5. This table shows that in
RICO, the 8Cat method requires approximately a factor of 8 fewer points to achieve a desired RMSE than the 2Cat-Cld method. The 2Cat-CldPcp method requires approximately a factor of 4 fewer sample points than the 2Cat-Cld method. In DYCOMS-II RF02, the reduction of necessary sample points for the given RMSE for the 8Cat method as compared to the others is a factor of approximately 1.6.

6.4 Computational Cost of the nCat Method

An important consideration among Monte Carlo integration methods is their computational cost. The cost of the new nCat method was tested against both the prior SILHS importance sampling method and the cost of CLUBB. Eight SILHS sample points were used in each simulation. Five RICO simulations were performed, and Table 7 shows the means and
standard deviations of the five simulations. Each time is a cumulative total of the respective component of the model over the entire simulation. The two nCat methods (2Cat-CldPcp and 8Cat) show no significant increase in computation time as compared to the original SILHS importance sampling method. All SILHS methods are about twice as expensive as CLUBB when eight sample points are used.

These costs may be compared with other costs in global climate simulations. To this end, Thayer-Calder et al. (2015) tested the cost of SILHS in the Community Atmosphere Model (Neale et al., 2012). They show that an adequate cloud climatology can be obtained with as few as 4 sample points (see Figs. 12 and 13 of Thayer-Calder et al. (2015)). The extra cost of computing 4 samples is $(1.89-1.69)/1.69 = 16\%$ (see Table 2 of Thayer-Calder et al.
(2015)).

7 Conclusions

We have developed a new ("nCat") method to sample subgrid variability in atmospheric models. The method divides the grid box sample space into N_{cat} categories and allows the modeler to prescribe the sampling probability in each category.

⁵ The most flexible variant of the nCat method that we consider here breaks the grid box into eight categories, depending on whether a parcel contains cloud droplets, rain droplets, or is within the first mixture component of the PDF. This "8Cat" variant allows a fine degree of control over where the samples are placed.

Another variant has been created by lumping the eight separate categories into two: one
¹⁰ that contains either cloud or precipitation, and one that contains neither cloud nor precipitation. This ("2Cat-CldPcp") variant is useful when the user does not have an estimate of the optimal sampling fraction for each of the eight categories.

We have tested the 8Cat and 2Cat-CldPcp methods on a drizzling cumulus case (RICO) and a drizzling stratocumulus case (DYCOMS-II RF02). The improvement we find relies on
¹⁵ two aspects of the method. One aspect is an algorithm that limits the weight of samples and thereby increases the number of samples in "unimportant" but large-probability categories. This helps prevent a user from becoming overzealous with importance sampling, thereby leaving excessive noise in "unimportant" categories. Another aspect is the choice of sampling variable to prescribe. We prescribe γ_j (see Eq. 21), which is related to the *density* of
²⁰ sample points in a category. This prescription allows the sampling to behave well as the cloud fraction and precipitation fraction vary widely between stratocumulus and cumulus cases.

The finer degree of control over the sampling in the nCat method allows us to improve sampling in evaporating (i.e., precipitating but non-cloudy) regions. This turns out to be a key
²⁵ to the improvement in the results. Evaporation of precipitation is an important process in the RICO case, but precipitation evaporates within only a small portion of a grid box, a portion that the nCat method can preferentially sample. Such fine-scale control of the sampling is

not possible in less flexible methods, such as the former method in SILHS, 2Cat-Cld, which does not allow importance sampling on precipitation.

Quantitative improvements are realized by the 2Cat-CldPcp and especially the 8Cat allocations. As compared to the 2Cat-Cld method, the 8Cat allocation allows a reduction in the
5 number of sample points, given equal accuracy in the tendency of autoconversion plus accretion plus rain evaporation. The reduction is approximately a factor of 1.6 in DYCOMS-II RF02 and a factor of 8 in RICO (see Figs. 1 and 5). This permits a factor of 1.6 to 8 fewer calls to the microphysics code. If a computationally expensive microphysical parameterization were used, this would result in a considerable reduction in computational cost.

10 **Appendix A: Derivation of Optimal Allocation**

The modified integral that is estimated by using importance sampling is given in Eq. (13). The goal is to minimize the centered variance of the integrand, $h(\boldsymbol{x})L(\boldsymbol{x})$, over the new sampling PDF $Q(\boldsymbol{x})$ (see Eq. 12). The variance of the integrand is given by:

$$\text{Var} = \left\{ \int [h(\boldsymbol{x})L(\boldsymbol{x})]^2 \, Q(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \right\} - \mu^2, \tag{A1}$$

15 where μ is the value of the integral in Eq. (13). The integral in curly brackets (call it I) needs to be minimized. The integral I can be split up over the N_{cat} categories:

$$I = \sum_{j=1}^{N_{\text{cat}}} \int_{C_j} [h(\boldsymbol{x})L(\boldsymbol{x})]^2 \, Q(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}. \tag{A2}$$

27

Then substituting for L (Eq. 9) and $Q = P/L$ (Eq. 12), we find

$$I = \sum_{j=1}^{N_{\text{cat}}} \int_{C_j} [h(\boldsymbol{x})]^2 \left(\frac{p_j}{S_j} \right)^2 \left(\frac{S_j}{p_j} \right) P(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$$

$$= \sum_{j=1}^{N_{\text{cat}}} \left(\frac{p_j}{S_j} \right) \int_{C_j} [h(\boldsymbol{x})]^2 P(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}. \tag{A3}$$

For convenience of notation, we make a substitution:

$$v_j = \frac{1}{p_j} \int_{C_j} [h(\boldsymbol{x})]^2 P(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}. \tag{A4}$$

Here, v_j is normalized by the probability p_j of category C_j. The quantity v_j represents the non-centered variance of the process rate $h(\boldsymbol{x})$, averaged over category C_j. Substituting Eq. (A4) into Eq. (A3), we find

$$I = \sum_{j=1}^{N_{\text{cat}}} \left(\frac{p_j^2}{S_j} \right) v_j. \tag{A5}$$

Because v_j is a within-category average, rather than a domain average, v_j may be large even when category C_j represents a small fraction of the domain. Because v_j depends on the process rate $h(\boldsymbol{x})$, v_j varies by grid box and time step.

We would like to find values of S_j that minimize Eq. (A5). Since the modified probabilities S_j must sum to one, we can use a Lagrange multiplier, which we will denote λ, and express the problem as a minimization of the following function I_λ:

$$I_\lambda \equiv \left[\sum_{j=1}^{N_{\text{cat}}} \left(\frac{p_j^2}{S_j} \right) v_j \right] + \lambda \left(\left[\sum_{j=1}^{N_{\text{cat}}} S_j \right] - 1 \right). \tag{A6}$$

28

Next we set the N_{cat} partial derivatives of I_λ with respect to each S_j to zero:

$$\left.\frac{\partial I_\lambda}{\partial S_j}\right|_{S_i, i \neq j} = 0. \tag{A7}$$

This yields

$$-\left(\frac{p_j^2}{S_j^2}\right)v_j + \lambda = 0. \tag{A8}$$

5 Rearranging, we find

$$S_j = \frac{p_j \sqrt{v_j}}{\sqrt{\lambda}}. \tag{A9}$$

Here, $\sqrt{v_j}$ may be interpreted as a sort of non-centered standard deviation of the process rate $h(\boldsymbol{x})$, averaged over category C_j. The S_j must sum to one, and so λ is determined to be

$$\lambda = \left(\sum_{j=1}^{N_{\text{cat}}} p_j \sqrt{v_j}\right)^2. \tag{A10}$$

Hence

$$S_j = \frac{p_j \sqrt{v_j}}{\sum_{i=1}^{N_{\text{cat}}} p_i \sqrt{v_i}}. \tag{A11}$$

Code availability

The CLUBB-SILHS code is freely available for non-commerical use after register-
15 ing for an account on the website http://clubb.larson-group.com. The specific ver-
sion of CLUBB-SILHS used in this paper is available in the SVN repository

located at http://carson.math.uwm.edu/repos/clubb_repos/tags/SILHS_flex_importance_sampling_paper_v2. In the repository is a file named `README_flexiblesampling` which gives instructions for reproducing the plots in this paper.

References

Barker, H. W., Pincus, R., and Morcrette, J.-J.: The Monte Carlo independent column approximation: application within large-scale models, in: Proceedings of the GCSS workshop, Kananaskis, Alberta, Canada, Amer. Meteor. Soc., 2002.

Barker, H. W., Cole, J. N. S., Morcrette, J.-J., Pincus, R., Räisänen, P., von Salzen, K., and Vaillancourt, P. A.: The Monte Carlo independent column approximation: an assessment using several global atmospheric models, Q. J. Roy. Meteor. Soc., 134, 1463–1478, 2008.

Boutle, I., Abel, S., Hill, P., and Morcrette, C.: Spatial variability of liquid cloud and rain: observations and microphysical effects, Q. J. Roy. Meteor. Soc., 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.

Chowdhary, K., Salloum, M., Debusschere, B., and Larson, V. E.: Quadrature methods for the calculation of subgrid microphysical moments, Mon. Weather Rev., 143, 2955–2972, 2015.

Colucci, P. J., Jaberi, F. A., Givi, P., and Pope, S. B.: Filtered density function for large eddy simulation of turbulent reacting flows, Phys. Fluids, 10, 499–515, 1998.

Gentle, J. E.: Random Number Generation and Monte Carlo Methods, 2nd Edn., Springer, New York, NY, USA, 2003.

Germano, M.: Turbulence: the filtering approach, J. Fluid Mech., 238, 325–336, 1992.

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, 2002.

Golaz, J.-C., Salzmann, M., Donner, L. J., Horowitz, L. W., Ming, Y., and Zhao, M.: Sensitivity of the aerosol indirect effect to subgrid variability in the cloud parameterization of the GFDL Atmosphere General Circulation Model AM3, J. Climate, 24, 3145–3160, doi:10.1175/2010JCLI3945.1, 2011.

Griffin, B. M. and Larson, V. E.: Analytic upscaling of local microphysics parameterizations, Part II: Simulations, Q. J. Royal Meteor. Soc., 139, 58–69, 2013.

Hill, P. G., Manners, J., and Petch, J. C.: Reducing noise associated with the Monte Carlo Independent Column Approximation for weather forecasting models, Q. J. Roy. Meteor. Soc., 137, 219–228, 2011.

Kalos, M. H. and Whitlock, P. A.: Monte Carlo Methods, 2nd Edn., Wiley-Blackwell, Hoboken, NJ, USA, 2008.

Khairoutdinov, M. and Kogan, Y.: A new cloud physics parameterization in a large-eddy simulation model of marine stratocumulus, Mon. Weather Rev., 128, 229–243, 2000.

Larson, V. E.: From cloud overlap to PDF overlap, Q. J. Roy. Meteor. Soc., 133, 1877–1891, doi:10.1002/qj.165, 2007.

Larson, V. E. and Golaz, J.-C.: Using probability density functions to derive consistent closure relationships among higher-order moments, Mon. Weather Rev., 133, 1023–1042, 2005.

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, Q. J. Roy. Meteor. Soc., 139, 46–57, 2013.

Larson, V. E. and Schanen, D. P.: The Subgrid Importance Latin Hypercube Sampler (SILHS): a multivariate subcolumn generator, Geosci. Model Dev., 6, 1813–1829, doi:10.5194/gmd-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.

Larson, V. E., Kotenberg, K. E., and Wood, N. B.: An analytic longwave radiation formula for liquid layer clouds, Mon. Weather Rev., 135, 689–699, 2007.

Larson, V. E., Schanen, D. P., Wang, M., Ovchinnikov, M., and Ghan, S.: PDF parameterization of boundary layer clouds in models with horizontal grid spacings from 2 to 16 km, Mon. Weather Rev., 140, 285–306, 2012.

Lebsock, M., Morrison, H., and Gettelman, A.: Microphysical implications of cloud-precipitation co-variance derived from satellite remote sensing, J. Geophys. Res.-Atmos., 118, 6521–6533, 2013.

Lemieux, C.: Monte Carlo and Quasi-Monte Carlo Sampling, Springer Science and Business Media, New York, NY, USA, 2009.

Leonard, A.: Energy cascade in large-eddy simulations of turbulent fluid flows, Adv. Geophys., 18A, 237–248, 1974.

McKay, M. D., Beckman, R. J., and Conover, W. J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics, 21, 239–245, 1979.

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.

Neale, R. B., A. Gettelman, S. Park, C.-C. Chen, P. H. Lauritzen, D. L. Williamson, A. J. Conley, D. Kinnison, D. Marsh, A. K. Smith, F. Vitt, R. Garcia, J.-F. Lamarque, M. Mills, S. Tilmes, H. Morrison, P. Cameron-Smith, W. D. Collins, M. J. Iacono, R. C. Easter, S. J. Ghan, X. Liu, P. J. Rasch, and M. A. Taylor, 2012: Description of the NCAR Community Atmosphere Model (CAM5.0). Technical Report NCAR/TN-486+STR, National Center for Atmospheric Research, Boulder, CO, USA.

Owen, A. B.: Quasi-Monte Carlo techniques, in: Siggraph 2003, Course 44, Association for Computing Machinery, San Diego, CA, 2003.

Pincus, R., Barker, H. W., and Morcrette, J.-J.: A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous cloud fields, J. Geophys. Res., 108, 4376, doi:10.1029/2002JD003322, 2003.

Pincus, R., Hemler, R., and Klein, S. A.: Using stochastically-generated subcolumns to represent cloud structure in a large-scale model, Mon. Weather Rev., 134, 3644–3656, 2006.

Pope, S. B.: Turbulent Flows, Cambridge University Press, 2000.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.: Numerical Recipes: The Art of Scientific Computing, 3rd Edn., Cambridge University Press, 2007.

Räisänen, P. and Barker, H. W.: Evaluation and optimization of sampling errors for the Monte Carlo independent column approximation, Q. J. Roy. Meteor. Soc., 130, 2069–2085, 2004.

Räisänen, P., Barker, H. W., Khairoutdinov, M. F., Li, J., and Randall, D. A.: Stochastic generation of subgrid-scale cloudy columns for large-scale models, Q. J. Roy. Meteor. Soc., 130, 2047–2067, 2004.

Räisänen, P., Barker, H. W., and Cole, J. N. S.: The Monte Carlo independent column approximation's conditional random noise: impact on simulated climate, J. Climate, 18, 4715–4730, doi:10.1175/JCLI3556.1, 2005.

Räisänen, P., Järvenoja, S., Järvinen, H., Giorgetta, M., Roeckner, E., Jylhä, K., and Ruosteenoja, K.: Tests of Monte Carlo independent column approximation in the ECHAM5 atmospheric GCM, J. Climate, 20, 4995–5011, doi:10.1175/JCLI4290.1, 2007.

Räisänen, P., Järvenoja, S., and Järvinen, H.: Noise due to the Monte Carlo independent-column approximation: short-term and long-term impacts in ECHAM5, Q. J. Roy. Meteor. Soc., 134, 481–495, 2008.

Rauber, R. M., Stevens, B., Ochs, H. T., Knight, C., Albrecht, B. A., Blyth, A. M., Fairall, C. W., Jensen, J. B., Lasher-Trapp, S. G., Mayol-Bracero, O. L., Vali, G., Anderson, J. R., Baker, B. A., Bandy, A. R., Burnet, E., Brenguier, J.-L., Brewer, W. A., Brown, P. R. A., Chuang, P., Cotton, W. R., Girolamo, L. D., Geerts, B., Gerber, H., Göke, S., Gomes, L., Heikes, B. G., Hudson, J. G., Kollias, P., Lawson, R. P., Krueger, S. K., Lenschow, D. H., Nuijens, L., O'Sullivan, D. W., Rilling, R. A., Rogers, D. C., Siebesma, A. P., Snodgrass, E., Stith, J. L., Thornton, D. C., Tucker, S., Twohy, C. H., Zuidema, P., Sperber, K. R., and Waliser, D. E.: Rain in shallow cumulus over the ocean: the RICO campaign, B. Am. Meteorol. Soc., 88, 1912–1928, 2007.

Stevens, B., Lenschow, D. H., Vali, G., Gerber, H., Bandy, A., Blomquist, B., Brenguier, J.-L., Bretherton, C. S., Burnet, F., Campos, T., Chai, S., Faloona, I., Friesen, D., Haimov, S., Laursen, K., Lilly, D. K., Loehrer, S. M., Malinowski, S. P., Morley, B., Petters, M. D., Rogers, D. C., Russell, L., Savic-Jovcic, V., Snider, J. R., Straub, D., Szumowski, M. J., Takagi, H., Thornton, D. C., Tschudi, M., Twohy, C., Wetzel, M., and Van Zanten, M. C.: Dynamics and chemistry of marine stratocumulus – DYCOMS-II, B. Am. Meteorol. Soc., 84, 579–593, 2003.

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.

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/gmdd-8-3801-2015, 2015.

Tonttila, J., Räisänen, P., and Järvinen, H.: Monte Carlo-based subgrid parameterization of vertical velocity and stratiform cloud microphysics in ECHAM5.5-HAM2, Atmos. Chem. Phys., 13, 7551–7565, doi:10.5194/acp-13-7551-2013, 2013.

Tonttila, J., Järvinen, H., and Räisänen, P.: Explicit representation of subgrid variability in cloud microphysics yields weaker aerosol indirect effect in the ECHAM5-HAM2 climate model, Atmos. Chem. Phys., 15, 703–714, doi:10.5194/acp-15-703-2015, 2015.

vanZanten, 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.

Wyant, M. C., Bretherton, C. S., Chlond, A., Griffin, B. M., Kitagawa, H., Lappen, C.-L., Larson, V. E., Lock, A., Park, S., de Roode, S. R., Uchida, J., Zhao, M., and Ackerman, A. S.: A single-column-model intercomparison of a heavily drizzling stratocumulus topped boundary layer, J. Geophys. Res., 112, D24204, doi:10.1029/2007JD008536, 2007.

Xiu, D.: Fast numerical methods for stochastic computations: a review, Commun. Comput. Phys., 5, 242–272, 2009.

Zhang, J., Lohmann, U., and Lin, B.: A new statistically based autoconversion rate parameterization for use in large-scale models, J. Geophys. Res., 107, 4750, doi:10.1029/2001JD001484, 2002.

Table 1. For each of the eight categories, this table lists (1) the category number; (2) whether the category is cloudy, in mixture component 1 or 2, or precipitating; (3) what inequalities must be satisfied for a sample point to lie within the category; and (4) the original probability mass associated with the category, p_j.

Category number	Category description	Category selection criteria	Category probability p_j
1	In cloud In mixt. comp. 1 In precip	$\chi > 0$ $u_{d+1} < \xi_{(1)}$ $u_{d+2} < f_{p(1)}$	$f_{c(1)} \times \xi_{(1)} \times f_{p(1)}$
2	In cloud In mixt. comp. 2 In precip	$\chi > 0$ $u_{d+1} \geq \xi_{(1)}$ $u_{d+2} < f_{p(2)}$	$f_{c(2)} \times \xi_{(2)} \times f_{p(2)}$
3	Out of cloud In mixt. comp. 1 In precip	$\chi \leq 0$ $u_{d+1} < \xi_{(1)}$ $u_{d+2} < f_{p(1)}$	$(1 - f_{c(1)}) \times \xi_{(1)} \times f_{p(1)}$
4	Out of cloud In mixt. comp. 2 In precip	$\chi \leq 0$ $u_{d+1} \geq \xi_{(1)}$ $u_{d+2} < f_{p(2)}$	$(1 - f_{c(2)}) \times \xi_{(2)} \times f_{p(2)}$
5	In cloud In mixt. comp. 1 Out of precip	$\chi > 0$ $u_{d+1} < \xi_{(1)}$ $u_{d+2} \geq f_{p(1)}$	$f_{c(1)} \times \xi_{(1)} \times (1 - f_{p(1)})$
6	In cloud In mixt. comp. 2 Out of precip	$\chi > 0$ $u_{d+1} \geq \xi_{(1)}$ $u_{d+2} \geq f_{p(2)}$	$f_{c(2)} \times \xi_{(2)} \times (1 - f_{p(2)})$
7	Out of cloud In mixt. comp. 1 Out of precip	$\chi \leq 0$ $u_{d+1} < \xi_{(1)}$ $u_{d+2} \geq f_{p(1)}$	$(1 - f_{c(1)}) \times \xi_{(1)} \times (1 - f_{p(1)})$
8	Out of cloud In mixt. comp. 2 Out of precip	$\chi \leq 0$ $u_{d+1} \geq \xi_{(1)}$ $u_{d+2} \geq f_{p(2)}$	$(1 - f_{c(2)}) \times \xi_{(2)} \times (1 - f_{p(2)})$

35

Table 2. Notable configuration settings for the RICO and DYCOMS-II RF02 simulations performed in this paper.

	RICO	DYCOMS-II RF02
Timestep (s)	300	60
Vertical levels	128	160
Vertical grid spacing (m)	25–250	10
Radiation	None	Analytic longwave (Larson et al., 2007)
Cloud droplet concentration (m^{-3})	70×10^6	55×10^6

Table 3. Estimated optimal sampling fractions (γ_j) for each importance category, averaged over the entire simulation, for the RICO and DYCOMS-II RF02 cases. These estimates were obtained by using SILHS to estimate the right-hand side of Eq. (21) for each category. Here "c" denotes "in-cloud", "nc" denotes "out of cloud", "p" denotes "in-precipitation", "np" denotes "out of precipitation", "1" denotes "in mixture component 1", and "2" denotes "in mixture component 2".

Category	1 (c,p,1)	2 (c,p,2)	3 (nc,p,1)	4 (nc,p,2)	5 (c,np,1)	6 (c,np,2)	7 (nc,np,1)	8 (nc,np,2)
RICO	0.539	0.004	0.223	0.203	0.033	0.004	0.000	0.000
DYCOMS-II RF02	0.351	0.143	0.238	0.061	0.140	0.070	0.000	0.000

Table 4. RICO: Percentage of sample points allocated to each category by each sampling method at the importance sampling level, time-averaged over the entire simulation. The more sample points placed in a particular category, the better the estimate of processes active in that category.

Category	8Cat	2Cat-CldPcp	2Cat-Cld	LH-only
(c,p,1)	12.1	5.7	13.6	0.3
(c,p,2)	0.03	0.03	0.04	0.02
(nc,p,1)	16.1	11.0	0.3	0.5
(nc,p,2)	7.9	18.8	0.4	0.7
(c,np,1)	14.8	15.4	36.0	0.7
(c,np,2)	0.2	0.3	0.4	0.2
(nc,np,1)	0.7	0.7	0.7	1.3
(nc,np,2)	48.2	48.2	48.7	96.3

Table 5. DYCOMS-II RF02: Percentage of sample points allocated to each category by each sampling method at the importance sampling level, time-averaged over the entire simulation. The more sample points placed in a particular category, the better the estimate of processes active in that category.

Category	8Cat	2Cat-CldPcp	2Cat-Cld	LH-only
(c,p,1)	57.3	35.0	35.1	35.1
(c,p,2)	41.2	63.0	63.0	63.1
(nc,p,1)	0.5	0.5	0.5	0.5
(nc,p,2)	0.2	0.3	0.3	0.3
(c,np,1)	0.6	0.7	0.7	0.7
(c,np,2)	0.2	0.4	0.4	0.4
(nc,np,1)	0	0	0.01	0.02
(nc,np,2)	0.03	0.04	0.07	0.07

Table 6. Number of sample points needed by each configuration of SILHS to achieve a given RMSE in estimating the sum of the three processes, for the RICO and DYCOMS-II RF02 cases. These numbers are estimated visually from Figs. 1 and 5. In RICO, the 2Cat-CldPcp method requires approximately a factor of 4 fewer sample points than the 2Cat-Cld method to achieve an RMSE of 10^{-8} kg kg^{-1}, and 8Cat method requires approximately a factor of 8 fewer sample points. In DYCOMS-II RF02, the 8Cat method requires approximately a factor of 1.6 fewer sample points than the 2Cat-CldPcp, 2Cat-Cld, and LH-only methods to achieve an RMSE of 4×10^{-9} kg kg^{-1}.

| Method	Samples for 10^{-8} RMSE (RICO)	Samples for 4×10^{-9} RMSE (DYCOMS-II RF02)
LH-only	~ 700	13
2Cat-Cld	65	13
2Cat-CldPcp	15	13
8Cat	8	8

Table 7. Cumulative run time of CLUBB and the different SILHS configurations over an 864-timestep RICO simulation. Each SILHS configuration uses 8 sample points. The means and standard deviations of five simulations are shown in the table. All times are in seconds. The two nCat methods (2Cat-CldPcp and 8Cat) show no significant increase in computation time as compared to the original SILHS importance sampling method. All SILHS methods, with eight sample points, are more expensive than CLUBB.

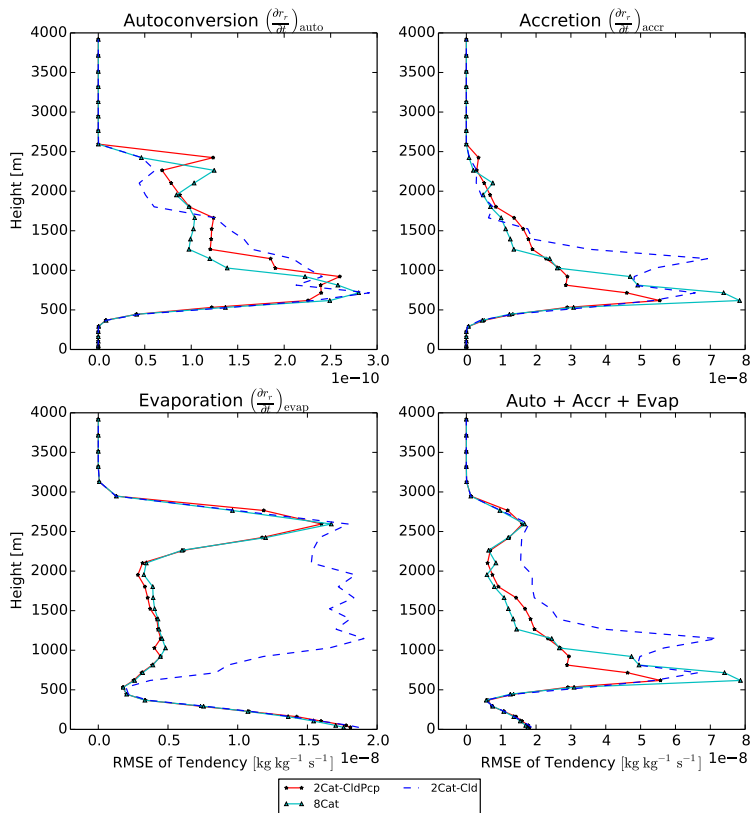|	Average Time [s]	Std. Dev [s]
CLUBB	0.311	0.007
SILHS (old)	0.691	0.013
SILHS (2Cat-CldPcp)	0.701	0.013
SILHS (8Cat)	0.698	0.007
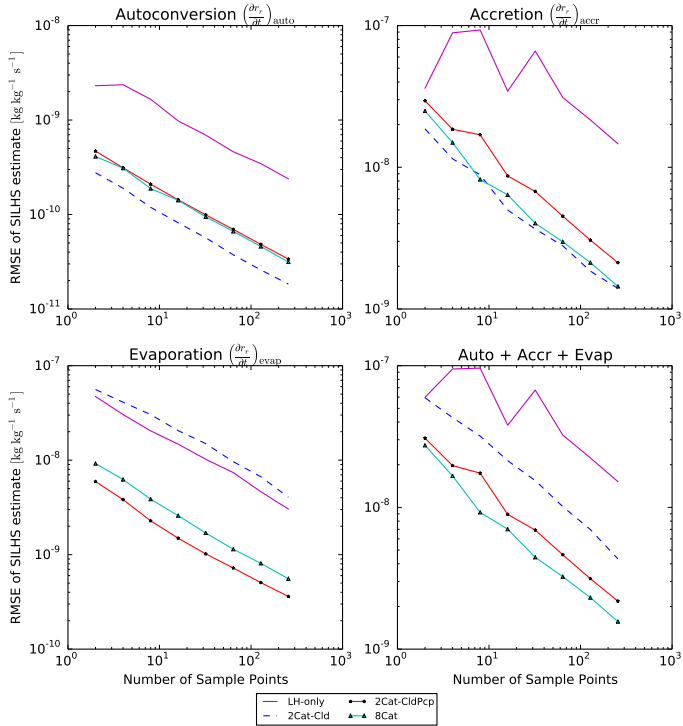
Figure 1. RICO: the root-mean-square error (RMSE) at the importance sampling level of SILHS simulations as a function of the number sample points, for the RICO cumulus case. The error is time-averaged over the entire simulation. The 2Cat-CldPcp and 8Cat methods show a large improvement over the 2Cat-Cld and LH-only methods in the estimate of evaporation, but not for autoconvesion and accretion, which are in-cloud processes. Nevertheless, the 8Cat and 2Cat-CldPcp methods both impove the estimate of the sum of the three processes.
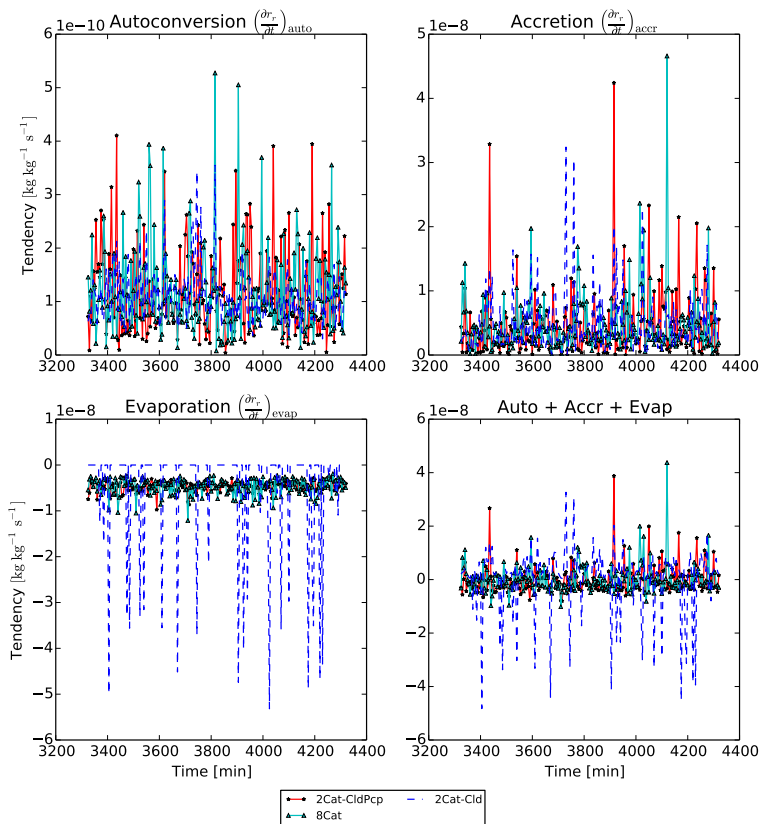
Figure 2. RICO: timeseries plots of the four tendencies at the importance sampling level. The simulations in these plots use 32 sample points, and the plots show minutes 3321 to 4320 of the simulations. To improve readibility, the LH-only method is not plotted. The evaporation tendencies are much more noisy in the 2Cat-Cld method than in the 2Cat-CldPcp or 8Cat methods.
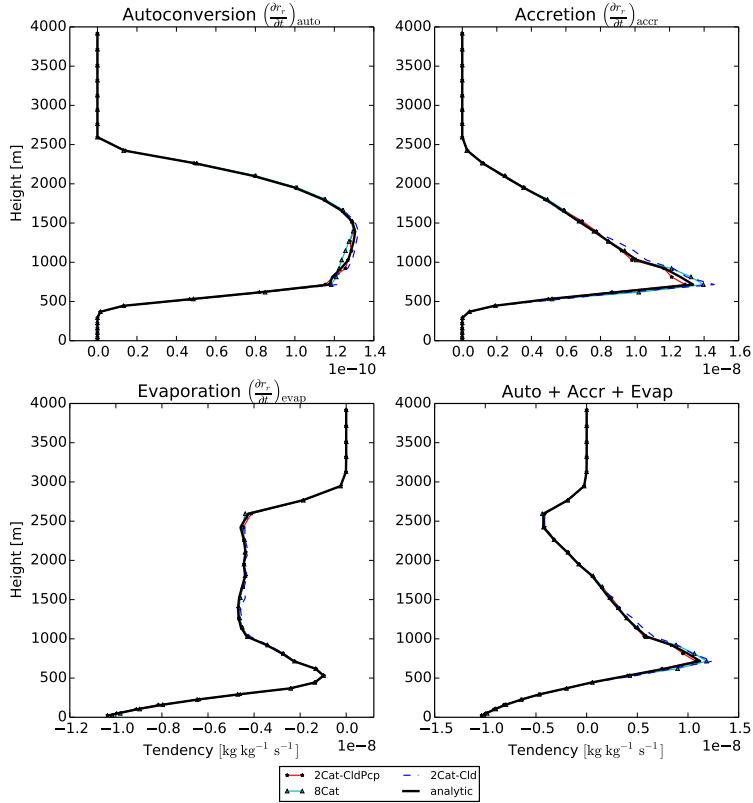
Figure 3. RICO: mean profile plots of the four tendencies. The simulations in these plots use 32 sample points. For each configuration, an ensemble of 12 simulations is used, each with a different seed. Profiles are averaged over all 864 timesteps of the simulation and all 12 ensemble members. It is seen that all SILHS sampling methods are clearly convergent at all height levels.

44

Autoconversion $\left(\frac{\partial r_r}{\partial t}\right)_{\mathrm{auto}}$

Accretion $\left(\frac{\partial r_r}{\partial t}\right)_{\mathrm{accr}}$

Evaporation $\left(\frac{\partial r_r}{\partial t}\right)_{\mathrm{evap}}$

Auto + Accr + Evap

RMSE of Tendency $[\mathrm{kg\,kg^{-1}\,s^{-1}}]$

Height [m]

2Cat-CldPcp
8Cat
2Cat-Cld

45

Figure 4. RICO: profile RMSE plots of the four tendencies. The simulations in these plots use 32 sample points. For each configuration, an ensemble of 12 simulations is used, each with a different seed. RMSE values are averaged over all 864 timesteps of the simulation and all 12 ensemble members. The 8Cat and 2Cat-CldPcp methods show improvement between 1000 m and 2500 m, where the improvement in sampling of the evaporation term is largest. All three methods suffer from extra noise below 1000 m, which is far away from the importance sampling level. The importance sampling level is just under 2000 m for most timesteps in the simulation.
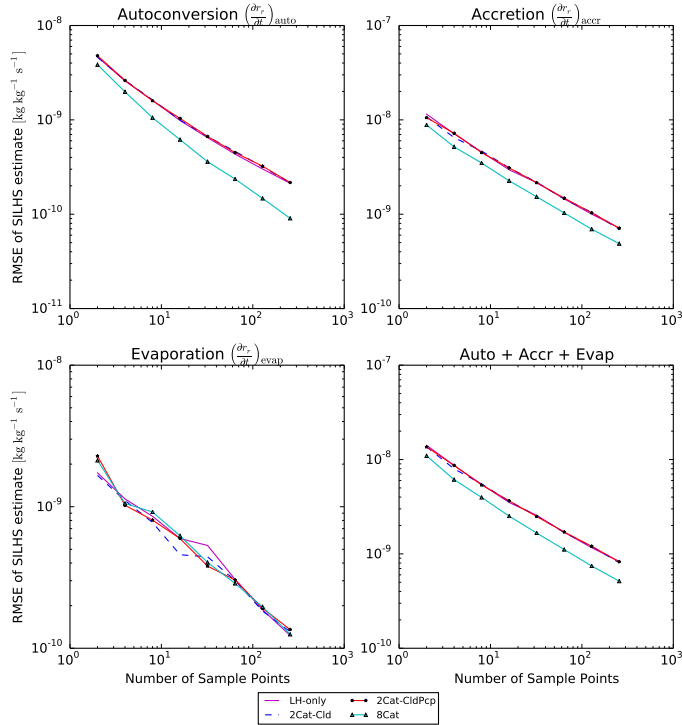
Figure 5. DYCOMS-II RF02: the root-mean-square error (RMSE) of SILHS simulations as a function of sample points for the DYCOMS-II RF02 stratocumulus case. The error is calculated at the importance sampling level and is averaged over all timesteps of the simulation. The LH-only, 2Cat-Cld, and 2Cat-CldPcp methods are expected to have roughly the same behavior in a case like DYCOMS-II RF02 that has cloud fraction near 100 %. The 8Cat method still improves the estimates of autoconversion and accretion because it is able to flexibly allocate points within the cloudy region of the sample space.

47
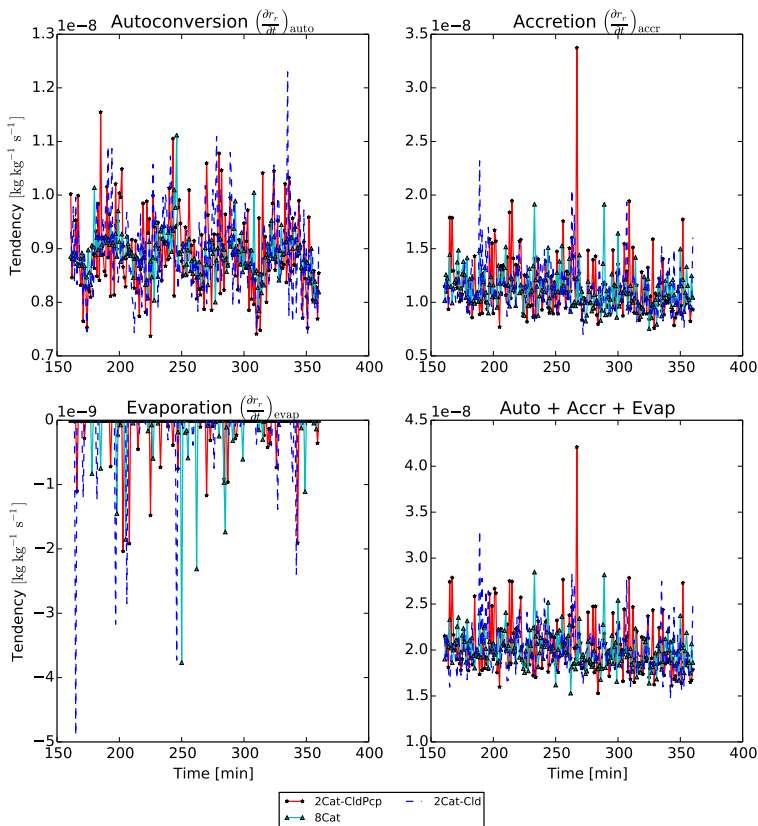
Figure 6. DYCOMS-II RF02: timeseries plots of the four tendencies at the importance sampling level. The simulations in these plots use 32 sample points. The time range plotted includes minutes 161 to 360 of the simulation. The evaporation process is poorly sampled in all three sampling methods, but it is a relatively small term and makes a much smaller contribution to the sum of the three processes than autoconversion and accretion.

48
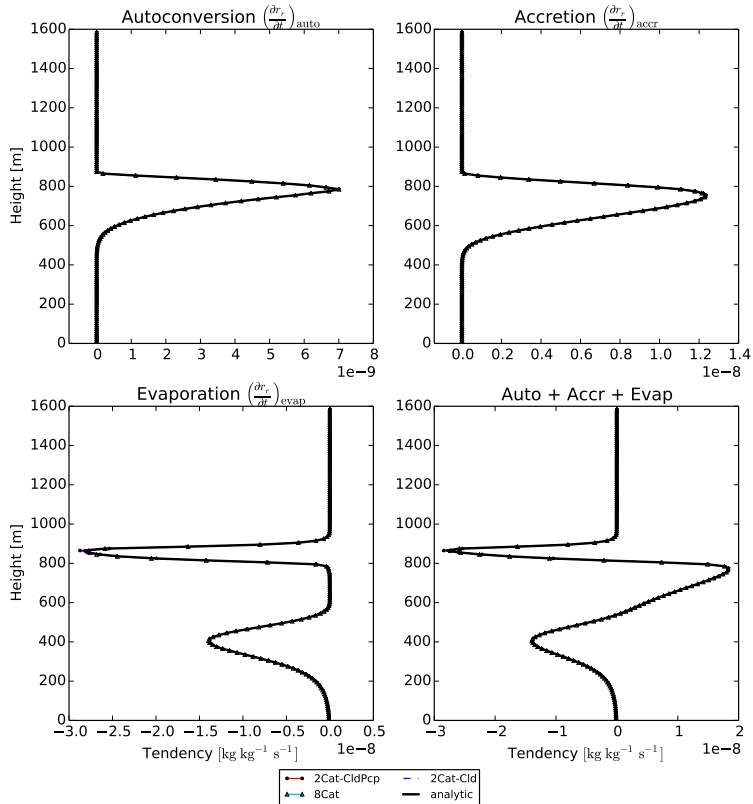
Figure 7. DYCOMS-II RF02: mean profile plots of the four tendencies. The simulations in these plots use 32 sample points. For each configuration, an ensemble of 12 simulations is used, each with a different seed. Profiles are averaged over all 360 timesteps of the simulation and all 12 ensemble members. It is seen that all SILHS sampling methods are clearly convergent at all height levels. All of the lines overlap well, indicating that all three processes are sampled well by all three sampling methods.
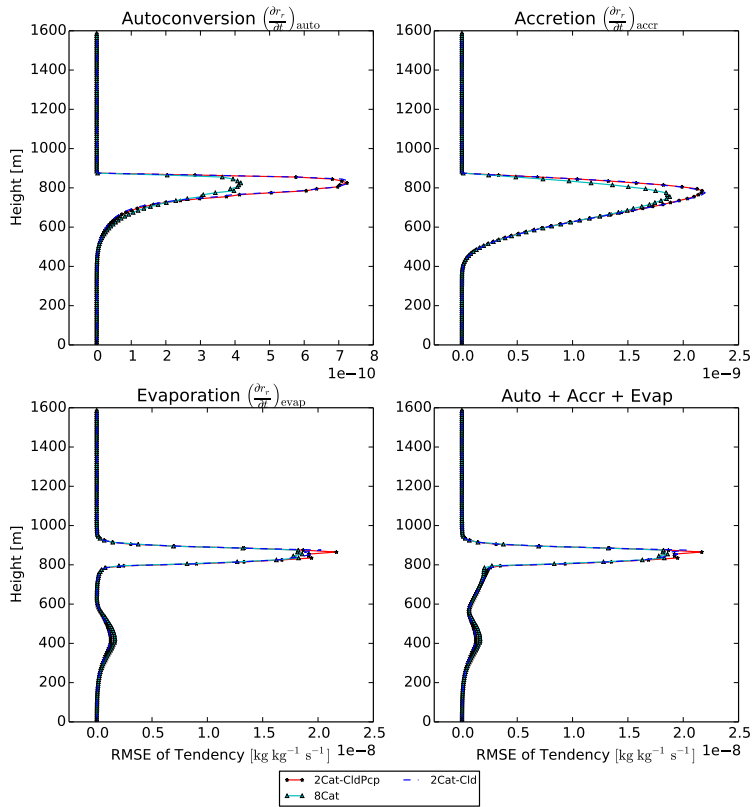
Figure 8. DYCOMS-II RF02: profile RMSE plots of the four tendencies. The simulations in these plots use 32 sample points. For each configuration, an ensemble of 12 simulations is used, each with a different seed. RMSE values are averaged over all 360 timesteps of the simulation and all 12 ensemble members. All sampling methods show the largest RMSE at around 800 m. At this level, the 8Cat method shows smaller error in autoconversion and accretion, but all methods show about the same error in evaporation.