Droplet size distributions in turbulent clouds: experimental evaluation of theoretical distributions

Kamal Kant Chandrakar1 | Izumi Saito2 | Fan Yang3 | Will Cantrell1 | Toshiyuki Gotoh2 | Raymond A. Shaw1

1Department of Physics, Michigan Technological University, Houghton, Michigan USA
2Department of Physical Science and Engineering, Nagoya Institute of Technology, Nagoya, Japan
3Brookhaven National Laboratory, Upton, New York USA

Correspondence
R. A. Shaw, Department of Physics, Michigan Technological University, Houghton, MI 49931, USA.
Email: rashaw@mtu.edu

Funding information
JSPS KAKENHI, 18K13611; MEXT KAKENHI, 15H02218; National Aeronautics and Space Administration, 80NSSC17K0449; National Science Foundation, AGS-1754244; US Department of Energy, DE-SC0012704

Abstract
Precipitation efficiency and optical properties of clouds, both central to determining Earth’s weather and climate, depend on the size distribution of cloud particles. In this work theoretical expressions for cloud droplet size distribution shape are evaluated using measurements from controlled experiments in a convective-cloud chamber. The experiments are a unique opportunity to constrain theory because they are in steady-state and because the initial and boundary conditions are well characterized compared to typical atmospheric measurements. Three theoretical distributions obtained from a Langevin drift-diffusion approach to cloud formation via stochastic condensation are tested: (a) stochastic condensation with a constant removal time-scale; (b) stochastic condensation with a size-dependent removal time-scale; (c) droplet growth in a fixed supersaturation condition and with size-dependent removal. In addition, a similar Weibull distribution that can be obtained from the drift-diffusion approach, as well as from mechanism-independent probabilistic arguments (e.g., maximum entropy), is tested as a fourth hypothesis. Statistical techniques such as the $\chi^2$ test, sum of squared errors of prediction, and residual analysis are employed to judge relative success or failure of the theoretical distributions to describe the experimental data. An extensive set of cloud droplet size distributions are measured under different aerosol injection rates. Five different aerosol injection rates are run both for size-selected aerosol particles, and six aerosol injection rates are run for broad-distribution, polydisperse aerosol particles. In relative comparison, the most favourable comparison to the measurements is the expression for stochastic condensation with size-dependent droplet removal rate. However, even this optimal distribution breaks down for broad aerosol size distributions, primarily due to deviations from the measured large-droplet tail. A possible explanation for the deviation is the Ostwald ripening effect coupled with deactivation/activation in polluted cloud conditions.

Keywords
aerosol indirect effects, cloud droplet size distribution, Ostwald ripening, stochastic condensation, turbulent clouds
**1 | INTRODUCTION**

The shape of the droplet size distribution (DSD) is central to understanding atmospheric clouds, with applications ranging from remote sensing retrieval to parametrization of precipitation and radiative transfer in climate models. For example, optical reflectivity of stratocumulus clouds, known as global reflectors, is related approximately to the second moment of the cloud DSD. Many empirical expressions for DSD shape have been proposed and are used in computational models as well as in remote-sensing retrievals, the latter even of relevance to investigation of exoplanet atmospheres. However, the uniqueness of a functional shape for the cloud DSD is questionable due to the existence of different physical mechanisms for droplet formation, growth and removal, and their significance in different regions of cloud or stages of cloud growth. Physical processes affecting the cloud DSD include the curvature and solute effects (and associated mechanisms such as spectral ripening and competing activation of cloud condensation nuclei of varying size, solubility, etc.) (Johnson, 1982; Korolev, 1995; Wood et al. 2002; Yang et al. 2018), turbulent fluctuations (Cooper, 1989; Paoli and Shariff, 2009; Sardina et al. 2015; Chandrakar et al. 2016; Siewert et al. 2017), entrainment and mixing (Baker et al. 1980; Lehmann et al. 2009; Yum et al. 2015), microphysical variability (Cooper, 1989; Desai et al. 2018), internal mixing of parcels of different growth history (Hudson and Yum, 1997; Lasher-Trapp et al. 2005), and droplet collision–coalescence (Pruppacher and Klett, 2010; Glienke et al. 2017; Witte et al. 2017). Cloud droplet growth prior to onset of collision–coalescence occurs through water vapour condensation, and therefore the central variable affecting the DSD is the water vapour supersaturation, including both its mean value and its variability (Ditas et al. 2012; Siebert and Shaw, 2017).

Often in applications of cloud physics, such as for climate and numerical weather prediction models and for remote sensing, empirical distribution shapes are used for obtaining relevant moments. Empirical functional forms commonly used for distribution shape are the gamma (Khrgian and Mazin distribution), log-normal, and exponential (Liu et al. 1995; Miles et al. 2000; Cotton et al. 2010; Pruppacher and Klett, 2010). Undoubtedly it would be a significant advance to identify theoretical distribution shapes from fundamental physical principles, even if restricted to some limiting atmospheric conditions. Several different theoretical models have been proposed to characterize the shape of cloud DSDs formed through the condensation process. Two approaches that have received considerable recent attention are systems-theory derivations based on the principle of maximum entropy (Liu and Hallett, 1998; Liu et al. 2002; Yano et al. 2016; Wu and McFarquhar, 2018) and derivations based on a Langevin equation representation of stochastic condensation (McGraw and Liu, 2006; Paoli and Shariff, 2009; Chandrakar et al. 2016; Siewert et al. 2017; Abade et al. 2018; Saito et al. 2019). By and large, however, these theoretical models have not been subjected to systematic experimental evaluation to validate the proposed functional forms for the DSD. This is primarily because the sparsity of measurements from field experiments and the natural variability of atmospheric clouds makes this evaluation process ambiguous: large uncertainties in field measurements do not effectively constrain distribution shape, especially in the important but often under-sampled distribution tails. In this article we describe well-controlled and characterized laboratory experiments to assess several predicted distribution shapes from a theoretical model of stochastic condensation. We use statistical techniques for goodness-of-fit analysis to judge the relative success or failure to describe the experimental data over a range of aerosol conditions, i.e., from relatively clean to polluted conditions. Furthermore, in addition to varying the aerosol concentration, we consider size-selected aerosols as well as a broad distribution of aerosol sizes. The experiments are conducted under conditions that are statistically steady in time and without the complications of collision–coalescence. This allows us to evaluate the theoretical models under conditions as close as possible in replicating the assumed simplifications in their derivation.

In the next section we describe the Langevin model for cloud droplet growth and the analytical solutions to the resulting Fokker–Planck equation that correspond to different assumptions regarding turbulence and droplet removal. The three distributions correspond to assumptions of: (a) stochastic condensation growth with size-independent removal rate, (b) stochastic condensation growth with size-dependent (Stokes settling) removal rate, and (c) condensation growth due to just mean supersaturated environment (no fluctuations) with droplet removal due to Stokes settling. For completeness, we also include a distribution resulting from the maximum-entropy (systems theory) approach, based on the constraint of conservation of liquid water. In Section 4 we outline the design of the experiment and the methodology for fitting analytical distributions to the observations, as well as the statistical tools used for evaluating suitability (via the $\chi^2$ test) and goodness of fit of the distributions. In Section 5 we present the data along with the fitted distributions for results from two sets of experiments: one with size-selected aerosol particles at several different aerosol concentrations, and one with a broad size range of aerosol particles, also at several different aerosol concentrations. We conclude with a discussion of the results and
their implications for our understanding of atmospheric clouds.

2 | SOLUTIONS TO THE FOKKER–PLANCK EQUATION FOR THE CLOUD DSD

Atmospheric clouds are inherently turbulent. This implies that if we could measure the supersaturation experienced by a population of droplets at a given instant (or alternatively, the supersaturation experienced by a droplet over time), it would be a highly variable quantity. The droplet condensation process is therefore stochastic, and at least approximately, we can draw an analogy between random “diffusion” of cloud droplet radius squared within size space, and Brownian molecular diffusion. The turbulence statistics, including the distribution of supersaturation and the Lagrangian correlation timescale of scalar fluctuations, drive this stochastic condensation process. Effectiveness of stochastic methods in modelling processes such as Brownian diffusion and turbulent combustion (Pope, 2000) has motivated the application of this mathematical technique to droplet condensation as well. The condensation process can be approximately represented as Brownian diffusion in the space of droplet surface area. Therefore, a Langevin equation for Brownian diffusion of droplet surface area (condensational growth of cloud droplets) with a drift due to mean supersaturation (5) experience by droplets and diffusion due to the supersaturation fluctuations (standard deviation $\sigma_s$) can be represented as (Gardiner, 2009; Jacobs, 2010):

$$dx = a \, dt + \sqrt{D} \, dw,$$

where, $a = 2\xi \xi \xi$, $\xi$ is the droplet growth factor and depends on thermodynamic quantities (Rogers and Yau, 1989), $D = \partial \sigma^2 \propto \sigma^2$, $x \equiv r^2$, and $r$ is the droplet radius. In Equation (1), the first term is the drift due to the mean supersaturation experienced by a group of cloud droplets; it represents mean $r^2$ growth, $\partial_t r^2 = 2 \xi \xi \xi$ (Rogers and Yau, 1989). The last term is the diffusion in $r^2$ space ($\partial_t \sigma^2$) due to supersaturation fluctuations with diffusivity $D$, and $dw$ is the differential of the Wiener process. A model for $r^2$ diffusivity can be obtained from the expression of droplet size dispersion growth presented in Chandrakar et al. (2016), where $\partial_t \sigma^2 \propto \sigma^2 \tau_r$. The time-scale governing droplet growth in the fluctuating system is $\tau_r$, 0.5 times the harmonic mean of the Lagrangian correlation time for supersaturation in the turbulent flow and the cloud phase relaxation time.

A Fokker–Plank equation describing the probability density function for droplet size can be obtained from the Langevin equation (1) with additional source and sink term due to droplet activation and removal (McGraw and Liu, 2006; Gardiner, 2009; Jacobs, 2010; Saito et al. 2019):

$$\frac{dp}{dt} = -a \frac{dp}{dx} + \frac{D}{2} \frac{\partial^2 p}{\partial x^2} + N_t \delta(x - x_{cr}) - \frac{p}{\tau_r}.$$  (2)

Here, the first term on the right-hand side represents drift in $r^2$-space (mean growth) due to the mean supersaturation, the second term signifies diffusion of cloud droplet size due to turbulence (with diffusivity $D$), the third term is probability gain due to activation of particle with critical size $x_{cr}$ (and the activation rate $N_t$), and the last term is the droplet loss or removal rate with time-scale $\tau_r$. We now proceed to describe three solutions for $p(r)$ from Equation (2) corresponding to different assumptions that could plausibly represent the conditions existing in a laboratory turbulent-cloud chamber, or in a natural cloud.

2.1 | Size distribution for mean and fluctuating supersaturation with size-independent removal

Saito et al. (2019) considered the simplifying assumptions of fixed critical size, equivalent to mono-disperse aerosol particles of the same composition, and a constant, size-independent rate of droplet removal, which is equivalent to a constant droplet lifetime (i.e., $\tau_r = \tau_r = constant$). Under those conditions, Equation (2) leads to the following steady-state solution for $x \geq x_{cr}$:

$$p(x) = A \, \exp(-A \, x).$$  (3)

In this equation the coefficient $A$ is

$$A = \sqrt{\frac{x_{m}^2 + 4\sigma_{x,m}^2 - x_{m}}{2\sigma_{x,m}^2}},$$  (4)

where, $x_m = 2 \xi \xi \xi \tau_r$, and $\sigma^2_{x,m} = D \tau_r / 2$. Alternately, in terms of droplet radius $r$, the size PDF can be expressed as:

$$p(r) = 2 \, A \, r \, \exp(-A \, r^2).$$  (5)

It is important to note here that the effect of curvature and solute terms on the equilibrium supersaturation are ignored. Ostwald ripening is therefore not accounted for in the derivation. Furthermore, because of the assumption of fixed activation size, cloud formation with a broad spectrum of aerosol sizes could lead to deviation from the measured (physical) distribution. Similarly, the assumption of
constant removal (or growth) time-scale \((\tau_r)\) could be violated if the primary removal mechanism is gravitational settling. Interestingly, a functional form of the steady-state solution similar to Equation (3), i.e., a Weibull distribution with a shape factor of 2 in radius space, can also be achieved by considering a balance between the mean drift and the diffusion terms (McGraw and Liu, 2006).

### 2.2 Size distribution for mean and fluctuating supersaturation with size-dependent removal

The constant-removal-rate assumption can be replaced with a combination of a size-independent removal time-scale (e.g., due to turbulent transport) and a size-dependent removal rate (e.g., due to gravitational settling). For a removal time-scale of the form \(\tau_{st}^{-1} = \tau_{st}^{-1} + \kappa x\), it is possible to derive an analytical solution of the corresponding Fokker–Plank equation. Here a linear form of the removal time-scale with droplet size \(x \equiv r^2\) corresponds to gravitational settling with Stokes drag. It is also possible to write a similar form of cloud droplet removal due to collision–coalescence process. An analytical solution of Equation (2) with this size-dependent removal time-scale (for \(x \geq x_{cr}\)) is:

\[
p(x) = C \exp(x_m \sigma_{x,m}^{-2} x/2) \times \text{Ai} \left( \frac{x_m^2 + 4\sigma_{x,m}^2 \tau_r^{-1} x_{st}^{-1} x_m \tau_{st}^{-1} x_m \sigma_{x,m}^{-2}}{4 \sigma_{x,m}^2 \tau_r^{-1} x_{st}^{-1} x_m \tau_{st}^{-1} x_m \sigma_{x,m}^{-2}} \right). \tag{6}
\]

Here \(x_m\) and \(\sigma_{x,m}\) are defined as in the preceding subsection, \(\tau_{st} \equiv (\kappa x_m)^{-1}\) is the average time-scale associated with the Stokes settling, and \(\text{Ai}\) is the Airy function of the first kind. The normalization constant is given by

\[
C = \left[ \int_{x_{cr}}^{\infty} \exp(x_m \sigma_{x,m}^{-2} z/2) \times \text{Ai} \left( \frac{x_m^2 + 4\sigma_{x,m}^2 \tau_r^{-1} x_{st}^{-1} x_m \tau_{st}^{-1} x_m \sigma_{x,m}^{-2}}{4 \sigma_{x,m}^2 \tau_r^{-1} x_{st}^{-1} x_m \tau_{st}^{-1} x_m \sigma_{x,m}^{-2}} \right) \, dz \right]^{-1} \tag{7}
\]

Fitting this function to data is cumbersome, and requires that the full size distribution is measured. We seek a simpler form that is valid for the large-droplet tail of the distribution, for which measurements are more reliable. An asymptotic form of Equation (6) can be obtained by Taylor series expansion of \(\text{Ai}(\alpha' x + \beta')\) (or by the steepest descent method) with the limit \(\alpha' x + \beta' \gg 1\). Here \(\alpha'\) and \(\beta'\) are constants and can be interpreted from Equation (6). We also assume zero mean supersaturation (i.e., \(a = 0\)) and just Stokes settling (i.e., \(\tau_r = (\kappa x)^{-1}\) and \(\tau_{st} = 0\)) for simplicity. The resulting distribution is

\[
p(x) \approx \frac{C}{2\sqrt{\pi}} \beta^{-1/2} \exp(-2/3\beta x^{3/2}), \tag{8}
\]

where \(\beta = 2\kappa/D\). Again it should be emphasized that this expression is valid for the right tail of the distribution. In terms of radius this can be expressed as

\[
p(r) \approx \frac{C}{\sqrt{\pi}} \beta^{-1/2} \sqrt{r} \exp(-2/3\beta r^3). \tag{9}
\]

### 2.3 Size distribution for fixed supersaturation with size-dependent removal

If we ignore supersaturation fluctuations in Equation (2) by setting \(D = 0\), we obtain an equation analogous to that familiar in the cloud physics literature (cf. equation 7.31 of Rogers and Yau 1989). The additional term describing removal of particles by gravitational settling is preserved, again assuming Stokes drag with \(\tau_r = (\kappa x)^{-1}\). Krueger (2019) showed that the resulting solution describes a DSD growing in a non-turbulent mixed parcel with gravitational settling as the only removal mechanism:

\[
p(x) = \sqrt{\frac{2\kappa}{\pi a}} \exp\left(-\left(\kappa/2a\right) x^2\right). \tag{10}
\]

In terms of droplet radius it is

\[
p(r) = 2r \sqrt{\frac{2\kappa}{\pi a}} \exp\left(-\left(\kappa/2a\right) r^4\right). \tag{11}
\]

This solution for the DSD in a fixed-supersaturation environment can reasonably be interpreted as the limiting solution when \(\tilde{s} \gg \sigma_s\).

Equations (8) and (10) provide two convenient extremes for comparison with observations; in the first case, only the supersaturation fluctuations determine the shape of the size distribution and in the second case only the mean supersaturation contributes. The Stokes-settling removal process is common in both cases. Equation (5) describes the simplest case in which turbulence is dominant, and for which droplet removal does not depend on particle size. We see already that the exponent in the distribution, which determines the shape of the large-droplet tail of the distribution, is different for the three derived distributions: \(r^2\) versus \(r^3\) versus \(r^4\). Moreover, the slope of the tail is inversely proportional to \(\sigma_s\) or \(\tilde{s}\) depending on the case.
3 SIZE DISTRIBUTIONS FROM THE PRINCIPLE OF MAXIMUM ENTROPY

Liu and Hallett (1998) proposed a functional form of the cloud DSD based on Shannon’s maximum entropy principle for a system of droplets of fixed mass and number concentration. This distribution does not have any information about the physical processes responsible for the cloud droplet growth. Instead, a maximum-likelihood size distribution is derived by maximizing the system entropy and with a constraint of conservation of liquid water content. The result in terms of droplet radius is

\[ p(r) = \frac{4\pi \rho n_d^3}{L} r^2 \exp \left( \frac{-4\pi \rho n_d}{3L} r^3 \right) \quad (12) \]

where \( n_d \) is the total droplet concentration and \( L \) is the liquid water content. This is a Weibull distribution with a shape factor of 3. In droplet-volume coordinates \( x \equiv r^3 \), the distribution is transformed to a simple exponential function of \( x \), i.e., \( p(x) \propto \exp(-x) \). Interestingly, a similar form of the DSD was obtained by Williams and Wojtowicz (1982) also using probabilistic arguments and the constraint of fixed cloud droplet concentration and liquid water content. Both derivations rest on the assumption that all droplet sizes allowed by the constraints have equal probability, and then consider the implications using the machinery of Boltzmann statistics.

Liu et al. (2002) proposed a generalized system theory which produces a Weibull distribution with a variable shape factor. The shape factor of the maximum likelihood distribution depends on the conserved variable, for example in the above case it is the liquid water content. Whether liquid water content, or some other variable like droplet surface area, is conserved, then becomes the central physical question. It was suggested that this scale factor may depend on the scale of averaging and the level of turbulent fluctuations Liu et al. (2002). If the averaging is below some threshold, the shape factor of the distribution should depend on the length-scale of averaging. Moreover, with decrease in turbulent fluctuations, the shape factor increases, and the distribution approaches toward a delta function corresponding to the zero fluctuation case. Connecting these general, probabilistic approaches to the diffusion-drift (Fokker–Planck) approach will be worth exploring.

4 EXPERIMENTS AND PROCEDURE FOR DATA ANALYSIS

4.1 Experimental description

Experiments are performed in the controlled environment of the Π cloud chamber. A simple schematic of the chamber is presented in the Figure 1. Here the thermodynamics and turbulent flow properties are regulated by maintaining the boundary conditions for temperature and water vapour concentration. Turbulent moist convection is driven by an imposed vertical temperature gradient, and the water-saturated top and bottom boundaries create supersaturated conditions via isobaric mixing. A steady-state cloud is produced through the continuous injection of aerosol particles (cloud condensation nuclei) into the chamber, balanced by the loss of cloud droplets by sedimentation. Technical details about the Π chamber, associated instrumentation, and typical experimental
configurations are described in more detail by Chang et al. (2016).

Conditions used in the current study are the same as the experiments reported in Chandrakar et al., 2016 (2016; 2017). The analysis of interstitial (nonactivated aerosol particles present in a cloud) and residual (aerosol samples obtained from collecting cloud droplets) aerosol distributions presented in Chandrakar et al., 2017 (2017, their figure 3) indicates the maximum supersaturation fluctuations in the chamber rarely exceed 0.5%. Moreover, the magnitude of supersaturation fluctuations (standard deviation) for these conditions is reported to be approximately 1.4% (Chandrakar et al. 2016). These measurements indicate that the mean supersaturation for the dataset used for the current study is very close to zero. Direct measurement of the absolute value of supersaturation is challenging with the currently available instrumentation.

A well-characterized aerosol distribution for injection (nearly a log-normal distribution with a mode diameter 50–70 nm and range 10–270 nm) is generated by atomizing a solution of NaCl and passing it through a diffusion dryer. The size distribution and concentration of aerosols in the injection system and the cloud interstitials are measured using a scanning mobility particle sizer (TSI). For the experiments with size-selected aerosols, the full distribution of aerosols from the atomizer is passed through a differential mobility analyzer (DMA) set to select 200 nm aerosol particles. The primary part of the distribution is in the approximate range 145–235 nm. There was a secondary mode at the larger size (≈ 235–350) associated with doubly charge particles, but their contributions to the overall concentration of aerosols were smaller than the primary mode. The primary and secondary modes referred to here correspond to the output of the Differential Mobility Analyzer (DMA) used to select monodisperse particles of size 200 nm in an ideal scenario. Aiming for a monodisperse size distribution is done in order to simplify the comparison of the experiments with theoretical distributions, as well as to isolate any effects of a wide range of aerosol sizes. In practice, the DMA produces a narrow distribution of aerosol sizes (instead of a single size) with the primary mode of the distribution at 200 nm and a secondary mode at 310 nm. The secondary mode appears due to doubly charged particles that result from the operational characteristics of a DMA.

The data used for this study were obtained when the microphysical conditions in the II chamber were in steady state, corresponding to a given aerosol injection rate. In other words, we establish the fixed thermodynamic boundary conditions and wait for transients to decay, then begin injecting aerosol particles and wait approximately an hour for microphysical properties to stabilize. For size-selected aerosols, five aerosol injection rates were used, i.e., five separate steady-state experiments carried out over a period of approximately one week (each sample is from 6–8 hr of steady-state data segments, and number of droplets used for the analysis were between 2,000 and 19,200). For the full-size-distribution aerosols, another six aerosol injection rates were used (each sample is from 6–20 hr of steady-state data segments, and the number of droplets used for the analysis were between 25,800 and 862,690).

Cloud droplet diameters are measured using a Phase-Doppler Interferometer (Dantec), and by averaging over time a size distribution can be estimated, subject to sampling uncertainty. For this study, droplet size bins below 7.5 μm are excluded during data fitting since there are indications that the Phase-Doppler system underestimates the droplet concentration at smaller sizes due to reduced signal-to-noise ratio. The minimum detectability of the instrument based on the manufacturer specifications for a configuration of transmitter and receiver optics used here is approximately 2.7 μm. However, this lower limit is for optimal conditions, whereas in the II-chamber experiments the transmitted and received signals must pass through windows, which often have some condensation present. As a conservative approach, we use this slightly higher cut-off; it is consistent with the emphasis on the large-droplet tail during the analysis, described next.

4.2 | Data analysis

The purpose of the analysis is to test whether the previously introduced functions are able to describe the measurements, to within acceptable statistical bounds. To that end, and with the capabilities of the available instrumentation in mind, the analysis is guided by the following principles:

1. Coordinates are defined so that the function is linear; i.e., as an exponential distribution on a semi-log plot.
2. Data are binned logarithmically by droplet size so that reasonably uniform counts occur within all bins.

Both of these principles serve to make the comparison to theory as direct as possible, and to place emphasis on the behaviour of the large-droplet tail of the size distribution. That is important because of the limitations in robustly measuring concentrations of small droplets. The approach also is intended to be as neutral as possible, allowing all functions to be compared to data in a way that is naturally suited to the distribution shape. The fitting functions, coordinates, and binning methods are summarized in Table 1. We discuss several more aspects of the two analysis principles in the following two paragraphs.
<table>
<thead>
<tr>
<th>Case label and equation</th>
<th>Assumptions</th>
<th>Functional form of fit</th>
<th>Binning method</th>
<th>Plotting coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D^2$ distribution; Equation (3)</td>
<td>Size-independent removal</td>
<td>$p_1(x) = A \exp(-Ax); x \equiv D^2$ (17)</td>
<td>Logarithmic bins in $D^2$</td>
<td>$D^2$ versus log ${p(x)}$</td>
</tr>
<tr>
<td>$D^3$ distribution; Equation (8)</td>
<td>Zero mean supersaturation and size-dependent removal</td>
<td>$p_2(x) = b_1 \sqrt{x} \exp(-b_2x^3); x \equiv D$ (18)</td>
<td>Logarithmic bins in $D$</td>
<td>$D^3$ versus log ${x^{-1/2}p(x)}$</td>
</tr>
<tr>
<td>$D^3$ distribution (Weibull); Equation (12)</td>
<td>System theory with the conservation of liquid water content</td>
<td>$p_3(x) = c_1 \exp(-c_2x); x \equiv D^3$ (19)</td>
<td>Logarithmic bins in $D^3$</td>
<td>$D^3$ versus log ${p(x)}$</td>
</tr>
<tr>
<td>$D^4$ distribution; Equation (10)</td>
<td>Zero supersaturation fluctuation and removal due to Stokes settling</td>
<td>$p_4(x) = d_1 \exp(-d_2x^2); x \equiv D^2$ (20)</td>
<td>Logarithmic bins in $D^2$</td>
<td>$D^4$ versus log ${p(x)}$</td>
</tr>
</tbody>
</table>

Column 1: label to be used in figure legends and the equation number from the text. Column 2: assumptions underlying the expression. Column 3: function to be fitted to the data. Column 4: method used for binning the data. Column 5: coordinates used in plotting the data. In all cases, the binning method and plotting coordinates are intended to provide a linear shape of the large-droplet tail of the distribution.
In order to express the four functions as simple exponential distributions, $D$, $D^2$, or $D^3$ coordinates are used. The coordinate transformation employed to simplify the theoretical distributions is

$$\hat{p}(x_2) = p(x_1) \left| \frac{dx_1}{dx_2} \right|. \quad (13)$$

Using this approach has a distinct advantage since the resulting distributions can be plotted as linear functions in semilog-$y$ coordinates, and therefore simplifies the fitting and evaluation processes. Significantly, the linear distributions also allow comparison of measurements with theoretical distributions, even though we are missing the smaller droplets range ($D < 7.5 \mu m$) in the measurements.

Droplet measurements are used to generate size distributions with the log binning described in Table 1. The logarithmic bins lead to a more uniform distribution of the data across bins, with the aim of maintaining relatively uniform sampling uncertainty for all data points (Barlow, 1989). The choice of binning method affects the droplet count and error in each bin, and therefore can influence the quality of fit and fitting residuals if not appropriately selected. For example, if a linear binning method is used, the extreme tails of the distribution have lower counts and higher errors compared to the distribution peak. Consequently, the distribution fitting could be biased if bin counts are too small. Given the emphasis on exponential tails in the theoretical distributions, logarithmic binning is the most natural choice. Finally, a minimum threshold of > 5 droplet counts per bin is applied for the distribution calculations.

In order to perform a goodness-of-fit test, it is necessary to have a reliable estimate of the measurement uncertainties. The uncertainties are estimated by breaking the time series of droplet size into smaller time segments and calculating the standard deviation per size bin. This method is used for estimating the uncertainties because simple counting (sampling) uncertainties represent the actual variability in the data. In the turbulent flow generated in the chamber, uncertainties in the measurements due to counting and instrumental errors are dominated by the physical variability in the system. The main driving factor for the physical variability is the turbulent convection and the associated large-scale circulation (Niedermeier et al. 2018). The time-scale associated with the large-scale circulation is the longest time-scale in the convective system. Therefore, the data sample is divided into segments of several circulation time-scales to estimate uncertainties (~ 45–60 times the large-scale circulation time-scale).

### 4.3 Weighted least-square fitting, $\chi^2$ test, and goodness-of-fit evaluation

The functions described by Equations (13)–(16) are fitted to the data using weighted least-squares (WLSQ) regression. Again, suitable coordinates are used such that the functions are linear when plotted, as summarized in Table 1. The inverse of the variances of counts for each bin are used as the weighting factors ($w_i = \sigma_i^{-2}$) for the least-squares fitting (Barlow, 1989).

The primary statistical tool we use for determining goodness of fit by the theoretical distributions to the measurements is the $\chi^2$ test. In this test, $\chi^2 \equiv \sum w_i(Y_i - \hat{Y}_i)^2$, i.e., the sum of squared error (SSE) and its probability is estimated for a fit. Here $Y_i$ represents a measurement value, $\hat{Y}_i$ is the corresponding value from the fit, and again, $w_i$ is the weighting factor. The expectation value (probability = 0.5) of the $\chi^2$ distribution is equal to the degrees of freedom (DOF) of a fit, where the DOF is the number of data points minus the number of free parameters of the fit (Barlow, 1989). If a fit produces $\chi^2 \gg$ DOF or the $\chi^2$ probability ($p_{\chi^2}$) is significantly lower than the desirable cut-off, the fit is considered as failing to describe the measurement. The SSE can be used for hypothesis testing (i.e., pass or fail), but it can also be considered as a fitting statistic specifying how close the fitted function values are to the dataset. In this work, we use it as a goodness-of-fit parameter.

We also apply residual analysis to obtain additional information about the goodness of a fit, and how it varies between different candidate fitting-functions. Unlike hypothesis testing or evaluations based on a single goodness-of-fit parameter (e.g., $\chi^2$), residual analysis accounts for whether the measurement points are above or below the fitted function. In our application, therefore, the fit residuals are plotted versus the droplet size bin. Based on the trend of the residual plot, relative appropriateness of a fit can be judged (Hughes and Hase, 2010). For example, in an ideal case, i.e., for a good fit, the fit residuals should be distributed randomly, being equally likely to be positive or negative, and their magnitudes should be small. However, if they have a monotonic (increasing or decreasing) trend, it means that there is a consistent deviation between the measurements and the fitted function. In our case, because the functions are linear and are fitted to the data, only a nonlinearity of the data would result in a consistent pattern (a ‘run’) of residuals with droplet size. A consistent change in the sign of residuals is easily interpreted as indicative of a poor fit. In this study, all three methods are used to evaluate the theoretical functions described in Table 1: the $\chi^2$ test, the relative comparison of the goodness-of-fit parameter (SSE), and the residual analysis.
5 | RESULTS

As indicated earlier, two different sets of experiments are compared to the four distribution functions: experiments with input of monodisperse, size-selected aerosol and experiments with a broad, polydisperse distribution of aerosol. The DSDs from both of the experiment sets, the WLSQ fitting results, and the statistical analysis as described in the previous section are presented next.

5.1 | Experiments with monodisperse aerosols

We begin with the analysis of cloud DSDs formed on size-selected aerosols (injection of a narrow, monodisperse distribution of aerosols with a mode diameter of around 200 nm). Five different aerosol concentration settings are used to achieve progressively increasing cloud droplet concentrations, corresponding to extremely clean to moderately clean cloud conditions (from about 10 to 72 cm$^{-3}$). The DSDs (blue circles with measured uncertainties) and fitted functions (red lines) are presented in Figures 2–5. Underneath each size-distribution plot is a plot of the corresponding fit residues. The blue circles are the residuals (difference between the measurement and the fitted value for each size bin), with dashed lines to aid in detecting trends, and the red line indicates a zero-residual, perfect fit to the data. The panels are arranged from top-left to bottom-right in order of increasing droplet number concentration.

Figure 2 shows the fitting of the $D^2$ function, Equation (13). This solution is for stochastic condensation with a constant (size-independent) droplet removal rate. The fits all appear to be reasonable, as judged by eye. The residual plots, however, show a modest but consistent trend with a concave-down shape, indicating that the theoretical function falls off somewhat too weakly in the large-droplet tail compared to the measurements.

Figure 3 shows the fitting of the $D^3$ function, Equation (14), corresponding to stochastic condensation with size-dependent droplet removal. Again, the fits appear quite reasonable by eye. The residual plots are also encouraging, showing no obvious trends, i.e., with residuals randomly distributed about the zero-line, and also with smaller magnitudes (mostly between ±0.1) than in the previous case.

Figure 4 shows the fitting of the $D^3$ Weibull distribution, Equation (15), which is obtained from probabilistic approaches based on assumption of fixed concentration and total volume. The fits are acceptable, given the uncertainties, as confirmed by the $\chi^2$ test presented later, but in this case it is evident even to the eye that, in the plotting coordinates that result in a linear theoretical function, the measurement points have a concave-up shape. This is readily confirmed in the residual plots, also showing a distinct, concave-up shape consistent for all cloud droplet concentrations. It should also be noted that the range of the residuals is considerably larger than in the previous two cases.

Figure 5 shows the fitting of the $D^4$ distribution, Equation (16), corresponding to growth of cloud droplets with fixed supersaturation and size-dependent removal. The situation here is similar to the $D^3$ Weibull distribution just discussed, with clearly identified nonlinear shape of the measurements in the $D^4$ coordinates. The theoretical function falls off too rapidly with increasing size compared to the measurements. As in the previous case, the residual plots confirm this, all showing a distinct, concave-up pattern. The range of residuals is largest in this case.

It is worth noting that there is no clear trend in the relative distribution shape between different aerosol injection rates, apart from the change in the distribution width. Based on the visual inspection and residual analysis made thus far, the $D^2$ (Equation 13) and $D^3$ (Equation 14) distributions appear to be most appropriate. That they can capture the size distribution shape over an order of magnitude change in droplet number concentration seems quite remarkable to us.

In order to make an additional, quantitative comparison between the suitability of the four functions for describing the measurements, we use the $\chi^2$ test. Figure 6 shows the sum of squared errors of prediction (SSE or $\chi^2$) for all functional fits and for the five different cloud droplet concentrations. In most of the cases (except, the fourth case where only the $D^2$ distribution and $D^3$ distribution for size-dependent removal pass), all distribution fits pass the $\chi^2$ test since the $\chi^2$ values are close to the DOF. By pass, we mean that SSE is less than DOF. Here it is worth recalling that SSE greater than DOF implies a poor fit, and low likelihood of the function describing the measurements. However, the opposite extreme, with SSE much less than DOF, implies that the fit is too good to be true, given the estimated uncertainties; this is usually an indication that the uncertainties of the data are overestimated. In this case, in four out of the five cases, all theoretical models technically pass the fit test. However, when we compare the relative SSE values as a goodness-of-fit parameter, the $D^3$ distribution is consistently the lowest and therefore the optimal fit. In four of the five cases, this is followed by the $D^2$ distribution. These quantitative conclusions are consistent with those reached from the residual analysis earlier in the section.

Furthermore, the residual plots suggest a Weibull distribution with a shape factor between 2 and 3 would fit the large-droplet tail of the distribution well. The
FIGURE 2  Weighted least-square fitting of the $D^2$ distribution, Equation (13), to the measured cloud droplet size distributions, for size-selected aerosol. The five panels (a) to (e) correspond to increasing aerosol injection rate and cloud droplet concentration. (f) to (j) show the corresponding residuals. The binning method and coordinates are defined in Table 1 [Colour figure can be viewed at wileyonlinelibrary.com].
FIGURE 3  As Figure 2 but for the $D^3$ distribution, Equation (14) [Colour figure can be viewed at wileyonlinelibrary.com].
FIGURE 4  As Figure 2 but for the $D^3$ Weibull distribution, Equation (15) [Colour figure can be viewed at wileyonlinelibrary.com].
FIGURE 5  As Figure 2 but for the $D^4$ distribution, Equation (16) [Colour figure can be viewed at wileyonlinelibrary.com].
residuals in Figure 2, corresponding to Weibull with shape factor 2, were concave-down, and the residuals in Figure 4, corresponding to Weibull with shape factor 3, were concave-up. The exponential in the $D^4$ distribution certainly decays too sharply. The $D^3$ distribution function corresponding to stochastic condensation with size-dependent removal apparently is the best fit: it has an additional $\sqrt{D}$ multiplied with the $D^3$-exponential, resulting in a tail that decays at a slightly slower rate than the $D^2$ Weibull distribution, but sharper than the $D^2$ Weibull distribution.

5.2 | Experiments with polydisperse aerosols

Experiments with a broad, polydisperse size distribution of aerosol input were carried out for three reasons: first, to allow larger cloud droplet concentrations to be achieved, thereby allowing more polluted cloud conditions to be explored, second, to obtain lower counting uncertainties as a result of the higher concentrations, thereby placing stronger constraints on the theoretical fits, and third, to explore any possible dependence of the cloud DSD on the shape of aerosol size distribution. In the six steady-state experiments, cloud droplet concentration is varied over two orders of magnitude, from very clean to very polluted (from about 21 to 1,790 cm$^{-3}$) conditions. The observed droplet size distributions and the four functional fits are summarized in Figures 7–10.

Figure 7 shows the fitting of the $D^2$ function, Equation (13). This solution is for stochastic condensation with a constant (size-independent) droplet removal rate. The fits at the lowest concentration (a) and at the fourth level of concentration (d) appear reasonable. Intriguingly, however, the two moderate-concentration cases in (b) and (c) show distinctly weaker tails in the measurements than in the theory, whereas the high-concentration cases in (e) and (f) show much stronger tails. The residual plots show that even the two fits that seemed reasonable do, in fact, display strong biases, with concave-down shape similar to the two moderate-concentration cases. The two high-concentration cases show a flip in the shape of the residual curve and display very high magnitudes indicative of a poor fit.

Figure 8 shows the fitting of the $D^3$ function, Equation (14), corresponding to stochastic condensation with size-dependent droplet removal. (a) to (d) show reasonable fits. The goodness-of-fit in (a), corresponding to the lowest droplet concentration, is further borne out in the residual plot with small magnitudes. The next three droplet concentration levels (b,c,d) show deviations in the tails indicated by the distinct, curved shape of the residual patterns. The residual patterns flip from concave-down to concave-up for these three panels, but the magnitudes of the residuals are reasonably small compared to other fits to the polydisperse-aerosol cases. The two highest droplet concentration cases (e,f) are clearly poor fits, with the measurements displaying pronounced large-droplet tails that are not captured by the theoretical distributions.

Figure 9 shows the fitting of the $D^3$ Weibull distribution, Equation (15), which is obtained from probabilistic approaches based on assumption of fixed concentration and total volume. Most of the fits can be seen to deviate from the measurements. (a) to (d), corresponding to the lowest droplet concentrations, show obvious, concave-up residual patterns, but the magnitudes are generally below 1. Once again, the two cases with highest droplet concentration (e,f) show obvious deviations in the tails, with the measurements displaying enhanced large-droplet concentrations.

Figure 10 shows the fitting of the $D^4$ distribution, Equation (16), corresponding to growth of cloud droplets with fixed supersaturation and size-dependent removal. The situation here is again similar to the $D^3$ Weibull distribution just discussed, with clearly identified nonlinear shape of the measurements in the $D^4$ coordinates. And once again, the deviations are more pronounced with increasing cloud droplet concentration: the theoretical function falls off too rapidly with increasing size.
FIGURE 7  Weighted least-square fitting of the $D^2$ distribution, Equation (13), to the measured cloud droplet size distributions, for polydisperse aerosol. (a) to (f) correspond to increasing aerosol injection rate and cloud droplet concentration, and (g) to (l) show the corresponding residuals. The binning method and coordinates are defined in Table 1 [Colour figure can be viewed at wileyonlinelibrary.com].
FIGURE 8  As Figure 7, but for the $D^3$ distribution, Equation (14) [Colour figure can be viewed at wileyonlinelibrary.com].
FIGURE 9  As Figure 7, but for the $D^3$ Weibull distribution, Equation (15) [Colour figure can be viewed at wileyonlinelibrary.com].
FIGURE 10  As Figure 7, but for the $D^4$ distribution, Equation (16) [Colour figure can be viewed at wileyonlinelibrary.com].
than in the measurements. As in the previous case, the residual plots confirm this, all showing a distinct, concave-up pattern. The range of residuals is largest in this case.

Figure 11 shows the SSE, normalized by the DOF, for all the fittings of the distributions from the different aerosol cases. In a relative sense, the results from the polydisperse aerosol cases are similar to the monodisperse aerosol experiments: in these cases as well, the \( D^3 \) and \( D^2 \) distributions are the most successful in describing the measurements. In the four cases with lowest cloud droplet concentration the \( D^3 \) distribution has the lowest SSE/DOF, whereas in the two most polluted cases the \( D^2 \) distribution is the best fit. Except for the last two cases, the \( D^3 \) distribution passes (or nearly passes) the \( \chi^2 \) test. The \( D^2 \) distribution passes in the lowest-concentration case. The \( D^3 \)-Weibull and \( D^4 \) distributions are reasonably close to passing the \( \chi^2 \) for the four lowest-concentration cases. However, it should be recalled that for most of these borderline \( \chi^2 \) test results, it is indeed true that there were significant deviations in the tail of the distribution for the measurements compared to the fit.

For the last two polluted cloud conditions, the \( \chi^2 \) test clearly rejects all distributions, with SSE/DOF values much greater than unity. This is consistent with what already was evident from visual inspection of the fits: all four theoretical distributions fail to faithfully capture the shape of the measured distribution in very polluted conditions. Unlike the size-selected, monodisperse aerosol cases, there is an evident trend in the deviation of the distribution shapes relative to the measurements as aerosol concentration is increased under polydisperse conditions.

6 | DISCUSSION

The purpose of this study was to evaluate a drift-diffusion framework for stochastic condensation that leads to a Fokker–Planck equation for the cloud DSD (Equation (2)). It can be solved to obtain analytical size distribution shapes under certain simplifying assumptions. Three solutions are explored: stochastic condensation with a constant cloud droplet removal rate (Saito et al. 2019), stochastic condensation with a removal rate proportional to the square of the particle diameter (relevant for settling with Stokes drag), and droplet growth in a fixed supersaturation field, with a similar diameter-squared removal rate (Krueger, 2019). In addition, a Weibull distribution with shape factor of 3 can be obtained from the Fokker–Planck equation with suitable assumptions for the drift or diffusion terms, and also results from mechanism-free probabilistic arguments (Williams and Wojtowicz, 1982; Liu et al. 2002). The functions are summarized in Table 1.

The \( \Pi \) chamber provides an ideal environment in which to test idealized distributions because the number of physical processes contributing to cloud droplet evolution is relatively small compared to the atmosphere: activation on cloud condensation nuclei, growth due to condensation in a fluctuating environment, and eventual removal due to settling or turbulent transport. The \( \Pi \) chamber configuration used here allows the boundary and aerosol conditions to be well characterized and steady in time. In addition, because the experiments are carried out in steady state, microphysics can be measured over long durations, allowing large-droplet tails to be well resolved.

Comparison of the SSE/DOF values along with analysis of fitting residuals suggests that the \( D^3 \) distribution corresponding to stochastic condensation with size-dependent droplet removal most successfully fits the observations. For these conditions, we can confidently state that the drift-diffusion approach and resulting Fokker–Planck equation are able to capture the behaviour of the cloud DSD over the parameter range explored. Application of the \( \chi^2 \) test by itself suggests that all four theoretical distributions can plausibly describe the observations under conditions of monodisperse aerosol injection. However, a careful comparison of the goodness-of-fit parameter (SSE/DOF) and residual analysis favours the \( D^3 \)
distribution for the size-dependent removal followed by the $D^3$ Weibull distribution. The $D^3$ Weibull distribution emerges from the assumption of uniform liquid water content. During any particular experiment with steady-state cloud conditions within the Π chamber, the liquid water content is statistically steady. However, there are fluctuations around the mean due to flow structure and dynamics and therefore likely they will lead to additional stochasticity in the system. This variability in liquid water content might be another source of deviation of the theoretical distribution and corresponding higher residuals and SSE/DOF.

In the experiments with polydisperse aerosol injection, the situation is more complex. Again, the $D^3$ distribution is most successful in fitting the measurements, except for the two experiments with highest aerosol injection rate, for which none of the functions are deemed acceptable by the $\chi^2$ test. In those cases the large-droplet tail of the measured size distributions far exceeds what can be described by any of the distribution functions. This suggests that there is missing physics in the governing equation. We can speculate on two causes. The first is the possibility that activation and deactivation of aerosol particles of different size with different efficiency may be the cause of the pronounced growth of the large-droplet tail. It stands to reason that this would be most effective when cloud droplet concentrations are high, with plentiful interstitial aerosol, all residing in a low mean supersaturated environment with turbulent fluctuations. Overall, this type of broadening is related to the Ostwald ripening effect (Korolev, 1995; Wood et al. 2002; Arabas and Shima, 2017; Jensen and Nugent, 2017; Yang et al. 2018). During polluted cloud conditions, the average supersaturation is expected to be nearly zero, but the presence of turbulent flow condition allows the supersaturation to fluctuate above and below the mean value. Therefore, there is a possibility of multiple activation–deactivation cycles. Due to the irreversibility of the droplet growth with the curvature and solute effects, size distribution broadening results. Obtaining analytical solutions when including the curvature and solute effects in the Fokker–Planck equation will remain a challenge for future work. The second possible explanation for the unaccounted broadening of the DSD in the most polluted cloud condition would be the effect of a highly skewed supersaturation distribution. In polluted cloud conditions the mean supersaturation in the chamber should be very close to zero. Although the phase relaxation time is expected to be very short compared to mixing times, there could be localized growth and activation at boundaries before complete mixing of the scalar fields. The possibility of strongly non-Gaussian behaviour in the supersaturation statistics is not accounted for in the stochastic condensation growth model.

Rejecting or accepting theoretical expressions rests heavily on the quality and uncertainty of the measured large-droplet tails. Higher statistical confidence than obtained here will require further reducing statistical uncertainties, mainly by measuring farther into the large-droplet tail. For a given cloud droplet concentration, that implies measuring for a longer time (or using an instrument with a larger sample volume). Longer measurement time places more severe constraints on the ability to maintain steady thermodynamic conditions in the Π chamber; most notably, drifts in the water-vapour boundary conditions over the period of days have been noted. Relying on rare droplet counts also places high demands on the ability of the droplet measurement to reject noise and ensure high probability of detection. Finally, reliable measurement ability in the small-droplet part of the distribution is sure to help substantially.

Lastly, we briefly consider implications for clouds in the atmosphere. The laboratory measurements favour the $D^3$ and $D^2$ distribution resulting from stochastic condensation, either with size-dependent removal or constant removal rate. However, when polydisperse aerosols were used, there was some deviation near the tail of the distribution possibly due to the significant effect of Ostwald ripening and associated deactivation and activation. In non-precipitating atmospheric clouds, the removal due to the gravitational settling is usually considered to be negligible. Therefore, the $D^2$ Weibull distribution would be a more appropriate form of DSD in such clouds unless there are other dominant and complicating processes like droplet collision–coalescence or entrainment effect near the cloud top. The stratocumulus cloud core before the start of significant drizzle formation would be an ideal location for applicability of this model. Exploring under what conditions the mean versus fluctuating supersaturation will be of primary importance, and when other factors such as Ostwald ripening will become significant, will be of great interest for future exploration.

ACKNOWLEDGEMENTS
This work was supported by National Science Foundation Grant AGS-1754244 and by NASA Headquarters under the NASA Earth and Space Science Fellowship Program grant 80NSSC17K0449. I.S. and T.G. were supported by MEXT KAKENHI grant 15H02218. I.S. was supported by JSPS KAKENHI grant 18K13611. F.Y. was supported by the US Department of Energy (DOE) under contract DE-SC0012704

ORCID
Kamal Kant Chandrakar https://orcid.org/0000-0003-2970-3458
REFERENCES


