



## RESEARCH ARTICLE

10.1029/2023JD038902

### Key Points:

- Statistical analysis of disdrometer data and cloud resolving model simulations reveal the same primary modes of rain drop size distribution variability
- The statistical analysis reveals six precipitation groups which are studied from the model for links to processes
- The statistical framework is applied to a case study from Mid-latitude Continental Clouds and Convection Experiment demonstrating the mutual benefit of this analysis technique

### Correspondence to:

B. Dolan,  
[bdolan@colostate.edu](mailto:bdolan@colostate.edu)

### Citation:

Dolan, B., Saleeby, S. M., Rutledge, S. A., van den Heever, S. C., & Van Valkenburg, K. (2023). A statistical framework for evaluating rain microphysics in model simulations and disdrometer observations. *Journal of Geophysical Research: Atmospheres*, 128, e2023JD038902. <https://doi.org/10.1029/2023JD038902>

Received 16 MAR 2023

Accepted 24 AUG 2023

### Author Contributions:

**Conceptualization:** Brenda Dolan, Susan C. van den Heever  
**Data curation:** Brenda Dolan  
**Formal analysis:** Stephen M. Saleeby  
**Funding acquisition:** Brenda Dolan  
**Investigation:** Kristen Van Valkenburg  
**Methodology:** Brenda Dolan  
**Resources:** Susan C. van den Heever  
**Software:** Stephen M. Saleeby  
**Supervision:** Steven A. Rutledge, Susan C. van den Heever  
**Writing – original draft:** Brenda Dolan  
**Writing – review & editing:** Stephen M. Saleeby, Steven A. Rutledge, Susan C. van den Heever

© 2023. The Authors.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

## A Statistical Framework for Evaluating Rain Microphysics in Model Simulations and Disdrometer Observations

Brenda Dolan<sup>1</sup> , Stephen M. Saleeby<sup>1</sup> , Steven A. Rutledge<sup>1</sup> , Susan C. van den Heever<sup>1</sup> , and Kristen Van Valkenburg<sup>1</sup>

<sup>1</sup>Colorado State University, Fort Collins, CO, USA

**Abstract** Statistical analyses of a large disdrometer data set and a diverse set of model simulations for convection using the Regional Atmospheric Modeling System were conducted, with the mutual goal of providing insights into precipitation formation and microphysical processes. We demonstrate that a two-moment bulk microphysical model successfully captures the dominant observed modes of variability in rainfall related to rainfall intensity and raindrop size distributions. The model reproduced the general distribution of observed precipitation groups (PGs) derived from Principal Component Analysis. The multi-variable analysis also uncovered some shortcomings in the model as well as limitations of the disdrometer data. The model solutions were constrained in their predicted drop size distributions (DSDs) due to the fixed DSD parameters assumed in a two-moment microphysics scheme. A case study from the Mid-latitude Continental Clouds and Convection Experiment field project demonstrated how model results can be used to contextualize the disdrometer observations which are limited in sample size, spatial coherence, and detection of small drops and low drop concentrations. The case study also showed that the spatial patterns of the statistically derived PGs revealed by the model are consistent with the hypothesized microphysical processes that determine surface rain DSDs. This work demonstrates how leveraging the strengths of observations and models together can improve our understanding and representation of rain microphysical processes.

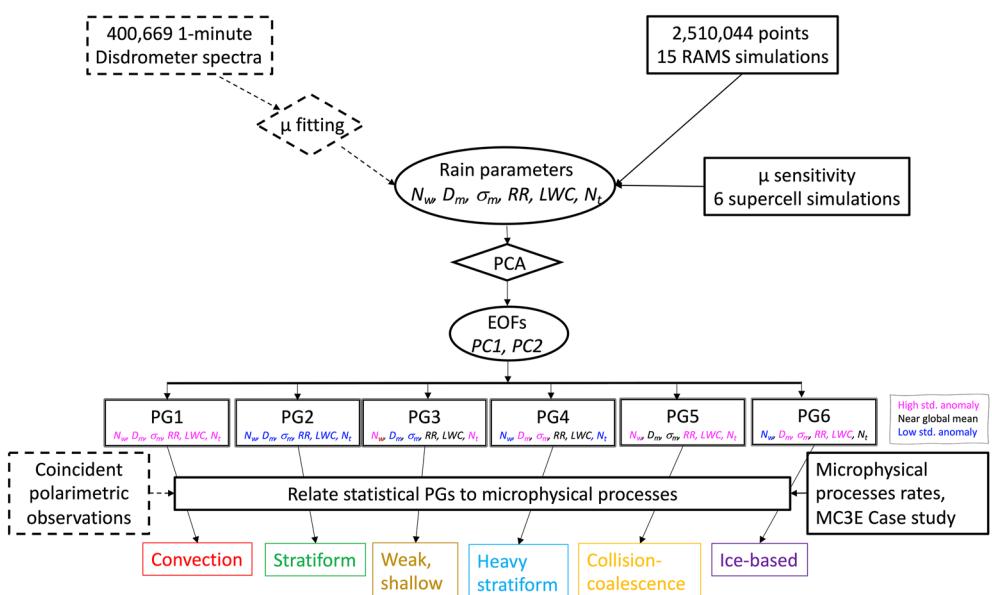
**Plain Language Summary** Understanding precipitation processes is key to the Earth's water cycle and regional and climate modeling efforts. Herein we demonstrate how multi-parameter statistical analyses can be used to evaluate model representations of precipitation variability, and how the model can subsequently provide information about types of processes resulting in precipitation at the surface.

### 1. Introduction

Understanding microphysical processes in warm, mixed, and cold clouds, and how these processes impact both surface precipitation and storm dynamics, is critical to validating model parameterizations in cloud resolving models (CRMs), and ultimately in global climate models. However, microphysical processes (accretion, melting, etc.) are not directly observed, and must be inferred from measured quantities (e.g., reflectivity, updraft speed) (Kumjian & Ryzhkov, 2008). Relating observations to the dominant microphysical processes contributing to surface precipitation not only provides a way to study the variability of such processes, but also a means for validating microphysical schemes implemented in models. While bringing observations and models together toward these mutual goals is critical, it remains onerous due to differences in spatial scales, temporal scales, and sample sizes, as well as contrasts in observed quantities (e.g., reflectivity, size distributions) and those routinely output by models (e.g., mixing ratios, 3D wind fields).

Model validation is most often undertaken as a comparison between model output and observations, regarding comparing radar reflectivity, total precipitation, vertical winds, and other quantities, most typically on a case-by-case basis (e.g., Iguchi et al., 2012; Li et al., 2009; Marinescu et al., 2020, 2016; Varble et al., 2014). However, these types of comparisons often do not contain information about the accuracy of various model parameterizations and do not lend themselves to validation of idealized simulations without coincident observations. Thus, a thorough and robust evaluation of model microphysics with observations across a wide range of convective regimes would be valuable.

Dolan et al. (2018; henceforth D18) applied Principal Component Analysis (PCA) to a worldwide disdrometer data set to examine rain drop size distribution (DSD) variability (Figure 1). This framework objectively identified the co-variability of six parameters related to rainfall. The findings revealed that the diverse set of observations



**Figure 1.** Flowchart depicting the framework for statistically relating inputs of 15 Regional Atmospheric Modeling System (RAMS) simulations (this study) and disdrometer spectra (D18) to microphysical processes through precipitation groups (PGs). The analysis focused in D18 is indicated by dashed boxes, while that of the present study is represented by the solid boxes and arrows in the upper half of the flows. The Principal Component Analysis (PCA) derived principal components (PCs) are used to determine six PGs with common co-variability of the six rain parameters. The color of variable within each PG box illustrates how the values in each group compare to the global means, with pink denoting significantly higher values, blue lower than average, and black with values near the global mean.

consistently had predictable modes of variability, which were objectively divided into six precipitation groups (PGs). Using co-located polarimetric radar analysis, these six PGs were hypothesized to be related to fundamental precipitation processes, such as convective and stratiform precipitation, and warm rain and ice-based precipitation mechanisms. However, more quantitative understanding of the microphysical processes contributing to these PGs can be gained by exploring the PGs with an observationally verified model that provides information on microphysical process rates, the conduits for transferring mass between hydrometeor categories.

The D18 objective analysis formulation provides a robust opportunity to intercompare models and observations. For model validation, rather than comparing a single parameter which may or may not be directly observed or predicted by the model, the multi-parameter space can assess if the model is able to capture the same *variability* evident in the observations. Importantly, by characterizing the co-variability of precipitation parameters, we can better understand processes, and statistically relate observations to models without having to use direct case study comparisons. For example, Saleeby et al. (2022b, henceforth S22) used a global set of disdrometer observations and a diverse set of model simulations from a variety of convective regimes using the Colorado State University Regional Atmospheric Modeling System (RAMS) and Weather Research and Forecasting (WRF) Model version 3.9.1 (Skamarock, 2006) to investigate the raindrop breakup parameterization. By comparing the diverse data sets from observations and models in the 2D  $N_w$ - $D_0$  parameter space described in D18, it was found that there was an abundance of drops occurring at a mean drop size of 1.1 mm, which is the diameter at which the raindrop self-collection efficiency curve becomes negative and breakup is most active.

Herein we build on the D18 framework using a diverse set of model simulations (S22) to evaluate the model-based rain multi-parameter variability and explore the statistical PGs within the context of the model, and finally apply the framework to a case study from the Mid-latitude Continental Clouds and Convection Experiment (MC3E) in order to evaluate both the model and disdrometer data (Figure 1). First, we outline the PCA and introduce the diverse disdrometer and model data sets used in the analysis. Next, we analyze the precipitation variability as determined by the observations and the model data sets, including exploring the impact of a fixed DSD shape parameter assumption in the model. We then leverage the framework to explore microphysical processes contributing to the PGs and conclude by applying the framework to a case study of a squall line from MC3E (Jensen et al., 2015).

## 2. Methods

### 2.1. Disdrometer Observations

In this study we use disdrometers as the primary source of rainfall observations. A variety of disdrometers spanning a diverse set of environments supporting different types of convection have been deployed across the world as parts of long-term installations or for field campaigns. The disdrometer data set used in this analysis includes many of the same mid- and low-latitude data from D18 (Table 1; see D18 for more details), with additional data from three recent Department of Energy (DOE) deployments: the East North Atlantic (ENA) site in the Azores (D. Wang & Bartholomew, 2014; J. Wang et al., 2022); the Clouds, Aerosol, and Complex Terrain Interactions (CACTI, Varble et al., 2021; D. Wang & Bartholomew, 2018) in Cordoba, Argentina; and the Green Ocean Amazon (GoAmazon; Martin et al., 2017) in Manaus, Brazil. Several NASA data sets are also included from the HYdrological cycle in Mediterranean EXperiment (HYMEX; Drobinski et al., 2014; Ducrocq et al., 2014) in France and Italy, and six disdrometers deployed in Huntsville, Alabama (Petersen et al., 2010; Thurai et al., 2011). Finally, two shipborne experiments, Propagation of Intra-Seasonal Tropical Oscillations (PISTON, Chudler et al., 2022; Sobel et al., 2021) and Salinity Processes in the Upper-ocean Regional Study (SPURS-2, Lindstrom et al., 2019; Rutledge et al., 2019) provide valuable open-ocean disdrometer measurements (Klepp et al., 2018). In total, these observations (OBS) encompass over 400,000 raining minutes. High latitude locations used in D18 were purposefully omitted from this analysis.

We note here that we are using a heterogeneous disdrometer data set, with three distinct instruments, each with different detection capabilities. Several studies have analyzed and compared the different platforms and found that the fits and integrated parameters are similar (Thurai et al., 2011; Tokay et al., 2001), although the minimum size detected by the disdrometers can have a larger impact on the derived parameters than missing the larger drops which are infrequent in the disdrometer data sets (Tokay et al., 2013). The disdrometers have different reliable minimum drop diameters of 0.2 mm for the 2DVD and 0.3 mm for the JWD (Tokay et al., 2013), 0.36–0.43 mm for the ODM (Klepp, 2015; Protat et al., 2019), and 0.2–0.25 mm for the Parsivels (Tokay et al., 2014). All disdrometer data were quality controlled in a similar manner by applying a minimum threshold of 100 drops and a minimum rain rate (RR) of 0.05 mm hr<sup>-1</sup> over a 1-min period (Thompson et al., 2015). Liquid water content (LWC) and RR are calculated from the distribution, and the data are fit to a gamma distribution (Thurai et al., 2014) to derive the gamma distribution parameters.

The OBS are broken into four geographic regions (Figure 2), representing different types of convection and environments leading to different precipitation processes. For example, oceanic convection tends to occur in environments comprised of low aerosol concentrations and very moist air which leads to deep warm-cloud depths and prolific warm-rain processes, while the continents are drier and more polluted in comparison, with robust mixed-phase and ice processes. D18 found variability in the PGs as a function of latitude, with the tropics (mid-latitudes) having the most warm-rain (ice) contributions. These four geographic regimes are continental mid-latitude, oceanic mid-latitude, tropical continent, and tropical ocean. Most deployments of ground instruments included herein are in the mid-latitude continents (primarily over the US), and the fewest are over the mid-latitude oceans (Figure 2).

### 2.2. RAMS Simulations

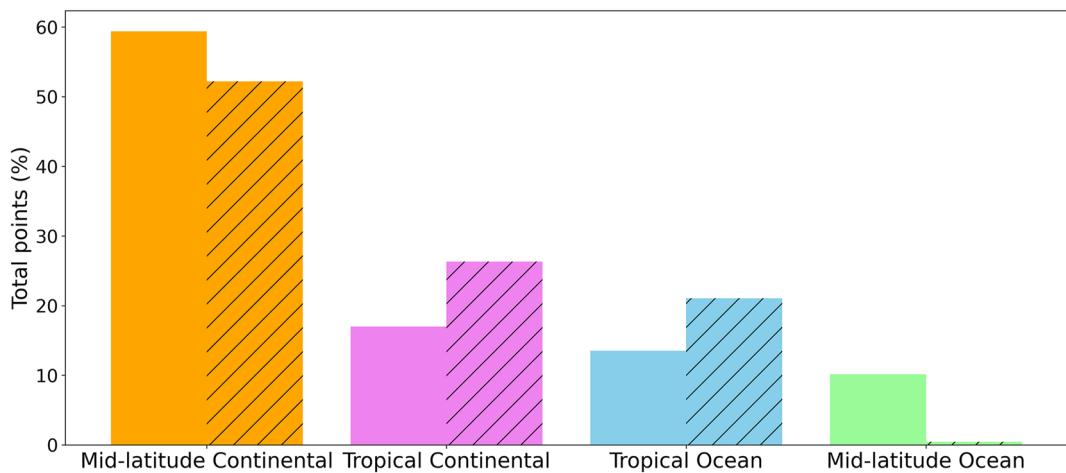
RAMS is a non-hydrostatic, fully compressible research and weather prediction model run on an Arakawa-C grid with a  $\sigma_z$  terrain following-vertical coordinate system (Cotton et al., 2003). RAMS has demonstrated success in simulating a range of cloud systems, such as supercell thunderstorms (e.g., Freeman et al., 2019; Grant & van den Heever, 2014b; Grasso, 2000), tropical and midlatitude mesoscale convective systems (e.g., Marinescu et al., 2016; McGee & van den Heever, 2014; Olsson & Cotton, 1997), tropical maritime systems (e.g., Riley-Dellaripa et al., 2018; Toms et al., 2020), maritime congestus clouds (Leung & van den Heever, 2022; Sheffield et al., 2015), and shallow, warm rain clouds (e.g., Cheng et al., 2009; Jiang & Feingold, 2006; Saleeby et al., 2015). For more details on the RAMS model and microphysics packages, please see Meyers et al. (1997), Saleeby and Cotton (2004), Saleeby and van den Heever (2013), and Walko et al. (1995). Each hydrometeor species is represented by a generalized gamma distribution with an a priori selected constant distribution shape parameter ( $\mu$ ).

For the PCA, it is vital to have a diverse set of precipitation processes represented in the model simulations. Additionally, for comparison with OBS, it is important to capture a similar range of convective environments

**Table 1**  
*Locations of Disdrometer Data Sets*

Name	Long name	Region	Location	Time frame	#	#/Type	References
iFloods	Iowa Flood Studies	CM	Eastern Iowa	6 Apr–6 Jun 2013	14,608	6 DVD	Petersen and Gatlin (2013) and Seo et al. (2015)
MC3E	Midlatitude Continental Clouds and Convection Experiment	CM	Central Oklahoma	23 Apr–1 Jun 2011	6,043	5 DVD	Jensen et al. (2015) and Petersen et al. (2011)
SGP	Southern Great Plains	CM	Central Oklahoma	28 Feb 2011–5 May 2016	39,592	12 DVD	D. Wang and Bartholomew (2011c)
IPHEX	Integrated Precipitation and Hydrology Experiment	CM	Western North Carolina	25 Apr–16 Jun 2014	10,718	5 DVD	Barros et al. (2014) and Petersen and Gatlin (2014)
Wallops	Wallop Island	CM	Eastern Virginia	2013, 9 Aug 2014–15 Oct 2015	42,238, 74,520	6 DVD, 5 DVD	Petersen et al. (2019)
HYMEX*	Hydrological cycle in Mediterranean Experiment	CM	Alés, France; Rome, Italy	30 Sept–12 Nov 2012	750, 2,635	2DVD	Drobinski et al. (2014), Ducrocq et al. (2014), and Petersen et al. (2016)
CACTI*	Clouds, Aerosol, and Complex Terrain Interactions	CM	Cordoba, Argentina	26 Sept 2018–23 Feb 2019	7,496, 11,056	2DVD, Parsivel2	VanBle et al. (2021) and D. Wang and Bartholomew (2018)
Huntsville*	—	CM	Huntsville, AL	19 Dec 2009–14 Oct 2011	28,775	3 DVD	Petersen et al. (2010) and Thurairaj et al. (2011)
ENAs*	Eastern North Atlantic	OM	Graciosa Island, Azores	31 Oct 2014–19 Jan 2018	37,522	12 DVD	D. Wang and Bartholomew 2018 and J. Wang et al. (2022)
Manus	Manus Island	OT	Manus Island	01 Dec 2011–04 Jul 2014	45,664	12 DVD	D. Wang and Bartholomew (2011d)
Gan Is.	ARM MJO Investigation Experiment	OT	Gan Island, Maldives	05 Oct 2011–06 Feb 2012	4,348	12 DVD	D. Wang and Bartholomew (2011b) and Yoneyama et al. (2013)
PISTON*	Propagation of Intra-Seasonal Tropical Oscillations	OT	Ship, West Pacific	15 Sept–10 Oct 2018	1,365	ODM	Klepp (2015), Klepp et al. (2018), and Sobel et al. (2021)
SPURS*	Salinity Processes in the Upper Ocean Regional Study 2	OT	Ship, East Pacific	20 Oct–17 Nov 2017	10,480	ODM	Klepp (2015), Klepp et al. (2018), and Lindstrom et al. (2019)
TWPICE	Tropical West Pacific—International Cloud Experiment	CT	Darwin, Australia	3 Nov 2005–10 Feb 2006	3,669	1 JWD	May et al. (2008)
Darwin	Darwin, Australia	CT	Darwin, Australia	1 Mar 2011–04 Jan 2015	42,248	12 DVD	D. Wang and Bartholomew (2011a)
GoAmazon*	Green Ocean Amazon Experiment	CT	Manacapuru, Brazil	24 Sept 2014–01 Dec 2015	16,942	Parsivel2	D. Wang et al. (2014)
Global	—	—	—	—	400,669	—	—

Note. CM, midlatitude continental; CT, tropical continental; OM, midlatitude oceanic; OT, tropical oceanic. The \* denotes data sets that were not included in D18.



**Figure 2.** Total frequency of points in each of the four regimes, with observations as solid bars and Regional Atmospheric Modeling System (RAMS) simulations as hatched bars. See Table 1 and S22 for more information about the data sets.

and types that are sampled by the disdrometers. To this end, 14 RAMS simulations that have been previously reported in the literature were selected for analysis (S22, their Table 1). The simulations broadly correspond to the four geospatial regimes represented within the disdrometer data set. The mid-latitude continental regime is represented by the 20 May squall line from MC3E (Marinescu et al., 2016), and an idealized supercell (Grant & van den Heever, 2014b). These simulations have deep and complicated ice processes, including strong convective motions and stratiform anvil development. Warm-rain only idealized simulations initialized using ATEX (Atlantic Trade Wind Experiment) conditions (Saleeby et al., 2015) result in shallow stratocumulus typical of the mid-latitude ocean. Select active and inactive phases of a 40-day long simulation of the boreal summer intra-seasonal oscillation (BSISO) in the Indo-Pacific include the range of convection from shallow warm rain to deep organized systems ubiquitous to the tropical ocean (Toms et al., 2020). Additionally, simulations initialized with a moist tropical Dunion and Velden (2004) sounding with different aerosol perturbations capture different aerosol effects in tropical deep convection (Saleeby et al., 2011). Lastly, tropical continental convection is represented by several idealized experiments using a radiative convective equilibrium to simulate linear and clustered convection (Grant et al., 2018), and polluted and clean sea breeze convection (Grant & van den Heever, 2014a). These sets of simulations mimic the relative distributions of observations in each of the four geographical regimes (Figure 2), and represent a diverse set of aerosol, ice, mixed-phase, moist convection, deep convection, and shallow warm rain processes.

### 2.3. Rain Parameters

A normalized gamma size distribution which accounts for varying LWC is often used to represent raindrop size distributions (DSD) at the surface (e.g., Bringi et al., 2003; Testud et al., 2001; Ulbrich, 1983; Willis, 1984). Normalized gamma DSDs are described by three parameters: the median drop diameter ( $D_0$ ), the intercept parameter of the normalized gamma ( $N_w$ ), and a shape parameter  $\mu$ . We note that in the modeling community, the shape parameter is often referred to as  $\nu$ , where  $\mu = 0$  ( $\nu = 1$ ) represents the exponential distribution. These formulations are detailed in Williams et al. (2014), D18, and S22. These parameters are readily derived from disdrometer data, which measures the number of drops as a function of size. The two-moment bulk microphysics scheme used by the RAMS simulations (Section 2.2) predicts the number concentration ( $n$ ) and mixing ratio ( $q$ ) for each hydrometeor species while *assuming* a  $\mu$  shape parameter a priori. All the simulations employed here assumed  $\mu = 1$  for rain (except in Section 3.2). From these parameters, the mass-weighted mean diameter  $D_m$  (and therefore  $D_0$ ),  $N_w$ , width of the mass spectrum ( $\sigma_m$ ), and total number of drops ( $N$ ) are derived.

We note here the confusing and often inconsistent terminology used across the literature in the modeling domain and the observational domain. Therefore, it is critical to define the nomenclature to place the observations and model results in the same spectrum space. While both consider gamma particle size distributions, the *number spectrum* is often the space of choice for the modeling community, while the *mass spectrum* is typically used for

observations. The modeling community often discusses “ $D_m$ ” as the diameter of the mean mass or the mean mass diameter or the volume-number mean diameter, which can be directly computed from the model parameters  $q_r$  (hydrometer mass) and  $N_r$  (hydrometer number) without having to know the characteristics of the entire spectrum. This can also be calculated from the ratio of the third moment to the zeroth moment of the number distribution:

$$D_{3,0} = \frac{\int_{D_{\min}}^{D_{\max}} N(D) D^3 dD}{\int_{D_{\min}}^{D_{\max}} N(D) dD} \quad (1)$$

We will refer to this herein as the “volume-number mean diameter.” Observations from disdrometers and radar use the mass-spectrum, where the median volume diameter,  $D_0$ , is the diameter that divides the total water content in half (Atlas et al., 1973) and the mass-weighted mean diameter is the ratio of the fourth to the third moment of the number distribution (Ulbrich, 1983; Williams et al., 2014):

$$D_{m4,3} = \frac{\int_{D_{\min}}^{D_{\max}} N(D) D^4 dD}{\int_{D_{\min}}^{D_{\max}} N(D) D^3 dD} \quad (2)$$

which is equivalent to that calculated from the first and zeroth moments of the mass distribution (Williams, et al., 2014):

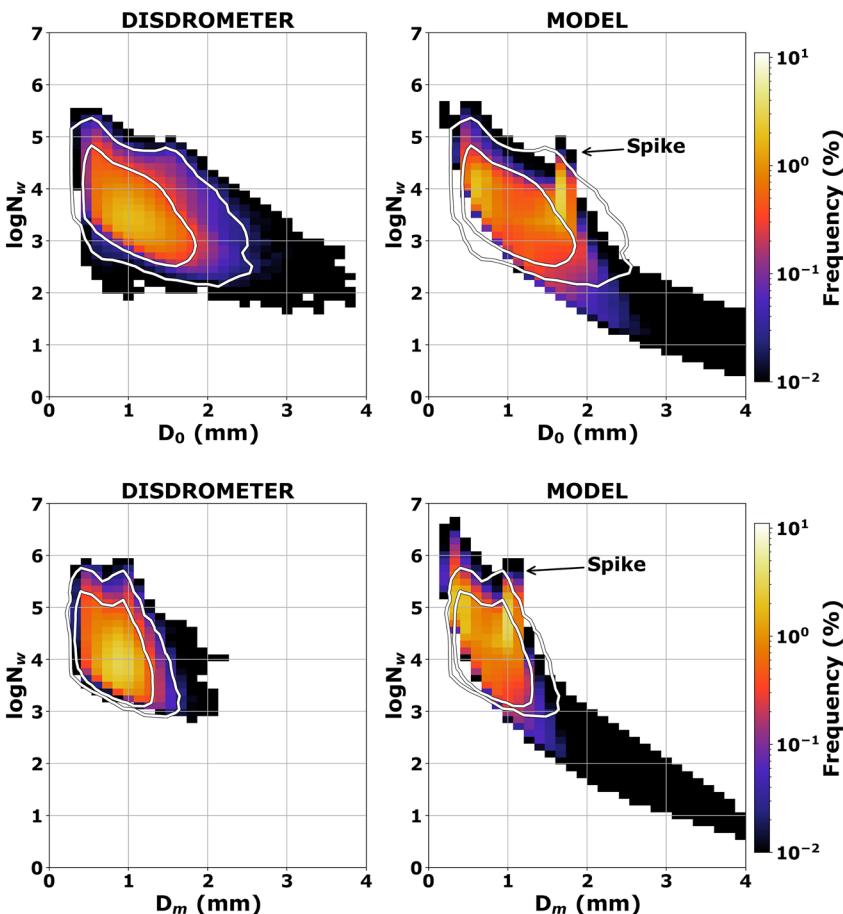
$$D_{m1,0} = \frac{\int_{D_{\min}}^{D_{\max}} m(D) D dD}{\int_{D_{\min}}^{D_{\max}} m(D) dD} \quad (3)$$

The median volume diameter is related to the mass-weighted mean diameter through  $D_{0(4,3)} = (3.67 + \mu)/(4 + \mu) D_{m(4,3)}$ . For comparison, the frequency distribution of the OBS and RAMS in the 2-D  $\log N_w$ - $D$  parameter space calculated with  $D_{m(3,0)}$  and  $D_{0(4,3)}$  are shown in Figure 3. In this space, the abscissa ( $D_m$ ,  $D_0$ ) is a measure of the mean size of raindrops in a distribution in mm, while the ordinate ( $\log N_w$ ) is related to the LWC and inversely proportional to  $D^4$ . This representation essentially describes how the available water is distributed, where in the lower right, there are low concentrations of large drops, and the upper left is high concentrations of small drops. In mass space (Figure 3, top row), the contributions from the larger diameters are accented, while in number space (Figure 3, bottom row) the distribution is more compact. Results described in S22 are framed in terms of  $D_{m(3,0)}$ , while those presented in D18 are in  $D_{0(4,3)}$ . For the remainder of this analysis, we will adopt the  $D_0 = D_{0(4,3)}$  and  $D_m = D_{m(4,3)}$  definitions related to the mass-weighted spectrum to align with the discussion and analysis in D18.

## 2.4. Principal Component Analysis

PCA is a statistical method for determining the co-variability of multiple parameters and is applied herein to understand the joint *variability* of rainfall parameters. While many studies evaluate model performance using single parameters (such as RR, accumulated rainfall, or reflectivity), PCA provides a deeper assessment of model microphysics and a model's ability to capture natural precipitation variability (represented herein by disdrometer observations; D18) through the use of multiple parameters. Following the methodology of D18, we use six parameters to define the surface rainfall:  $\log N_w$ ,  $D_m$ ,  $\sigma_m$ , RR, LWC, and  $N_r$ . These six parameters are normalized by their means to get a standard anomaly of each parameter. We use the log form to calculate the empirical orthogonal basis vectors (EOFs) of the latter three parameters to account for their large dynamic range. Each datapoint can be described by a linear combination of the resulting EOFs, the coefficients of which are called the principal components (PCs). The PCs are essentially a measure of how much a given point resembles an EOF. Each EOF has a positive and negative vector which are equal and opposite but arbitrary. For consistency, we will follow the D18 definition of the positive mode of EOF1 as all positive standard anomalies of the six parameters, and positive EOF2 as positive standard anomalies of  $\log N_w$  and negative standard anomalies of  $D_m$ . These definitions are used to group the results as described below.

D18 showed that EOF1 was broadly an “intensity” mode—that is, all standard anomalies were either positive (larger than mean values) or all negative (smaller than mean values). This mode, explaining roughly 58% of the variability in the global DSD, equates to a broad DSD containing many drops that are larger than average and comprise heavy rain (positive EOF1); conversely, low concentrations of smaller than average drops and a narrow DSD are associated with lighter than average RRs. These two modes were shown to be consistent with



**Figure 3.** Density plot of the (top)  $\log N_w$ - $D_{m(3,0)}$  space and (bottom)  $\log N_w$ - $D_{0(4,3)}$  space for (left) OBS and (right) Regional Atmospheric Modeling System (RAMS).

convective (strong vertical motions) and stratiform (weak vertical motions) processes. The second mode of variability (EOF2) was found to be related to the size distribution, with the largest standard anomalies associated with  $\log N_w$ ,  $D_m$ ,  $\sigma_m$ , and  $N_t$ , but smaller variability in the integrated parameters of RR and LWC. About 30% of the variability in the D18 global data set was found to be explained by EOF2. In this situation, there was either a large concentration of moderate drops with a narrow DSD or a broad DSD with low concentrations of larger than average drops. D18 argued that these differences were predominantly related to the warm versus ice-based rain processes. Higher EOFs accounted for 12% of the variability and were not further explored by D18.

D18 defined six PGs based on PC thresholds of the first two EOFs (Figure 1). While the imposed thresholds are somewhat arbitrary, higher thresholds pick out points at the extrema, resulting in groups which are more distinct from each other, but with fewer samples and more ambiguous points. For this analysis, we selected a threshold of PC1 and PC2  $> |0.9|$  for both OBS and simulations as this yielded the most similar distribution of the six groups. D18 studied associated radar profiles for cases of each of the six groups and labeled them by inferred microphysical characteristics: PG1 is associated with convection; PG2 with stratiform processes; PG3 is weak, shallow convection; PG4 is heavy stratiform likely formed through aggregation processes; PG5 is robust warm rain processes such as collision-coalescence; and PG6 is ice-based processes producing rain via melted hail and graupel (Figure 1). These PGs will be shown and analyzed further in Section 3.1.

## 2.5. Discussion of Model and Observation Samples

For this analysis, we note that it is not necessary to match individual points or even storm characteristics within OBS and simulations, but in order to understand the *variability*, we need sufficient samples from all of the

different types of precipitation processes and throughout the storm lifecycles in both data sets. While disdrometers offer high temporal-resolution and often multi-year temporal sampling, a single disdrometer is limited to whatever precipitation passes over a single point and therefore may not capture the full lifecycle of a system or even sample the most intense precipitation core. This is compounded by the fact that convective cores are generally highly compact in area and short-lived compared to broad, long-lived stratiform areas, a point which is highlighted by the relative sampling of the PGs in D18 (much more stratiform compared to ice-based or convection). Conversely, many of the model simulations are either case events or shorter-term (~1 day) simulations which fully encapsulate a storm lifecycle with thousands of surface points, but which may not fully capture environmental influences or the seasonality of a multi-year disdrometer installation (the exception is the BSISO simulations which encompass 40 days). To address these data difference issues, we appeal to a large statistical sample size. For the disdrometers, by selecting a large variety of locations, especially those with long-term data sets, presumably every phase of a convective lifecycle will be sampled. Additionally, some of the projects such as MC3E and IFloodS deployed disdrometer arrays (five to six disdrometers over  $<100 \text{ km}^2$  domain) which enhance, at least to some degree, the spatial sampling within a given storm. We selected model simulations which cover similar types of precipitation (e.g., warm rain, ice-based, stratiform, convective, organized, isolated) to the OBS but with a variety of environments. Although these selections are somewhat qualitative, we tried to quantitatively match the relative sample size within each of the four geospatial convective regimes between OBS and RAMS (Figure 2). Often the simulations have millions of points per simulation, whereas the OBS may have much fewer. Therefore, to achieve a similar distribution to the OBS, we only sampled every second or third model point at the surface spatially (temporally). On a convective scale, this sampling is on the order of several kilometers (minutes), and therefore all types of precipitation and phases of lifecycle are still represented.

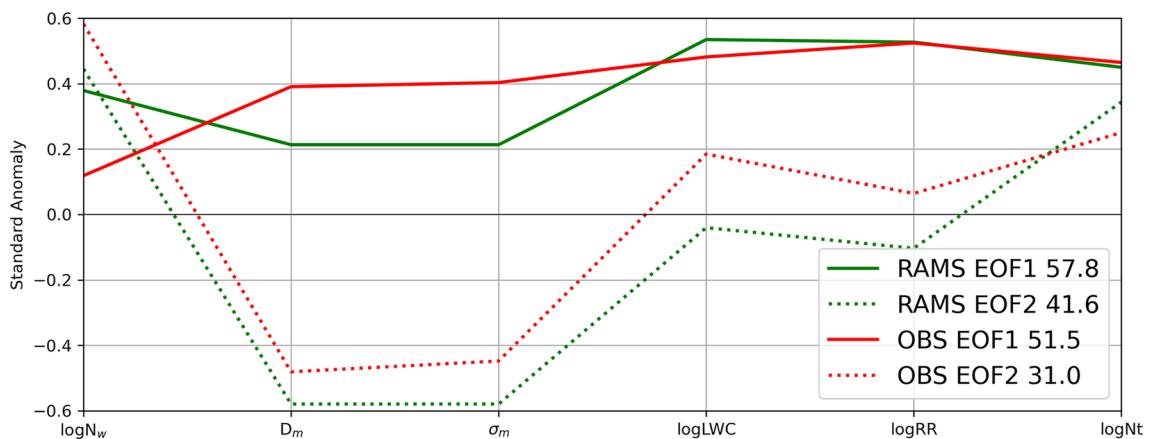
The temporal and spatial sampling of the DSDs from the disdrometer and the model must also be considered. The disdrometer DSDs are typically measured over 10s of  $\text{cm}^2$ , with an integration time period of 1-min (Rees & Garrett, 2020; Schönhuber et al., 2008), while the model DSDs are calculated over areas of 300–1,000  $\text{m}^2$  and represent an instantaneous DSD within the grid cell at each 2–5 min output interval. The types of disdrometers used in this study also have a minimum reliable detection limit of 0.2–0.5 mm, thereby under-sampling the smallest rain drops (Tokay et al., 2013). Large drops ( $>5 \text{ mm}$ ) are also likely under-sampled by disdrometers due to the small sampling cross-section and the relative infrequency of large drops (Gatlin et al., 2015). Finally, the two-moment bulk scheme used in these RAMS simulations keeps the rain drop shape parameter ( $\mu$ ) fixed, thereby reducing the degrees of freedom in the rain spectra. All of these factors must be considered when making the OBS to RAMS comparisons and drawing conclusions about the modeling and observational platforms and the processes being observed. However, these sampling issues also underpin the two-way street upon which we built this framework: the observations can help to evaluate and understand model results and parameterizations, and the simulations can contextualize the observations and processes producing them.

### 3. Results of the Multi-Parameter Framework

#### 3.1. PCA Results Applied to RAMS Database

The resulting EOFs from RAMS and OBS show minor differences between the first two EOFs (Figure 4). Both show large positive anomalies of all six parameters in EOF1, while EOF2 has large positive anomalies in  $\log N_w$  and  $\log N_t$  and negative anomalies in  $D_m$  and  $\sigma_m$ . Interestingly, the anomalies of  $\log RR$  and  $\log LWC$  are positive in OBS and negative in RAMS, but in both cases the amplitude of the standard anomaly is much smaller than the other four variables. From this we conclude that RAMS captures the same variability as is evident in the OBS with respect to the six parameters of the rain DSD. This suggests that the underlying physics of rainfall and precipitation development are largely well-simulated by the RAMS microphysical parameterizations. Consider EOF1, broadly described as the intensity mode with positive standard anomalies in all parameters. Strong vertical motions (characteristic of convection) generate heavy rain from numerous relatively large drops. Conversely, small drop sizes, low number concentrations, and low liquid water contents will generate lower RRs seen in the negative EOF1. For EOF2, D18 found this mode to be consistent with variability in size—that is, either large concentrations of moderately sized drops (as with the positive mode), consistent with warm rain processes, or low concentrations of large drops which form from melting ice and mixed phase processes, such as melting hail and graupel.

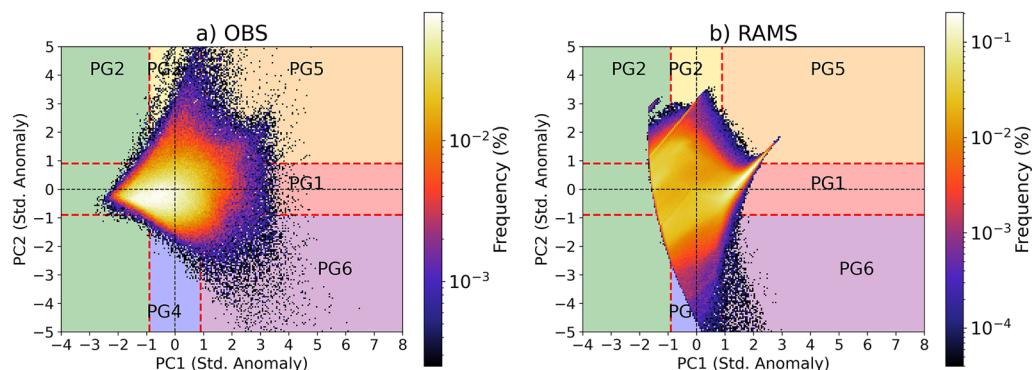
However, differences in the amount of variability explained by each of the EOFs is telling. In both RAMS and OBS, the first mode of variability, associated with rainfall intensity, explains over 50% of the variance (57.8%



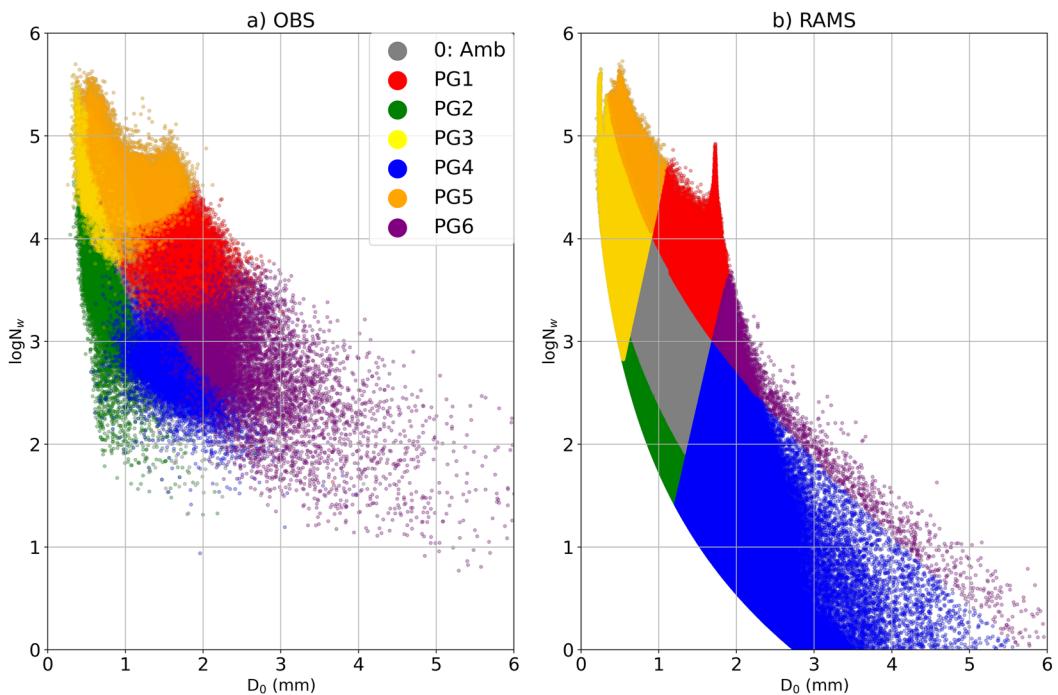
**Figure 4.** The first (solid line) and second (dotted) empirical orthogonal basis vectors (EOFs) for Regional Atmospheric Modeling System (RAMS) (green) and OBS (red) from the Principal Component Analysis (PCA). The sign of the EOF is arbitrary but we assign the vectors shown as “positive.” The amount of variance explained by each EOF is given in the legend.

and 51.5% for RAMS and OBS, respectively). However, EOF2 explains nearly the remainder of the variability (41.6%) in RAMS, with the higher order EOFs accounting for only 0.6% of the variability in this parameter space. On the other hand, OBS shows that EOF2 explains a much smaller portion of the variability (31%), with more variability (17.5%) in the higher order EOFs. While we do not consider the higher order EOFs, we hypothesize the differences could be related to the constraints imposed in the model parameterizations, or perhaps limited by the case study selections comprising the simulation database.

If we now examine the 2D space occupied by PC1 and PC2 for OBS and RAMS, we see notable differences (Figure 5). The most striking difference is the correlation between PC1 and PC2 in RAMS, also corresponding to the highest density of points (Figure 5b) extending into the red (PG1) and orange (PG5) regions. OBS has a much wider spread in values, and the highest density of points is centered closer to (0,0), although with slightly more negative PC1 values indicative of more frequent stratiform sampling, as expected from single point disdrometer measurements (Figure 5a). Both OBS and RAMS have notable discrete boundaries which are likely related to physical minimum values, such as RR = 0 and LWC = 0. The upper boundary at high PC2/low PC1 is likely related to the minimum detectable diameter in OBS, and similarly in RAMS is related to the minimum rain drop size used in the simulations. The boundary in RAMS negative PC1 sweeps out an arc and is also likely related to natural minimum values in the parameters. It is worth noting that the RAMS two-moment microphysics scheme, like many bulk microphysics schemes, fixes the DSD shape parameter,  $\mu$ . This could artificially constrain the breadth of the prognostic parameters and the resulting DSD, and this feature may explain some of the noted correlations and boundaries. This will be explored briefly in Section 3.2.



**Figure 5.** Frequency of occurrence of PC1 versus PC2 for (a) OBS and (b) Regional Atmospheric Modeling System (RAMS). Shaded boxes separated represent the principal component (PC) range of values associated with each precipitation group. Dashed red lines indicate the thresholds (0.9) used to isolate each precipitation group. The zero lines are indicated by dotted black lines.



**Figure 6.** Scatter plot of precipitation groups (colored) in  $\log N_w$ - $D_0$  space for (a) OBS and (b) Regional Atmospheric Modeling System (RAMS).

The representation of the PGs derived from the PCs in  $\log N_w$ - $D_0$  space is enlightening and provides insights into the strengths and limitations of both observations and the simulations (Figure 6). While the PGs are in the same relative locations in this space, they do not necessarily occupy the same absolute values of  $D_0$  and  $\log N_w$ . This means that RAMS successfully captures distinct PGs that have the same relative differences in the precipitation parameters as OBS—shallow convection has more numerous small drops and lower RRs compared to ice-based precipitation—but that the actual values of the parameters within those groups differ. The “spike” discussed by S22 is a notable feature of the RAMS PG1 (Figure 6b), and it was found to result from the drop-breakup parameterization. While there is a small bump in OBS in a similar  $D_0$  to the RAMS PG1 spike, this region is identified with PG5. This bump in OBS is likely related to the equilibrium distribution (Zawadzki & Antonio, 1988) where there is a balance between collision-coalescence and drop breakup in heavily raining deep warm rain cores. The RAMS parameterization appears to be too aggressively forcing this equilibrium toward a preferred drop size and artificially constraining the DSD. The imposed minimum  $N_t$  threshold in the OBS data limits the extent of PG4 and PG6 to low  $\log N_w$  compared to RAMS. In RAMS, these are very low concentrations of large drops from melted hail or graupel that have not undergone breakup. However, PG6 has larger  $D_0$  values and higher  $\log N_w$  than RAMS. PG2 sits primarily below  $\log N_w = 3$  in RAMS, but at higher  $\log N_w$  in OBS. Additionally, RAMS has very few points extending to the area of combined high  $D_0$  ( $>2$  mm) and high  $\log N_w$  ( $>2$ ). Finally, the RAMS PGs are spatially separated by a large region of ambiguous points, whereas there is significant overlap in OBS in this space. This highlights the role of the other four variables used in the PCA in defining the PGs, notably  $\mu$  which provides more degrees of freedom than simply using  $\log N_w$  and  $D_0$ .

The tails of the RAMS  $D_0$  spectrum highlight the limitations in the observational data. Reliable measurements of drops smaller than 0.2 mm, as discussed in Section 2.1, are lacking with the types of measurements comprising OBS, and therefore the points at high  $\log N_w$  and low  $D_0$  may be underrepresented by OBS. Similarly, at the large sizes and low (concentrations)  $\log N_w$  the OBS data are limited to  $N_t > 100$  drops in order to ensure enough data for the curve-fitting and prevent errant non-precipitation points from entering the data set (Thompson et al., 2015). Additionally, as discussed in Section 2.5, convective precipitation formed through melting graupel and hail, which we argue forms PG6, is limited in space and time and likely under sampled by disdrometers. In fact, the frequency distribution (Figure 3) demonstrates that these points have the lowest frequency of occurrence in both OBS and RAMS. Thus, differences in PG4 and PG6 at very low  $\log N_w$  and large  $D_0$ , could be due to

compounding sampling problems related to the transient and infrequent nature of points with that distribution and the threshold of  $N_t = 100$  drops in OBS. In these extreme ends of the DSD parameter space, RAMS is able to extend observations beyond the sampling limits to fill out the tails.

We conclude from the PCA analysis of the global data sets that the RAMS model is able to successfully reproduce the rain variability and the six PGs seen in OBS. We interpret this to mean that the underlying microphysical parameterizations generating rainfall are generally capturing the main mechanisms seen in observations. However, notable differences between the OBS and RAMS in the  $\log N_w$ - $D_0$  space suggest that certain microphysical parameterizations may need modifications, such as the drop breakup (S22). The assumed fixed shape parameter that two-moment schemes, such as the RAMS scheme used herein, may also contribute to some of the noted differences by constraining the breadth of DSDs that are allowed. The sensitivity of the model solutions to changes in the shape parameter are investigated next.

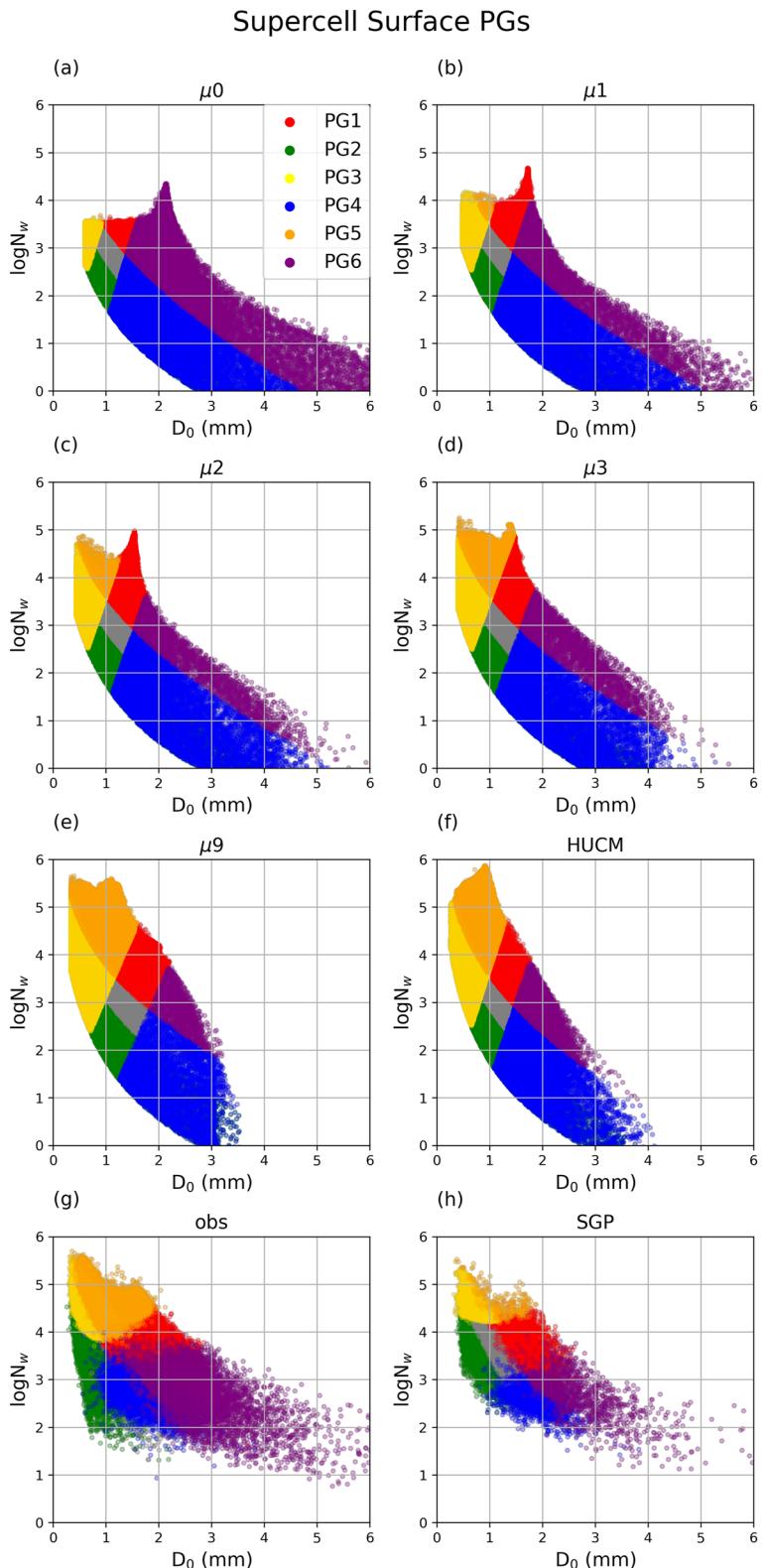
### 3.2. Investigation of the Influence of Fixed $\mu$ Assumption

The assumption of a fixed  $\mu$  value in many two-moment bulk microphysics schemes removes one degree of variability and could therefore influence the results of the PCA. To test this, we examine the impacts of a change in shape parameter on a deep convective supercell case (Grant & van den Heever, 2014b), in which a variety of precipitation formation modes, ranging from complex ice and mixed phase to warm rain, are simulated. For the sensitivity experiments, the same model setup was used but the value of the assumed rain shape parameter was permuted:  $\mu = 0, 1, 2, 3, 9$ , where  $\mu = 0$  is the exponential,  $\mu = 1$  is the default value for RAMS, and  $\mu = 9$  is an extremely narrow size distribution. Previous studies such as Morrison et al. (2012) and Freeman et al. (2019) have shown that the shape parameter assumption has important feedbacks to the total precipitation, microphysics, dynamics, storm structure, and timing of the simulated storm. Here we investigate how the PCA PGs change with different  $\mu$  values, noting that the results are also influenced by the above factors when  $\mu$  is permuted. For each sensitivity test, the six variables used in the PCA are normalized with the full RAMS database, and the data are projected onto the EOFs resulting from the full RAMS database. These steps ensure that all the different modes of variability are captured in the PCA in order to be intercomparable and consistent with the observations. Additionally, to remove the impacts of fixing  $\mu$ , we simulated the same supercell case with the Hebrew University Cloud Model (HUCM, Khain et al., 2004) Spectral Bin Microphysics (SBM) that is implemented in RAMS (Igel & van den Heever, 2017). For reference, observations from a similar convective regime represented by the Southern Great Plains (SGP) observations over north central Oklahoma are included (Table 1). Although these comparisons are not one-to-one as in a case study, the OBS serve as bounds against which we can compare the results of the  $\mu$ -sensitivities.

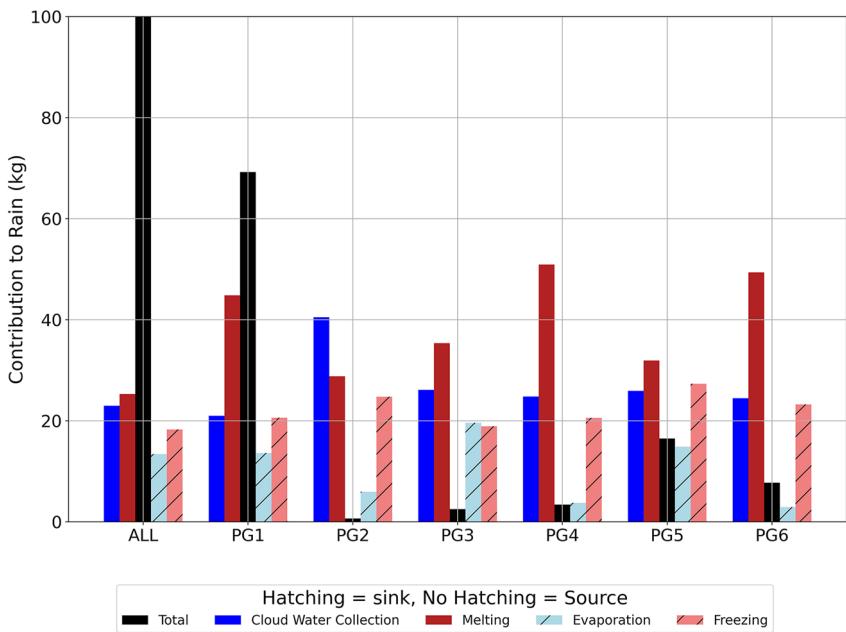
Changing the default RAMS  $\mu$  value for rain from a broad gamma distribution with  $\mu = 1$  (Figure 7b) to an exponential ( $\mu = 0$ , Figure 7a) results in many more parts of the spectrum classified as PG6 (ice processes, low  $\log N_w$  large  $D_0$ ), and virtually no PG5 (warm rain, high LWC and  $\log N_w$ , moderate  $D_0$ ) points. The largest values of  $\log N_w$  are associated with the spike (high  $\log N_w$  at  $D_0 \sim 1.6$  mm, S22), and a consequential number of points are in the very low  $\log N_w$  large  $D_0$  region identified as PG4 and PG6. Increasing  $\mu$  increases the regions of the parameter space identified as PG5 while decreasing the prominence of the spike. We hypothesize this is because as the raindrop size spectrum narrows, there are more drops of similar size which do not readily collide, thus, decreasing the influence of the drop breakup parameterization. Simultaneously, with larger  $\mu$ , the region of low  $D_0$  and high  $\log N_w$  increases while the maximum  $D_0$  decreases. The HUCM scheme, which allows the shape of the raindrop size distribution to vary, broadly resembles the narrowest distribution ( $\mu = 9$ ), with large values of  $\log N_w$  at small  $D_0$  and a limited maximum  $D_0$ , but with a relatively broad breadth of values with moderate  $\log N_w$  and  $D_0$  which is similar to the observations both in the global and SGP data sets (Figures 7g and 7h). Interestingly, in this parameter space the  $\mu = 3$  results more closely resemble the observed distributions and PG absolute values than the default value of  $\mu = 1$  most-often used for rain in RAMS simulations. Also, the value of  $\mu = 3$  is much closer to the global mean of  $\mu = 4.2$  found in D18.

## 4. Inferred Microphysical Processes

To investigate the microphysical processes contributing to each of the derived PGs we will use the RAMS  $\mu = 3$  supercell simulation (which most closely resembled the observations, Figure 7). It is important to recognize that



**Figure 7.** Precipitation groups in  $\log N_w$ - $D_0$  space for bulk simulations of a supercell with rain  $\mu = 0, 1, 2, 3, 9$  (a–e), and Hebrew University Cloud Model (HUCM) (f); and global disdrometer observations (g), and Southern Great Plains (SGP) disdrometer observations (h).



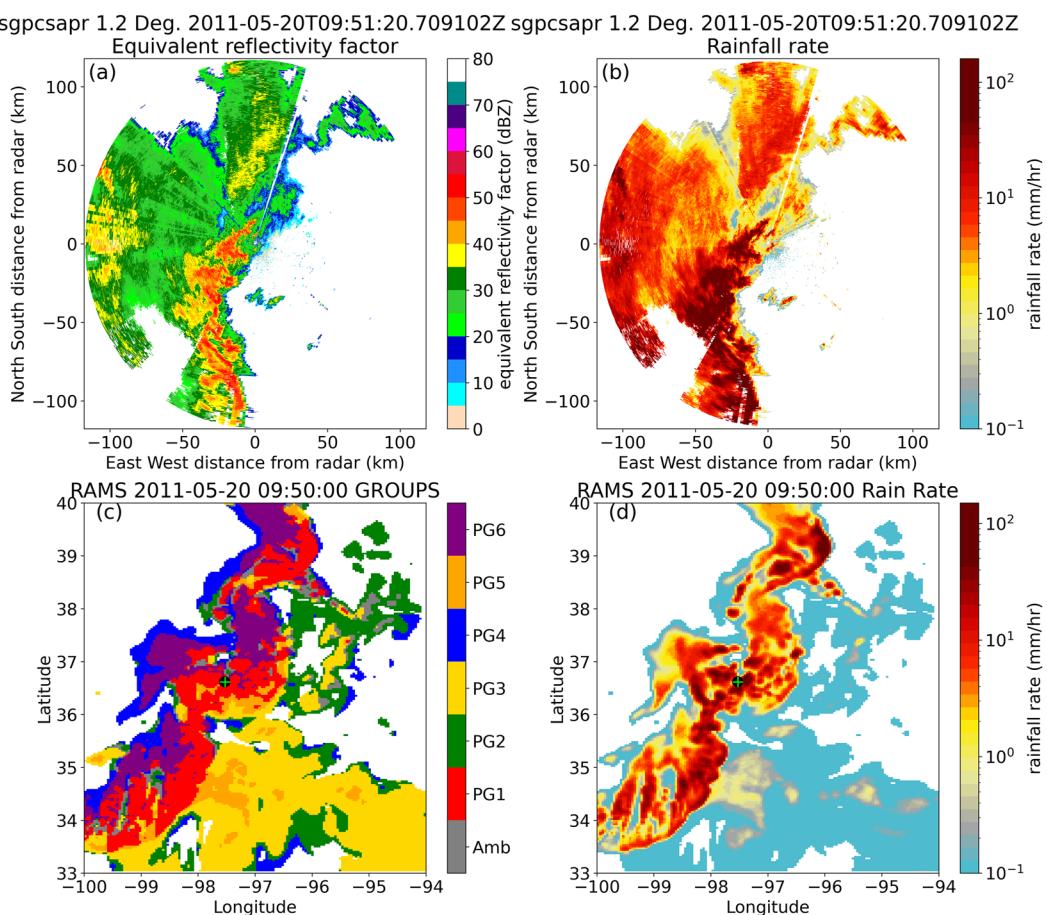
**Figure 8.** The relative contributions to the total rain budget (kg) for each precipitation group (PG) in a supercell case. Sources of rain are represented by unhatched bars and the sinks of rain represented using the hatched bars, and the total (source + sink) in black.

a single case may not fully capture the PG variability within the database, but this analysis is a step toward understanding processes associated with different PGs.

We will focus on the processes involved with generating or removing rain mass within a precipitating column. While RAMS tracks the process rates for different species (such as freezing of rain and aggregates melting to rain, etc.), to simplify the analysis we will consider warm phase processes of cloud water collection (rain source) and evaporation (rain sink), as well as ice phase processes of melting from all frozen species (rain source) and transfer of rain to all ice categories (rain sink). Condensation and evaporation are represented as a single combined term where positive (negative) values correspond to net condensation (evaporation), but in the column integrated budgets the evaporation term dominates the condensation term due to high evaporation in the subcloud air and small condensational growth of rain in saturated conditions. The process rates are initially cumulative mixing ratios for each model column in units of  $\text{g kg}^{-1} \text{ s}^{-1}$ , but are shown as a relative percentage of the total amount of rain mass (kg) generated in each PG.

Most notably, PG1 generates the largest amount of net rainfall (Figure 8, black bar), while PG2 the least, consistent with convective and stratiform classifications, respectively. Within PG2, cloud droplet collection is the greatest contributor to rain, but melting of ice is similar in magnitude. The greatest loss of rain in this PG is through its transfer to ice hydrometeors by freezing or other mechanisms. In PG1, more rain is produced by melting ice, likely because there is much more ice in the column in convection compared to stratiform, and it will eventually melt. Rain generation in PG4 and PG6 is dominated by ice melting, and loss of rain occurs primarily via freezing at supercooled temperatures. While these two groups have similar rain budgets, the total amount of rain coming from PG4 is less than PG6, potentially indicative of a convective versus stratiform mode as suggested by D18. PG3, classified as shallow or weak convection, has a larger contribution from melting compared to cloud water collection, but slightly more mass is removed from the warm phase process of evaporation compared to freezing. PG5, which D18 hypothesized was related to strong collision-coalescence, has similar contributions to rain generation from cloud water collection and melting. The loss of rain in this group is dominated by freezing instead of evaporation.

While this analysis provides some insights into the origins of the PGs, the complexity of the supercell mixed-phase microphysics obscure any dominant process for several of the groups. Here we remember that the PGs were derived from a large and diverse set of data encompassing many environments and modes of precipitation



**Figure 9.** May 20, 2011 Mid-latitude Continental Clouds and Convection Experiment (MC3E) squall line at around 0950 UTC. (a) C-SAPR reflectivity at 1.2° elevation and (b) C-SAPR rain rates at 1.2° elevation, (c) Regional Atmospheric Modeling System (RAMS) Principal Component Analysis (PCA) groups, and (d) RAMS rain. The location of SN25 is given by the black circle, and VSN with a green “+” in panels (c) and (d).

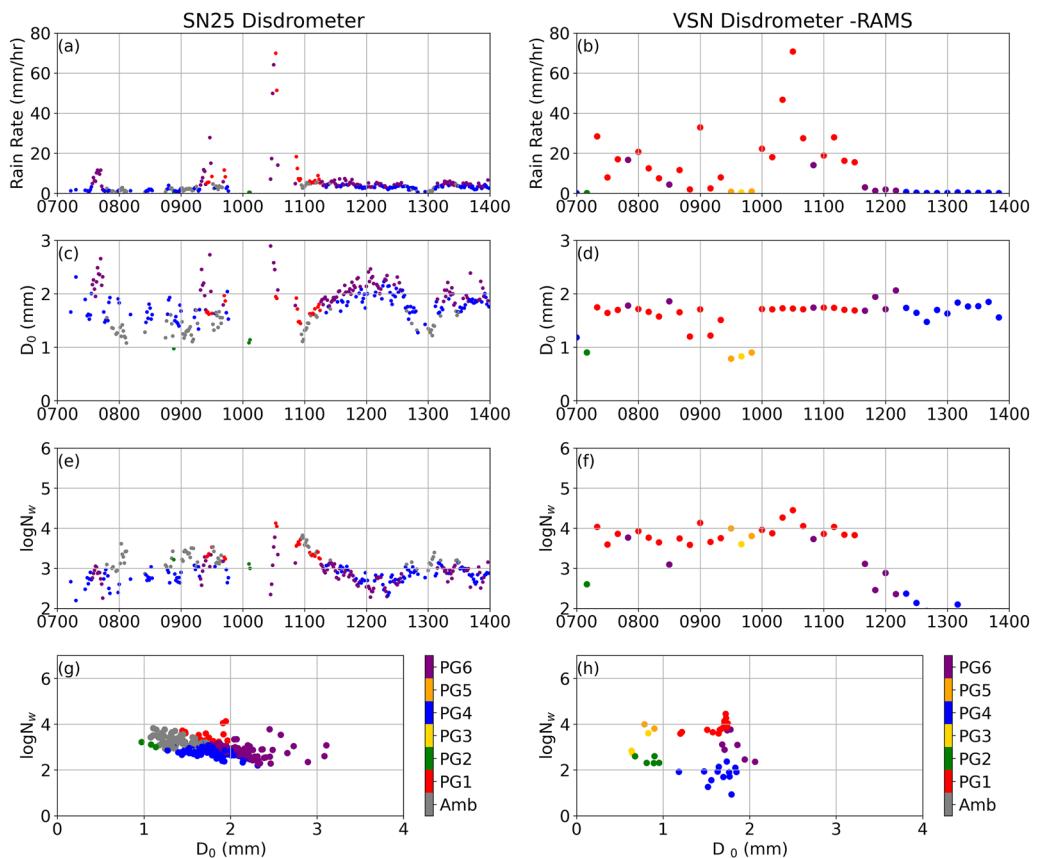
variability. Therefore, a single case may not isolate these distinct modes, and the same analysis would be required across all simulations to definitively conclude the origins of the PGs.

## 5. MC3E Case Intercomparison Using PCA Framework

We have seen that through a statistical comparison of large and diverse simulations and observations, RAMS generally reproduces the variability evident in the observational data. To understand these results in a more detailed way, we now look at individual storm-scale comparisons to try to understand both the processes behind the model and the findings based on observations. Both the RAMS and OBS data sets include the MC3E field campaign which took place in Oklahoma in 2011 (Jensen et al., 2015). On May 20, 2011, a squall line propagated eastward across the MC3E domain, which contained several NASA disdrometers deployed. This squall line had a region of stratiform precipitation with intense linear convection (Figure 9); the squall line reached the disdrometers and the DOE ARM C-SAPR polarimetric radar around 1000 UTC (Figures 9a and 9b). The case and RAMS simulation were documented in detail by Marinescu et al. (2016) and assumes the default  $\mu = 1$  for rain.

Application of the PCA to the RAMS surface data provides a spatial representation of the PGs which is not possible using the disdrometer observations (Figure 9c). The weaker leading stratiform is generally characterized by PG3 and PG2, which have small mean drop sizes and low RRs (Figure 9d). PG1 is found in the heaviest raining cores (Figures 9c and 9d), which is perhaps not unexpected because this group has the highest LWC, the largest RR, and large  $D_0$  and  $\log N_w$  values. The relative locations of PG4 and PG6 are notable in Figure 9a. Points associated with the PG6 are on the upwind edges of the convective cores, suggesting that group could be

MC3E May 20 2011



**Figure 10.** Time series of (a and b) rain rate, (c and d) median diameter, (e and f)  $\log N_w$ , and (g and h)  $\log N_w-D_0$  for the NASA SN25 2DVD (left column) and a disdrometer in a similar location in the Regional Atmospheric Modeling System (RAMS) domain, VSN (right column). Points in all panels are color-coded by the precipitation group, following the legends in (g) and (h).

associated with fallout of large precipitation ice from the main updrafts as D18 speculated and supported by the large contribution from melting to the rain budget (Figure 8). Finally, PG4 is found around the rear edges of the storm system, primarily behind the convective line. Radar analysis by D18 showed that this group was associated with strong bright band signatures in stratiform regions indicating precipitation formation through melting of larger snow, also supported by the large contribution to the rain budget from melting (Figure 8), but with significantly less rain mass than PG6.

The consistency between the inferred PGs and the simulated storm structure is an important step toward understanding the link between observations of precipitation and microphysical processes. The simulated PGs show there are coherent regions of rain with similar properties, and that these groups are aligned within areas of a storm which might be subject to different dominant microphysical processes acting to shape the surface DSD, highlighting the value of using the model to investigate and support the underlying microphysical processes.

Finally, we can compare the results between a disdrometer (SN25) and a “virtual” disdrometer (VSN) located similarly within the RAMS simulation (Figure 10). The exact location of the RAMS virtual disdrometer was selected to represent a RR rate time series similar to NASA’s 2DVD “SN25.” Both OBS and RAMS RRs traces show the convective line passage after 1000 UTC, with peak RRs around  $80 \text{ mm hr}^{-1}$ . Early in the case around 0700 UTC, a smaller peak in RR reached  $15 \text{ mm hr}^{-1}$  ( $30 \text{ mm hr}^{-1}$ ) at SN25 (VSN), and another peak around 0900 UTC to around  $30 \text{ mm hr}^{-1}$  (SN25 and VSN). However, significant differences arise when looking at the DSD traces. SN25 shows  $D_0$  at around 1.5 mm, with deviations to much larger sizes ( $>2.5 \text{ mm}$ ) associated with the peaks in RR. On the other hand, VSN shows  $D_0$  remaining relatively constant at around 1.8 mm with small deviations to

smaller sizes in the time preceding the passage of the squall line. Interestingly, there is no change in the RAMS  $D_0$  associated with the peak RR. In terms of  $\log N_w$ , SN25 shows a range of values from 3.5 to 4.5, with peak values at the passage of the squall line. VSN  $\log N_w$  values are generally above 3.5 until after the squall line passage, dropping significantly after 1130 UTC. The VSN  $\log N_w$  values fluctuate with the RR, and show the highest values (4.5) coincident with the largest RR. Since  $D_0$  does not change with RR and  $\log N_w$ , this is a clear sign of the drop breakup constraint where drops are forced to breakup in high LWC due to high collision efficiencies, resulting in numerous drops around  $d_c$  (S22).

In terms of the PGs, SN25 mostly shows PG4 and PG6 with PG1 during the passage of the convective line. A few PG2 points were identified in the hour preceding the squall line. VSN shows mainly PG1 before 1130 UTC, with PG5 prior to the squall line passage. The end of the period after 1130 UTC is classified as PG6, then PG4 similar to SN25.

This case study has shown the power of this statistical framework in which model output is effectively and appropriately compared with observational data. Application of PCA to RAMS data allowed for visualization of the spatial variation of the PGs, which is not possible with point-measurements collected by disdrometers. The identifications were spatially cohesive and roughly consistent with the conceptual model of a leading stratiform squall line, with stratiform (PG2) and shallow (PG3) in the leading stratiform regions, convection (PG1) in the cores, and strong aggregation (PG4) and melting ice (PG6) trailing the main convective line. Although the RR traces were very similar between the observational disdrometer and one placed similarly in the simulation, there were large differences in the variability of the DSDs during the passage of the convective line. The disdrometer data showed that the heaviest RR had the largest drop sizes, while the virtual disdrometer was limited in size by the parameterizations which constrained drop growth. Both showed a significant decrease in  $\log N_w$  behind the squall line and similar low RR.

## 6. Conclusions

We applied a statistically based technique to assess the primary modes of precipitation variability in a large disdrometer data set and an equally large suite of convective simulations. By comparing the PCA derived variability seen in the observations with that of the model results, we were able to better understand how the model represents precipitation, as well as leverage the model results to contextualize the observations toward a better understanding of precipitation processes.

Importantly, we found that the model was able to reproduce the first two modes of variability as determined by PCA, therefore capturing the natural co-variability of parameters related to rainfall intensity and raindrop size seen in the disdrometer data set. However, while the amount of variability explained by the first EOF was similar between OBS (~52%) and RAMS (58%), RAMS found that almost all the remaining variability (42%) was explained by the second EOF. In contrast, the OBS showed that the second EOF accounted for 31% of the variability in the data set, leaving the remaining 18% of the variability in higher order EOFs. This suggests that higher order processes found in nature may not be represented by the model microphysics, selection of cases, or limited area simulations. Examination of the PC values of the first two EOFs revealed that the model simulations show a strong correlation between positive PC1 and PC2 values that is not in the observations, suggestive of necessary constraints linking parameters of the DSD in the model's bulk microphysics scheme. RAMS-OBS comparisons also highlighted limitations of the disdrometer data set, including the inability to detect small drops (<0.2 mm) and the sampling of low concentrations of drops, as well as the challenge of observing deep convective cores which are limited in spatial and temporal extent.

We also explored the constraints imposed by assuming a single shape parameter on the PCA results. A series of sensitivity studies permuting the value of  $\mu$  in a supercell, as well as with a bin microphysics model where  $\mu$  freely evolves, were compared. Increasing the value of  $\mu$  broadened the  $\log N_w$ - $D_0$  spectrum and reduced the drop-breakup spike as drops became similar in size, decreasing their collision efficiency. The  $\mu = 3$  simulation had the closest resemblance to the observed distributions, suggesting that selection of the fixed  $\mu$  value in two-moment schemes should be carefully considered.

Microphysical process rates available in the simulations were used to investigate the contributions of processes to the rain budgets of different PGs. PG1, hypothesized to be general convection, contributed the most to the rainfall budget with dominant contributions from ice processes of riming and freezing, while PG4 and PG6, which

are thought to be dominated by ice-processes, had the largest contributions from melting. PG2, considered to be stratiform in nature, had the most rain generated from cloud water collection, and PG3, associated with weak and shallow convection, had a larger component of rain removed due to evaporation compared to riming. Rain generated in PG5, hypothesized to be dominated by warm-rain processes, still had large contributions to the rain budget from ice-processes, highlighting the complex nature of precipitation formation in a supercell case.

Finally, the framework was applied to OBS and RAMS data from a squall line case study from MC3E. Application of PCA to RAMS surface rainfall parameters revealed coherent spatial groupings within the storm structure. The leading stratiform area was generally classified as weak/shallow and stratiform, while the convective line was identified as the D18 convective category. Finally, behind the convective line were the ice-based and strong aggregation PGs, suggestive of the fallout of larger melted precipitation ice like hail and graupel, or size sorting. These structures are consistent with expected storm dynamics and microphysical processes, and they demonstrate how the model provides context for the disdrometer point-measurements and underpin the dominant microphysical processes that were hypothesized in D18. Comparison of a synthetic disdrometer trace placed in RAMS to a comparable OBS disdrometer revealed how similar RRs could result from different rain DSDs, especially in the heaviest raining cores due to constraints imposed by the drop breakup parametrization.

We have demonstrated a new methodology to statistically compare a large disdrometer data set and a set of wide-ranging cloud simulations toward the mutual benefit of understanding precipitation processes in both models and observations. This analysis reveals both strengths and limitations in the rainfall parameterizations and the observational data. Such a statistical approach to model and observational comparisons provides a more sophisticated and meaningful means for model validation, while leveraging the strengths of CRMs to contextualize observations and better understand processes which cannot be directly observed. Such a framework could be applied to other types of weather and climate models with different microphysics schemes, toward the same analysis given a diverse set of simulations representing a variety of cloud types.

## Data Availability Statement

The disdrometer data used in this study are available from NASA (Farrar, 2023; Petersen et al., 2011; Petersen & Gatlin, 2013, 2014; Petersen et al., 2010, 2016, 2019; NASA, 2023) and ARM DOE (D. Wang & Bartholomew, 2011a, 2011b, 2011c, 2011d, 2018). CSAPR data shown in Figure 9 is available from Lindenmaier et al. (2011) and plotted using the DOE-PyART software (Helmus & Collis, 2016). The archive of the available files and tools related to the analyses, as well as model Namelists and code necessary to reproduce numerical simulations, is available in Saleeby et al. (2022a). The Python and Jupyter notebooks to create the figures are available on GitHub (Dolan, 2023).

## Acknowledgments

The authors would like to extend gratitude to those whose prior hard work allowed this unique opportunity to bring together a large database of simulations, including Dr. Leah Grant, Dr. Peter Marinescu, Dr. Ben Toms, and Dr. Adele Igel, all who provided high-quality archived simulations used herein. We would also like to thank Dr. Brody Fuchs for his input and discussions on this project. We also thank the two anonymous reviewers whose comments improved the manuscript. This research was supported primarily by the U.S. Department of Energy's Atmospheric System Research, an Office of Science Biological and Environmental Research program, under DE-SC0017977.

## References

- Atlas, D., Srivastava, R. C., & Sekhon, R. S. (1973). Doppler radar characteristics of precipitation at vertical incidence. *Reviews of Geophysics*, 11, 1–35. <https://doi.org/10.1029/rg011i001p00001>
- Barros, A. P., Tao, J., Wu, D., Nogueira, M., Peters-Lidard, C. D., Gourley, J. J., et al. (2014). *The GPM GV Integrated Precipitation and Hydrology Experiment (IPHEX) in the southern appalachians—Focus on water cycle processes*. American Geophysical Union, Fall Meeting.
- Bringi, V. N., Chandrasekar, V., Hubbert, J., Gorgucci, E., Randeu, W. L., & Schoenhuber, M. (2003). Raindrop size distribution in different climatic regimes from disdrometer and dual-polarized radar analysis. *Journal of the Atmospheric Sciences*, 60(2), 354–365. [https://doi.org/10.1175/1520-0469\(2003\)060<0354:RSIDC>2.0.CO;2](https://doi.org/10.1175/1520-0469(2003)060<0354:RSIDC>2.0.CO;2)
- Cheng, W. Y., Carrión, G. G., Cotton, W. R., & Saleeby, S. M. (2009). Influence of atmospheric aerosols on the development of precipitating trade wind cumuli. *Journal of Geophysical Research*, 114(D8), D08201. <https://doi.org/10.1029/2008JD011011>
- Chudler, K., Rutledge, S. A., & Dolan, B. (2022). Unique radar observations of large raindrops in tropical warm rain during PISTON. *Monthly Weather Review*, 150(10), 2719–2736. <https://doi.org/10.1175/mwr-d-21-0298.1>
- Cotton, W. R., Pielke, R. A., Walko, R. L., Walko, R. L., Liston, G. E., Tremback, C. J., et al. (2003). RAMS 2001: Current status and future directions. *Meteorology and Atmospheric Physics*, 82(1–4), 5–29. <https://doi.org/10.1007/s00703-001-0584-9>
- Dolan, B. (2023). bdolan44/JGR\_Manuscript\_2023, version 1.0 [Software]. Zenodo. <https://doi.org/10.5281/zenodo.8342559>
- Dolan, B., Fuchs, B., Rutledge, S. A., Barnes, E., & Thompson, E. (2018). Primary modes of global rain drop-size distribution variability. *Journal of the Atmospheric Sciences*, 75(5), 1453–1476. <https://doi.org/10.1175/JAS-D-17-0242.1>
- Drobinski, P., Ducrocq, V., Alpert, P., Anagnostou, E., Béranger, K., Borga, M., et al. (2014). HyMeX: A 10-year multidisciplinary program on the Mediterranean water cycle. *Bulletin of the American Meteorological Society*, 95(7), 1063–1082. <https://doi.org/10.1175/BAMS-D-12-00242.1>
- Ducrocq, V., Braud, I., Davolio, S., Ferretti, R., Flamant, C., Jansa, A., et al. (2014). HyMeX-SOP1: The field campaign dedicated to heavy precipitation and flash flooding in the northwestern Mediterranean. *Bulletin of the American Meteorological Society*, 95(7), 1083–1100. <https://doi.org/10.1175/bams-d-12-00244.1>
- Dunion, J. P., & Velden, C. S. (2004). The impact of the Saharan Air Layer on Atlantic tropical cyclone activity. *Bulletin of the American Meteorological Society*, 85(3), 353–365. <https://doi.org/10.1175/bams-85-3-353>

- Farrar, J. T. (2023). SPURS-2 disdrometer data [Dataset]. PO.DAAC, CA. <https://doi.org/10.5067/SPUR2-DISDR>
- Freeman, S. W., Igel, A. L., & van den Heever, S. C. (2019). Relative sensitivities of rainfall prediction to fixed shape parameters and collection efficiencies. *Quarterly Journal of the Royal Meteorological Society*, 145(722), 2181–2201. <https://doi.org/10.1002/qj.3550>
- Gatlin, P. N., Thurai, M., Bringi, V. N., Petersen, W., Wolff, D., Tokay, A., et al. (2015). Searching for large raindrops: A global summary of two-dimensional video disdrometer observations. *Journal of Applied Meteorology and Climatology*, 54(5), 1069–1089. <https://doi.org/10.1175/jamc-d-14-0089.1>
- Grant, L. D., Lane, T. P., & van den Heever, S. C. (2018). The role of cold pools in tropical oceanic convective systems. *Journal of the Atmospheric Sciences*, 75(8), 2615–2634. <https://doi.org/10.1175/jas-d-17-0352.1>
- Grant, L. D., & van den Heever, S. C. (2014a). Aerosol-cloud-land surface interactions within tropical sea breeze convection. *Journal of Geophysical Research: Atmospheres*, 119(13), 8340–8361. <https://doi.org/10.1002/2014jd021912>
- Grant, L. D., & van den Heever, S. C. (2014b). Microphysical and dynamical characteristics of low-precipitation and classic supercells. *Journal of the Atmospheric Sciences*, 71(7), 2604–2624. <https://doi.org/10.1175/jas-d-13-0261.1>
- Grasso, L. D. (2000). The dissipation of a left-moving cell in a severe storm environment. *Monthly Weather Review*, 128(8), 2797–2815. [https://doi.org/10.1175/1520-0493\(2000\)128<2797:tdalm>2.0.co;2](https://doi.org/10.1175/1520-0493(2000)128<2797:tdalm>2.0.co;2)
- Helmus, J. J., & Collis, S. M. (2016). The Python ARM Radar Toolkit (Py-ART), a library for working with weather radar data in the Python Programming Language. [Software]. *Journal of Open Research Software*, 4, e25. <http://doi.org/10.5334/jors.119.1>
- Igel, A. L., & van den Heever, S. C. (2017). The role of the gamma function shape parameter in determining differences between condensation rates in bin and bulk microphysics schemes. *Atmospheric Chemistry and Physics*, 17(7), 4599–4609. <https://doi.org/10.5194/acp-17-4599-2017>
- Iguchi, T., Matsui, T., Tokay, A., Kolllias, P., & Tao, W.-K. (2012). Two distinct modes in one-day rainfall event during MC3E field campaign: Analyses of disdrometer observations and WRF-SBM simulation. *Geophysical Research Letters*, 39(24), L24805. <https://doi.org/10.1029/2012GL053329>
- Jensen, M. P., Petersen, W. A., Bansemer, A., Bharadwaj, N., Carey, L. D., Cecil, D. J., et al. (2015). The midlatitude continental convective clouds experiment (MC3E). *Bulletin of the American Meteorological Society*, 97(9), 1667–1686. <https://doi.org/10.1175/BAMS-D-14-00228.1>
- Jiang, H., & Feingold, G. (2006). Effect of aerosol on warm convective clouds: Aerosol-cloud-surface flux feedbacks in a new coupled large eddy model. *Journal of Geophysical Research*, 111(D1), D01202. <https://doi.org/10.1029/2005JD006138>
- Khain, A., Pokrovsky, A., Pinsky, M., Seifert, A., & Phillips, V. (2004). Simulation of effects of atmospheric aerosols on deep turbulent convective clouds using a spectral microphysics mixed-phase cumulus cloud model. Part I: Model description and possible applications. *Journal of the Atmospheric Sciences*, 61(24), 2963–2982. <https://doi.org/10.1175/JAS-3350.1>
- Klepp, C. (2015). The oceanic shipboard precipitation measurement network for surface validation—OceanRAIN. *Atmospheric Research*, 163, 74–90. <https://doi.org/10.1016/j.atmosres.2014.12.014>
- Klepp, C., Michel, S., Protat, A., Burdanowitz, J., Albern, N., Kähnert, M., et al. (2018). OceanRAIN, a new in-situ shipboard global ocean surface-reference dataset of all water cycle components. *Scientific Data*, 5(1), 180122. <https://doi.org/10.1038/sdata.2018.122>
- Kumjian, M. R., & Ryzhkov, A. V. (2008). Polarimetric signatures in supercell thunderstorms. *Journal of Applied Meteorology and Climatology*, 47(7), 1940–1961. <https://doi.org/10.1175/2007jamc1874.1>
- Leung, G., & van den Heever, S. C. (2022). Controls on the development and circulation of terminal versus transient congestus clouds and implications for midlevel aerosol transport. *Journal of the Atmospheric Sciences*, 79(11), 3083–3101. <https://doi.org/10.1175/JAS-D-21-0314.1>
- Li, X., Tao, W. K., Khain, A. P., Simpson, J., & Johnson, D. E. (2009). Sensitivity of a cloud-resolving model to bulk and explicit bin microphysical schemes. Part I: Comparisons. *Journal of the Atmospheric Sciences*, 66(1), 3–21. <https://doi.org/10.1175/2008jas2464.1>
- Lindenmaier, I., Feng, Y.-C., Matthews, A., Wendler, T., & Castro, V. (2011). C-Band Scanning ARM Precipitation Radar (CSAPRSEC) [Dataset]. Atmospheric Radiation Measurement (ARM) User Facility. <https://doi.org/10.5439/1984392>
- Lindstrom, E. J., Edson, J. B., Schanze, J. J., & Shcherbina, A. Y. (2019). SPURS-2: Salinity Processes in the Upper-ocean Regional study 2. The eastern equatorial Pacific experiment. *Oceanography*, 32(2), 15–19. <https://doi.org/10.5670/oceanog.2019.207>
- Marinescu, P. J., Kennedy, P. C., Bell, M. M., Drager, A. J., Grant, L. D., Freeman, S. W., & van den Heever, S. C. (2020). Updraft vertical velocity observations and uncertainties in High Plains supercells using radiosondes and radars. *Monthly Weather Review*, 148(11), 4435–4452. <https://doi.org/10.1175/MWR-D-20-0071.1>
- Marinescu, P. J., van den Heever, S. C., Saleeby, S. M., & Kreidenweis, S. M. (2016). The microphysical contributions to and evolution of latent heating profiles in two MC3E MCSs. *Journal of Geophysical Research: Atmospheres*, 121(13), 7913–7935. <https://doi.org/10.1002/2016jd024762>
- Martin, S. T., Machado, L., Manzi, A. O., Souza, R. A. F., Schumacher, C., Wang, J., et al. (2017). The Green Ocean Amazon experiment (GoAmazon2014/5) observes pollution affecting gases, aerosols, clouds, and rainfall over the rain forest. *Bulletin of the American Meteorological Society*, 98(5), 981–997. <https://doi.org/10.1175/bams-d-15-00221.1>
- May, P. T., Mather, J. H., Vaughan, G., Jakob, C., McFarquhar, G. M., Bower, K. N., & Mace, G. G. (2008). The tropical warm pool international cloud experiment. *Bulletin of the American Meteorological Society*, 89(5), 629–646. <https://doi.org/10.1175/bams-89-5-629>
- McGee, C. J., & van den Heever, S. C. (2014). Latent heating and mixing due to entrainment in tropical deep convection. *Journal of the Atmospheric Sciences*, 71(2), 816–832. <https://doi.org/10.1175/jas-d-13-0140.1>
- Meyers, M. P., Walko, R. L., Harrington, J. Y., & Cotton, W. R. (1997). New RAMS cloud microphysics parameterization. Part II. The two-moment scheme. *Atmospheric Research*, 45(1), 3–39. [https://doi.org/10.1016/s0169-8095\(97\)00018-5](https://doi.org/10.1016/s0169-8095(97)00018-5)
- Morrison, H., Tessendorf, S. A., Ikeda, K., & Thompson, G. (2012). Sensitivity of a simulated midlatitude squall line to parameterization of raindrop breakup. *Monthly Weather Review*, 140(8), 2437–2460. <https://doi.org/10.1175/mwr-d-11-00283.1>
- NASA/LARC/SD/ASDC. (2023). PISTON 2018 Research Vessel (RV) Thompson ship datasets of ocean, atmosphere, and air-sea interaction [Dataset]. NASA Langley Atmospheric Science Data Center DAAC. <https://doi.org/10.5067/SUBORBITAL/PISTON2018-ONR-NOAA/RVTOMPSON/DATA001>
- Olsson, P., & Cotton, W. R. (1997). Balanced and unbalanced circulations in a primitive equation simulation of a midlatitude MCC. Part I: The numerical simulation. *Journal of the Atmospheric Sciences*, 54(4), 457–478. [https://doi.org/10.1175/1520-0469\(1997\)054<0457:baucia>2.0.co;2](https://doi.org/10.1175/1520-0469(1997)054<0457:baucia>2.0.co;2)
- Petersen, W., Carey, L., Bringi, V. N., Tokay, A., & Gatlin, P. (2011). GPM ground validation two-dimensional video disdrometer (2DVD) MC3E [Dataset]. NASA EOSDIS Global Hydrology Resource Center Distributed Active Archive Center Huntsville, Alabama, U.S.A. <https://doi.org/10.5067/GPMGV/MC3E/2DVD/DATA301>
- Petersen, W., & Gatlin, P. (2013). GPM ground validation two-dimensional video disdrometer (2DVD) IFloodS [Dataset]. NASA EOSDIS Global Hydrology Resource Center Distributed Active Archive Center Huntsville, Alabama, U.S.A. <https://doi.org/10.5067/GPMGV/IFLOODS/2DVD/DATA301>

- Petersen, W., & Gatlin, P. (2014). GPM ground validation two-dimensional video disdrometer (2DVD) IPHEX [Dataset]. NASA EOSDIS Global Hydrology Resource Center Distributed Active Archive Center Huntsville, Alabama, U.S.A. <https://doi.org/10.5067/GPMGV/IPHEX/2DVD/DATA301>
- Petersen, W. A., Tokay, A., Gatlin, P., & Wingo, M. T. (2010). GPM ground validation two-dimensional video disdrometer (2DVD) NSSTC [Dataset]. NASA EOSDIS Global Hydrology Resource Center Distributed Active Archive Center Huntsville, Alabama, U.S.A. <https://doi.org/10.5067/GPMGV/NSSTC/2DVD/DATA201>
- Petersen, W. A., Tokay, A., Gatlin, P. N., & Wingo, M. T. (2016). GPM ground validation two-dimensional video disdrometer (2DVD) HYMEX [Dataset]. NASA Global Hydrology Resource Center DAAC, Huntsville, Alabama. <https://doi.org/10.5067/GPMGV/HYMEX/2DVD/DATA301>
- Petersen, W. A., Tokay, A., Gatlin, P. N., & Wingo, M. T. (2019). GPM ground validation two-dimensional video disdrometer (2DVD) WFF V1 [Dataset]. NASA Global Hydrology Resource Center DAAC, Huntsville, Alabama. <https://doi.org/10.5067/GPMGV/WFF/2DVD/DATA201>
- Protat, A., Klepp, C., Louf, V., Petersen, W. A., Alexander, S. P., Barros, A., et al. (2019). The latitudinal variability of oceanic rainfall properties and its implication for satellite retrievals: 1. Drop size distribution properties. *Journal of Geophysical Research: Atmospheres*, 124(23), 13291–13311. <https://doi.org/10.1029/2019jd031010>
- Rees, K., & Garrett, T. J. (2020). Effect of disdrometer sampling area and time on the precision of precipitation rate measurement. *Atmospheric Measurement Techniques & Discussions*, 1–16.
- Riley Dellaripa, E. M., Maloney, E. D., & van den Heever, S. C. (2018). Wind-flux feedbacks and convective organization during the November 2011 MJO event in a high-resolution model. *Journal of the Atmospheric Sciences*, 75(1), 57–84. <https://doi.org/10.1175/jas-d-16-0346.1>
- Rutledge, S. A., Chandrasekar, V., Fuchs, B. R., George, J., Junyent, F., Dolan, B., et al. (2019). SEA-POL goes to sea. *Bulletin of the American Meteorological Society*, 100(11), 2285–2301. <https://doi.org/10.1175/BAMS-D-18-0233.1>
- Saleeby, S. M., & Cotton, W. R. (2004). A large-droplet mode and prognostic number concentration of cloud droplets in the Colorado State University Regional Atmospheric Modeling System (RAMS). Part I: Module descriptions and supercell test simulations. *Journal of Applied Meteorology*, 43(1), 182–195. [https://doi.org/10.1175/1520-0450\(2004\)043<0182:almapn>2.0.co;2](https://doi.org/10.1175/1520-0450(2004)043<0182:almapn>2.0.co;2)
- Saleeby, S. M., Dolan, B., & Bukowski, J. (2022a). Data associated with “assessing rain drop breakup parameterizations using disdrometer observations” [Dataset and Software]. <https://doi.org/10.25675/10217/235397>
- Saleeby, S. M., Dolan, B., Bukowski, J., van Valkenburg, K., van den Heever, S. C., & Rutledge, S. A. (2022b). Assessment and limitations of parameterized raindrop breakup in numerical models of cloud microphysics. *Journal of the Atmospheric Sciences*, 79(11), 2949–2963. <https://doi.org/10.1175/jas-d-21-0335.1>
- Saleeby, S. M., Herbener, S. R., van den Heever, S. C., & L’Ecuyer, T. (2015). Impacts of cloud droplet-nucleating aerosols on shallow tropical convection. *Journal of the Atmospheric Sciences*, 72(4), 1369–1385. <https://doi.org/10.1175/jas-d-14-0153.1>
- Saleeby, S. M., Twohy, C. H., van den Heever, S. C., & DeMott, P. J. (2011). Impacts of Saharan dust on the microphysical processes in tropical convection. AGU Fall Meeting. A11C-0107. *Aerosol Convection Interactions: Microphysics, Radiation, and Dynamics*, San Francisco, CA. 4–9 December.
- Saleeby, S. M., & van den Heever, S. C. (2013). Developments in the CSU-RAMS aerosol model: Emissions, nucleation, regeneration, deposition, and radiation. *Journal of Applied Meteorology and Climatology*, 52(12), 2601–2622. <https://doi.org/10.1175/jamc-d-12-0312.1>
- Schönhuber, M., Lammer, G., & Randeu, W. L. (2008). The 2D-video-disdrometer. In S. Michaelides (Ed.), *Precipitation: Advances in measurement, estimation and prediction* (pp. 3–31). Springer. [https://doi.org/10.1007/978-3-540-77655-0\\_1](https://doi.org/10.1007/978-3-540-77655-0_1)
- Seo, B.-C., Dolan, B., Krajewski, W., Rutledge, S. A., & Petersen, W. A. (2015). Comparison of single- and dual-polarization-based rainfall estimates using NEXRAD data for the NASA Iowa Flood studies project. *Journal of Hydrometeorology*, 16(4), 1658–1675. <https://doi.org/10.1175/jhm-d-14-0169.1>
- Sheffield, A. M., Saleeby, S. M., & van den Heever, S. C. (2015). Aerosol-induced mechanisms for cumulus congestus growth. *Journal of Geophysical Research: Atmospheres*, 120(17), 8941–8952. <https://doi.org/10.1002/2015jd023743>
- Skamarock, W. C. (2006). Positive-definite and monotonic limiters for unrestricted-time-step transport schemes. *Monthly Weather Review*, 134(8), 2241–2250. <https://doi.org/10.1175/MWR3170.1>
- Sobel, A. H., Sprintall, J., Maloney, E. D., Martin, Z. K., Wang, S., de Szeike, S. P., et al. (2021). Large-scale state and evolution of the atmosphere and ocean during PISTON 2018. *Journal of Climate*, 34(12), 5017–5035. <https://doi.org/10.1175/jcli-d-20-0517.1>
- Testud, J., Oury, S., Black, R. A., Amayenc, P., & Dou, X. K. (2001). The concept of “normalized” distribution to describe raindrop spectra: A tool for cloud physics and cloud remote sensing. *Journal of Applied Meteorology and Climatology*, 40(6), 1118–1140. [https://doi.org/10.1175/1520-0450\(2001\)040<1118:TCNDT>2.0.CO;2](https://doi.org/10.1175/1520-0450(2001)040<1118:TCNDT>2.0.CO;2)
- Thompson, E. J., Rutledge, S. A., Dolan, B., & Thurai, M. (2015). Drop size distributions and radar observations of convective and stratiform rain over the equatorial Indian and West Pacific Oceans. *Journal of the Atmospheric Sciences*, 72(11), 4091–4125. <https://doi.org/10.1175/jas-d-14-0206.1>
- Thurai, M., Petersen, W. A., Tokay, A., Schultz, C., & Gatlin, P. (2011). Drop size distribution comparisons between Parsivel and 2-D video disdrometers. *Advances in Geosciences*, 30, 3–9. <https://doi.org/10.5194/adgeo-30-3-2011>
- Thurai, M., Williams, C. R., & Bringi, V. N. (2014). Examining the correlations between drop size distribution parameters using data from two side-by-side 2D-video disdrometers. *Atmospheric Research*, 144, 95–110. <https://doi.org/10.1016/j.atmosres.2014.01.002>
- Tokay, A., Kruger, A., & Krajewski, W. F. (2001). Comparison of drop size distribution measurements by impact and optical disdrometers. *Journal of Applied Meteorology and Climatology*, 40(11), 2083–2097. [https://doi.org/10.1175/1520-0450\(2001\)040<2083:codsdm>2.0.co;2](https://doi.org/10.1175/1520-0450(2001)040<2083:codsdm>2.0.co;2)
- Tokay, A., Petersen, W. A., Gatlin, P., & Wingo, M. (2013). Comparison of raindrop size distribution measurements by co-located disdrometers, 2013. *Journal of Atmospheric and Oceanic Technology*, 30(8), 1672–1690. <https://doi.org/10.1175/JTECH-D-12-00163.1>
- Tokay, A., Wolff, D. B., & Petersen, W. A. (2014). Evaluation of new version of the laser-optical disdrometer, OTT Parsivel2. *Journal of Atmospheric and Oceanic Technology*, 31(6), 1276–1288. <https://doi.org/10.1175/JTECH-D-13-00174.1>
- Toms, B. A., van den Heever, S. C., Riley Dellaripa, E. M., Saleeby, S. M., & Maloney, E. D. (2020). The boreal summer Madden-Julian oscillation and moist convective morphology over the maritime continent. *Journal of the Atmospheric Sciences*, 77(2), 647–667. <https://doi.org/10.1175/JAS-D-19-0029.1>
- Ulbrich, C. W. (1983). Natural variations in the analytical form of the raindrop size distribution. *Journal of Applied Meteorology and Climatology*, 22(10), 1764–1775. [https://doi.org/10.1175/1520-0450\(1983\)022,1764:NVITAF2.0.CO;2](https://doi.org/10.1175/1520-0450(1983)022,1764:NVITAF2.0.CO;2)
- Varble, A. C., Nesbitt, S. W., Salio, P., Hardin, J. C., Bharadwaj, N., Borque, P., et al. (2021). Utilizing a storm-generating hotspot to study convective cloud transitions: The CACTI experiment. *Bulletin of the American Meteorological Society*, 102(8), E1597–E1620. <https://doi.org/10.1175/BAMS-D-20-0030.1>

- Varble, A., Zipser, E. J., Fridlind, A. M., Zhu, P., Ackerman, A. S., Chaboureau, J.-P., et al. (2014). Evaluation of cloud-resolving and limited area model intercomparison simulations using TWP-ICE observations: 2. Precipitation microphysics. *Journal of Geophysical Research: Atmospheres*, 119(24), 13919–13945. <https://doi.org/10.1002/2013JD021372>
- Walko, R. L., Cotton, W. R., Meyers, M. P., & Harrington, J. Y. (1995). New RAMS cloud microphysics parameterization: Part I. The single-moment scheme. *Atmospheric Research*, 38(1–4), 29–62. [https://doi.org/10.1016/0169-8095\(94\)00087-t](https://doi.org/10.1016/0169-8095(94)00087-t)
- Wang, D., Bartholomew, M., & Shi, Y. (2014). Laser disdrometer (LD) [Dataset]. Atmospheric Radiation Measurement (ARM) User Facility. <https://doi.org/10.5439/1779709>
- Wang, D., & Bartholomew, M. J. (2011a). Video disdrometer (VDIS) Darwin [Dataset]. Atmospheric Radiation Measurement (ARM) User Facility. <https://doi.org/10.5439/1025315>
- Wang, D., & Bartholomew, M. J. (2011b). Video disdrometer (VDIS) Gan [Dataset]. Atmospheric Radiation Measurement (ARM) User Facility. <https://doi.org/10.5439/1025315>
- Wang, D., & Bartholomew, M. J. (2011c). Video disdrometer (VDIS) SGP [Dataset]. Atmospheric Radiation Measurement (ARM) User Facility. <https://doi.org/10.5439/1025315>
- Wang, D., & Bartholomew, M. J. (2011d). Video disdrometer (VDIS) TWP. [Dataset]. Atmospheric Radiation Measurement (ARM) User Facility. <https://doi.org/10.5439/1025315>
- Wang, D., & Bartholomew, M. J. (2014). Video disdrometer (VDIS) during ENA. [Dataset]. Atmospheric Radiation Measurement (ARM) User Facility. <https://doi.org/10.5439/1025315>
- Wang, D., & Bartholomew, M. J. (2018). Video disdrometer (VDIS) during CACTI. [Dataset]. Atmospheric Radiation Measurement (ARM) User Facility. <https://doi.org/10.5439/1025315>
- Wang, J., Wood, R., Jensen, M. P., Chiu, J. C., Liu, Y., Lamer, K., et al. (2022). Aerosol and Cloud Experiments in the Eastern North Atlantic (ACE-ENA). *Bulletin of the American Meteorological Society*, 103(2), E619–E641. <https://doi.org/10.1175/BAMS-D-19-0220.1>
- Williams, C. R., Bringi, V. N., Carey, L. D., Chandrasekar, V., Gatlin, P. N., Haddad, Z. S., et al. (2014). Describing the shape of raindrop size distributions using uncorrelated raindrop mass spectrum parameters. *Journal of Applied Meteorology and Climatology*, 53(5), 1282–1296. <https://doi.org/10.1175/jamc-d-13-076.1>
- Willis, P. T. (1984). Functional fits to some observed drop size distributions and parameterization of rain. *Journal of the Atmospheric Sciences*, 41(9), 1648–1661. [https://doi.org/10.1175/1520-0469\(1984\)041,1648:FFTSOD.2.0.CO;2](https://doi.org/10.1175/1520-0469(1984)041,1648:FFTSOD.2.0.CO;2)
- Yoneyama, K., Zhang, C., & Long, C. N. (2013). Tracking pulses of the Madden-Julian oscillation. *Bulletin of the American Meteorological Society*, 94(12), 1871–1891. <https://doi.org/10.1175/bams-d-12-00157.1>
- Zawadzki, I., & Antonio, M. D. A. (1988). Equilibrium raindrop size distributions in tropical rain. *Journal of the Atmospheric Sciences*, 45(22), 3452–3459. [https://doi.org/10.1175/1520-0469\(1988\)045<3452:ersdit>2.0.co;2](https://doi.org/10.1175/1520-0469(1988)045<3452:ersdit>2.0.co;2)