This is a preliminary PDF of the author-produced manuscript that has been peer-reviewed and accepted for publication. Since it is being posted so soon after acceptance, it has not yet been copyedited, formatted, or processed by AMS Publications. This preliminary version of the manuscript may be downloaded, distributed, and cited, but please be aware that there will be visual differences and possibly some content differences between this version and the final published version.

The DOI for this manuscript is doi: 10.1175/JAS-D-17-0242.1

The final published version of this manuscript will replace the preliminary version at the above DOI once it is available.

If you would like to cite this EOR in a separate work, please use the following full citation:


© 2018 American Meteorological Society
Primary modes of global drop-size distributions

B. Dolan*¹, B. Fuchs¹, S. A. Rutledge¹, E. A. Barnes¹
¹Dept. Atmos. Sci., Colorado State University

E. J. Thompson²
²Applied Physics Laboratory – University of Washington

August 2017
Revision 1: December 2017

*Corresponding author address:
Brenda Dolan, Colorado State University
E-mail: bdolan@atmos.colostate.edu
Abstract

Understanding drop-size distribution (DSD) variability has important implications for remote sensing and numerical modeling applications. Twelve disdrometer datasets across three latitude bands are analyzed in this study, spanning a broad range of precipitation regimes: light rain, orographic, deep convective, organized mid-latitude, and tropical oceanic. Principal component analysis (PCA) is used to reveal comprehensive modes of global DSD spatial and temporal variability. Although the locations contain different distributions of individual DSD parameters, all locations are found to have the same modes of variability.

Based on PCA, six groups of points with unique DSD characteristics emerge. The physical processes that underpin these groups are revealed through supporting radar observations. Group 1 (Group 2) is characterized by high (low) liquid water content (LWC), broad (narrow) distribution widths, and large (small) median drop diameters ($D_0$). Radar analysis identifies Group 1 (Group 2) as convective (stratiform) rainfall. Group 3 is characterized by weak, shallow radar echoes and large concentrations of small drops, indicative of warm rain showers. Group 4 identifies heavy stratiform precipitation. The low-latitudes exhibit distinct bimodal distributions of normalized intercept parameter ($N_w$), LWC, and $D_0$, and are found to have a clustering of points (Group 5) with high rain rates, large $N_w$, and moderate $D_0$, a signature of robust warm rain processes. A distinct group associated with ice-based convection (Group 6) emerges in the mid-latitudes. Although all locations exhibit the same co-variance of parameters associated with these groups, it is likely that the physical processes responsible for shaping the DSDs vary as a function of location.
1. Introduction

Understanding the variability of drop size distributions (DSDs) around the globe is important for remote sensing of precipitation, retrieving distributions of latent heating, and parameterizing microphysical processes in numerical models. Remote sensing retrievals routinely make assumptions about DSDs to relate observations to physical quantities (Munchak et al. 2012). For example, the DSD fundamentally determines the relationship between radar reflectivity $Z$ and rainfall $R$, with individual relationships varying with location and storm type (e.g. Battan 1973, Atlas et al. 1999, Ulbrich and Atlas 2007). Additionally, many bulk numerical model parametrizations employ two moment microphysical schemes where the number concentration and mass mixing ratio of each hydrometeor type are predicted and mean size is diagnosed, thus requiring that a fixed shape parameter ($\mu$) be assumed if a gamma distribution is used (e.g. Meyers et al. 1997; Saleeby and Cotton 2004). However, $\mu$ assumptions may introduce undesirable effects regarding the development of precipitation (Uijlenhoet et al. 2003, Milbrandt and Yau 2005).

DSD variability is determined by cloud-scale processes as well as environmental characteristics (Cotton et al. 2011). For example, in convection, strong updrafts can lead to supercooled liquid water in the mixed phase region, thereby promoting the growth of graupel and hail via accretion. Upon melting, large drops are formed which may fall out or undergo additional coalescence growth below the melting level. Evaporation may occur below cloud base, which preferentially removes smaller drops. In contrast, weak vertical motions in stratiform precipitation allow for ice crystals to grow initially via deposition followed by aggregation and possibly riming (Rutledge and Houze 1987; Houze 1997).
Previous studies have investigated DSDs in different locations around the world. Bringi et al. (2003; BR03) and Bringi et al. (2009; BR09) initially separated disdrometer data into convective and stratiform DSDs using a threshold on standard deviation of the rain rate (over five consecutive 2-minute samples). Distinct clustering was identified within the parameter space of mass weighted mean diameter ($D_m$) and normalized intercept ($N_w$). The clusters were identified as continental and maritime convection as well as stratiform rain. BR09 also found the same convective and stratiform clusters of DSD parameters when $D_m$ and $N_w$ were derived from polarimetric radar, leading to the rendering of a continuous separator line for segregating convective and stratiform rain. Previous studies have noted a significant overlap between convective and stratiform populations in Z-R space at low R, suggesting that a true separation between these rain types requires additional parameters (Yuter and Houze 2002, Atlas et al. 2000).

Thompson et al. (2015; T15) built upon these studies by investigating tropical oceanic rainfall, and found that while stratiform populations were similar to other regimes, convective precipitation had high $N_w$ but relatively low median drop diameters ($D_0$). This is in contrast with continental convective regimes, which can attain larger $D_0$ at lower $N_w$ (Atlas and Ulbrich 2000, BR03, Ulbrich and Atlas 2007, BR09, Thurai et al. 2010). For instance, ubiquitous shallow, weak convection identified by T15 over the tropical oceans did not conform to the BR09 convective-stratiform (C-S) separation line, which was derived from continental and coastal rain samples. T15 determined an updated C-S line for the tropical oceanic regime defined by constant normalized number concentration ($\log N_w$).

The goal of this study is to examine a larger, global disdrometer dataset to investigate DSD variability in space and time, and place some of the previous findings in a larger context.
(Most of the studies reviewed above focused on specific locations and/or regimes, thereby limiting generalizations of their findings.) To this end, we employ the statistical analysis technique of principal component analysis (PCA) to help interpret and understand trends in the data. PCA is a powerful tool for analyzing large and complex datasets because it yields the most significant modes of variability in a dataset without requiring any *a priori* information. PCA is commonly used in climate analysis to reveal spatial relationships or patterns in atmospheric quantities but it also been applied to investigate the variability of DSDs relative to environmental variables (Munchak et al. 2012). Hannachi et al. (2007) provides a detailed discussion on the various uses of PCA in atmospheric science.

This study describes the results of applying PCA to a large disdrometer dataset from diverse locations across the globe, ranging from low to high latitudes, including continental and maritime rainfall. The datasets and PCA methodology are outlined in Sec. 2. The different locations are compared and the results of the PCA are presented in Sec. 3. Polarimetric radar data are used to attribute physical processes to six distinct groups resulting from the PCA. The overall results are discussed and synthesized in Sec. 4.

2. Methodology

a. Disdrometer datasets

Recent campaigns by the Department of Energy (DoE) Atmospheric Radiation Measurement Program (ARM, Ackerman and Stokes 2003) and the National Aeronautics and Space Administration (NASA) Global Precipitation Measurement (GPM, Hou et al. 2014) Ground Validation (GV) program have greatly expanded DSD observations globally. We have compiled 12 disdrometer datasets from diverse locations and meteorological regimes, spanning the tropics to the high-latitudes that are supported by polarimetric radar observations (Table 1).
Two-dimensional video disdrometer (2DVD) datasets from tropical ocean (Manus Island, Gan Island), tropical coastal (Darwin, Australia), mid-latitude continental (Southern Great Plains, SGP), and high-latitude continental (Finland) locations were provided by the DoE ARM program. These datasets were collected in collaboration with several DoE field campaigns: TWP-ICE in 2006 (Darwin), AMIE in 2011 (Gan and Manus Is.), and BAECC in high latitude boreal forest (Finland). Collocated disdrometer and polarimetric radar datasets from coastal orographic high-latitude, continental high-latitude, mid-latitude orographic, and mid-latitude continental locations were provided by five recent NASA GV experiments (LPVEx-Finland, MC3E-Great Plains, IFloodS-Iowa, IPHEx-southern Appalachian Mountains, and OLYMPEx-Washington State). Additionally, a mid-latitude coastal disdrometer dataset was collected at NASA Wallops Island, VA. The locations of the datasets are shown in Fig. 1, and a summary of the data and experiments is given in Table 1. These datasets ranged in length from six weeks (LPVEx) to five years (SGP), and sampled from 2080 raining minutes (~35 hours; LPVEx) to 78124 raining minutes (~1300 hours; OLYMPEx). While the NASA GV experiments deployed several instruments to sample spatial variability, all raining minutes from all described instruments are considered together for each campaign. Field campaign data from the same geographic location and/or time are grouped with corresponding longer-term datasets resulting in eight datasets: IFloodS, SGP, IPHEx, Finland, Tropical Ocean, Darwin, OLYMPEx, and Wallops. For ease of discussion and analysis herein, data have been grouped by latitude (Fig. 1): high (≥ 45°), middle (25°-45°), and low (≤25°), based on the similarities between datasets in

---

1 In OLYMPEx, 13 Parsivels were deployed at various topographic heights. The entire dataset is over 300000 raining minutes, larger than all the other datasets combined. Therefore, we selected three APUs with minimal contamination from frozen precipitation (Fish Hatchery, Amanda Park, and Bishop CRN) for the analysis.
common latitude bands (not shown). This grouping results in approximately 90000 raining
minutes in the high and low latitudes, and 150000 minutes in the mid-latitudes, with each band
comprising both field experiments and long-term installations.

Most of the observations used are derived from a 2DVD (Schönhuber et al. 2008), which
use perpendicular lasers to image drops that fall through the square catchment area (100 cm²).
During TWP-ICE, a Joss-Waldvogel impact disdrometer was used (JWD, Joss and Waldvogel
1967). During OLYMPEX, five 2DVDs were deployed along with 13 Automated Parsivel Units
(APU, Battaglia et al. 2010). During OLYMPEX, the 2DVDs were found to under-sample the
number of drops due to high drop concentrations in this complex, orographic environment. For
this reason, we opted to use the APU optical disdrometers for OLYMPEX.

Following the methodology in T15, 1-minute data in every location were thresholded on
total number of drops (> 100) and rain rate (> 0.05 mm h⁻¹). This threshold removed on average
30% of the data but less than 1% of total rainfall (the drop count threshold was responsible for
removal of most observations). These thresholds help prevent small sample sizes from skewing
DSD estimates (Smith et al. 1993, Smith 2016). The 2DVDs directly measure liquid water
content (LWC, g m⁻³) and rain rate (R, mm h⁻¹), while an empirical fall speed relationship is used
for the JWD and APUs to obtain LWC and R (Tokay et al. 2001). A “deadtime correction”
(Sheppard and Joe 1994) was applied to the JWD data to account for recovery time of the
instrument transducer following drop impacts (Williams et al., 2000). Although it would be ideal
to have common platforms across all locations, the availability of these datasets necessitates
using three different instruments. Despite different limitations of each instrument such as limited
drop sizes (JWD), overestimates of drop size in heavy rain (Parsivels) and beam mis-match
(2DVD), previous studies have analyzed the comparative performance of JWD, 2DVD, and
Parsivels and generally found agreement in distribution fits and integrated parameters (e.g. Tokay et al. 2001, Thurai et al. 2011). We performed a sensitivity study (not shown) where analysis with co-located instruments (e.g. 2DVD and APU in OLYMPEEx) confirm that the results presented herein do not significantly change based on instrument choice.

The mass spectrum can be derived directly from disdrometer measurements (Williams et al. 2014):

\[ m(D) = \frac{\pi}{6 \times 10^7} \rho_w N(D) D^3 \]  

(1)

where \( N(D) \) is measured by the disdrometer as a function of the size bins. The rain DSD can then be described by the mean mass diameter \( D_m \), which is the first moment of the mass spectrum, and the mass standard deviation \( \sigma_m \), the square root of the second moment, which represents the breadth of the mass spectrum. For more details of this formulation, see Williams et al. (2014).

Rain DSDs have often been described with a modified gamma distribution (Ulbrich 1983), which is nominally defined by an intercept, median size, and shape factor that describes the breadth and slope of the distribution. A special case of the four-parameter modified gamma distribution is the normalized gamma with three free parameters (Petty and Huang 2013). The normalized gamma distribution accounts for varying LWC (Willis 1984), and is described by the intercept parameter \( N_w \) (m\(^{-3}\) mm\(^{-1}\); herein we will use \( \log N_w \) for ease), median drop diameter \( D_0 \) (mm), and shape parameter \( \mu \):

\[ N(D) = N_w f(\mu) \left( \frac{D}{D_m} \right) \mu \exp \left[ -(4 + \mu) \frac{D}{D_m} \right] \]  

(2)

where \( N_w \) is defined as:

\[ N_w = \frac{3.67^{4}10^{3}LWC}{\pi \rho_w D_0^3} \]  

(3)
Additionally, $D_m$ is related to the median drop diameter through $\mu$:

$$D_m = \frac{4+\mu}{3.67+\mu} D_0$$ (4)

Although the assumption of a normalized gamma fit to the DSD may not fully capture the true spectrum of rain DSDs (Thurai et al. 2017), we employ it here to place our results in context with the wide body of previous literature that invoked this assumption.

Further details of the gamma distribution can be found in Bringi and Chandrasekar (2001). To derive the parameters of the normalized gamma DSD from each minute of disdrometer data, the Thurai et al. (2014) methodology was adopted which fits the gamma parameters through a $\mu$-search technique. For further information about DSD formulations and processing, see Bringi and Chandrasekar (2001), Thurai et al. (2014), and T15. Data were restricted to $\mu$ values in the range -4 to 15 to ensure the gamma fit is a reasonable assumption. This screening process removed $< 1\%$ of data points in our dataset.

Several datasets included times when frozen precipitation was present (e.g. LPVEx, OLYMPEX, SGP, Finland). For the shorter GV field projects with multiple 2DVDs, instruments and dates were excluded from analysis when snow-contamination was identified (LPVEx, OLYMPEX; A. Tokay and J. Zagrodnik personal communication, 2016). However, for the longer datasets (SGP, Finland) a more automated method was developed to discard snow-contaminated observations. A distinct population of data exhibited lower-than-expected $R$ for a given LWC. Since $R$ is fundamentally related to the fall speed and size of each drop, this indicates a population of particles with fall speeds lower than that of rain (e.g. snow; Yuter et al. 2006). Thus, any days that had 10 or more points that fell below $R = 9.0$ LWC$^{1.1}$ (determined subjectively based on the anomalous population and examination of those days against ASOS observations of snow) were excluded from the analysis to remove snow contamination.
b. Principal Component Analysis (PCA)

PCA is a technique often used in atmospheric science to simplify the analysis of large and complex datasets. In essence, PCA uses linear regression to explain the main modes of variability of a dataset. Hence the variability of a dataset is distilled into its most important components. PCA can be thought of as a type of pattern or cluster analysis that explains the covariance of parameters simultaneously. Additionally, PCA is empirical in that the results are based solely on the dataset. For example, in the BR03 study, convective and stratiform DSD were stratified based on rain rate. In PCA, such assumptions are unnecessary.

PCA results in a set of vectors (also called empirical orthogonal functions or EOFs) forming an orthogonal set of basis vectors and are ordered by their ability to explain the variability in the dataset. The first EOF is the vector that explains the largest amount of variance. The second EOF is orthogonal to the first EOF and describes the largest fraction of the remaining variance, after the variance from the first vector has been removed. This process continues until the collective EOFs explain all of the variance (including noise). For example, if there are six input parameters, there will be six resulting EOFs, which will collectively explain 100% of the variance in the sample dataset. In practice, some studies may concentrate on the first few leading EOFs, since successive EOFs may yield little additional physical interpretation. For these reasons, we only present the leading two EOFs for our analysis. A linear combination of the EOF vectors describes each point in the dataset, the coefficients of which are known as the principal components (PCs). The PCs can be thought of as a measure of the resemblance between a particular data point and an EOF vector.

To prepare the disdrometer data for PCA, we constructed M (number of attributes or quantities describing the DSD) arrays of length N (corresponding to the number of DSD data
points). For the PCA analysis we selected $D_m$, $N_w$, and $\sigma_m$ to describe the DSD, and the integral rain parameters $R$, $LWC$ and $N_t$ (total number of drops) to describe the DSD variability. These parameters are selected because they are commonly used to describe precipitation, they can be easily calculated from disdrometer measurements, and they provide meaningful information toward understanding the physics of rainfall. We have chosen to use the mass spectrum width $\sigma_m$ and the mean mass diameter $D_m$ instead of the gamma shape parameter $\mu$ and the median drop diameter $D_0$ because the former two variables can be calculated from the disdrometer measurements without the need to assume a gamma distribution. We use $N_w$ because it provides information about how the LWC is distributed across diameter space (Eq. 3). The covariance matrix among the six parameters for each latitude band and the global dataset are available in the supplemental materials (Fig. S1). The parameters of $N_w$, $N_t$, $R$, and $LWC$ are log-normally distributed (e.g. Bringi and Chandrasekar 2001, T15), and are therefore are included in the PCA in log form ($\log N_w$, $\log N_t$, $\log R$, $\log LWC$). The data are normalized to standard anomalies of each characteristic quantity by subtracting the mean and dividing by the standard deviation. This was done to prevent quantities with large variances from dominating the EOFs, which could obscure scientifically relevant results. Additionally, deviations from the mean indicate when values are anomalously high or low, making simultaneous interpretation of multiple quantities easier. Once each quantity is standardized, the arrays are combined into a matrix of $N$ rows x $M$ columns. The PCA returns $M$ orthogonal EOF vectors, with corresponding values of fractional variance explained, as well as $N$ PC values (corresponding to each data point) for each EOF vector.

The orthogonality constraint imposed by PCA can cause problems in certain situations. First, information about a particular process may be included in multiple EOFs because physical
processes are not necessarily orthogonal. Second, each PC series can have positive and negative values. Since the sign of each EOF is arbitrary, each mode of variability has exactly two opposite components. This can complicate physical interpretation of PCA results since physical processes are not always linear and orthogonal. Nonetheless, PCA provides an objective lens through which we can look at a large disdrometer dataset to gain novel insights about precipitation formation processes.

3. Results

a. Comparisons of datasets

We begin by examining the variability of DSDs across latitude bands projected in logNw-D0 space (Fig. 2a-c). The BR09 and T15 C-S separation lines are included for context. In all locations, the most frequent values are centered near D0 =1 mm, with the median logNw higher in the high latitudes (3.8 - 4), and lower in the mid-latitudes (3-3.5). The mid-latitudes have broader ranges of D0 and Nw compared to high and low-latitudes. The characteristic bimodality in the tropics of logNw is evident, which was noted by previous studies. The tropical distribution of logNw peaks in terms of frequency of occurrence at 3.4 and also at 4, and these maxima correspond to convective and stratiform populations (Ulbrich and Atlas 2007, BR09, Thurai et al. 2010, Bringi et al. 2012, T15). Indeed, this distinction serves as the basis for the T15 C-S separation line. We find it curious that the tropical datasets exhibit such a distinct bimodality in logNw and D0 associated with C-S rain. This bimodality is lacking in other datasets such as the mid-latitudes, despite significant convective and stratiform precipitation components. In fact, the T15 line appears to cut through the mode of the logNw-D0 distribution in the high latitudes, suggesting this method is only appropriate for characterizing C-S precipitation within the tropical
oceans. It is possible that the intense ice-based precipitation in the mid-latitudes somehow disrupts the clear separation seen in the tropics, where ice processes are likely weaker, or play a lesser role in overall precipitation. Thus, the oceanic convection is size-limited (e.g. constrained D₀) but dependent on LWC, which directly drives logNₖ (Eq. 3). Another striking feature of the tropics is the moderate density of points clustered around D₀ = 1.5 mm and logNₖ = 4-5, a population that is not readily evident in the other locations. This is associated with a peak in LWC > 1 g m⁻³ at about D₀=1.7 (Fig. 2f). Again, the tropics have two distinct clusters in D₀-LWC space that have been noted in many previous studies (Tokay and Short 1996, Yuter and Houze 1997, T15) as being associated with convective (upper branch) and stratiform (lower branch) precipitation. The mid-latitude datasets have a noticeable extension of large mean diameters for moderate LWC (Fig. 2e). The high latitudes rarely exceed 1 g m⁻³, with the most frequent values being 0.1-0.5 g m⁻³ (Fig. 2d). Interestingly, these median values are higher than the most frequent values in the other locations (Fig. 2d-f), which could be due to the atmospheric rivers affecting the OLYMPEx region (which dominate the high-latitude sample) bringing ample moisture to the domain (Houze et al. 2017).

It is obvious from Fig. 2 that DSDs vary by location (individual locations shown in supplemental material Fig. S2-S3), and as in the case of C-S separation lines, there may be important underlying microphysical processes that are responsible for this variance. It is also clear that no one location fully captures the entire spectrum of global DSD variability. Given this regional DSD variability, we are motivated to ask: What are the primary modes of temporal and spatial DSD variability across the globe? Can different surface DSDs be explained and linked to microphysical processes shaping them aloft? We explore these questions next.

b. PCA Results
Fig. 3 shows the first two primary EOFs associated with the DSD characteristic parameters 
(logNw, Dm, σm, logLWC, logR, and logNt). The standard anomaly associated with each 
parameter for each of the resulting EOFs is indicated. Importantly, despite some spread in the 
magnitude of the standard anomalies, every location shows the same modes of co-variance in the 
first two EOFs. Although we present the aggregate of datasets in a given latitude band, this 
finding holds true for every individual dataset (supplemental material Fig. S4). Again, we note 
that the sign of each PC is arbitrary, and that mathematically the positive and negative modes are 
extact opposites. We have neglected the EOFs beyond EOF 2, which collectively explain a much 
smaller amount of variability and may relate to measurement noise or higher-order temporal 
variability rather than underlying physical processes (Larsen and O’Dell 2016). It is important to 
note that the first two EOFs explain ~88% of the variability across all latitude bands (Table 2). 
The positive mode of EOF 1 is associated with relatively high logLWC and logR, large logNw, 
large D0, and high σm. This EOF explains 57-62% of the variability in each location, and 58% 
globally (Table 2). EOF 2, which explains 20-31% of the variability (30% globally), is 
characterized by relatively large logNt and logNw, small D0 and σm in the positive mode, but 
relatively little variability in logLWC or logR.

In order to equally compare resulting EOFs from different locations, each of which may or 
may not capture the full breadth of DSD variability, we normalize all data against the global 
dataset (i.e. all data combined) by subtracting the global mean and dividing by the standard 
development. All subsequent results are shown relative to this global normalization and global 
EOFs.

It is most useful to examine these orthogonal modes of variability by investigating data 
points that are most characteristic of each mode. This is accomplished by isolating points with
the largest PC values, which represent where the points project most strongly onto an EOF. To
this end, unless otherwise indicated, we select a threshold of $|1.5| (\sigma_{1.5})$ for all PCs. The selection
of points which do not meet these criteria will be referred to as ‘ambiguous’ because they bear
resemblance to multiple EOFs. Here we note that these threshold values are somewhat arbitrary
and the high thresholds used herein have been selected to highlight differences between
populations, but the overall results do not change with the prescribed threshold; larger values
result in more distinct separation between opposing modes, while smaller values result in more
overlap.

The distributions of PC1 and PC2 for each latitude band are shown in Fig. 4. The $\sigma_{1.5}$
thresholds are illustrated with dashed lines. While the bulk of points (~77%) are contained below
the thresholds (by design), 23% of points meet one or more of these thresholds; in fact, there is a
noticeable population with large absolute PC1 and PC2 values. Leveraging these PC
distributions, we define six distinct groups. Illustrated by colored boxes in Fig. 4, these groups
are: positive PC1 but ambiguous PC2 (Group 1, red), negative PC1 (Group 2, green), positive
PC2 but ambiguous PC1 (Group 3, yellow), negative PC2 and ambiguous PC1 (Group 4, blue),
positive PC1 and positive PC2 (Group 5, orange) and positive PC1 and negative PC2 (purple).

Already, clear latitudinal differences emerge from this PC-based grouping (Fig. 4). For
example, the high latitudes have a large number of points in Group 3, but few points in Group 5,
Group 1, or Group 6 at the $\sigma_{1.5}$ threshold. In contrast, the low-latitudes have a number of points
in both Group 1 and Group 5. The mid-latitudes have the largest number of points in Group 6 of
any latitude band. The relative percentages of each group in each band is shown in supplemental
material (Fig. S5). We will probe the characteristics of each of these six populations in the next
section to understand their potential origins from a microphysical perspective and possible implications.

c. DSD parameters

Here we will cast the results of the PCA into different formulations and 2D distributions for physical interpretation and comparison with previous studies. For example, B09 and T15 found distinct populations in log\(N_w\)-\(D_0\) and LWC-\(D_0\) space, which they attributed to convective and stratiform rainfall.

Mean gamma DSD, determined by inserting the mean values for \(N_w\), \(D_0\), and \(\mu\) into Eq. 2, for each of the six groups identified above contrast the characteristic size distributions associated with the associated co-variance among the DSD parameters (Fig. 5a). Fig. 5b illustrates the normalized (Sec. 2b) mean values of different parameters for each of the six groups. For reference, the global mean values of LWC, \(R\), \(N_t\), \(\sigma_m\) and the gamma parameters (log\(N_w\), \(D_0\), \(\mu\)) are annotated in Fig. 5a, and mean values of the gamma parameters are listed in Tables 3-5.

Group 1 (red) is associated with \(R\) significantly larger than 4 mm h\(^{-1}\) (the global mean), shape parameters \(\mu < 4\), and mean drop diameters \(D_0 > 1.1\) mm, resulting in a broad size distribution. In contrast, Group 2 (green) is characterized by narrow size distributions (Fig. 5a) and relatively low \(R\) and LWC (Fig. 5b). Both Groups 4 (blue) and 6 (purple) have strong negative anomalies in \(\mu\) (that is, smaller than the global mean of 4), resulting in more exponential-type (broad) size distributions (Fig. 5a). While having anomalously large \(D_0\) and low log\(N_w\), Group 6 also has anomalously high log\(R\), logLWC, and drop concentrations (log\(N_t\)). Here we recall that \(N_w\) is directly proportional to LWC but inversely proportional to \(D_0^4\) (Eq. 3). The resulting anomalously low log\(N_w\) is due to the exceptionally high \(D_0\) anomalies even though the total number of drops and LWC are high. Group 3 has a high log\(N_w\) consistent with the smallest mean
mass-weighted diameter, indicating a population of numerous small drops with little size
dispersion. This population also has logLWC and logR values near the global mean. However,
DSDs in Group 5 (orange) have the highest anomalies of logR and logLWC, greatest total
number of drops, and highest logN_w.

These PCA-guided groups result in distinct clusters in 2D parameter space (Fig. 6-7). Again,
we note that although we only display the aggregate results for the three latitude bands, these
clusters are repeatable in every dataset considered (see supplemental material Fig. S6). Perhaps
most striking, but maybe not unexpected due to the orthogonality of EOF analysis, is the
distinction in logN_w-D_0 space (Fig. 6 a-c). Group 1 (red) resides in the moderate to high D_0 and
moderate to low logN_w, but it is clear this group contains the highest LWC observations in our
dataset, along with Group 5 (Fig. 7a-c). This population of points is associated with relatively
low, even negative, values of μ (Fig. 6, d-i), indicating broad size distributions (Fig. 5a). The
surprising conformity of this cluster in the tropics to the robust C-S lines proposed by T15 and
BR09 underpins Group 1 as convective. Conversely, the points in Group 2 (green) are confined
to relatively small D_0 values and low logN_w due to the combined effect of low LWC and small
D_0. The bulk of the population falls in the stratiform classification for both BR09 and T15 C-S in
the tropics. Additionally, this cluster has generally high μ values (>6), indicating relatively
narrow size-distributions (Fig. 6d-f, 5a), or perhaps an indication that a normalized gamma
distribution is inadequate to describe the distribution of these particular samples (Thurai et al.
2017).

Group 3 (yellow) points are largely restricted to logN_w values larger than 4.5 due to a wide
range of LWC but small D_0 (Fig. 7a-c), with the bulk of the population having small (< ~1.0
mm) D_0 (Table 3). Group 3 points span nearly all μ (Fig. 6d-i, Fig. 7 g-i), especially in the high
latitudes, but the majority are larger than $\mu=5$ (Table 5). The Group 3 population falls in the region of the spectrum that both BR09 and T15 attribute to weak, shallow radar echoes. Group 4 (blue) is associated with large $D_0$ (> 1.5 mm, up to 3 mm) and low $\log N_w$ (< 3.5, as low as 1), which is a function of both the large sizes and the relatively low LWC (Fig. 7). Group 4 has $\mu$ values generally < 3, centered around -1 to 0 (Fig. 6d-f, Tables 3-5).

The two groups with large absolute values of PC1 and PC2 (Group 5 and Group 6, orange and purple respectively) have notable properties. Group 5 is more common in the tropics compared to the mid-latitudes and high latitudes. Group 5 is characterized by the largest total number of drops and the largest LWC despite $D_0$ near the global mean of 1.13 mm. These points are exclusively above the BR09 and the T15 convective lines in the tropics. In contrast, the mid-latitudes have the most points in Group 6, which consist of relatively high LWC and moderate $N_t$, but very large sizes, leading to low $\log N_w$ values (Fig. 6b). This group also has the lowest $\mu$ due to the large number of drops and large mean $D_0$.

d. Radar Observations

We have demonstrated distinct clustering of surface DSD parameters and speculated about possible physical processes producing these distributions. To further investigate the possible microphysical processes contributing to the DSDs in each of the six Groups, we examined the vertical structure of radar echoes over specific disdrometer locations using polarimetric radar observations. In each case, Range Height Indicator (RHI) and/or Plan Position Indicator (PPI) scans were used to geolocate the radar gates above the disdrometer locations, creating a vertical profile over the disdrometer. These vertical profiles were then time-matched to the disdrometer PCA Group classifications. This analysis provides an independent method to
study the characteristics of each Group because the radar and the disdrometer are separate measurements.

Mean reflectivity ($Z$) and differential reflectivity ($Z_{dr}$) profiles were created from RHI data for each disdrometer PCA group (Fig. 8). It should be noted that for the radar analysis in this section, the PC thresholds have been lowered for illustration purposes. That is, the matched radar-disdrometer times may not have many points in that meet the $\sigma_{1.5}$ thresholds, since only 22% of points in the entire disdrometer database meet these thresholds. In order to make this problem tractable, we will focus on specific locations: OLYMPEEx to represent the high latitudes; MC3E, IFloodS, and IPHEx for the mid-latitudes; and Gan Is. and Darwin for the tropics, all providing coincident polarimetric data over nearby disdrometers (Table 6). The disdrometer PC values spanning five minutes around the radar RHI time were averaged and used to assign the radar profiles to a particular Group. Six RHI gates over each point are averaged, and RHI gate heights were interpolated to a common set of heights for comparison. To account for clutter at the lowest radar bins, the heights start at 0.2 km AGL. The analysis resulted in 1466 (mid latitudes), 1336 (low latitude), and 3320 (high latitude) disdrometer / radar matches with precipitation (Table 6). Mean $Z$ and $Z_{dr}$ for each match were smoothed with a Gaussian filter in height. In each profile, the 0 dBZ echo top height (ETH) was calculated and plotted against the reflectivity at 2.5 km (Fig. 8g-i) as an estimate of the mean intensity and depth of precipitation in each group. We recognize the complexities of linking vertical information to surface observations due to advection and point-to-volume comparisons, and these effects may dilute the signal in the analysis below. However, we have attempted to minimize these effects through spatial and temporal averaging of the radar and disdrometer data.
Although the systematic increase in ETH with decreasing latitude is evident, (perhaps due to the annual mean tropopause height; Price and Rind 1993), there are clear trends within each group across latitudes (Fig. 8). Group 1 and Group 6 have the highest mean reflectivity values (for all latitudes) throughout the depth of the column, and significant ETH. Group 2 and Group 3 have the lowest mean reflectivities and lowest ETH. Mean ETH in Group 3 remain below 5 km in the high and mid latitudes and mean reflectivities are below 20 dBZ in all latitude bands, suggesting a designation of shallow, weak warm rain showers for this group. Group 2 has similar mean Z and ETH values to Group 3, and a relative maximum in \(Z_{dr}\) near the environmental freezing level (e.g. the radar bright band signature). We suggest that this group is likely associated with weak vertical motions indicative of stratiform precipitation, where particles grow by vapor deposition and aggregation and then melt as they fall below the 0°C level (Houze 1997) or, are detrained from deep convection and continue to grow in mesoscale updrafts as they fallout (Rutledge and Houze 1987). These processes fundamentally limit the maximum drop diameters and rain rates (Yuter and Houze 2002). Group 4 has a very distinct signature, especially in the mid-latitude samples. Group 4 has high ETH and modest peak reflectivities (<30 dBZ). At sub-freezing temperatures, radar reflectivities increase toward the melting layer while \(Z_{dr}\) is slightly positive aloft and then decreases toward the melting layer, reaching a minimum near 0 dB just above the environmental melting layer. While this is behavior is evident in all mid-latitude \(Z_{dr}\) profiles, it is most notable in Group 4. This is consistent with small oriented ice crystals aloft (Kennedy and Rutledge, 2011; evident by weak reflectivity and positive \(Z_{dr}\)) that begin to aggregate to form larger particles that increases reflectivity and decreases \(Z_{dr}\) to near zero (associated with very low bulk density and small degrees of oblateness). However once these particles begin to melt, they produce a sharp increase.
in both reflectivity and $Z_{dr}$ (e.g. the radar bright band signature). We postulate that Group 4 is consistent with deposition-aggregation-driven stratiform precipitation. There were too few coincident RHIs in Group 5 in the mid-latitude samples (<20), so they are not shown. At Gan (OLYMPEEx), the mean Group 5 ETH reach 8 (5) km, compared to the deeper Group 1 and 6, at ~9 (6) km. In both locations, $Z$ values are larger than Group 2 or 3, particularly below the melting layer. In the low-latitudes, $Z_{dr}$ remains constant or increases slightly below the melting layer while $Z$ increases, a signature of robust collision-coalescence in the warm layer (Kumjian and Pratt 2014). The Group 5 signature in OLYMPEEx is associated with decreasing $Z_{dr}$ in the shallow warm layer. This signature has been attributed to drop breakup (Kumjian and Pratt 2014), however this mechanism is unlikely in the OLYMPEX regime due to the short fall distance of drops and generally small drop sizes. Therefore the physical processes underpinning Group 5 in the high latitudes is unclear. The $Z_{dr}$ profiles of Group 6 exhibit increases below the melting layer, especially in the mid-latitudes, which had the largest number of samples in this Group. The high ETH and large reflectivities above the melting layer are consistent with strong convective vertical motions, and are similar to the median convective vertical profiles of mid-latitude convection illustrated in Carr et al. (2017). The larger $Z$ throughout the column and warm-layer $Z_{dr}$ values of Group 6 is consistent with a signature of ice-based convection and melting below the 0$^\circ$ C layer. The increasing $Z_{dr}$ and $Z$ toward the surface suggests continued coalescence growth below the melting layer (Kumjian and Pratt 2014).

To put these mean samples into a larger spatial context, we examined time-height series for several cases in each latitude band. AMIE/Gan and OLYMPEX experiments both employed high temporal resolution RHI scanning, so time-height series were created by averaging 30 gates in range at each disdrometer gate for each elevation over the disdrometer point. Time-height
21

series for IFloodS and Darwin used PPI data by averaging 10 gates in the azimuthal direction and
30 gates in range over the disdrometer location for each elevation, similar to the quasi-vertical
profile (QVP) method described by Ryzhkov et al. (2016) but averaged over a smaller area.
Representative reflectivity time-height series for each latitude band for different groups are
shown in Figs. 9-11.

An example from Darwin on 16 January illustrates a region of stratiform precipitation
followed by a convective cell passing near the disdrometer at 1205 UTC (Fig. 9a). The majority
of the time series is classified as Group 4 by the disdrometer during clearly stratiform
precipitation, evidenced by a persistent bright band around 5 km. The passage of the convective
core results in a brief classification of Group 1. A second tropical example from Gan Is. shows
widespread stratiform region with continuous bright band evident at the environmental melting
level (~5 km). As an embedded convective element passes by around 0420-0450 UTC (Fig. 9e),
reflectivities are enhanced below the bright band, and the Group switches from 1 to 5 as PC2
values increase (Fig. 9g). After this time, the PCA is ambiguous before switching to Group 6 as
the heavy stratiform sets in (Fig. 9e), and finally Group 4 and 2 as the stratiform echo weakens.
This example is also illustrated in T15 and shows the transition from convective (Group 1 and 5)
to stratiform (Group 2 and 4). The last tropical example, also from Gan Is., shows what is termed
“weak, shallow convection” by T15. Reflectivities are largely below 25 dBZ and echo top
remains below the melting layer in an area of weak echo (Fig. 9i). The disdrometer PCA shows
small PC1 and positive PC2, mostly classified as Group 2 and 3.

The first example from IFloodS, representing the mid-latitudes, shows the passage of an
intense deep convective core, with peak reflectivities > 50 dBZ and 40 dBZ echoes reaching 10
km, followed by a fledgling stratiform region (Fig. 10a). As the core passes over the disdrometer,
large PC1 values result in a Group 1 classification, followed by a narrow band of Group 6 on the back edge of the core, then Group 4 and finally Group 2 as the stratiform sets in after 1015 UTC. The second IFloodS example (Fig. 10e) shows a small, shallow convective cell passing around 1840 UTC, followed by a region of ‘lumpy stratiform’, with a clear bright band signature but some enhanced reflectivity above the bright band between 1855 and 1955 UTC. The small convective cell with echo heights only reaching 6 km and reflectivities ~ 35 dBZ is associated with a limited number of Group 5 identifications. Group 4 is identified where the bright band has strong reflectivities (> 30 dBZ), and the most intense bright band with reflectivities approaching 40 dBZ is identified as Group 6 and even some Group 1. As the stratiform weakens after 1955 UTC, the disdrometer PCA identifies mostly Group 2. The last example also illustrates a period of “lumpy stratiform”, with a clear bright band signature but enhanced reflectivity aloft (Fig. 10i). Group 1 is identified early in the time period, but switches to Group 6 as 30 dBZ reflectivities reach 6-8 km accompanied by large reflectivities in the bright band (> 45 dBZ). As the brightband and reflectivities aloft weaken, the classification switches to Group 4 and some Group 2 (both identifying stratiform precipitation).

The last examples illustrate the unique coastal mountain high-latitude regime in OLYMPEX. Group 1 and Group 6 are identified during a period of strong low-level reflectivities, upwards of 40 dBZ below 3 km, the approximate height of the environmental melting layer during this time. In a 2 km deep layer above the melting level, reflectivities are up to 35 dBZ, and echo tops are ~ 8 km (Fig. 11a). A contrasting example (Figs. 11e) illustrating mostly Group 6 and some Group 4 shows shallower echo tops (around 6 km), but similar strong reflectivities in the warm layer (below 2 km in this case), and some more intense periods with reflectivities above 45 dBZ (e.g. 1145 UTC). These intense periods are classified as Group 6. In contrast to
the first OLYMPEEx example, reflectivities fall off near the surface. As the reflectivity aloft 
weakens after 1250 UTC, Group 4 and Group 2 are evident. Finally, a period of weak, shallow 
reflectivity is illustrated by the example in Fig. 11i. Echo tops are lower in comparison to the 
previous two cases (< 4 km), and reflectivities are limited to < 30 dBZ. The majority of these 
points are classified as Group 3, similar to the weak, shallow echoes accompanying Group 3 in 
the tropical example (Fig. 9g).

These specific examples serve to support our hypothesis that Group 1 encompasses 
convection, with stronger reflectivities and deeper echoes (Fig. 9a, b, 10a, 11a). The 
hypothesized stratiform nature of Group 2 is also supported by these examples, as this group is 
generally identified by the disdrometer when radar echoes display a weak to moderate bright 
band and reflectivities remain below 30 dBZ (Fig. 9a, 10e, 11i). Radar examples support the 
hypothesis that Group 3 is associated with weak, shallow echoes (Fig. 9i, 11i), while Group 4 
appears to be associated with more intense stratiform, where bright band reflectivities are > 35 
dBZ (Fig. 9a, e, 10e, i 11e). According to the radar analysis, Group 5 appears when reflectivities 
in the warm layer are significant (>30 dBZ), suggesting the presence of robust coalescence 
processes (Fig. 9e). Group 6 is generally associated with the most intense reflectivities, 
sometimes on the edge of a convective core (Fig. 10a), sometimes in strong, lumpy stratiform 
(Fig. 10e, i, 11a, e), suggestive of larger melted ice particles.

4. Discussion and Summary

We have presented an objective PCA-based framework for examining the spatial and 
temporal variability of DSDs and applied it to a large dataset spanning the deep tropics to the 
high latitudes. We used PCA to examine comprehensive modes of variability between quantities
describing DSDs derived from disdrometers. Importantly, the leading two EOFs revealed the same co-variance in all datasets (Fig. 3), despite differences in breadth of individual DSD parameters (Fig. 2). Based on the PCA analysis, six different groups of points with similar DSD characteristics were designated. However, the physical nature of these six DSD populations varied as a function of latitude, indicating that the processes contributing to the formation of each Group may have different underlying physics.

In the low latitudes, Group 1 and Group 2 (roughly associated with the positive and negative modes of EOF 1 respectively), showed surprising conformity to the previously identified convective-stratiform lines proposed by BR09 and T15 (Fig. 6a-c). Group 1 is exclusively associated with R > 10 mm h\(^{-1}\) at the prescribed \(\sigma_{1.5}\) thresholds (Fig. 7d-f), and are characterized by significant ETH and strong reflectivities above the melting layer (Fig. 8), further supporting the convective nature of this population (Tokay et al. 1999, Atlas et al. 2000, Yuter and Houze 2002). The low latitudes have a pronounced population of points with moderate mean drop sizes (1.5-2 mm) and large \(\log N_w\) (> 4) which mathematically emerge from PCA as having both high PC1 and high PC2 values (Group 5), a feature that is suppressed in the mid- and high latitudes. These points yield some of the highest rain rates, have the highest drop concentrations, exhibit high reflectivities in the warm layer (Fig. 9,10) and have ETH above the environmental melting level, but below Group 1 and 6 (Fig. 8g-i). We postulate this characteristic DSD is associated with warm rain processes and prolific collision-coalescence, where large LWC and deep warm cloud depths facilitate collision-coalescence growth. In low R (<10 mm h\(^{-1}\)), many previous studies have grappled with the difficulty in separating convective and stratiform processes due to the similarities and overlap in DSD parameters in 2D space (Tokay and Short 1996, Yuter and Houze 1997). This is evident in the overlap between Group 2
and Group 3 (Fig. 6-7), which have similar ranges of µ and D₀ values. However, PCA allows for
covariance in multiple dimensions, thereby providing a more robust separation between Groups
2 and 3 through differing N₇, LWC, and N₁ (Fig. 5b). Group 3 is associated with a large
population of points with very small median sizes, but numerous drops. Radar analysis revealed
that the radar echoes producing the surface DSD observations for Group 3 are shallow and weak
in all three latitude locations. Perhaps in these cases the physical depth of the cloud is
insufficient to support coalescence (Berry and Reinhardt 1974, Cotton et al. 2010, Lau and Wu
2003).

In the mid-latitudes, a significant number of points are associated with large D₀ but low
logN₇, which fall into Groups 4 and 6. However, while Group 4 has R and LWC values that do
not stray far from the global mean of ~ 4 mm h⁻¹ and 0.23 g m⁻³, Group 6 has comparably large
deviations of these quantities, indicating significant LWC coupled with large drops (D₀ up to 5
mm). This could indicate convective ice processes and/or continued coalescence in the warm
cloud within Group 6. Group 1 in the mid-latitudes has a lower mean logN₇ and higher mean D₀
compared to the low latitudes (Table 3-4), consistent with the maritime and continental regimes
described by BR03. Radar analysis revealed larger mean reflectivities aloft and higher ETH, as
well as very large (1.5 dB) mean Zdr (Fig. 8).

The high latitudes are notable for the significant population of points with small D₀ (0.68
mm), high logN₇ (4.81), and high µ (6.75) (Table 3-5). Many of these points are captured by the
variability distinguished in Group 3. It is clear that the T15 C-S separation line developed for the
tropical ocean is not applicable to the high latitudes (nor was it intended to be), as it cuts directly
through the highest density of points in logN₇-D₀ space in the high latitudes (Fig. 2c). Group 5 is
notably absent, while Groups 1 and 6 are limited in logN₇-D₀ space, suggesting that there may
There are differences in the microphysical processes shaping the surface DSD despite sharing commonalities with the convective populations in the mid and low-latitudes.

These DSD groupings have important implications for estimating rain rates from radar, as they result in a spectrum of Z-R relationships (Fig. 7d-f). Group 2 features modest Z but the smallest R values, generally < 1 mm h\(^{-1}\) (at the \(\sigma_{1.5}\) threshold). There is overlap with Group 3, which tends to have larger R and larger Z due to the increased LWC and drop concentrations. Interestingly, Group 1 points are associated with R > 10 mm h\(^{-1}\), a threshold considered to be exclusively convective precipitation based on earlier studies (e.g., Atlas et al. 2000, Yuter and Houze 2002). Most of the Group 3 points have R < 10 mm h\(^{-1}\), although Group 5 yields some of the highest R values along with Group 1. It is interesting to note that Group 3 spans a relatively narrow range of Z-R relationships, particularly in the tropics (Fig. 7f), and similarly for Group 5. On the contrary, Group 1 does not follow a distinct Z-R relationship, but is associated with both high Z and high R in all latitude bands. Group 4 and 6 have much lower R for a given Z compared to the Group 3 and 5 distributions. Interestingly, there is increased scatter between Z and R with increasing latitude, and there is a larger fraction of ambiguous points which do not meet the \(s_{1.5}\) threshold, particularly for the light rain rates. It is possible that additional modes of DSD variability may be explained by the higher order EOFs that are not described by the six groups derived from EOF1 and EOF2. It is also possible that more data from high latitude, light rain rate conditions are needed in order to account for DSD variability in this regime. Thus, our analysis reveals important information about how the DSD variability relates to radar-based rainfall retrievals.

To summarize our findings, we present a conceptual model based on our Group determination in \(N_w-D_0\) space (Fig. 12). Group 1, Group 3, Group 5 and Group 6 are
characterized by *convective precipitation processes*, where particles are assumed to grow as they are lifted by a convective updraft, then after sufficient growth, fall back through the updraft growing to even larger sizes. We note that convection is a continuum, where Group 5 represents convection dominated by warm rain processes, such as where warm cloud depths are especially deep, supporting enhanced growth by collision and coalescence. Group 6 is dominated by ice-based convective precipitation, indicated by large $D_0$ and low $\log N_w$, where strong vertical motions support robust mixed-phase production of graupel and possibly hail, leading to large raindrops at the surface owing to melting. Group 1 has some components of ice-based and warm rain growth processes. The progression from warm rain to ice-based processes in the convective group also follows the maritime / continental trends noted in previous studies (e.g. BR03, Ulbrich and Atlas 2007, BR09, T15). Groups 1, 5 and 6 generally conform to the C-S separation proposed by BR09 (Fig 6a-c). Group 3 is also convective in nature, with numerous but small drops, shallow ETH, and generally weak reflectivities corresponding to warm rain showers. In the tropics, this group is associated with weak convective motions, but in a topographically forced location such as OLYMPEx, this may be a signature of orographic enhancement. Groups 2 and 4 are *stratiform precipitation processes*, with increasing $D_0$ and decreasing $N_w$ being correlated with bright band intensity, representing an evolution from melted vapor-grown particles to aggregation and riming processes. The light rain rates and small drop diameters associated with Group 2, as well as the vertical profile of $Z$ and $Z_{dr}$, indicate this group is weak stratiform precipitation with small to modest sized ice particles grown by vapor deposition entering the melting layer leading to bright band reflectivities only reaching $20-25$ dBZ. Group 4 has much larger drop sizes but relatively low $\log N_w$ (and also comparatively low drop concentrations), accompanied by a distinct radar-based signature of aggregation aloft, especially
in the IFloodS data. In this group, small ice particles form aloft, undergo growth by deposition, followed by aggregation. Distinct and sharp radar bright bands (> 30 dBZ) form as the aggregates begin their initial stages of melting.

While the ultimate goal of this work was to link different global DSD modes to the physical processes shaping them, we recognize this is a complex problem with many different aspects including thermodynamic, microphysical and dynamical feedbacks. To truly understand the variability in each region and the multitude of factors shaping the six DSD groups described herein requires synergistic modeling and observational studies. We have outlined a framework that allows for the statistical, objective separation of raining points by their DSD characteristics in six-dimensional space, and have demonstrated that six distinct populations with similar characteristics can be grouped around the globe. While these populations vary in frequency and breadth across latitudes, they have important implications for radar-based rainfall retrieval and model assumptions surrounding rain and DSDs. An important continuation of the work will be to include environmental factors to understand their role in DSD variability. We also acknowledge the dataset may not have captured all rain regimes, and therefore datasets from underrepresented locations (e.g. high latitude oceans) should be considered in the future in order to more comprehensively address regional DSD variability. Seasonal and diurnal cycles of DSDs could also be examined from the multi-year datasets used in this study. Higher order EOFs could be examined to identify more modes of variability.

Acknowledgements

We would like to thank Dr. Merhala Thurai (CSU ECE) for her assistance in processing the disdrometer datasets and discussions about the analysis, and Dr. Ali Tokay (NASA GSFC) for providing the NASA disdrometer datasets. We also thank Drs. Walt Petersen (NASA MSFC),
Matt Schwaller and Ramesh Kakar for their leadership of the NASA GV program. Joe Zagrodnik and Drs. Robert Houze and Lynn McMurdie (U. Washington) provided valuable insights for this work. Datasets were provided by NOAA Earth Science Research Laboratory, U.S. Department of Energy Atmospheric Radiation Measurement (ARM) Climate Research Facility, and Australia’s Bureau of Meteorology (BOM). This work was supported by the DOE Atmospheric System Research program (DE-SC0014371), the NASA Precipitation Measurement Mission (PMM) (NNX16AD85G), and the NASA PMM Ground Validation program (NNX16AI11G). The Python ARM Radar Toolkit (Py-ART) was used in the analysis performed herein (JJ Helmus and SM Collis, JORS 2016, doi: 10.5334/jors.119). We also thank two anonymous reviewers and Dr. Joe Munchak for constructive comments that improved the analysis.


BRINGI, V. N., V. CHANDRASEKAR, J. HUBBERT, E. GORGUCCI, W. L. RANDEU, and M. SCHÖNHUBER, 2003: Raindrop size distribution in different climatic regimes from disdrometer and dual-


doi:http://dx.doi.org/10.5067/GPMGV/IFLOODS/2DVD/DATA301

Petersen, W. and P. Gatlin, 2014: GPM Ground Validation Two-Dimensional Video Disdrometer (2DVD) IPHEX. Data set available online [http://ghrc.nsstc.nasa.gov/] from the NASA EOSDIS Global Hydrology Resource Center Distributed Active Archive Center Huntsville, Alabama, U.S.A. doi:
http://dx.doi.org/10.5067/GPMGV/IPHEX/2DVD/DATA301


Table 1: Description of disdrometer datasets used in this study. The *, †, #, and † denote datasets that were grouped due to co-spatial / co-temporal or similar variability, and will be referred to by the identifier name. The raining minutes presented are after the QC described in the text.

<table>
<thead>
<tr>
<th>Name</th>
<th>Long Name</th>
<th>Location</th>
<th>Identifier</th>
<th>Time Frame</th>
<th># raining points</th>
<th># disrometers / type</th>
<th>Notes</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>iFloods</td>
<td>Iowa Flood Studies</td>
<td>Eastern Iowa</td>
<td>IFloodS</td>
<td>6 Apr – 16 Jun 2013</td>
<td>14608</td>
<td>6 2DVDs</td>
<td>Deployed along NPOL radial</td>
<td>Petersen and Gatlin 2013, Seo et al. 2015</td>
</tr>
<tr>
<td>MC3E*</td>
<td>Midlatitude Continental Clouds and Convection Experiment</td>
<td>Central Oklahoma</td>
<td>SGP</td>
<td>23 Apr – 1 Jun 2011</td>
<td>6043</td>
<td>5 2DVDs</td>
<td>Deployed in 25 km² area</td>
<td>Petersen et al. 2011, Jensen et. al. 2015</td>
</tr>
<tr>
<td>SGP*</td>
<td>Southern Great Plains</td>
<td>Central Oklahoma</td>
<td>SGP</td>
<td>28 Feb 2011 – 5 May 2016</td>
<td>39592</td>
<td>1 2DVD</td>
<td>Data from snow cases removed</td>
<td>Petersen and Gatlin 2014, Barros et al. 2014</td>
</tr>
<tr>
<td>IPHEx</td>
<td>Integrated Precipitation and Hydrology Experiment</td>
<td>Western North Carolina</td>
<td>IPHEx</td>
<td>25 Apr – 16 Jun 2014</td>
<td>10718</td>
<td>5 2DVD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LPVEx†</td>
<td>Light Precipitation Validation Experiment</td>
<td>Helsinki, Finland</td>
<td>Finland</td>
<td>9 Sept – 20 Oct 2010</td>
<td>2080</td>
<td>3 2DVD</td>
<td>Some days removed due to frozen precip reaching ground</td>
<td>L’Ecuyer et al. 2010, Petersen et al. 2010</td>
</tr>
<tr>
<td>BAECC†</td>
<td>Biogenic Aerosols – Effects on Clouds and Climate</td>
<td>Hyytiala, Finland</td>
<td>Finland</td>
<td>15 Feb – 11 Sept 2014</td>
<td>13502</td>
<td>1 2DVD</td>
<td>Data from snow cases removed</td>
<td>Petäjä et al. 2016</td>
</tr>
<tr>
<td>Manus*</td>
<td>Manus Island</td>
<td>Tropical Ocean</td>
<td>Manus</td>
<td>01 Dec 2011 – 04 Jul 2014</td>
<td>45506</td>
<td>1 2DVD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiment</td>
<td>Location</td>
<td>Start Date – End Date</td>
<td>Duration</td>
<td>Number of Days</td>
<td>Notes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------------</td>
<td>----------------------------</td>
<td>----------</td>
<td>----------------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMIE, Gan*</td>
<td>Gan Island, Maldives</td>
<td>05 Oct 2011 – 06 Feb 2012</td>
<td>4348</td>
<td>1 2DVD</td>
<td>Long et al. 2011</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TWPICE†</td>
<td>Darwin, Australia</td>
<td>3 Nov 2005 – 10 Feb 2006</td>
<td>3669</td>
<td>1 JWD</td>
<td>May et al. 2008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Darwin†</td>
<td>Darwin, Australia</td>
<td>1 Mar 2011 – 04 Jan 2015</td>
<td>42248</td>
<td>1 2DVD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLYMPEX</td>
<td>Olympic Peninsula OLYMP Ex</td>
<td>31 Oct 2015 – 16 Jan 2016</td>
<td>78124</td>
<td>3 APUs*</td>
<td>Houze et al. 2017</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global</td>
<td></td>
<td></td>
<td>334958</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2: Percent variance explained by the first 2 EOFs.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>EOF1</th>
<th>EOF2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>58</td>
<td>30</td>
<td>88</td>
</tr>
<tr>
<td>High Latitude</td>
<td>57</td>
<td>31</td>
<td>88</td>
</tr>
<tr>
<td>Mid Latitude</td>
<td>59</td>
<td>20</td>
<td>89</td>
</tr>
<tr>
<td>Low Latitude</td>
<td>62</td>
<td>26</td>
<td>88</td>
</tr>
</tbody>
</table>
Table 3: Mean $D_0$ values for the six Groups using the PC thresholds in Fig. 4.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>All</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
<th>Group 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>1.13</td>
<td>1.79</td>
<td>0.67</td>
<td>0.81</td>
<td>2.16</td>
<td>1.35</td>
<td>2.62</td>
</tr>
<tr>
<td>High</td>
<td>1.0</td>
<td>1.71</td>
<td>0.59</td>
<td>0.68</td>
<td>1.93</td>
<td>0.98</td>
<td>2.25</td>
</tr>
<tr>
<td>Mid</td>
<td>1.18</td>
<td>2.05</td>
<td>0.68</td>
<td>0.70</td>
<td>2.21</td>
<td>1.23</td>
<td>2.71</td>
</tr>
<tr>
<td>Low</td>
<td>1.18</td>
<td>1.62</td>
<td>0.70</td>
<td>1.07</td>
<td>2.24</td>
<td>1.39</td>
<td>2.74</td>
</tr>
</tbody>
</table>

Table 4: Mean $\log(N_w)$ values for the six Groups using the PC thresholds in Fig. 4.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>All</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
<th>Group 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>3.95</td>
<td>4.19</td>
<td>3.64</td>
<td>4.73</td>
<td>2.87</td>
<td>4.69</td>
<td>3.04</td>
</tr>
<tr>
<td>High</td>
<td>4.12</td>
<td>4.14</td>
<td>3.93</td>
<td>4.81</td>
<td>3.08</td>
<td>4.94</td>
<td>3.21</td>
</tr>
<tr>
<td>Mid</td>
<td>3.79</td>
<td>3.85</td>
<td>3.53</td>
<td>4.71</td>
<td>2.78</td>
<td>4.80</td>
<td>2.96</td>
</tr>
<tr>
<td>Low</td>
<td>3.94</td>
<td>4.34</td>
<td>3.46</td>
<td>4.63</td>
<td>2.82</td>
<td>4.66</td>
<td>3.05</td>
</tr>
</tbody>
</table>

Table 5: Mean $\mu$ values for the six Groups using the PC thresholds in Fig. 4.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>All</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
<th>Group 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>4.20</td>
<td>0.64</td>
<td>6.91</td>
<td>6.23</td>
<td>-0.02</td>
<td>2.14</td>
<td>-0.79</td>
</tr>
<tr>
<td>High</td>
<td>4.44</td>
<td>0.23</td>
<td>6.82</td>
<td>6.75</td>
<td>-0.51</td>
<td>2.36</td>
<td>-1.07</td>
</tr>
<tr>
<td>Mid</td>
<td>4.35</td>
<td>0.30</td>
<td>7.26</td>
<td>7.18</td>
<td>0.20</td>
<td>2.70</td>
<td>-0.61</td>
</tr>
<tr>
<td>Low</td>
<td>3.70</td>
<td>0.99</td>
<td>6.48</td>
<td>4.61</td>
<td>-0.36</td>
<td>2.08</td>
<td>-1.20</td>
</tr>
</tbody>
</table>
Table 6: Summary of the radar characteristics and disdrometer matches.

<table>
<thead>
<tr>
<th>Project</th>
<th>Radar</th>
<th>RHI elevations</th>
<th>PPI elevations (#, highest)</th>
<th>Resample time</th>
<th>Disdrom.</th>
<th>Distance to Disdrom.</th>
<th># Raining matches</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLYMPEx</td>
<td>NPOL</td>
<td>0-45º</td>
<td>--</td>
<td>20 mins</td>
<td>APU03, APU05, APU08</td>
<td>19, 32, 40 km</td>
<td>3320</td>
</tr>
<tr>
<td>IFloodS</td>
<td>NPOL</td>
<td>0-20º</td>
<td>12, 5-8º</td>
<td>2-5 mins</td>
<td>SN36, SN37</td>
<td>25, 47 km</td>
<td>1112</td>
</tr>
<tr>
<td>IPHEx</td>
<td>NPOL</td>
<td>0-60º</td>
<td></td>
<td>2-5 mins</td>
<td>SN35, SN36</td>
<td>22, 42 km</td>
<td>354</td>
</tr>
<tr>
<td>MC3E</td>
<td>NPOL</td>
<td>0-40º</td>
<td>28, 35º</td>
<td>1-5 mins</td>
<td>SN25, SN70</td>
<td>33, 28 km</td>
<td>1919</td>
</tr>
<tr>
<td>Gan</td>
<td>SPOL</td>
<td>0-60º</td>
<td>--</td>
<td>15 mins</td>
<td>2DVD</td>
<td>8 km</td>
<td>519</td>
</tr>
<tr>
<td>Darwin</td>
<td>CPOL</td>
<td>0-45º</td>
<td>17, 42º</td>
<td>10 mins</td>
<td>JWD</td>
<td>23 km</td>
<td>817</td>
</tr>
</tbody>
</table>
Figure Captions

Fig. 1: a) Locations of disdrometer observations used herein. See Table 1 for individual datasets that fit into the larger eight locations. b) Distribution of raining minutes as a function of latitude band.

Fig. 2: Two-dimensional normalized frequency of occurrence as a function of latitude band for logNw-D0 (a-c) and LWC-D0 (d-f). The Bringi et al. (2009) and Thompson et al. (2015) convective – stratiform lines are presented with gray dashed and solid lines, respectively, in panels a-c.

Fig. 3: The first two EOFs for each latitude band and the global dataset (black). Solid lines indicate the positive modes and dashed lines indicate the negative modes (where the sign is arbitrary).

Fig. 4: Frequency density of joint distributions of PC1 and PC2. Dashed red lines illustrate the σ1.5 thresholds, and the dashed black lines indicate the 0 value for PC1 and PC2. Shaded boxes represent the six groups of points with similar characteristics based on PCA.

Fig. 5: a) Global characteristic size distributions using the mean values for each identified Group. Global means of all data points are indicated. b) Normalized mean values of parameters for the six Groups.

Fig. 6: Two-dimensional distributions of each Group in the high, middle and low latitude bands for a-c) logNw-D0, d-f) µ-D0 and g-i) logNw-µ space. The gray dots encompass the entire space for each dataset and represent the points that did not meet PC thresholds (“ambiguous”); red is Group 1, green is Group 2, yellow is Group 3, blue is Group 4, orange is Group 5 and purple is Group 6. Contours highlight smoothed Gaussian 1-σ points for each group. In panel a-c, the solid black line is the Thompson et al. (2015) convective – stratiform separation line for the tropical oceans, the dashed black line is the Bringi et al. (2009) convective – stratiform separation.

Fig. 7: Two-dimensional distributions of each Group in the high, middle and low latitude bands for a-c) LWC-D0, d-f) Z-R and g-i) R-µ space. The gray dots encompass the entire space for each dataset and represent the points that did not meet PC thresholds.

Fig. 8: Vertical structure of a-c) reflectivity and d-f) differential reflectivity (Zdr) for each of the six groups from coincident radar RHIs for the high, middle and low latitude bands. Panels g-i display the mean reflectivity at 2.5 km as a function of the mean radar 0 dBZ echo top height as a function of group determined by the surface disdrometer. Here the low latitudes are comprised from Darwin and Gan data, the mid latitudes from IFloodS, MC3E, and IPHEX, and the high latitudes from OLYMPEX radar data. Note that the numbers comprising each group indicated in the legends of panel a-c do not match those given in Table 6 due to the ambiguous points, which are not shown. PC thresholds of σ1.0 was used for this analysis.

Fig. 9: Time-height cross-sections constructed from CPOL PPI data from Darwin (a) and SPOL RHI data from Gan (e, i). Radar reflectivity is color-contoured in panels a), d), e), h), i), and l) where d), h) and l) represent a PPI at the time of the vertical line and the disdrometer location denoted by a black dot. Reflectivity scale is the same as the time-height, and the approximate scale of the PPI is given at the bottom. The start times of the PPIs/RHIs used to construct the time-height cross-sections are illustrated by hatch marks along x=0 in...
The disdrometer Group classifications are shown as colored dots along the time series. For reference the disdrometer points are shown in logN_w-D_0 space in panels b), f), j). The panels below the reflectivity time-height panel (c, g, and k) illustrate the PC values as a timeseries, with the PC thresholds indicated by dashed horizontal lines.

Note: For illustration purposes, the PC σ-threshold has been lowered to σ_1.0.

Fig. 10: Same as Fig. 9, but time-height cross-sections constructed from NPOL PPI data from IFloodS.

Fig. 11: Same as Fig. 9, but time-height cross-sections constructed from NPOL RHI data from OLYMPEEx.

Fig. 12: Conceptual model illustrating the dominant mechanisms for the six groups objectively determined from the surface disdrometers using PCA in logN_w-D_0 space. The grey dots represent every point in the global dataset. Solid contours represent smoothed Gaussian 1-σ regions encompassing each of the six groups (determined using thresholds on PC1 and PC2 values), while inferred processes are in black. Dashed dark red and cyan contours represent groups related to convective and stratiform processes, respectively.
Fig. 1: a) Locations of disdrometer observations used herein. See Table 1 for individual datasets that fit into the larger eight locations. b) Distribution of raining minutes as a function of latitude band.
Fig. 2: Two-dimensional normalized frequency of occurrence as a function of latitude band for log\(N_w-D_0\) (a-c) and LWC-\(D_0\) (d-f). The Bringi et al. (2009) and Thompson et al. (2015) convective – stratiform lines are presented with gray dashed and solid lines, respectively, in panels a-c.
Fig. 3: The first two EOFs for each latitude band and the global dataset (black). Solid lines indicate the positive modes and dashed lines indicate the negative modes (where the sign is arbitrary).
Fig. 4: Frequency density of joint distributions of PC1 and PC2. Dashed red lines illustrate the $\sigma_{1.5}$ thresholds, and the dashed black lines indicate the 0 value for PC1 and PC2. Shaded boxes represent the six groups of points with similar characteristics based on PCA.
Fig. 5: a) Global characteristic size distributions using the mean values for each identified Group. Global means of all data points are indicated. b) Normalized mean values of parameters for the six Groups.
Fig. 6: Two-dimensional distributions of each Group in the high, middle and low latitude bands for a-c) logN_w-D_0, d-f) µ-D_0 and g-i) logN_w-µ space. The gray dots encompass the entire space for each dataset and represent the points that did not meet PC thresholds (“ambiguous”); red is Group 1, green is Group 2, yellow is Group 3, blue is Group 4, orange is Group 5 and purple is Group 6. Contours highlight smoothed Gaussian 1-σ points for each group. In panel a-c, the solid
black line is the Thompson et al. (2015) convective – stratiform separation line for the tropical oceans, the dashed black line is the Bringi et al. (2009) convective – stratiform separation.
Fig. 7: Two-dimensional distributions of each Group in the high, middle and low latitude bands for a-c) LWC-D₀, d-f) Z-R and g-i) R-μ space. The gray dots encompass the entire space for each dataset and represent the points that did not meet PC thresholds.
Fig. 8: Vertical structure of a-c) reflectivity and d-f) differential reflectivity ($Z_{dr}$) for each of the six groups from coincident radar RHIs for the high, middle and low latitude bands. Panels g-i display the mean reflectivity at 2.5 km as a function of the mean radar 0 dBZ echo top height as a function of group determined by the surface disdrometer. Here the low latitudes are comprised from Darwin and Gan data, the mid latitudes from IFloodS, MC3E, and IPHEX, and the high latitudes from OLYMPEX radar data. Note that the numbers comprising each group indicated in the legends of panel a-c do not match those given in Table 6 due to the ambiguous points, which are not shown. PC thresholds of $\sigma_{1.0}$ was used for this analysis.
Fig. 9: Time-height cross-sections constructed from CPOL PPI data from Darwin (a) and SPOL RHI data from Gan (e, i). Radar reflectivity is color-contoured in panels a), d), e), h), i), and l) where d), h) and l) represent a PPI at the time of the vertical line and the disdrometer location.
denoted by a black dot. Reflectivity scale is the same as the time-height, and the approximate scale of the PPI is given at the bottom. The start times of the PPIs/RHIs used to construct the time-height cross-sections are illustrated by hatch marks along x=0 in panels a), e), and i). The disdrometer Group classifications are shown as colored dots along the time series. For reference the disdrometer points are shown in logNw-D0 space in panels b), f), j). The panels below the reflectivity time-height panel (c, g, and k) illustrate the PC values as a timeseries, with the PC thresholds indicated by dashed horizontal lines. Note: For illustration purposes, the PC σ-threshold has been lowered to σ1,0.
Fig. 10: Same as Fig. 9, but time-height cross-sections constructed from NPOL PPI data from IFloodS.
Fig. 11: Same as Fig. 9, but time-height cross-sections constructed from NPOL RHI data from OLYMPEEx.
Fig. 12: Conceptual model illustrating the dominant mechanisms for the six groups objectively determined from the surface disdrometers using PCA in log\(N_w\)-\(D_0\) space. The grey dots represent every point in the global dataset. Solid contours represent smoothed Gaussian 1-\(\sigma\) regions encompassing each of the six groups (determined using thresholds on PC1 and PC2 values), while inferred processes are in black. Dashed dark red and cyan contours represent groups related to convective and stratiform processes, respectively.