Evaluation of the Microphysical Assumptions within GPM-DPR Using Ground-Based Observations of Rain and Snow

Randy J. Chase 1,*, Stephen W. Nesbitt 1 and Greg M. McFarquhar 2

1 Department of Atmospheric Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA; snesbitt@illinois.edu
2 Cooperative institute of Mesoscale Meteorological Studies and School of Meteorology, University of Oklahoma, Norman, OK 73072, USA; mcfarq@ou.edu

* Correspondence: randyjc2@illinois.edu

Received: 30 April 2020; Accepted: 9 June 2020; Published: 11 June 2020

Abstract: The Global Precipitation Measurement Dual-Frequency Precipitation Radar (GPM-DPR) provides an opportunity to investigate hydrometeor properties. Here, an evaluation of the microphysical framework used within the GPM-DPR retrieval was undertaken using ground-based disdrometer measurements in both rain and snow with an emphasis on the evaluation of snowfall retrieval. Disdrometer measurements of rain show support for the two separate prescribed relations within the GPM-DPR algorithm between the precipitation rate ($R$) and the mass weighted mean diameter ($D_m$) with a mean absolute percent error (MAPE) on $R$ of 29% and 47% and a mean bias percentage (MBP) of −6% and −20% for the stratiform and convective relation, respectively. Ground-based disdrometer measurements of snow show higher MAPE and MBP values in the retrieval of $R$, at 77% and −52%, respectively, compared to the stratiform rain relation. An investigation using the disdrometer-measured fall velocity and mass in the calculation of $R$ and $D_m$ illustrates that the variability found in hydrometeor mass causes a poor correlation between $R$ and $D_m$ in snowfall. The results presented here suggest that $R - D_m$ retrieval is likely not optimal in snowfall, and other retrieval techniques for $R$ should be explored.

Keywords: snowfall rate; rainfall rate; radar retrievals

1. Introduction

Global retrievals of precipitation that falls as rain and snow at the surface and aloft are important for quantifying several key components of the Earth’s water cycle. At the surface, the amount of snow that falls within a watershed located in mountainous terrain is vital for water supply and water resource management [1]. Furthermore, even though most of the Earth’s precipitation reaches the surface as rain, more than 60% of precipitation on average can be connected to ice phase processes aloft [2]. Thus, in order to obtain a complete quantitative understanding of the hydrologic cycle, an accurate retrieval of snowfall is required. Moreover, quantifying the total amount of ice mass in the atmospheric column (i.e., the ice water path (IWP)) has direct implications on the Earth’s radiative balance and is important for constraining general circulation models [3,4].

One method of retrieving snowfall properties is the use of radar. Since the first use of radar in atmospheric sciences, studies have shown that snowfall mass—and thus the precipitation rate—is related to the power scattered back to the radar [5]. In this work, the abbreviation $R$ is used for the precipitation rate and is used to refer both to the rainfall and snowfall rate. Current spaceborne radars, such as CloudSat [6] or the Global Precipitation Measurement mission’s Dual-Frequency Precipitation Radar (GPM-DPR [7]), attempt to retrieve snowfall properties such as $R$, ice water content (IWC), and in
the case of GPM enabled by its dual frequency measurements, the mass-weighted mean diameter ($D_m$) and characterize the spatial distribution of precipitation rate and microphysical properties.

As the first spaceborne radar designed to observe clouds and precipitation at high latitudes, CloudSat has been used to retrieve global distributions of near-surface retrieved snowfall. In [8,9], the authors were the first to investigate global retrievals of snowfall which contained large uncertainty (e.g., a factor of 10; see figure 1 in [10]) since the basis of the retrieval required the use of a power-law relation between the measured equivalent radar reflectivity factor ($Z_e$) and the corresponding snowfall rate (also known as a $Z_e - R$ relation), which is sensitive to the assumed particle type used in its formulation. In order to reduce uncertainty, renewed investigations of the global distribution of snowfall [11–14] use the CloudSat 2C-SNOWPROFILE product, which uses an optimal estimation technique [15] to retrieve snowfall rate and produce an uncertainty estimate [16,17]. Since the release of the 2C-SNOWPROFILE, studies have compared the CloudSat retrieval of $R$ to ground-based techniques and show encouraging results [18–21], enhancing confidence in the use of the CloudSat 2C-SNOWPROFILE to retrieve global snowfall properties.

In an effort to expand NASA’s earth observing capabilities with an emphasis on global precipitation, GPM-DPR was launched into orbit in early February 2014. With the new spaceborne radar, new retrievals of global rainfall and snowfall properties could be obtained (e.g., [22,23]). Since GPM-DPR is a scanning Ku and Ka-band radar, while CloudSat is a non-scanning W-band radar, it requires its own suite of retrievals that make a range of different assumptions. Once a sufficient sample size of near-coincidences between satellites was obtained, the authors in [24] investigated GPM-DPR’s ability to detect snowfall events using CloudSat as a reference. They found that GPM-DPR does not detect more than 90% of the detected snowfall events by CloudSat, attributing this to GPM-DPR’s lack of sensitivity. Furthermore, in [24], the authors showed that GPM-DPR’s retrievals of $R$ in falling snow have a large low bias. The bulk statistical evaluation of the GPM-DPR and CloudSat retrieved products in falling snow by the authors in [14] showed that the global average retrieved GPM-DPR near-surface snowfall rate is about 43% lower compared to CloudSat even after carefully considering the differences in sampling (e.g., footprint size), sensitivity and operating frequency. Regional differences show much larger disparity (see Figure 7 in [14]). The low bias in the GPM-DPR snowfall retrieval is also shown by [25], where the GPM-DPR’s snowfall retrieval was considerably lower than relationships derived from measurements collected during the Olympic Mountains Experiment (OLYMPEX [26]), the Global Precipitation Measurement Cold Season Experiment (GCPEX [27]) and a mass flux technique that uses mass continuity through the melting layer and more constrained retrievals in rain. Since various investigators and methods have shown consistent results, it is evident that the current GPM-DPR retrieval in snowfall is consistently biased towards low values of $R$. Furthermore, the results in [14] suggest that the likely reason for this low bias is the GPM-DPR retrieval algorithm itself and not the satellite hardware (e.g., calibration, sensitivity, footprint size, scanning/non-scanning) or orbit (e.g., inclination, sun synchronous/non-sun synchronous).

A primary microphysical assumption made within the GPM-DPR retrieval algorithm is that there is a prescribed empirical relationship between $R$ and $D_m$. The $R - D_m$ relation used is derived ultimately from ground-based disdrometer measurements in the Tropics [28,29] and allows the simultaneous retrieval of $R$ and $D_m$ for a measured $Z_e$. This framework was first adopted in version 4 of the GPM-DPR algorithm (released in 2016) and is used currently (version 6) for all precipitating echoes regardless of hydrometeor phase.

Evaluations of GPM-DPR retrieval products within rain have been ongoing. Studies have considered the direct evaluation of GPM-DPR retrievals with surface-based rain gauges [30], radars [31] and disdrometers [32]. Results generally show a good agreement and meet the Level 1 Scientific Requirements of the GPM mission (see Section 1.1.3 of [33]). However, there have been no investigations of the empirical relation itself and whether it can encapsulate observations of rainfall
from various meteorological regimes. Furthermore, no studies have verified that the same $R - D_m$ empirical relation applies to snowfall.

The analysis provided here will assess the empirical relation prescribed between $R$ and $D_m$ using surface observations of rainfall from various geographic locations and will quantify its error. Then, observations of surface snowfall will be contextualized in the liquid equivalent $R - D_m$ space for a direct evaluation of the $R - D_m$ relations used in the GPM-DPR retrieval algorithm. The structure of the rest of the manuscript is as follows: Section 2 describes the datasets used in this study and how the data were collected and discusses how PSD parameters are derived from both liquid and solid-phase PSDs. Section 3 presents the results of how well $R - D_m$ relations generated using PSDs measured in both rain and snow compare to the prescribed $R - D_m$ relation in the GPM-DPR algorithm, quantifies error in retrievals emulating the GPM-DPR algorithm and investigates the correlation between $R$ and $D_m$. Section 4 presents a brief summary and states the overall conclusions and some future research ideas.

2. Data and Methodology

2.1. Measurements

2.1.1. Rainfall Data

In order to evaluate the applicability of the prescribed $R - D_m$ relationships within the GPM-DPR algorithm in rain, data collected as part of NASA Ground Validation (GV) field campaigns and Department of Energy-Atmospheric Radiation Measurement (DOE-ARM) mobile facility deployments are used. The NASA GV field campaigns have accumulated numerous surface-based rainfall PSDs over numerous geographic locations around the United States of America. Specifically, the campaign data used here are from the following projects: Mid-Latitude Continental Convective Clouds Experiment (MC3E, Oklahoma [34]), the Iowa Flood Studies (IFloodS, Iowa), Integrated Precipitation and Hydrology Experiment (IPHEx, North Carolina [35]), and the Olympic Mountain Experiment (OLYMPEX, Washington State [26]). Data collected at two sites that were not a part of official campaigns but nevertheless are included in the NASA GV database are also used here, namely data collected in Huntsville, Alabama and at Wallops Air Force Base, Virginia. In order to add samples from outside the United States, data are included form several DOE-ARM campaigns and fixed field sites of DOE-ARM. Specifically, data from the following international sites are used: Cloud, Aerosol and Complex Terrain Interactions (CACTI, Cordoba, Argentina [36]), Tropical West Pacific (TWP, Darwin, Australia and Manus Island, Papua New Guinea [37]), Dynamics of the Madden-Julian Oscillation / Cooperative Indian Ocean experiment (DYNAMO/CINDY2011, Gan Island, Maldives [38]) and Eastern North Atlantic (ENA, Graciosa Island in the Azores, Portugal [39]). Finally, a long-term record from the Southern Great Plains (SGP) site at Lamont, Oklahoma is used [40].

For all the aforementioned campaigns and field sites, data from a two-dimensional video disdrometer (2DVD [41]) are used. The 2DVD is commonly used as a reference disdrometer (e.g., [42–44]) that measures particle maximum dimension, shape and fall velocity. The minimum particle size that can be sampled reliably by the 2DVD is 0.2 mm [41]. The specific data used here are the processed datafiles that remove particles that have a terminal fall velocity less than 50% of the predicted fall velocity based on size following the relation in [45]. This prevents data from being contaminated by secondary drops (i.e., two drops in the sample volume at the same time). Furthermore, only time periods where the rain rate is greater than 0.01 mm hr$^{-1}$ and had more than 10 drops within 1 min were considered in the analysis. Then, measured particle size distributions (PSDs) are scaled from the 1 min averages to 5 min averages. The result of the amalgamation of all the 2DVD measurements is approximately 1.67 years of raining PSDs. Overview statistics from all 2DVD data and their respective campaigns are found in Table 1.
Table 1. Statistics of disdrometer measurements. The equations to calculate the parameters below are found in the following section, and their means (when $R \geq 0.01 \text{ mm hr}^{-1}$) are reported in the table. The number of samples refers to the number of 5 min particle size distributions (PSDs). The means of $N_w$ and $Z_{ku}$ were calculated in linear units and then converted to the logarithmic units shown. The underlined text in the first column denotes the data source and probe.

<table>
<thead>
<tr>
<th>Campaign or Field Site</th>
<th>Number of Samples</th>
<th>$R$ $\text{mm hr}^{-1}$</th>
<th>$D_m$ $\text{mm}$</th>
<th>$N_w$ (\log\text{(m}^{-3} \text{ mm}^{-1}))</th>
<th>$Z_{ku}$ $\text{dBZ}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NASA: 2DVD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC3E</td>
<td>2357</td>
<td>2.10</td>
<td>1.36</td>
<td>3.65</td>
<td>35.74</td>
</tr>
<tr>
<td>IFloodS</td>
<td>8565</td>
<td>2.41</td>
<td>1.27</td>
<td>3.74</td>
<td>35.64</td>
</tr>
<tr>
<td>IPHEx</td>
<td>3926</td>
<td>3.54</td>
<td>1.25</td>
<td>3.63</td>
<td>37.88</td>
</tr>
<tr>
<td>OLYMPEX</td>
<td>22,401</td>
<td>2.58</td>
<td>1.15</td>
<td>3.97</td>
<td>32.12</td>
</tr>
<tr>
<td>Huntsville, AL</td>
<td>9810</td>
<td>2.78</td>
<td>1.26</td>
<td>3.77</td>
<td>35.49</td>
</tr>
<tr>
<td>Wallops, VA</td>
<td>32,447</td>
<td>1.93</td>
<td>1.12</td>
<td>3.81</td>
<td>32.92</td>
</tr>
<tr>
<td>DOE: 2DVD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CACTI</td>
<td>6151</td>
<td>1.68</td>
<td>0.86</td>
<td>4.72</td>
<td>32.94</td>
</tr>
<tr>
<td>Darwin, AU</td>
<td>8868</td>
<td>6.10</td>
<td>1.57</td>
<td>3.54</td>
<td>42.37</td>
</tr>
<tr>
<td>Manus isl, PNG</td>
<td>13,196</td>
<td>4.44</td>
<td>1.38</td>
<td>3.61</td>
<td>38.95</td>
</tr>
<tr>
<td>Gan isl, MV</td>
<td>1583</td>
<td>3.77</td>
<td>1.29</td>
<td>3.72</td>
<td>37.48</td>
</tr>
<tr>
<td>ENA</td>
<td>34893</td>
<td>1.45</td>
<td>0.90</td>
<td>4.28</td>
<td>30.64</td>
</tr>
<tr>
<td>SGP</td>
<td>31871</td>
<td>2.11</td>
<td>1.21</td>
<td>3.88</td>
<td>36.22</td>
</tr>
<tr>
<td>NASA: PIP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAECC</td>
<td>2573</td>
<td>0.53</td>
<td>0.72</td>
<td>4.83</td>
<td>18.44</td>
</tr>
</tbody>
</table>

2.1.2. Snowfall Data

From January until April 2014, a joint field campaign between NASA, the University of Helsinki, DOE-ARM and the Finnish Meteorological Institute was conducted with an intensive observation period designed explicitly to study snowfall [46]. The campaign, named the Biogenic Aerosols—Effect on Clouds and Climate experiment (BAECC), collected data with several radars (C-band, X-band, Ka-band and W-band; dual-polarized and Doppler; scanning and non-scanning), surface-based meteorological instruments, vertically pointing microwave radiometers and surfaced-based automated measurements of snowfall and its PSD. Snowfall rate was measured directly through two OTT Hydromet Pluvio2 weighing bucket gauges enclosed in a Double Fence Intercomparison Reference (DFIR) and a Precipitation Imaging Package (PIP)—the successor to the Snowflake Video Imager [47] and not to be confused with the commonly used aircraft probe, the Precipitation Imaging Probe.

The PIP is a high-speed camera pointed at a light source that is 2 m away. This allows for any particle falling within its field of view (48 mm × 64 mm) and between the light source and the camera to cast a shadow. From these videos of shadowed particles, the particle size and fall velocity can be diagnosed. Since the terminal fall velocity is explicitly measured for each particle, the derivation of the mass of each particle can be deduced if atmospheric base state information (i.e., pressure, temperature) is known. Effectively, the mass can be estimated through hydrodynamic theory by considering the drag on the particle, the buoyancy of the particle in the fluid (i.e., air) and gravity [48,49]. This method has been used on other ground-based disdrometers [50–53] as well as the dataset used herein [54–56]. For a more complete discussion of the retrieval of the mass and its intricacies, consult section 3a from [54]. After the BAECC campaign concluded its snowfall IOP, the PIP and weighting bucket gauges remained at the measurement site in southern Finland and are currently still collecting observations. Thus, the ongoing record is now five complete winter seasons. Since the processing of the PIP data and retrieval of mass is not trivial, at the time of writing this manuscript, only the data from February 2014–March 2015 were available. Summary statistics of the PIP data are shown in Table 1 for comparison against the 2DVD datasets.
2.1.3. Data Availability

All rainfall data are free for use can be found in two main locations, the NASA GHRC [57] and ARM DOE data discovery [58]. The processed snowfall observations, including the retrieved mass, can be found on github [59]. For convenience, some data and the current version of the Algorithm Theoretical Basis Document (at the time of submission) is found on the github site [60].

2.2. Particle Size Distribution Parameters

2.2.1. Rainfall PSD Parameters

The precipitation rate for rain can be determined from a measured PSD by

\[ R = \sum_{i=0}^{n_{\text{bins}}} v(D_i) \ vol(D_i) \ N(D_i) \ \Delta D_i \]

where \( \text{vol}(D_i) \) is the volume of liquid water, \( v(D_i) \) is the terminal fall velocity and \( N(D_i) \) is the number distribution function for particles with a maximum dimension of \( D_i - (\frac{\Delta D_i}{2}) < D_i < D_i + (\frac{\Delta D_i}{2}) \), where \( \Delta D_i \) is the bin width and the subscript \( i \) indicates the \( i^{th} \) bin of the measured PSD. The characteristic size of mass weighted mean diameter \( (D_m) \) can also be determined from a measured PSD as

\[ D_m = \frac{\sum_{i=0}^{n_{\text{bins}}} m(D_i) \ D_i \ N(D_i) \ \Delta D_i}{\sum_{i=0}^{n_{\text{bins}}} m(D_i) \ N(D_i) \ \Delta D_i} \]

where \( m(D_i) \) is the mass of the particle with a maximum dimension of \( D_i \). The bounds of the summation for Equations (1) and (2) are the minimum \( (D_{\text{min}}) \) and maximum \( (D_{\text{max}}) \) trusted size bin of the 2DVD, which are 0.2 mm and 10 mm, respectively. If the raindrops are assumed to be spherical, then the volume and mass of a raindrop with size \( D_i \) can be substituted into Equations (1) and (2) as follows:

\[ R = \sum_{i=0}^{n_{\text{bins}}} v(D_i) \ \frac{\pi}{6} D_i^3 \ N(D_i) \ \Delta D_i \]

\[ D_m = \frac{\sum_{i=0}^{n_{\text{bins}}} D_i^4 \ N(D_i) \ \Delta D_i}{\sum_{i=0}^{n_{\text{bins}}} D_i^3 \ N(D_i) \ \Delta D_i} \]

In reality, raindrops larger than 1 mm are not observed to be spherical [61]. On average, if \( R \) and \( D_m \) are calculated from the PSDs using a sphere versus a spheroid with the axis ratio predicted in [62], there is approximately a 10\% and 3\% overestimation of \( R \) and \( D_m \), respectively (not shown). While this is not a trivial amount of error, it is a noted limitation of the work, and assessing the raindrop shape assumption within GPM-DPR is a topic of future research.

Under the Rayleigh assumption (generally applicable in S and C-band weather radar retrievals) where the wavelength of the radar is much greater than the maximum dimension of the particle, the equivalent radar reflectivity factor \( (Z_e) \) is given by

\[ Z_e = \sum_{i=0}^{n_{\text{bins}}} D_i^5 \ N(D_i) \ \Delta D_i \]

Since the GPM-DPR consists of Ku and Ka-band radars, there could be instances in which Rayleigh scattering is not a valid assumption. To avoid additional inaccuracies on the wavelength dependence, the Ku and Ka-band \( Z_e \) are calculated as

\[ Z_e = \frac{\lambda^4}{\pi^5 |K|^2} \sum_{i=0}^{n_{\text{bins}}} \sigma_{\text{bsc}}(D_i, \lambda) \ N(D_i) \ \Delta D_i \]
where $\lambda$ is the radar frequency, $|K|^2$ is the dielectric constant (0.93) and $\sigma_{\text{bsc}}(D_i, \lambda)$ is the backscatter cross-section of a spherical raindrop of maximum dimension $D_i$. The $\sigma_{\text{bsc}}(D_i, \lambda)$ is determined using T-matrix theory [63] as implemented by the pytmatrix python package ([64] https://github.com/jleenonen/pytmatrix). As noted previously, the spherical assumption may not always be correct, and assuming a sphere results in an average overestimation of $Z_e$ by approximately 6%.

In order to classify PSD points into convective or stratiform (Section 3.1), additional PSD parameters are required. One of the additional parameters is the normalized intercept parameter ($N_w$), which is defined as

$$N_w = \frac{4^4 \ LWC}{\pi \rho_l \ D_m^4}$$

where $\rho_l$ is the density of liquid water and $LWC$ is the liquid water content, defined as

$$LWC = \sum_{i=0}^{n_{\text{bins}}} m(D_i) \ N(D_i) \ \Delta D_i.$$  (8)

The last PSD parameter needed is known as the median volume diameter ($D_0$), which, when assuming the presence of a gamma size distribution shape, can be written as

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

where $\mu$ is the shape parameter of the three-parameter gamma distribution. The relation from [65] between the standard deviation of the mass distribution ($\sigma_m^2$) and $D_m$ is used to define $\mu$ by

$$\mu = \frac{D_m^2}{\sigma_m^2} - 4,$$  (10)

where $\sigma_m^2$ is defined as

$$\sigma_m^2 = \sum_{i=0}^{n_{\text{bins}}} (m(D_i) - D_m)^2 m(D_i) \ N(D_i) \ \Delta D_i \ / \ \sum_{i=0}^{n_{\text{bins}}} m(D_i) \ N(D_i) \ \Delta D_i.$$  (11)

2.2.2. Snowfall PSD Parameters

The conversion from a solid precipitation rate to liquid equivalent snowfall rate is straightforward and not novel to this study. Equation (1) is re-written with $m(D_i)$ and $\rho_l$, where $m(D_i)$ represents the retrieved mass estimate of a particle with $D_i$ from the PIP,

$$R = \sum_{i=0}^{n_{\text{bins}}} v(D_i) \ \frac{m(D_i)}{\rho_l} \ N(D_i) \ \Delta D_i.$$  (12)

Calculating a liquid equivalent $D_m$ is not as direct and not commonly done in the literature. The liquid equivalent diameter ($D_{\text{melted}}$) needs to be considered in Equation (2), which—if a sphere is assumed as the melted shape—can be determined by

$$D_{\text{melted}} = 3 \sqrt{\frac{6 m(D_i)}{\pi \rho_l}}.$$  (13)

Then, Equation (2) becomes

$$D_m = \frac{\sum_{i=0}^{n_{\text{bins}}} m(D_i) \ D_{\text{melted}} \ N(D_i) \ \Delta D_i}{\sum_{i=0}^{n_{\text{bins}}} m(D_i) \ N(D_i) \ \Delta D_i}.$$  (14)
The conversion of \( N(D_i) \) and \( \Delta D_i \) to a melted equivalent (e.g., \( N(D_{melted,i}) \); \( \Delta D_{melted,i} \)) is not done for the following reason. Consider Figure 1, where two sample volumes are depicted: one with frozen hydrometeors and the other with the liquid equivalent. If it is assumed that all particles within some bin are of the same particle type and that the melted version of the particle forms a sphere (i.e., no breakup), then the particles simply translate to smaller, higher density versions of themselves. This then preserves the original number of particles, and thus the product of \( N(D_i)\Delta D_i \) is equivalent to \( N(D_{melted,i})\Delta D_{melted,i} \), and the integrated parameters such as LWC and total number concentration are preserved.

In reality, all particles found in one size bin are probably not the same particle type (e.g., same habit, same degree of riming). To account for this, separate PSDs for different habits would need to be known. Since the variability of habits and particle densities within each size bin are not well known and are difficult to determine, the assumption of a single particle type is used.

In order to calculate \( Z_e \) for snowfall, Equation (6) is used. The only change is in how \( 
bsc(D_i, \lambda) \) is determined. To avoid the issues of underestimating the \( 
bsc(D_i, \lambda) \) when using Mie or T-matrix theory at higher operating frequencies (Ka-, W-band [66]), \( 
bsc(D_i, \lambda) \) is derived following the technique in [67]. The authors of that work simulated several different degrees of rimed aggregates, and their corresponding scattering properties were determined from the Discrete Dipole Approximation (DDA [68]). Since particles can have a variety of degrees of riming, and thus a variety of masses with the same maximum dimension, Equation (6) becomes

\[
Z_e = \frac{\lambda^4}{2\pi^5 |K|^2} \sum_{i=0}^{n_{bins}} \sigma_{bsc}(D_i, \lambda, m_i) N(D_i) \Delta D_i
\]

(15)

where \( \sigma_{bsc}(D_i, \lambda, m_i) \) is determined from the \( D_i \) and \( m(D_i) \) retrieved by the PIP. This is done using the kd-tree search algorithm from Scipy [69], which efficiently searches the [67] database of particles for the particle with the most similar \( D_i \) and \( m(D_i) \). This method was chosen instead of taking the median or mean over all particles simulated in order to allow the natural variability in \( \sigma_{bsc}(D_i, \lambda, m_i) \) to propagate into the forward calculation of \( Z_e \).

Figure 1. Schematic diagram to illustrate the conversion of the measured solid phase size distribution of particles to the liquid equivalent. The box is an instantaneous sample volume, and the particle diagrams are from ARTS ([70] https://zenodo.org/record/1175573). Each number corresponds to the same bin.
2.3. GPM-DPR Algorithm

A complete description of version 6 of the GPM-DPR algorithm can be found in the Algorithm Theoretical Basis Document (ATBD, [71]). The microphysical basis of the algorithm assumes the \( R \) and \( D_m \) are well correlated, and thus a mathematical relation can be formulated between the two parameters. Specifically, the relations used to relate \( R \) and \( D_m \) in the GPM-DPR algorithm are

\[
R = 1.370e^{4.258} D_m^{5.420} \tag{16}
\]

for convective precipitation and

\[
R = 0.401e^{4.649} D_m^{6.131} \tag{17}
\]

for stratiform precipitation. In both expressions, \( e \) is a diagnosed parameter that is constant in the atmospheric column and is used to reconcile estimates of path integrated attenuation (PIA) and the retrieved PSD parameters. The default value is \( e = 1 \), but this is logarithmically varied between 0.2 and 5. The relationship between \( Z_e, R \) and \( D_m \) can be calculated with some assumptions about the PSD.

In the GPM-DPR algorithm, the form of the PSD is assumed to be the normalized three-parameter \((N_w; D_m; \mu)\) gamma distribution [72]:

\[
N(D) = N_w \frac{6(\mu + 4)^{\mu+4}}{4 \Gamma(\mu + 4)} \left( \frac{D}{D_m} \right)^\mu \exp \left( \frac{(\mu + 4)D}{D_m} \right) \tag{18}
\]

The \( \mu \) parameter in the GPM-DPR algorithm is assumed to be three; thus, the PSD can be described using two parameters: \( N_w \) and \( D_m \). Equation (1) can then be rewritten with the GPM-DPR assumptions in integral form as

\[
R = \int_0^\infty v(D) vol(D) N_w \frac{6(\mu + 4)^{\mu+4}}{4 \Gamma(\mu + 4)} \left( \frac{D}{D_m} \right)^\mu \exp \left( \frac{(\mu + 4)D}{D_m} \right) dD \tag{19}
\]

where \( v(D) \) in the GPM-DPR algorithm follows [73]

\[
v(D) = 4.854D \exp (-0.195D). \tag{20}
\]

Similarly, Equation (6) can be rewritten as

\[
Z_e = \frac{\lambda^4}{\pi^5|K|^2} \int_0^\infty v(D) \sigma_{bsc}(D, \lambda) N_w \frac{6(\mu + 4)^{\mu+4}}{4 \Gamma(\mu + 4)} \left( \frac{D}{D_m} \right)^\mu \exp \left( \frac{(\mu + 4)D}{D_m} \right) dD \tag{21}
\]

where particles are assumed to be spherical and Mie theory is used to determine \( \sigma_{bsc}(D, \lambda) \). Since the relation between \( R \) and \( D_m \) is prescribed (Equations (16) and (17)), Equation (19) can be algebraically manipulated to solve for \( N_w \)

\[
N_w = \frac{R}{\int_0^\infty v(D) vol(D) \frac{6(\mu+4)^{\mu+4}}{4 \Gamma(\mu + 4)} \left( \frac{D}{D_m} \right)^\mu \exp \left( \frac{(\mu + 4)D}{D_m} \right) dD}. \tag{22}
\]

Then, all parameters needed to calculate \( Z_e \) for each \( R \) and \( D_m \) pair are known. In order to solve the integration of the indefinite integrals (Equations (21) and (22)), quadrature is used. Specifically, the Scipy [69] quadrature is used, which uses the FORTRAN QUADPACK [74] to perform automatic integration. The result of solving for \( Z_e \) is shown in Figure 2, assuming particles are in the liquid phase. Furthermore, the result of varying \( e \) within the \( R - D_m \) relation is shown in Figure 2b. Effectively, as expected, the \( Z_e \) value increases away from the origin (i.e., larger characteristic sizes and larger precipitation rates).
Figure 2. Precipitation rate ($R$) and mass weighted mean diameter ($D_m$) theoretical diagrams. (a) The default ($\epsilon = 1$) $R - D_m$ relation for both convective (solid red line) and stratiform (solid blue line) regimes with the equivalent radar reflectivity factor ($Z_e$) in logarithmic units (dBZ) is contoured (dashed rainbow lines) assuming rain. The $Z_e$ contours are determined using the same assumptions as those in the GPM-DPR algorithm for a range of $\epsilon$ values. The dashed black line is the 12 dBZ contour which is the GPM-DPR Ku-band minimum sensitivity. (b) Similar to (a), the rainbow contours are the $Z_e$, while the other lines are the stratiform relation, but now the $\epsilon$ parameter is varied. The default relation ($\epsilon = 1$) is in black, while $\epsilon > 1$ is in red shades and $\epsilon < 1$ is in blue shades.

The retrieval process can be found in the ATBD [71] but is described here briefly. Assuming that GPM-DPR has an observed profile containing precipitation echoes, the algorithm diagnoses whether the profile of $Z_e$ is convective or stratiform based on the vertical structure of $Z_e$. This is most simply done by considering if a melting layer can be detected in $Z_e$ (i.e., radar bright-band). Additional methods are used and can be found in Section 3.5 in the ATBD [71]. Then, the phase of each radar gate is assigned using temperature information from numerical weather prediction and the location of the radar bright-band if present. Radar gates found 500 m above the radar bright-band or at temperatures less than 0 $^\circ$C are designated as solid phase. The solid phase designation assumes that the particles contained within the gate are solid spherical particles with an effective density of 0.1 g cm$^{-3}$. Radar gates found below 500 m of the radar bright-band or temperatures greater than 0 $^\circ$C are designated as liquid phase (i.e., spherical with effective densities of 1 g cm$^{-3}$). After the phase is determined, the observed $Z_e$ and the initial estimate of $\epsilon$ ($\epsilon = 1$) are used to search for where the $R - D_m$ relation intersects the contoured $Z_e$, thus simultaneously retrieving $R$ and $D_m$. Once $R$ and $D_m$ are retrieved for all radar gates with precipitation echoes, the PIA and Ka-band $Z_e$ can be calculated. This process is repeated with different values of $\epsilon$, and the optimal value of $\epsilon$ is chosen by minimizing the difference of the retrieved PIA with the estimated PIA and the retrieved Ka-band $Z_e$ with the measured Ka-band $Z_e$ if available. If only Ku-band $Z_e$ is measured, then only the PIA error is used in the optimization. For a more complete description of the algorithm, consult the ATBD.

It should be emphasized that an $R - D_m$ relation derived from surface rain observations is used for the entire atmospheric column, regardless of whether the gates are identified to contain ice or snow. The only differences between the solid phase retrieval and the liquid phase retrieval is the complex index of refraction needed for the determination of $\sigma(D, \lambda)$, the $\nu(D)$ and a slightly modified version of Equation (18) (ATBD page 68). To the authors’ knowledge, the effect of using the rain $R - D_m$ relation on snow has not been investigated. It is hypothesized here that the use of a rainfall relation in snowfall retrieval could potentially account for the large low bias reported in [14].
3. Results and Discussion

3.1. Rainfall

The GPM-DPR algorithm documentation is not explicit regarding the origin of the $R - D_m$ relation, but it can be found in [28]. In that work, it was noted that the $R - D_m$ relations were derived using the methodology in [29] on a limited disdrometer dataset that would be tested on more PSD observations in future work. It is noted that the data used in [29] are derived from an impact disdrometer described in [75], which has been shown to potentially undercount small and large drops relative to the 2DVD [76]. In order to assist in these efforts, an investigation of the $R - D_m$ relationship in rainfall is conducted here. The 2DVD dataset provides an opportunity to evaluate whether the $R - D_m$ relation used in the GPM-DPR algorithm is general enough to apply to PSDs applicable to global precipitation measurements. All PSD data from the 2DVDs measured in rain are considered in a bulk sense in Figure 3, without separation into convective or stratiform regimes. The 2D histogram of the density of observations (Figure 3a) shows that the majority of points follow the general shape of the $R - D_m$ relations and lie between the convective–stratiform curves for $\epsilon = 1$ (Figure 3a, red and blue curves). The mode of all the $R - D_m$ observations (Figure 3b), also lies between the convective–stratiform relations, with it being centered on the stratiform curve for $D_m < 1.5$ mm and centered on the convective curve for $D_m > 1.5$ mm. This is consistent with expectation that convective rainfall generally produces larger $R$ and larger $D_m$.

The relationship with Ku-band $Z_e$ is shown in Figure 3c, where $Z_e$ increases with increasing $R$ and $D_m$. Overall, the 2DVD observations collected around the world support the use of the $R - D_m$ relation in retrievals of rainfall and are consistent with the current $R - D_m$ framework used within the GPM-DPR algorithm. Since the Ku-band $Z_e$ is calculated for each raining PSD, the retrieval of PSD parameters using the GPM-DPR microphysical assumptions (Section 2.3) can be emulated as if they were observed radar reflectivity values. This is done by using the relations illustrated in Figure 2a ($\epsilon = 1$). Then, for each raining PSD, the calculated $Z_e$ can be used to retrieve $R$ and $D_m$. Using these retrieved $R$ and $D_m$ values, error metrics can be derived by comparing the retrieved values to the calculated values from the PSD. Specifically, the mean absolute error ($MAE$), mean absolute percent error ($MAPE$), mean bias percentage ($MBP$) and root mean squared error ($RMSE$) are used, which are formulated as follows:

$$MAE = \frac{\sum_{i=1}^{n} |x_{r_i} - x_{c_i}|}{n};$$

$$MAPE = \frac{100 \times \sum_{i=1}^{n} |x_{r_i} - x_{c_i}|}{\sum_{i=1}^{n} x_{r_i}};$$

$$MBP = \frac{100 \times \sum_{i=1}^{n} \frac{x_{r_i} - x_{c_i}}{x_{r_i}}}{n};$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (x_{r_i} - x_{c_i})^2}{n}};$$

where $x_{r_i}$ is the $i^{th}$ retrieved value and $x_{c_i}$ is the $i^{th}$ calculated value. The calculation of the aforementioned error metrics for the entire 2DVD dataset is found in Table 2. Without the classification of the PSDs (i.e., convective or stratiform), the error is 40–50% and 20–30% for $R$ and $D_m$, respectively.
Figure 3. Investigating the relationship between \( R \) and \( D_m \) from all the ground-based 2DVD disdrometers in rain. (a) Counts of observations on the \( R - D_m \) plane normalized by the total number of observations and converted to percentages. Only bins with at least 10 observations are shown. The same relations of \( R - D_m \) and the minimum sensitivity from Figure 2a are shown for reference. (b) Same data as (a), but the data are normalized to the total number of observations in each bin of \( R \). A value near one implies that it is near the mode of observations in that bin of \( R \). (c) Same 2D histogram as (a), but colored by the bin median value of Ku-band calculated \( Z_e \) from the PSD assuming spheres and using a Tmatrix.

Table 2. Error statistics for the entire 2DVD dataset. MAE: mean absolute error; MAPE: mean absolute percentage error; MBP: mean bias percentage; RMSE: root mean square error.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Retrieval Error Metrics</th>
<th>Stratiform Relation</th>
<th>Convective Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R )</td>
<td>MAE, [mm hr(^{-1})]</td>
<td>1.28</td>
<td>1.77</td>
</tr>
<tr>
<td></td>
<td>MAPE, [%]</td>
<td>54</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>MBP [%]</td>
<td>-23</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>RMSE, [mm hr(^{-1})]</td>
<td>4.94</td>
<td>6.80</td>
</tr>
<tr>
<td>( D_m )</td>
<td>MAE, [mm]</td>
<td>0.21</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>MAPE, [%]</td>
<td>19</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>MBP [%]</td>
<td>-4</td>
<td>-22</td>
</tr>
<tr>
<td></td>
<td>RMSE, [mm]</td>
<td>0.31</td>
<td>0.37</td>
</tr>
</tbody>
</table>

From Figure 3, it is not immediately apparent that there should be two distinct \( R - D_m \) relations. To assess if two relations should be used, a convective–stratiform partitioning of the data could be used. Since thermodynamic profiles are unavailable for all raining instances in the database and reanalysis would likely struggle to diagnose the observed buoyancy on the spatial and temporal scale needed to diagnose convection, there is no way to categorize the data based on thermodynamic variables, and coincident radar data are not available to perform a radar-based separation [77]. Thus, the only viable way to categorize the data into convective and stratiform regimes is to use relationships derived from the PSD parameters themselves, such as the relationship between \( N_w \) and \( D_0 \), as used in [78–80]. In [79], the authors showed that PSDs with a \( \log(N_w) \) greater than that given by the expression

\[
\log(N_w) = -1.6D_0 + 6.3
\]  

were diagnosed as convectively produced, while PSDs with \( \log(N_w) \) less than Equation (27) were produced by stratiform vertical motions. In [80], the authors continued this work and added an additional constraint for weak convective instances, stating that \( \log(N_w) > 3.85 \) corresponds to convective and \( \log(N_w) < 3.85 \) to stratiform regions; however, the work presented in [78] noted that a convective–stratiform separation based solely on \( N_w \) may only apply in warm-rain-dominated convection over tropical oceans. In [78], the authors used Empirical Orthogonal Functions (EOF) to diagnose stratiform and convective rain and created polygons enclosing the relevant regions on the \( N_w - D_0 \) plane. Here, the polygons in that work are used to determine the classification (Figure 4a).
The result of separating the observations shown in Figure 3 into stratiform and convective categories is shown in Figure 4. The observations show a clear separation, with the convective-labeled points closer to the convective $R - D_m$ relation, while the stratiform-labeled points are indeed closer to the stratiform $R - D_m$ relation (Figure 4b). Similarly, the convective and stratiform $R - D_m$ relations lie close to the mode of their respective classified points (Figure 4c). The same error metrics from Table 2 are recalculated for the category-specific data and shown in Table 3. All metrics improve by sub-setting the data, except for the convective rainfall rate retrieval. This is likely a result of most of the points being located at a smaller $R$ than the $R - D_m$ relation, centered around $D_m = 1.2$ mm for the convective labeling. However, the MAPE is similar, at 43% and 47% for the “not classified” and “classified” data, respectively.

![Figure 4](https://example.com/figure4.png)

**Figure 4.** Convective–stratiform partitioning based on PSD parameters. (a) Normalized intercept parameter ($N_w$) median volume diameter ($D_0$) plane of all raining PSDs. The red points are labeled as convective and blue are labeled as stratiform based on the classification in [78]. The polygons from [78] are drawn in solid blue (stratiform) and dashed red (convective). (b) Same as Figure 3a, but with the separated categories. Red is convective, blue is stratiform. The lighter the color, the higher the density of points. (c) Same as Figure 3b, but with the separated categories. Colors are the same as (b).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Retrieval Error Metrics</th>
<th>Stratiform Relation</th>
<th>Convective Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R$</td>
<td>MAE, [mm hr$^{-1}$]</td>
<td>0.14</td>
<td>3.37</td>
</tr>
<tr>
<td></td>
<td>MAPE, [%]</td>
<td>29</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>MBP, [%]</td>
<td>6</td>
<td>−20</td>
</tr>
<tr>
<td></td>
<td>RMSE, [mm hr$^{-1}$]</td>
<td>0.43</td>
<td>8.57</td>
</tr>
<tr>
<td>$D_m$</td>
<td>MAE, [mm]</td>
<td>0.16</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>MAPE, [%]</td>
<td>18</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>MBP, [%]</td>
<td>−14</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>RMSE, [mm]</td>
<td>0.22</td>
<td>0.27</td>
</tr>
</tbody>
</table>

### 3.2. Snowfall

In this section, the entirety of the snowfall PSDs observed in Finland are analyzed in the same bulk way that was done for the rainfall results presented above. Figure 5 shows the results of calculating the liquid equivalent $R$ and liquid equivalent $D_m$ from the snowfall PSDs. The highest density of observations is found at a smaller $D_m$ and larger $R$ than prescribed by the $R - D_m$ relation with $\epsilon = 1$ (Figure 5a). Furthermore, the ranges of both $R$ and $D_m$ in snow (Figure 5) are less than that of rainfall (Figure 3). The mode of the snowfall distribution is closest to the stratiform curve for $0.5 \text{ mm} < D_m < 0.75 \text{ mm}$ but then deviates from both empirical relations at a larger $D_m$ (Figure 5b). The calculated Ku-band $Z_v$ (Figure 5c) increases with increasing $R$ and $D_m$, similar to the...
rainfall data analysis shown above. The snowfall PSD observations from Finland suggest that the \( R - D_m \) relationship with \( \epsilon = 1 \) does not fit well and could be a large source of error in the algorithm.

Figure 5. As in Figure 3, but now for the snowfall PSDs collected in Finland. The x and y axis are in their liquid equivalent values (see Section 2.2.2) (a) Counts of observations on the \( R - D_m \) plane normalized by the total number of observations and converted to percentages. The same relations of \( R - D_m \) Figure 2a are shown for reference. (b) Same data as (a), but the data are normalized to the total number of observations in each bin of \( R \). A value near one implies that it is near the mode of observations in that bin of \( R \). (c) Same 2D histogram as (a), but colored by the bin median value of Ku-band calculated \( Z_e \) from the PSD using the method described in Section 2.2.2.

In order to quantify error in the same manner as before, the GPM-DPR algorithm is emulated, but now using the solid phase assumptions. The retrieval metrics are shown in Table 4. There are errors of 77% and 49% for the retrieval of \( R \) using the stratiform and convective relation, respectively. This is a +47% increase in error when compared to the stratiform relation and stratiform-classified PSDs in rain. The error for the retrieval of \( D_m \) is not as large as the retrieval of \( R \), at 24% and 20% for the stratiform and the convective relation, respectively.

Table 4. Error metrics from the snowfall data measured in Finland.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Retrieval Error Metrics</th>
<th>Stratiform Relation</th>
<th>Convective Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R )</td>
<td>MAE, [mm hr(^{-1})]</td>
<td>0.24</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>MAPE, [%]</td>
<td>77</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>MBP, [%]</td>
<td>-52</td>
<td>-2</td>
</tr>
<tr>
<td></td>
<td>RMSE, [mm hr(^{-1})]</td>
<td>0.47</td>
<td>0.54</td>
</tr>
<tr>
<td>( D_m )</td>
<td>MAE, [mm]</td>
<td>0.21</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>MAPE, [%]</td>
<td>24</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>MBP, [%]</td>
<td>18</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>RMSE, [mm]</td>
<td>0.25</td>
<td>0.18</td>
</tr>
</tbody>
</table>

One potential way to improve retrievals could be to fit new coefficients to Equations (16) and (17), but the snowfall PSD data suggest that an \( R - D_m \) framework may not be advantageous for retrievals in snowfall because of the lack of correlation between \( \log(R) \) and \( D_m \) (Table 5). Snowfall PSDs have a correlation value of 0.29, which is approximately half of the correlation value derived from rain, at 0.61. The poor correlation in snow persists regardless of whether the conversion of \( D_m \) to liquid equivalent is considered (Table 5). Thus, even fitting new coefficients to Equations (16) and (17) might not result in optimal snowfall property retrieval.
Table 5. Pearson ρ correlation values between log (R) and Dm for the entire rain dataset and the snowfall measurements in Finland.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Pearson ρ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rain</td>
<td>0.61</td>
</tr>
<tr>
<td>Snow</td>
<td></td>
</tr>
<tr>
<td>Liquid Eq. Dm</td>
<td>0.29</td>
</tr>
<tr>
<td>Frozen Dm</td>
<td>0.24</td>
</tr>
</tbody>
</table>

3.3. Investigation of Poor Correlation between R and Dm in Snowfall

Two hypotheses are formulated to explain the poor correlation between R and Dm in snow. One hypothesis is that the velocity–diameter relation and its natural variability in snow causes a reduction in the correlation between R and Dm. The second hypothesis is that the mass–diameter relation of snow and its natural variability causes a reduction in the correlation between R and Dm. Both hypotheses are tested by imposing velocity and mass relations derived from snowfall into the calculation of R and Dm in rain. The goal of replacing the velocity and mass relations in this way is to test the sensitivity of the correlation of R and Dm and whether the poor R − Dm correlation in snow can be reproduced by the rain dataset. The snowfall speed and mass relations are derived from the PIP snowfall observations, where the particle’s observed maximum dimension is related to fall velocity and mass. The mean fall velocity and mass for the entire dataset is shown in Figure 6 (blue) alongside the default assumptions for the rainfall (red) and some parameterizations from previous studies (black). In order to test the first hypothesis, the fall velocity (e.g., v(Di) in Equation (1)) is sampled from a normal distribution with a mean and standard deviation of that in snow; in other words, the fall velocity at any diameter Di is randomly sampled from the normal distribution with the mean and standard deviation determined from the data in Figure 6a. Similarly, in order to test the second hypothesis, the mass (e.g., m(Di) in Equations (1) and (2)) is sampled from the mean and standard deviation of snow particle masses.

Figure 6. Fall speed and mass diameter relations for rain and snow. (a) Particle fall velocity for rain (red) determined by the empirical relation from [45]. Mean (blue centerline) and one standard deviation (blue bars) particle fall velocity for snow measured by the PIP and the predicted fall velocity at −1 °C from [81] (black). (b) Particle mass for rain (red) determined by assuming a spherical shape. Particle mass for snow as retrieved from the PIP and predicted by the relation in [82].

For these tests, the following substitutions are made: replace v(Di), replace m(Di) and replace both v(Di) and m(Di). The results of these three experiments are shown in Figure 7, and their respective correlations are reported in Table 6. Replacing only the fall speed relation results in a 0.08 reduction in
the correlation compared to the original rainfall analysis \((\rho = 0.61; \text{Figure 3a})\). As expected, there is a reduction in the magnitude of \(R\) since the fall velocity magnitude has been decreased and \(D_m\) has remained the same because \(D_m\) is not a function of \(v(D_i)\) (Figure 7a). Substituting the mass relation results in a much larger magnitude reduction in the correlations \((-0.41)\), yielding similar correlation values \((0.2)\) to those found from the snowfall data. Figure 7b shows that there is now not only a reduction in the magnitude of \(R\) but also in \(D_m\). Finally, replacing both relations results in the same correlation as swapping the \(m(D_i)\). It should be noted that the experiments shown in Figure 7 are applied separately and randomly (i.e., through the normal distribution) despite knowing that the \(m(D_i)\) is not mutually exclusive from \(v(D_i)\) (e.g., more massive particles fall faster, absence of drag). Thus, these experiments are likely missing the co-variability between \(v(D_i)\) and \(m(D_i)\) that could improve the calculated correlation values.

![Figure 7. Recalculating \(R\) and \(D_m\) for rain using measured snowfall relationships. (a) Difference in the normalized counts between the new calculation of \(R\) and \(D_m\) using the measured snowfall velocity relationship (shown in Figure 6a) and the original calculation (same as Figure 3a). (b) Same as (a), but with \(R\) and \(D_m\) calculated with the measured snow mass relation (shown in Figure 6b).](image)

<table>
<thead>
<tr>
<th>Snow (v(D))</th>
<th>Snow (m(D))</th>
<th>Snow (v(D)) and (m(D))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson (\rho)</td>
<td>0.53</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Substituting the mean mass relation from the PIP observations above is effectively using one \(a\) and \(b\) value in the common parameterization of ice particle mass:

\[
m(D) = aD^b.
\]  

The effective curve fit values of \(a\) and \(b\) from all the snowfall data presented here are \(0.1\ \text{g cm}^{-b}\) and \(2.58\), respectively, determined by the Scipy [69] curve fit method. In order to determine how the correlation varies across a wide range of possible \(a\) and \(b\), the formulation for mass from Equation (28) is adopted, but \(a\) and \(b\) are systematically varied when inserted into Equation (1) and Equation (2). It should be noted that typical values of \(a\) and \(b\) in ice and snow are \(10^{-3}\ \text{g cm}^{-b}\leq a \leq 10^{-1}\ \text{g cm}^{-b}\) and \(1 \leq b \leq 3\), respectively [83]. The fall velocity relation used here is the same as the original rainfall analysis [45]. First, the result allowing no variability in the mass (i.e., the mass is exactly prescribed by Equation (28)) is investigated (Figure 8a) and shows that the correlation between \(R\) and \(D_m\) is only a function of \(b\), which can be expected, as the \(a\) parameter will cancel out in Equation (2). Note that the magnitude of the correlation for the fits \(a\) and \(b\) for the snowfall data, which is between 0.55 and 0.575, is much larger than that reported in Figure 7b (0.20). This shows that the magnitude of the mass is less important when determining the correlation of \(R\) and \(D_m\), while the natural variability of mass for any particle size determines the correlation between \(R\) and \(D_m\). To illustrate this further, the \(a\) and \(b\) values are systematically varied again, but now the natural variability in hydrometeor
mass is added in a similar to that done previously by sampling a normal distribution with the mean predicted using Equation (28) and the standard deviation determined by the snowfall observations (Figure 6b). The results are shown in Figure 8b and show that, for larger \( a \) values \((a \geq 0.75 \text{ g cm}^{-b})\) a similar relationship to Figure 8a is found. But for small \( a \) values, where many empirical fits for snow and ice have been found \((10^{-3} \text{ g cm}^{-b} \leq a \leq 10^{-1} \text{ g cm}^{-b} \text{ and } 1 \geq b \geq 3; \text{ see [83] Figure 1})\), the correlation is low for any value of \( b \). Thus, for a majority of ice and snow, it is likely that the correlation between \( R \) and \( D_m \) is poor. Doing the same type of analysis on the snowfall PSDs (e.g., removing the natural variability and replacing it with the rainfall relations) resulted in improved correlations between \( R \) and \( D_m \) to about 0.5, which is the same value achieved for rain with static \( a \) and \( b \) values (Figure 8a).

![Figure 8](image_url)

**Figure 8.** Investigating the variance of the correlation between \( R \) and \( D_m \) with various \( a \) and \( b \) values in the calculation of mass \((m(D) = aD^b)\). (a) No variability in Equation (28) is used. The hexagon is the fit \( a \) and \( b \) values to the entirety of the Finland snowfall data, while the circle is the values for a liquid sphere. (b) Same as (a), but now the mass is randomly sampled from a normal distribution with the mean predicted by Equation (28) and the standard deviation given from the observations shown in Figure 6b.

4. Conclusions

Since being launched in 2014, the Global Precipitation Measurement (GPM) Dual-Frequency Precipitation Radar (DPR) has collected copious equivalent radar reflectivity factor \((Z_e)\) measurements of both rain and snow. Microphysical parameters of interest, such as the precipitation rate \((R)\) and mass weighted mean diameter \((D_m)\), are retrieved and published. An evaluation of \( R \) retrieved for snowfall by GPM-DPR reported the current status of the GPM-DPR algorithm, showing an approximately 50% low bias on the global mean snowfall rate compared to CloudSat [14]. In order to investigate the potential causes of this low bias, the main microphysical assumption that the particle size distribution (PSD) follows an \( R - D_m \) retrieval framework was evaluated. The principal conclusions are as follows:

1. Assuming that raindrops are spherical, ground-based 2D-video disdrometer measurements of rainfall from various geographic locations show good agreement with the current GPM-DPR algorithm framework, with the majority of observations being found near the prescribed GPM-DPR relation and a Pearson-\( \rho \) correlation coefficient of 0.61 between \( \log(R) \) and \( D_m \) (Figure 3, Table 5).
2. The classification of PSDs as convective or stratiform according to the method in [78] (e.g., based on the normalized intercept parameter \((N_w)\) and the median volume diameter \((D_0)\)) is consistent with and supports the use of two separate \( R - D_m \) relations for each rainfall class, as they reduce error compared with the use of a single \( R - D_m \) relation (Figure 4, Table 3).
3. Ground-based Precipitation Imaging Probe measurements of snowfall in Finland do not show the same consistency as the \( R - D_m \) retrieval framework compared to rainfall. The error using
the GPM-DPR stratiform relation is much larger (+47% comparing Table 3 to Table 4), and the Pearson-ρ correlation between \( \log(R) \) and \( D_m \) is considerably lower (Pearson-ρ = 0.29; Figure 5; Table 5).

4. The variability in the mass of particles with a similar maximum dimension likely causes most of the poor correlation between \( R \) and \( D_m \) in snow (Figures 7 and 8).

The analysis provided here suggests that the use of an \( R - D_m \) relation derived in rain is inappropriate for use in snowfall. A potential alternative is to default to old techniques for deriving snowfall rate, such as the power-law fit between \( Z_e \) and \( R \). While there can be an order of magnitude of uncertainty on single frequency \( Z_e - R \) relations depending on the assumed particle type [10], this is more physically based than the current GPM-DPR algorithm (version 6) in snowfall. One advantage that GPM-DPR has over its predecessors is a second operating frequency (Ka-band). Having a second frequency should allow for improved retrieval results compared to a single frequency, as shown by [84]. As a potential avenue for retrieving \( R \) and \( D_m \), the authors are pursuing the use of neural networks to allow for an unsupervised approach to retrieve snowfall parameters. This was done first in [85] but deserves renewed investigation since machine learning methods have improved and new multiple frequency datasets exist for retrieval implementation and evaluation (e.g., GCPEX [27]; OLYMPEX/RADEX [26]).

**Author Contributions:** Conceptualization, R.J.C., S.W.N. and G.M.M.; methodology, R.J.C., S.W.N. and G.M.M.; software, R.J.C.; validation, R.J.C.; formal analysis, R.J.C.; investigation, R.J.C., S.W.N. and G.M.M.; resources, R.J.C.; data curation, R.J.C.; writing—original draft preparation, R.J.C.; writing—review and editing, R.J.C., S.W.N. and G.M.M.; visualization, R.J.C.; supervision, S.W.N. and G.M.M.; project administration, S.W.N. and G.M.M.; funding acquisition, S.W.N. and G.M.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** Funding for this research was provided to the University of Illinois by NASA Precipitation Measurement Missions grant 80NSSC19K0713 and NASA Earth System Science Fellowship 80NSSC17K0439.

**Acknowledgments:** The authors thank the two anonymous reviewers and the assistant editor for taking the time to review and enhance the manuscript. We also acknowledge the fruitful scientific discussions with Shinta Seto and Toshio Iguchi. Furthermore, we thank all the participants of the field campaigns used here for their tireless effort in collecting the data used in this study. Additionally, we would like to thank Patrick Gatlin for providing some of the NASA processed 2DVD data. Finally, we would like to thank Annakaisa von Lerber for her dedication and effort in providing the PIP dataset processing that was vital to the conclusions of this manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**


58. ARM DOE Data Discovery. Available online: https://adc.arm.gov/discovery/ (accessed on 10 June 2020).


80. Thompson, E.J.; Rutledge, S.A.; Dolan, B.; Thurai, M. Drop size distributions and radar observations of convective and stratiform rain over the equatorial Indian and West Pacific Oceans. *J. Atmos. Sci.* 2015, 72, 4091–4125. [CrossRef]


