Correspondence to:

Y. You,
yyou@umd.edu

Evaluation of Rainfall-Snowfall Separation Performance in Remote Sensing Datasets

Yalei You[1] ![ORCID], **Christa Peters-Lidard[2]** ![ORCID], **Sarah Ringerud[3]**, and **John M. Haynes[4]** ![ORCID]

[1]Earth System Science Interdisciplinary Center/Cooperative Institute for Climate and Satellites, University of Maryland, College Park, MD, USA, [2]NASA Goddard Space Flight Center, Greenbelt, MD, USA, [3]Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA, [4]Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, USA

Abstract The first step to accurately measure global snowfall is to separate rainfall from snowfall correctly (i.e., precipitation phase discrimination). This study first evaluates the phase discrimination performance in four remote sensing datasets, including observations from ground radar, spaceborne radars, and spaceborne radiometer, relative to ground observations. Results show that the snowfall discrimination accuracy varies greatly among these datasets ranging from 42% to 96%, dependent on whether and how the temperature information are considered. For example, over half of the snowfall from the Global Precipitation Measurement Mission (GPM) spaceborne radar is actually rainfall at the surface since it detects snowfall in the air without considering the temperature information close to the surface. Second, we evaluate the discrimination performance using the temperature information from four reanalysis datasets. It is found that MERRA2 temperature close to the surface is colder than the other three datasets, leading to more rainfall being misclassified as snowfall.

Plain Language Summary Satellite remote sensing provides the only means of measuring rainfall/snowfall on the global scale. Misclassifying the precipitation phase (i.e., rainfall as snowfall, or vice versa) could lead to the estimated precipitation rate being one order of magnitude smaller or larger. Our results reveal that the snowfall discrimination accuracy varies greatly among four remote sensing datasets ranging from 42% to 96%. For example, over half of the snowfall from the state-of-the-art precipitation product based on the Global Precipitation Measurement radar is rainfall at the surface without considering the temperature information close to surface. Additionally, the temperature discrepancy among different reanalysis datasets also greatly affects precipitation phase discrimination. Our results show that MERRA2 temperature close to the surface is colder than the other three major datasets, leading to more rainfall pixels being misclassified as snowfall pixels.

1. Introduction

Accurate global snowfall measurement is needed for many applications including water resources management (Gergel et al., 2017), water budget evaluation (Sheffield et al., 2009), and long-term climate change monitoring (OGorman, 2014). Satellite remote sensing provides the only means for the snowfall measurement on the global scale. The CloudSat Cloud Profiling Radar (CPR) observations (Stephens et al., 2002) have been used widely to investigate global snowfall features (Kulie & Bennartz, 2009; Kulie et al., 2016; Liu, 2008; Milani et al., 2018). Snowfall characteristics have also been characterized by the dual frequency precipitation radar (DPR) on board the Global Precipitation Measurement Mission (GPM) Core Observatory Satellite (Adhikari et al., 2018; Hou et al., 2014; Skofronick-Jackson et al., 2017, 2019), though the GPM DPR has a limited capability to measure light snowfall events with detection sensitivity at about 12 dBZ (Hamada & Takayabu, 2016; Panegrossi et al., 2017; Skofronick-Jackson et al., 2019). In addition, passive microwave radiometers are also commonly used for global snowfall estimation (Ebtehaj & Kummerow, 2017; Kidd et al., 2016; Kongoli et al., 2003; Kummerow et al., 2015; Meng et al., 2017; You et al., 2015). Ground radar observations also provide valuable snowfall information on the regional and continental scale. For example, the Multi-Radar Multi-Sensor (MRMS) provides snowfall estimation over the continental United States at 2-min and about 1-km resolution (Kirstetter et al., 2012; Zhang et al., 2016).

It remains extremely challenging to accurately measure falling snow from both ground and spaceborne radars (Chen, Gourley, et al., 2016; Kulie et al., 2021; Pettersen et al., 2020; Skofronick-Jackson et al., 2019), and microwave radiometers (Kummerow et al., 2015; Meng et al., 2017; Milani et al., 2021; Wang et al., 2013; You et al., 2017). A major obstacle is how to accurately determine the precipitation phase (i.e., separating rainfall from snowfall). Precipitation phase misclassification (e.g., rainfall to snowfall, or vice versa) could lead to the estimated precipitation rate from spaceborne radar being one order of magnitude smaller or larger (Sims & Liu, 2015). On the other hand, there exists very limited information in the remote sensing measurements themselves for precipitation phase determination. Specifically, a bright-band in the radar profiles may indicate rainfall at the surface (Ryzhkov & Zrnic,, 1998), while not all radar profiles have a bright-band. Furthermore, the uncontaminated bin height from spaceborne radar observations can vary from ~500 m to ~3 km above the ground level (Casella et al., 2017; Maahn et al., 2014), while the precipitation phase may change from that height to the surface. The brightness temperature (TB) observations from passive microwave radiometers do not directly indicate the precipitation phase at the surface. Very cold TB may be associated with hail events (Ferraro et al., 2015; Mroz et al., 2017). However, it is difficult to distinguish snowfall from rainfall by using TB only.

Researchers have long sought to exploit temperature-related parameters for precipitation phase discrimination. For example, early studies based on limited station observations using the surface air temperature revealed that the rainfall-snowfall transition threshold value is close to 2°C (Auer, 1974; United States Army Corps of Engineers, 1956). Later studies showed that other variables besides the surface temperature can also be used to constrain the rainfall-snowfall separation, including surface pressure (Dai, 2008; Sims and Liu, 2015), land surface elevation (Ding et al., 2014; You et al., 2016), relative humidity (Behrangi et al., 2018; Heymsfield et al., 2021; Jennings et al., 2018; Matsuo et al., 1981), and temperature lapse rate (Haynes et al., 2009; Sims and Liu, 2015). The mean temperature (or thickness) from the surface to the low or mid-troposphere has also been widely used to determine the precipitation type (Bourgouin, 2000; Wagner, 1957). In addition, Jennings et al. (2018) pointed out that the snowfall-rainfall transition temperature varies from −0.4°C to 2.4°C across the Northern Hemisphere, further adding complexity and difficulty for accurate precipitation phase partitioning. Recent work by Pettersen et al. (2021) showed that video disdrometer observations can provide accurate precipitation phase information.

The first objective of this study is to evaluate precipitation phase discrimination performance in four commonly used precipitation remote sensing datasets. They are the CloudSat CPR precipitation product (2C-PRECIP-COLUMN) (Haynes et al., 2009), GPM DPR precipitation product (2ADPR) (Hamada & Takayabu, 2016; Iguchi, 2020; Le et al., 2017), GPM Microwave Imager (GMI) precipitation product (2AG-PROFGMI) (Kummerow et al., 2015), and the MRMS precipitation product in the Continental United States (CONUS) (Zhang et al., 2016). More details regarding the precipitation phase determination in these products will be provided in the following sections. In addition, several phase segregation methods exist in the literatures, and temperature related parameters often are obtained from different reanalysis datasets. Therefore, the second objective is to assess the phase discrimination accuracy by using different methods and reanalysis datasets.

2. Datasets and Methodology

2.1. Datasets

The reference ("truth") data set for this study is the precipitation phase code reports (i.e., rainfall or snowfall) from the NOAA National Centers for Environmental Information (NCEI)'s Integrated Surface Database (ISD), which consists of global hourly and synoptic observations (3-hourly and 6-hourly observations) from over 35,000 stations worldwide (Smith et al., 2011). This data set has also been used to train precipitation phase discrimination schemes (Behrangi et al., 2018; Dai, 2008; Liu, 2008; Sims & Liu, 2015; You et al., 2016) and compute precipitation frequencies (Dai, 2001; Petty, 1995).

The precipitation phase information in the remote sensing datasets to be compared in this study is from CloudSat CPR precipitation product (2C-PRECIP-COLUMN, version 5) (Haynes et al., 2009), GPM DPR precipitation product (2ADPR, version 6) (Hamada & Takayabu, 2016; Iguchi, 2020; Le et al., 2017), GPM

Microwave Imager (GMI) precipitation product (2AGPROFGMI, version 5) (Kummerow et al., 2015), and the MRMS precipitation (operational version) (Zhang et al., 2016).

Specifically, for CPR 2C-PRECIP-COLUMN, we obtain the "Precip_flag" variable, which groups the precipitation phase into one of three categories: rain, snow, and mixed. For DPR, we obtain the "phaseNearSurface" and "flagSurfaceSnowfall." The first variable is determined by the temperature at the near surface level and the bright-band information (Iguchi, 2020). The second variable is computed from the difference between Ku-band PR (KuPR) and Ka-band PR (KaPR) at the clutter-free height, which varies from 0.5 to 3 km over different surface types and scan angles (Casella et al., 2017; Le et al., 2017). For GPM Microwave Imager (GMI) retrieval from the Goddard Profiling Algorithm (GPROF), we obtain surface precipitation rate ("surfacePrecipitation") and snowfall rate ("frozenPrecipitation"). For MRMS, we obtain the precipitation flag variable ("pcp_flag").

To collocate the remote sensing datasets with the surface reference data, we use the threshold values of 10 min and 10 km for CPR, DPR, and GMI observations. That is, when remote sensing observations and the surface reference data are less than 10 min apart in time and less than 10 km away in distance, they are considered as coincident observations. These two threshold values (10 min and 10 km) are selected by considering the trade-off between the sample size (primarily for the CPR nadir-only observations) and the accuracy of coincident observations. It is worth mentioning that the spatial resolution for CPR, DPR, and GMI precipitation products are about 1.6 km, 5.2 km, and 14.3 km, respectively. The ancillary temperature information used in these products for precipitation phase determination is often much coarser than these spatial resolutions. For MRMS, we simply find the nearest time and the closest distance with the surface data due to its high spatial (about 1 km) and temporal (2 min) resolutions. For the collocation period, we use the full CPR observation record from 2006 to 2017. For DPR and GMI products, we use observations from March 2014 (launch of the GPM satellite) to December 2018. For MRMS, we only use the data in 2016 with large enough sample size (>1 million collocated samples) since the ground radars used to generate the MRMS precipitation data set often are on the same location with the ground gauges, from where the precipitation phase reports are obtained.

As mentioned previously, the temperature information for the phase determination in remote sensing products is often obtained from model outputs. Our second objective is to evaluate the phase discrimination accuracy when the ancillary temperature information is obtained from different sources. To this end, this study compares four global reanalysis datasets, including the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA2) (Gelaro et al., 2017), the ECMWF Reanalysis fifth generation (ERA5) (Hersbach et al., 2020), the Japanese 55-year Reanalysis (JRA55) (Kobayashi et al., 2015), and the Global Forecast System (GFS) (Kanamitsu et al., 1991). The spatial resolutions are $0.5° \times 0.625°$ for MERRA-2, $0.25° \times 0.25°$ for ERA5, $1.25° \times 1.25°$ for JRA55, and $0.5° \times 0.5°$ for GFS, respectively. The temporal resolutions for the surface temperature, 2-m air temperature, and the surface pressure are hourly for MERRA2 and ERA5, and 6-hourly for JRA55 and GFS, respectively. The temporal resolutions for the temperature profile and geopotential height profile are hourly for ERA5, 3-hourly for MERRA2, and 6-hourly for JRA55 and GFS, respectively. For all these model datasets, we match the surface weather report data with the closest grid and linearly interpolate the temperature information in the temporal dimension.

2.2. Evaluation Metric

We first evaluate the precipitation detection performance from these four remote sensing datasets relative to the ground observations. The commonly used detection statistical metrics are computed, including Probability of Detection (POD), False Alarm Rate (FAR), and Heidke Skill Score (HSS).

Second, we assess the phase discrimination performance in the four remote sensing datasets using the snowfall or rainfall success percentage. We take the snowfall and the CPR as an example to show the definition of the success percentage. The snowfall success percentage is computed as the number of snowfall observations from CPR divided by the number of the snowfall observations from the reference. Similar computations are applied to rainfall success percentage and to other remote sensing datasets. We would like to emphasize that the snowfall or rainfall success percentage is computed when both the remote sensing instruments and the surface reference data set detect precipitation. This requirement normalizes for any
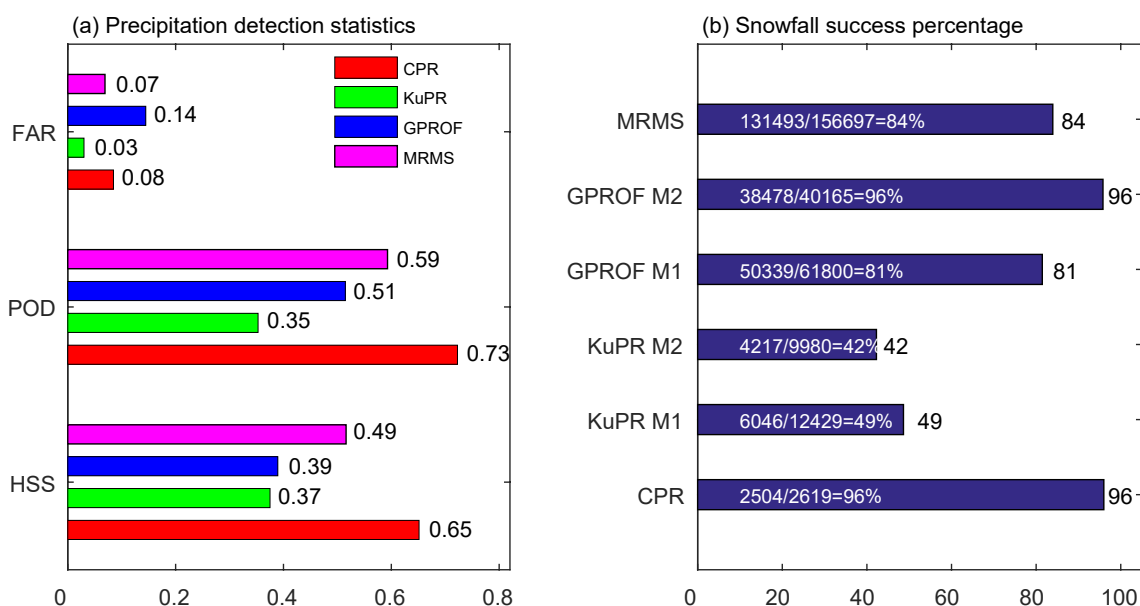
Figure 1. (a) The precipitation detection statistics, including Heidke Skill Score (HSS), Probability of Detection (POD), and False Alarm Rate (FAR), from four precipitation datasets (CloudSat CPR, Global Precipitation Measurement Mission (GPM) Ku-band PR (KuPR), GPROF, and MRMS) validated against surface observations; (b) The snowfall success percentage from four precipitation remote sensing datasets, including CloudSat Cloud Profiling Radar (CPR), GPM Dual Frequency Precipitation Radar (DPR), GPM GMI precipitation retrieval results from the Goddard Profiling Algorithm (GPROF) algorithm, and the Multi-Radar Multi-Sensor (MRMS). The denominator on each bar is the snowfall number from the ground weather report, while the numerator is the number from each remote sensing data set. DPR M1 and DPR M2 represent two methods used in DPR for the precipitation phase discrimination. GPROF M1 and DPR M2 represent two methods used in GPROF retrieval for the precipitation phase discrimination (see corresponding text for more details). MRMS data are only available over the Continental United States (CONUS).

differences in detection capabilities. For example, it is known that CPR has a better detection sensitivity compared with both DPR and ground radars in MRMS datasets (Skofronick-Jackson et al., 2019; Wang et al., 2018).

3. Results

This section first evaluates the precipitation detection performance from these four precipitation datasets. Then, we use a contingency table analysis to determine the snowfall success percentage among the four datasets. This analysis also explains why the snowfall success percentages differ greatly across these datasets. Finally, we analyze the snowfall (rainfall) success percentages from six different phase determination schemes using ancillary parameters from four reanalysis datasets.

3.1. Precipitation Detection Performance

Figure 1a shows the precipitation detection statistics from four aforementioned precipitation datasets relative to the ground observations. It is immediately clear that CPR has the best detection performance indicated by the largest HSS value of 0.65 when validated against the ground observations. This is an expected result because CPR has the best detection capability with the minimum detection reflectivity at~−15 dBZ (Durden et al., 2011; Haynes et al., 2009; Stephens & Haynes, 2007). In contrast, the precipitation detection performance from KuPR and GPROF are poor, with HSS values being 0.37 and 0.39, respectively. With a minimum detection reflectivity of ~12 dBZ (Hamada & Takayabu,, 2016), KuPR misses many light precipitation events, leading to a small POD value of 0.35. Previous work also showed the poor detection capability of KuPR for light precipitation (e.g., Casella et al., 2017). The large misidentified precipitation occurrence indicated by the large FAR value of 0.14 is primarily responsible for the poor performance of GPROF. The detection performance form MRMS is in between of CPR and KuPR/GPROF with a HSS value of 0.49. It is

worth mentioning that similar statistical values are obtained when only using observations over CONUS (i.e., the MRMS covered region) from CPR, KuPR, and GPROF.

3.2. Snowfall Success Percentage

As mentioned previously, there are two methods used in the DPR product to determine the precipitation phase. The first method (hereafter referred to as DPR M1) separates the snowfall from rainfall using temperature information at the clutter-free height (varying from 0.5 to 3 km), combined with bright-band information if it exists in the radar reflectivity profile. The second method uses the KuPR and KaPR radar reflectivity difference at the clutter-free height (varying from 0.5 to 3 km, hereafter referred to as DPR M2).

Figure 1b shows the snowfall success percentage from CPR, DPR, GPROF, and MRMS. It is immediately clear that both DPR methods show rather poor performance with snowfall success percentages of less than 50%, which means that more than half of the snowfall indicated by DPR is rainfall or mixed precipitation at the surface, which is somewhat expected since both DPR methods diagnose the precipitation phase a the clutter-free height (varying from 0.5 to 3 km), instead of at the surface. The temperature can increase significantly from the clutter-free height to the ground. Consequently, snow/ice particles detected in the atmosphere may transition to liquid raindrops. In contrast, CPR demonstrates a much larger snowfall success percentage of about 96%. Over land, CPR uses a conservative classification scheme where pixels are classified as snow if the maximum temperature in the column is less than 0°C, rain if the maximum temperature in the column is greater than 2°C, and an undetermined "mixed precipitation" in all other cases. This partially explains why CPR has a much better snowfall success percentage, compared with DPR. The more accurate temperature information from the ECMWF analysis data set used in the CPR precipitation phase determination may also contribute to its better performance (more details in the next section).

Additionally, the near-surface bin height from nadir-looking CPR is lower than that from scanning KuPR over the edge scan positions (Arulraj & Barros, 2021; Casella et al., 2017; Maahn et al., 2014), which also contributes to the better performance of CPR. Specifically, near-surface bin height from CPR is about 1.44 km above the ground level over land. In contrast, the near-surface bin height (i.e., clutter-free height) from KuPR can vary from ~0.5 km (nadir and near-nadir scans) to ~2.2 km (furthest off-nadir scans) above ground level. The varying near-surface bin height from KuPR has a significant impact on the snowfall success percentage computation. For example, the snowfall success percentage for DPR (Method 1) using nadir and near-nadir scans is 0.67, while it is only 0.32 when using the furthest off-nadir scans. This result is expected because the near surface bin height over the edge scan positions is about 1.5 km higher than that from the central scan positions. Therefore, the precipitation phase more likely changes from snowfall to rainfall over the edge scan positions because the temperature may increase more significantly from the near surface to the ground. Further analyses reveal that the snowfall success percentage for CPR also varies depending on the near surface bin height, although the degree of the variation is much smaller because CPR uses the temperature profile information in the phase determination process.

In the GPROF GMI retrievals, there are two precipitation rate variables (i.e., precipitation rate and snowfall rate). First, we select snowfall rates greater than 0 to compute the snowfall success percentage (hereafter referred to as GPROF M1). Second, we select pixels with a snowfall rate greater than 0 and the difference between snowfall rate and precipitation rate is less than 0.01 mm/hr (hereafter referred to as GPROF M2). Figure 1b shows that GPROF M2 has a much higher snow success percentage of 96% than that from GPROF M1 of 81%. The much larger success percentage from GPROF M2 is directly determined by how GPROF calculates the snowfall rates. Specifically, GPROF produces the snowfall rates by multiplying the precipitation rates with the snowfall probability determined by the 2-m wet bulb temperature (Sims & Liu, 2015). In the GPROF M2, we intentionally make these two variables close (i.e., the difference between snowfall rate and precipitation rate less than 0.01 mm/hr). As expected, choosing different threshold values (e.g., 0.1 or 0.001 mm/hr) will affect the snowfall success percentage values. In fact, the snowfall success percentage is 99% (90%) with the difference being 0.001 mm/hr (0.1 mm/hr). The unrealistic small precipitation value (0.001 mm/hr) is an artifact from the Bayesian retrieval algorithm employed by GPROF. It does not imply that GPROF can retrieve this light precipitation. It is worth mentioning that evaluating the snowfall success percentage from GPROF is actually equivalent to evaluating the snowfall probability computed from 2-m wet bulb temperature based on Sims and Liu (2015).

Table 1

Snowfall Success Percentage (%) From Six Methods (M1 to M6) and Four Reanalysis Datasets (MERRA2, JRA55, ERA5, and Global Forecast System [GFS])

Method	Variables	MERRA2	JRA55	ERA5	GFS
M1	T2m	94	93	95	95
M2	Tw	97	95	96	96
M3	Tw, Ts	97	94	96	95
M4	Tw, Γ	97	95	96	96
M5	Tw, Γ, Ts	97	95	96	96
M6	Tmax	91	90	93	91

Note. The first five approaches are from Sims and Liu (2015), which use 2-m air temperature (T2m), 2-m wet bulb temperature (Tw), 2-m wet bulb temperature (Tw) and surface temperature (Ts), 2-m wet bulb temperature (Tw) & temperature lapse rate (Γ), and 2-m wet bulb temperature (Tw) and temperature lapse rate (Γ) and surface temperature (Ts), respectively. The sixth approach is from Haynes et al. (2009), which requires the maximum temperature (Tmax) in the temperature profile being less than 0°C.

For MRMS, the snowfall success percentage is 84%, while the vast majority of the other 16% of data (i.e., surface indicates snowfall while MRMS judges them as rainfall) is misclassified as "cold stratiform" rainfall. Using independent precipitation type reports from the citizen-scientists, Chen, Hong, et al. (2016) also noticed that MRMS tends to misidentify snowfall as rainfall. The authors pointed out that the temperature threshold values or the uncertainties in the model-output temperature information may be responsible for the misidentification. Further analysis shows that the temperature information used in MRMS agrees very well with ground observation. Therefore, we conclude that the most likely reason for the misclassification (i.e., snowfall as rainfall) in MRMS is caused by the "colder" threshold values. Finally, it is worth mentioning that the rainfall success percentages from all four datasets are greater than 94% with no clear differences.

We also compute the snowfall success percentage by only using data over CONUS for CPR, DPR, and GPROF datasets. It is found that the major conclusion (i.e., the rank of the snowfall success percentage) remain unchanged, although the absolute values of the snowfall success percentage vary when only using data over CONUS.

3.3. Phase Discrimination Based on Different Methods and Different Reanalysis Datasets

In this section, we compute the snowfall and rainfall success percentages, using six phase discrimination methods and four global reanalysis datasets. The first five approaches are from Sims and Liu (2015), which uses 2-m air temperature, 2-m wet bulb temperature, 2-m wet bulb temperature and surface temperature, 2-m wet bulb temperature & temperature lapse rate, and 2-m wet bulb temperature and temperature lapse rate and surface temperature, respectively. It worth mentioning that the GPROF precipitation product utilizes the second method (i.e., 2-m wet bulb temperature) for the phase discrimination. These five methods, in order, are referred to as M1, M2, ..., M5. The sixth method (referred to as M6) is based on Haynes et al. (2009), which is the phase segregation method for CloudSat precipitation products. As mentioned previously, CPR judged a pixel as the snowfall pixel over land when the maximum temperature in the temperature profile less than 0°C (i.e., all temperature values are less than 0°C in the profile). We do not include the MRMS phase discrimination method since its threshold values are based on the temperature information from the continental-scale Rapid Refresh modeling system (Zhang et al., 2016). Additionally, we only utilize the ground weather reports in 2016 since there are about 2.2 million precipitating reports in total.

Table 1 shows the snowfall success percentage from these six different methods and four different datasets. The results show that using 2-m wet-bulb temperature (M2) generates slightly better phase discrimination performance than that using 2-m temperature (M1), regardless of the reanalysis datasets. This is because that the wet-bulb temperature is closer to the hydrometeors' temperature than the temperature itself since it combines the temperature and moisture information (Sims & Liu, 2015). Recent work by Heymsfield et al. (2021) showed that the ice-bulb temperature may provide further improvement for precipitation phase determination. More importantly, we do not notice additional discriminant capability being added when including more variables (e.g., lapse rate and surface temperature), by comparing M2 with M3–M5. The lack of a significant relationship between the lapse rate/surface temperature snow-rain separation found here is interesting, and something we will explore in future work. Further, the snowfall success percentage from M6 is slightly lower than those from M1 to M5, regardless of the reanalysis datasets. For example, the success percentage from M6 is 93% using ERA5 (Table 1, seventh row and fifth column), while it is 95% from M1 (Table 1, second row and fifth column). The reason why M6 generates slightly lower success percentage is because M6 requires the maximum temperature in the profile being less than 0°C, which is colder than the temperature threshold values used from M1 to M5 (e.g., 2-m air temperature being 1.6°C in M1).

Table 2
Rainfall Success Percentage (%) From Six Methods (M1–M6) and Four Reanalysis Datasets (MERRA2, JRA55, ERA5, and GFS)

Method	Variables	MERRA2	JRA55	ERA5	GFS
M1	T2m	89	95	96	96
M2	Tw	88	94	96	96
M3	Tw, Ts	88	95	97	96
M4	Tw, Γ	88	96	97	97
M5	Tw, Γ, Ts	88	97	97	97
M6	Tmax	92	91	93	92

Note. The first five approaches are from Sims and Liu (2015), which use 2m air temperature (T2m), 2-m wet bulb temperature (Tw), 2-m wet bulb temperature (Tw) and surface temperature (Ts), 2-m wet bulb temperature (Tw) and temperature lapse rate (Γ), and 2-m wet bulb temperature (Tw) and temperature lapse rate (Γ) and surface temperature (Ts), respectively. The sixth approach is from Haynes et al. (2009), which requires the maximum temperature (T_{max}) in the temperature profile greater than 2°C.

Consequently, a small percentage of observed snowfall is misclassified as mixed precipitation.

Table 2 shows that the rainfall success percentage based on MERRA2 is smaller than those from other reanalysis datasets for M1 to M5, because MERRA2's temperature profile from about 200m to the surface is noticeably colder than those from the other three datasets, shown in Figure 2a. The colder temperature from MERRA2 is further corroborated from the 2-m air temperature plot in Figure 2b, which demonstrates the 2-m air temperature difference between the reference and those from each reanalysis datasets. Clearly, 2-m air temperature from MERRA2 is about 1°C colder than those from other datasets. Statistical test shows that this 1°C colder temperature is statistically significant at the 95% confidence level. The colder MERRA2 temperature misidentifies the rainfall pixels as the snowfall pixels. The colder surface temperature from MERRA2 is also reported by Draper et al. (2018).

Rainfall success percentage from M6 is about 92% using all four model outputs (Table 2). This number (92%) is about 4% higher than those from M1 to M5 when using MERRA2, which is because M6 requires the maximum temperature in the temperature profile greater than 2°C. Even though MERRA2 is noticeably colder than observed, M6 achieves a slightly better success percentage by using a slightly higher temperature threshold value than those from M1 to M5 (e.g., 2-m air temperature being 1.6°C in M1). In contrast, M6 generates a slightly lower success percentage comparing with M1 to M5 when using the other model outputs, which again can be attributed to the 2 °C threshold value used in M6, which aims to exclude purely snowing scenes while allowing a larger unknown "mixed" category. Basically, M6 classifies a small percentage of rainfall as "mixed precipitation."

4. Conclusions

This study evaluated the precipitation phase discrimination performance in four widely used remote-sensing precipitation datasets using global ground precipitation phase reports. It is found that more than half of the snowfall indicated by GPM DPR is actually rainfall on the ground because DPR determines the precipitation phase of the hydrometers in the air, instead of on the ground. In contrast, CloudSat CPR shows a much better snowfall classification because it considers the temperature profile information. The lower
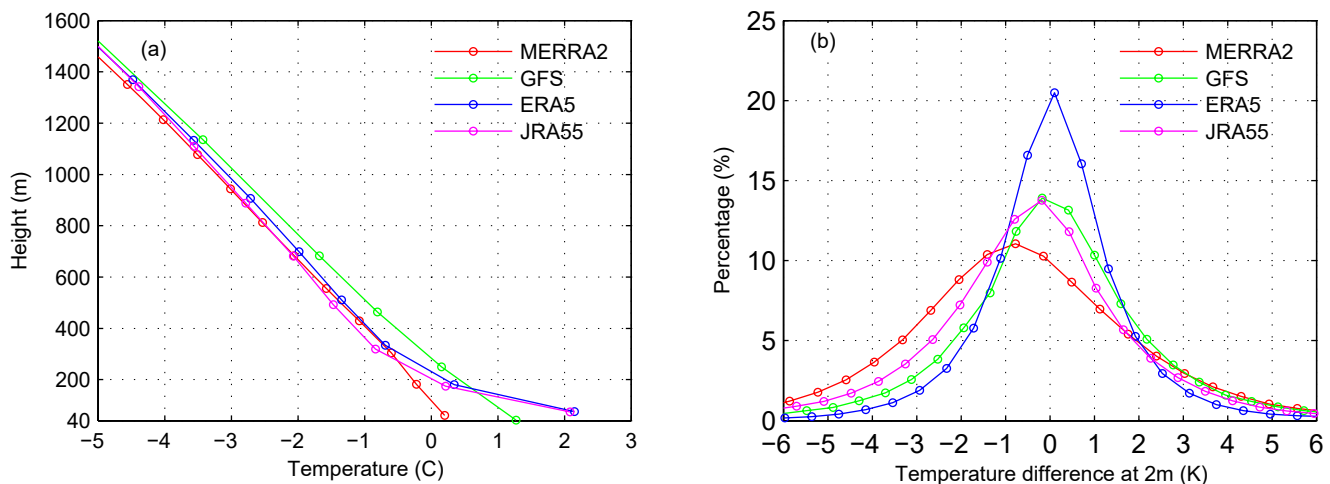

Figure 2. (a) Temperature profiles below 1,600 m from four reanalysis datasets, including MERRA2, Global Forecast System (GFS), ERA5, and JRA55. (b) The histograms of the 2-m air temperature differences between surface observations and each reanalysis data set. All data are from January 2016 to December 2016, regardless of the surface precipitation status.

near surface bin height (i.e., smaller blind zone) is another important factor responsible for the better performance of the nadir-looking CPR's snowfall classification, compared with the scanning DPR. In addition, the snowfall rate in GPM GMI precipitation product is computed as the product of the snowfall probability and the precipitation rate. Therefore, the snowfall discrimination accuracy is determined by the snowfall probability accuracy calculated by Sims and Liu's method from the 2-m wet bulb temperature (Sims & Liu, 2015). Further, MRMS misclassifies some snowfall as cold stratiform rainfall likely due to the "colder" threshold values used in the MRMS phase determination scheme. An analysis of four different reanalysis datasets shows that the MERRA2 temperature profile is noticeably colder close to the surface than those from the other three global reanalysis datasets, which results in the rainfall being misclassified as snowfall. Finally, it is found that using the 2-m wet-bulb temperature is adequate for snowfall determination regardless of which reanalysis datasets is chosen.

The comparison analysis reveals that different remote sensing datasets use very different approaches for precipitation phase determination, leading to significantly different accuracy performance. Understanding this feature is critically important for correctly using these datasets and future satellite precipitation data set evaluations. This study also highlights the challenges and obstacles in precipitation phase discrimination, an issue that must be addressed to generate more accurate global snowfall climatology.

Data Availability Statement

GPM data were downloaded from NASA PPS at https://storm.pps.eosdis.nasa.gov/storm/. CloudSat data were download from CloudSat data processing center at https://www.cloudsat.cira.colostate.edu/data-products. MRMS precipitation data were downloaded from National Centers for Environmental Prediction (NCEP) (http://mrms.ncep.noaa.gov/data/). Surface precipitation phase reports were downloaded from National Center for Environmental Information (NCEI) (https://www.ncdc.noaa.gov/isd). MERRA2 data were downloaded from NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC) (https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/). JRA55 data were downloaded from National Center for Atmospheric Research (United States) (http://rda.ucar.edu/datasets/ds628.0/), which is a mirror site of Data Integration & Analysis System (Japan) (http://search.diasjp.net/en/dataset/JRA55). ERA5 data were downloaded from European Centre for Medium-Range Weather Forecasts (https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5). GFS data were downloaded from National Centers for Environmental Information (https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs).

References

Adhikari, A., Liu, C., & Kulie, M. S. (2018). Global distribution of snow precipitation features and their properties from 3 years of GPM observations. *Journal of Climate*, *31*(10), 3731–3754. https://doi.org/10.1175/jcli-d-17-0012.1

Arulraj, M., & Barros, A. (2021). Automatic detection and classification of low-level orographic precipitation processes from space-borne radars using machine learning. *Remote Sensing of Environment*, *257*, 112355. https://doi.org/10.1016/j.rse.2021.112355

Auer, A. H., Jr (1974). The rain versus snow threshold temperatures. *Weatherwise*, *27*(2), 67. https://doi.org/10.1080/00431672.1974.9931684

Behrangi, A., Yin, X., Rajagopal, S., Stampoulis, D., & Ye, H. (2018). On distinguishing snowfall from rainfall using near-surface atmospheric information: Comparative analysis, uncertainties and hydrologic importance. *Quarterly Journal of the Royal Meteorological Society*, *144*, 89–102. https://doi.org/10.1002/qj.3240

Bourgouin, P. (2000). A method to determine precipitation types. *Weather and Forecasting*, *15*(5), 583–592. https://doi.org/10.1175/1520-0434(2000)015<0583:amtdpt>2.0.co;2

Casella, D., Panegrossi, G., Sanò, P., Marra, A. C., Dietrich, S., Johnson, B. T., & Kulie, M. S. (2017). Evaluation of the GPM-DPR snowfall detection capability: Comparison with CloudSat-CPR. *Atmospheric Research*, *197*, 64–75. https://doi.org/10.1016/j.atmosres.2017.06.018

Chen, S., Gourley, J. J., Hong, Y., Cao, Q., Carr, N., Kirstetter, P.-E., et al. (2016). Using citizen science reports to evaluate estimates of surface precipitation type. *Bulletin of the American Meteorological Society*, *97*(2), 187–193. https://doi.org/10.1175/bams-d-13-00247.1

Chen, S., Hong, Y., Kulie, M., Behrangi, A., Stepanian, P. M., Cao, Q., et al. (2016). Comparison of snowfall estimates from the NASA CloudSat cloud profiling radar and NOAA/NSSL multi-radar multi-sensor system. *Journal of Hydrology*, *541*, 862–872. https://doi.org/10.1016/j.jhydrol.2016.07.047

Dai, A. (2001). Global precipitation and thunderstorm frequencies. Part I: Seasonal and interannual variations. *Journal of Climate*, *14*(6), 1092–1111. https://doi.org/10.1175/1520-0442(2001)014<1092:gpatfp>2.0.co;2

Dai, A. (2008). Temperature and pressure dependence of the rain-snow phase transition over land and ocean. *Geophysical Research Letters*, *35*(12). https://doi.org/10.1029/2008gl033295

Ding, B., Yang, K., Qin, J., Wang, L., Chen, Y., & He, X. (2014). The dependence of precipitation types on surface elevation and meteorological conditions and its parameterization. *Journal of Hydrology*, *513*, 154–163. https://doi.org/10.1016/j.jhydrol.2014.03.038

Draper, C. S., Reichle, R. H., & Koster, R. D. (2018). Assessment of MERRA-2 land surface energy flux estimates. *Journal of Climate*, *31*(2), 671–691. https://doi.org/10.1175/jcli-d-17-0121.1

Durden, S. L., Tanelli, S., Epp, L., Jamnejad, V., Perez, R., Prata, A., et al. (2011). A cloud and precipitation radar system concept for the ACE Mission, in: *IEEE international geoscience and remote sensing society (IGRSS)* (p. 4). Vancover, Canada.

Ebtehaj, A., & Kummerow, C. (2017). Microwave retrievals of terrestrial precipitation over snow-covered surfaces: A lesson from the GPM satellite. *Geophysical Research Letters*, *44*(12), 6154–6162. https://doi.org/10.1002/2017gl073451

Ferraro, R., Beauchamp, J., Cecil, D., & Heymsfield, G. (2015). A prototype hail detection algorithm and hail climatology developed with the Advanced Microwave Sounding Unit (AMSU). *Atmospheric Research*, *163*, 24–35. https://doi.org/10.1016/j.atmosres.2014.08.010

Gelaro, R., McCarty, W., Sua´rez, M. J., Todling, R., Molod, A., & Takacs, L., et al. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/jcli-d-16-0758.1

Gergel, D. R., Nijssen, B., Abatzoglou, J. T., Lettenmaier, D. P., & Stumbaugh, M. R. (2017). Effects of climate change on snowpack and fire potential in the western USA. *Climatic Change*, *141*(2), 287–299. https://doi.org/10.1007/s10584-017-1899-y

Hamada, A., & Takayabu, Y. N. (2016). Improvements in detection of light precipitation with the global precipitation measurement dual-frequency precipitation radar (GPM DPR). *Journal of Atmospheric and Oceanic Technology*, *33*(4), 653–667. https://doi.org/10.1175/jtech-d-15-0097.1

Haynes, J. M., L'Ecuyer, T. S., Stephens, G. L., Miller, S. D., Mitrescu, C., Wood, N. B., & Tanelli, S. (2009). Rainfall retrieval over the ocean with spaceborne W-band radar. *Journal of Geophysical Research*, *114*(D8). https://doi.org/10.1029/2008jd009973

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Hora´nyi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Heymsfield, A. J., Bansemer, A., Theis, A., & Schmitt, C. (2021). Survival of snow in the melting layer: Relative humidity influence. *Journal of the Atmospheric Sciences*, *78*(6), 1823–1845. https://doi.org/10.1175/jas-d-20-0353.1

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., et al. (2014). The global precipitation measurement mission. *Bulletin of the American Meteorological Society*, *95*(5), 701–722. https://doi.org/10.1175/bams-d-13-00164.1

Iguchi, T. (2020). Dual-frequency precipitation radar (DPR) on the global precipitation measurement (GPM) missions core observatory. In *Satellite precipitation measurement* (pp. 183–192). Springer. https://doi.org/10.1007/978-3-030-24568-9_11

Jennings, S. S., Winchell, B., Livneh, P., & Molotch, N. (2018). Spatial variation of the rain–snow temperature threshold across the Northern Hemisphere. *Nature Communications*, *9*(1), 1–9. https://doi.org/10.1038/s41467-018-03629-7

Kanamitsu, I., Alpert, J. C., Campana, K. A., Caplan, P. M., Deaven, D. G., Iredell, M., et al. (1991). Recent changes implemented into the global forecast system at NMC. *Weather and Forecasting*, *6*(3), 425–435. https://doi.org/10.1175/1520-0434(1991)006<0425:rciitg>2.0.co;2

Kidd, C., Matsui, T., Chern, J., Mohr, K., Kummerow, C., & Randel, D. (2016). Global precipitation estimates from cross-track passive microwave observations using a physically based retrieval scheme. *Journal of Hydrometeorology*, *17*(1), 383–400. https://doi.org/10.1175/jhm-d-15-0051.1

Kirstetter, P., Hong, Y., Gourley, J. J., Chen, S., Flamig, Z., Zhang, J., et al. (2012). Toward a framework for systematic error modeling of spaceborne precipitation radar with NOAA/NSSL ground radar–based National Mosaic QPE. *Journal of Hydrometeorology*, *13*(4), 1285–1300. https://doi.org/10.1175/jhm-d-11-0139.1

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., et al. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II.*, *93*(1), 5–48. https://doi.org/10.2151/jmsj.2015-001

Kongoli, C., Pellegrino, P., Ferraro, R. R., Grody, N. C., & Meng, H. (2003). A new snowfall detection algorithm over land using measurements from the Advanced Microwave Sounding Unit (AMSU). *Geophysical Research Letters*, *30*(14). https://doi.org/10.1029/2003gl017177

Kulie, M. S., & Bennartz, R. (2009). Utilizing spaceborne radars to retrieve dry snowfall. *J. Appl. Meteor. Climatol.*, *48*(12), 2564–2580. https://doi.org/10.1175/2009jamc2193.1

Kulie, M. S., Milani, L., Wood, N. B., Tushaus, S. A., Bennartz, R., & LEcuyer, T. S. (2016). A shallow cumuliform snowfall census using spaceborne radar. *Journal of Hydrometeorology*, *17*(4), 1261–1279. https://doi.org/10.1175/jhm-d-15-0123.1

Kulie, M. S., Pettersen, C., Merrelli, A. J., Wagner, T. J., Wood, N. B., Dutter, M., et al. (2021). Snowfall in the northern great lakes: Lessons learned from a multi-sensor observatory. *Bulletin of the American Meteorological Society*, *102*, 1–61. https://doi.org/10.1175/bams-d-19-0128.1

Kummerow, C. D., Randel, D. L., Kulie, M., Wang, N.-Y., Ferraro, R., Joseph Mun- chak, S., & Petkovic, V. (2015). The evolution of the Goddard profiling algorithm to a fully parametric scheme. *Journal of Atmospheric and Oceanic Technology*, *32*(12), 2265–2280. https://doi.org/10.1175/jtech-d-15-0039.1

Le, M., Chandrasekar, V., & Biswas, S. (2017). An algorithm to identify surface snowfall from GPM DPR observations. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(7), 4059–4071. https://doi.org/10.1109/tgrs.2017.2687420

Liu, G. (2008). Deriving snow cloud characteristics from CloudSat observations. *Journal of Geophysical Research*, *113*(D8). https://doi.org/10.1029/2007jd009766

Maahn, M., Burgard, C., Crewell, S., Gorodetskaya, I. V., Kneifel, S., Lhermitte, S., et al. (2014). How does the spaceborne radar blind zone affect derived surface snowfall statistics in polar regions? *Journal of Geophysical Research*, *119*(24), 13–604. https://doi.org/10.1002/2014jd022079

Matsuo, T., Sasyo, Y., & Sato, Y. (1981). Relationship between types of precipitation on the ground and surface meteorological elements. *Journal of the Meteorological Society of Japan*, *59*(4), 462–476. https://doi.org/10.2151/jmsj1965.59.4_462

Meng, H., Dong, J., Ferraro, R., Yan, B., Zhao, L., Kongoli, C., et al. (2017). A 1DVAR-based snowfall rate retrieval algorithm for passive microwave radiometers. *Journal of Geophysical Research*, *122*, 6520–6540. https://doi.org/10.1002/2016jd026325

Milani, L., Kulie, M. S., Casella, D., Kirstetter, P. E., Panegrossi, G., Petkovic, V., et al. (2021). Extreme lake-effect snow from a GPM microwave imager perspective: Observational analysis and precipitation retrieval evaluation. *Journal of Atmospheric and Oceanic Technology*, *38*(2), 293–311. https://doi.org/10.1175/jtech-d-20-0064.1

Milani, L., Kulie, S., Casella, D., Dietrich, S., L'Ecuyer, T. S., Panegrossi, G., et al. (2018). CloudSat snowfall estimates over Antarctica and the Southern Ocean: An assessment of independent retrieval methodologies and multi-year snowfall analysis. *Atmospheric Research*, *213*, 121–135. https://doi.org/10.1016/j.atmosres.2018.05.015

Mroz, K., Battaglia, A., Lang, T. J., Cecil, D. J., Tanelli, S., & Tridon, F. (2017). Hail-detection algorithm for the GPM Core Observatory satellite sensors. *Journal of Applied Meteorology and Climatology*, *56*(7), 1939–1957. https://doi.org/10.1175/jamc-d-16-0368.1

OGorman, P. A. (2014). Contrasting responses of mean and extreme snowfall to climate change. *Nature*, *512*(7515), 416–418.

Panegrossi, G., Rysman, J.-F., Casella, D., Marra, A. C., San`o, P., & Kulie, M. S. (2017). CloudSat-based assessment of GPM Microwave Imager snowfall observation capabilities. *Remote Sensing*, *9*(12), 1263. https://doi.org/10.3390/rs9121263

Pettersen, C., Bliven, L. F., Kulie, M. S., Wood, N. B., Shates, J. A., Anderson, J., et al. (2021). The Precipitation Imaging Package: Phase Partitioning Capabilities. *Remote Sensing*, *13*(11), 2183. https://doi.org/10.3390/rs13112183

Pettersen, C., Kulie, M., Bliven, L., Merrelli, A., Petersen, W., Wagner, T., et al. (2020). A composite analysis of snowfall modes from four winter seasons in Marquette, Michigan. *Journal of Applied Meteorology and Climatology*, *59*(1), 103–124. https://doi.org/10.1175/jamc-d-19-0099.1

Petty, G. W. (1995). Frequencies and characteristics of global oceanic precipitation from shipboard present-weather reports. *Bulletin of the American Meteorological Society*, *76*(9), 1593–1616. https://doi.org/10.1175/1520-0477(1995)076<1593:facogo>2.0.co;2

Ryzhkov, V., & Zrnic, D. S. (1998). Discrimination between rain and snow with a polarimetric radar. *Journal of Applied Meteorology*, *37*, 1228–1240. https://doi.org/10.1175/1520-0450(1998)037<1228:dbrasw>2.0.co;2

Sheffield, J., Ferguson, C. R., Troy, T. J., Wood, E. F., & McCabe, M. F. (2009). Closing the terrestrial water budget from satellite remote sensing. *Geophysical Research Letters*, *36*(7). https://doi.org/10.1029/2009gl037338

Sims, M., & Liu, G. (2015). A parameterization of the probability of snow–rain transition. *Journal of Hydrometeorology*, *16*(4), 1466–1477. https://doi.org/10.1175/jhm-d-14-0211.1

Skofronick-Jackson, G., Kulie, M., Milani, L., Munchak, S. J., Wood, N. B., & Levizzani, V. (2019). Satellite estimation of falling snow: A global precipitation measurement (GPM) core observatory perspective. *J. Appl. Meteor. Climatol.*, *58*(7), 1429–1448. https://doi.org/10.1175/jamc-d-18-0124.1

Skofronick-Jackson, G., Petersen, W. A., Berg, W., Kidd, C., Stocker, E. F., Kirschbaum, D. B., et al. (2017). The Global Precipitation Measurement (GPM) mission for science and society. *Bulletin of the American Meteorological Society*, *98*(8), 1679–1695. https://doi.org/10.1175/bams-d-15-00306.1

Smith, A., Lott, N., & Vose, R. (2011). The integrated surface database: Recent developments and partnerships. *Bulletin of the American Meteorological Society*, *92*(6), 704–708. https://doi.org/10.1175/2011bams3015.1

Stephens, G. L., & Haynes, J. M. (2007). Near global observations of the warm rain coalescence process. *Geophysical Research Letters*, *34*, L20805. https://doi.org/10.1029/2007GL030259

Stephens, G. L., Vane, D. G., Boain, R. J., Mace, G. G., Sassen, K., Wang, Z., et al. (2002). The CloudSat mission and the A-Train: A new dimension of space-based observations of clouds and precipitation. *Bulletin of the American Meteorological Society*, *83*(12), 1771–1790. https://doi.org/10.1175/bams-83-12-1771

United States Army Corps of Engineers. (1956). *Summary report of the snow investigation Snow hydrology* (p. 437). North Pacific Division report.

Wagner, A. J. (1957). Mean temperature from 1000 mb to 500 mb as a predictor of precipitation type. *Bulletin of the American Meteorological Society*, *38*(10), 584–590. https://doi.org/10.1175/1520-0477-38.10.584

Wang, Y., Liu, G., Seo, E.-K., & Fu, Y. (2013). Liquid water in snowing clouds: Im- plications for satellite remote sensing of snowfall. *Atmospheric Research*, *131*, 60–72. https://doi.org/10.1016/j.atmosres.2012.06.008

Wang, Y., You, Y., & Kulie, M. (2018). Global virga precipitation distribution derived from three spaceborne radars and its contribution to the false radiometer precipitation detection. *Geophysical Research Letters*, *45*(9), 4446–4455. https://doi.org/10.1029/2018gl077891

You, Y., Wang, N.-Y., & Ferraro, R. (2015). A prototype precipitation retrieval algorithm over land using passive microwave observations stratified by sur- face condition and precipitation vertical structure. *Journal of Geophysical Research*, *120*, 5295–5315. https://doi.org/10.1002/2014jd022534

You, Y., Wang, N.-Y., Ferraro, R., & Meyers, P. (2016). A prototype precipitation retrieval algorithm over land for ATMS. *Journal of Hydrometeorology*, *17*(5), 1601–1621. https://doi.org/10.1175/jhm-d-15-0163.1

You, Y., Wang, N.-Y., Ferraro, R., & Rudlosky, S. (2017). Quantifying the snowfall detection performance of the GPM microwave imager channels over land. *Journal of Hydrometeorology*, *18*(3), 729–751. https://doi.org/10.1175/jhm-d-16-0190.1

Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., & Tang, L., & . others. (2016). Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bulletin of the American Meteorological Society*, *97*(4), 621–638. https://doi.org/10.1175/bams-d-14-00174.1