THESIS

SATELLITE OBSERVATIONS OF OCEANIC HIGH-LATITUDE DRIZZLE USING A COMBINED RADAR-RADIOMETER RETRIEVAL

Submitted by

Spencer R. Jones

Department of Atmospheric Science

In partial fulfillment of the requirements

For the degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2024

Master's Committee:

Advisor: Christian Kummerow

Christine Chiu
Venkatachalam Chandrasekaran
Christopher Grassotti

ABSTRACT

SATELLITE OBSERVATIONS OF OCEANIC HIGH LATITUDE DRIZZLE USING A COMBINED RADAR-RADIOMETER RETRIEVAL

The high latitude oceans are problematic for satellite estimations of precipitation due to the high frequency of occurrence of light drizzle and snowfall. Passive microwave radiometric observations are sensitive to integrated cloud water path and provide good sampling for robust statistics but have little skill in distinguishing precipitation onset from cloud water and cloud ice due to a lack of sensitivity to drop sizes when they are small. Spaceborne precipitation radars to date have lacked sensitivity to drizzle, and cloud radars have suffered from both the uncertainties inherent in Z-R relations and poor sampling due to nadir-only scans. This study combines coincident active and passive microwave observations from CloudSat's Cloud Profiling Radar (CPR) and the Advanced Scanning Microwave Radiometer (AMSR2) to resolve cloud and hydrometeor distribution parameters and to force consistency between the two independent sets of coincident observations. Consistency between the radar and radiometer is found by using an optimal estimation (OE) retrieval algorithm, a physics-based technique that simultaneously resolves the most likely atmospheric state given both radar and radiometer observations as well as *a priori* information.

The OE algorithm uncertainties are estimated using a method that attempts to emulate the departure in observation space of retrieved states from the unknown true state. The focus on observational uncertainties and the accuracy obtained by using nondiagonal observational error covariance matrices allows the algorithm both to resolve states that are radiatively consistent and

to reduce the level of nonuniqueness found in dealing with passive observations alone. The result is an estimation of drizzle frequency and intensity that are consistent with both the CPR and AMSR2 observations for the high latitude oceans. We find that zonal means of retrieved high-latitude drizzle below 0.25 mm hr^{-1} from these combined observations (0.263 mm day^{-1}) falls slightly above those of CloudSat estimates (0.244 mm day^{-1}), provided by the 2C-RAIN-PROFILE and 2C-SNOW-PROFILE products (Lebsock 2018; Wood and L'Ecuyer 2018), and far below that of radiometer-only estimates (0.920 mm day^{-1}) provided by GPROF (Kummerow et al. 2015).

ACKNOWLEDGEMENTS

I would like to sincerely thank my advisor, Dr. Christian Kummerow, for his invaluable guidance and personal involvement throughout this project, as well as his kind encouragement, and for always expecting and allowing me to do my very best. I would also like to thank my master's committee members, with a special thank-you to Dr. Christine Chiu, whose expertise and openness I have benefitted from many times. Thank-you's are also due the entire Kummerow research group—especially Eric Goldenstern, whose advice and willingness to be both a sounding board for research ideas and golfing partner I have enjoyed; Ryan Gonzales for being a great friend and colleague; Rick Schulte, who laid much of the foundation for this problem, and who has graciously provided me with source code and advice when needed; Paula Brown, who provided me with data and much needed advice at times; and Simon Pfreundschuh, whose expertise on optimal estimation and the intricacies of cloud radars I have found to be helpful on several occasions.

I would also like to thank the CSU Department of Atmospheric Science for fostering a community of mutual support and friendship as well as an atmosphere that is simply electric with scientific curiosity.

Last, but certainly not least, I would like to thank my wonderful wife, Steffani, for her unrelenting love, encouragement, and support throughout this endeavor, and to whom I owe gratitude for more reasons than space here allows.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Clouds and precipitation remain some of the most highly variable and uncertain parameters in global climate trend forecasting (Ban et al. 2014). The advancement of satellite atmospheric remote sensing techniques over the last half century have contributed much to the current understanding of changing precipitation trends and extremes on the global scale (Adler et al. 2017). Better understanding of the distribution and spatial patterns of precipitation has informed every field in Earth science, including the creation and improvement of climate and weather forecasting models, the Earth energy budget, the hydrologic cycle, and climate monitoring and predictions. With much ongoing effort to improve the synergy of climate forecasting by reducing uncertainties and to understand and reduce negative human impacts on the Earth system, the observation of global precipitation and clouds remains a relevant frontier of Earth system science. The current understanding of global precipitation trends is summarized by the "dry gets drier and wet gets wetter" hypothesis (Liu et al. 2012). However, much progress remains to be made to reduce uncertainties in order to form a consistent precipitation climate data record from which these signals can be teased.

Over oceans, where most precipitation occurs and where in-situ observations are sparse, satellite retrieval algorithm validation is especially difficult. Despite more global coverage than ever before, and despite continual improvements in retrieval algorithms, uncertainties remain in global satellite precipitation estimations on the order of tenths of millimeters per day (Petković et al. 2023). One of the main contributors to these uncertainties is the disagreement between operational satellite precipitation climate products on the retrieval of light rain (Schulte and Kummerow 2022).

For the global scale, the best effort so far in satellite observations of precipitation has been the Global Precipitation Measurement Mission (GPM) (Hou et al. 2014), with over a decade of continuous passive microwave (PMW) and active radar observations. While PMW imagers provide large spatial coverage and high temporal resolution in the middle and high-latitude oceans appropriate for robust statistics, the skillful separation of nonprecipitating cloud water and drizzle is difficult with PMW observations alone due to the nonuniqueness of scattering and absorption signals in the observed brightness temperature (Tb) vector (Berg et al. 2006; Stephens and Kummerow 2007). Combining passive and active observations, however, has been shown to effectively constrain the separation of cloud and precipitation water, such as in the GPM Combined Algorithm (hereafter GPM-CMB; Grecu et al., 2016), which optimally combines Dual-frequency Precipitation Radar (DPR) and GPM Microwave Imager (GMI) observations to retrieve atmospheric profiles of hydrometeor and cloud water content. Unfortunately, the DPR's low-reflectivity detection threshold of ~12 dBZ, corresponding to roughly 0.25 mm hr^{-1} at the surface, means that the lowest precipitation rates within the drizzle regime are missed (Schulte et al. 2022). The GPM PMW scheme, the Goddard Profiling Algorithm (GPROF; Kummerow et al. 2015), is a Bayesian retrieval scheme that estimates the most likely precipitation using an *a priori* database of observed radar profiles. While skillful over a range of rainfall rates, the source of the *a priori* information which forms the database for precipitation retrievals is the DPR and thus it also misses the light precipitation observations.

Climatologically, the problem of drizzle is particularly relevant in the high latitude oceans, where light precipitation is meteorologically prevalent and where few sensors capable of detecting low precipitation rates have sampled (Behrangi et al. 2012). This problem is illustrated in Figure 1, which shows high-latitude zonal mean precipitation climatologies, as estimated by

several different operational satellite precipitation products and one reanalysis product. The first three products shown are three different operational precipitation retrieval algorithms from CloudSat, a spaceborne W-band cloud radar. The other estimates shown in Figure 1 come from the GPM-CMB algorithm, GPROF, and from the ECMWF Reanalysis 5 (ERA5; Hersbach et al. 2020). CloudSat's 2C-PRECIP-COLUMN algorithm (Haynes 2018) estimates the attenuation of the radar beam by hydrometeors to estimate surface precipitation. The 2C-RAIN-PROFILE (Lebsock 2018) and 2C-SNOW-PROFILE (Wood and L'Ecuyer 2018) algorithms retrieve precipitation by estimating the volumetric precipitation water and ice content and their particle sizes from the observed radar reflectivity. GPM-CMB uses observations from the GPM Dual-frequency Precipitation Radar and coincident passive observations from the GPM Microwave Imager (GMI) together for the best estimates. GPROF estimates come from the aforementioned Bayesian method of retrievals on PMW observations from GMI. Finally, in contrast to the methods mentioned above, precipitation in ERA5 is created by the ECMWF model, which assimilates the raw radiance measurements from PMW sensors instead of creating precipitation estimates directly from the observations (Lavers et al. 2022). From each of these techniques we see not only a disagreement in overall zonal means of precipitation, but also a stark separation of drizzle accumulations as well. This hints at fundamental differences from these products in the frequency distributions as well as the intensity estimations of light drizzle, which is defined for this study as at or below the DPR detection threshold of 0.25 mm hr^{-1}.

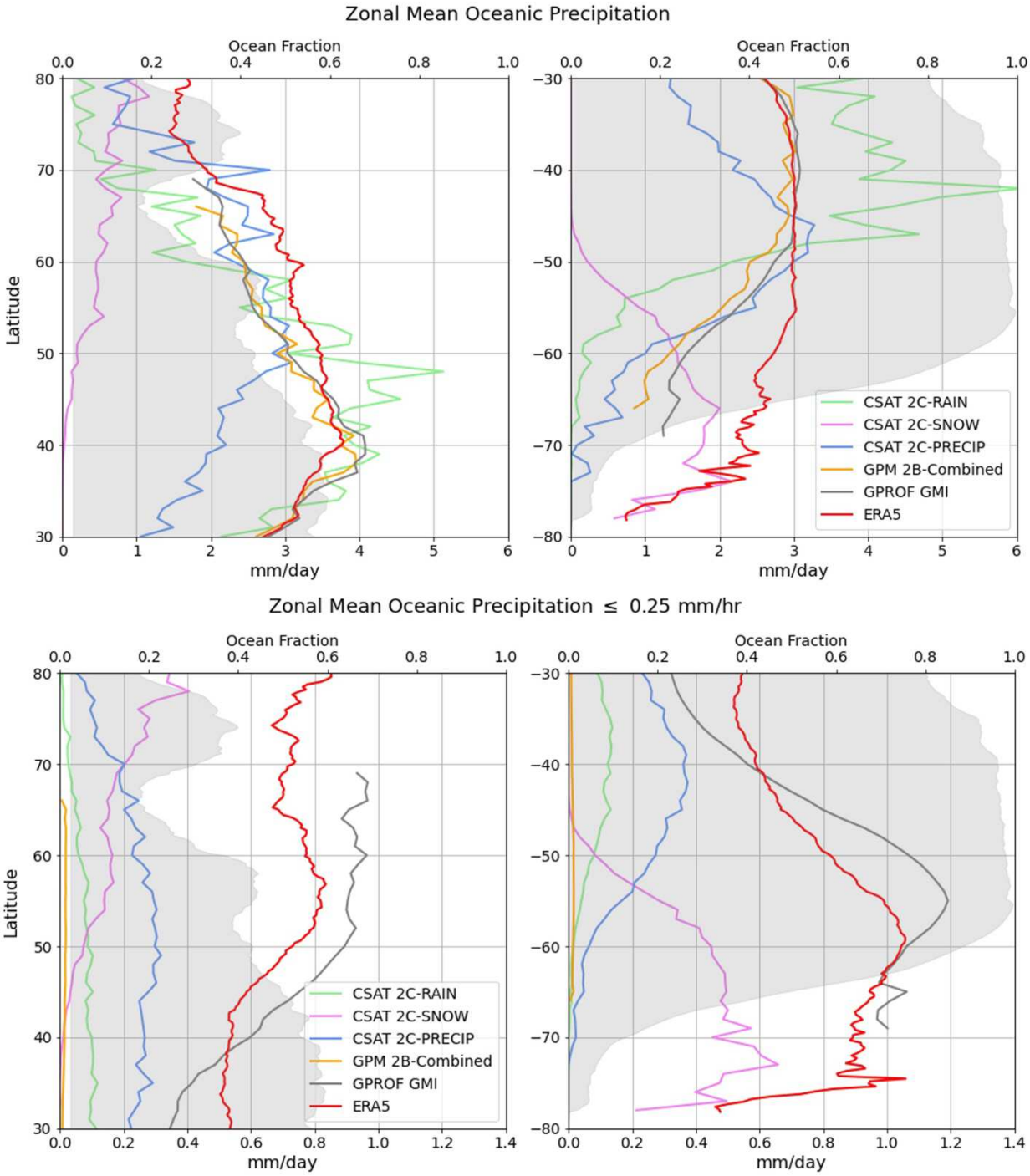

Figure 1: Top two panels: zonal mean precipitation for northern hemisphere (left) and southern hemisphere (right) oceans from ECMWF Reanalysis 5 (ERA5), The GPM Level 2B Combined Algorithm, Goddard Profiling Algorithm (GPROF), and three operational CloudSat retrieval algorithms: 2C-RAIN-PROFILE, 2C-SNOW-PROFILE, and 2C-PRECIP-COLUMN. Bottom two panels: zonal mean estimations of drizzle by accumulating only precipitation rates less than or equal to 0.25 mm hr^{-1}. All non-gridded products are gridded into 1° boxes to reduce resolution dependency. Gray shading indicates ocean fraction as a function of latitude (top axis).

While the overall mean precipitation estimates are far more sensitive to higher intensities, fundamental disagreements in the accumulations for the high latitude oceans can also be attributed to light precipitation and its high frequency of occurrence there. It is well-known that current sensors miss an unknown but significant portion of the total rain volume in the high latitudes due to the low intensity rates which dominate the region and the inadequacy of PMW, radar, and IR estimation techniques to capture them (Behrangi et al. 2012, 2014). The Ocean Rainfall And Ice-phase precipitation measurement Network (OceanRAIN; Klepp et al., 2018) field campaign, using shipborne disdrometers, found that in all cases of precipitation occurrence poleward of 35° latitude, intensity rates below 0.5 mm hr^{-1} formed the greatest fraction (54.7-93.8%) of observations for both rain and snow. Those findings were consistent with previously constructed ocean precipitation climatology maps using present-weather ship reports by Petty (1995). Even though the drizzle thresholds as defined in those studies are higher than the maximum drizzle rate considered in this study (0.25 mm hr^{-1}), the importance of light drizzle to both the frequency of occurrence and accumulation in these regions from a climate perspective is clear, and the poor constraints of high-latitude oceanic precipitation by current methods can be at least partially attributed to these observational limitations.

For this study, since the observation of precipitation rates higher than 0.25 mm hr^{-1} are considered well-constrained by the GPM-CMB algorithm, we combine radar and radiometer observations from CloudSat's Cloud Profiling Radar (CPR) and the Advanced Microwave Scanning Radiometer 2 (AMSR2), to constrain the problem of drizzle over the high latitude oceans. Using this technique, hereafter referred to as the AMSR2-CloudSat combined retrieval algorithm (A2CS), observations of precipitation rates below the DPR detection threshold are possible, and we are able to retrieve both hydrometeor content profiles and particle size

distribution (PSD) characteristics for these lightly precipitating scenes. These profiles could be beneficial in expanding the current operational GPROF *a priori* database to include directly observed drizzle if physical consistency were reached between the CPR and a PMW radiometer similar to the GPM-CMB for higher precipitation rates.

Previous studies have shown that the CPR's sensitivity to clouds, with noise floors typically around -26 dBZ, makes it particularly suitable for the observation of the onset of precipitation. Light drizzle and snow intensities from CPR observations have been estimated by a number of authors (Lebsock and L'Ecuyer 2011; L'Ecuyer and Stephens 2002; Liu 2008a; Stephens 2002). Behrangi et al. (2014) merged CloudSat rainfall estimates from the 2C-RAIN-PROFILE algorithm (Lebsock and L'Ecuyer 2011) with those from the Tropical Rainfall Measuring Mission Precipitation Radar (TRMM-PR) to fill out the zonal distribution of rain below the TRMM-PR detection threshold and to extend the distribution into the high latitude oceans. So far, however, no effort has been made to force consistency between CPR observations and PMW observations to constrain the estimation of light drizzle with sufficient sampling for robust statistics as has been done by Grecu et al. (2016) for higher precipitation rates. Since radar algorithms typically rely on assumed particle size distribution (PSD) models and ancillary data in order to constrain the solution, a retrieved scene from radar observations only has no guarantee of producing simulated microwave radiances consistent with any radiometer measurements.

Primarily using path-integrated attenuation (PIA) of the radar beam to detect precipitation occurrence has been shown to be effective in Haynes and Stephens (2007). The detection of precipitation-sized particles using reflectivity thresholds with PIA as a constraint is the basis of the latest rainfall algorithm from CloudSat (Lebsock and L'Ecuyer 2011). However, as Figure 1 illustrates, there remains disagreement even between the current operational CloudSat products.

We therefore assume that the CloudSat radar algorithms are good for the detection of precipitation onset within the cloud, but the intensity estimate is not as well-constrained. By optimally combining both passive and active observations from AMSR2 and CloudSat, we are able to directly retrieve the PSD parameters and vertical distribution of water content for each scene that are consistent with both the radiometer and radar observations.

CHAPTER 2: DESCRIPTION OF DATA AND TECHNIQUES

2.1 Data used

2.1.1 Observational data

The CloudSat radar observations used in this study are the Level 2B-GEOPROF CPR volumetric reflectivity estimations (Marchand et al. 2008). This algorithm output is used due to its known accuracy and skillful separation of reflectivity factors most likely caused by cloud particles from noise. These reflectivity profiles are also used because they are uncorrected for gaseous and hydrometeor attenuation, which are significant at the 94.0 GHz operating frequency of the radar (Haynes et al. 2007), though the 2B-GEOPROF algorithm does give an estimate for the correction factor using the collocated water vapor profile from ECMWF. Using the uncorrected radar profile allows the radiative transfer model of the A2CS algorithm to estimate the attenuation for direct comparison to the observed reflectivity profile.

There are three current operational precipitation algorithms for CloudSat: 2C-PRECIP-COLUMN (Haynes 2018), 2C-RAIN-PROFILE (Lebsock 2018), and 2C-SNOW-PROFILE (Wood and L'Ecuyer 2018). Both 2C-RAIN-PROFILE and 2C-SNOW-PROFILE use output data from 2C-PRECIP-COLUMN primarily for precipitation detection and phase determination. 2C-RAIN-PROFILE and 2C-SNOW-PROFILE are used together in the A2CS algorithm for precipitation detection as discussed in Section 2.3.4. 2C-RAIN-PROFILE estimates the effective, or unattenuated, volumetric reflectivity factor of hydrometeors, as well as the liquid and ice contents for each sampled volume at the native 240-m vertical resolution. 2C-RAIN-PROFILE is considered the premier rainfall algorithm for CloudSat, and 2C-SNOW-PROFILE is the analogous snowfall product. Together, they form the basis for the separation of cloud-only and

precipitating scenes in the A2CS algorithm, and therefore the probability of precipitation for A2CS is directly related to the overall frequency of occurrence as observed by the CPR.

The passive microwave observations used are the Level 1C brightness temperatures (Berg et al. 2016) from AMSR2, flying aboard the Global Change Observation Mission-Water satellite (GCOM-W1). The conically-scanning radiometer measures naturally emitted microwave radiation at 7 frequencies and at both horizontal and vertical polarizations with a nominal incidence angle of 55 degrees. These observed brightness temperatures are cross-calibrated for maximum consistency with other PMW sensors across the GPM constellation. An additional small algorithm-specific calibration adjustment was necessary to reduce biases initially present in the A2CS algorithm and is discussed further in section 3.1. PMW channel frequencies are selected to take advantage of their differential sensitivities to changes in various physical parameters. Relative sensitivity values for each frequency are defined as the ratio of the change in Tb to the change in physical parameters, and they can be calculated using a method such as Kilic et al. (2021). A list of measured frequencies, their ground footprints, and their physical primary sensitivities over ocean is given in Table 1.

It should be noted that not all of the measured AMSR2 channels are used in the retrieval. The 6.9 and 7.3 GHz channels are not used due to their very large FOV and their exclusion from the original GPM constellation cross-calibration project as they had no analog channels in the reference sensor (GMI). Additionally, AMSR2 double-samples the 89.0 GHz channel to reduce observational gaps that would occur due to its very small FOV, resulting in 486 pixels per scan for the 89.0 GHz vertically-polarized (89 V) and horizontally-polarized (H) channels instead of the 243 per scan for the other channels. The 89 V and H observations used in the A2CS algorithm are taken to be the ones most closely collocated with pixel centroid of the 18.7 GHz

channel, and the FOV from this channel is taken to be the nominal footprint of the retrieval. Additionally, the 89.0 GHz B-scan is neglected due to the locations of its pixels that are typically between scans of the other channels.

Table 1: Columns: AMSR2 measured frequency, polarizations measured, nominal ground footprint size (along-track direction x across-track direction), whether or not these observations are used in the A2CS retrieval, and the relative sensitivity of each frequency to sea surface temperature (SST), ocean surface wind speed, integrated water vapor, and integrated cloud liquid water. Specifications taken from AMSR2 Data Users' Manual (2016). Sensitivity values are taken from AMSR2 Channel Specifications found at https://www.ospo.noaa.gov/Products/atmosphere/gpds/about_amsr2.html.

Frequency [GHz]	Polarization	FOV [km]	Used	Weighted Sensitivity			
				SST	Wind	Water Vapor	Cloud
6.925	V/H	62 x 35	no	1.0	0.4	0.0	0.0
7.3	V/H	62 x 35	no	1.0	0.4	0.0	0.0
10.65	V/H	42 x 24	yes	0.7	0.4	0.1	0.1
18.7	V/H	22 x 14	yes	0.1	0.5	0.5	0.2
23.8	V/H	26 x 15	yes	0.0	0.6	0.9	0.2
36.5	V/H	12 x 7	yes	0.0	0.6	0.5	0.5
89.0 (A-scan)	V/H	5 x 3	yes	0.0	0.9	1.0	1.0
89.0 (B-scan)	V/H	5 x 3	no	0.0	0.9	1.0	1.0

2.1.2 Ancillary data

The use of ancillary data is common in many retrievals, including GPM-CMB (Grecu et al. 2016). Ancillary data is considered to be any data that is needed to supplement the retrievable parameters in order to define a physical profile for use in the radiative transfer simulations. Using combined observations, however, it is possible to retrieve many of the parameters that would normally be considered ancillary. For instance, a spaceborne radar algorithm may use

model-derived surface wind speed in order to calculate the expected surface echo, which is useful in estimating the PIA of the radar beam (e.g. Haynes et al. 2009). Wind speed in A2CS is a retrievable quantity, since it is used in the radiative transfer model to calculate the surface emissivity, thus allowing it to be well-constrained by the passive observations. The ancillary data used in A2CS is minimal and consists of the air temperature profile and pressure profile; these are taken from ERA5. The corresponding pressure at each height that forms the 500-meter vertical layer boundaries are calculated from the geopotential height at each pressure level given by ERA5. This is done in logarithmic space to obtain a natural distribution of pressure with height. Any uncertainty that might arise from the use of this data is addressed in the observational error covariance matrix construction. This is discussed further in section 2.3.3.

2.1.3 Validation data

The high-latitude oceans are severely under-sampled by traditional in-situ observation data sources commonly used for satellite precipitation validation, such as rain gauges. While OceanRAIN disdrometer observations could be a reliable source of drizzle validation, the number of coincidences between CloudSat and OceanRAIN ships in precipitating conditions are likely so small that they would be statistically unusable. This problem necessitates the use of a more indirect approach for evaluating retrieval performance in this region. The line of reasoning proposed here is that validation of retrieved ocean surface parameters and integrated water vapor indicates that the algorithm seems to be converging on the correct physical environment while fitting the observations, which is at least an indirect way of assessing whether or not the precipitation parameters retrieved are physically consistent with the observations. For in-situ, there are a small number of matchups of A2CS retrievals with buoy and shipborne observations,

which provide valuable insight into retrieval performance regarding ocean surface parameters. For better sampling, however, comparisons can be made with other, independent retrieval algorithms that have undergone extensive validation globally. While not considered validation, since we are directly comparing A2CS retrievals with another satellite retrieval method, good performance in the environmental variables could indicate convergence in the precipitation parameters as mentioned above. This method is useful for clear-air and mildly cloudy scenes, but again these solutions become nonunique in the presence of precipitation. The in-situ data used consists of buoy and ship observations of SST and surface wind speed from OceanSITES (www.oceansites.org). The independent retrieval data is Remote Sensing Systems (RSS) ocean retrievals of SST, surface wind speed, and total precipitable water (TPW) on observations from AMSR2 (Wentz et al. 2014) and GMI (Wentz et al. 2015).

2.2 Domain of study and coincidence

For this study, we are interested in constraining the high latitude drizzle intensity estimations over oceans. We therefore limit the retrievals and uncertainty estimations in the algorithm to poleward of 40° latitude. The limitation of the uncertainty estimations to this domain is done to keep the retrieval sensitive to the physical environment of the high-latitude oceans and to avoid the uncertainty estimates being dominated by the warm ocean surface, calmer winds, and high-intensity convective precipitation common in the Tropics. Uncertainty estimation in the A2CS algorithm is discussed further in section 2.3.3. We use the period of observations from 1 January through 31 December 2015, since throughout this period, AMSR2, aboard GCOM-W1, and CloudSat shared the same orbital path in the A-Train (Stephens 2002).

This provides a continuous year of coincident AMSR2 and CPR observations with good coverage of the high latitudes due to their ~98° orbital inclination.

A common problem in radar-radiometer combined retrievals is the large spatial differences between the relatively small radar beam widths and the large radiometer pixel footprints. This issue is especially problematic in dealing with CloudSat data due to the large discrepancy between the CPR's ~1.4 × 1.7 km ground footprint compared with the ~14 × 22 km AMSR2 pixel footprint from the 18.7 GHz channel. Further complicating this issue is the radiometer antenna pattern and the two-dimensional (2D) gaussian nature of the pixel which results in a nonlinear falloff of contribution to the measured radiance farther away from the pixel centroid. This mismatch in sampling volumes causes a large disparity between the frequency distributions of precipitation between CloudSat and other products as a direct result of the resolution dependency of precipitation observations. It is possible however, to homogenize these sampling volumes at the statistical level by averaging the CPR observations along the orbital track. Although no unique proper solution to this method is known, there is a similar probability of precipitation between CloudSat and other observational products if the averaging length is chosen to be much longer than the horizontal length of the radiometer footprint (Behrangi et al. 2012). Stephens et al. (2010) used TRMM-PR observations and model data to test the sensitivity of the frequency of occurrence to the one-dimensional (1D) averaging length of radar data and found that a factor of 3 times the length of the 2D model gridbox area was optimal for the 1D averaging length, but acknowledged that this becomes problematic when dealing with the spatial correlation of the observations. Forced to make a compromise between statistical congruence and spatial correlation in this scene-by-scene retrieval, we choose an averaging length of 30 km,

centered at the location of the AMSR2 pixel centroid. The implications of this assumption are examined in section 2.3.5.

2.3 Algorithm description

This section contains a brief description of the method of optimal estimation in general, as well as the specific techniques used to tailor the algorithm to the meteorological domain of interest. Section 2.3.1 provides the background of optimal estimation theory, and the following sections, 2.3.2 through 2.3.5 deal primarily with the specifics of the A2CS combined algorithm.

2.3.1 Optimal Estimation

The method of optimal estimation (OE) is a physically constrained, mathematical technique utilized in many retrieval algorithms in satellite remote sensing. A few examples are Boukabara et al. (2011), Duncan and Kummerow (2016; hereafter DK16), Grecu et al. (2016), and Lebsock et al. (2011). Based in Bayesian statistical theory, this method's main advantages are 1) the ability to provide an estimate for a given atmospheric scene based on all available observations, in addition to any *a priori* information regarding the physics and statistics of the system and 2) to retrieve a state using a process that has rigorously considered the uncertainties in both the observations and the *a priori* information. A full mathematical treatment of OE can be found in numerous other works, including Rodgers (2000), so the discussion of OE here is kept brief. We combine the Level 1C cross-calibrated AMSR2 observed Tbs (Berg et al. 2016) and the coincident Level 2B CPR reflectivity profiles (Marchand et al. 2008) into the observation vector **y** and attempt to minimize the following cost function:

$$J(\mathbf{x}) = \big(\mathbf{y} - F(\mathbf{x},\mathbf{b})\big)^T \mathbf{S_y^{-1}} \big(\mathbf{y} - F(\mathbf{x},\mathbf{b})\big) + (\mathbf{x} - \mathbf{x}_a)^T \mathbf{S_a^{-1}} (\mathbf{x} - \mathbf{x}_a) \quad (1)$$

where **x** is the state vector, whose elements describe the physical state of the emissive surface and of the atmosphere through which the radiation passed, and its associated ancillary data **b**. $\mathbf{x_a}$ is the *a priori* state, F is the forward model, operating on **x** to produce simulated observations, and $\mathbf{S_y}$ and $\mathbf{S_a}$ are the observational and *a priori* error covariance matrices, respectively. The minimum of J(**x**), found by Gauss-Newton iteration, represents the most likely state, given a multivariate Gaussian probability density function of possible outcomes. The following represents a stepwise movement from any arbitrary state to the minimum of the cost function:

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \left(\mathbf{K}_i^T \mathbf{S}_y^{-1} \mathbf{K}_i + \mathbf{S}_a^{-1}\right)^{-1} \left[\mathbf{S}_y^{-1}\left(\mathbf{y} - F(\mathbf{x}, \mathbf{b})\right) + \mathbf{S}_a^{-1}(\mathbf{x} - \mathbf{x}_a)\right] \quad (2)$$

where **K** is the Jacobian matrix representing the sensitivity of F(**x**,**b**) to changes in elements of **x**. **K** is estimated by applying perturbations to the elements of **x**, and then running the forward model on the results. The elements of **K** are then directly calculated as $\mathbf{K}_{\mathrm{i,j}} = \partial F_{\mathrm{i}}(\mathbf{x}, \mathbf{b}) / \partial \mathbf{x}_{\mathrm{j}}$, where i and j are indices corresponding to the rows and columns of **K**. This method provides a good estimate of the gradient of the multivariate function J(**x**) as long as it is not too nonlinear at small distances from the current state. To help ensure that the linear slopes calculated for **K** are good approximations, perturbation sizes are kept relatively small (generally 2% or less).

The state vector **x** consists of the retrievable variables that describe the emissivity of the ocean surface and the physical state of the atmosphere through which the radiation passed. The 30-layer profiles of liquid and ice contents are retrieved for each layer independently, therefore they exist in the state vector as elements 1-60. Their *a priori* values and retrievability are set based on whether the radar reflectivity for that layer is determined to be above the algorithm's assumed noise floor of -26 dBZ. If the reflectivity for a layer is above the noise floor, the first guess for that layer is set in the a priori state vector at 0.01 g m^{-3}. If the reflectivity for the layer is below the noise floor, the layer is assumed free of liquid and solid hydrometeors, and the

algorithm does not attempt to retrieve a nonzero value for liquid or ice in that layer. The remaining variables are retrieved as single quantities that are considered effective for the entire scene, namely, the residual cloud water path (discussed further in Section 2.3.4), the first three EOFs of water vapor, the retrievable PSD parameters (also discussed in section 2.3.4), SST, and surface wind speed. The remaining ancillary data, **b**, are necessary for physical forward model simulations and are read in a priori, but they are held fixed in the retrieval. A list of retrieved variables, ancillary data, and their a priori uncertainties is given in Table 2.

Table 2: Columns: variables used in the retrieval, whether they are retrieved or fixed to their *a priori* value, the first guess value (or the origin of the first guess) and its *a priori* uncertainty, and the observational instrument primarily sensitive to the radiation signal.

Variable	Retrieved or fixed	First guess and *A priori* uncertainty (σ^2)	Observational sensitivity
Sea Surface Temperature	Retrieved	From Reynolds SST (Reynolds et al. 2007), 0.5625 K^2	AMSR2
Air temperature profile	Fixed	From ERA5, none	N/A
Surface Wind Speed	Retrieved	From ERA5, 4.0 m^2 s^{-2}	AMSR2
Sea Surface Salinity	Fixed	35.0 psu, none	N/A
Water Vapor EOF1	Retrieved	From SST-dependent mean profile, 1.34	AMSR2
Water Vapor EOF2	Fixed	N/A (see section 2.3)	N/A
Water Vapor EOF3	Fixed	N/A	N/A
Residual Cloud Water Path	Retrieved	10 g m^{-2}, $\log_{10}(2.0)$ g^2 m^{-4}	AMSR2
Liquid Precipitation Water Content	Retrieved	0.001 g m^{-3}, 2.0 g^2 m^{-6}	both
Ice Water Content	Retrieved	0.001 g m^{-3}, 2.0 g^2 m^{-6}	both
Gamma PSD shape parameter μ	Retrieved	1.5, 0.5	both
Ice particle density	Retrieved	1.0 g cm^{-3}, 0.04 g^2 cm^{-6}	both

Convergence on the minimum of J(**x**), or the most likely state given the observation vector, is determined when the residual distance δ^2 between **y** and the simulated observations produced by the retrieved state,

$$\delta^2 = \left(\mathbf{y} - F(\mathbf{x}, \mathbf{b})\right)^T \mathbf{S}_\mathbf{y} \left(\mathbf{y} - F(\mathbf{x}, \mathbf{b})\right) \quad (3)$$

becomes small. A small value for δ^2 indicates that there is little difference between the simulated observations produced by the retrieved state and the observation vector. Chi squared, another common convergence metric, includes the fit of the converged state vector and the *a priori* state:

$$\chi^2 = \left(\mathbf{y} - F(\mathbf{x}, \mathbf{b})\right)^T \mathbf{S_y^{-1}} \left(\mathbf{y} - F(\mathbf{x}, \mathbf{b})\right) + (\mathbf{x} - \mathbf{x}_a)^T \mathbf{S_a^{-1}} (\mathbf{x} - \mathbf{x}_a) \quad (4)$$

When using the Gauss-Newton iteration method, the algorithm penalizes excessive departure from both the *a priori* state and from the observations, therefore the minimum of chi squared is also the minimum of the cost function. Since minimizing chi squared is implicit in the iteration method used in this algorithm, we use the residual distance δ^2 to determine convergence, similar to Pfreundschuh et al. (2020). The convergence threshold in the A2CS algorithm is taken to be when the value of δ^2, normalized by the number of observations, is less than or equal to 2.0. Conversely, the retrieval is considered to fail when the algorithm reaches 20 iterations and δ^2 is higher than 2.0.

An additional consideration in a combined retrieval is the asymmetric weighting of the fit of the radar profile and the fit of the Tbs between simulated and observed measurements. It is common for the radar profile to dominate δ^2 due to the imbalance between the large number of highly correlated radar range gates and the much smaller number of radiometer channels. This issue is addressed in two ways. Firstly, averaging the radar reflectivities from the CPR's native resolution of 240 m to 500 m per range bin reduces the number of radar bins from 125 per profile to 30 extending from the surface to 15 km. This also helps to reduce the accumulation of noise in the forward model that could occur by tracing scattering and absorption through too many atmospheric layers given the information content of the sensors. The second way is by utilizing an observational error covariance matrix that includes the estimations of cross-correlated errors between each element of **y** (described in section 2.3.3).

2.3.2 Forward Model

The forward model is analogous to the mapping function in a variational data assimilation scheme, as it directly maps the state space to the observation space. In the case of the OE algorithm, this is done by using a physically based numerical radiative transfer model to simulate both the active and passive observations. The model calculates the radar reflectivity factors for each radar volume and the top of atmosphere (TOA) microwave radiances at both horizontal and vertical polarizations that correspond to each of the AMSR2 channels. The effective radar reflectivity factor Z_e is the theoretical maximum reflectivity factor and is first calculated from the PSD as

$$Z_e = \frac{\lambda^4}{\pi^5 |K|^2} \int_{D=D_{min}}^{D=D_{max}} n(D)\sigma_b(D)dD \quad (5)$$

where λ is the operating wavelength of the radar, $|K|^2$ is the dielectric constant of cloud particles, n(D) is the number concentration of particles, in mm^{-1} m^{-3} (given by the PSD), and $\sigma_b(D)$ is the backscatter cross-section of the individual particle at diameter D. $|K|^2$ is wavelength and phase dependent, but is assumed to be 0.75 for the W-band. This value is the one assumed for an operating frequency of 94.0 GHz in QuickBeam, an operational radar simulator often used for studies involving CloudSat (Haynes et al. 2007).

The passive observations are simulated using an Eddington plane-parallel, two-stream approximation. Surface emissivity and extinction coefficients are calculated for gaseous absorption at each channel frequency, along with the Mie scattering parameters for all cloud and hydrometeor particles. The resulting total atmospheric optical depth is used to calculate the final TOA radiance for each channel frequency and polarization as well as the PIA of the radar beam. The PIA is not used as an explicit constraint in this algorithm as it is in the 2C-RAIN-PROFILE

algorithm, but the volume-by-volume attenuation is estimated for directly matching the retrieved state's simulated reflectivities with the observed reflectivities.

2.3.3 Uncertainties

The construction of the error covariance matrices, $\mathbf{S_a}$ and $\mathbf{S_y}$, is not straightforward and no unique solution for estimating their values is known to exist. These uncertainties are the primary drivers of the location of the posterior state within the multivariate space and are used to place varying amounts of constraint on the final solution. Low uncertainties in the *a priori* essentially assumes that the true state is likely not far from the mean background. Low observational uncertainties tightly constrain the solution to the observations and allow for more deviation from the *a priori*. It is therefore of great importance that uncertainty estimates contained in the error covariance matrices are well-suited to the problem. We use a method similar to DK16 for the estimation of $\mathbf{S_y}$, and we assume diagonality for $\mathbf{S_a}$.

The diagonal elements of $\mathbf{S_a}$ essentially describe the variance of the departure of the retrieved state elements from the a priori knowledge of the state, and since the true state is unknown beforehand, the problem is not well-posed. In other works, these values have been estimated with multiple techniques and with varying amounts of importance placed on their estimation. In Boukabara et al. (2011), a distinction is made between the first guess, or the starting point for iteration, and the a priori. Theoretically, the location of first guess should have no impact on the location of the solution in the multivariate space, if there exists a single, unique minimum. The variances and cross-covariances for $\mathbf{S_a}$ in their algorithm are carefully derived from numerical weather prediction (NWP) model analyses in order to ensure that retrieved states are close to the *a priori*. In A2CS, we assume that the *a priori* is not as well-known, since the

high-latitude oceans are severely under-observed. By focusing on more careful estimation of the observational and forward model uncertainties, we can ensure that the retrieval is driven primarily by the information contained in the observations.

For this study, the diagonal elements of $\mathbf{S_a}$ that describe the variances in SST, wind speed, water vapor, and residual cloud water path are derived from values estimated for other works, namely DK16 and Schulte and Kummerow (2019). These values can again be found listed as variances along with their physical units in Table 1. The layer-dependent liquid and ice water content have relatively larger uncertainties (1.41 g m^{-3}) compared with the first guess (0.01 g m^{-3}). This is due to the fact that volumetric liquid and ice water content varies widely in nature, and larger *a priori* uncertainties for rain and snow water content allows the observational error covariances a more important role in determining the final solution (Schulte and Kummerow 2022). These first guesses and variances were also selected in order to optimize the algorithm's iteration process and to allow it to find sensitivity and calculate meaningful gradients without taking large steps into unphysical space. The variances for the liquid and solid retrievable PSD parameters μ and ρ_{ice} were selected in a similar way but have relatively smaller uncertainties compared with the first guess. While in theory it is relatively straightforward to estimate the *a priori* uncertainties in PSD parameters, PSDs for high-latitude oceanic drizzle are rarely observed, and the disdrometer observations in OceanRAIN fail to capture drop sizes below 0.392 mm that are important for fully characterizing drizzle PSDs.

It is acknowledged that the *a priori* error variances for liquid and ice water content and the two PSD parameters, as well as the assumption of diagonality for this matrix, could be areas of focus for future improvements to the algorithm, for instance, by using a process similar to Boukabara et al. (2011) to estimate these uncertainties. This is considered outside the scope of

this study, and the assumption of a diagonal $\mathbf{S_a}$ matrix has been used here in order to allow the observational uncertainties to dominate the total prior retrieval uncertainty. Furthermore, radar bin correlation is considered in the formulation of the $\mathbf{S_y}$ matrix, therefore it is ultimately considered in the formulation of the posterior estimation of the state according to equation (2).

The construction of $\mathbf{S_y}$ is very important to the final fit of the retrieval to both sets of observations and therefore requires more careful attention. $\mathbf{S_y}$ encompasses the uncertainty in the measurements and in the forward model, together forming the total observational uncertainties. The measurement uncertainties are given in terms of the noise equivalent differential temperature of the radiometer and the calibration error of the radar and are relatively small. The overall magnitudes of the observational uncertainties are driven by assumptions necessary in the forward model computations. These errors are much larger and must be accurately estimated in order for the retrieval to converge on physically realistic states. As mentioned before, we use a method similar to DK16 in order to estimate both the diagonal and off-diagonal elements of $\mathbf{S_y}$, with some important differences due to the retrieval of the vertical liquid and ice profile, as well as the inclusion of radar observations along with the PMW observations. The remainder of this section contains an overview of how the $\mathbf{S_y}$ matrix is estimated, and it is again based on the work of DK16 with important modifications and added complexities due to the complications inherent in the problem of combining radar and PMW observations together. More background information on this method can be found in DK16.

In order to estimate the uncertainties in both passive and active observations simultaneously, we simulate the departure from the true profile that a successful retrieval would produce. Atmospheric profiles from ERA5 of temperature, pressure, humidity, and cloud and hydrometeor content are averaged and interpolated from their native resolution of 27 pressure

levels between 1000 and 100 hPa to 30 layers of 500 m each between the surface and 15 km to emulate the vertical resolution of the retrieval. The cloud water, cloud ice, and liquid and frozen hydrometeor contents are given in the ERA5 profiles, and PSD parameters are randomly assigned using the models described in section 2.4. These profiles represent the "true" profiles, and they are forward modeled to simulate the "error-free" observations. To maintain the sensitivity of $\mathbf{S_y}$ to the high-latitude ocean regime, only profiles over ocean and poleward of and including $\pm 40°$ latitude are considered.

The same profiles are then reconstructed in the format used by the retrieval, and then these profiles are forward-modeled as well. For water vapor, this is done in EOF space by deconstructing the ERA5 water vapor profile and reconstructing it according to its first three EOFs. In DK16, retrieving the water vapor profile in EOF space was found to successfully constrain the vertical profile in the presence of clouds, where the vertical resolution of the retrieved profile depended on the number of channels measured that were sensitive to water vapor emission. Generally, reconstruction of a profile is done as $\boldsymbol{w}_{retr} = \bar{\boldsymbol{w}} + c_1\boldsymbol{e}_1 + c_2\boldsymbol{e}_2 + \cdots + c_n\boldsymbol{e}_n$, where $\boldsymbol{w}_{retr}$ is the water vapor in retrieval resolution, i.e. 30 layers, $\bar{\boldsymbol{w}}$ is the mean profile, in this case derived from an analysis of ERA5, c is the corresponding EOF coefficient, and $\boldsymbol{e}$ is the corresponding EOF. This technique was originally applied to GMI, and three EOFs were found to be well-retrieved due to its high-frequency sounding channels. However, since AMSR2's water vapor sensitivity lies primarily in the two 23 GHz channels, we find that the first EOF, directly related to the total precipitable water, is all that can be skillfully retrieved in the presence of clouds and precipitation. Therefore, the water vapor profile in the retrieval is vertically constructed based using the SST-dependent mean mixing ratio, derived from the original analysis of humidity profiles from ERA5 done for DK16, and a first guess of 0.01 is

given for the first EOF coefficient. The algorithm is then allowed to modify the EOF coefficient with each iteration, reconstructing the vertical profile of water vapor in the radiative transfer model by $\boldsymbol{w}_{retr} = \bar{\boldsymbol{w}} + c_1\boldsymbol{e}_1$. Since only the first EOF is used in the reconstruction, the profile has a relatively fixed overall vertical shape, but it is able to be adjusted in EOF space to account for the correct total integrated value. Figure 2 shows the mean mixing ratio values for each SST bin, along with the first three EOFs in units of mixing ratio (g kg^{-1}). Again, only the first EOF is retrieved in A2CS due to reduced information content with respect to water vapor, but three were used in the uncertainty estimation due to the method being originally applied to sounders. The SST-dependent mean profile and EOF vectors for EOFs 1-3 are shown in Figure 2 for reference.

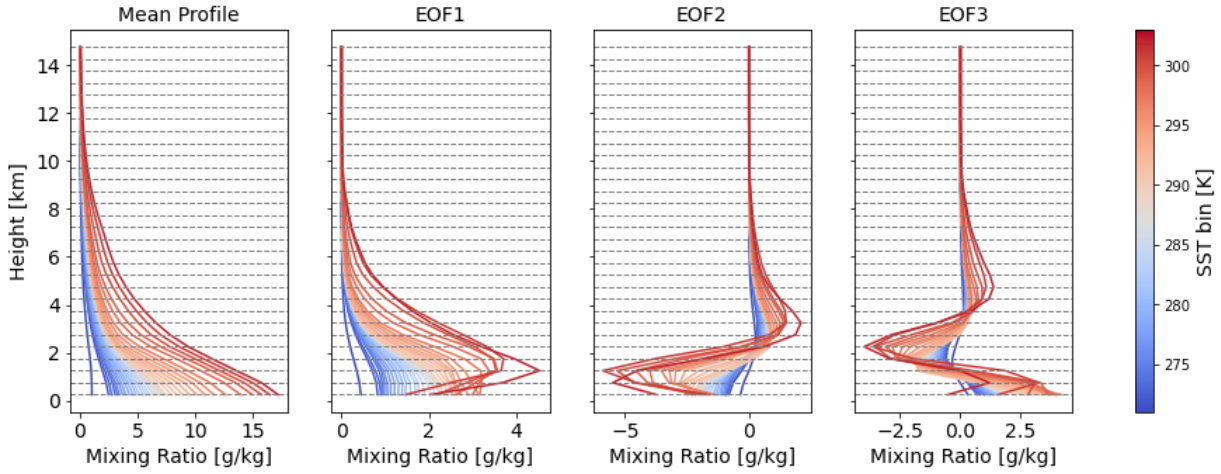

Figure 2: Mean profile of water vapor by SST bin (left panel), and the first three EOFs of water vapor, reprojected into mixing ratio space.

Adding noise to the temperature profile attempts to account for the departure of ERA5 from the unknown "true" profile. The noise added to the temperature profile is drawn from a gaussian distribution, with an uncertainty taken to be the DK16 value of 2 K. ERA5 cloud water content is integrated and redistributed evenly between the cloud base and the freezing level, which is the same process the algorithm uses for vertically distributing cloud water that does not contribute to the precipitation content (see section 2.3.4 for more detail). We further assume that

the forward-model uncertainties caused by liquid and ice hydrometeor scattering increases with the amount present. A scaled noise is added to the hydrometeor contents as $\rho_{hyd,retr} = \rho_{hyd,ERA}(1+\epsilon)$, where ϵ is a random noise value drawn from a gaussian distribution with a standard deviation taken to be 0.03 g m^{-3}. Finally, PSDs are remodeled by randomly reassigning moveable parameters μ and ρ_{ice} within their retrieval bounds and by randomly varying the fixed *a priori* parameters ($N_{0,liq}$ and $N_{0,ice}$) by a maximum of one order of magnitude in either direction to attempt to approximate the effect of using any fixed PSD model and its variance from the "true" (natural) PSD.

These resulting reconstructed profiles are again forward modeled, and the difference between the two sets of simulated observations, $\mathbf{y}_{sim,ERA5} - \mathbf{y}_{sim,retr}$ is deemed the "error." To calculate the elements of $\mathbf{S_y}$, we simply calculate the covariance of $\mathbf{Y}_{sim,ERA5} - \mathbf{Y}_{sim,retr}$, where $\mathbf{Y}$ is a matrix of dimensions [m x n], with m being the sampling dimension and n being the number of elements in $\mathbf{y}$. Finally, the NEDT values for AMSR2 are added to the diagonal elements of $\mathbf{S_y}$ that correspond to the AMSR2 channels. These are taken from the World Meteorological Organization's specifications for AMSR2 (2022).

As stated before, estimating $\mathbf{S_y}$ as accurately as possible is paramount to producing successful and accurate retrievals. A degradation in retrieval quality is found when $\mathbf{S_y}$ is ill-suited for describing the errors in a given scene—for example, including error contributions based on highly uncertain ice scattering in the scene when there are no clouds present. We therefore allow the algorithm to use a different version of $\mathbf{S_y}$ based on whether the scene is predetermined to be clear, cloudy, or precipitating. This predetermination is a unique feature of using passive and active observations together, for we could not make such a discrimination uniquely using only

passive observations. The scene is determined to be clear when all CPR reflectivities are at or below -26 dBZ. Scenes are determined to be cloudy, but not precipitating, when the CloudSat along-track averaged surface precipitation rate (a combination of 2C-RAIN-PROFILE and 2C-SNOW-PROFILE for the best estimates) for the AMSR2 pixel is below 0.01 mm hr^{-1}, and scenes are considered to be precipitating when it is above or equal to this threshold. The only function of this predetermination between precipitating and cloud-only scenes is to allow the algorithm to use the most appropriate version of $\mathbf{S_y}$, and no further constraint is made to produce an appreciable precipitation rate on any scene; this allows a smooth transition in the retrieval statistics between the cloudy and drizzling regimes.

The different versions of $\mathbf{S_y}$ are constructed by including or excluding profiles in the error covariance calculation described above based on whether the profiles are considered to be clear, cloudy, or precipitating by ERA5. Each of these matrices are shown in Figure 3, where it can be seen that uncertainties generally increase with the amount of water content in each layer. A noticeable drop in correlation between radar bins can be seen in the cloudy $\mathbf{S_y}$ when compared with the clear and precipitating versions of $\mathbf{S_y}$. This is likely due to the natural variablility of nonprecipitating clouds, especially high-level cirrus or stratus clouds, which are vertically thin and tend to occur in just one layer. Precipitating clouds, on the other hand, such as cumulonimbus or even nimbostratus, often stretch through many vertical layers and lead to the higher correlation observed between radar bins in the precipitating version of $\mathbf{S_y}$. Similarly, nominally clear scenes, which can in reality contain a very small amount of cloud water, will lead to very high correlation between radar bins.

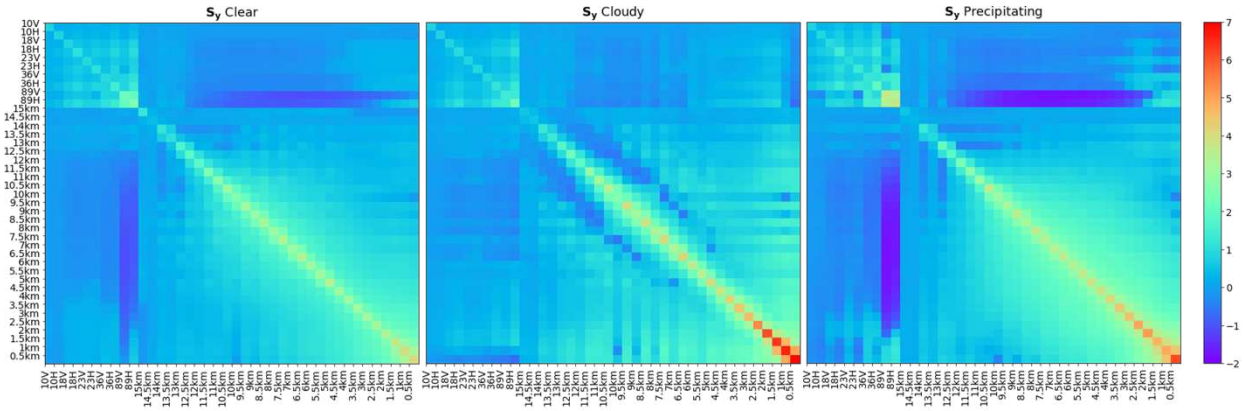

Figure 3: Observational error covariance matrices for clear (left), cloudy (center), and precipitating (right) scenes. Errors are plotted using the signed square root of the variances in K for the AMSR2 channels and in dBZ for the radar height bins.

The last form of uncertainty to consider exists in the form of the posterior, or retrieval, uncertainty. This uncertainty, a scalar value, is directly estimated at retrieval time for each retrieved state and is expressed in terms of the square root of the variance, or standard deviation, of the final solution:

$$\sigma = \left(\mathbf{S_a^{-1}} + \mathbf{K}^T\mathbf{S_y^{-1}}\mathbf{K}\right)^{-0.5} \tag{6}$$

This uncertainty metric is generally valid for physical parameters that are understood to be approximately Gaussian in nature, such as SST and wind speed, but fails to capture the true retrieval uncertainty in any meaningful way if the problem is non-Gaussian. Rain and snow water content, as well as cloud water and ice, are extremely non-Gaussian and vary widely in their natural distributions. Therefore, the posterior uncertainties for these quantities are not a good estimate of the true retrieval uncertainty. An example of this problem is illustrated in Figure 4, where the example retrieved precipitation rate profiles are within the drizzle regime as defined in this study (below 0.25 mm hr^{-1}), but the uncertainties in the precipitation rates seem to indicate that the true rate could be a wide range of values from near zero to well into the DPR's detection

capability. Furthermore, pixel-level retrieval uncertainties disappear in the zonal accumulations, and given this study's focus on drizzle climatology, they will hereafter be neglected.

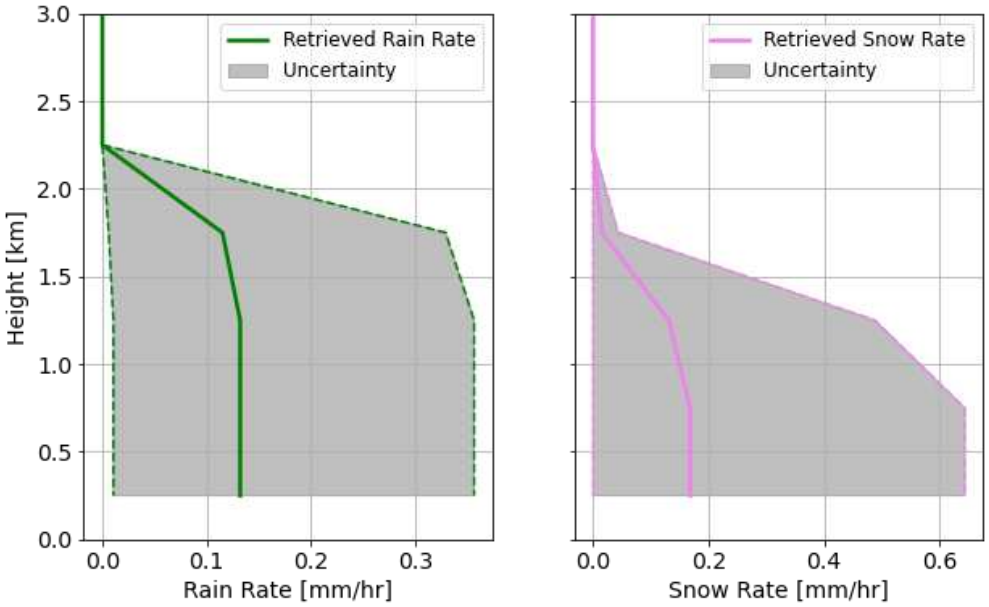

Figure 4: Example retrieved precipitation rate profiles (solid lines) with uncertainties (gray shading) plotted. The uncertainty limits (dashed lines) in precipitation rate are found by calculating the precipitation rate on the retrieved rain and snow water content for each layer, plus and minus one standard deviation of retrieval uncertainty as defined in equation (6).

2.3.4 Cloud and hydrometeor PSD models

Various PSD models have been used for clouds and precipitation in both observation and modeling studies, with the most common being some form of the three-parameter gamma distribution (Ulbrich 1983). Most often, three moveable parameters are insufficiently constrained for most observational methods, and various assumptions are used to reduce the number of free parameters. CloudSat's 2C-RAIN-PROFILE algorithm uses a modified form of the gamma distribution, where additional constraints, taken from Abel and Boutle (2012), reduce the number of free parameters from three to one (Lebsock 2018). The GPM-CMB algorithm uses a normalized gamma distribution (Testud et al. 2001) and reduces the number of free parameters by assuming a fixed value for one of the parameters (Grecu et al. 2016). Schulte and Kummerow

(2022), however, showed that retrieval errors are sensitive to assumptions necessary to sufficiently constrain the PSD models, and therefore it is important that as general a PSD as possible be used in the retrieval to provide enough flexibility in the PSD model to reduce uncertainties stemming from these assumptions. Using combined observational methods, two moveable parameters are possible to provide maximum flexibility. This section describes the PSD models assumed in the A2CS algorithm, separated into subsections based on how the algorithm handles clear, cloudy, and precipitating retrievals. Both cloudy and precipitating scenes use a two-parameter gamma distribution (described in equation (7)), based on the fundamental assumption that there is no difference between the type of PSD model used to describe clouds and precipitation—only that the parameters may be different. This assumption is used due to this study's focus on low precipitation rates and small particle sizes important for drizzle and is discussed further in section 2.3.4.2.

2.3.4.1 Clear retrievals

Scenes are determined to be clear whenever the CPR reflectivity profile shows all bins at or below the algorithm's noise floor of -26 dBZ. While the scene is nominally clear, there may be some cloud water that is undetectable by the radar due to small drop sizes but may be contributing emission signals to the passive measurements and to the attenuation of the radar beam. With no appreciable backscatter signal to constrain the drop sizes, this cloud water is modeled as monodisperse, with all cloud droplet diameters taken to be a single value D_{cld}. Also, due to a lack of any vertical sensitivity, cloud water is then evenly distributed between the freezing level and the surface bin. Retrieved residual cloud water paths in clear scenes are low and are most often less than 0.02 mm.

2.3.4.2 Cloudy retrievals

Scenes are determined to be cloudy and not precipitating whenever CPR reflectivities are above the noise floor, but the 30 km along-track averaged CloudSat precipitation rate is less than 0.01 mm hr^{-1}. Since there is signal in this case for both the cloud water path from the radiometer and the particle sizes from the radar, the algorithm distributes cloud water or cloud ice according to a prescribed PSD. In this combined algorithm, we assume no fundamental difference in the PSD model (only that the parameters can change) before and after drizzle onset. The liquid PSD is modeled as a nonnormalized gamma distribution:

$$n(D) = N_{0,liq} D^{\mu} \exp(-\Lambda D) \quad (7)$$

where n(D) is the number concentration, in m^{-3} mm^{-1}, of drops at diameter D, $N_{0,liq}$ is the intercept parameter, μ is the shape parameter, which is allowed to vary between 0 and 2.5, and Λ is the slope parameter in mm^{-1}. In the algorithm μ and the liquid water content are retrievable, while $N_{0,liq}$ is fixed at 1.1×10^5 m^{-3} $mm^{-\mu}$. This value for $N_{0,liq}$ was derived from an analysis of high-latitude drizzle in-situ observations from OceanRAIN, and the combination of this value along with varying μ was determined to be flexible enough to capture the natural variability of drizzle PSDs. This is illustrated in Figure 5, where 5-minute averaged PSD observations from the OceanRAIN disdrometers are shown with two example PSD model fits. While the disdrometer cannot detect drizzle droplets smaller than a diameter of 0.392 mm, and therefore the true behavior of natural PSDs at these sizes is unknown, the two-parameter liquid gamma PSD offers additional flexibility to capture the behavior of natural PSDs at these sizes with the limitations imposed by assuming that μ can vary only between values of 0.0 and 2.5.

As stated before, μ and $N_{0,liq}$ are considered valid for the entire scene, while the slope parameter Λ is constrained by the liquid water content in each vertical radar bin. Since the slope

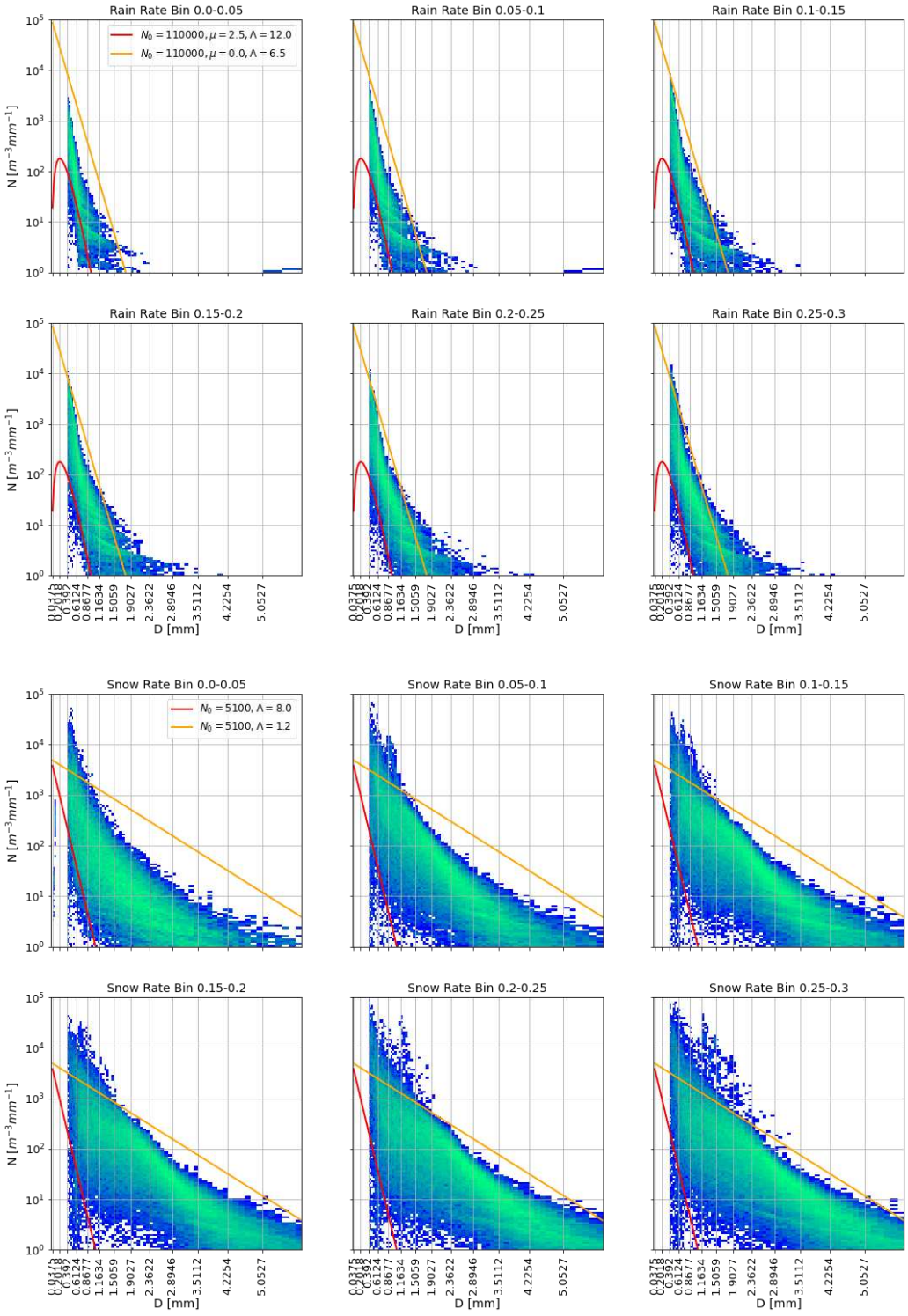

Figure 5: 5-minute averaged PSD observations from OceanRAIN shipborne ODM disdrometers for high latitude oceans. Color from blue to green indicates increasing density of observations. Also plotted are two extreme fits of the PSD models used in the algorithm to show sufficient flexibility to capture these observed PSDs.

parameter is not able to be derived from the liquid water content analytically, it is found using a high-resolution lookup table in the algorithm that was made using the following relationship:

$$LWC = \int_{D_{min}}^{D_{max}} n(D)V(D)\rho_l dD \quad (8)$$

where n(D) is the PSD described above, V(D) is the volume of a single drop, and ρ_l is the density of liquid water, taken to be 1.0 g cm^{-3}.

In cloudy and precipitating retrievals, there is often a small amount of residual cloud water that is needed to correctly simulate the absorption and emission throughout the atmospheric column that is either not captured by the liquid PSD or that does not contribute to the radar reflectivities. This cloud water is also modeled as monodisperse with the same value for D_{cld} as in the clear case, and it is vertically evenly distributed between the freezing level and the cloud base. This monodisperse residual cloud water is used here to mimic CloudSat's 2C- RAIN-PROFILE algorithm that accounts for some attenuation of the radar beam and that is also not part of the hydrometeor PSD (Lebsock 2018). In 2C-RAIN-PROFILE, the cloud water amount is constrained to be a function of the surface rain rate. Instead, in the A2CS algorithm, the residual cloud water path is retrievable and is always small. More than 96% of cloudy retrievals were found to have retrieved residual cloud water paths of less than 0.2 mm.

Both precipitation and cloud ice particles are modeled as spherical for radiative transfer computations and are distributed according to a simple inverse exponential PSD:

$$n(D) = N_{0,ice} \exp(-\Lambda D) \quad (9)$$

where $N_{0,ice}$ is the intercept parameter, derived from in-situ observations of high latitude oceanic snow (again from OceanRAIN), and is fixed at 5100 mm^{-1} m^{-3}. The inverse exponential PSD is a special case of the gamma distribution where $\mu = 0$ and has been shown to be effective in

capturing both cloud ice and snow particles to correctly simulate W-band radar observations (Liu 2008a), and it is additionally consistent with aircraft observations of snow clouds (Heymsfield et al. 2008). The slope parameter Λ is then constrained by the ice water content and is calculated as

$$\Lambda = \left(\frac{N_{0,ice}\rho_{ice}}{IWC}\pi\right)^{\frac{1}{4}} \tag{10}$$

Unlike liquid drops, however, an additional factor that must be considered that is radiatively important in the microwave spectrum is the ice particle density ρ_{ice}, which strongly affects microwave scattering via changes in the Mie scattering parameters. This is done in the forward model, which uses the "fluffy sphere" approximation by modeling ice particles as pure ice spheres with elliptical air inclusions. The Maxwell-Garnett formula is then used to calculate the effective dielectric properties for the particles at each diameter, and the Mie scattering parameters are calculated for the entire PSD for each layer. Therefore, ρ_{ice}, constrained to remain between 0.05 and 0.4 g cm^{-3}, is a retrievable variable in the algorithm and is considered to be an effective particle density valid for all cloud layers to characterize the scene. In reality, ρ_{ice} is likely not uniform with height, but by assuming it is constant with height throughout all frozen layers, it allows the algorithm to avoid unphysical vertical gradients in ρ_{ice} that would result from from algorithm instability due to too many free parameters.

2.3.4.3 Precipitating retrievals

A precipitating retrieval is performed on scenes where the CloudSat along-track mean precipitation rate is greater than or equal to 0.01 mm hr^{-1}. Liquid and solid precipitation particles are modeled according to the PSDs described above, with the additional step of calculating a liquid and solid precipitation rate for each bin. The precipitation rate for liquid drops is:

$$R_{liq} = \int_{D_{min}}^{D_{max}} n_{liq}(D) V(D) v(D) dD \quad (11)$$

where $n_{liq}(D)$ is the number concentration of drops at diameter D, given by the PSD, V(D) is the volume, and v(D) is the terminal velocity of individual rain drops, calculated according to equation (1) of Villermaux and Eloi (2011). For snow, a similar scheme is implemented, but the liquid equivalent precipitation rate follows from the liquid equivalent volume of the snow particles, or:

$$R_{ice} = \int_{D_{min}}^{D_{max}} n_{ice}(D) \frac{m_{ice}(D, \rho_{ice})}{\rho_l} v(D, \rho_{ice}) dD \quad (12)$$

where m_{ice} is the mass of the ice particles at diameter D, and ρ_l is the density of liquid water. The terminal velocity of snow particles is difficult to estimate directly and can be parameterized in many different ways (Matrosov 1998). We use the parameterization $v(D, \rho_{ice}) = 8.8\sqrt{D(\rho_{ice} - \rho_{air})}$, valid in the DPR algorithm for snow particles with densities between 0.05 g cm^{-3} and 0.3 g cm^{-3} (Iguchi et al. 2021). Ice particle density values allowable in A2CS retrievals range from aggregate-like (0.05 g cm^{-3}) to graupel-like (0.4 g cm^{-3}) and have a strong effect on the liquid equivalent precipitation rate. Thus, having combined-observational constraints on the retrieval of ice particle density produces precipitation rates that are consistent with both sets of observations.

One main assumption in the algorithm that is necessary to fill in the profile below the lowest good radar bin and to reduce surface clutter is the implementation of ERA5 cloud base heights. Surface echo contamination of the lower radar range gates is routinely found in spaceborne radar observations of all frequencies, and the CPR suffers from this problem as well. For instance, Lamer et al. (2020) found that the CPR cannot directly resolve the cloud base in approximately half of shallow marine cloud observations. The 2B-GEOPROF algorithm does

provide an estimated correction factor for the surface echo contamination, but this correction factor is not used in the A2CS algorithm since it is based on an estimate of surface wind speed that comes directly from ancillary data and would therefore introduce more assumptions to be carried over into A2CS. The CPR therefore cannot be reliably used to detect the cloud base in cloudy and precipitating retrievals. Instead, the radar bin above the ERA5 cloud base height is used as the lowest resolved precipitating bin in the retrieval. Residual cloud water, however, is allowed to extend into the bin that contains the cloud base but does not contribute to the precipitation rate. This process improves convergence in the radar profile and is very effective in reducing the number of profiles contaminated with unresolved surface clutter, which can be the source of inflated precipitation rates in CloudSat.

2.3.5 Testing of algorithm assumptions

To assess the dependency of retrievals on the assumptions necessary in the algorithm, several sensitivity tests were performed. As mentioned before, the along-track averaging of CPR reflectivities is necessary to approximate equivalent sampling between the radiometer and the radar. In addition to the 30-km averaging length assumed in the retrieval, we also tested a 15-km and 45-km averaging length to get a sense of the dependence of the precipitation accumulations on the averaging length. We found a 5.0% decrease in mean precipitation for a month of retrievals when using a 15-km averaging length compared to the assumed 30-km averaging length. Similarly, using a 45-km averaging length showed an increase of 3.4%. This suggests that the bulk statistics are not overly sensitive to the averaging length to within a reasonable value based on the along-track radiometer field-of-view (FOV) length.

Inhomogeneities in the FOV are an additional potential source of retrieved precipitation rate uncertainty. Nonuniform beamfilling (NUBF) is a well-known source of underestimation in both radar and PMW retrievals at the climate scale (Kummerow 1998; Schulte et al. 2023). Although a full characterization of and a correction for NUBF errors is beyond the scope of this study, a simple synthetic retrieval was performed to get a first-order estimate of how sensitive these precipitation retrievals might be to an inhomogeneous cloud field. We hypothesized that the sensitivity of the retrieval to NUBF might not be as great for drizzle as it would be for higher rain rates and cloud water contents. Figure 6 shows the relatively linear behavior of the AMSR2 Tbs in this regime with the notable exception of 89H for liquid water. The CPR reflectivities show extremely nonlinear behavior in both liquid and ice, but the saturation effect due to the liquid water attenuation is not as prominent in ice clouds.

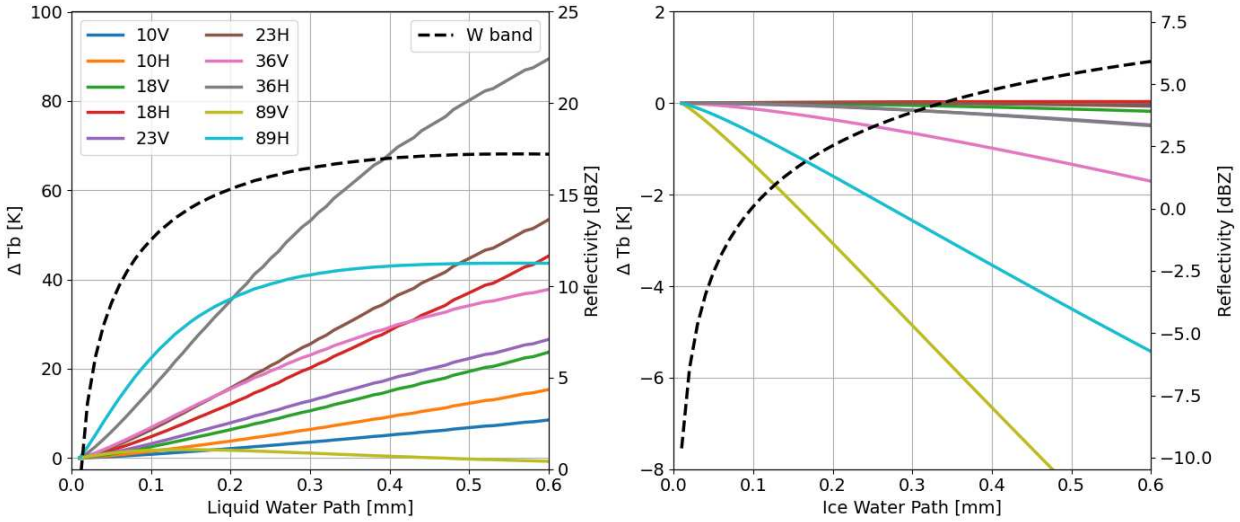

Figure 6: Change in Tb with respect to the clear-scene Tb as a function of cloud liquid water path (left panel) and cloud ice water path (right panel) for AMSR2 channels calculated using example high-latitude profiles and an assumed 1-3 km cloud. Also plotted is cloud top reflectivity for the CPR operating frequency (right axis). Upper limit of 0.6 mm corresponds to an approximate 90% probability of precipitation using 30-km CloudSat estimates in Figure 6.

An inhomogeneous cloud field was created by filling a synthetic AMSR2 pixel with an assumed liquid cloud from 1-2 km that increased in cloud water content horizontally across the FOV from 0.025 to 0.05 to 0.1 g m^{-3}, each occupying one-third of the scene. Radar observations were simulated at the 1-km radar resolution and averaged along the orbital track through the scene to produce the 30-km radar "observation." The "observed" Tbs for this scene were created by using a 2D gaussian weighted average of the Tbs simulated at the radar resolution across the AMSR2 FOV for this scene. The mean radar precipitation rate for the inhomogeneous cloud field of 0.22 mm hr^{-1} was determined by along-track averaging the precipitation rates for each radar observation calculated according to equation (7). A combined retrieval was then performed on these simulated observations and produced a precipitation rate of 0.28 mm hr^{-1}. We compared this answer to a retrieval using a homogeneous cloud field that completely filled the FOV with the 2D gaussian weighted average cloud water content (0.056 g m^{-3}) for this pixel. Once again, both radar and radiometer observations were simulated for this scene, and the subsequent retrieval resulted in a precipitation rate of 0.27 mm hr^{-1}. This test is far from a full investigation into NUBF errors, but the results of this simple experiment seem to indicate that the radiometer is largely able to correct for the NUBF from the radar because of the linear relationships in this portion of the cloud/rain water Tb space.

We also tested the sensitivity of the simulated top-of-atmosphere radiances to the choice of the monodisperse residual cloud water diameter, D_{cld}, by modeling liquid cloud water in varying amounts from 0.05 to 0.25 g m^{-3} and with varying values of D_{cld} from 10 to 100 μm. We found that there is no appreciable effect on the top-of-atmosphere microwave radiances from the choice of D_{cld} between these bounds, therefore it is taken to be 10 μm.

The assumption of spherical ice particles is usually unphysical in reality, but it is common in many operational retrieval algorithms, including GPM-CMB (Grecu et al. 2016). More recent work has shown that the "fluffy sphere" approximation may be a significant source of microwave scattering calculation error when it is used to approximate scattering by more complex ice habits (Kuo et al. 2016). However, fully parameterizing the radiative effects of different particle shapes is beyond the scope of this study, but it is likely that snowfall estimates are sensitive to them. The main sensitivity to these signals are at higher frequencies, such as the 89 GHz AMSR2 channels and the 94 GHz CPR operating frequency, for which particle orientation is also currently thought to modify scattering parameters (Brath et al. 2020). These effects, however, as well as multiple scattering for these frequencies, only become important at large particle sizes and generally at heavier precipitation rates (Haynes et al. 2009) that are beyond the drizzle size regime focused on in this study. Furthermore, introducing particle shape parameterizations may increase the number of free parameters and cause instabilities when only using a single-frequency radar and a radiometer without high-frequency channels. Reduction of the possible error introduced by the "fluffy sphere" approximation is an improvement reserved for future work as discussed in Chapter 4.

CHAPTER 3: RESULTS

3.1 Algorithm validation

As already mentioned, validation over oceans, especially in the high latitudes, is challenging due to the relatively small number of retrieval matchups with sparse in-situ observations. This makes establishing a "truth" dataset difficult, if not impossible. We can, however, compare the AMSR2-CloudSat combined retrieval (A2CS) to other independent and well-validated retrievals. Here we present comparisons with Remote Sensing Systems (RSS) (Wentz et al. 2014), since we can directly compare A2CS to RSS AMSR2 retrievals of SST, water vapor, and surface wind speed for clear and cloudy scenes. Additionally, a relatively small number of coincident CloudSat overpasses for the period of study shows that the algorithm validates well against buoy and ship observations.

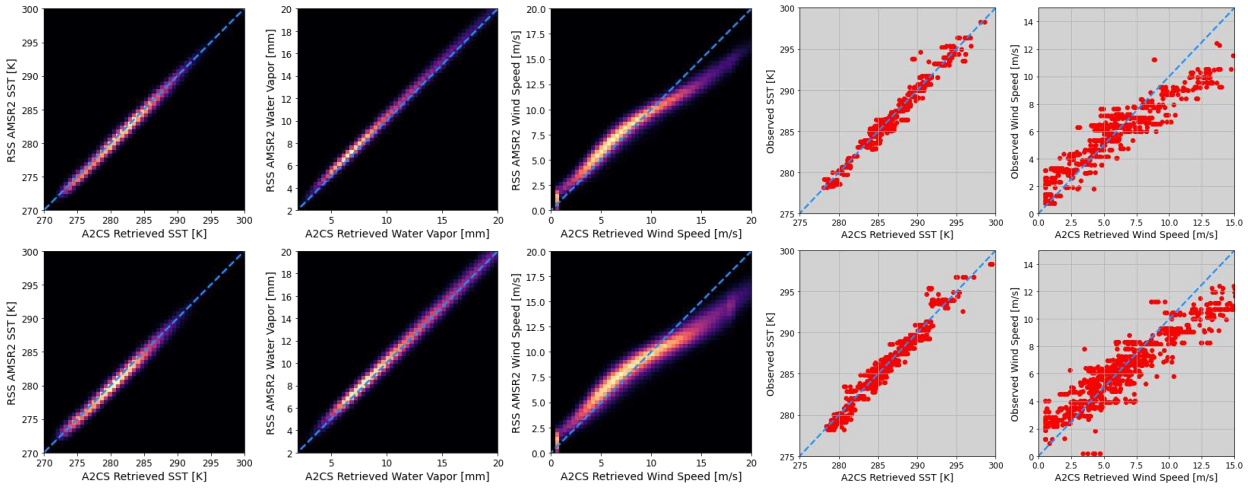

Figure 7: A2CS retrievals of surface parameters compared with RSS AMSR2 (left six panels) retrievals and with buoy and ship observations (right four panels). Top row shows clear scene retrievals and bottom row shows cloudy scene retrievals.

Initial comparisons showed excellent correlation but relatively large biases in surface variables when compared with in-situ observations and RSS. We attributed this to biases in the forward model simulations of AMSR2 Tbs causing the retrieval to converge on biased states. This was addressed by using offsets applied to the observed Tb vector before its ingestion into the algorithm. This algorithm-specific calibration is sometimes used in operational PMW ocean surface retrievals, including RSS (Gentemann and Hilburn 2015), to improve validation against in-situ data. To get sufficient sampling, again a challenge in this domain, we estimated the Tb offsets by using an independent sensor, GMI, and by comparing the retrieved surface variables from the A2CS algorithm in clear scenes to RSS GMI (Wentz et al. 2015). Using an independent, well-calibrated radiometer helps ensure that the forward model biases can be isolated from the sensor information content, and tuning the offsets to compare well against RSS ensures good comparisons with in-situ observations with better sampling than using in-situ data alone. The offsets are manually adjusted until the retrieved surface variables are unbiased with respect to RSS GMI, indicating that the forward model bias has been correctly accounted for. These same offsets calculated from RSS GMI are applied to the AMSR2 channels that correspond to GMI channels, and the offset for the 23.8 GHz horizontally polarized channel (23H) is adjusted until retrievals of integrated water vapor aligns well with RSS AMSR2. The implementation of these Tb offsets, relatively small for all channels except for 23H (for which there is no GMI analog), resulted in much better comparisons with in-situ surface observations and with RSS water vapor both in clear and cloudy scenes.

Some consistent biases in wind speed seem to remain, however. These are most likely due to differing surface emissivity models between A2CS and RSS AMSR2. At higher wind speeds (greater than 7.5 m s^{-1}), the increasing departure from one-to-one correlation means that

higher wind speeds than are observed are needed by the forward model to simulate the correct surface emissivity due to the likely presence of ocean surface foam in the scene. The result is that any retrieved wind speed above 7.5 m s^{-1} must be treated as an effective wind speed that is nearly always higher than the physical wind speed at the surface. Given the focus here on precipitation, however, the surface wind speed bias should not be overly problematic given that the Tbs converge appropriately.

3.2 Improved constraints

The clear-air retrievals of SST, water vapor, cloud water, and surface wind speed with PMW observations are known to be robust (Gentemann et al. 2010). The following are some selected cases of cloudy retrievals to show the effect of having radar observations to constrain the vertical distribution of cloud particles and hydrometeors. In radiometer-only retrievals, the vertical distribution of cloud water must be assumed, since the PMW observations are only sensitive to the integrated absorption and emission throughout the atmospheric column. PMW retrievals of total liquid and ice water path tend to converge well until the particle sizes move well into the larger precipitation size regime and begin to scatter the microwave radiation, causing the solutions to become non-unique (Weng 2007). Normally, this signal manifests itself in an inflation of chi squared, indicating the algorithm is unable to find convergence on the scene. Some operational PMW retrieval algorithms, such as NOAA's Microwave Integrated Retrieval System (MiRS) (Boukabara et al. 2011), use the signal in chi squared to flag the scene for possible precipitation. This approach works well if the hydrometeor particle sizes are well into the scattering regime for the wavelengths of interest, but for drizzle-sized particles, it is difficult to rely on this signal to discriminate precipitation from clouds. Retrieved liquid water

paths associated with relatively large probabilities of precipitation often occur with chi squared values below the convergence limit as illustrated in Figure 8.

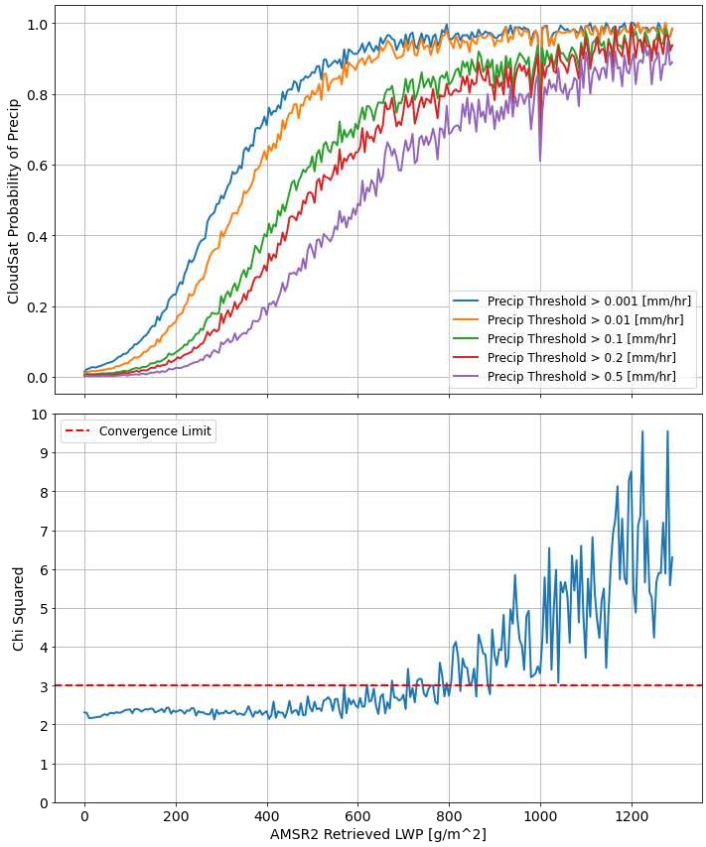

Figure 8: CloudSat along-track averaged probability of precipitation using several precipitation thresholds and final chi squared values from AMSR2 PMW retrievals. Convergence limit is taken to be the MiRS convergence threshold of 3.0.

By constraining the solution with coincident radar observations and estimating the PSD parameters, the scattering in the forward model can be prescribed to the correct vertical levels. Figure 9 shows the evolution of the simulated observation vector F(**x**,**b**) throughout the retrieval iterations from three examples of cloudy retrievals. The rows show three different observed cloud scenarios: a shallow ice cloud, a shallow liquid cloud, and a mixed phase cloud. The columns represent different types of retrieval: from left to right, these are the combined retrieval, a retrieval using only the AMSR2 Tbs, and a retrieval using only the CPR reflectivity profile.

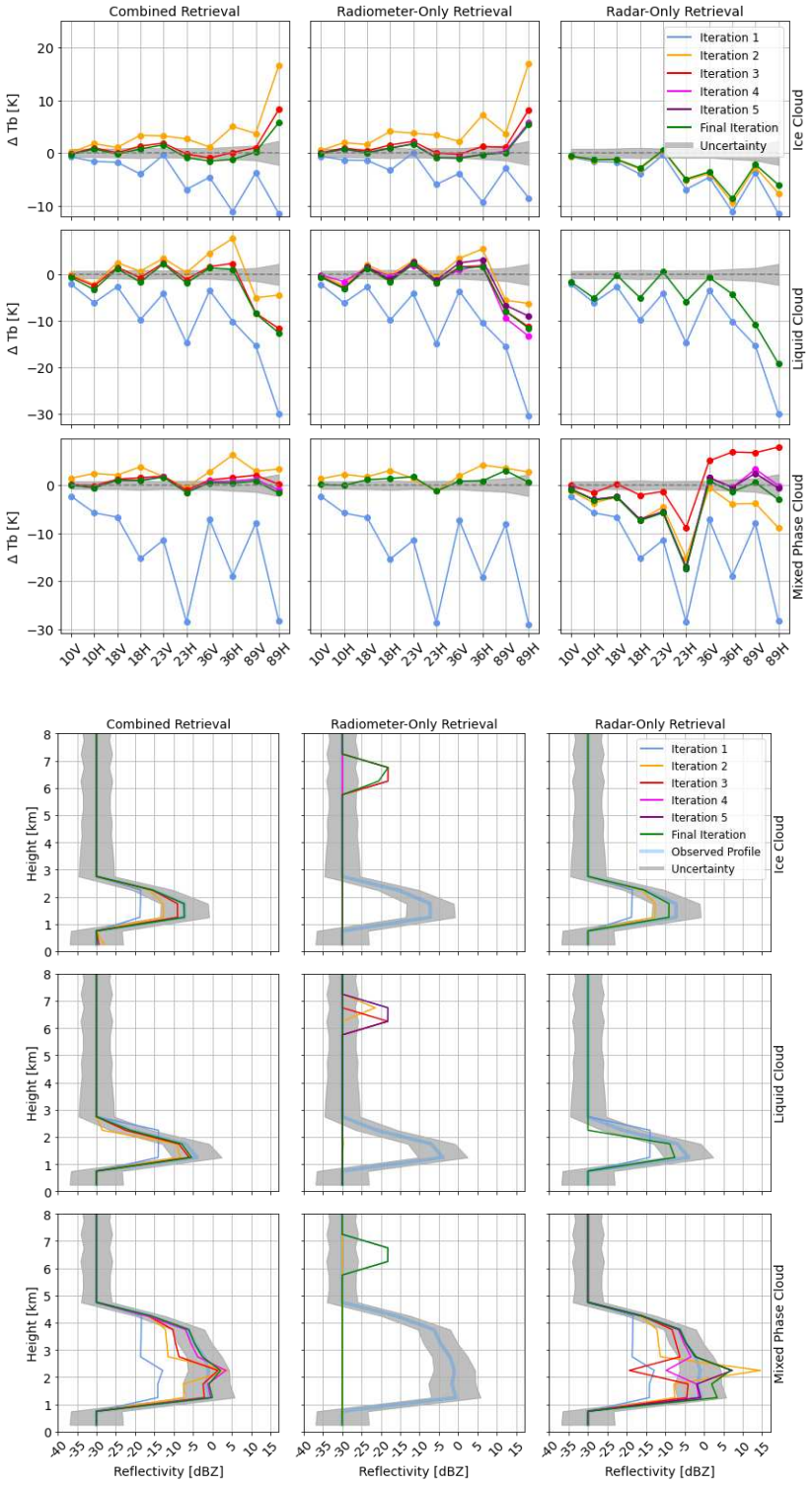

Figure 9: Evolution of the simulated observation vector F(x) through three cloudy retrieval cases (rows) with three different retrieval methods (columns). Top nine panels show residual Tb, or difference between simulated Tb and observed Tb throughout the retrieval. Lower nine panels show simulated and observed radar profile. Uncertainties are diagonal elements of $\mathbf{S_y}$ matrix and are plotted in terms of standard deviation to aid in interpretation.

The combined retrieval (left column) shows the best fit between the observation vector and the simulated observations from the retrieved state in all three cases. Here we emulate a typical radiometer-only retrieval (middle column) by making the necessary *a priori* assumptions, namely the vertical distribution of ice and liquid water and their PSD parameters. We assume an ice cloud layer that is between 6 and 7 km and a fixed ice particle density, allowing the ice water path only to be retrievable. Similarly, cloud water is assumed monodisperse and distributed between the freezing level and the cloud base, both taken *a priori* from ERA5. For the radar-only retrieval (right column), there is no signal in SST, wind speed, or water vapor (except for attenuation, which alone is poorly constrained in the presence of cloud water), and these are also taken as fixed ancillary data. The ice particle density is again assumed fixed, as well as the gamma PSD shape parameter μ, in order to avoid instabilities in the algorithm due to too many free parameters. For the radiometer-only retrieval, the radar observational uncertainties in $\mathbf{S_y}$ are set to be very large and uncorrelated, which allows the algorithm to ignore the radar observations, and vice versa.

In all three cases, we see the best fit between the final iteration and the observations from the combined retrieval. The relatively larger differences between the simulated and observed Tbs at 89 GHz are due to the higher uncertainties in these channels produced by ice scattering. For both the ice cloud and the mixed phase cloud cases, the PMW observations are sensitive to the presence of ice scattering, but not to the vertical location of cloud ice. This is illustrated by the fact that the radiometer-only retrieved state produces nearly identical Tbs to those from the retrieved state in the combined retrieval. In the liquid cloud case (second row), the PMW observations correctly show no ice scattering in the scene, as the *a priori* ice cloud is removed by the algorithm to get convergence once the correct amount of absorption by the monodisperse

cloud water is achieved. In this case, sufficient absorption by the monodisperse cloud water was enough to lead to convergence, even though this cloud water had no appreciable backscatter signal to produce simulated radar reflectivities above the noise floor (-26 dBZ). In all cases, the radar-only final retrieved state shows significant differences in the simulated Tbs that are well outside their uncertainty estimates. An extreme example is the nearly 20-K difference between the simulated and observed Tb for the 23H in the mixed phase cloud case. This difference does not exist in either the combined or radiometer-only retrieval due to water vapor being constrained by the observations.

A noticeable degradation of the fit of the radar profile is seen in the radar-only retrieval in all three cases. This is most likely due to the fact that the error covariance matrices are optimized for using both the Tbs and the radar reflectivities together. A radar-only algorithm might, for instance, use smaller uncertainties for the *a priori* data, assuming the first guess is good, or even use a lower convergence threshold, essentially forcing more iterations until the radar profile is fit nearly exactly.

Statistically, we see much better correlation between observed Tbs and final simulated Tbs from retrieved profiles when we apply the constraint of the radiometer observations in Figure 10. The remaining biases, though consistent, are again within the observational and forward model uncertainty estimations for these channels. Additionally, 25% of cloudy scenes that had converged in the non-precipitating radiometer-only retrieval were found to be precipitating due to the constraint of the CPR profile on the PMW observations.

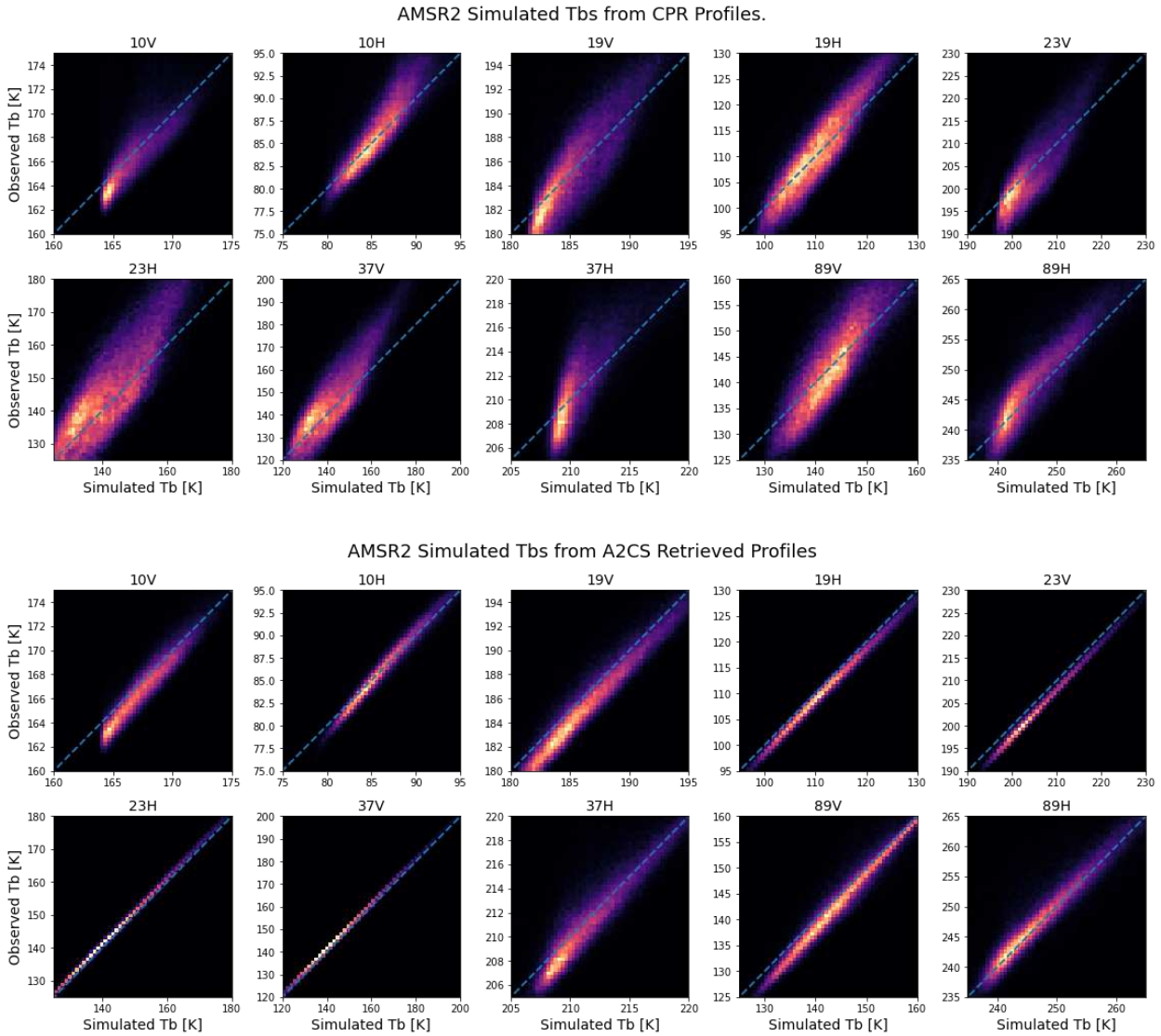

Figure 10: 2D histograms of simulated AMSR2 channel Tbs compared with observed Tbs using the radar-only retrieved CPR profile and ancillary data (top two rows) and the combined retrieved profile (bottom two rows) for one month of coincident cloudy observations. Color from dark to light indicates the number of profiles.

3.3 Revisiting zonal means of drizzle

Accumulating retrieved precipitation estimations from the A2CS retrieval shows general agreement with CloudSat precipitation estimates. Figure 11 shows latitudinal means of three CloudSat precipitation algorithms along with two climatology products, ERA5 and GPROF GMI, throughout the period of study for low precipitation rates (below 0.25 mm hr^{-1}). We see a

large gap between ERA5 and any of the CloudSat estimates, with CloudSat being the only sensor shown that is reliably capable of directly observing drizzle. Some of the drizzle in GPROF GMI is a mathematical artifact caused by the averaging of convective precipitation as observed by the DPR at the radar resolution into the much larger GMI footprint, and the rest of the drizzle over oceans is added by blending MiRS retrievals from GMI observations with DPR observations within the *a priori* GPROF database (Pfreundschuh et al. 2024). This brings GPROF estimations of drizzle into good agreement with ERA5 but far from CloudSat.

As latitudes get higher, the increasing reliance on snow retrievals as the dominant form of precipitation estimation becomes apparent as the accumulations of liquid surface rain get smaller and A2CS begins to follow the CloudSat 2C-SNOW-PROFILE accumulations. This trend is apparent in both hemispheres. Where liquid precipitation becomes more important towards the midlatitudes, the zonal means of drizzle follow the approximate sum of the means of CloudSat 2C-RAIN-PROFILE and 2C-SNOW-PROFILE. The agreement between A2CS and CloudSat operational algorithms is a surprising result as there is no constraint on the algorithm to produce intensity estimates that align with CloudSat, for CloudSat estimates are used on a first-order basis for precipitation detection only in the A2CS algorithm. This agreement also increases confidence that the A2CS drizzle estimates are not spurious retrieval artifacts as in GPROF.

However, it is important to point out the small, but consistent, overestimation with respect to CloudSat, especially in the Southern Ocean at latitudes south of -55°. This small increase of 7.6% in zonal means is likely a signal of the radiometer's tendency to correct the radar's intensity estimates by constraining the total water path and integrated water vapor. It is also important to note the stark departure of A2CS drizzle from 2C-PRECIP-COLUMN accumulations, which generally follow, but overestimate, those of 2C-RAIN-PROFILE. An

overestimation by A2CS of 19.3% was found compared to 2C-PRECIP. Currently, CloudSat's operational radar algorithms rely on ancillary data (water vapor profiles from ECMWF) to constrain the separation of radar beam attenuation due to water vapor from that due to cloud water and hydrometeors. Having the PMW observations to constrain this separation in A2CS results in an overall higher accumulation of drizzle compared to CloudSat, indicating the possibility that using ECMWF water vapor alone leads to an underestimation bias in CloudSat's radar-only derived drizzle estimates in the high-latitude oceans.

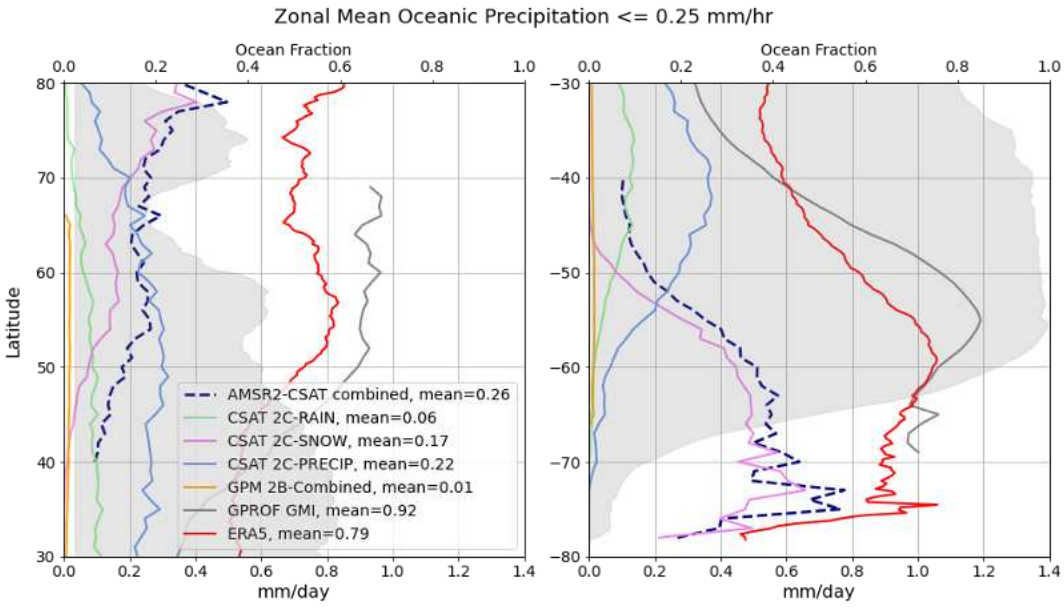

Figure 11: A reproduction of the bottom two panels of Figure 1, with the A2CS drizzle estimates added (dashed navy blue). Figure legend contains total latitude-weighted mean precipitation, in mm day^{-1}, within the domain of study. Period of study is calendar year 2015.

CHAPTER 4: SUMMARY AND DISCUSSION

We used an optimal estimation scheme to constrain observations of light drizzle over the high latitude oceans (±40° latitude and poleward) using a combination of radar and radiometer observations. Collocating the radiometer observations from AMSR2 and the radar observations from the CPR is challenging since the measurements are representative of very different spatial scales. This problem is addressed statistically by decreasing the horizontal and vertical resolution of the CPR observations. The OE algorithm successfully forces the solution to be radiatively consistent with both the CPR reflectivity profile and the AMSR2 Tbs. This is important since precipitation onset cannot be determined case-by-case with PMW observations alone, nor can a CPR profile alone produce the observed Tbs for a non-clear scene. The inverted state from the combined retrieval includes the distribution of cloud particles and hydrometeors, resulting in a retrieved precipitation rate that is consistent with both sets of observations. Specification of a two-parameter PSD is possible in both liquid and frozen cloud and hydrometeor retrievals due to the additional constraints provided by both the active and passive measurements.

We also find that relying primarily on the observational uncertainties as the primary driver of the total algorithm uncertainties produces physically-sound retrievals on their own for this regime and at these low precipitation rates. Failed convergence is most likely an indicator of heavy scattering and seems to be associated with very high precipitation estimates from CloudSat that are well beyond the saturation threshold of the radar beam. These scenes, while not retrievable by the A2CS algorithm, theoretically could be retrievable by a radar with an operating wavelength that is less prone to total attenuation, but this is outside the scope of the current study. This fact does, however, make the case for future missions with multiple-frequency

spaceborne coincident radars to correctly capture all possible precipitation intensities via additional constraints to the radiometer observations.

The A2CS algorithm is able to successfully constrain the discrimination between drizzling and nonprecipitating clouds, producing zonal means of light precipitation rates that are not only consistent with both observations, but also within the bounds of estimations by current operational satellite products. Using CloudSat native estimates for frequency of occurrence only, we see good agreement between the A2CS algorithm and CloudSat combined rain and snow estimates for drizzle. A slight overestimation of 7.6% with respect to the combined means of 2C-RAIN and 2C-SNOW, as well as the much larger 19.3% increase with respect to the mean of 2C-PRECIP, however, likely shows the correction that the radiometer observations apply to the radar's intensity estimates in the bulk statistics. This general agreement with CloudSat increases confidence in retrieval skill, although validation in this meteorological regime remains difficult.

One line of future study has already been mentioned. Incorporation of A2CS retrieved profiles into the GPROF *a priori* database would greatly improve GPROF drizzle estimations for all PMW sensors. Since radiometric consistency with cross-calibrated PMW observations are already established for these drizzling profiles, and since the algorithm-specific calibration was originally estimated using GMI, it can be speculated that the merger of this CloudSat-informed drizzle regime and the existing GPM-CMB database is not only possible, but practical. Of further value is the relatively small, but significant, number of GPM and CloudSat/AMSR2 coincidences for validation. The resolution and sampling differences between the nadir-only CPR and the phased-array DPR are a non-trivial obstacle, but data from coincident overpasses may provide some insight into how to proceed with the transition between drizzling and raining regimes. The smooth transition in precipitation estimates from the W-band to the Ka-band to the Ku-band

might be replicated with similarly gradual transitions in Tb space between AMSR2 and GMI in these profiles.

Possible improvements to the A2CS algorithm that were outside the scope of this study will also be addressed in future work. The deficiencies in high wind speed retrievals seen in Figure 7 are to be addressed by either using a post-retrieval correction for the surface emissivity model, which validation data seem to indicate can possibly be done with a simple linear correction function, or by simply using an emissivity model that validates better with in-situ data at high wind speeds, such as the emissivity model used by RSS. Improvements in forward-model ice scattering parameter estimations could also be possible with the incorporation of non-spherical ice particle scattering tables, such as Liu (Liu 2008b). These current findings, however, are encouraging, and future investigations of drizzle observation using combined techniques could likely provide valuable insight into the nature of this sensitive precipitation mode and its implications in a warming climate.

REFERENCES

Abel, S. J., and I. A. Boutle, 2012: An improved representation of the raindrop size distribution for single-moment microphysics schemes. *Quarterly Journal of the Royal Meteorological Society*, **138**, 2151–2162, https://doi.org/10.1002/qj.1949.

Adler, R. F., G. Gu, M. Sapiano, J.-J. Wang, and G. J. Huffman, 2017: Global Precipitation: Means, Variations, and Trends During the Satellite Era (1979-2014). *Surveys in Geophysics*, **38**, 679–699, https://doi.org/10.1007/s10712-017-9416-4.

Ban, N., J. Schmidli, and C. Schär, 2014: Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations. *Journal of Geophysical Research: Atmospheres*, **119**, 7889–7907, https://doi.org/10.1002/2014JD021478.

Behrangi, A., M. Lebsock, S. Wong, and B. Lambrigtsen, 2012: On the quantification of oceanic rainfall using spaceborne sensors. *Journal of Geophysical Research*, **117**, https://doi.org/10.1029/2012JD017979.

——, G. Stephens, R. F. Adler, G. J. Huffman, B. Lambrigtsen, and M. Lebsock, 2014: An Update on the Oceanic Precipitation Rate and Its Zonal Distribution in Light of Advanced Observations from Space. *Journal of Climate*, **27**, 3957–3965, https://doi.org/10.1175/JCLI-D-13-00679.1.

Berg, W., T. L'Ecuyer, and C. Kummerow, 2006: Rainfall Climate Regimes: The Relationship of Regional TRMM Rainfall Biases to the Environment. *Journal of Applied Meteorology and Climatology*, **45**, 434–454, https://doi.org/10.1175/JAM2331.1.

——, and Coauthors, 2016: Intercalibration of the GPM Microwave Radiometer Constellation. *Journal of Atmospheric and Oceanic Technology*, **33**, 2639–2654, https://doi.org/10.1175/JTECH-D-16-0100.1.

Boukabara, S.-A., and Coauthors, 2011: MiRS: An All-Weather 1DVAR Satellite Data Assimilation and Retrieval System. *IEEE Transactions on Geoscience and Remote Sensing*, **49**, 3249–3272, https://doi.org/10.1109/TGRS.2011.2158438.

Brath, M., R. Ekelund, P. Eriksson, O. Lemke, and S. A. Buehler, 2020: Microwave and submillimeter wave scattering of oriented ice particles. *Atmospheric Measurement Techniques*, **13**, 2309–2333, https://doi.org/10.5194/amt-13-2309-2020.

Duncan, D. I., and C. D. Kummerow, 2016: A 1DVAR retrieval applied to GMI: Algorithm description, validation, and sensitivities. *Journal of Geophysical Research: Atmospheres*, **121**, 7415–7429, https://doi.org/10.1002/2016JD024808.

Gentemann, C. L., and K. A. Hilburn, 2015: In situ validation of sea surface temperatures from the GCOM-W1 AMSR2 RSS calibrated brightness temperatures. *Journal of Geophysical Research: Oceans*, **120**, 3567–3585, https://doi.org/10.1002/2014JC010574.

Gentemann, C. L., F. J. Wentz, M. Brewer, K. Hilburn, and D. Smith, 2010: Passive Microwave Remote Sensing of the Ocean: An Overview. *Oceanography from Space*, Springer.

Grecu, M., W. S. Olson, S. J. Munchak, S. Ringerud, L. Liao, Z. Haddad, and B. L. Kelley, 2016: The GPM Combined Algorithm. *Journal of Atmospheric and Oceanic Technology*, **33**, 2225–2245, https://doi.org/10.1175/JTECH-D-16-0019.1.

Haynes, J. M., 2018: CloudSat 2C-PRECIP-COLUMN Data Product Process Description and Interface Control Document, Product Version: P1 R05.

——, and G. L. Stephens, 2007: Tropical oceanic cloudiness and the incidence of precipitation: Early results from CloudSat. *Geophysical Research Letters*, **34**, https://doi.org/10.1029/2007GL029335.

——, R. T. Marchand, Z. Luo, A. Bodas-Salcedo, and G. L. Stephens, 2007: A Multipurpose Radar Simulation Package: QuickBeam. *Bulletin of the American Meteorological Society*, **88**, 1723–1728, https://doi.org/10.1175/BAMS-88-11-1723.

——, T. S. L'Ecuyer, G. L. Stephens, S. D. Miller, C. Mitrescu, N. B. Wood, and S. Tanelli, 2009: Rainfall retrieval over the ocean with spaceborne W-band radar. *Journal of Geophysical Research: Atmospheres*, **114**, D00A22, https://doi.org/10.1029/2008JD009973.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Heymsfield, A. J., P. Field, and A. Bansemer, 2008: Exponential Size Distributions for Snow. *Journal of the Atmospheric Sciences*, **65**, 4017–4031, https://doi.org/10.1175/2008JAS2583.1.

Hou, A. Y., and Coauthors, 2014: The Global Precipitation Measurement Mission. *Bulletin of the American Meteorological Society*, **95**, 701–722, https://doi.org/10.1175/BAMS-D-13-00164.1.

Iguchi, T., and Coauthors, 2021: GPM/DPR Level-2 Algorithm Theoretical Basis Document.

Kilic, L., C. Prigent, C. Jimenez, and C. Donlon, 2021: Technical note: A sensitivity analysis from 1 to 40 GHz for observing the Arctic Ocean with the Copernicus Imaging Microwave Radiometer. *Ocean Science*, **17**, 455–461, https://doi.org/10.5194/os-17-455-2021.

Klepp, C., and Coauthors, 2018: OceanRAIN, a new in-situ shipboard global ocean surface-reference dataset of all water cycle components. *Scientific Data*, **5**, https://doi.org/10.1038/sdata.2018.122.

Kummerow, C., 1998: Beamfilling Errors in Passive Microwave Rainfall Retrievals. *Journal of Applied Meteorology and Climatology*, **37**, 356–370, https://doi.org/10.1175/1520-0450(1998)037<0356:BEIPMR>2.0.CO;2.

Kummerow, C. D., D. L. Randel, M. Kulie, N.-Y. Wang, R. Ferraro, S. J. Munchak, and V. Petković, 2015: The Evolution of the Goddard Profiling Algorithm to a Fully Parametric Scheme. *Journal of Atmospheric and Oceanic Technology*, **32**, 2265–2280, https://doi.org/10.1175/JTECH-D-15-0039.1.

Kuo, K.-S., and Coauthors, 2016: The Microwave Radiative Properties of Falling Snow Derived from Nonspherical Ice Particle Models. Part I: An Extensive Database of Simulated Pristine Crystals and Aggregate Particles, and Their Scattering Properties. *Journal of Applied Meteorology and Climatology*, **55**, 691–708, https://doi.org/10.1175/JAMC-D-15-0130.1.

Lamer, K., P. Kollias, A. Battaglia, and S. Preval, 2020: Mind the gap – Part 1: Accurately locating warm marine boundary layer clouds and precipitation using spaceborne radars. *Atmospheric Measurement Techniques*, **13**, 2363–2379, https://doi.org/10.5194/amt-13-2363-2020.

Lavers, D. A., A. Simmons, F. Vamborg, and M. J. Rodwell, 2022: An evaluation of ERA5 precipitation for climate monitoring. *Quarterly Journal of the Royal Meteorological Society*, **148**, 3152–3165, https://doi.org/10.1002/qj.4351.

Lebsock, M. D., 2018: Level 2C RAIN-PROFILE Product Process Description and Interface Control Document.

——, and T. S. L'Ecuyer, 2011: The retrieval of warm rain from CloudSat. *Journal of Geophysical Research: Atmospheres*, **116**, https://doi.org/10.1029/2011JD016076.

L'Ecuyer, T. S., and G. L. Stephens, 2002: An Estimation-Based Precipitation Retrieval Algorithm for Attenuating Radars. *Journal of Applied Meteorology and Climatology*, **41**, 272–285, https://doi.org/10.1175/1520-0450(2002)041<0272:AEBPRA>2.0.CO;2.

Liu, C., R. P. Allan, and G. J. Huffman, 2012: Co-variation of temperature and precipitation in CMIP5 models and satellite observations. *Geophysical Research Letters*, **39**, https://doi.org/10.1029/2012GL052093.

Liu, G., 2008a: Deriving snow cloud characteristics from CloudSat observations. *Journal of Geophysical Research: Atmospheres*, **113**, D00A09, https://doi.org/doi:10.1029/2007JD009766.

——, 2008b: A Database of Microwave Single-Scattering Properties for Nonspherical Ice Particles. *Bulletin of the American Meteorological Society*, **89**, 1563–1570, https://doi.org/10.1175/2008BAMS2486.1.

Marchand, R., G. G. Mace, T. Ackerman, and G. Stephens, 2008: Hydrometeor Detection Using Cloudsat—An Earth-Orbiting 94-GHz Cloud Radar. *Journal of Atmospheric and Oceanic Technology*, **25**, 519–533, https://doi.org/10.1175/2007JTECHA1006.1.

Matrosov, S. Y., 1998: A Dual-Wavelength Radar Method to Measure Snowfall Rate. *Journal of Applied Meteorology and Climatology*, **37**, 1510–1521, https://doi.org/10.1175/1520-0450(1998)037<1510:ADWRMT>2.0.CO;2.

Petković, V., P. Brown, W. Berg, D. L. Randel, S. R. Jones, and C. D. Kummerow, 2023: Can We Estimate the Uncertainty Level of Satellite Long-Term Precipitation Records? *Journal of Applied Meteorology and Climatology*, **62**, 1069–1082, https://doi.org/10.1175/JAMC-D-22-0179.1.

Petty, G. W., 1995: Frequencies and Characteristics of Global Oceanic Precipitation from Shipboard Present-Weather Reports. *Bulletin of the American Meteorological Society*, **76**, 1593–1616, https://doi.org/10.1175/1520-0477(1995)076<1593:FACOGO>2.0.CO;2.

Pfreundschuh, S., C. Guilloteau, P. J. Brown, C. D. Kummerow, and P. Eriksson, 2024: GPROF V7 and beyond: assessment of current and potential future versions of the GPROF passive microwave precipitation retrievals against ground radar measurements over the continental US and the Pacific Ocean. *Atmospheric Measurement Techniques*, **17**, 515–538, https://doi.org/10.5194/amt-17-515-2024, 2024.

Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax, 2007: Daily High-Resolution-Blended Analyses for Sea Surface Temperature. *Journal of Climate*, **20**, 5473–5496, https://doi.org/10.1175/2007JCLI1824.1.

Rodgers, C. D., 2000: *Inverse Methods For Atmospheric Sounding*. World Scientific Publishing Co. Pte. Ltd., 240 pp.

Schulte, R. M., and C. D. Kummerow, 2019: An Optimal Estimation Retrieval Algorithm for Microwave Humidity Sounding Channels with Minimal Scan Position Bias. *Journal of Atmospheric and Oceanic Technology*, **36**, 409–425, https://doi.org/10.1175/JTECH-D-18-0133.1.

——, and ——, 2022: Can DSD Assumptions Explain the Differences in Satellite Estimates of Warm Rain? *Journal of Atmospheric and Oceanic Technology*, **39**, 1889–1901, https://doi.org/10.1175/JTECH-D-22-0036.1.

——, ——, S. M. Saleeby, and G. G. Mace, 2023: How Accurately Can Warm Rain Realistically Be Retrieved with Satellite Sensors? Part II: Horizontal and Vertical Heterogeneities. *Journal of Applied Meteorology and Climatology*, **62**, 155–170, https://doi.org/10.1175/JAMC-D-22-0051.1.

Stephens, G. L., 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**, 1771–1790, https://doi.org/10.1175/BAMS-83-12-1771.

——, and C. D. Kummerow, 2007: The Remote Sensing of Clouds and Precipitation from Space: A Review. *Journal of the Atmospheric Sciences*, **64**, 3742–3765, https://doi.org/10.1175/2006JAS2375.1.

——, and Coauthors, 2010: Dreary state of precipitation in global models. *Journal of Geophysical Research: Atmospheres*, **115**, D24211, https://doi.org/10.1029/2010JD014532.

Testud, J., S. Oury, R. A. Black, P. Amayenc, and X. Dou, 2001: The Concept of "Normalized" Distribution to Describe Raindrop Spectra: A Tool for Cloud Physics and Cloud Remote Sensing. *Journal of Applied Meteorology and Climatology*, **40**, 1118–1140, https://doi.org/10.1175/1520-0450(2001)040<1118:TCONDT>2.0.CO;2.

Ulbrich, C. W., 1983: Natural Variations in the Analytical Form of the Raindrop Size Distribution. *Journal of Climate and Applied Meteorology*, **22**, 1764–1775.

Villermaux, E., and F. Eloi, 2011: The distribution of raindrops speeds. *Geophysical Research Letters*, **38**, https://doi.org/10.1029/2011GL048863.

Weng, F., 2007: Advances in Radiative Transfer Modeling in Support of Satellite Data Assimilation. *Journal of the Atmospheric Sciences*, **64**, 3799–3807, https://doi.org/10.1175/2007JAS2112.1.

Wentz, F. J., T. Meissner, C. Gentemann, K. A. Hilburn, and J. Scott, 2014: Remote Sensing Systems Daily Environmental Suite on 0.25 deg grid, Version 8.2. www.remss.com/missions/amsr. (Accessed December 21, 2023).

——, ——, J. Scott, and K. A. Hilburn, 2015: Remote Sensing Systems GPM GMI Daily Environmental Suite on 0.25 deg grid, Version 8.2. *Remote Sensing Systems, Santa Rosa, CA.*,.

Wood, N. B., and T. S. L'Ecuyer, 2018: Level 2C Snow Profile Process Description and Interface Control Document, Product Version: P1 R05.

World Meteorological Organization, 2022: OSCAR: Observing Systems Capability Analysis and Review Tool: AMSR2. https://space.oscar.wmo.int/instruments/view/amsr2 (Accessed December 5, 2023).

2016: Data Users' Manual for the Advanced Microwave Scanning Radiometer 2 (AMSR2) onboard the Global Change Observation Mission 1st - Water "SHIZUKU" (GCOM-W1).

APPENDIX: ALGORITHM CONVERGENCE STATISTICS

Table 3: Algorithm convergence statistics: lists retrieval runtime statistics for clear, cloudy, precipitating scenes, and overall during period of study.

	Clear	Cloudy	Precipitating	Overall
Successful	721 551	1 118 453	373 811	2 213 815
Failed	36 304	153 306	59 475	249 085
Total	757 855	1 271 759	433 286	2 462 900
Success rate	95.2%	87.9%	86.3%	89.9%
Failure rate	4.8%	12.1%	13.7%	10.1%
Average Number of Iterations	2.73	3.81	5.99	3.97
Average δ^2	0.32	0.55	0.62	0.49
Average CPU time per orbit				124.5 sec