DISSERTATION

EXPLORING THE LIMITS OF VARIATIONAL PASSIVE MICROWAVE RETRIEVALS

Submitted by

David Ian Duncan

Department of Atmospheric Science

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2017

Doctoral Committee:

Advisor: Christian D. Kummerow

Sid-Ahmed Boukabara
Christopher W. O'Dell
Steven C. Reising
Steven A. Rutledge
Russ S. Schumacher

ABSTRACT

EXPLORING THE LIMITS OF VARIATIONAL PASSIVE MICROWAVE RETRIEVALS

Passive microwave observations from satellite platforms constitute one of the most important data records of the global observing system. Operational since the late 1970s, passive microwave data underpin climate records of precipitation, sea ice extent, water vapor, and more, and contribute significantly to numerical weather prediction via data assimilation. Detailed understanding of the observation errors in these data is key to maximizing their utility for research and operational applications alike. However, the treatment of observation errors in this data record has been lacking and somewhat divergent when considering the retrieval and data assimilation communities. In this study, some limits of passive microwave imager data are considered in light of more holistic treatment of observation errors.

A variational retrieval, named the CSU 1DVAR, was developed for microwave imagers and applied to the GMI and AMSR2 sensors for ocean scenes. Via an innovative method to determine forward model error, this retrieval accounts for error covariances across all channels used in the iteration. This improves validation in more complex scenes such as high wind speed and persistently cloudy regimes. In addition, it validates on par with a benchmark dataset without any tuning to in-situ observations. The algorithm yields full posterior error diagnostics and its physical forward model is applicable to other sensors, pending intercalibration.

This retrieval is used to explore the viability of retrieving parameters at the limits of the available information content from a typical microwave imager. Retrieval of warm rain, marginal sea ice, and falling snow are explored with the variational retrieval. Warm rain retrieval shows

some promise, with greater sensitivity than operational GPM algorithms due to leveraging CloudSat data and accounting for drop size distribution variability. Marginal sea ice is also detected with greater sensitivity than a standard operational retrieval. These studies ultimately show that while a variational algorithm maximizes the effective signal to noise ratio of these observations, hard limitations exist due to the finite information content afforded by a typical microwave imager.

ACKNOWLEDGMENTS

At the culmination of a long educational journey, I am humbled to reflect back on the numerous people and institutions who have supported, encouraged, and challenged me along the way. My parents have been tireless supporters of my education, in ways great and small, all my life. This is dedicated to them and the public education institutions that I have been fortunate enough to attend, from Pennsylvania to Illinois to Colorado.

My advisor, Professor Chris Kummerow, deserves special gratitude. His advising allowed me great freedom in much of the research that went into this dissertation, but he was always available if I needed guidance or a good argument to break up the day. It is hard to imagine an advising style that could have fit me better, and I am grateful to Chris for all the opportunities he has provided me.

I extend thanks to the other members of my doctoral committee and the valuable feedback they have given in meetings both formal and informal: Sid Boukabara, Chris O'Dell, Steve Reising, Steve Rutledge, and Russ Schumacher.

The Kummerow research group—past and present—has been like a second family to me in Fort Collins. It has been helpful in too many ways to count, from coding support to bike rides to infusions of sugar, and I am very thankful to have been a part of this group. Special thanks are extended to Wes Berg, Dave Randel, and Veljko Petkovic for technical assistance and discussions that contributed to this work. Kummerow group alumni Greg Elsaesser and Joe Munchak also contributed helpful discussions for which I am grateful.

Similarly, I would like to broadly thank the Department of Atmospheric Science at Colorado State University. It is an esteemed department that has provided a welcoming

atmosphere conducive to fostering research, and it is an honor to be a part of it. Thanks to the departmental staff for travel and administrative help, the faculty, and my fellow students and colleagues for all that they have taught me.

This work was supported by a grant from the National Aeronautics and Space Administration, NNX13AN44G. I am very grateful for the public support of science that is a hallmark of the USA.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Passive microwave inversion

Space-borne passive microwave remote sensing of planet Earth has a history that goes back to the 1960s and has been recognized as an operational priority since the late 1970s (Staelin 1969, Njoku 1982). The development of passive microwave sensors has been bifurcated since the beginning, focusing either on frequencies near oxygen and water vapor absorption bands (Staelin et al. 1975), or frequencies in window parts of the spectrum that see the surface of the Earth (Wilheit 1977); these sensors became known as microwave sounders and imagers, respectively. Microwave sounders and imagers have become an indispensable part of the Earth observing system for climate monitoring (Mears et al. 2003), numerical weather prediction (Andersson et al. 1994), and research.

To make use of radiometric observations of the Earth, radiances are inverted to determine characteristics of the atmosphere or Earth's surface that are of interest. Frequencies selected for Earth observation are of course chosen because of their sensitivity to specific parameters. However, at no frequency does a microwave radiance have a perfect one-to-one relationship with any one geophysical parameter. This interdependency of microwave radiances and geophysical parameters makes the inversion problem challenging. In addition, nearby frequencies can have similar behavior with respect to surface or atmospheric constituents. Thus both the radiometric characteristics of retrieved parameters and the radiances themselves can be correlated.

Inversion methods have evolved over the decades as inversion theory, spectroscopy, and computational power have progressed. Staelin (1977) noted that "the [passive microwave] inversion problem is adequately approximated as linear with jointly Gaussian statistics, and, thus, a linear retrieval performs well." However, limitations both computational and relating to radiative transfer modeling caused early retrieval algorithms to rely on statistical relationships or regressions between microwave brightness temperatures (T_B) and geophysical parameters such as rainfall (Wilheit et al. 1977, Spencer et al. 1989) or sea ice (Cavalieri et al. 1984) to make the problem manageable. Most such techniques attempted to minimize the impact of atmospheric effects to better sense the surface parameters, and vice versa for atmospheric parameters. Ultimately, early retrievals had a singular focus on retrieval of one parameter and tried to ignore other radiometrically significant constituents via channel differencing or regressions. It is argued in Chapter 4 that this vestige of decades past still exists for a reason; however, a more holistic retrieval approach has many benefits.

The first problem that demonstrated a need for more codified treatment was retrieval of atmospheric profiles of temperature and water vapor from microwave and infrared sounders (Rodgers 1976). The Microwave Sounding Unit (MSU), for instance, had four channels sensitive to emission from oxygen with weighting functions that peaked at different altitudes. It would be folly to treat different parts of the profile as independent, as profiles must be continuous with some autocorrelation length. Thus the channels were used together, fitting the observations to known atmospheric profile shapes, while taking into account prior knowledge of atmospheric profiles as well as instrument errors from MSU itself. By assuming Bayesian statistics and Gaussian distributed errors for both the *a priori* and the observations, the variational technique

known as optimal estimation (OE) emerged as a method suited to this type of retrieval problem (Rodgers 2000).

The optimal estimation methodology balances observations and prior knowledge (known as the *a priori*) and iterates to find an optimal solution that balances the two by minimizing the cost function below:

$$\Phi = \big(y - f(x,b)\big)^T {S_y}^{-1}\big(y - f(x,b)\big) + (x - x_a)^T {S_a}^{-1}(x - x_a)$$

In this formulation y is the observation vector, x is the state vector, f is the forward model function, b represents all non-retrieved variables assumed in the forward model, x_a is the *a priori* vector, and $\mathbf{S_y}$ and $\mathbf{S_a}$ are error covariance matrices for the observation and *a priori* vectors, respectively. The maximum probability solution of x given measurement y is found where the gradient of Φ with respect to x is zero. This methodology is predicated on Gaussian-distributed parameters and errors, and thus works best for moderately linear and Gaussian inversion problems (Rodgers 2000).

The inversion is possible through Bayes' Theorem (Bayes and Price 1763), which relates the conditional probability of a state A given B to the probabilities of A, B, and the conditional probability of B given A:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

In the language of probability distribution functions (PDFs), the inversion uses the measurement PDF to update the *a priori* PDF and calculate a posterior PDF. For the satellite retrieval problem, the inversion takes knowledge of the distribution of T_Bs and a parameter of interest, and their conditional probability, to determine the conditional PDF of that parameter from the T_B vector observed.

1.2 Parallel advances in data assimilation

In parallel to the development of retrieval algorithms, many advances in numerical weather prediction (NWP) have come from the use of satellite and other data to anchor the analysis state, the initialization upon which NWP models run (English et al. 2000). NWP has progressed steadily in concert with the rise in capability of global satellite observations, though certainly other advances have contributed to an increase in global NWP skill (Bauer et al. 2015). These advances mirror the retrieval problem. The process of establishing an initial state for the model requires the assimilation of staggering amounts of data, now an established subfield known as data assimilation (DA). Not only are the bulk of these data from satellite platforms, but DA relies on the same mathematical framework as an atmospheric profile retrieval.

Both problems involve prior knowledge and observations, each with their own errors, through which an optimal estimate of the current state should be possible. The only real differences are the number of dimensions at play and the origin of the *a priori* state. Retrievals that employ variational techniques are usually still one-dimensional (1DVAR), whereas variational DA systems take all three spatial dimensions into account (3DVAR) and may even include so-called flow-dependence by also accounting for variations in time (4DVAR). Throughout this document, variational techniques and optimal estimation are terms used interchangeably.

Since the early 1990s, the data volume of satellite radiances ingested into operational DA systems for NWP initialization has increased by orders of magnitude (e.g. Reinecker et al. 2011). Passive microwave radiances have constituted a large fraction of the satellite data assimilated (Andersson et al. 1994). However, despite movement towards hyperspectral sensors, especially for infrared sensors, relatively few microwave frequencies are actually assimilated by operational

NWP centers (Geer et al. 2010, Lean et al. 2017). The examples cited are from the European Centre for Medium-range Weather Forecasting (ECMWF) and represent the cutting edge; American NWP centers assimilate fewer frequencies from microwave sensors, for example. Furthermore, in areas of cloud or precipitation—known as 'all-sky' to differentiate from clear-sky conditions—some of the frequencies that are assimilated use assumed errors so large that a lot of the information content is lost (Geer and Bauer 2011).

The main reason for not using available satellite data is that these unassimilated frequencies cannot be characterized well enough to have a positive impact on the assimilation. Put another way, the models cannot simulate certain microwave frequencies with enough fidelity to reliably add information to the analysis state. Assigning large error variances for some channels is a workaround method (Geer et al. 2010, Kazumori et al. 2015, Lean et al. 2017) but this greatly degrades the weight placed on the observations. Discarding these data or assuming very large errors is due to two factors: a lack of physical understanding to underpin the radiative transfer modeling, and an inability of the model to represent everything affecting the radiances.

1.3 Retrieval errors

In contrast to the DA community, the retrieval community has had the opposite problem—over-fitting to the observed radiances. This has happened for simple reasons. NWP models are penalized for adding bad information as that decreases forecast skill. Conversely, retrievals are judged on their ability to yield an answer but are not often penalized for a lack of error estimates for a given result, and so tend to use as much information as may be useful.

Most retrieval algorithms either assume no observation error or uncorrelated errors that are simply sensor noise values. Typically, regressions and semi-physically based algorithms

neglect observation error entirely (Alishouse et al. 1990, Aonashi et al. 2009, Comiso et al. 2003, Greenwald et al. 1993, Wentz 1997) while variational algorithms often assume uncorrelated sensor noise to be the only observation error because no better estimate is available (Deblonde and English 2003, Liu and Weng 2005, Boukabara et al. 2011, Munchak et al. 2016). Notable exceptions exist, such as the WindSat retrieval which included correlated errors (Bettenhausen et al. 2006), or Elsaesser and Kummerow (2008) which attempted to quantify observation error variances that were caused by forward model parameter error.

In theory, the addition of information will always improve a retrieval, but only if the errors of the measurements are fully understood. Running a radiative transfer model that has inherent errors, without quantifying those errors, will necessarily mean that the result is an over-fit to the observations. For example, consider a static bias for one channel, such as an emissivity model bias that causes simulated T_B to be too large for one polarization. Such a bias could cause high biases in retrieved wind speed and cloud water over ocean if unaccounted for. In fact, it's possible that the retrieval would fare better if run without that channel, depending on the severity of the bias. In contrast, if one channel was known to have large emissivity model errors, its inclusion would still add information and stability to a variational retrieval if the error is properly quantified. Even a sophisticated OE retrieval such as that described by Munchak et al. (2016) implicitly assumes a perfect forward model if it assumes uncorrelated sensor noise as the observation error and will thus over-fit to the observed radiances.

One of the goals of this document is to argue that proper treatment of observation errors—and their covariances—is not trivial and does in fact lead to better retrievals. This is touched on in Sections 2.5, 3.6, and 4.7. Representation of observation errors is necessary to maximize the information content of observations that is used, in essence getting the most out of

the past and present global observing system. The DA community is beginning to realize the importance of treating correlated errors as such (Weston et al. 2014), but treatment of correlated observation errors is still largely an area of research. It is similar in the retrieval community, where observation errors are often an afterthought if they are considered at all.

1.4 Information content limitations

One diagnostic from the variational framework is the averaging kernel, or A-matrix,essentially the retrieval's sensitivity to the true state (Rodgers 2000). This diagnostic sheds light on the information content available to the retrieval. The trace of this matrix represents the degrees of freedom for signal (DFS), a value that conveys the number of parameters that can be effectively retrieved. The non-raining retrieval described in Chapter 2 exhibits 3.5 to 5 DFS, depending mostly on the region and type of scene; the raining retrieval described in Chapter 3 exhibits around 3 to 4 DFS for raining pixels. With the caveat that DFS is dependent on the forward model, this is a sobering result—it dictates limits to the independent pieces of information that can be theoretically retrieved by a microwave imager with a channel complement typical of current sensors.

The retrieval paradigm for microwave imagers is thus a balancing act between variables for which there is signal and variables that impact radiances but cannot be accurately retrieved. This engenders a need to decrease the dimensionality of the retrieval problem. Though intelligent use of ancillary data can ameliorate this problem to some degree, it does not solve it. Principal component analysis can aid in this regard, such as collapsing profile variability to a few dimensions. This technique is common in the literature (e.g. Boukabara et al. 2011, Munchak et al. 2016) and leveraged in Chapters 2 and 3.

Retrieval of non-raining parameters over ocean, the focus of Chapter 2, is possible with a microwave imager channel complement and moderate use of ancillary data. Treatment of *a priori* and observation error covariances is shown to improve retrieval accuracy relative to earlier work in the literature (Wentz 1997, Elsaesser and Kummerow 2008), for which so-called 'crosstalk' issues are known to plague downstream data products (Seethala and Horvath 2010, Lebsock and Su 2014, Elsaesser et al. in review). In short, the microwave imager retrieval problem is tractable over ocean in non-scattering conditions, and a variational retrieval can perform well if adequately constrained.

With regard to retrieval of precipitation, the limits of information content from typical microwave imagers are well known. For example, Petty and Li (2013) collapsed a nine channel imager to three 'pseduochannels' for precipitation retrieval, implicitly recognizing that there are around 3 DFS available. Perhaps the most widely used passive precipitation retrieval, the Goddard Profiling algorithm (GPROF; Kummerow et al. 2015), is predicated upon the concept of limited information content. Instead of explicitly modeling the radiances to find the maximum probability solution, GPROF calculates a Bayesian average precipitation rate because of the multiplicity of solutions that can be contained within the sensor noise bounds of an imager's T_B vector. This is encapsulated in the first figure from Kummerow et al. (2011), reproduced here as Fig. 1.1, showing hydrometeor profiles that exhibit nearly identical top of atmosphere radiances but have different surface rain rates.

The problem of precipitation retrieval from passive microwave sensors is therefore underconstrained (Stephens and Kummerow 2007) and can even be described as "notoriously ill-constrained" (Boukabara et al. 2011). In most cases there are more parameters of interest than

DFS with which to retrieve them—a contrast to the non-scattering oceanic scenes mentioned above. Reasons for this are discussed in Chapter 3.

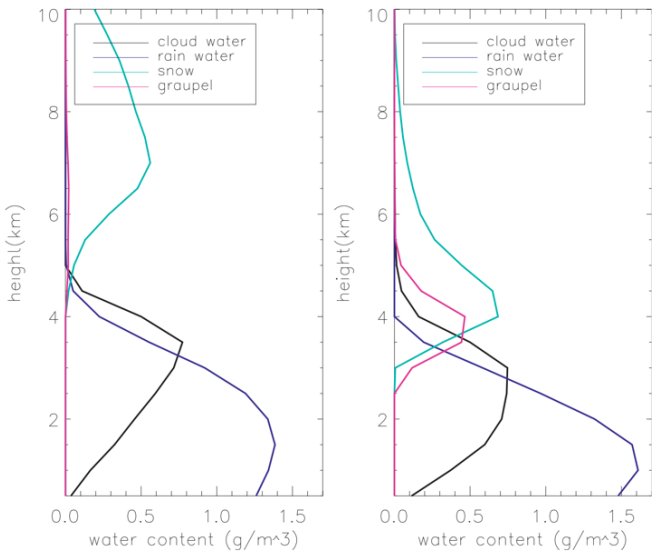

Figure 1.1 Simulated hydrometeor profiles from a cloud-resolving model with similar radiances at typical microwave imager frequencies but different surface rain rates. Reproduced from Kummerow et al. (2011).

The underconstrained nature of the problem is a great challenge for a variational retrieval, and indeed only warm rain retrieval is attempted in this work. However, there is motivation to attack this problem despite its inherent difficulty. Active sensors are limited in their spatiotemporal sampling, much more so than the constellation of passive sensors currently available. In addition, active sensors are blind in the lowest altitudes of the troposphere and have minimum detectable signals that may miss a significant percentage of precipitation in some regions. Especially at high latitudes where light but frequent precipitation dominates, improved passive retrievals could be complementary to current efforts that disagree on the magnitude of precipitation (Behrangi et al. 2016). As discussed above, variational retrievals that accurately account for observation errors maximize the available information content, and thus a variational retrieval could theoretically outperform GPROF or other algorithms in this area.

1.5 Beyond non-scattering ocean scenes

From the early days of microwave remote sensing, sensing of the ocean was recognized as a strength (Wilheit 1977). The ocean is radiometrically cold and polarizing, which contrasts well with warm hydrometeor emission signals in the atmosphere. Emissivity models for open ocean are quite advanced, able to characterize the ocean's radiometric properties with high accuracy from the inputs of temperature, salinity, wind speed, and wind direction (Meissner and Wentz 2012, Kazumori and English 2015). While not without errors, modeled emissivities over a pure ocean surface are multiple times more accurate than modeled ice, coast, or land surface emissivities (Ferraro et al. 2013), though emissivities can also be retrieved with some degree of skill (Boukabara et al. 2011, Lee and Sohn 2015).

The challenge of modeling sea ice, coast, and land surface emissivities, combined with their radiometrically warmer surfaces, are the cause for limited development of variational algorithms over non-ocean surfaces (Ferraro et al. 2013). Historically, this caused the different retrieval communities to run wholly separate algorithms despite their inherent interdependencies on other radiometrically significant species. This separation is discussed and examined in Chapter 4. While the work presented in Chapter 4 stops short of a global variational retrieval, something already extant (Boukabara et al. 2011), it argues for a more holistic approach to microwave imager data in the retrieval community.

Chapter 4 investigates a few limits of a variational retrieval over ocean and how it can complement other types of retrievals. For instance, sea ice monitoring has long been recognized as a strength of passive microwave observations (Wilheit 1977, Comiso and Nishio 2008). The emission signal of sea ice is stark against that of the relatively cold and polarized signal from the ocean surface, making its identification particularly easy. However, the large footprint of passive

microwave sensors and the complex nature of the sea ice edge make marginal ice detection a compelling problem—a problem that traditional sea ice algorithms are not suited to study. Such research applications are ideal for a variational algorithm to explore, as its higher effective signal to noise ratio (SNR) can shed light on complex scenes that are otherwise ignored.

Similar to the sea ice retrieval problem, precipitation retrieval can theoretically be improved with a variational framework to maximize the available information content. However, as with sea ice, there are forward model limitations that have guided algorithm development in the past to simpler solutions (e.g. Kummerow et al. 1996, Wentz and Spencer 1998). For sea ice, this is driven by the high variability in sea ice emissivity due to snow, ice age, melt ponds, and other factors. For precipitation, the high variability in hydrometeors, their radiometric characteristics, and largely unknown size distributions have driven algorithm development. The drop size distribution (DSD) has a large impact on the resultant rain rate (Comstock et al. 2004) though solving for the shape of this distribution typically requires dual-polarization radar (Bringi et al. 2003) or a combined radar and radiometer approach (e.g. Munchak and Kummerow 2011). From Lebsock and L'Ecuyer (2011): "By far the most troubling assumption that must be made in any... algorithm is that of the rain DSD."

The assumption of a rain DSD is so troubling because it impacts the surface rain rate estimates but is not well constrained by typical space-borne measurements. This is especially true for passive-only retrievals, and is one possible reason why previous passive microwave precipitation algorithms forwent explicit forward modeling of rain. Chapter 3 examines the forward model errors engendered by DSD assumptions for a passive-only retrieval of warm rain without explicitly solving for the DSD itself.

Lastly, the retrieval of falling snow is briefly examined in Chapter 4, comparing GPROF with the non-scattering variational algorithm of Chapter 2. This is not examined in much depth as the viability of this problem has been investigated elsewhere (Kulie et al. 2010, Skofronick-Jackson et al. 2013). The scattering properties of snowfall and mixed phase hydrometeors are highly variable, making their retrieval from a passive-only sensor very challenging. This particular retrieval problem is near or beyond the limit of what is reasonably retrievable from a passive microwave imager, a theme which is explored in one way or another throughout this document.

Note that Chapter 2 was previously published in a lightly modified form in the *Journal of Geophysical Research: Atmospheres* in June 2016 as Duncan and Kummerow (2016), while a lightly modified form of Chapter 4 is currently in review with the *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

CHAPTER 2

A 1DVAR RETRIEVAL APPLIED TO GMI: ALGORITHM DESCRIPTION, VALIDATION, AND SENSITIVITIES [1]

2.1 Introduction

The coordination of satellite sensors that make up the Global Precipitation Measurement (GPM; Hou et al. 2014) constellation represents an unprecedented step forward in the capability of the Earth observing system, providing a powerful tool for studying the global hydrologic cycle. The GPM Microwave Imager (GMI) on the GPM core satellite is the anchor of this constellation, a multispectral passive sensor to which all other constellation members are intercalibrated.

Precipitation is not, however, the only aspect of the Earth's hydrologic cycle for which GPM can improve our knowledge. GMI also presents an opportunity for development of next generation retrievals of so-called 'non-raining parameters' because of its channel suite. Previous retrievals of oceanic winds, cloud liquid water path (CLWP), and total precipitable water (TPW) have typically focused on 5- or 7-channel algorithms (e.g. Wilheit and Chang, 1980; Wentz, 1997; Kazumori et al., 2012), largely due to the residual influence of the seminal Special Sensor Microwave/Imager (SSM/I) series of sensors that spans from the late 1980s through 2016. Careful analyses of patterns and trends in global oceanic winds, clouds, and water vapor are of great importance for observational climate data records and climate model evaluation (Parkinson, 2003; Trenberth et al., 2005; Wentz et al., 2007; Hartmann et al., 2013; Manaster et al., 2017). An holistic understanding of the Earth's hydrologic cycle, and how it is changing, demand

1. Previously published in modified form as Duncan and Kummerow (2016)

accurate knowledge of water in both liquid and vapor forms over the global ocean; microwave imagers are unique in the global observing system in that they can provide near-global coverage of these parameters on a daily basis, resolve features that global models currently cannot, and retrieve in cloud. Water vapor is the most powerful greenhouse gas and is expected to increase with global temperatures (O'Gorman and Muller, 2010), though separating the signal from the noise is difficult in the observational record (Sherwood et al., 2010). Similarly, feedbacks associated with clouds are still the biggest uncertainty in model projections of future climate (Boucher et al., 2013), with cloud water and ice content difficult to determine globally from observations (O'Dell et al., 2008; Eliasson et al., 2013).

While satellites provide the best platform for global analysis of the hydrologic cycle, these observations are not without drawbacks. Satellites' orbits can drift over time, sensors degrade, and while the GPM constellation is intercalibrated via methods of relative calibration, the absolute calibration of radiometers is an ongoing topic of research. Radiometric observations carry inherent uncertainty due to limitations of the sensors themselves, which translates into uncertainty in retrieved geophysical parameters. In addition, microwave imager retrievals are typically 'under-constrained,' in that the information content of the radiometric observations alone is insufficient to retrieve a unique solution for all parameters that affect the measurements (Stephens and Kummerow, 2007). Some amount of ancillary data is thus required to constrain the retrieval, but should be chosen with care so as not to bias the final answer.

The oceanic parameters of TPW, 10m wind speed, and CLWP are termed 'non-raining' because, historically, microwave radiative transfer models can quickly and fairly accurately model radiances in an absorbing/emitting atmosphere, but the scattering of microwave radiation by precipitation is more difficult and remains an ongoing problem (Weng, 2007). In fact, most

satellite observations in areas of precipitation, and indeed cloud, are not assimilated into global weather models and constitute an active area of research (Bauer et al., 2006; Bennartz and Greenwald, 2011). The non-raining parameters are thus geophysical variables retrievable from a passive microwave platform in clear-sky or cloudy conditions, but not in precipitating or scattering conditions; to retrieve in raining conditions, assumptions have to be made about partitioning cloud water and rainwater, as in Hilburn and Wentz (2008), and the radiative transfer is more complicated.

Ability to retrieve the non-raining parameters from passive microwave radiances derives from the emission/absorption characteristics of water vapor and liquid water at microwave frequencies and the effect of wind speed on the emissivity of the ocean. To first order, the emission of the ocean surface at microwave frequencies is dependent on wind speed and temperature alone; other factors like ocean salinity and wind direction are second-order effects (Meissner and Wentz, 2002). By assuming a vertical profile of water vapor and cloud water, as well as a temperature profile, the atmosphere can be adequately modeled at microwave frequencies, and thus the non-raining parameters can be retrieved. Highly variable atmospheric constituents such as aerosols and ozone are ignored due to their small radiative impacts.

Retrieval algorithms for microwave imagers have historically relied upon statistical methods (Wilheit and Chang, 1980; Alishouse et al., 1990), and semi-physical methods (Liu et al., 1992; Greenwald et al., 1993; Wentz, 1997) to determine some or all of the non-raining parameters over ocean. Though empirical and semi-physical methods are still widely used, retrieval algorithms have generally moved towards simultaneous retrieval of all non-raining parameters. In addition to increased computational power, this is due to the interdependence of microwave frequencies on surface emissivity and emission from both water vapor and cloud

water, and is thus the best way to ensure that the resultant output is consistent with the observed radiances. Elsaesser and Kummerow (2008), hereinafter EK08, utilized the mathematical framework outlined in Rodgers (2000) to iteratively solve for all parameters simultaneously, and it is this approach that is built upon here.

In the last decade or more, efforts have moved from sensor-specific approaches to an emphasis on intercalibration and algorithms that can provide consistent time series of geophysical constituents from multiple satellite sensors (Berg et al., 2013; Robertson et al., 2014; Hou et al., 2014). The approach detailed here follows previous studies that iteratively solve for geophysical parameters by forward modeling the atmosphere and finding a solution that closely matches observed radiances, balancing observations with prior knowledge of the state vector (Deblonde and English, 2003; Bettenhausen et al., 2006; EK08; Boukabara et al., 2011; Munchak et al., 2016).

The 1D variational (1DVAR) approach is generalizable to other sensors due to its fully physical forward model. The process of forward modeling and iterating to find a solution that minimizes differences between observed and simulated radiances, while taking prior information into account, adds constraints to the under-constrained problem. The forward model must contain all atmospheric components to which the radiances are significantly sensitive and simulate the atmospheric profile of absorbing/emitting constituents with enough fidelity to satisfy the weighting functions of all channels. Any combination of channels may be used in the iteration, the radiative transfer and emissivity models may be swapped, and prior information can be weighted heavily or not at all, lending the approach great flexibility.

Alongside increasing complexity in retrieval algorithms, passive microwave instruments are increasingly multispectral, with frequencies for both sounding and imaging capabilities. The

channel suite of GMI provides sensitivity to mid-tropospheric water vapor via high frequency channels near the strong water vapor absorption line at 183.31GHz, as well as sensitivity to cloud water and the ocean surface from low frequency channels. GMI builds upon the success of previous imagers such as the Advanced Microwave Scanning Radiometer for EOS (AMSR-E) and the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI), which had channels from 6.9 GHz through 89 GHz. Since the channels available on GMI are common to many other imagers currently in use, such as AMSR2 or SSMIS, the CSU 1DVAR is easily adaptable to other sensors.

The remainder of this paper is organized as follows. Section two lists the data sources employed for algorithm development and validation; section three gives a brief overview of 1DVAR theory, details the CSU 1DVAR's forward model, and describes calculation of the covariance matrices; section four contains results and validation against independent datasets; section five discusses sensitivities of the algorithm; and section six contains discussion and some conclusions.

2.2 Data sources

The GMI sensor aboard the GPM satellite is a passive microwave radiometer with 13 channels of vertical and horizontal polarization between 10.65 GHz and 183±7 GHz (see Table 2.1), with spatial resolution on the Earth's surface ranging from 25 km at 10 GHz to 6 km at 183 GHz (Draper et al., 2015). The GPM satellite was launched on February 27, 2014 into a non-sun-synchronous orbit at an inclination of 65° and at a nominal altitude of 407 km, giving GMI an effective latitudinal range of 68°N-68°S due to its swath width of 885 km. GMI is a conically

scanning radiometer, viewing Earth at an average angle of 52.8° for frequencies 10.65 GHz to 89.0 GHz and 49.1° for higher frequency channels.

Table 2.1. GPM Microwave Imager (GMI) sensor specifications.

Frequency (GHz)	**10.65**	**18.70**	**23.8**	**36.5**	**89.0**	**166.0**	**183.31±3**	**183.31±7**
Polarization	V, H	V, H	V	V, H	V, H	V, H	V	V
NEDT (K)	0.78	0.63	0.51	0.42	0.32	0.70	0.56	0.47
Nominal Earth Incidence Angle	52.8°					49.1°		

The brightness temperature (T_B) dataset used in this study is the Level 1C-R version 4 product (GPM Science Team, 2016). This dataset co-locates the center of each channel's field of view (FOV) and is calibrated according to the guidance of the GPM working group on intercalibration. Only pixels that have the highest quality flag, i.e. those determined to be free of radio frequency interference (RFI), with good geolocation data, and no sun glint or warm load intrusion, are used in the retrieval.

European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Reanalysis, ERA-Interim (Dee et al., 2011), data are used to construct a priori matrices, to determine channel error and covariance values, and as the prior. These are analyses from the 6-hourly, N128 (about 78km at the Equator) resolution global product, subset in the vertical dimension to match the forward model of the retrieval. The Modern-Era Retrospective Analysis for Research and Applications, MERRA (Rienecker et al., 2011), was also used to test sensitivity to the prior, and is interpolated to the same grid as ERA-Interim for consistency. For sea surface temperatures (SST), the retrieval reads in the Optimally Interpolated SST product described in Reynolds et al.

(2007), also known as OISST; this is done to avoid potential contamination of SST data with other microwave retrievals.

Validation is based on analysis of matchups between GMI overpasses coincident with a few data sources. International Comprehensive Ocean-Atmosphere Data Set, ICOADS (Woodruff et al., 2011), Release 2.5 buoys are used for validation of wind speeds. Radiosonde observations are used for validation of retrieved TPW. The Remote Sensing Systems (RSS) GMI v8.1 product is used for comparison purposes (Hilburn and Wentz, 2008).

2.3 Methods

2.3.1 1DVAR

At the heart of the retrieval is the mathematical framework known as a one-dimensional variational (1DVAR) technique, also known as optimal estimation. While the mathematics of this technique is described in detail in numerous places (Rodgers, 2000; Bettenhausen et al., 2006; EK08; Boukabara et al., 2011), it is critical to the retrieval and thus warrants a quick overview.

1DVAR is a matrix-based inverse method predicated upon Bayes' Theorem. It blends observations, prior environmental knowledge, and knowledge of the errors in both to invert the measurement vector **y** to determine the state vector **x**. In this case, ***y*** is a vector of brightness temperatures from GMI and ***x*** is a vector that includes CLWP, wind speed, and atmospheric water vapor. The measurements are related to the geophysical state by a forward model, **f(x,b)**, that is dependent upon elements in the state vector and also assumed geophysical properties, **b**. In this case, the forward model is described in Section 2.3.2 and **b** constitutes the atmospheric temperature profile, wind direction, surface salinity, and everything else not solved for explicitly.

The covariances in the measurement vector are described by matrix $\mathbf{S_y}$, and covariances in the a priori vector, $\boldsymbol{x_a}$, that describe prior knowledge of the state vector is given by matrix $\mathbf{S_a}$. The inversion is solved iteratively via Newton's method by assuming Gaussian-distributed errors and a moderately linear response of the measurements to changes in the state vector (Rodgers 2000). To solve the inverse problem, we minimize a cost function that weights both measurements and prior knowledge in accordance with their uncertainties:

$$\Phi = (\mathbf{x}-\mathbf{x_a})^T\mathbf{S_a}^{-1}(\mathbf{x}-\mathbf{x_a}) + [\mathbf{y}-f(\mathbf{x},\mathbf{b})]^T\mathbf{S_y}^{-1}[\mathbf{y}-f(\mathbf{x},\mathbf{b})]. \quad \text{(Eq. 2.1)}$$

The minimum gradient of Φ with respect to $\mathbf{x}$ describes the maximum probability solution of $\mathbf{x}$ given measurement $\mathbf{y}$. This formalism produces a posterior probability density function (PDF) of the retrieved state, described by

$$\mathbf{S_x} = (\mathbf{K}^T\mathbf{S_y}^{-1}\mathbf{K} + \mathbf{S_a}^{-1})^{-1} \quad \text{(Eq. 2.2)}$$

where $\mathbf{K}$ is the weighting function matrix or Jacobian that describes the first derivative of each forward modeled element to changes in the state vector. Elements of the $\mathbf{S_x}$ matrix describe the estimated variances and covariances of errors for retrieved parameters.

Convergence is defined by a minimized cost function and simulated T_Bs that change very little between iterations, given as equation 5.33 in Rodgers (2000). Convergence is typically reached in two iterations for clear-sky scenes and 3-5 iterations for cloudy scenes. A normalized chi-squared metric is used to examine quality of convergence,

$$\chi^2 = [y-f(x,b)]^T S_y^{-1}[y-f(x,b)] / N_{chan} \quad \text{(Eq. 2.3)}$$

where N_{chan} is the number of satellite channels used. The χ^2 cost function referred to throughout the paper is thus the part of Φ that signals the fit to the observations and is independent of the prior. χ^2 is normalized by N_{chan} to allow greater adaptability to other sensors and channel combinations.

2.3.2 Forward model

The forward model employs the Community Radiative Transfer Model, CRTM (Liu and Weng 2013), to compute simulated radiances at the GMI frequencies. Surface emissivities come from the FASTEM6 ocean surface emissivity model (Kazumori and English, 2015). The CSU 1DVAR version described here uses CRTM Release 2.2.3. The forward model has 16 vertical layers in pressure, from 100hPa to the surface. The lowest layer's depth varies in accordance with sea level pressure from the analysis data, while the other layers are static.

The EOF-based approach of retrieving the water vapor profile is a compromise between allowing the channels' weighting functions to guide the retrieved distribution of water vapor and requiring prior information to make the problem viable. This is a common method to reduce dimensionality in 1DVAR retrievals (Boukabara et al., 2011; Munchak et al., 2016). To guide the retrieval, the a priori profiles are subset by SST. The EOFs are defined as variations around a mean profile, calculated offline from ERA-Interim data. The 1DVAR process solves for the coefficient of each EOF, which may be positive or negative, yielding a profile that best matches the T_B vector,

$$WV_{ret} = WV_{SST} + c_1*EOF_{SST,1} + c_2*EOF_{SST,2} + c_3*EOF_{SST,3}.$$

An advantage of EOFs is no covariance between the principal components, by definition, setting off-diagonal elements of the $\mathbf{S_a}$ matrix to zero. According to analysis of ERA-Interim data, the first three EOFs of water vapor mixing ratio account for 90-98% of the total variability, depending on the SST regime. The number of EOFs employed could be increased or decreased for different channel combinations, but for GMI using three EOFs is optimal.

Both wind direction and the atmospheric temperature profile are taken from ERA-Interim. Wind direction has a small but significant effect on T_Bs that increases with wind speed and

depends on frequency. Instead of solving for wind direction, which has a weak radiometric signal (Wentz, 1992), wind direction is taken from ERA-Interim. Its inclusion removes cross-scan artifacts that are present if a static wind direction is assumed. The channel suite of GMI does not contain any temperature sounding channels, and therefore the temperature profile is also taken from the model. The CSU 1DVAR can be run without this model guidance, but assumed channel errors need to be increased if using a climatological lapse rate. Sea surface salinity is taken from monthly climatologies derived from Aquarius mission data (Le Vine et al., 2015).

Other assumptions in the forward model include a fixed cloud layer, no scattering due to clouds or aerosols, a fixed cloud drop size distribution, and a plane-parallel atmosphere. The significance of the plane-parallel assumption is not assessed here but has been explored elsewhere (Rapp et al., 2009; Bennartz and Greenwald, 2011); FOV inhomogeneity is a source of retrieval error for T_Bs that have not been convolved to a common resolution. This is potentially a significant source of error for broken cloud fields in particular, given the size differential between 36 GHz and 166 GHz FOVs, for example. Placement of the cloud layer has little impact on the retrieval due to the increasing emissivity of cloud drops with decreasing temperature (Mätzler et al., 2010), causing the effective emission to be somewhat independent of cloud height but tied strongly to the total amount of liquid water in the column. Pixels for which the non-scattering assumption is invalid typically lead to non-convergence and often signal areas of precipitation, explored in Section 2.4. If the cloud is non-precipitating, due to cloud drops being in the Rayleigh regime for microwave frequencies, emission is proportional to column liquid water mass (Bennartz, 2007), and thus the droplet size assumption is not significant.

2.3.3 Determination of S_y and S_a

The matrices $\mathbf{S_a}$ and $\mathbf{S_y}$ especially have a large impact on the 1DVAR retrieval. $\mathbf{S_a}$ represents the best estimate of prior knowledge of the state vector, and can thus come from a variety of sources. For example, Boukabara et al. (2011) use output from various global and mesoscale models, whereas EK08 defines global variances of each retrieved parameter from climatology and assumes zero covariances for all. Determination of $\mathbf{S_y}$ is much more nuanced.

Elements of $\mathbf{S_y}$ essentially inform the algorithm how much weight each channel should be given in the inversion. This includes sensor noise, forward model parameter error, and the error introduced by the forward model. In EK08 this was estimated in a piecewise manner, adding sensor noise (noise-equivalent differential temperature, NEDT) to estimates of channel sensitivity taken from perturbations to parts of the forward model. In contrast, Boukabara et al. (2011) model radiances through ECMWF model output and then scale the resultant uncertainties by uncertainties in the geophysical input values themselves. This also serves as a method of bias correction, essentially forcing the radiances to match the model, a method common in satellite data assimilation (Liu and Boukabara, 2014), though variational methods are now common for bias correction in data assimilation as well (Auligne et al., 2007).

In this study, $\mathbf{S_a}$ is largely determined via analysis of 6-hourly ERA-Interim data. For wind speed, climatological maps of standard deviations yield location- and month-dependent values. These values are calculated at the N128 model grid resolution and then interpolated and smoothed to produce a high-resolution global grid. In contrast, the EOFs of water vapor are allowed to move more freely, constrained weakly by the variability observed in the reanalysis data, but subset only by SST so as not to alias any regional biases of the model into the retrieval. Each EOF coefficient has maximum and minimum values so as to limit unphysical behavior.

For liquid water there is less constraint, with the prior and variance globally defined in logarithm space; this permits the retrieval to solve for a multitude of clear-sky scenes without biasing against high CLWP regimes, as the global distribution of CLWP is far from Gaussian but is more so in logarithm space. The relatively strong radiometric signal of liquid water thus guides the inversion instead of prior information, with no reliance upon model data. Winds and water vapor, in contrast, have subtler radiometric signals and therefore benefit from greater constraint. The only off-diagonal elements of $\mathbf{S_a}$ that are significant, and thus taken into account, are covariances between CLWP and the EOFs of water vapor, which come from the reanalysis. The results are fairly insensitive to whether $\mathbf{x_a}$ comes from a model or climatology, but assumed errors need to be increased if using climatology and the results are noisier.

Calculation of $\mathbf{S_y}$ is more complex. First, T_Bs are forward modeled using ERA-Interim data at their native vertical resolution, 37 pressure levels, including ice and water clouds as well as scattering in the radiative transfer. This is then compared to simulated radiances, again from ERA-Interim, but using the simplified forward model of the retrieval—16 vertical pressure levels, the EOF-based water vapor profile described in Section 2.3.2, a fixed cloud layer, and no cloud ice. To account for forward model parameter error—the uncertainty of prescribed elements (**b**) in the forward model—noise is added to the parameters of SST, salinity, wind direction, and the temperature profile when running the simplified forward model. Specifically, the uncertainties are taken to be 0.62K for SST, 0.5psu for salinity, 20 degrees for wind direction, and 2K for temperature. Around a hundred million points are forward modeled using both sets of assumptions to get a large statistical sample.

The difference in T_B, simulated versus simulated, is then a synthetic dataset that speaks to forward model error, including forward model parameter error, in an imperfect but tractable way.

This method attempts to account for the largest forward model error sources; however, some error sources, such as absorption model or FOV inhomogeneity errors, cannot readily be quantified. After careful screening for bad matchups and contamination from precipitation or sea ice, the covariances of these simulated versus simulated T_B differences yield an estimate of the forward model errors. This approach also yields forward model biases, typically on order ~0.1K for low frequencies and 1-4K for the 183 GHz channels, which are applied to the observed T_Bs prior to processing.

In order to account for emissivity model errors, a separate analysis compared the T_B response of FASTEM6 at GMI frequencies against another state-of-the-art emissivity model, Meissner and Wentz (2012). Covariances of the difference between the two models are taken as a proxy for emissivity model uncertainty. These covariances are added to those from the previous analysis. Lastly, NEDT values from Draper et al. (2015) are added to the diagonal elements of $\mathbf{S_y}$ from the analysis to yield the channel error covariance matrix used in the algorithm, shown in Fig. 2.1.

It should be stressed that small differences in $\mathbf{S_y}$ can cause large differences in the performance of the retrieval. Decreasing the diagonal elements by say 20% markedly decreases the number of pixels that reach convergence, whereas overestimating $\mathbf{S_y}$ will cause the retrieval to over-weight the prior. Typically, this sensitivity most affects convergence in cloudy scenes, consistent with Bormann et al. (2011), whereas convergence can occur in some clear-sky scenes when assuming that $\mathbf{S_y}$ is simply NEDT. The method described for calculating $\mathbf{S_y}$ yields channel errors and their covariances related to the forward model, but it does not take every possible error source into account and thus may need to be adjusted for different applications and especially for other sensors that may not be as well calibrated as GMI. As seen in Fig. 2.2,

histograms of T_B residuals (observed minus simulated T_B) for each channel are largely Gaussian and exhibit maxima near zero.

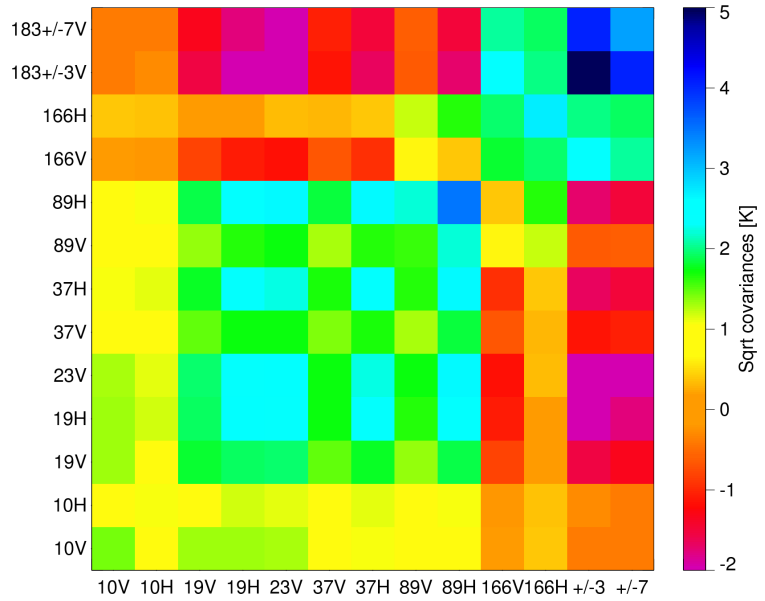

Figure 2.1 Channel error covariance matrix, given as the square root of the covariances so as to be in units of Kelvin. Negative covariances are shown as -1 times the square root of the absolute value of the covariance so as to aid interpretation. This represents the calculated forward model error plus the published NEDT values along the diagonal.

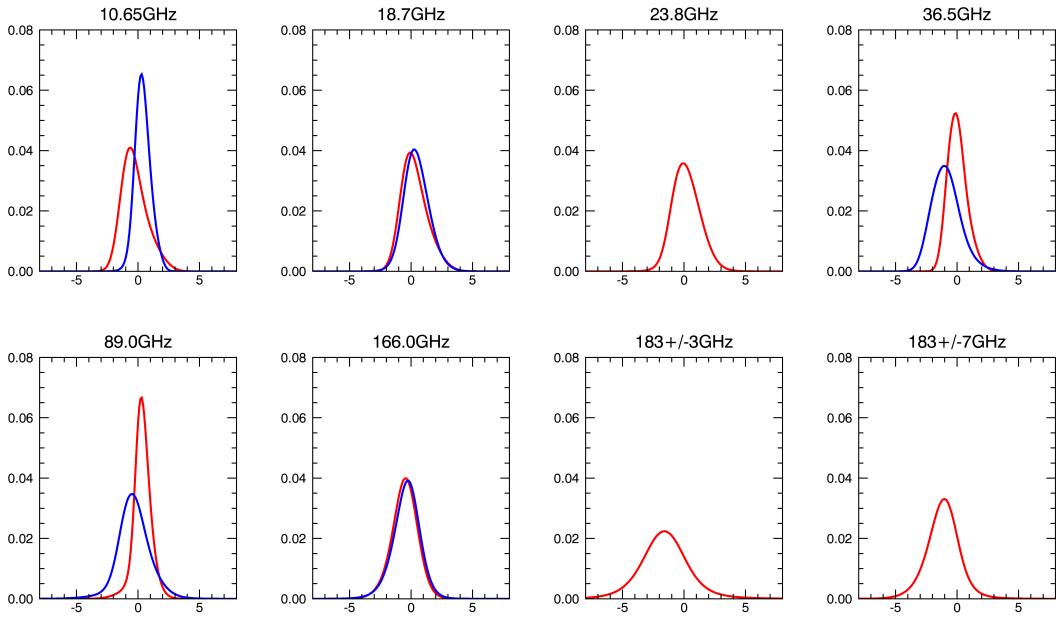

Figure 2.2 Histograms of T_B residuals (observed minus simulated T_B) for 13 channels of GMI, comprised of all converged pixels over one month. Red signifies V polarization, blue H polarization.

2.4 Results

The results presented comprise almost two years of the CSU 1DVAR run for GMI, spanning March 2014 through January 2016. Fig. 2.3 gives an example of pixel-level data from an orbit on November 30th 2014, with retrieved variables, the χ^2 cost function, a single channel T_B residual, and rainfall data from the GMI 2A rainfall product (Kummerow et al., 2015).

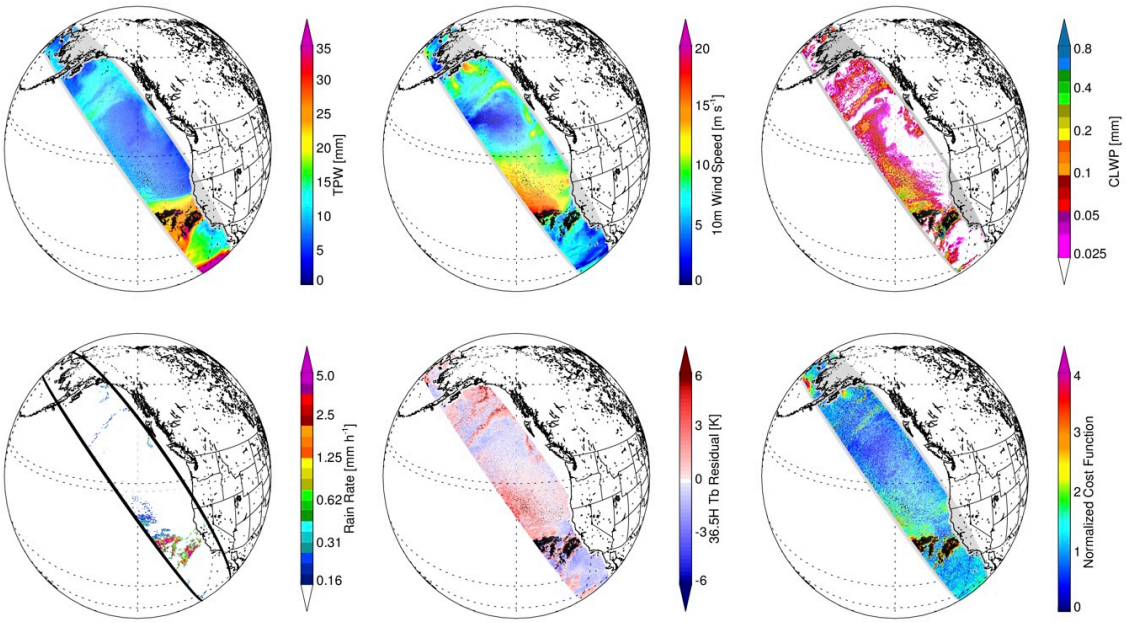

Figure 2.3 Panels of swath-level data for TPW, 10m-wind speed, CLWP, GMI 2A rain rate, 36.5H T_B residual, and the normalized cost function. Black pixels indicate non-convergence. Overpass time is approximately 04Z on Nov. 30th 2014, orbit granule 004286.

This snapshot contains a few aspects of the retrieval worth highlighting. The range extends to high latitudes not observed from GPM's predecessor TRMM, bringing thicker ice clouds more into play. This is possibly the cause of the small area of unsuccessful retrievals off the western coast of Alaska that the rain algorithm does not define as precipitating. Clear-sky regions have low χ^2 values, 0.5 or less, whereas cloudy or high wind regions typically have higher χ^2 values due to assumptions of the forward model being less valid or the emissivity model not being adequate at high wind speeds. The areas of precipitation off California cause

non-convergence and are set to missing. Broken cloud fields with light precipitation are a challenge for the algorithm, visible northwest of the main precipitation feature by California; here, χ^2 is higher, CLWP is high but variable, and there is a tendency for the algorithm to trade TPW for CLWP in some cases (explored in Section 2.6).

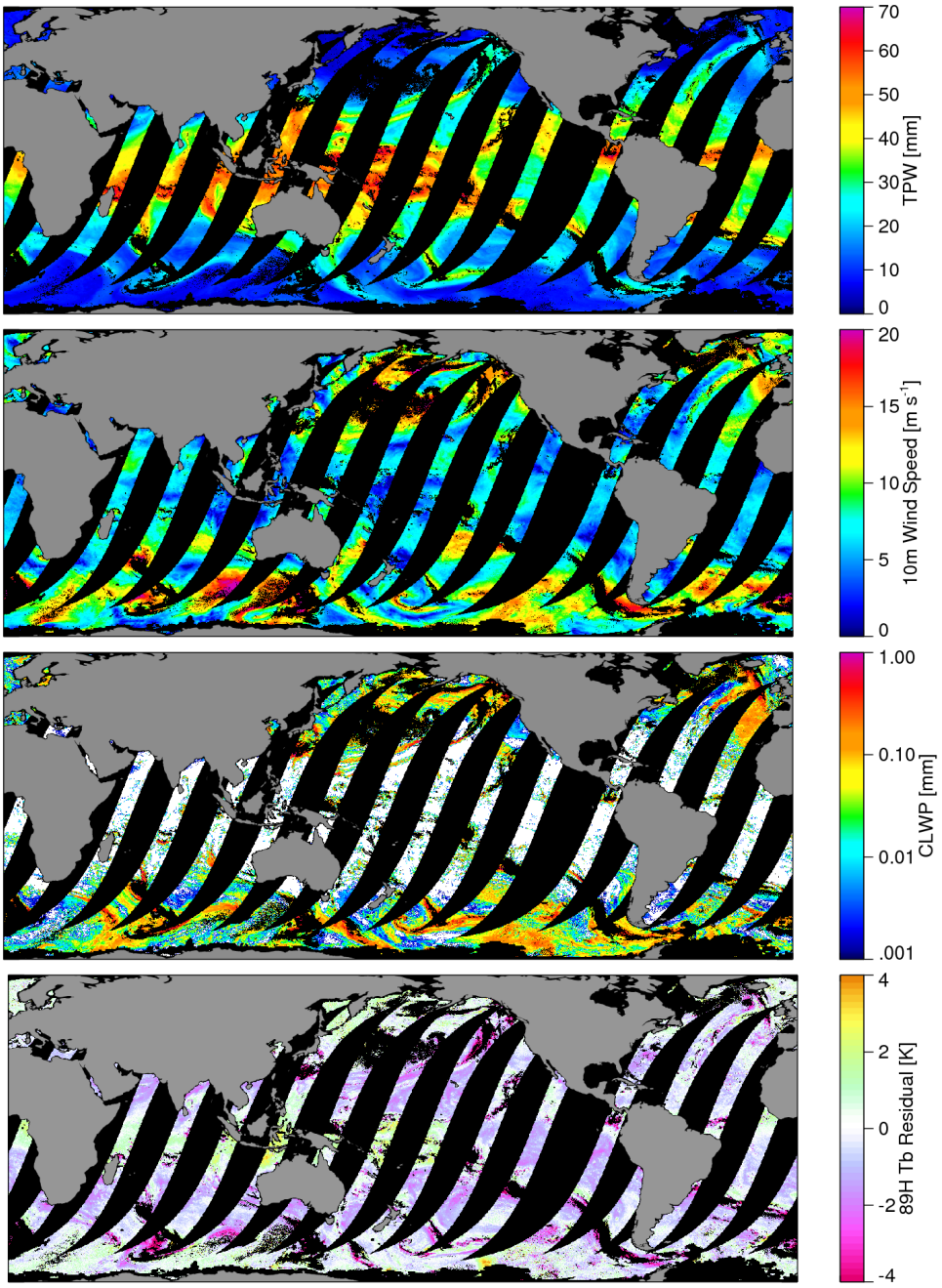

Figure 2.4 Panels showing ascending orbits for 1/21/2015 of TPW, 10m wind speed, CLWP, and T_B residual for 89H. Areas of no data, sea ice, and non-convergence are shown in black. Land is shown in grey.

To compute global grids from these data, all pixel-level data that meet the criteria for high quality retrievals are averaged on a quarter degree grid. Fig. 2.4 displays the ascending orbits from January 21st 2015, showing the non-raining parameters and the T_B residual for one channel. The grid is limited to 70°N to 70°S due to GPM's orbit. The low amount of cloud water in the Tropics is a consequence of the non-raining nature of the retrieval. The panel showing the T_B residual for 89H shows fairly consistent agreement and a slight high bias in the forward model at this frequency.

Figs. 2.5 and 2.6 show validation against buoy and radiosonde observations, respectively. For buoys, matches are defined as lying within a space-time window of 30min and 10km. For the radiosonde matches, the space-time window is 120min and 50km to allow a larger number of matches, driven by the number of coastal radiosonde sites. Vaisala RS92 radiosondes are used exclusively, as they have been shown to have a near-zero bias and ≤0.9mm RMS error when compared to co-located GPS-based TPW observations (Yu et al., 2015). The matches discussed are also required to have a normalized cost function (χ^2) of less than 1.0, signaling a high quality retrieval.

Table 2.2 shows the mean bias and root mean squared error (RMSE) of all matches for both wind speed and TPW. The CSU 1DVAR is also directly compared to gridded RSS data, comparing very well especially for retrievals with low χ^2. For the matchups shown here, the RSS RMSE values were found to be 2.46mm and 1.12m/s for TPW and wind speed, respectively. Interestingly, the difference between retrieval results and the validation data for both wind speed and TPW is highly correlated with the difference between RSS results and the validation ($r \approx 0.85$), true at individual sites and when considered altogether. Considering that the two retrievals use independent methodologies, this could be a consequence of the matchup criteria or

indicate systematic biases in the validation data, though sensor noise and undiagnosed RFI could be factors.

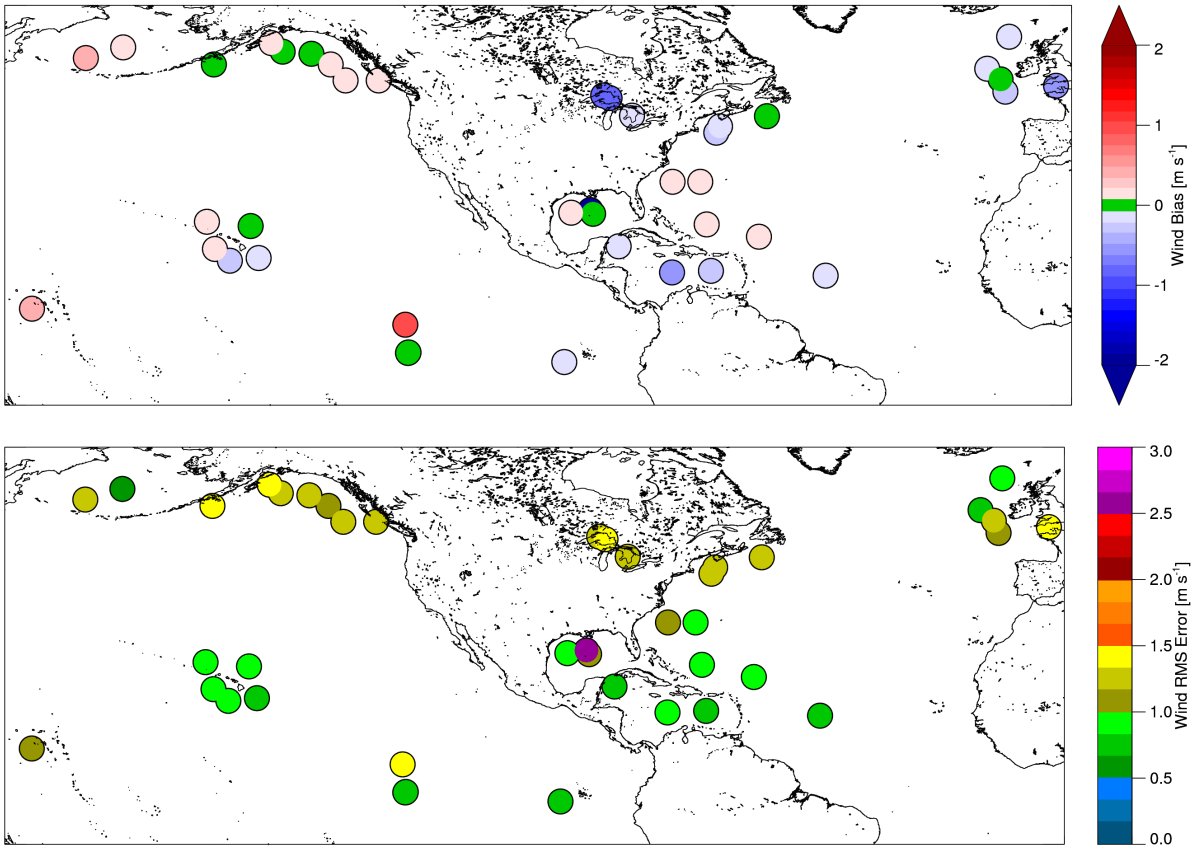

Figure 2.5 Validation of winds against ICOADS buoys. The top panel gives bias relative to all the buoy observations at a particular site, and the bottom panel gives the RMS difference. Sites are shown only if they represent more than 10 matchups.

Table 2.2 CSU 1DVAR output for GMI matched up with radiosondes and ocean buoys for validation. All matched pixels have χ^2<1.0. Matchups are within space-time windows of 10km and 30min for buoys, 60km and 120min for radiosondes, spanning a twenty-two month period.

	CSU 1DVAR		**CSU 1DVAR (Clear-sky)**		**Prior (ERA-Interim)**	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
TPW [mm]	+0.35	2.61	+0.29	2.75	-0.41	3.46
# Matches	1730		1092		1730	
Wind [m s^{-1}]	+0.04	1.20	-0.14	1.09	-0.36	1.74
# Matches	9408		5874		9408	

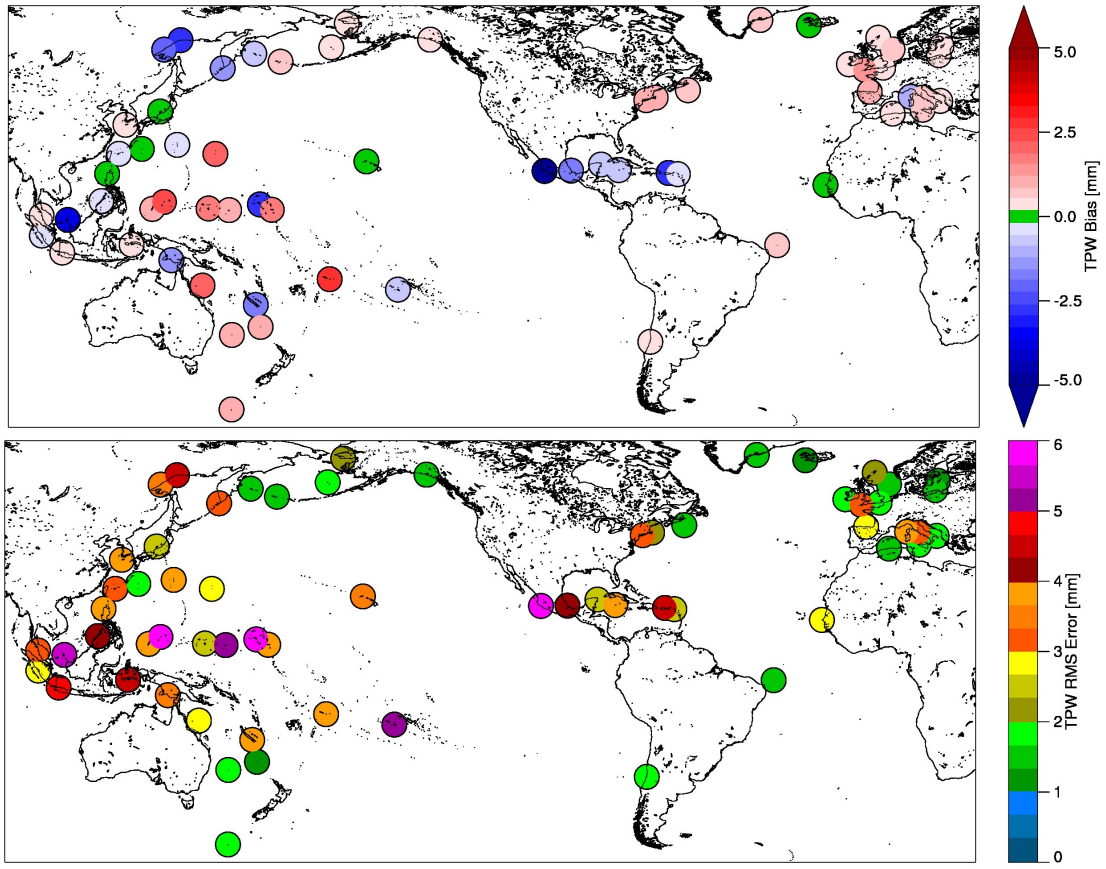

Figure 2.6 Validation of columnar water vapor against radiosonde observations, limited to sites using Vaisala RS92 radiosondes. Sites with 10 or more matches are shown. The top panel gives bias relative to all the observations at a particular site, and the bottom panel gives the RMS difference.

A few conclusions can be drawn from Figs. 2.5 and 2.6 regarding the retrieval skill for TPW and 10m wind speed at individual sites. This breakdown shows interesting patterns and some anomalous sites; many sites have near-zero biases, though some sites do stand out. For instance, the wind retrieval appears much worse at one station in the Gulf of Mexico than at two other adjacent sites; this buoy sits on a tight SST gradient caused by coastal upwelling and OISST errors could cause the bias in the retrieved winds. The TPW retrieval has trouble with a few sites in the Pacific warm pool, which may be caused by unscreened land contamination by small islands. Matchups less likely to have land influences tend to validate better.

An advantage of the 1DVAR framework is the full description of retrieval errors it provides. To test the validity of these posterior errors, the retrieval error relative to the validation data is assessed against the posterior error values. Since posterior errors are defined as the $\pm 1\sigma$ bounds, 68% of the validation data points should lie within the error bounds given by the retrieval. For wind speed and TPW the validation data are within the posterior error bounds 58% and 64% of the time, respectively. Given that the validation data and the spatiotemporal matchups are imperfect, this signals that the posterior errors typically do a good job of describing the retrieval's errors.

In contrast to TPW and wind speed, validation of CLWP is more problematic. Observations of cloud water content are almost exclusively from other satellite retrieval products, with scant field campaigns to provide robust validation (Turner et al., 2007). Thin and warm-phase clouds are not modeled well (King et al., 2015) and remain a significant source of uncertainty in climate projections (Boucher et al., 2013). However, absolute validation for a global CLWP product is not currently feasible and is not the focus of this study. In-depth comparison studies of CLWP products can be found elsewhere, but can only address relative agreement instead of absolute accuracy. For these reasons the examination of CLWP here is limited to a comparison of retrieved CLWP against RSS, shown in Fig. 2.7. This was done using 0.25° gridded data with raining areas (RSS rain rate >2 mm hr^{-1}) screened out. RSS tends to underestimate CLWP on the high end due to that algorithm's cloud/rain partitioning for CLWP>180 g m^{-2} (Hilburn and Wentz, 2008). The CSU 1DVAR is more likely to assign clear sky—defined here as CLWP<1 g m^{-2}—than RSS and is biased lower in the range of 25 g m^{-2} <CLWP<400 g m^{-2}, also in line with the finding that RSS overestimates CLWP in almost every regime when compared to an independent satellite product (Seethala and Horvath, 2010). The

mean posterior error for CLWP is around 20 g m^{-2} for most regimes, so these differences with RSS are indeed significant. See Wentz (1997) for why RSS retrieves negative CLWP values.

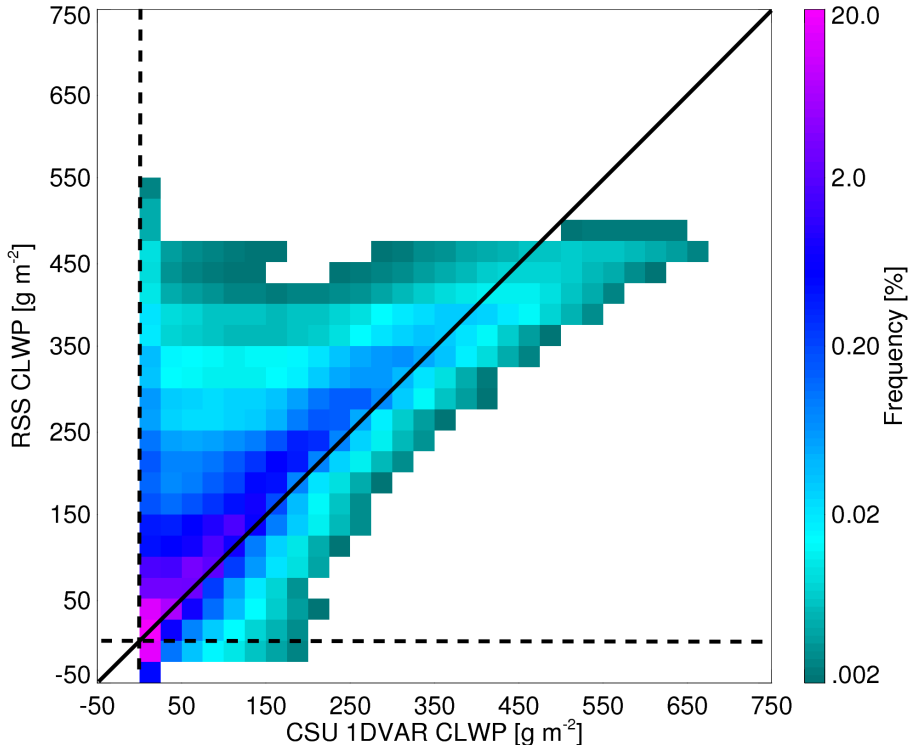

Figure 2.7 2D histogram of CLWP from the CSU 1DVAR applied to GMI and co-located points from RSS v8.1 GMI gridded data. The solid line represents the one to one line while the dashed lines represent CLWP=0.

Finally, as a validation of the non-raining aspect of the CSU 1DVAR, GMI footprints were matched to the GPM Dual-frequency Precipitation Radar, DPR (Hou et al., 2014) to assess the performance of the retrieval in light precipitation. Using an effective FOV equal to that of the 23.8 GHz channel, Ku-only DPR data were averaged within each GMI footprint and compared over a month of retrievals. The results are shown in Fig. 2.8. At rain rates above 1 mm hr^{-1}, a large majority of retrievals fail, a result expected due to the non-scattering assumption made in the forward model. However, the retrieval fails to converge for 1.9% of pixels for which DPR observed a rain rate of zero, and it is not until 0.4 mm hr^{-1} that more pixels fail to converge than succeed. This points to the CSU 1DVAR being fairly good at screening out precipitation even

though it may converge in very light rain rates. A full 86% of pixels with zero rain rate are considered high quality retrievals (χ^2 <1.0), dropping to <60% for any nonzero rain rate, signifying that χ^2 is indeed a good indicator of rain-free conditions, in agreement with EK08. The 1.9% of non-raining pixels for which the retrieval fails may be scenes with significant cloud ice, sea ice within the FOV, unscreened RFI, or any number of things not fully represented in the forward model.

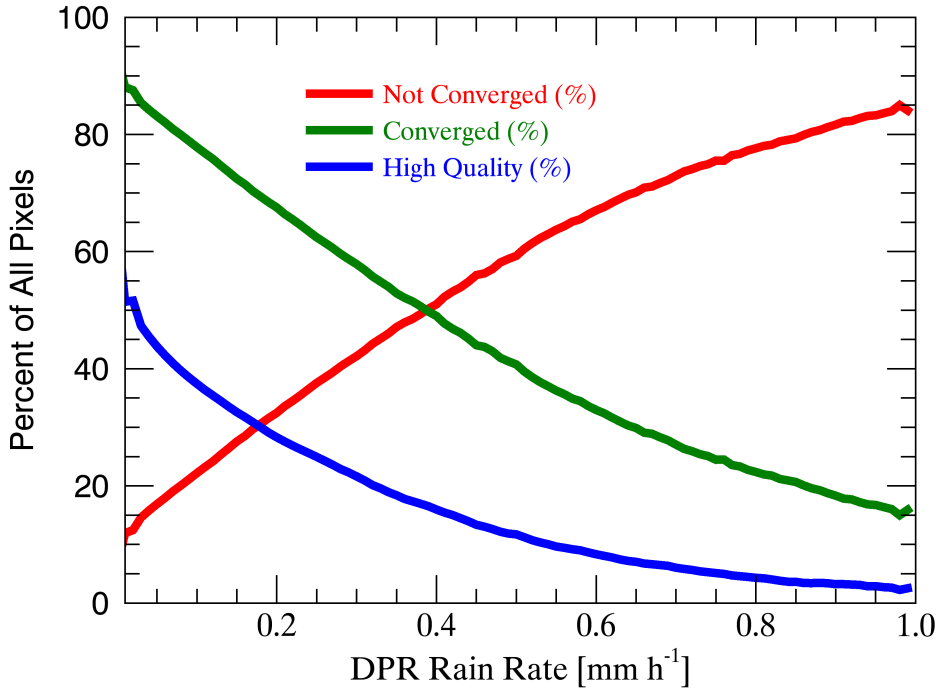

Figure 2.8 Histogram from one month (November 2014), contrasting all pixels where convergence was achieved (green), not achieved (red), and high quality convergence was achieved (blue) against rain rates from DPR.

2.5 Sensitivities

Representation of water vapor's vertical distribution is one of the subtlest and most significant components of the algorithm. The physical reason for this is that signals from wind speed and cloud water may be somewhat aliased onto water vapor, whose emission signal may

weakly or strongly affect most of GMI's 13 channels (explored in the following section). Other studies have assumed a static water vapor profile (EK08) and relied upon historical radiosonde data to inform a single layer atmosphere (Wentz, 1997; Hilburn and Wentz, 2008). Boukabara et al. (2011) solve for the whole profile by relying on model guidance and using a reduced parameter space by employing EOFs, not dissimilar from this approach. Given the information content of a microwave imager and specifically GMI with its water vapor sounding channels, the approach of solving for the moisture profile via EOFs is a compromise. Biases on synoptic scales driven by the distribution of water vapor in the column, say ahead of versus behind a frontal boundary, disappeared after moving to an EOF-based retrieval.

A goal of this study was to develop an algorithm that did not require fine-tuning of calibration to achieve good validation. Apart from accounting for biases directly attributable to the forward model (see Section 2.3.3), no offsets were applied to GMI L1CR T_Bs for processing, a calibration standard based on first principles only (Draper et al., 2015). Addressing the sensitivity of the CSU 1DVAR to various calibration changes would necessitate a separate study. Crucially, if the calibration is significantly off or the forward model is not adequate at simulating the atmosphere and ocean surface, convergence will not be consistently achieved.

Development and processing of the CSU 1DVAR relies on ERA-Interim, though any global analysis could be used. To test the sensitivity of the retrieval to the prior, MERRA was used in place of ERA-Interim. This has a small effect, mainly impacting upper level water vapor while winds and CLWP are unaffected. The information content of GMI limits the effect of the prior on all but upper level moisture. This translates into a small mean decrease in TPW and a worse fit with the forward model for the 183±3 GHz channel, increasing the mean T_B residual bias from -1.5 K to -2.3 K. We conclude that the CSU 1DVAR performs similarly with MERRA

as the prior, showing that the choice of global analysis has a small impact on the retrieval's output.

Retrieval of SST is also possible with the CSU 1DVAR. However, including SST as a variable within the forward model did not conclusively demonstrate skill when compared to using OISST. Smaller assumed errors for the 10 GHz channels and better prior information on SST covariances would potentially render retrieval of SST more skillful, but that has yet to be explored.

2.5.1 Diagonal S_y experiment

Most 1DVAR microwave retrievals employ diagonal $\mathbf{S_y}$ matrices, essentially assuming that there is no covariance of errors between the channels (EK08; Boukabara et al., 2011; Munchak et al., 2016); the Bettenhausen et al. (2006) retrieval for WindSat is a notable exception. While this is a fairly good assumption for instrument errors, it is not necessarily a good assumption when considering forward model errors. For instance, if the forward model inaccurately models the middle troposphere, errors at the 183 GHz channels will be slightly or even strongly correlated. Observation error covariances for hyperspectral infrared sounders have recently been included in data assimilation schemes and shown to improve forecast skill (Weston et al., 2014), but as yet microwave observations have been assumed to have no error covariance.

An experiment was run using the same $\mathbf{S_y}$ matrix given in Fig. 2.1 but setting all off-diagonal elements to zero. This caused slightly smaller mean T_B residuals, smaller average χ^2 values, and a 6% increase in converged pixels. However, this experiment caused a wind speed bias of +0.25 m s^{-1} and a TPW bias of +0.13 mm when compared to the standard algorithm, with a small decrease in CLWP as well. The areas of most significant impact are high latitudes and high CLWP or wind speed regimes, where the diagonal $\mathbf{S_y}$ yields many more converged pixels

but decreased retrieval skill. This result indicates some utility in including observation error covariances, in some situations especially, but it does increase computational cost. This result is implicitly in line with findings in Bormann et al. (2011), in that forward model error exhibits greater covariance in cloudy scenes.

2.5.2 Decreased reliance on prior experiment

By increasing the assumed errors on the prior, the retrieval is forced to rely more on the observations, lending increased weight to the T_Bs and less weight to the prior. For this experiment $\mathbf{S_a}$ was quadrupled, effectively doubling the assumed standard deviation of each parameter. The retrieval required more iterations on average and converged for fewer pixels. For pixels where both the standard retrieval and this experiment converged, the standard retrieval had larger mean T_B residuals and slightly larger χ^2 values, an expected result given the greater weight on observations. Mean TPW was 0.13 mm higher while CLWP and winds were unchanged. This is an encouraging result, as it speaks to the CSU 1DVAR's relative independence from model guidance and robustness with respect to assumptions of the prior, but also underlines how prior information helps guide the retrieval to convergence.

2.5.3 Nine channel experiment

This experiment limits the GMI channel suite to that of TMI. Running the CSU 1DVAR without the high frequency channels of GMI causes it to run faster but also incurs some loss of skill. Without the information content to resolve middle tropospheric water vapor, there is slightly less convergence while TPW is biased 0.32 mm higher and exhibits slightly less overall variability. There is some crosstalk present that causes CLWP to be 16% lower, though winds are unaffected in the mean. While these are generally small impacts, it both demonstrates the

utility of including the high frequency channels of GMI and reinforces that the 19-89 GHz channels contain the greatest information content (Fig. 2.9). This experiment indicates the adaptability of the CSU 1DVAR to other microwave imagers that lack high frequency channels (TMI, SSMI, AMSR2, etc.) while retaining most of the retrieval's skill.

2.6 Discussion

This paper describes a 1DVAR retrieval applied to GMI that shows significant skill in retrieving oceanic wind speeds and columnar water vapor in non-raining scenes. The fully physical forward model could be applied to any microwave sensor in the GPM constellation, utilizing any combination of channels. No calibration adjustments have been made to achieve the results presented; rather, the forward model adequately simulates the geophysical state within the error bounds prescribed by $\mathbf{S}_\mathbf{y}$ to achieve good validation. The inclusion of observation error covariances has shown value in scenes where the forward model assumptions are less valid, such as scenes with thick cloud, strong winds, and at high latitudes. The CSU 1DVAR is demonstrably robust, as changes to the prior, channel suite, and weighting of prior knowledge have small effects on retrieved parameters.

A poorly understood source of error for microwave retrievals is crosstalk error—misinterpretation of the radiometric signal causes an error in estimating one variable that can affect others (Lin and Rossow, 1994). Crosstalk errors tend to occur when the sensitivity of the channel suite is similar for multiple forward model parameters. For instance, Fig. 2.9 shows that while the sensitivity of wind in the forward model is largely independent, some EOFs of the water vapor profile may be confused with CLWP. The 1DVAR framework provides posterior

error estimates not only for individual parameters, but also their covariances, thus yielding a way to analyze crosstalk errors.

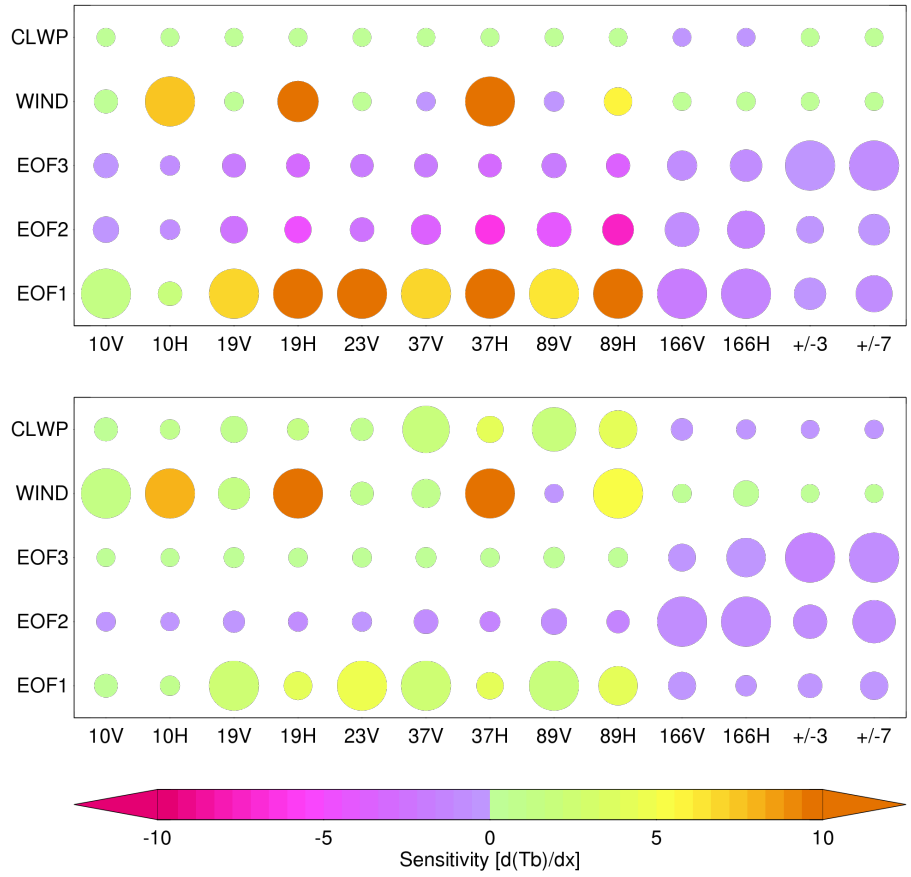

Figure 2.9. Jacobians for two converged pixels. The top panel shows a tropical pixel (TPW=49 mm, wind=6.1 m s^{-1}, CLWP=0 g m^{-2}), while the bottom panel shows a mid-latitude pixel (TPW=26 mm, wind=12.4 m s^{-1}, CLWP=180 g m^{-2}). Sensitivities are given as dT_B/dx, where x is the retrieved variable in its normal units, however the wind and CLWP values are scaled by 10 and 100, to units of K/(10 m s^{-1}) and K/(100 g m^{-2}), respectively, to aid comparison. The size of each dot is proportional to the relative sensitivity of each channel to retrieved variables, with the biggest dot in each column signifying that channel's greatest sensitivity.

For the two example pixels in Fig. 2.9, posterior error covariances for wind and each water vapor EOF are near zero, spanning -0.02 to 0.03 m s^{-1}. For both example pixels, crosstalk errors between EOF coefficients are small, -0.03 to 0.07. The error covariance of CLWP with EOF2, however, is 74 g m^{-2} for the tropical pixel. The posterior error value given for EOF2, 0.32, translates into a large potential error in CLWP, dominating the relatively large posterior error

given for CLWP of 18 g m^{-2} even with a reasonable fit to observed radiances (χ^2=0.76) and CLWP=0.0. This result is reinforced by analysis of the Jacobian, which shows similar T_B sensitivities for EOF2 and CLWP. The mid-latitude pixel exhibits less crosstalk and consequently a smaller posterior error in CLWP of 15 g m^{-2}, in spite of the significant 184 g m^{-2} of cloud water retrieved. We conclude that crosstalk errors are most prevalent between CLWP and water vapor, especially in high TPW regimes. Crosstalk errors vary greatly depending on regime, and the 1DVAR framework proves useful in their identification; amelioration of crosstalk errors should be possible via better representation of a priori error covariances in $\mathbf{S_a}$.

Dynamic treatment of the vertical distribution of water vapor in the atmospheric column is an important aspect of the retrieval. EOF analysis is leveraged to make the problem viable, with at most three degrees of freedom from the observations versus 16 vertical levels in the forward model. Thus, while total column water vapor is well validated, the vertical structure of water vapor from the retrieval is far too coarse to be of use for research purposes. It is reliant upon the variability of model data and, with only two channels in GMI's 183 GHz complex, lacks the information content to resolve fine vertical structure. Analysis of clear-sky matchups with a sensor that does have the necessary information content, the Atmospheric Infrared Scanner, AIRS (Aumann et al., 2003), showed that the CSU 1DVAR has a consistent low bias for water vapor in the 700-850 mb levels and a high bias for 300-700 mb and 850-925 mb levels relative to AIRS. As the AIRS retrieval has been validated against radiosonde profiles (Dang et al., 2012), we conclude that the CSU 1DVAR's water vapor profiles constitute an area of potential improvement.

CHAPTER 3

TOWARDS VARIATIONAL RETRIEVAL OF WARM RAIN FROM PASSIVE MICROWAVE OBSERVATIONS

3.1 Introduction

Global observation of precipitation depends heavily on passive measurements of hydrometeors at microwave wavelengths. Active sensors possess certain advantages relative to passive sensors, but a full global picture of precipitation is currently impossible from active sensors alone as they yield limited spatial coverage and may miss near-surface precipitation due to ground clutter effects. While landmasses may be covered by ground radar networks, a satellite platform is effectively required for global observation of rainfall. Accurate observation of the hydrologic cycle at a high spatiotemporal resolution is a worthy goal (Hou et al. 2014), and a task that realistically requires passive microwave rainfall retrievals.

Retrieval of precipitation from passive microwave observations is an under-constrained problem (Stephens and Kummerow 2007). This is due to many factors, including limitations in radiative transfer modeling, beam-filling errors, lack of vertical information, limited channels with non-independent information content, and unknown distributions of ice, mixed-phase, and liquid hydrometeors, to name a few. In effect, there are more unknowns than pieces of independent information, and thus many assumptions are necessary to make the problem tractable. This has historically been done via algorithms built on empirical relationships (e.g. Wilheit and Chang 1980; Hilburn and Wentz 2008) or algorithms based on Bayesian principles with Gaussian-distributed parameters (Iturbide-Sanchez et al. 2011; Kummerow et al. 2015).

The patterns and magnitude of precipitation over the tropical oceans are largely agreed upon, partly a result of the coordinated study of tropical precipitation from the Tropical Rainfall Measuring Mission (TRMM; Kummerow et al. 2000). In contrast, the high latitude oceans remain an area of disagreement between different observing platforms and between global models (Stephens et al. 2010; Behrangi et al. 2012; Behrangi et al. 2016). The launch of the Global Precipitation Measurement (GPM) Core Observatory in 2014 (Hou et al. 2014) increased the observational capability of sensing precipitation at higher latitudes, but since the Dual-frequency Precipitation Radar (DPR) has limited sensitivity to the light precipitation so prevalent at high latitudes, a large degree of uncertainty remains (Skofronick-Jackson et al. in press). In theory, a variational retrieval is sensitive to rainfall below the detectability threshold of DPR, and is also not susceptible to ground clutter that may obscure shallow precipitation. Thus the effective detection threshold of a passive-only algorithm may be better suited to retrieval of light rain rates characteristic of high latitude oceans.

GPM's operational passive algorithm, the Goddard Profiling Algorithm (GPROF; Kummerow et al. 2015), leverages the synergy of co-located radar and radiometer observations from GPM to calculate a Bayesian average precipitation rate for all GMI observations, and by extension all GPM constellation radiometers. The Bayesian scheme uses the T_B vector to find an average set of atmospheric profiles that match what the radar would have seen, based on the *a priori* database. One weakness of this approach is that hydrometeors below the detectability threshold of the radar will necessarily be missed by the passive algorithm, even if the T_Bs exhibit signal where the radar does not. Further, while this type of approach gives a satisfactory average answer, it does not explicitly model radiation coming from the surface and atmosphere, blunting

the effective signal to noise ratio by including many surface states and cloud types in the Bayesian average.

Warm rain—precipitation driven primarily by collision-coalescence below the freezing level—is particularly challenging to sense from satellite platforms. Passive microwave algorithms are built to exploit the differential signals of emission from liquid drops and scattering from large drops and mixed phase or frozen hydrometeors, but in the absence of very significant emission or scattering, the signal may be from cloud alone or a combination of factors (Stephens and Kummerow 2007). In spite of these challenges, warm rain is not an insignificant player in the global hydrologic cycle. Warm rain constitutes a majority of precipitating clouds in some stratocumulus and high latitude regions (Lebsock and L'Ecuyer 2011) and 20% of total rainfall over Tropical oceans is from warm clouds (Liu and Zipser 2009).

The operational data assimilation (DA) community is also invested in passive microwave radiances in precipitating conditions. Successful assimilation of 'all-sky' radiances from microwave radiometers can yield a more accurate analysis state from which numerical weather prediction (NWP) models can run (Bauer et al. 2006). However, the same factors that cause the retrieval problem to be under-constrained are relevant for DA schemes as well. Thus, microwave radiances from raining or cloudy pixels are often not included in the data assimilation. Or, if they are included, it is with observation errors as large as 40K (Lean et al. 2017), greatly diminishing the information content actually added to the analysis state. Furthermore, errors are assumed independent of one another—a poor assumption for nearby microwave frequencies especially.

The main impediments to retrieval of precipitation over ocean can be distilled down to three factors. First, it is difficult to differentiate between cloud water and rainwater from radiances alone, necessitating an assumed partition between cloud water and rainwater in the

absence of significant scattering. Second, passive radiances contain almost no information on the vertical structure of hydrometeors. Third, the T_Bs do not contain enough information to solve for the DSD parameters, but the scattering properties, fall velocity, and resultant rain rate of hydrometeors are dependent upon their size distribution; also, the degree to which the hydrometeors' distribution impacts the radiances depends on the concentration of hydrometeors in the atmospheric column.

This study builds upon the ocean algorithm developed for the GMI described by Duncan and Kummerow (2016), the CSU 1DVAR, with several augmentations. The satellite instruments and datasets used in this study are detailed in Section 3.2. Section 3.3 addresses each of the aforementioned impediments to a variational precipitation retrieval. Section 3.4 describes the CSU 1DVAR and the algorithm's innovations that permit retrieval of warm rain. Section 3.5 presents a few case studies of GMI overpasses at high latitude ocean locations, compared against independent rainfall estimates from space-borne and ground radars. The paper closes with conclusions and discussion in Section 3.6.

3.2 Data

The GPM Core Observatory holds two instruments, the GPM Microwave Imager (GMI) and the Dual-frequency Precipitation Radar (DPR). GPM is on a non-Sun synchronous orbit at an inclination of 65° and was launched in February 2014. Compared to its predecessor, TRMM, the higher inclination orbit allows for observation of latitudes well outside the Tropics. GMI is a 13 channel passive microwave radiometer containing channels from 10 GHz to 183 GHz at horizontal and vertical polarizations. DPR is a dual-frequency precipitation profiling radar

observing at Ku (13.6 GHz) and Ka (35 GHz) bands with a 12 dBZ sensitivity threshold. This study uses GPM V05 brightness temperatures and level 2 products.

The CloudSat mission's payload is a 94 GHz cloud profiling radar (Stephens et al. 2002). CloudSat was launched in 2007 and flies in the A-Train constellation (L'Ecuyer and Jiang 2010). At a higher frequency than DPR, CloudSat has sensitivity to clouds and light rain not seen by DPR, though its signal can attenuate in moderate to heavy precipitation. CloudSat's small footprint permits highly limited spatial sampling. For light precipitation, CloudSat provides the best observational record currently available from satellite, and is thus complementary to GPM observations. CloudSat's overpasses coincident with GPM were determined using the CloudSat-GPM Coincidence Dataset version 1C (Turk 2016).

The warm rain retrieval from CloudSat (Lebsock and L'Ecuyer 2011) is used to construct *a priori* states usable by the CSU 1DVAR. This algorithm and the associated data product, 2C-Rain-Profile, yields profiles of rain water content, cloud water content, and precipitating ice water content as well as surface rain rate. 2C-Rain-Profile uses a variational approach to match observed radar reflectivities with a two-stream forward model that includes multiple scattering. It employs a variable DSD chosen specifically for its applicability to warm rain scenes that are dominated by small drops. CloudSat's single frequency radar is supplemented by visible optical depth information from another A-Train sensor to constrain the retrieval of cloud water path. To analyze observed hydrometeor profiles over the global oceans, CloudSat is the best observing platform available.

The GPM Ground Validation team collects data from certain NEXRAD (Next Generation Radar) sites matched with GPM overpasses (gpm-gv.gsfc.nasa.gov). NEXRAD operates a dual-pol radar site on the island of Middleton, Alaska at 59°N. This radar site is ideal for comparisons

against GPM observations due to its essentially oceanic location at a high latitude location. This ground radar will be referred to as PAIH, its station identifier, hereafter. Ground radar rain rates used in the analysis are from the polarimetric Z-R algorithm (Bringi et al. 2004).

3.3 Impediments

In this section, the key impediments to a successful variational precipitation retrieval are enumerated, described, and given solutions. Each is directly tied to an element of the retrieval as described in the following section.

3.3.1 Partitioning non-scattering liquid emission

At GMI wavelengths, cloud drops are firmly within the Rayleigh scattering regime. Liquid clouds interact with microwave radiation as pure emitters if there is any non-negligible emission from cloud drops (Bennartz 2007). Mie theory dictates that scattering is proportional to the size parameter ($X=2\pi r/\lambda$) to the fourth power. For instance, going from a cloud drop of r=10 μm to a drizzle drop of r=100 μm, scattering can approach the Mie regime at higher microwave frequencies. But even an effective radius of 100 μm yields X=0.19 at 89 GHz, just on the verge between the Rayleigh and Mie scattering regimes. Thus for many drizzle cases, the GMI frequencies will not diverge significantly from a forward model that neglects scattering.

A simple absorbing/emitting forward model can be run due to the lack of scattering from cloud and drizzle drops. In fact, the predominant lack of scattering from drizzle holds for pristine and polluted regimes, as cloud top effective radii usually remain less than 30 μm even for precipitating clouds (Lebsock et al. 2008). However, because non-raining and raining clouds exhibit similar signals, this requires an assumption of partitioning between cloud and rain water emission from passive microwave algorithms. In contrast, a radar algorithm such as that used by

CloudSat is more skillful at differentiating between cloud and rain drops because radar backscatter is very sensitive to drop size.

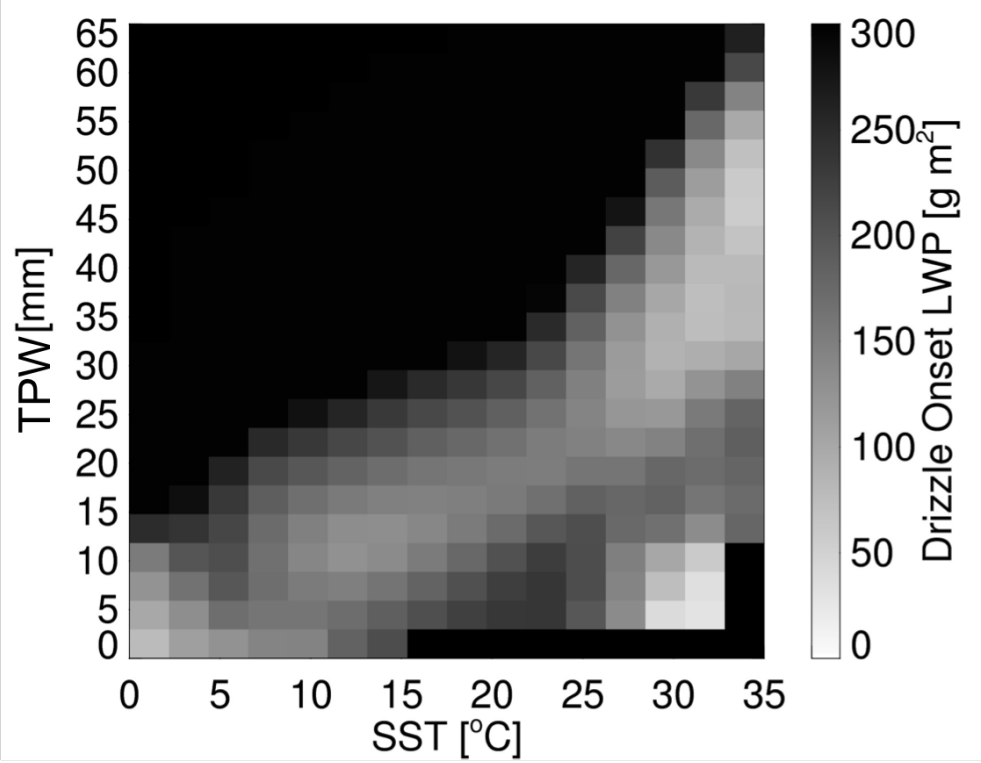

Figure 3.1 Drizzle onset value, separated by SST and TPW. Regimes with no data were assigned the maximum value, 300g m^{-2}.

To calculate regime-dependent values for the onset of drizzle from liquid clouds, CloudSat data were used. Fig. 3.1 shows the drizzle onset values of liquid water path (LWP) used in this study, separated by total precipitable water (TPW) and sea surface temperature (SST). Precipitation frequency observed by CloudSat in all oceanic regimes was analyzed and compared to the distribution of LWP as retrieved by the CSU 1DVAR non-raining retrieval. This approach implicitly assumes that oceanic warm clouds with higher LWP are more likely to be precipitating, an assumption broadly true in studies of A-Train data (Stephens and Haynes 2007; L'Ecuyer et al. 2009; Chen et al. 2011). Taking non-convergent retrievals into account and assuming them to be precipitating, the precipitation frequency from CloudSat was matched to a

LWP value from GMI. These values are in general agreement with some in the literature (Wentz and Spencer 1998; Lebsock et al. 2008; Chen et al. 2011) and lower than some others (Kida et al. 2008; Iturbide-Sanchez et al. 2011), though direct comparison is difficult due to the subdivision by environmental regime done here.

3.3.2 Profiles of hydrometeors

Profiles of hydrometeor species are required to run a realistic radiative transfer scheme as part of the forward model. Further, the surface rain rate depends on the rainwater content in the lowest atmospheric level, not a column total. Vertical information is however effectively nonexistent in the T_B vector, as emissivity is not strongly tied to temperature or pressure. Global model data are insufficient to aid in vertical constraints due to the spatiotemporal heterogeneity of clouds and precipitation. Principal component (PC) analysis can reduce the dimensionality of the problem, simplifying treatment of hydrometeor profiles in the retrieval.

Two years of data from the CloudSat 2C-Rain-Profile product (Lebsock and L'Ecuyer 2011) were analyzed to determine the principal components that best describe hydrometeor profile variability for warm rain. These are separated by SST and lightly smoothed, with the first PC of rain water content (RWC) and precipitating ice water content (PIWC) shown in Fig. 3.2. The first PCs of RWC and PIWC describe 63% and 51% of the total variability, respectively. Covariances between the PCs of RWC and PIWC are also calculated and used in the *a priori* covariance matrix of the retrieval in raining scenes. It is only the shapes of profiles from the 2C-Rain-Profile product that are used in the retrieval.

Attempting to retrieve more than one PC of each species is unproductive and often leads to non-convergent retrievals. The second PC of each species is effectively a vertical redistribution of the first PC in altitude, i.e. more RWC near the surface and less RWC higher up

or vice versa. Since the T_B vector is, to first order, sensitive to total columnar liquid, inclusion of the second PCs is not useful for a passive retrieval. Sensitivity experiments conducted with additional PCs yielded approximately the same number of degrees of freedom for signal (DFS) in the retrieval, determined via the trace of the A-matrix (Rodgers 2000), demonstrating that retrieval of additional profile parameters is not possible with the information content available. Raining pixels typically exhibit 3.0 to 3.5 DFS.

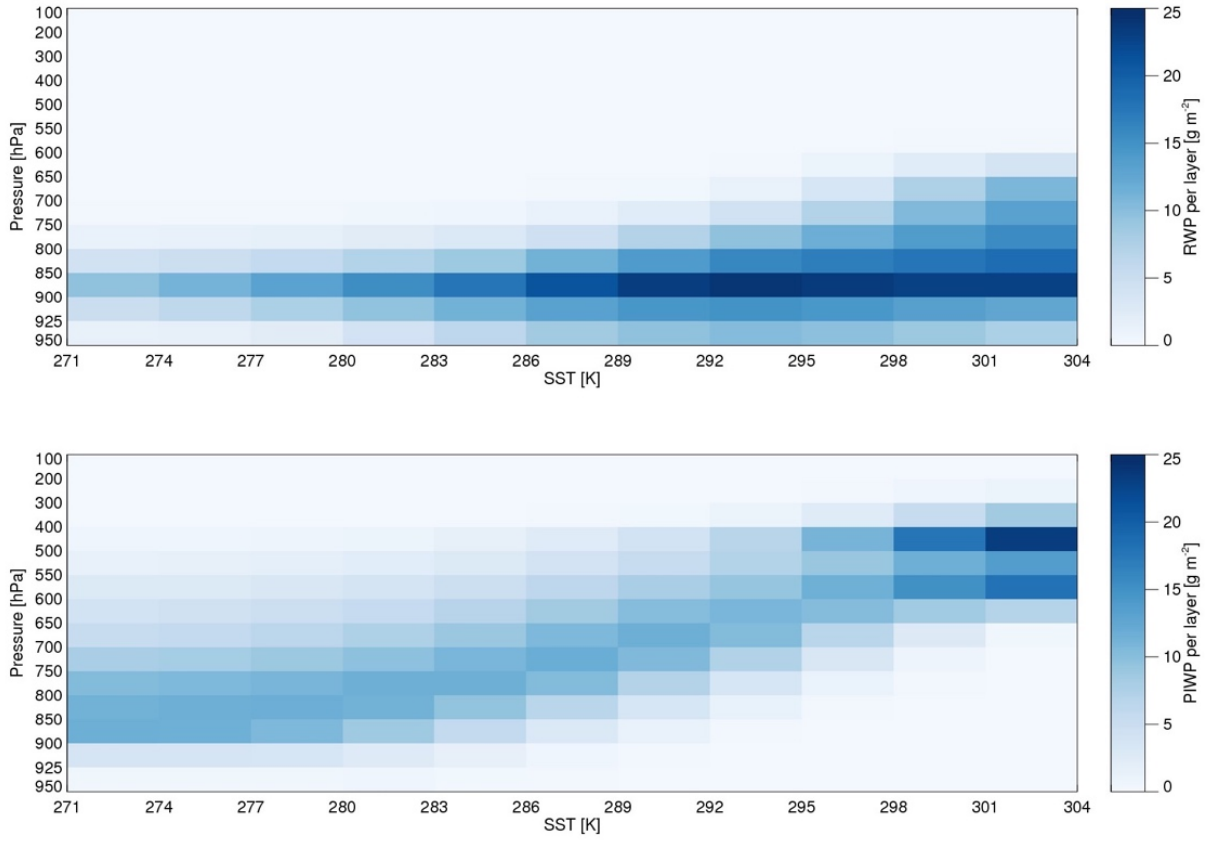

Figure 3.2 First principal components of RWC (top) and PIWC (bottom) from CloudSat for warm rain.

As in Duncan and Kummerow (2016), the liquid cloud layer is fixed in altitude and has a set thickness. Results are largely insensitive to the placement of the cloud layer, partly because the emissivity of cloud drops increases slightly with decreasing temperature, so emission from a

liquid cloud is fairly independent of cloud height or depth. Cloud drops are assumed to have an effective radius of 12 μm.

3.3.3 Drop size distributions

Variational algorithms require realistic estimates of the errors for both the *a priori* state and observation vector. Estimating the error covariances of the *a priori* state is more straightforward, as shown in the previous subsection. Error covariances for the observation vector are more complex, as they should include instrument noise, forward modeling error, and also forward model parameter error as explored by Duncan and Kummerow (2016). Radiometer sensor noise is measured prior to launch and is smaller than forward model error in most conditions for microwave sensors. For a raining retrieval, the assumption of a DSD is likely the largest source of error for the forward model but difficult to quantify because the true DSD is almost never known. This is effectively a forward model parameter error, assuming that the DSD is not somehow retrieved (this topic is explored in the Appendix).

Retrieval of hydrometeors from passive measurements is "notoriously ill-constrained" (Boukabara et al. 2011) largely due to the uncertainty inherent in assuming a DSD. As shown by Lebsock and L'Ecuyer (2011), choosing an inappropriate DSD can greatly impact the final results of a retrieval, as variations in drizzle rates over ocean are largely explained by variations in drop number concentrations (Comstock et al. 2004). Unfortunately, the distribution of drops in the forward model significantly affects the resultant rain rate and has an effect on the T_B vector, but is not readily retrievable—a topic discussed in the conclusions section. In an attempt to circumvent this issue while accounting for the inherent forward model uncertainty of assuming a DSD, such errors are quantified in a way intended to reduce the dimensionality of the problem without ignoring it.

For this study, the normalized gamma distribution is used to characterize rain drop distributions (Ulbrich 1983). This functional form, given below, approximates DSDs found in nature with fidelity (Bringi et al. 2003). The normalized gamma distribution allows comparison of DSDs with different rain rates due to the generalized intercept parameter (N_w) and mass-weighted mean diameter (D_m), with the median volume diameter (D_0) related to D_m via the shape parameter (μ).

$$N(D) = N_w f(\mu)\left(\frac{D}{D_m}\right)^{\mu} e^{-(4+\mu)D/D_m} \quad \text{(Eq. 3.1)}$$

$$f(\mu) = \frac{6}{4^4}\frac{(4+\mu)^{\mu+4}}{\Gamma(\mu+4)} \text{ and } \frac{D_0}{D_m} = \frac{3.67+\mu}{4+\mu} \quad \text{(Eqs. 3.2, 3.3)}$$

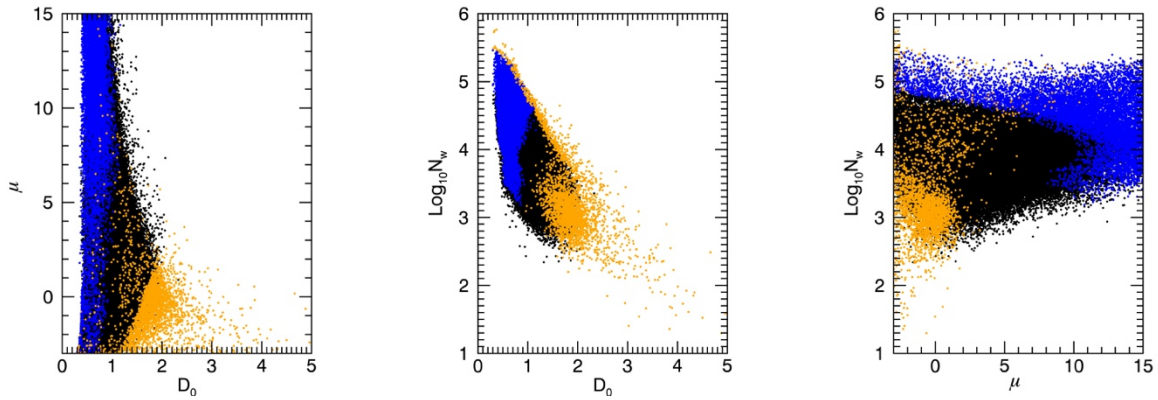

Figure 3.3 Disdrometer data from two high latitude field experiments, analyzed per gamma distribution parameters (Eqs. 3.1, 3.3). Blue and orange represent data points determined via EOF analysis to be similar, analogous to stratiform and convective DSDs, respectively.

In-situ disdrometer measurements from GPM Ground Validation field campaigns were used to quantify the error in forward modeled T_Bs given a range of DSDs. Given the focus on high latitude warm rain the sites chosen were near Seattle and Helsinki, from the OLYMPEX (Houze et al. 2017) and LPVEx (pmm.nasa.gov/science/ground-validation/lpvex) campaigns, respectively. With the parameters that describe a gamma distribution provided along with liquid

water content, the disdrometer data were used in a simple radiative transfer model with Mie scattering modules. The RT model was run with a fixed atmosphere and 150 g m^{-2} liquid cloud from 925-850 hPa, with the disdrometer data setting how much RWC exists between cloud and surface.

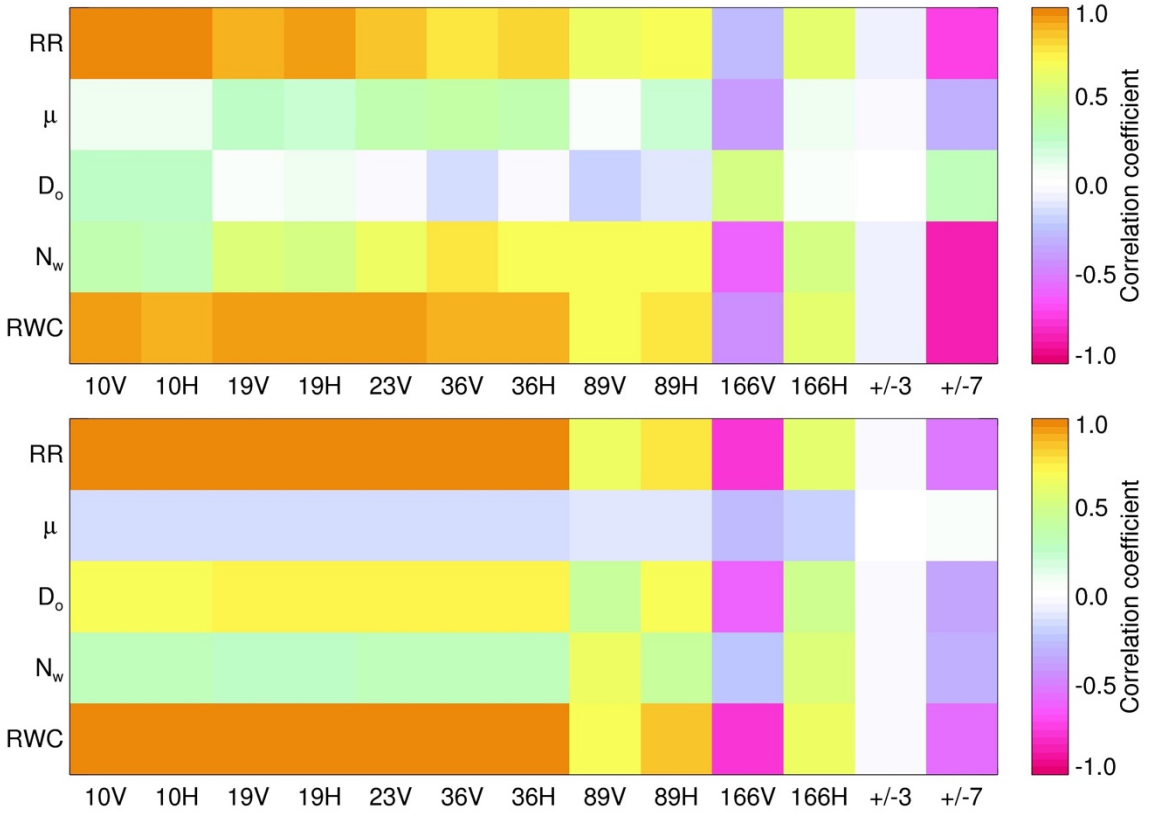

Figure 3.4 Correlations of T_B at GMI frequencies with the gamma DSD parameters (Eq. 1) as well as rain rate and RWC, as derived from disdrometer measurements run through a radiative transfer model. Convective DSDs (top) and stratiform DSDs (bottom) are shown.

From the work of Dolan et al. (personal comm.), PC analysis of the disdrometer data reveals leading modes of variability in the gamma distribution parameters that suggest convective and stratiform regimes of rainfall, colored in Fig. 3.3. Given the different radiometric characteristics of stratiform and convective DSDs as diagnosed by the RT model, the analysis here and the retrieval described later delineate between the two. This functions as a way to simplify the problem without treating all DSDs as the same, in line with there being limited

signal to solve for the DSD but some information related to the DSD existing in the passive measurements. In this formulation, the forward model parameter error stemming from an assumed drop distribution is defined as:

$$\sigma_{conv}(\nu) = \text{stddev}(T_B[\nu,DSD_{conv}] - T_B[\nu,DSD_{actual}]) \quad \text{(Eq. 3.4)}$$

$$\sigma_{strat}(\nu) = \text{stddev}(T_B[\nu,DSD_{strat}] - T_B[\nu,DSD_{actual}]) \quad \text{(Eq. 3.5)}$$

Fig. 3.4 shows the correlation between Tb at GMI frequencies and three of the gamma DSD parameters (μ, D_0, N_w) as well as RWC and rain rate, broken up into stratiform and convective regimes as shown in Fig. 3.3. The correlation between T_B and these parameters confirms what precipitation retrievals have leveraged for decades, namely that RWC is tightly tied to emission at lower frequencies while both scattering and emission influence the higher frequencies. The strong positive correlations between low frequency T_B and RWC reveal why it makes more sense to retrieve RWC than any of the DSD parameters, which exhibit weaker correlations that are more channel dependent. The strong differences in scattering and emission signals from the two DSD types are enumerated via a simple model in Table 3.1.

Table 3.1 Effects on T_B at top of atmosphere from cloud and rain water alone. Surface conditions are SST=281 K and wind=5 m s^{-1}, with water vapor and temperature profiles representative of such an ocean scene. Liquid water path is 100 g m^{-2}, residing between 925-975 hPa, for both rain and cloud water. GMI's 183 GHz channels are not included due to the invariance of water vapor here. All units are ΔK except for the top row which is in K.

	10V	**10H**	**19V**	**19H**	**23V**	**37V**	**37H**	**89V**	**89H**	**166V**	**166H**
Clear Sky [K]	160.16	82.8	178.53	104.79	199.25	204.04	133.86	243.35	192.73	269.36	261.03
Cloud	+0.4	+0.65	+1.02	+1.77	+1.29	+2.73	+5.29	+5.28	+13.09	+1.4	+3.64
Stratiform rain	+0.58	+0.96	+1.73	+3.02	+2.27	+5.02	+10.37	+5.58	+26.11	+0.12	+4.72
Convective rain	+2.1	+3.61	+4.1	+8.3	+4.17	+5.33	+16.37	+2.18	+16.95	+0.29	+2.83
Stratiform rain (emission-only)	+0.59	+0.97	+1.8	+3.11	+2.42	+5.9	+11.45	+15.17	+37.65	+3.1	+8.16
Convective rain (emission-only)	+2.44	+4.02	+6.34	+10.98	+7.7	+13.67	+26.54	+11.69	+29.0	+1.83	+4.79
Stratiform rain (scattering-only)	-0.01	-0.01	-0.07	-0.09	-0.15	-0.88	-1.08	-9.59	-11.54	-2.98	-3.44
Convective rain (scattering-only)	-0.34	-0.41	-2.24	-2.68	-3.53	-8.34	-10.17	-9.51	-12.05	-1.54	-1.96

Fig. 3.5 translates the results from this experiment into error covariance matrices usable by the retrieval. Shown are error covariance matrices calculated for both stratiform and convective DSD observations. The assumed DSD parameters are μ=9 and D_0=0.75 mm for the stratiform case, and μ=-1 and D_0=1.8 mm for the convective case. N_w is solved for given μ, D_0, and RWC. While not the arithmetic means of the observational data, these values are close to the medians and yield near-zero differences with respect to rain rate, N_w versus observed N_w, and difference of T_B at most frequencies when considering the whole distribution of disdrometer observations. This is a key point, as assumed DSD parameters should not bias the resultant distribution of rain rates. Similarly, the distribution of simulated T_Bs should not have a strong bias or skew, as the variational retrieval assumes Gaussian errors from the forward model (Rodgers 2000).

The result of this analysis is an estimate of forward model error at GMI frequencies caused by the assumption of a DSD for rain in each regime. Since this analysis used the observed variability of the gamma distribution parameters for given RWC values, the resultant error covariance matrices can be scaled as a function of RWC in the retrieval without further assumptions. The inclusion of covariances between channels' errors (off-diagonal elements in the matrix) is a key element of this analysis, as many of the errors caused by assuming a DSD are highly correlated between channels, especially channels of similar frequency.

3.4 Retrieval description

3.4.1 Non-raining algorithm

The CSU 1DVAR (Duncan and Kummerow 2016) was originally developed as a non-scattering retrieval for the so-called 'ocean suite' parameters over water: columnar water vapor,

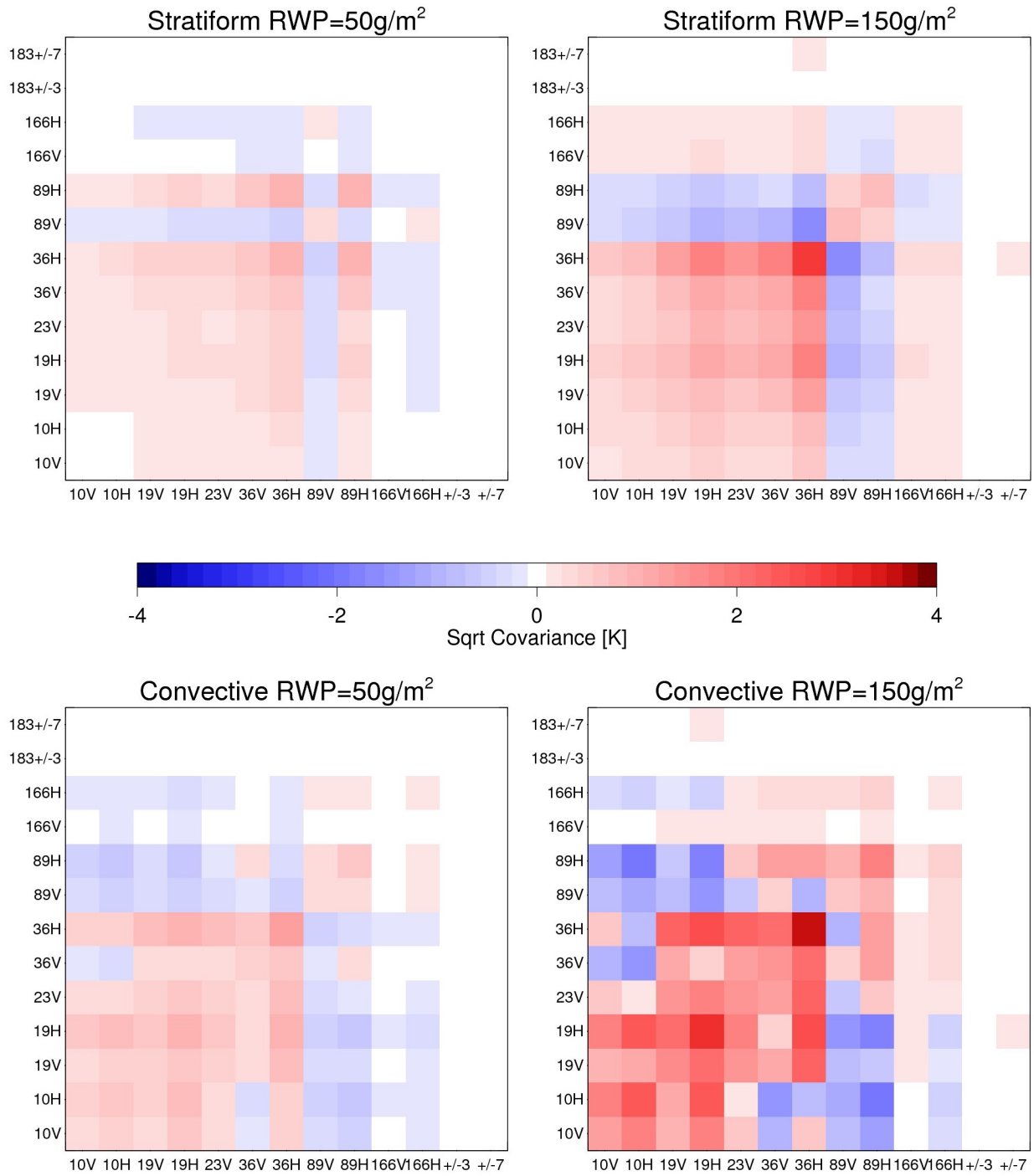

Figure 3.5 Error covariances due to DSD variability at GMI frequencies. Covariances are expressed in units of K, with negative covariances given as $-\sqrt{|\boldsymbol{S}_y|}$, to aid interpretation.

10m wind speed, cloud liquid water path, and sea surface temperature. It is a variational (optimal estimation) algorithm that iterates to find an optimal geophysical state that best matches the observed T_B vector within the bounds of *a priori* knowledge of the geophysical state (Rodgers 2000). This is done via a physical forward model tailored to the radiometric sensitivities of the variables being retrieved, coupled to a radiative transfer model. Mathematically, the iterative process endeavors to find a state vector (x) that minimizes a cost function (Φ) and yields a metric of fit (χ^2) to the observed radiances:

$$\Phi = \big(y - f(x,b)\big)^T {S_y}^{-1}\big(y - f(x,b)\big) + (x - x_a)^T {S_a}^{-1}(x - x_a) \quad \text{(Eq. 3.6)}$$

$$\chi^2 = \big(y - f(x,b)\big)^T {S_y}^{-1}\big(y - f(x,b)\big) \quad \text{(Eq. 3.7)}$$

Here y is the observation vector, f is the forward model, b contains all non-retrieved elements of the forward model, x_a is the *a priori* state vector, and $\mathbf{S_a}$ and $\mathbf{S_y}$ represent the error covariance matrices of the *a priori* and observation vectors, respectively. S_y for the non-raining retrieval is the same as that given in Fig. 2.1. The cost function balances knowledge of the prior state with confidence in the observations to find an optimal retrieved state.

The non-raining CSU 1DVAR solves for six parameters: wind speed, liquid water path, SST, and coefficients of the first three PCs of water vapor. As described in Section 3.3.2, PCs are used to reduce the dimensionality of the water vapor profile. To make the problem more Gaussian, LWP is retrieved in log space but with effectively no constraint by the prior. The *a priori* states for SST, wind, and water vapor come from a global model, as do sea level pressure and the temperature profile.

3.4.2 Drizzle

Conditions in which the non-raining (non-scattering) retrieval converges with a high quality of fit (low χ^2) are not necessarily non-raining for the reasons mentioned in Section 3.3.1.

Thus, if retrieved LWP is greater than the drizzle onset threshold (Fig. 3.1), LWP is partitioned into cloud and rain water. Not all extra water is partitioned into drizzle, with some of the extra water remaining as cloud water as discussed in Wentz and Spencer (1998). Rain water path (RWP) is defined below, with LWP_{driz} determined by SST and TPW state from Fig. 3.1:

$$RWP = \partial LWP * (1 - \frac{\sqrt{\partial LWP}}{\partial LWP}) \quad \text{(Eq. 3.8)}$$

$$\text{where } \partial LWP = LWP - LWP_{driz} \quad \text{(Eq. 3.9)}$$

The resultant rain rate is a function of RWP. Because no information exists on the distribution of drops or their altitude, a simple regression relationship derived from the 2C-Rain-Profile dataset is used to calculate a rain rate, linearly related to RWP. This relationship is subset by SST so as not to mix regimes.

The largely non-scattering scenario with drizzle is quite common, especially at high latitudes, and forms a majority of all scenes with retrieved rain. Shallow clouds with high liquid water contents often converge well in the non-scattering retrieval if there is a lack of significant snow or mixed phase hydrometeors present. The high frequency channels on GMI, 166 GHz and higher, are sensitive to scattering from frozen hydrometeors. Because this retrieval is for warm rain only, a significant departure between observed and simulated T_B at 166V, 166H, and 183±7 GHz is a sign that the warm rain retrieval should not be run.

3.4.3 Warm rain

For cases in which the non-scattering retrieval fails, or cases in which $LWP > LWP_{driz}$ but $\chi^2 > 1.0$, indicating a fit to the T_Bs that exceeds assumed errors, the stratiform retrieval is run subsequent to the non-scattering retrieval. The number of retrieved parameters drops from six to four: PC1 of RWC, PC1 of PIWC, LWP, and PC1 of water vapor. This is necessary due to the limited information content afforded by the T_B vector in raining conditions. Wind and SST from

the non-raining retrieval are held constant; attempting to retrieve wind speed or SST degrades the retrieval of the other parameters by being effectively free variables. Even with four variables, the *a priori* errors on LWP and PC1 of water vapor are decreased so as to discourage unphysical behavior in the retrieval. Raining scenes typically exhibit around 3 DFS given these four retrieved parameters.

A key element of the raining retrieval is its dynamic observation error covariance matrix. In theory, $\mathbf{S_y}$ for a raining pixel should contain all the known uncertainties of the forward model, forward model parameters, and instrument noise. In practice, this means adding the non-scattering retrieval's error covariances with the errors associated with a given RWP. As described in Section 3.3.3, the forward model error caused by assuming a drop size distribution was calculated earlier and is a function of RWP. Though some care needs to be taken to ensure that $\mathbf{S_{y,non\text{-}scat}} + \mathbf{S_{y,rain}}\textit{(RWP)}$ is in fact a positive definite matrix, this sum of matrices yields the observation error covariance matrix used in the iteration. Because RWP is retrieved, the matrix is updated with every iteration. While this complicates the iteration process and typically requires more iterations, it is based in the physics of the situation—heavier rain begets larger uncertainties.

If the raining retrieval fails for the stratiform case, or if it converges with a poor fit, or if simulated T_Bs at 37 GHz and 89 GHz are far too high (indicating insufficient scattering in the forward model), the convective case is run. This is treated the same as the stratiform case—only the DSD parameters and observation errors have changed.

For both the stratiform and convective cases, the resultant rain rate is calculated in the same way. As with the drizzle case, surplus LWP is converted to RWP. The majority of the final rain rate comes from the calculated DSD and resultant rain rate from RWC in the lowest altitude

layers of the forward model. This includes the standard assumption that the drops reach their terminal fall velocity. No explicit evaporation model is included due to the lack of true vertical information.

3.5 Case studies

3.5.1 Comparison to space-borne radars

CloudSat's high sensitivity to light rain rates makes it a useful point of comparison, although the orbits of GPM and CloudSat mean that high quality matchups are somewhat limited. This section includes one case that contains GMI, DPR, and CloudSat observations in the North Atlantic, and one case with GMI and CloudSat off the coast of France.

Fig. 3.6 compares the CSU 1DVAR, GPROF, DPR, and CloudSat rain rates for a coincident overpass a few hundred kilometers west of Ireland on June 1st 2015. The figure's projection orients the CloudSat ribbon horizontally, with CloudSat reflectivities shown at the top of the figure. GPROF and DPR underestimate rainfall whereas the CSU 1DVAR estimates are closer in magnitude to CloudSat, as seen in the line plot within Fig. 3.6. DPR misses the majority of the raining pixels seen by CloudSat, as the reflectivites are generally below DPR's detection threshold. From 48°N to 51.5°N, the CloudSat 2C-Rain-Profile product has a mean rain rate of 1.43 mm hr^{-1} whereas GPROF and DPR NS measure 0.58 mm hr^{-1} and 0.09mm hr^{-1}, respectively. The CSU 1DVAR mean for the same pixels is 1.95 mm hr^{-1}, though there are a few pixels that failed to converge. This is an encouraging result, showing that warm rain from the variational algorithm is of the same order as that from CloudSat.

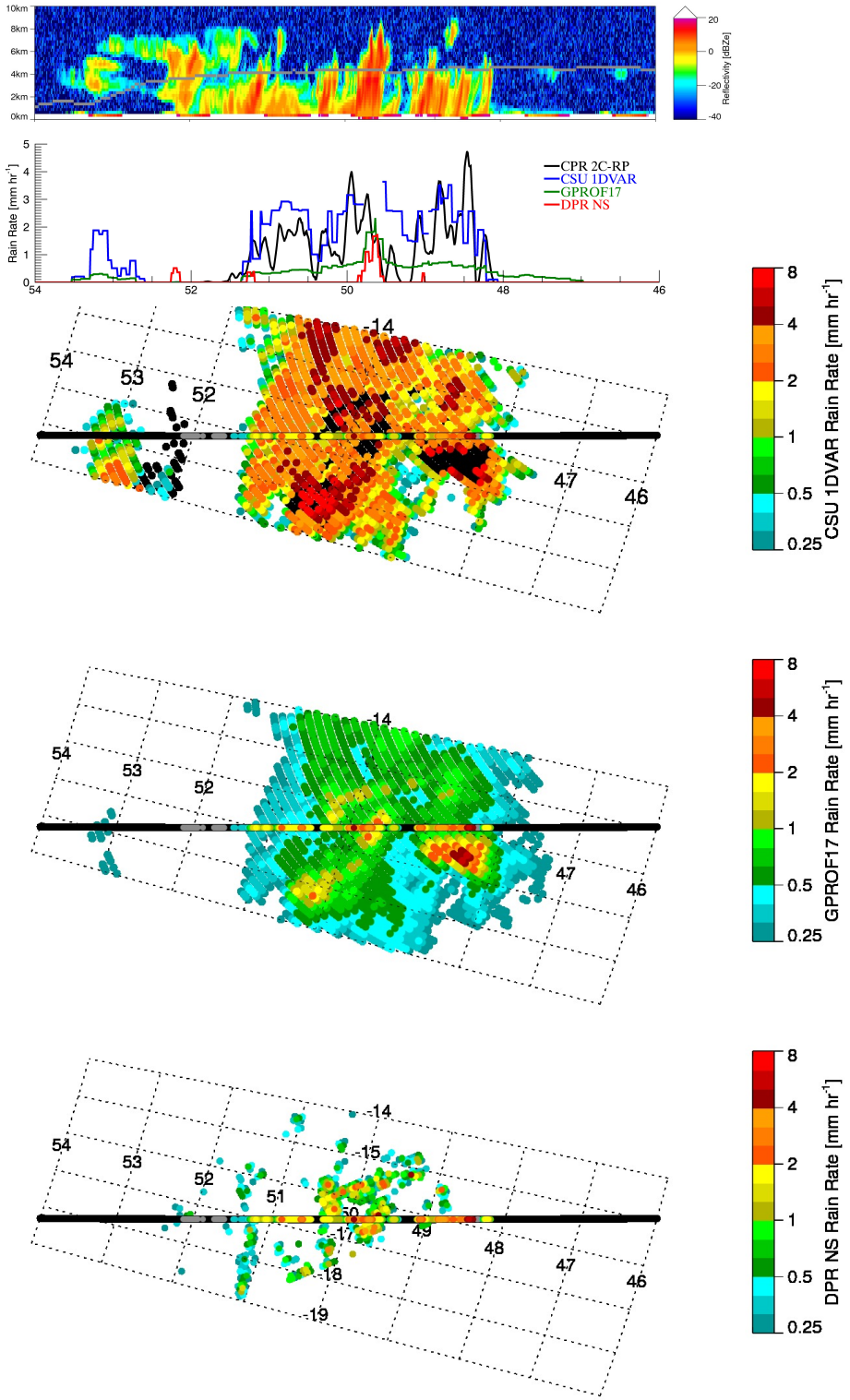

Figure 3.6. GMI and CloudSat rain rates for an overpass in the North Atlantic, June 1st 2015. The top panel shows CloudSat reflectivities with a gray line indicating the freezing level. The second panel indicates rainfall rates from the four products along the CloudSat ribbon. The bottom three panels show estimated rain rates compared with CloudSat, where black indicates no CloudSat rain and gray indicates snow or mixed phase precipitation present.

The freezing level is denoted by a grey line in the top panel of Fig. 3.6, calculated from ancillary data. It is above most of the cloud tops seen by CloudSat, indicating that most of the clouds are probably liquid. The CSU 1DVAR converges for many of these pixels, excepting a few near 52°N and 50°N, where CloudSat shows stronger convection and more cloud above the freezing level. The GPM and CloudSat overpasses were 10.5 minutes apart, which may explain some incongruity in the pixels that converged, especially with regard to convective clouds.

Fig. 3.7 provides a closer look of a raining system in the Atlantic, a scene from March 30th 2016 off the coast of France. In this figure, CloudSat reflectivities show a complex scene with multiple cloud layers and cloud depths ranging from 1km to 8km. The second panel holds profile retrieval results from 2C-Rain-Profile, color coded to differentiate between liquid and ice portions of the cloud, and showing significant rainwater content near the surface that translates into rain rates of about 4 mm hr^{-1}. This is in contrast to the GPROF rain rates, which are all less than 0.5 mm hr^{-1}. As with the previous case, this is not surprising because GPROF's *a priori* database is based upon DPR and most of the CloudSat reflectivities seen from 46°N to 47°N in Fig. 3.7 are too small to be seen by DPR. This raining system is on the edge of the GMI swath, so no direct DPR data are available.

The CSU 1DVAR mostly performs well in this scene. On the right of the figure where the clouds are shallow and mostly liquid, it retrieves rain rates on the order of CloudSat and much higher than GPROF. As the cloud deepens and non-liquid hydrometeors dominate, it fails to converge—the forward model is insufficient due to the transition away from warm rain, and the retrieval performs as it should. In fact, the apparent overestimation of rain rates on the right side of the figure may be due to CloudSat missing some rainwater; GMI senses total columnar

liquid, whereas CloudSat is mostly blind in the lowest kilometer of the atmosphere and thus may be missing some near surface rainwater.

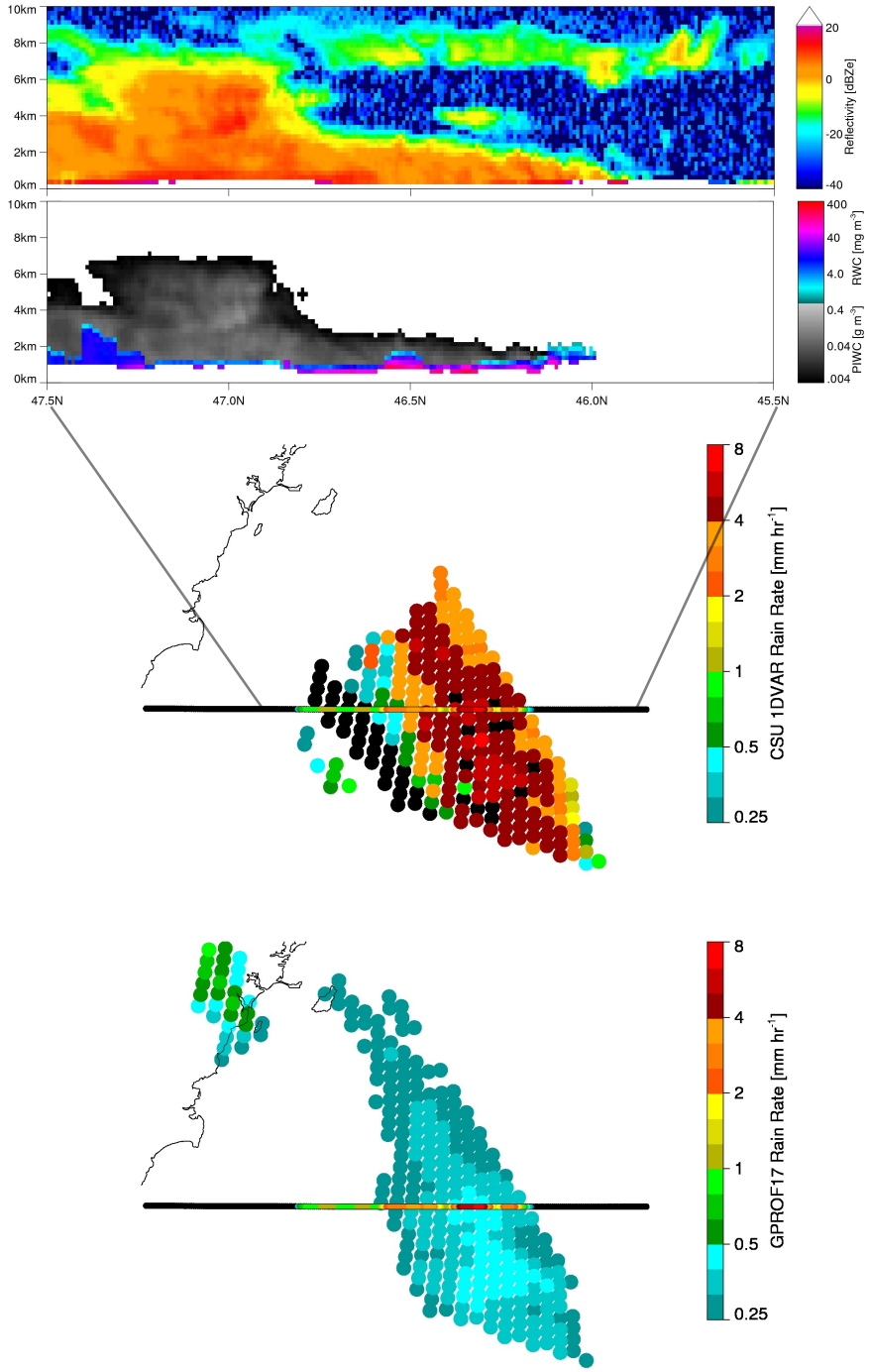

Figure 3.7 Shallow rain and mixed phase cloud off the coast of France, March 30th 2016. The top panel is CloudSat reflectivities while the second panel shows 2C-Rain-Profile RWC and PIWC retrievals from the same scene. The bottom two panels contrast CloudSat rain rates with those of the CSU 1DVAR and GPROF.

However, on the northern edge of the retrieved rain band in Fig. 3.7, there is a transition zone with low retrieved rain rates in an area with moderate CloudSat rainfall. This violates the forward model, but not strongly enough to cause non-convergence. Instead, the scattering signal of mixed phase hydrometeors appears to cancel out the rain drops' emission, and the algorithm reaches convergence with limited rainfall, albeit with a fairly poor fit to the T_Bs. As with the previous case, about 9 minutes elapsed between the overpasses, so the characteristics of the clouds and precipitation may have evolved. These discrepancies at the edges of the rain band signal some limitations of the algorithm and its discrete methodology.

Moving beyond a single case study, two months of the 1DVAR retrieval were compared against DPR rain estimates to assess the representativeness of the two cases analyzed above. Focusing on warm season high latitude precipitation, February 2015 and July 2014 were analyzed for the Southern and Northern hemispheres, respectively. Only pixels within the DPR Ku-band swath and outside the Tropics (defined as 30°N-30°S) were considered, and DPR pixels were averaged into the GMI 23 GHz FOV via a spatial weighting. The results of this experiment are shown in Table 3.2.

Table 3.2 Frequency of raining pixels observed by CSU 1DVAR and DPR Ku-band averaged to the same GMI footprint. All values are given as a percent of all oceanic pixels run by both retrievals in a month that meet the given rain rate cutoff. 'Both' includes pixels where RR_{Ku} > cutoff and either RR_{1DVAR} > cutoff or the 1DVAR did not converge.

	Cutoff [mm hr^{-1}]	Ku rain	1DVAR rain	Ku only	1DVAR only	Both	Either	1DVAR missed
Southern Hemisphere	0.05	10.6%	14.9%	8.1%	10.3%	6.2%	23.0%	2.5%
	0.50	3.5%	10.5%	2.7%	6.8%	3.1%	13.3%	0.08%
Northern Hemisphere	0.05	8.1%	16.6%	5.2%	11.4%	6.4%	21.9%	1.0%
	0.50	3.2%	11.4%	2.4%	7.5%	3.0%	14.0%	0.03%

The results in Table 3.2 are broken up by hemisphere, with percentages of all ocean pixels with rain rates above two cutoff values shown. Cutoffs were necessary as many pixels

display positive, near-zero rain rates from DPR averaging. In both hemispheres and with both rain rate cutoffs, the 1DVAR retrieves rainfall in a significant percentage of pixels where DPR sees no precipitation, approximately 7% of total pixels at a 0.5 mm hr^{-1} rain rate threshold. Pixels counted as '1DVAR missed' are a subset of 'Ku only' raining pixels, counting only if reasonable convergence ($\chi^2<2$) was reached, screening out scenes that deviated strongly from the forward model. The 1DVAR misses a relatively small percentage of definite raining cases from DPR and effectively none at the higher threshold, an encouraging result that echoes Fig. 2.8. Without ascribing truth to either sensor or trying to screen for precipitation type or environment (i.e. warm rain), this analysis points to a complementary relationship between the two sensors' precipitation detection abilities; it does however appear that the 1DVAR sees precipitation too frequently, as Ku-band radar should see most rain rates greater than 0.5 mm hr^{-1}. It should be noted that these precipitation frequencies may not match those of other studies, as GPM's orbit weights this analysis to pixels near 65°N and 65°S.

3.5.2 Comparison to ground radar

In this section two GPM overpasses of the PAIH ground radar are examined. Due to GPM's orbit and the radar's location south of Alaska, it is an ideal location for comparisons between high latitude oceanic GPM observations and a polarimetric ground radar. For this analysis, the focus is on precipitation away from the coastline, as the passive signal from land nearby is a contaminating factor in precipitation retrievals; indeed, the CSU 1DVAR does not run if nearby pixels contain land contamination.

The first case, shown in Fig. 3.8, is from an overpass on July 12th 2015 with scattered showers visible from PAIH. DPR does a fairly good job of seeing these showers, although it

misses some of the lightest raining pixels observed by PAIH. GPROF picks up the strongest area of rain but underestimates the rain rate relative to PAIH and misses the weaker showers.

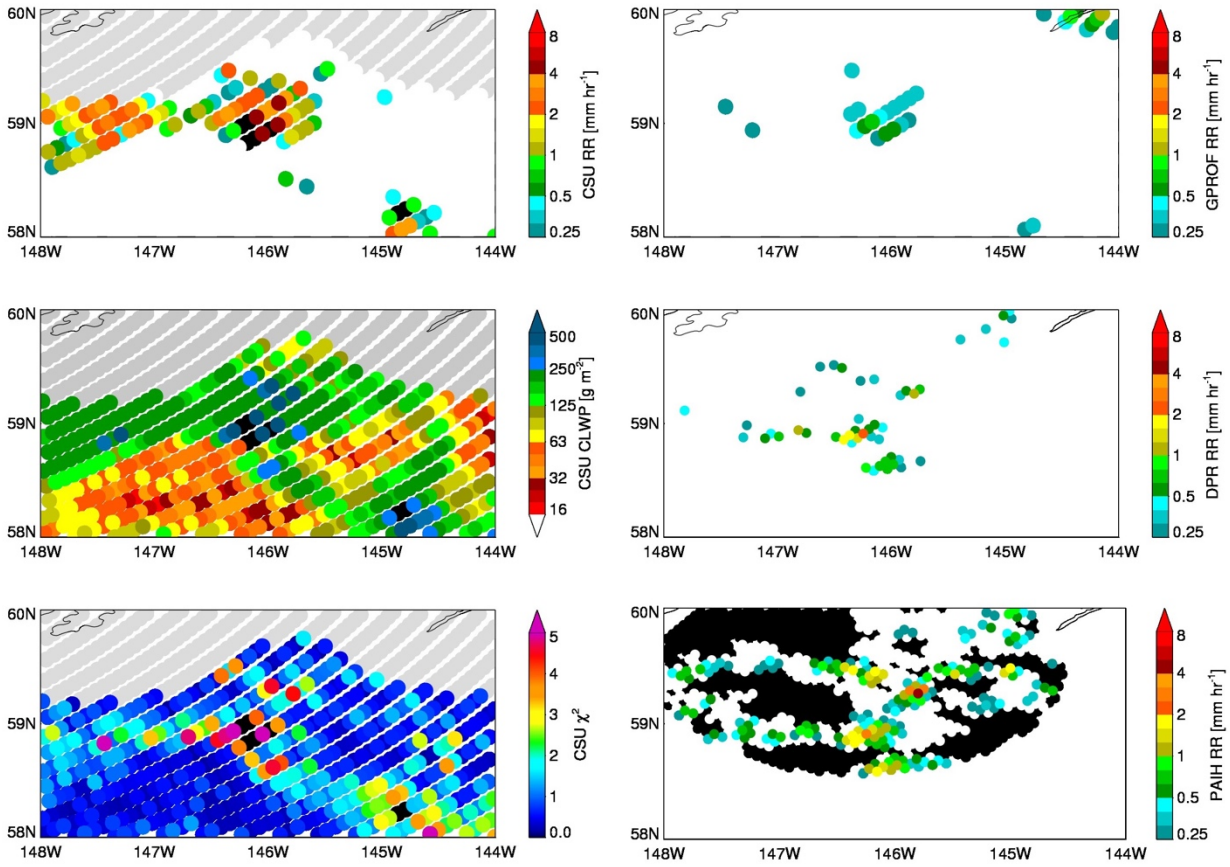

Figure 3.8 Middleton Island (PAIH) radar compared with GPM products and CSU 1DVAR retrievals from July 12th 2015. The left column contains CSU 1DVAR retrievals of rain rate, cloud liquid water path, and quality of fit (χ^2). The right column has rain rates from GPROF, DPR, and the ground radar.

This scene proves challenging for the CSU 1DVAR as well. This region is covered with retrieved liquid cloud, including some pixels above the drizzle onset threshold given in Fig. 3.1 that fit the forward model well. Contrasting these pixels with PAIH, some are not raining to the surface while others are below the drizzle threshold but do indeed seem to be raining. The few pixels raining hardest according to PAIH, DPR, and GPROF do not converge in the iteration, in line with significant mixed phase or frozen hydrometeors present and echo top heights of 3 km to

5 km observed by DPR. So while this scene is nearly ideal for the CSU 1DVAR rain retrieval, in that it rarely violates the forward model, the assumption of a drizzle onset threshold proves to be too simplistic to accurately capture drizzling versus non-drizzling liquid clouds in this scene.

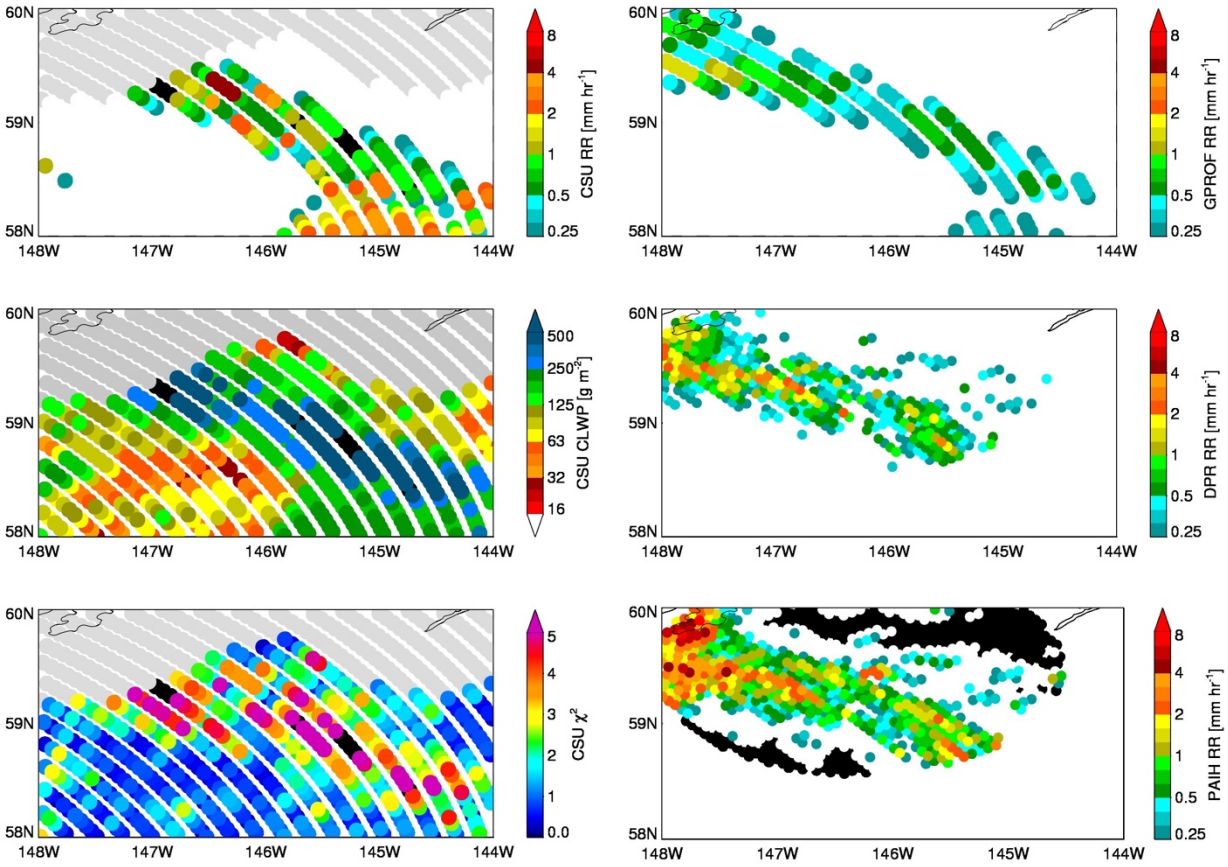

Figure 3.9 Middleton Island (PAIH) radar compared with GPM products and CSU 1DVAR retrievals from June 29th 2015. The left column contains CSU 1DVAR retrievals of rain rate, cloud liquid water path, and quality of fit (χ^2). The right column has rain rates from GPROF, DPR, and the ground radar.

Fig. 3.9 shows a second ground radar matchup with GPM, from June 29th 2015. A stronger band of rain is identified consistently by DPR and GPROF, and they agree on the general magnitude of precipitation, but PAIH is slightly higher. The CSU 1DVAR gets the right general shape of this rain band but mostly overestimates the rain rates compared to the other

estimates. Examination of the fit metric, χ^2, shows that much of this band had relatively poor fits to the observations.

Further analysis of the DPR and PAIH data in Fig. 3.9 shows that the forward model assumptions were violated for many of these raining pixels. DPR retrieved echo top heights of 1.5 km to 4.0 km and bright bands evident in most pixels between 1.6 km and 1.8 km. The existence of these bright bands picked up by DPR demonstrates that there were significant areas of mixed phase hydrometeors present, something not in the forward model. Additionally, most of the raining pixels in the figure reached convergence with the non-scattering forward model but not with the warm rain turned on, so the solution is effectively heavy drizzle but with a poor fit to the observations. This type of mischaracterization is due to the highly limited forward model, but also points to the utility of χ^2 as a marker of trustworthiness for the retrieved parameters.

3.6 Conclusions

There are advantages and disadvantages to the variational approach when it comes to precipitation retrieval. As shown in the comparisons against radar estimates, the retrieval described here compares favorably in some cases and fails to converge in others, sometimes for observations mere tens of kilometers apart. This is a function of the simple forward model's ability or inability to adequately represent all radiometrically significant atmospheric constituents that are associated with oceanic rainfall. However, the simplicity of the forward model is dictated by the limited information content from the observed T_B vector. This is the fundamental catch-22 of precipitation retrieval with limited information.

Ultimately, this study demonstrates two main points. First, that explicit forward modeling of warm rain in a passive-only variational algorithm can indeed work if constructed and

constrained properly. Second, observed microwave radiances can be matched well to modeled radiances in raining scenes if DSD variability is taken into account. It is worth stressing that algorithms which do not account for the forward model error of DSD variability, which often assume uncorrelated channel errors or simply account for sensor noise (e.g. Iturbide-Sanchez et al. 2011), are certainly overfitting to the measured T_Bs.

This study has shown that DSD effects on forward model error can be dealt with, but the other impediments of partitioning liquid water path and the lack of vertical information are the main cause of errors with respect to radar rainfall estimates. A globally-derived drizzle onset threshold can cause high and low biases side by side (Figs. 3.8 and 3.9), as the T_Bs cannot necessarily convey information on cloud lifecycle stage, microphysics, or environmental regime that will affect whether or not a cloud is raining. Similarly, because GMI lacks profile information, there is no evaporation model as in Lebsock and L'Ecuyer (2011), nor a physical model for drizzle rate, and thus very light rain rates are likely underrepresented by the algorithm presented. These aspects could conceivably be improved by more extensive use of ancillary data.

It is implicit in the approach presented that the retrieval of rainfall should be done in multiple steps. Boukabara et al. (2011) include large hydrometeors in the forward model only after non-convergence of the non-raining retrieval, then perform post-processing to calculate a regression-based rain rate (Iturbide-Sanchez et al. 2011) without rerunning the variational scheme. The approach given here differs in that two DSD assumptions are possible for rain in the atmospheric column, whereas Boukabara et al. (2011) does not allow for DSD variability. The reason for the multistep approach presented can be reduced to the importance of the *a priori* state and first guess in an under-constrained retrieval.

The algorithm requires tight constraints on some parameters and a reasonable first guess to converge on a sensible and physical solution. In fact, the raining retrieval is more likely to converge if using the RWP from the GPROF retrieval rather than the mean RWP from analysis of CloudSat data. Additionally, the variational framework requires smooth movement of the retrieved variables while iterating, and therefore abrupt changes to the DSD assumption function better with a multistep framework. If rain water content were allowed in the first pass of the retrieval, alongside the non-raining variables, it would jeopardize their retrieval due to the limited information content of the T_B vector.

Assimilation of satellite radiances is typically done with prescribed and uncorrelated errors, although there has been movement towards including correlated observation errors (Bormann et al. 2011; Bormann et al. 2015). Satellite radiances which diverge too much from those of the background state are not assimilated. To avoid the issues associated with forward model uncertainties in precipitation, all-sky schemes may boost assumed channel variances, but this also blunts the potential information content gained. Fig. 3.10 shows observed minus simulated radiances for two GMI channels, with little difference between raining and non-raining cloudy pixels. This level of agreement would likely have a positive impact on all-sky radiance assimilation schemes. Results given here on the effects of DSD variability on forward model errors may be of interest to the data assimilation community as a way to wring more information out of T_Bs from existing sensors in all-sky conditions.

Furthermore, dynamic adjustment of observation errors ($\mathbf{S_y}$) based on the retrieved scene's characteristics is not commonly done in either retrievals or DA schemes. A proxy for cloud amount is used to determine errors by Lean et al. (2017), a scheme akin to a dynamic error assignment though not specific to DSD assumptions. Interestingly, the largest errors described in

Lean et al. (2017) are at the 19H and 37H channels for GMI, in line with the results of Fig. 3.4 for large RWC. The errors assumed here are, however, multiple times smaller and thus should retain more information content that would otherwise be discarded by many DA schemes.

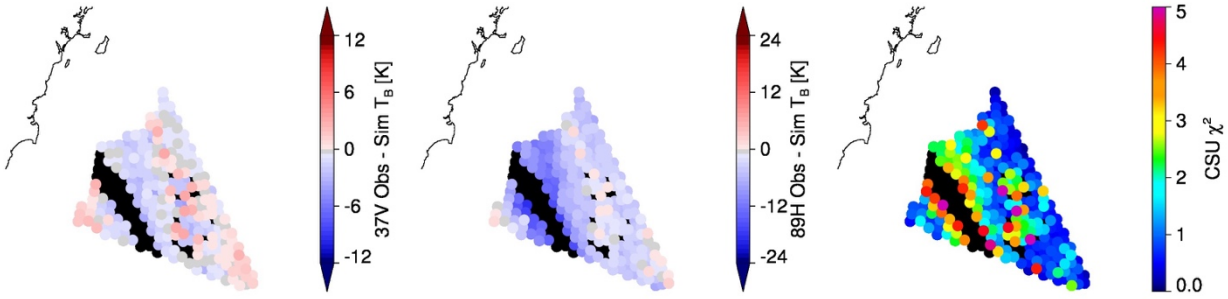

Figure 3.10 The same scene as Fig. 3.7, showing the difference of observations and simulated radiances, specifically at 37V (left) and 89H (middle) channels, and quality of fit χ^2 (right).

The feasibility of a more global version of this retrieval is an open question. The method given in Section 3.3.3 for determining forward model error from assuming a DSD for rain is applicable to any disdrometer dataset, and expanding that analysis to locations that are representative of other oceanic rainfall regimes seems a natural next step. With more disdrometer data from research vessels, a fuller characterization of DSD variability for various environments could yield greater constraints for both passive and active rainfall algorithms. The difficulty would lie in determining which rainfall regime to assume, or adjusting that assumption to one more suitable in the case of non-convergence.

Characterization of mixed phase and frozen hydrometeors in the forward model is perhaps the most difficult aspect of developing a global variational retrieval of precipitation. Mixed phase and frozen hydrometeors have highly variable physical and radiometric characteristics (Skofronick-Jackson et al. 2013; Gong and Wu 2017), making the representation of rain DSDs seem simple in contrast. Ancillary data can inform as to the freezing level and

near-surface dewpoint, or on bulk characteristics such as stability or convective potential, but no environmental variable can reliably yield information on rimed particles or supercooled water drops due to their spatiotemporal heterogeneity. And while hydrometeor profiles may be constrained by coincident active measurements, a passive-only algorithm has to make myriad assumptions. Furthermore, proper treatment of forward model errors from mixed phase and frozen hydrometeors is in no way a simple proposition.

For the stated reasons, a Bayesian retrieval such as GPROF still has advantages over a variational scheme for operational global products. A blended Bayesian/variational approach is conjectured to be preferable for current generation radiometers, as warm rain is handled better with a variational algorithm while anything beyond is better handled by a Bayesian algorithm. Hyperspectral passive microwave sensors could provide better observational constraint for a variational algorithm in the future (Birman et al. 2017), but current sensors' information content dictates that sensing precipitation from a passive satellite platform requires many compromises yet.

CHAPTER 4

AN INTEGRATED EXAMINATION OF THE AMSR2 PRODUCTS OVER OCEAN [2]

4.1 Introduction

Passive microwave radiometry is sensitive to many constituents of Earth's atmosphere and surface, making such measurements very desirable for research and data assimilation alike. Sea ice concentration, sea surface temperature, cloud water content, precipitation, soil moisture, snow water content, and water vapor are a few of the global measurements that passive microwave satellite instruments such as the Advanced Microwave Scanning Radiometer-2 (AMSR2) provide for the Earth sciences community. This is a great strength but also a challenge to invert the observations to retrieve specific geophysical quantities of interest.

The numerous sensitivities of passive microwave radiometry have been known for decades (Staelin 1969) and exploited in earnest via satellite platforms since the late 1970s (Njoku 1982). Observations by microwave imagers and sounders have proven increasingly important for numerical weather prediction (Anderson et al. 1994; English et al. 2000; Geer et al. 2010) as well as research purposes. The Special Sensor Microwave Imager (SSMI) series of sensors ushered in a new era of remote sensing of cloud water, precipitation, water vapor, sea ice, and ocean winds but lacked lower frequency channels suited more for other surface measurements (Hollinger et al. 1990).

Launched in 2002, the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) aboard NASA's Aqua satellite (Parkinson 2003) featured higher

2. Previously published in modified form as Duncan et al. (in review)

spatial resolution than SSMI and frequencies ranging from 6.925 to 89.0 GHz. With its low frequency channels sensitive to soil moisture and sea surface temperature (SST) as well as higher frequency channels sensitive to emission and scattering from hydrometeors, AMSR-E provided a vehicle ideally suited to studying the hydrologic cycle, complemented by other sensors in the A-Train constellation (L'Ecuyer and Jiang 2010). A follow-on sensor, AMSR2 aboard the Global Change Observation Mission Water-1 (GCOM-W1) satellite, extends the data record of AMSR-E but features slightly greater spatial resolution and two further low frequency channels for improved noise mitigation over land (Imaoka et al. 2010). GCOM-W1 was launched on May 17th 2012 by the Japanese Aerospace Exploration Agency (JAXA) and joined the A-Train on June 29th 2012.

A radiometer measures a set of brightness temperatures (T_Bs) for a given scene. To decipher what the satellite sensor sees—to determine what contributes to the T_B vector from the observed scene—is accomplished through the retrieval algorithm. Each science discipline typically developed its own algorithms independently of other disciplines; this occurred for historical reasons as well as physical ones. For instance, inversion of the T_B vector to yield a rain rate or atmospheric profile is very different from determining where the sea ice edge is located. Each discipline contends with its own signal to noise ratio (SNR) and often with a different channel subset. The U.S. AMSR-E Science Team had five separate retrieval algorithms, none of which interacted despite having many of the same sensitivities (Comiso et al. 2003; Kelly et al. 2003; Njoku et al. 2003; Wentz and Meissner 2000; Wilheit et al. 2003). This has understandably led to somewhat divergent efforts.

Due to the numerous sensitivities of passive microwave frequencies, the standard AMSR-E retrieval algorithms used screening criteria and ancillary data to make the respective inversion

problems tractable. However, with an integrated retrieval of all geophysical quantities of interest, the need for screening and ancillary information is more limited, making an integrated algorithm attractive. The integrated approach removes the possibility of contradictory retrievals and helps to limit the number of ambiguous retrievals. This approach is however a compromise of sorts, in that an ideal retrieval would simultaneously retrieve all variables via physical radiative transfer modeling of all radiatively significant species. This has been attempted by others (Boukabara et al. 2011), but current limitations of modeling hydrometeors and surface emissivities at microwave frequencies dictate a more piecemeal approach for the time being.

The remainder of this paper is split into two main parts. The first part gives details of the integrated approach for AMSR2 retrievals. The second part focuses on science objectives over ocean surfaces, namely retrieving light precipitation and detecting the sea ice edge, which have historically been difficult to accurately retrieve.

4.2 Data

The AMSR2 sensor aboard JAXA's GCOM-W1 satellite is a conically-scanning radiometer with frequencies from 6.925 GHz to 89.0 GHz (Imaoka et al. 2010). Its 2.0 m reflector dish enables greater spatial resolution than its predecessor, AMSR-E, which had a 1.6 m reflector. As a microwave imager, it provides excellent spatial coverage of the Earth within its 1450 km swath. GCOM-W1 is in a sun-synchronous orbit at a nominal inclination of 98.2° and altitude of 700 km, crossing the equator at approximately 13:30 local time on ascending orbits as a member of the A-Train constellation.

The input data common to all the science algorithms discussed are the AMSR2 L1R dataset (Maeda et al. 2016). The L1R product contains brightness temperatures, geolocation

information, and land percentage within the satellite's field of view (FOV), among other fields. Crucially, L1R provides T_Bs already co-registered and convolved to the FOV of choice using the well-established Backus-Gilbert method (Backus and Gilbert 1970). It should be noted that T_Bs are never de-convolved in L1R, in that the original spatial resolution of each frequency is never artificially enhanced, a process that inherently adds noise. Thus, for instance, T_Bs convolved to the 10 GHz resolution means that T_Bs are averaged into the 10 GHz FOV from 18 GHz to 89 GHz, but the 6.9 GHz and 7.3 GHz channels remain at their native resolution. These differences are key, as the native resolutions of AMSR2 vary significantly by frequency, as seen in Table 4.1. The use of convolved T_Bs is important for physical retrievals of spatially heterogeneous fields such as precipitation, cloud water, or sea ice (e.g. Rapp et al. 2009).

Table 4.1 AMSR2 sensor characteristics. All frequencies include both vertically and horizontally polarized channels, and 89 GHz has A and B scans. For which algorithms use each channel, O = ocean suite, P = precipitation, I = sea ice, and an asterisk signals that only the V-pol channel is used.

Frequency [GHz]	**6.925**	**7.3**	**10.65**	**18.7**	**23.8**	**36.5**	**89.0**
FOV [km]	62 x 35	58 x 34	42 x 24	22 x 14	26 x 15	12 x 7	5 x 3
NEDT [K]	0.34	0.43	0.70	0.70	0.60	0.70	1.2
Algorithm(s)	O*	-	O, P	O, P, I	O, P*, I*	O, P, I*	O, P, I

In lieu of each algorithm using its own land mask, the integrated algorithms use the L1R land percentage field at the FOV size of choice. The only exception is a mask of Antarctic coastlines, for which more up-to-date maps are used. For the ancillary data used by the ocean suite and precipitation algorithms, see Duncan and Kummerow (2016) and Kummerow et al. (2015), respectively.

In addition to AMSR2 data, this study also uses CloudSat data products. CloudSat is a nadir-pointing radar operating at 94 GHz, a frequency suited to cloud and light precipitation

detection (Stephens et al. 2002). Due to CloudSat's position in the A-Train constellation closely following GCOM-W1, CloudSat observations occur a few minutes later, near the center of the AMSR2 swath. In this study we use the CloudSat 2C-Rain-Profile (Lebsock and L'Ecuyer 2011) and 2C-Snow-Profile (www.cloudsat.cira.colostate.edu/data-products/level-2c/2c-snow-profile) products.

4.3 The Individual retrieval algorithms

The following are very brief descriptions of the three individual algorithms for AMSR2 that are discussed in this study: precipitation, sea ice, and ocean suite. For further information and technical details, see the referenced publications.

4.3.1 Precipitation

The precipitation algorithm for AMSR2 is the latest available version of the Goddard Profiling algorithm (GPROF; Kummerow et al. 2015), GPROF2014v2, the same algorithm used operationally for the Global Precipitation Measurement (GPM) mission (Hou et al. 2014). GPROF performs a Bayesian inversion, meaning that it calculates a Bayesian average precipitation rate by assuming that precipitation in the past has the same radiometric characteristics that it does now. For this to work, GPROF relies upon a massive *a priori* database of T_Bs and hydrometeor profiles from coincident radiometer and radar observations made by the GPM core satellite. The inversion is then simply a weighted average of hydrometeor profiles, and associated precipitation rates, from pixels that had a T_B vector similar to that observed by AMSR2. This is shown in Eq. 4.1 as the expected value (E) of rain rate x given measurement y, forward model $f(x)$, and observation error covariance matrix S_y.

$$E(x) = \sum_i x_i e^{-0.5(y-f(x_i))^T S_y^{-1}(y-f(x_i))} \Big/ \sum_i e^{-0.5(y-f(x_i))^T S_y^{-1}(y-f(x_i))} \quad \text{(Eq. 4.1)}$$

The database is stratified by water vapor and skin temperature to make the prior more representative and to further constrain the retrieval. The weighting includes the distance between the observed T_B vector and profiles in the *a priori* database as well as the inverse of the channel uncertainties—effectively sensor noise plus an estimate of forward model error. GPROF currently uses nine AMSR2 channels in the Bayesian inversion (channels that approximate those on the GPM core satellite): 10V/H, 18V/H, 23V, 36V/H, and 89V/H.

4.3.2 Sea ice

Sea ice concentration (SIC) from AMSR2 is computed by the NASA Team 2 (NT2) algorithm, effectively unchanged from the operational algorithm used for AMSR-E (Comiso et al. 2003), except for adjustment of T_Bs to maintain consistency with the original AMSR-E coefficients (see Meier and Ivanoff (in review) for details). Presence of sea ice on ocean surfaces presents a warm and unpolarized signal against the comparatively cold and polarized ocean background due to ice's higher emissivity. NT2 exploits the differential emission signals of V and H polarizations, with presence of sea ice increasing H-pol T_Bs more than V-pol. This is done via the so-called polarization ratio (PR) and spectral gradient ratio (GR), shown in Eq. 4.2 and Eq 4.3 respectively, with ν denoting frequency and *p* polarization.

$$PR(\upsilon) = (T_{B,V}(\upsilon) - T_{B,H}(\upsilon))/(T_{B,V}(\upsilon) + T_{B,H}(\upsilon)) \quad \text{(Eq. 4.2)}$$

$$GR_p(\upsilon_1, \upsilon_2) = (T_{B,p}(\upsilon_1) - T_{B,p}(\upsilon_2))/(T_{B,p}(\upsilon_1) + T_{B,p}(\upsilon_2)) \quad \text{(Eq. 4.3)}$$

These equations account implicitly for the effect of sea ice temperature on the emission signal but can be sensitive to contamination by weather signals such as precipitation, thick clouds, or high wind speeds (Meier et al. 2015). For this reason, an iterative approach is used to match the observations with a lookup table of modeled values derived from a radiative transfer

model for 12 different atmospheres (Markus and Cavalieri 2009). The NT2 sea ice algorithm uses six of the AMSR2 channels: 18V/H, 23V, 36V, 89V/H.

4.3.3 Ocean suite

The "ocean suite" constitutes a group of parameters that are readily retrievable over water surfaces in the absence of significant precipitation. These include total precipitable water (TPW), wind speed, SST, and cloud liquid water path (LWP). The presence of hydrometeors larger than cloud drops causes scattering, which is difficult to model accurately at microwave frequencies. Thus the ocean suite parameters are successfully retrieved in non-precipitating scenes only. The ocean suite algorithm for AMSR2 is the Colorado State University 1-dimensional variational algorithm, or CSU 1DVAR (Duncan and Kummerow 2016). The variational approach, also known as optimal estimation, blends prior knowledge, observations, and the uncertainties in both to retrieve an optimal solution (Rodgers 2000). The 1DVAR determines a state vector x that minimizes the cost function Φ and yields fit metric χ^2:

$$\Phi = \big(y - f(x,b)\big)^T {S_y}^{-1}\big(y - f(x,b)\big) + (x - x_a)^T {S_a}^{-1}(x - x_a) \quad \text{(Eq. 4.4)}$$

$$\chi^2 = \big(y - f(x,b)\big)^T {S_y}^{-1}\big(y - f(x,b)\big) \quad \text{(Eq. 4.5)}$$

where y is the observed T_B vector, f is the forward model, b includes assumed elements of the forward model, x_a is the *a priori* state vector, and $\mathbf{S_y}$ and $\mathbf{S_a}$ are the observational and *a priori* error covariance matrices, respectively.

The CSU 1DVAR uses the Community Radiative Transfer Model (CRTM v2.2.3) and FAST microwave Emissivity Model (FASTEM-6) to forward model the atmospheric and ocean surface state, iterating to find an optimal geophysical state that best matches the observed T_B vector from AMSR2 (Liu and Weng 2013; Kazumori and English 2015). Unlike many variational retrievals that assume non-correlated observation errors, the CSU 1DVAR takes co-

varying forward model errors into account, as seen in Fig. 4.1. Prior information comes from reanalysis. For the variational inversion the CSU 1DVAR currently uses eleven AMSR2 channels: 6V, 10V/H, 18V/H, 23V/H, 36V/H, 89V/H.

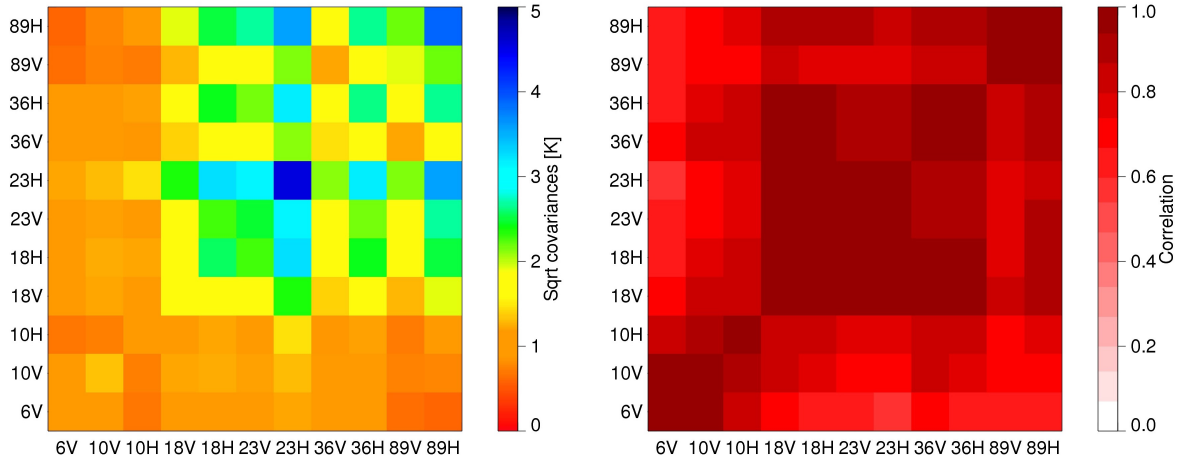

Figure 4.1 Observation error covariance matrix (left) and associated correlations (right) for the AMSR2 ocean suite algorithm, CSU 1DVAR.

4.4 The Integrated retrievals

4.4.1 First pass

The guiding philosophy for integrating the retrievals was to modify historical retrievals as little as possible, keeping the code modular so as to accommodate updates to the algorithms quickly and easily. This was achieved by converting the individual algorithms to callable subroutines for the integrated algorithm. The algorithms are all run using L1R T_Bs convolved to the 18/23 GHz FOV unless otherwise stated.

Precipitation is run first, as neither sea ice nor ocean suite can accurately retrieve in pixels with significant precipitation present. Surface types are initially assigned according to classes developed for the operational GPM algorithm (Kummerow et al. 2015), but can be supplanted and GPROF rerun if the other algorithms disagree with that surface type. Observed

pixels that exhibit precipitation rates greater than 2 mm hr^{-1} with probabilities of precipitation greater than 50% are not run by sea ice.

Ocean suite runs next. Ocean suite is very sensitive to contamination by land or sea ice due to the accuracy expected from the forward model; a sea ice or land intrusion of 1 to 2% may be noticeable in the retrieval, with even side-lobe emission signals outside the pixel's nominal FOV (defined by half power beam-width, see Fig. 4 in Draper et al. (2015)) significant near the sea ice edge and coastal regions. For instance, a 2% sea ice concentration produces approximate jumps in the 6V and 10H channels of +2 K and +3 K compared to open ocean, well within the sensitivity of the algorithm. Because of this, ocean suite uses a very conservative screen for pixels with land nearby. The algorithm runs for all pixels that meet the criteria for being land-free.

Sea ice is run last. The original screens for contamination by weather signals (Comiso et al. 2003) are replaced with retrieved precipitation and cloud values from the other algorithms. The sea ice algorithm is run for pixels with SST below 278 K and 275 K in the Northern and Southern hemispheres, respectively, if the percentage of land within the FOV is below 50%. These SST data come from the ocean suite algorithm or its *a priori* if convergence was not achieved.

4.4.2 Consistency checks

Once the three algorithms have been run—largely without modification—initial inconsistencies are examined (Fig. 4.2). Checks are performed to see if there are any physical inconsistencies, such as nonzero SIC and a converged ocean suite retrieval, or precipitation- and ice-free pixels where ocean suite did not converge.

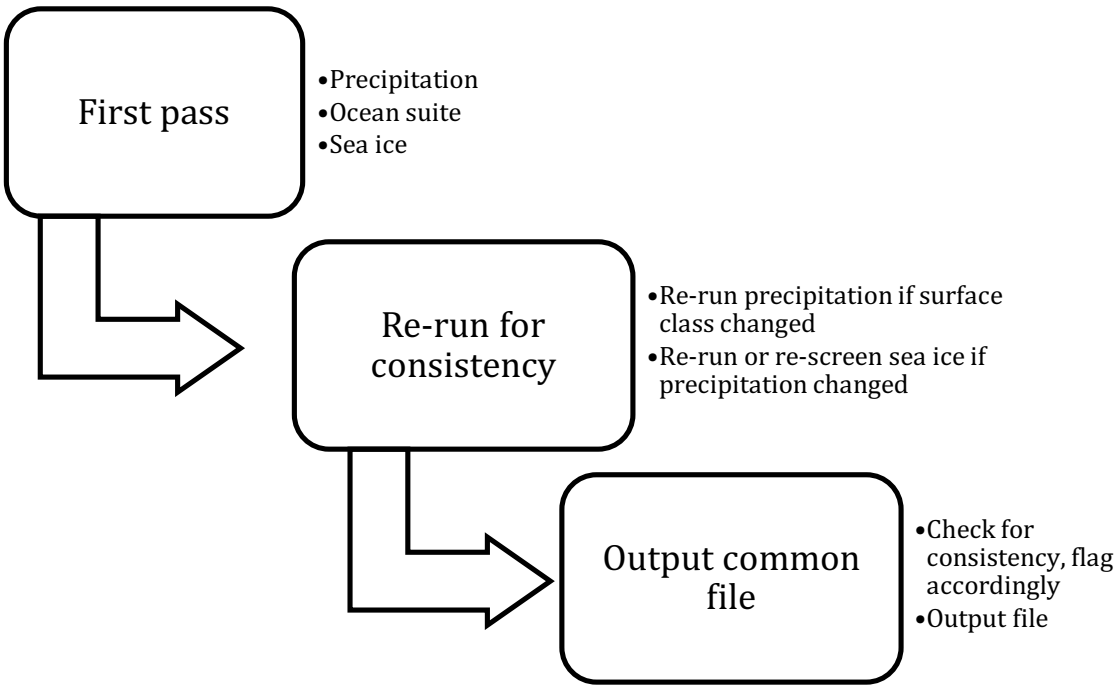

Figure 4.2 Flow chart for processing the integrated retrievals.

The precipitation algorithm is rerun for pixels where the surface classification differs from that of the initial run. This could be due to the sea ice algorithm giving a different sea ice edge than the operational GPM scheme, or if ocean suite converged for a pixel that was not defined as open ocean previously. Because retrieved precipitation is used as a screen for the sea ice algorithm, and retrieved precipitation is dependent upon the surface type defined, the sea ice algorithm is rerun or results re-screened depending on the new precipitation results. This step ensures better physical consistency of the retrieval results, an improvement on each algorithm being run individually.

4.4.3 Ambiguous areas

Beyond the simpler checks for consistency, there are necessarily gray areas where the algorithms disagree but are not contradictory. For instance, due to the CSU 1DVAR's use of the 6V and 10V/H channels, it is sensitive to sea ice outside the nominal 18 GHz FOV used by the

sea ice algorithm. Since the 6 GHz FOV is significantly larger than the 18 GHz FOV (see Table 4.1), there will necessarily be disagreement between the ocean suite and sea ice algorithms as to the location of the sea ice edge. The sea ice algorithm is generally limited to SIC≥15% and can therefore sometimes underestimate sea ice extent when compared to other satellite products (Meier et al. 2015), whereas the CSU 1DVAR is sensitive to SIC of around 1% to 3%. These ambiguous pixels between definite sea ice and definite open ocean are flagged. This topic is examined further in Section 4.6.

Even when uncontaminated by sea ice or nearby land, open ocean pixels can result in disagreements between the ocean suite and precipitation algorithms. This is often due to light or frozen precipitation, but can be caused by radio frequency interference (RFI) or other factors not included in the retrievals' forward model. RFI typically affects the lower frequencies most (Zabolotskikh et al. 2015), causing some pixels near a populated coastline to fail for the ocean retrieval. GPROF over ocean is typically unaffected by RFI due to larger assumed errors for low frequencies, as is NT2 since RFI is rare near the poles. The CSU 1DVAR is very sensitive to RFI because of its small assumed errors at 6V and 10V/H (Fig. 4.1), however any significant RFI will cause the cost function to spike and result in a failed retrieval. RFI is consequently not a main driver of discrepancies found in the integrated retrievals.

Detection of light rainfall is problematic from a satellite platform due to the smooth continuum consisting of liquid cloud, drizzle, and rain. If drizzle droplets are sufficiently small, they produce negligible scattering and are almost indistinguishable from cloud drops at AMSR2 frequencies. This continuum of warm cloud to drizzle is largely linear in T_B space with increasing water content, making it difficult to determine areas of light precipitation definitively. The current version of GPROF is based upon observations from the GPM space-borne radar, the

Dual-frequency Precipitation Radar (DPR; Hou et al. 2014). Any precipitation below the sensitivity limits of DPR, approximately 0.2 to 0.5 mm hr^{-1} depending on regime, is therefore not included in the *a priori* database of GPROF. In contrast, the CSU 1DVAR will readily ascribe light drizzle to cloud liquid water if the liquid water path is not constrained. These issues are further explored in the following section.

Detection of snowfall is even more difficult for a sensor like AMSR2, as the scattering signal from frozen hydrometeors can be small (Bennartz and Bauer 2003; Skofronick-Jackson et al. 2013) and highly dependent on ice particle type and distribution (Bennartz and Petty 2001). In addition, supercooled water can suppress the scattering signal from frozen hydrometeors. These properties are useful for seeing through most ice clouds, but make snow detection difficult.

Precipitation detection is important for the integrated retrieval as both ocean suite and sea ice retrievals are compromised by certain precipitation intensities, though each has different sensitivity to precipitation. Even heavier precipitation presents difficulty, as it manifests differently in the precipitation and ocean suite algorithms depending on regime. From the continuum of increasing precipitation rate and probability from the Bayesian averaging of GPROF, and decreasing quality-of-fit (increasing χ^2) values from the ocean suite algorithm, there is no ideal method to demarcate which pixels are precipitating or not. For this reason, these diagnostic variables are output and some degree of interpretation is left to the user.

4.5 High latitude precipitation

Tropical precipitation has received a lot of attention from the satellite remote sensing community due to the fact that a large majority of global precipitation falls in the Tropics (Kummerow et al. 2000). However, as work continues to better characterize precipitation around

the globe, light rain, mixed phase precipitation, and snowfall at high latitudes are gaining attention (Hou et al. 2014; Behrangi et al. 2016). Remote sensing of precipitation at high latitudes is a great challenge for passive (Munchak and Skofronick-Jackson 2013) and active (Kulie and Bennartz 2009) sensors alike, even given the observational advances brought by the launch of the GPM core satellite in February 2014 (Skofronick-Jackson et al. 2017). The following two case studies demonstrate the research utility of integrated AMSR2 retrievals over the ocean as related to precipitation at high latitudes.

4.5.1 Shallow rain

The first case study is of light rainfall from warm, shallow clouds near the Faroe Islands in the north Atlantic on June 27th 2015. Shallow, warm rainfall is challenging to sense from a satellite platform but also problematic for models (King et al. 2015), and thus worthy of investigation here. Fig. 4.3 shows shallow rain near the Faroe Islands as seen from CloudSat and AMSR2. The CloudSat curtain image from the 2C-Rain-Profile product exhibits liquid clouds with tops near 2 km. The rain rates retrieved by CloudSat should be detectable from GPROF, and indeed GPROF does detect rain in the two areas where CloudSat sees >1 mm hr^{-1} rain rates. However, GPROF underestimates these rain rates by approximately 50% to 80%. This is symptomatic of Version 4 GPM products, in that they significantly underestimate precipitation at high latitudes due to an excessive reliance on DPR precipitation, which misses much of this precipitation (Skofronick-Jackson et al. 2017).

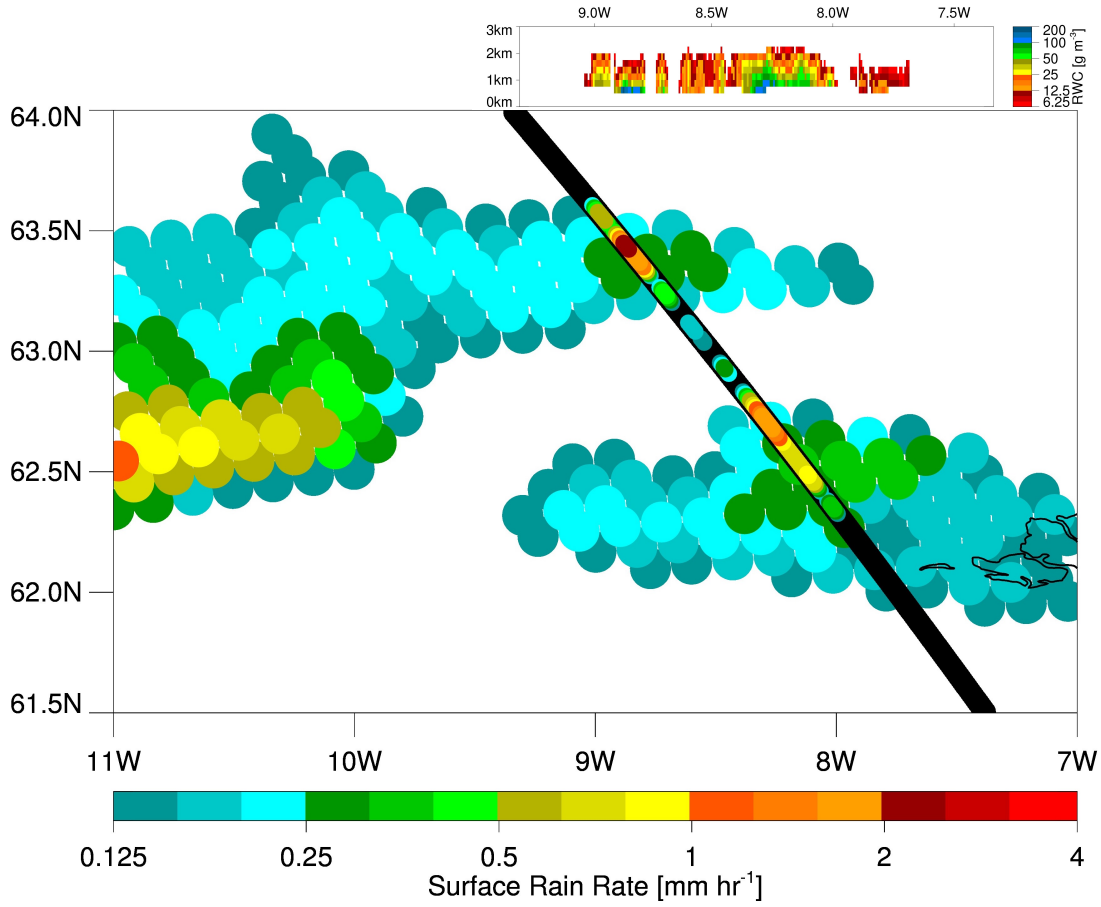

Figure 4.3 Warm rain near the Faroe Islands, June 27th, 2015. The colors denote surface rain rates as estimated by the precipitation algorithm for AMSR2 and the CloudSat 2C-Rain-Profile algorithm. The associated CloudSat profile of rain water content (RWC) is shown in the top right.

Fig. 4.4 shows the perspective from ocean suite, with observed and simulated T_B matching reasonably well. These T_B residuals are scaled by the diagonal elements from Fig. 4.1 to better convey fit relative to assumed forward model error. Apart from an area of heavier precipitation on the west side of the figure, simulated and observed T_B are within about 2σ of assumed errors. The quality of forward model fit in areas shown to be raining from CloudSat demonstrates that the drop sizes are small enough to be largely non-scattering. The viability of retrieving precipitation from non-scattering drizzle such as this with the variational retrieval is discussed later in Section 4.7.

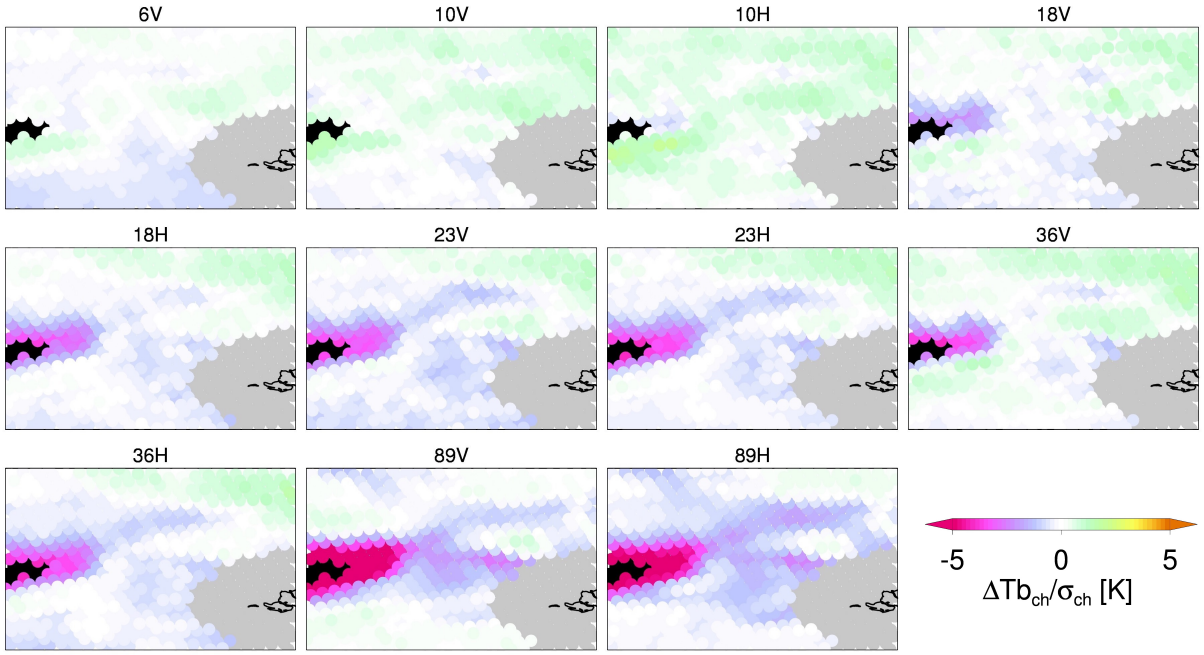

Figure 4.4 Ocean suite T_B residuals (observed minus simulated T_B) for the eleven AMSR2 channels used in the retrieval for the warm rain case, June 27th, 2015. Values are scaled by the assumed forward model uncertainty, seen in Fig. 4.1.

Compared to the following case from the Southern Ocean, this scene near the Faroe Islands is relatively uncomplicated. With SST around 280 K, wind speeds of 4 to 10 m s^{-1}, and the islands themselves accounted for, the scene should not be a tough task for either retrieval. However, although the shallow rainfall is detected by the precipitation algorithm, it is underestimated relative to CloudSat. One possible reason for this is that the model-derived TPW field used within GPROF to subset the *a priori* is 3 to 7 mm lower than retrieved by ocean suite (not shown). Ocean suite TPW may be biased high due to the clouds present, but a low-biased TPW used by GPROF could explain some of the rainfall underestimation (Kummerow et al. 2015). Another possible reason is that some of the profiles included in the Bayesian average were below the DPR detectability limit, bringing the average down.

For this scene, the integrated perspective shows that GPROF is largely successful in detecting precipitation but its estimate is biased low relative to CloudSat. The CSU 1DVAR

retrieves thick, largely non-scattering clouds because it does not partition columnar liquid water into cloud and drizzle. To improve rain rate estimates from GPROF, a more accurate TPW value could be used, especially for cases in which retrieved TPW exceeds that of a saturated column. This could apply even for shallow rain that DPR may struggle to detect or differentiate from ground clutter. The integrated retrievals show that future improvements for detection and estimation of warm rain at high latitudes may be aided by running a variational retrieval in tandem with GPROF, an idea explored in Section 4.7.

4.5.2 Snow

The second case study is from an overpass of a Southern Hemisphere extratropical cyclone on Oct 31st 2014 near the South Georgia and Sandwich Islands. This case presents a frontal system occurring near the CloudSat-observed southern hemisphere precipitation maximum around 55°S, a region in which observational precipitation datasets and reanalyses alike have marked disagreements (Behrangi et al. 2016). Containing high wind speeds, the Antarctic sea ice edge, and precipitation in multiple phases, this region offers a complex case to analyze with the integrated retrievals.

Fig. 4.5 shows the ocean suite retrievals and precipitation retrievals, along with the CloudSat intersect and sea ice boundary. The vertical cross-section from the 2C-Snow-Profile product (Fig. 4.6) shows that there are many CloudSat pixels whose total IWP is 0.5 kg m^{-2} or more, values that are at least theoretically visible from an AMSR2-like sensor (Skofronick-Jackson et al. 2013), depending on predominant ice habit. GPROF completely misses the snowfall seen by CloudSat, which exceeds 1 mm hr^{-1} in some areas. CloudSat determines pixels further north in the transect to be either mixed phase or non-precipitating, whereas GPROF assigns light snowfall albeit with middling probabilities of precipitation.

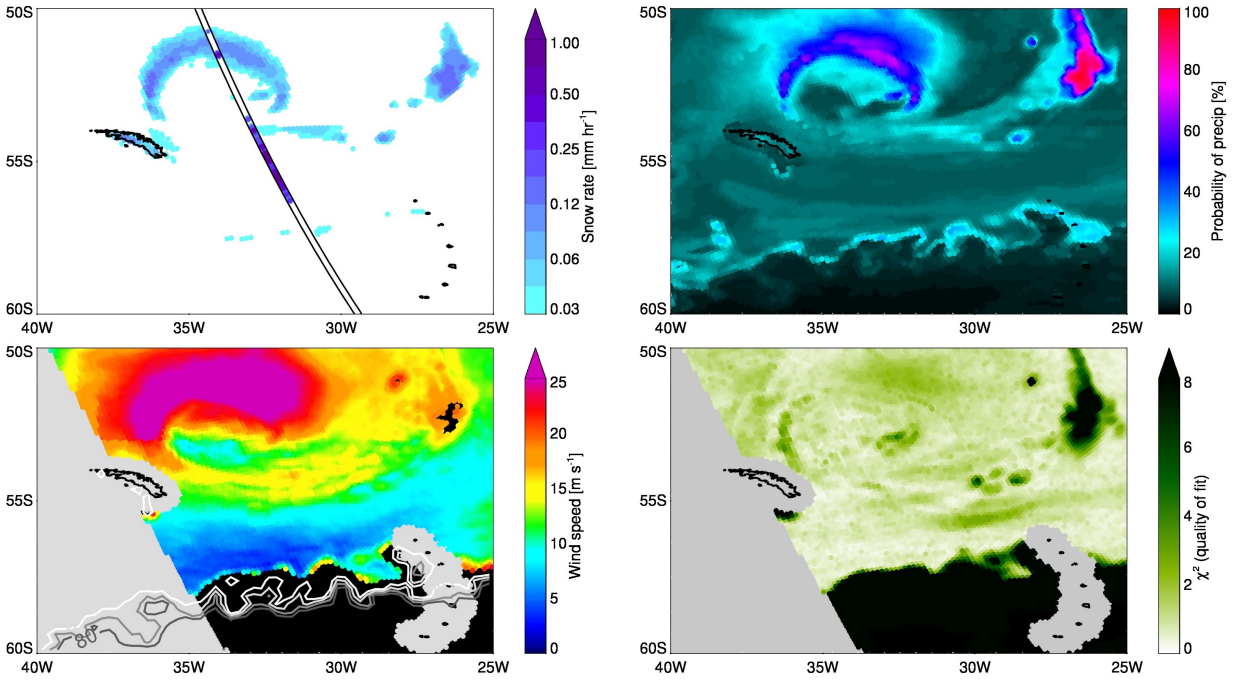

Figure 4.5 Snow and mixed-phase precipitation near South Georgia and the Sandwich Islands in the South Atlantic, Oct. 31st, 2014. Panels show: snow rate from AMSR2 and CloudSat (top left), probability of precipitation from GPROF (top right), wind speed with SIC contoured at 20% intervals (bottom left), ocean suite quality of fit (bottom right).

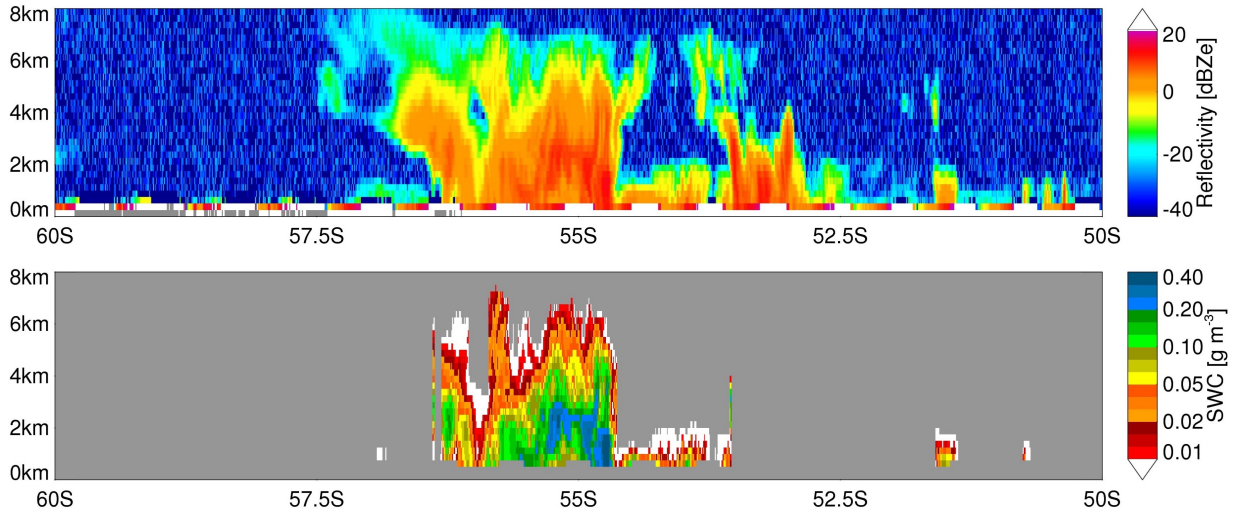

Figure 4.6 Profile data from CloudSat 2B-GEOPROF and 2C-Snow-Profile products. Top panel shows radar reflectivity and bottom panel shows snow water content (SWC).

The ocean suite retrieval finds high wind speeds of 20 to 30 m s^{-1} on the north side of the cyclone and converges fairly well throughout the scene, apart from an area of heavier precipitation north of the Sandwich Islands. This is a difficult scene for ocean suite to parse,

evidenced by what appears to be a spatial artifact just north of South Georgia Island where a sharp change in the *a priori* is manifest. The gray pixels on the west of the figure were not run by ocean suite due to standard sun glint screening. Fig. 4.7 shows T_B residuals scaled by assumed forward model errors. Low simulated T_B at low frequencies in the vicinity of high winds points to an emissivity model bias that could be corrected. Significantly, the snow band observed by CloudSat appears only in the 89V T_B residual and not very strongly—it is similar in magnitude to low frequency emissivity model errors, underscoring the difficulty of snowfall retrieval with AMSR2.

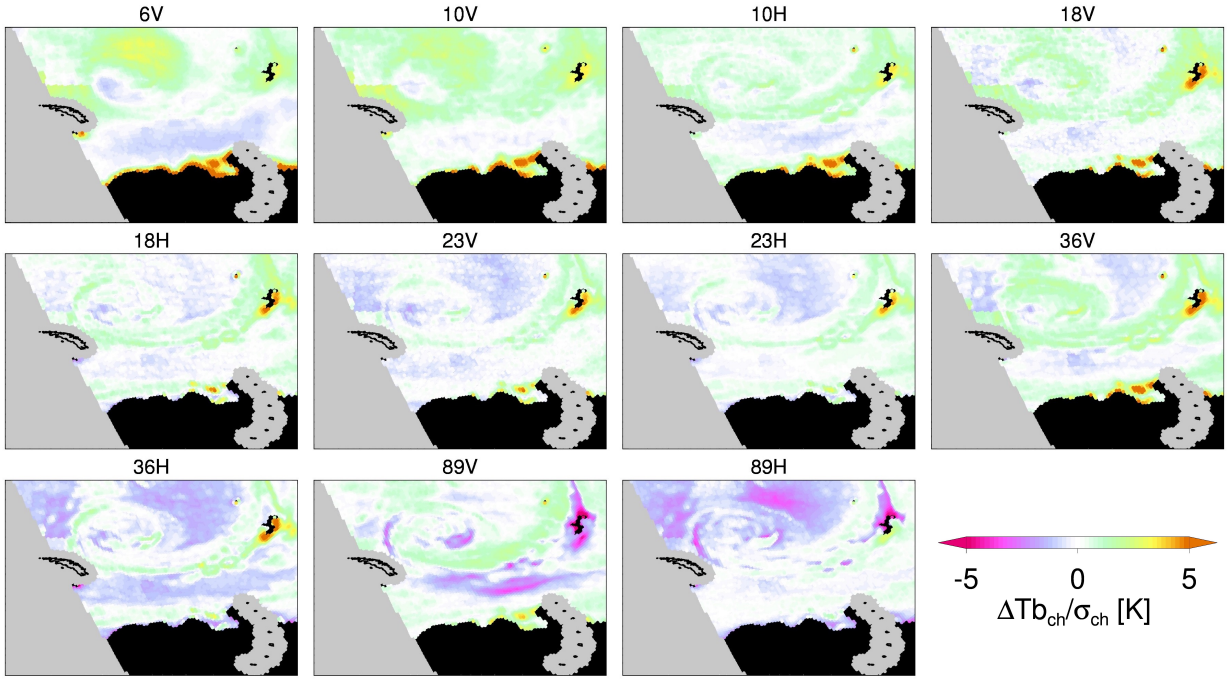

Figure 4.7 Ocean suite T_B residuals, observed minus simulated T_B, for AMSR2 channels used in the retrieval for the snow case, Oct. 31st, 2014. Values are scaled by the assumed forward model uncertainty, seen in Fig. 4.1.

It appears that GPROF is aliasing high wind speeds on the north side of the cyclone to a precipitation signal and missing a significant snow band on the south of the cyclone. The latter is an area where CloudSat is confident in retrieving snowfall whereas the northern part of the cyclone is a region where CloudSat sees no definite precipitation and ocean suite converges

fairly well, although CloudSat does not converge due to it being a mixed-phase regime. The snow on the south side of the cyclone missed by GPROF is, however, consistent with theory in that it is near the edge of detectability at AMSR2's high frequency channels for a deep, synoptic cloud system (Skofronick-Jackson et al. 2013).

Interestingly, GPROF's probability of precipitation field is at its minimum in the region CloudSat shows it snowing hardest. The nature of GPROF's Bayesian averaging explains both why the snow signal is given such a low probability of precipitation and why a high wind speed might be associated with precipitation. A large majority of snowing scenes at high latitudes are missed by DPR (Munchak and Skofronick-Jackson 2013), the sensor providing GPROF's *a priori* database of hydrometeor profiles, and GPROF currently contains no on/off precipitation determination (Kummerow et al. 2015). So, it is unsurprising that the result yields a low probability of precipitation and consequently miniscule snowfall rates due to the averaging.

The relationship between high wind speeds and precipitation from GPROF may come from two effects. GPROF searches the *a priori* database for all profiles within the model-derived TPW/SST regime and then weights these profiles by the T_B vector alone. This weighting implicitly includes a large range of wind speeds, as the profiles originally come from observations by the GPM core satellite that necessarily include a range of wind speeds. Therefore, since the database is unaware of wind speeds and resultant ocean emissivities, the best-matching profiles may be profiles containing a mixture of liquid water emission and scattering from snow, not inconsistent with a mixed phase cloud and high wind speeds. Secondly, if many profiles in the database were cases in which high wind speeds were coincident with precipitation seen by DPR, the simple correlation will be manifest in the result since GPROF does not account for wind speeds explicitly. This represents a clear case in which the integrated

retrievals could benefit one retrieval specifically, as there is potentially useful information available that is not currently used.

A key finding here is that GPROF lacks sensitivity in certain regimes, but there remains information content in the T_B vector that can be utilized. Although the scattering characteristics of precipitating ice vary widely by species (Skofronick-Jackson et al. 2013) and are difficult to forward model with fidelity (Bennartz and Petty 2001), the observed T_B vector's departure from an emitting/absorbing atmosphere is notable and could plausibly be brought to bear for improved precipitation retrievals. Certainly, accounting for surface emissivity changes due to wind speed could improve the Bayesian average, as could the inclusion of hydrometeor profiles such as those observed by CloudSat that exist below DPR's sensitivity limit.

4.6 Sea ice edge

Passive microwave measurements have been a critical component of sea ice monitoring efforts for nearly four decades (Comiso and Nishio 2008). Polar regions can have very persistent cloud cover, with average cloud fraction of about 74% over the Arctic ocean (Chernokulsky and Mokhov 2012), rendering microwave imagery invaluable for detecting sea ice extent in the Arctic and Antarctic, complementary to infrared and visible polar-orbiting satellite sensors.

The large footprint of passive microwave sensors can smear out significant areas of sea ice or miss smaller areas of open ocean (Meier et al. 2015). AMSR2 features lower frequency channels and a higher effective spatial resolution than most microwave imagers, making it a good testbed for research on improved methods for sea ice edge detection. The two cases that follow exemplify difficulties in sea ice edge detection from a microwave sensor which benefit from integrated retrievals and independent observations to better understand the full picture.

4.6.1 Sea of Okhotsk

Fig. 4.8 shows a visible image from the Japanese Meteorological Agency's Himawari-8 geostationary satellite (Bessho et al. 2016) from March 29th, 2016 at 2:50Z, near the island of Hokkaido. This image presents a compelling case for an integrated retrieval, containing swirls of sea ice in the Sea of Okhotsk and clouds at various levels. Fig. 4.9 shows approximately the same domain as seen from AMSR2, which passed over between 2:43Z and 2:48Z, with SIC, χ^2 from ocean suite, and outlines of retrieved liquid clouds shown. This figure contains pixel-level retrievals averaged together on a 0.15° grid to aid interpretation, as AMSR2 observations are dense near the swath's edge.

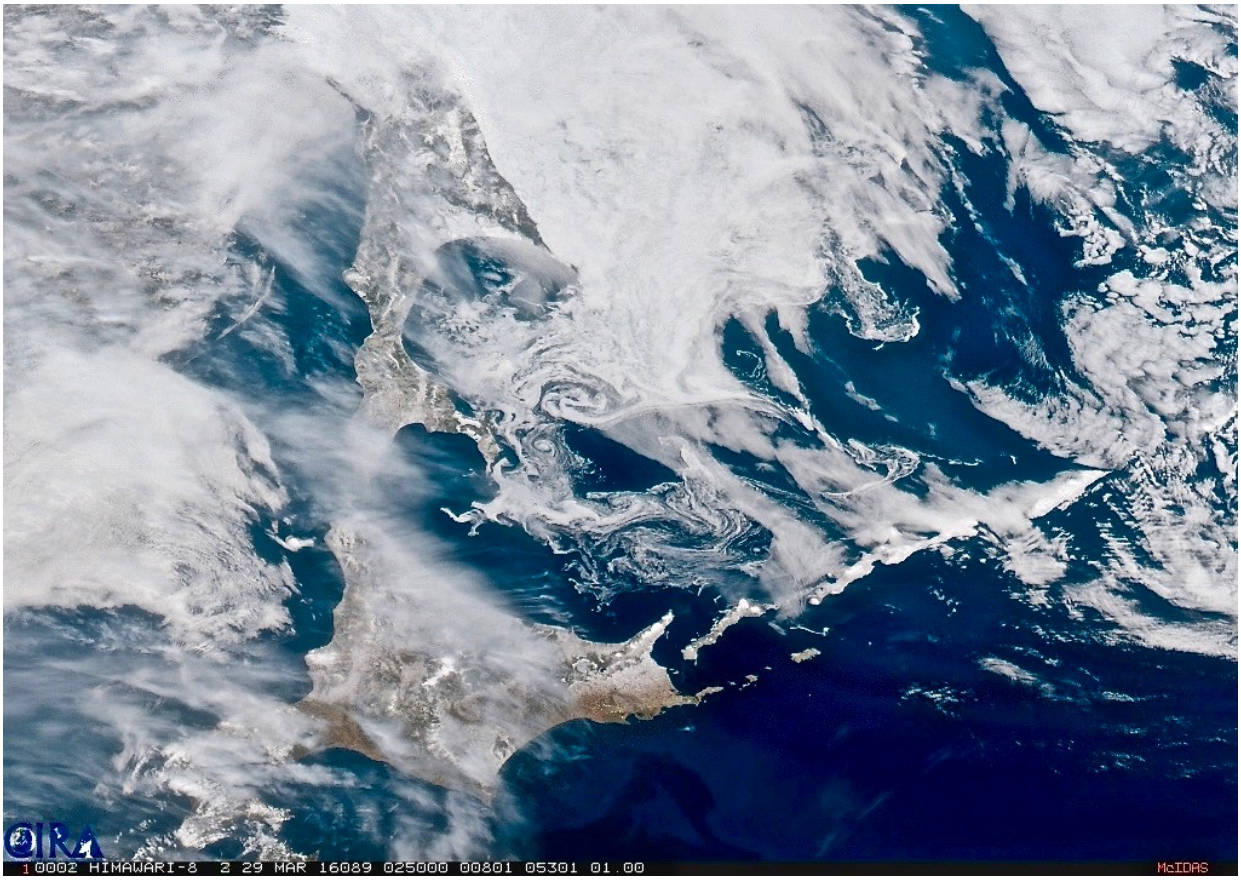

Figure 4.8 Visible image from the Himawari-8 geostationary satellite sensor, March 29th, 2016 at 2:50Z. Image courtesy of JMA/NOAA/CIRA.

As seen in Fig. 4.9, the sea ice and ocean suite retrievals effectively parse the emission signals from sea ice, liquid clouds, and areas with clear sky or thin ice clouds. Some areas near the ice edge with no ocean suite convergence or high χ^2 values deserve a closer look, as some of these appear to be swirls of sea ice in the Himawari-8 image. There was no significant precipitation in this scene from GPROF, hence the focus on sea ice edge alone.

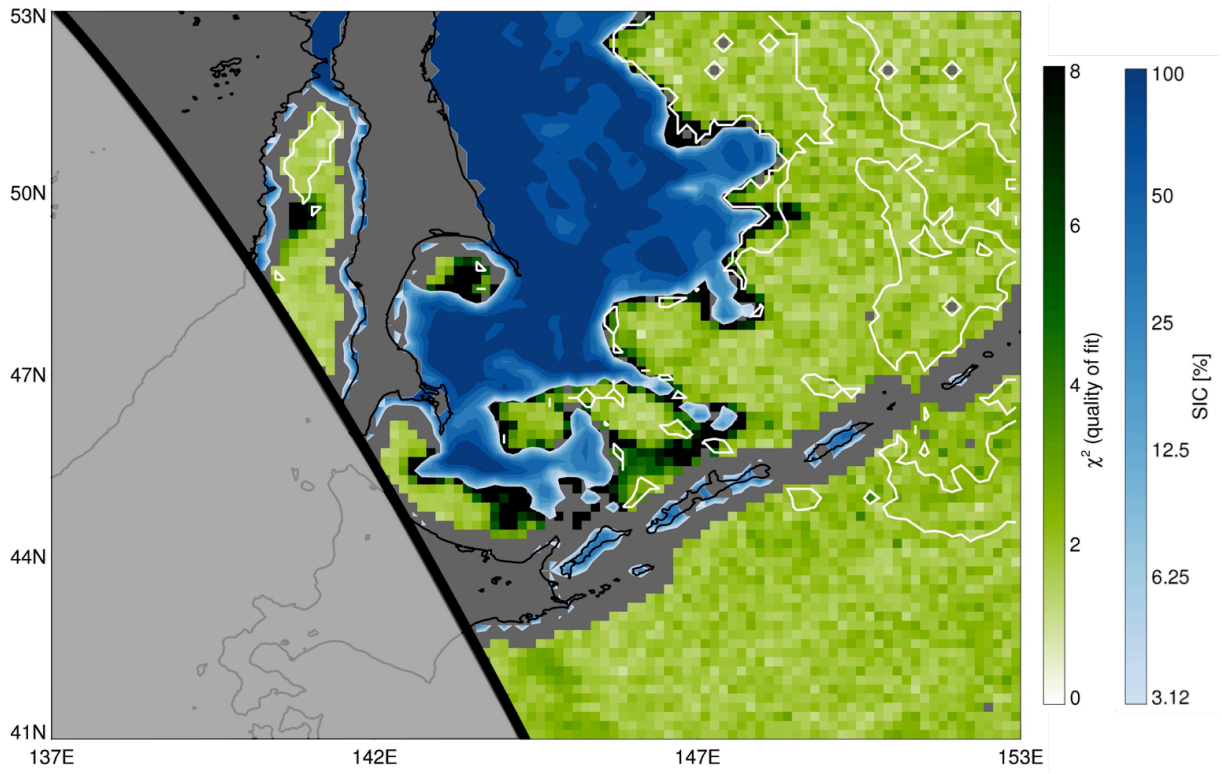

Figure 4.9 AMSR2 overpass of approximately the same region shown in Fig. 4.8, the Sea of Okhotsk on March 29th, 2016 at 0245Z. Ocean suite χ^2 is shown in green, SIC is shown in blue, and liquid clouds are shown in white contours.

In an attempt to resolve the sea ice edge more accurately, the 1DVAR is run again, but this time with no 6 GHz or 10 GHz channels and SST now fixed in the iteration. This limits the impact of emission signals from outside the nominal 18/23 GHz FOV and should yield better agreement with the sea ice algorithm. We then zoom in to the swirls of sea ice northeast of

Hokkaido. These pixel-level retrievals and the corresponding zoomed-in Himawari-8 image are given together in Fig. 4.10.

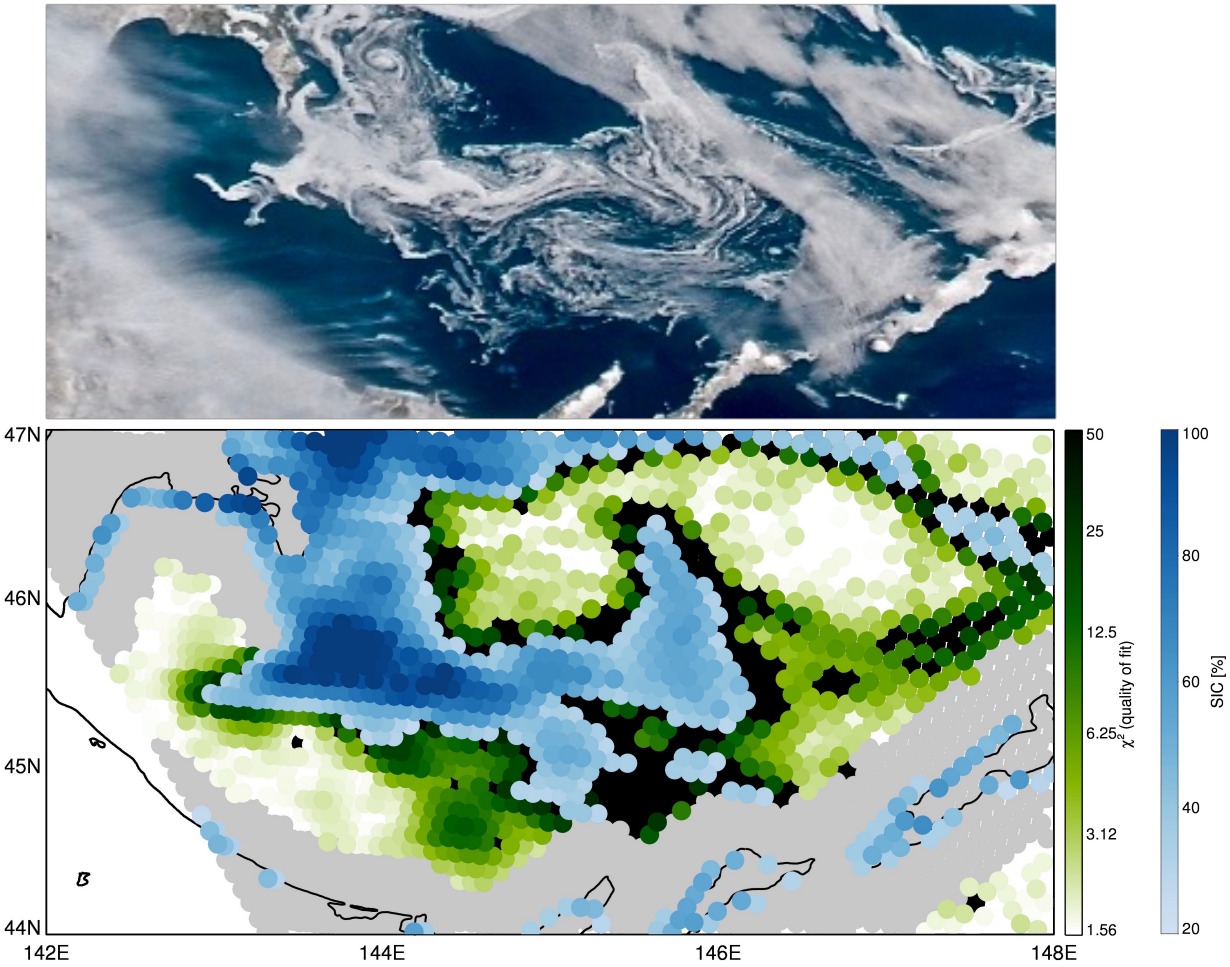

Figure 4.10 A zoomed in view of Figs. 4.8 and 4.9, centered on a sea ice gyre northeast of Hokkaido. The top panel is the zoomed in Himawari-8 image. The bottom panel shows SIC and ocean suite χ^2 at the AMSR2 pixel level. Gray pixels signify pixels where ocean suite was not run due to land contamination.

The close-up view of sea ice and ocean suite retrievals northeast of Hokkaido (Fig. 4.10) makes it clear that the lower spatial resolution of the low frequency channels is not the cause for discrepancy, but instead sensitivity to low sea ice concentrations is to blame. This is not an unknown cause for error in standard sea ice retrievals (Meier and Ivanoff (in review)) due to their being tuned to match validation data, and the need to have a cutoff at some concentration. Since passive microwave sea ice algorithms are primarily concerned with the large signal of ice

extent in the Arctic or Antarctic, the fact that low concentrations near Hokkaido, Japan are missed is not unexpected. This is a good example of the high sensitivity exhibited by the CSU 1DVAR to an inhomogeneous ocean scene. So despite resulting in a gray area of uncertain sea ice concentration, such scenes of marginal sea ice are still better depicted by the integrated algorithm than if using one algorithm by itself.

4.6.2 2012 Arctic ice loss

The rapid loss of Arctic sea ice over the past few decades is a strong manifestation of climate change and a positive feedback on the climate system (Flanner et al. 2011). Arctic sea ice minima typically occur in September each year and are closely watched for purposes of climate monitoring. Local weather can play a key role in these events, perhaps mostly dramatically in the two lowest minima recorded to date, those of 2007 and 2012, an increasing consequence of thinner ice comprising more of the total area of sea ice (Kay et al. 2008; Parkinson and Comiso 2013). The causes and monitoring of these minima are of interest for operational and research purposes, as such events may become more common in the future as multiyear ice rapidly declines (Comiso 2012).

Prior to the 2012 minimum, early August witnessed the rapid loss of about $4x10^5$ km^2 of sea ice, an area roughly the size of California. Along with the abundance of thin, first-year ice, the 2012 Arctic sea ice minimum was affected by an early August storm that brought warm, moist air and strong winds (Parkinson and Comiso 2013). With such an event, thick cloud cover and precipitation make it difficult to assess the extent of sea ice, and indeed there were large discrepancies between estimates from AMSR2 and those from other datasets (Meier et al. 2015). The integrated AMSR2 retrieval provides a unique view of such complex events, as it captures the interplay of precipitation, clouds, winds, and sea ice.

Integrated AMSR2 retrievals from August 16th 2012 are shown in Fig. 4.11. This date was chosen as indicative of the period of large-scale breakup of sea ice extent north of the Chukchi Sea (between Russia and Alaska), and because it was a period of extreme disagreement in sea ice extent between AMSR2 and the operational Multisensor Analyzed Sea Ice Extent (MASIE) product (nsidc.org/data/masie), as shown by Meier et al. (2015). The integrated retrieval sheds some light on this complex scene that includes precipitation, high winds, and an unclear sea ice edge. A strong cyclone in the bottom of the figure, north of Alaska, brings rain and warm air deep into the Arctic, speeding up the sea ice breakup. This is the area of interest, and one where the integrated retrievals can be of use.

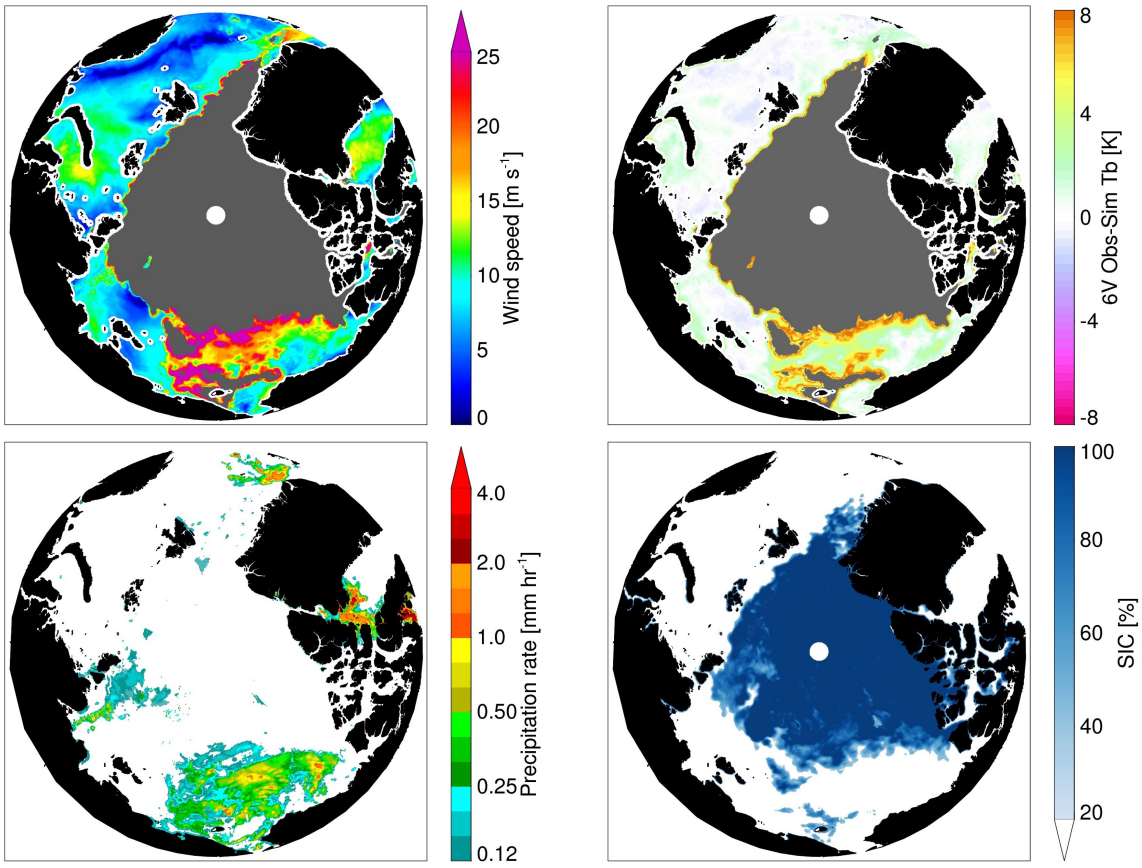

Figure 4.11 Polar projection view of all AMSR2 pixels from August 16th, 2012. Panels show: wind speed (top left), 6V T_B residual from ocean suite (top right), surface precipitation rate (bottom left), and SIC (bottom right). For the top panels, gray signifies pixels with no convergence below $\chi^2 = 10$ for ocean suite. For sea ice, gray signifies pixels not run due to atmospheric contamination.

GPROF finds significant precipitation north of Alaska in an area deemed to be ice-free. However, examination of the 6V T_B residual from ocean suite within the precipitating area exhibits a +5 K signal in some pixels (Fig. 4.11), indicating that low sea ice concentrations are likely present, as the 6V channel is largely unaffected by precipitation. Contrast in the clarity of the sea ice edge is stark—it is very well defined north of Eurasia but messy north of the Chukchi Sea—and demonstrates the utility in using 6.925 GHz for sensing sea ice. It should be noted that the high wind speeds found in marginal sea ice cases, as in Fig. 4.11 or the ice edge in Fig. 4.5, are almost certainly mistaken, as the signal from low SIC is hard to distinguish from high wind speeds because both preferentially increase emission in the H-pol channels.

The integrated retrievals ultimately reveal a middle ground of ice extent that falls between the discordant NT2 and MASIE estimates as investigated by Meier et al. (2015). Rather than being ice-free or remaining ice-covered, the largescale breakup is ongoing, with high wind speeds, TPW, columnar liquid water, and areas with likely ice concentrations below the NT2 algorithm's sensitivity threshold. This more nuanced view of a particularly complex scene is afforded by leveraging AMSR2's multispectral information content and all algorithms' strengths. Further exploration of this scene's areas of marginal ice is given in the following section.

4.7 Discussion

An integrated approach has been presented for the AMSR2 sensor over ocean. It brings together three separate retrieval algorithms—precipitation, sea ice, and ocean suite—to provide a powerful product with more information and fewer discrepancies than any one individually. Four case studies have been discussed that demonstrate benefits of this approach but also challenges for microwave remote sensing that require further work and investigation. This analysis focused

on high latitude precipitation and sea ice extent, subjects of interest for global modeling and climate monitoring in which microwave observations necessarily play a key role.

In each case presented, the integrated retrieval yields information potentially valuable for physical interpretation. Ambiguous scenes with light drizzle or marginal sea ice concentrations are not definitively resolved by the three algorithms, and snowfall remains difficult to detect. Whereas in some instances the analysis may point toward improved methods for future algorithm development, other instances demonstrate real physical limits to microwave remote sensing with the AMSR2 frequencies.

Improved screening for precipitation aids the sea ice algorithm, and ocean suite can override erroneous retrievals that mistake cloud or wind emission signals for emission by sea ice. However, marginal cases of sea ice remain a gray area, and not one easily remedied. A few retrieval alternatives exist for this scenario. One approach could involve forward modeling marginal ice concentrations via CRTM; however, the variability in ice emissivity is difficult to model, and the signal is hard to distinguish from that of high wind speeds. Another alternative is direct retrieval of emissivity and then backing out the SIC as in (Kongoli et al. 2011), though even that approach has a lower limit of 12% SIC and thus would not solve the problems encountered here. Lastly, we could use T_B residuals at low frequencies from ocean suite to flag for probable marginal ice, an idea explored in this section. The explicit forward modeling of radiation by the CSU 1DVAR is a potential asset for detecting both marginal sea ice and light precipitation. Using the cases presented in Sections 4.5.1 and 4.6.2, we explore solutions to these problems next.

4.7.1 Shallow warm rain

To better detect light rainfall from a passive sensor, it helps to leverage information from an active instrument such as CloudSat. This was done in two ways: determining likely LWP thresholds for drizzle onset, and deriving a mean relationship between rain water path (RWP) and surface rain rate.

Determination of LWP values for drizzle onset was achieved using CloudSat statistics for frequency of liquid precipitation after smoothing to approximate a microwave imager footprint (Wood 2016). Precipitation frequency in various regimes, broken up by TPW and SST, was matched with the probability distribution of LWP as seen by AMSR2, accounting for non-convergence due to precipitation. From this an appropriate value for drizzle onset was calculated—one consistent with CloudSat statistics for each regime. This works best for warm, light rain, as large drops or mixed phase drops would cause significant scattering not accounted for in the forward model. LWP values for drizzle onset vary from 100 to 240 g m^{-2} for typical TPW/SST regimes encountered.

To determine a rain rate from a given RWP, assumptions have to be made regarding its vertical extent, the drop size distribution, and the fall velocity. Since none of these are known, they need to be assumed either explicitly or implicitly. Assuming those parameters implicitly, a best-fit relationship between RWP and surface rain rate was calculated from CloudSat data (Munchak 2016).

$$RR = 5 * RWP^{1.1} \quad \text{(Eq. 4.6)}$$

This enables us to calculate a rain rate after partitioning total column liquid into cloud LWP and rain LWP, a method not uncommon in microwave remote sensing (Wilheit et al. 1977; Wentz and Spencer 1998). It is worth emphasizing two things. First, the LWP retrieval assumes

non-scattering hydrometeors, so this is only applicable to warm clouds with no large drops present. Second, this is akin to an upper bound on drizzle, as the retrieval attributes all water above the drizzle threshold to RWP, whereas LWP is typically partitioned into cloud and rain gradually above a chosen threshold (Wentz and Spencer 1998).

We then reexamine the scene near the Faroe Islands discussed in Section 4.5.1. Fig. 4.12 replaces the GPROF rain rate of Fig. 4.3 with that derived from the CSU 1DVAR and Eq. 4.6. The magnitude of these rain rates agrees well with CloudSat in spite of the simplistic assumptions. The fact that it does agree well despite being an upper bound estimate is likely due to the very shallow nature of the cloud. A thicker cloud with the same RWP, consequently with lower rain water content on average, would likely exhibit a lower rain rate and not agree as closely with this estimate.

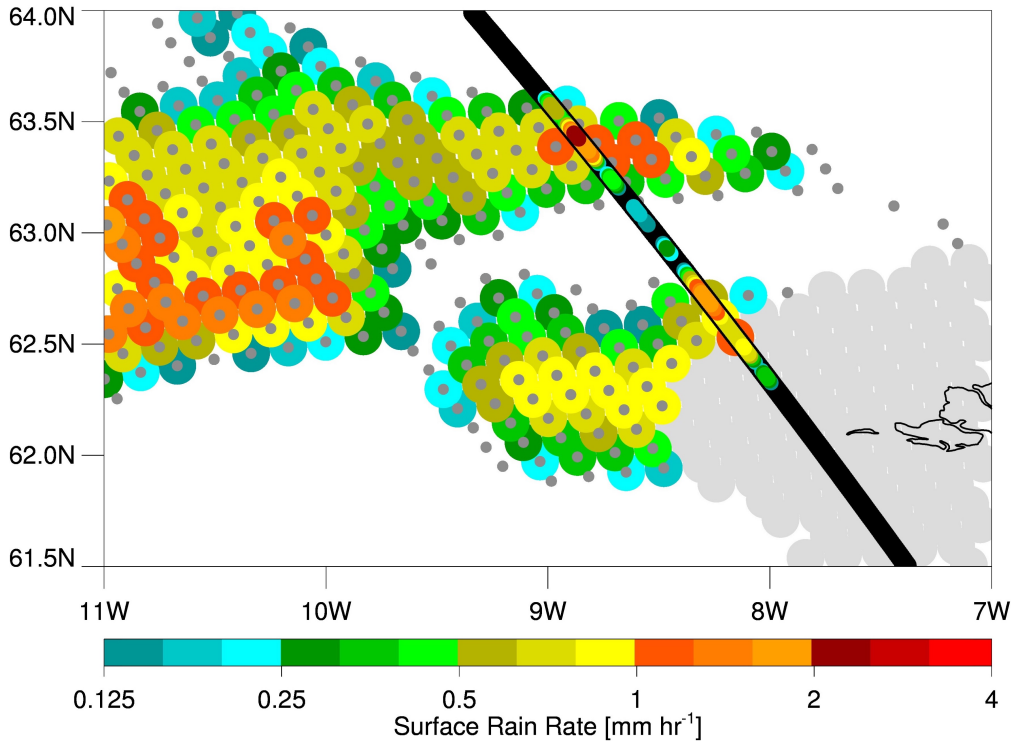

Figure 4.12 Recreating Fig. 4.3, but replacing GPROF rain rates with those calculated from the 1DVAR via Eq. 4.6. Gray dots signify pixels whose columnar liquid is greater than or equal to the drizzle threshold as determined from CloudSat data, i.e. pixels assumed to be precipitating as per Fig. 3.1.

4.7.2 Marginal sea ice

As shown in Section 4.6, areas of marginal sea ice provide a small but detectable signal for passive microwave measurements, especially at low frequencies. However, as with the example given above for light precipitation, explicit forward modeling of the atmospheric and oceanic state is necessary to effectively diminish the uncertainty and increase the SNR.

The large-scale breakup of Arctic sea ice, given in Section 4.6.2, exhibited large areas of ambiguity north of Alaska where operational and microwave-only estimates diverged. Low frequency observations from AMSR2, however, hold information able to elucidate the state further. 6.925 GHz is problematic at fine scales due to its very large footprint, but is largely insensitive to water vapor, cloud water, and precipitation, making it ideal for sensing the surface state. In fact, 6V T_B will increase approximately +1 K for 1% of SIC increase in the FOV, and 10V and 10H will increase 1% and 1.5%, respectively, making concentrations of 2% or greater theoretically detectable given the assumed observational uncertainties (Fig. 4.1). We therefore define likely marginal sea ice of greater than 3% SIC but less than that detectable by NT2 thus:

$$\Delta Tb_{6V} > 3\ K;\ \ \Delta Tb_{10H} > 4.5\ K;\ \ (\Delta Tb_{10H} - \Delta Tb_{10V}) \leq 0\ K \quad \text{(Eq. 4.7)}$$

where ΔT_B is the T_B residual from the ocean suite algorithm, observed minus simulated T_B. This accounts for significantly increased emission at low frequencies and also the differential polarization signal expected from sea ice. As with NT2, possible marginal ice is limited to SST<278 K in the Northern Hemisphere.

For this inequality to be satisfied for an ice-free scene (barring forward model error) SST would need to be about 6 K higher than retrieved as well as windier than retrieved; alternately, sensor noise at 9σ for 6V and 6σ at 10H could cause an erroneously labeled marginal ice scene (Table 4.1). So despite inexact sensor calibration, absorption errors, and positively correlated

emissivity model errors at low frequencies, this is a fairly conservative expression for marginal ice, as assumed forward model errors are $\sigma_{err,6v}$ = 0.99 K and $\sigma_{err,10h}$ = 1.00 K. The negative difference of 10H and 10V will hold for scenes where the 1DVAR cannot increase H-pol emission enough relative to V-pol, a situation indicating likely ice presence. This inequality could plausibly be satisfied in very heavy precipitation, but is strongly tied to sea ice at these latitudes.

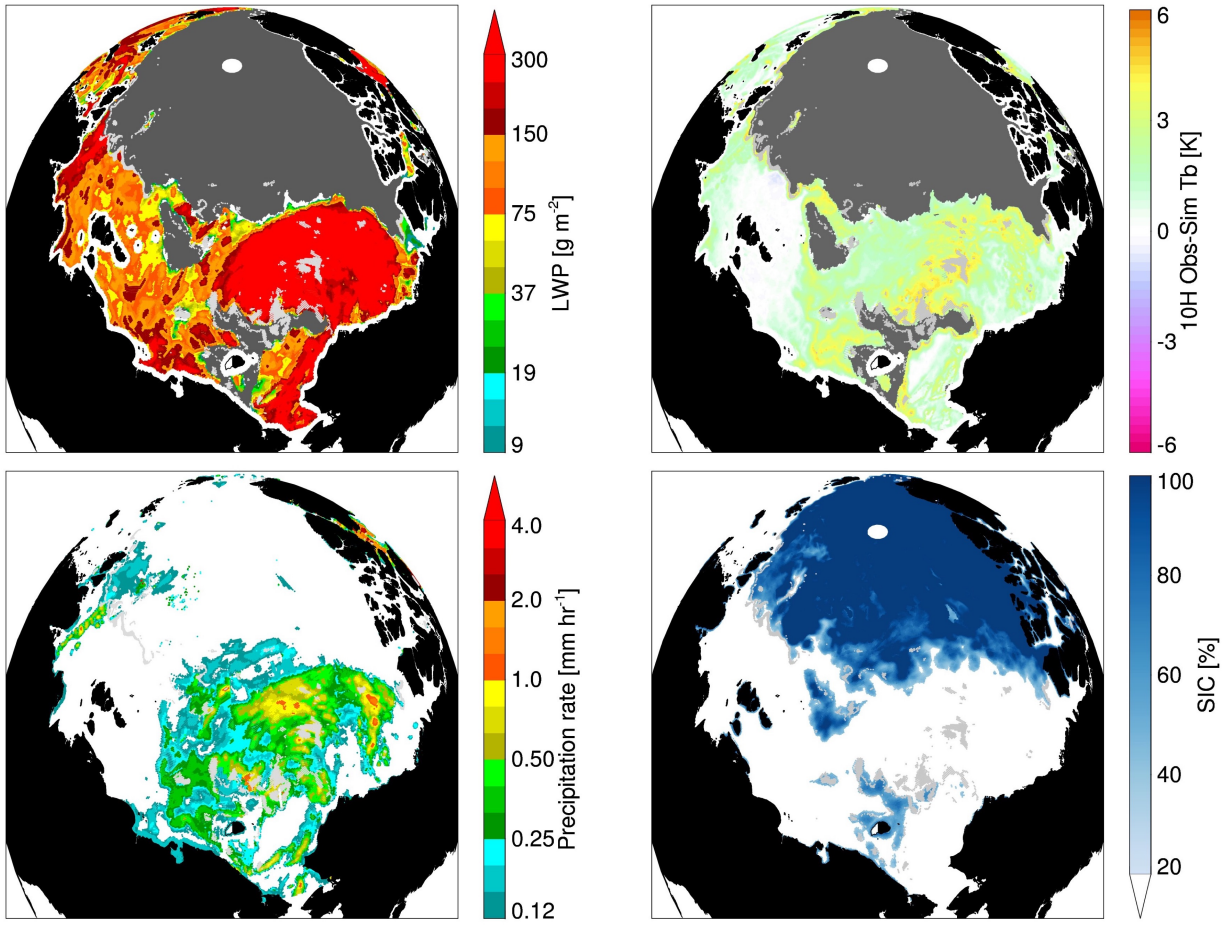

Figure 4.13 As with Fig. 4.11, a near-polar view of Arctic sea ice breakup on Aug. 16th 2012. Panels show: LWP (top left), 10H T_B residual from ocean suite (top right), surface precipitation rate (bottom left), and SIC (bottom right). For the top panels, dark gray signifies pixels with no convergence below χ^2 = 10 for ocean suite. For all panels, light gray shows pixels denoted as marginal sea ice as described by Eq. 4.7.

Fig. 4.13 is similar to Fig. 4.11, but with the projection rotated down towards the Chukchi Sea in which the sea ice breakup is ongoing. The pixels that satisfy Eq. 4.7 are shown in light gray, signaling marginal sea ice. Additionally, the top panels are changed from wind speed and Δ6V to LWP and Δ10H, respectively, to demonstrate that the resultant marginal ice area is not tied to clouds or precipitation. Perhaps unsurprisingly, marginal ice is found near the edge of higher SIC areas as determined by NT2. Both behind the cyclone—near Wrangel Island—and along the path of the cyclone, marginal ice concentrations are left behind. A trail of marginal ice concentrations stretches from Wrangel Island north into the center of the cyclone, covered by clouds and precipitation but visible in the lower frequencies of AMSR2.

Lastly, we test the robustness of the marginal ice signal to assess significance of the larger patterns. Most of the marginal ice appears to be small concentrations and sensitive to the chosen thresholds. A 33% increase (decrease) in the 6V and 10H thresholds causes approximately a 4-fold decrease (increase) in the number of pixels that satisfy the inequality, though the spatial patterns largely hold. The marginal ice also changes if maximum allowable wind speed in the 1DVAR is lowered, causing an increase of 40% in marginal ice pixels if wind speed is limited to 20 m s^{-1}. This happens because the retrieval tends to increase wind speed if ice is present due to their similar radiometric effects. This is not a practical limit for the standard retrieval, however, since wind speeds greater than 20 m s^{-1} do exist in nature.

Due to the large sensitivity of designated marginal ice pixels to the thresholds chosen, these results should be interpreted with caution. The average marginal ice pixel exhibits SST 1.5 K higher than the *a priori* value and a mean wind speed of 22 m s^{-1}. This indicates that the retrieval is working hard to match the T_Bs, increasing SST and wind speed as much as possible but still off by >3 K at low frequencies; conversely, it can indicate that the marginal ice

designation is mistaking abnormally warm, windy ocean scenes for sea ice, exacerbated by larger than expected forward model errors. Because of these ambiguities, the marginal ice designation is better suited as a qualitative tool than a quantitative measurement.

4.7.3 Conclusions

These explorations of methods for light precipitation and marginal sea ice detection are not yet intended for operational use. They do, however, investigate topics much more suited to an integrated retrieval than a simple augmentation of any one individual algorithm. This demonstration argues for a more holistic approach to microwave imager retrievals and shows the power of AMSR2 for water cycle studies.

Shallow rain detection and retrieval can be aided by a variational approach to augment the Bayesian precipitation algorithm, as shown in Fig. 4.12. Increased sensitivity from direct forward modeling of the state effectively diminishes uncertainty and increases the SNR due to fewer assumptions being necessary. In contrast, the retrieval of snowfall rates from passive microwave measurements will remain a challenging problem because of the lack of significant scattering at lower frequencies and the combination of microphysics and radiative transfer uncertainty diminishing the SNR.

Detection of sea ice below the minimum detectable concentrations of NT2 is possible due to the higher sensitivity of a variational algorithm. A fairly simple, physically-based heuristic was presented that attempted to find areas of marginal sea ice by using low frequency channels exclusively. This heuristic can shed light on a complex scene such as sea ice breakup under thick cloud cover, though is imperfect and not intended to quantify low concentrations.

The integrated approach improves the output of each algorithm, but was not a focus of the analysis presented. For instance, pixel to pixel sea ice variability is important to capture for

using the correct surface type in GPROF, and indeed removes some spatial artifacts from the GPROF retrieval because otherwise ancillary data grids for ice are necessary. Since many GPM constellation radiometers contain the frequencies used by NT2, it is possible to imagine embedding NT2 in the next version of GPROF to improve precipitation estimates in polar regions by the GPM constellation.

Ultimately the integrated retrievals approach will be extended to the AMSR2 land algorithms as well, including snow and soil moisture as well as precipitation. This work is ongoing, hence the focus on ocean algorithms here. Full reprocessing of a combined AMSR-E and AMSR2 data record as AMSR Unified, or AMSRU, featuring integrated retrievals and consistent calibration could provide a powerful climate dataset to assess changes over time in the global hydrologic cycle.

CHAPTER 5

CONCLUSIONS

This document centers around a variational retrieval that was developed for microwave imagers over ocean. Using this retrieval, information content from passive microwave satellite observations was investigated via two state-of-the-art sensors, assessing limitations of the current observing system. Forward model errors are posited to have a large impact on retrieval results, the conveyed uncertainties of retrieved parameters, and the ultimate utility of the microwave data record. To summarize the conclusions of the preceding studies, this final chapter is divided into main findings, implications, and future work.

5.1 Main findings

Variational retrievals are not new, nor is their application to microwave imagers; each chapter of this document cites numerous examples of variational retrievals from the literature, studies upon which this dissertation rests. As explained in Chapter 1, a key difference between variational methods applied within the satellite data assimilation and retrieval communities lies in the assumption of observation errors, especially in areas of cloud and precipitation. Treatment of these observation errors is a focal point of Chapters 2 and 3. In reality, it may be impossible to quantify all possible errors of an observing system. However, the fullest possible characterization of errors will maximize the information content that can be wrung out of an observing system. Unrealistic assumed errors constitute wasted information or an over-fit to the measurements.

This is not a concept exclusive to satellite meteorology, but it has historically been neglected in this field.

Section 2.2.3 developed a methodology for quantifying covarying forward model errors for microwave imagers. This methodology is not perfect and neglects certain sources of error while possibly overestimating others. It could be augmented to be situation-dependent or account for spatial inhomogeneity, for instance. The methodology is novel in its attempt to quantify forward model error as fully as possible, including forward model parameter error. As in Bettenhausen et al. (2006), inclusion of off-diagonal observation error terms was concluded to improve the channel weighting and in turn improve the retrievals in areas of otherwise ambiguous state vector signals. This was examined in Section 2.5.1 and is in line with conclusions on correlated forward model error in DA schemes (Bormann et al. 2011).

Retrieval of non-scattering ocean scenes was shown to be successful with the variational algorithm presented in Chapter 2, simulating radiances with fidelity and validating well against other datasets. This demonstrated that a retrieval with a fully physical forward model could compete with a benchmark product (Hilburn and Wentz 2008) with no tuning required. It outperforms other published variational imager retrievals (Elsaesser and Kummerow 2008; Boukabara et al. 2010) as well as the reanalysis *a priori* (Table 2.2). While the CSU 1DVAR does not yet outperform the so-called Wentz products, its full description of posterior errors and fit to the T_B vector provide valuable information to users.

As outlined in Sections 1.4, 2.6, 3.6, and 4.5, variational retrievals must be guided by the limitations imposed by a finite number of channels with non-independent information. The information content contained in a given T_B vector places a hard limit on what can realistically be retrieved. While ancillary data can offer useful constraints in some circumstances, there is no

substitute for more spectral information. In the case of raining scenes, for instance, a typical imager's information content limitations dictate that only 3 or 4 variables can be retrieved, necessitating assumptions about the drop size distribution and vertical profile (Section 3.3). Again in this regard the variational framework is helpful, guiding algorithm development through use of the A-matrix as discussed in Sections 3.3.2 and the Appendix. For this reason, retrieval of precipitation other than warm rain is deemed too underconstrained for a passive-only variational algorithm with current sensors' limitations.

Despite these stated caveats about limited information content and retrieved parameters, Chapters 3 and 4 highlighted cases in which the variational retrieval outperforms other operational algorithms or proves valuable for interpreting a complex scene. Chapter 4 contends that the higher effective signal to noise ratio (SNR) afforded by explicit forward modeling can be leveraged to shed light on regions of marginal sea ice, warm rain, and falling snow. The same argument is used in Chapter 3 to argue for the CSU 1DVAR's higher sensitivity to warm rain when compared to GPROF. These cases exemplify the thesis that a more holistic treatment of observation errors leads to extracting more information from the microwave data record.

5.2 Implications

The topic of sensor calibration, mentioned in Chapter 2 and briefly in Chapter 4, was not a focus of the work presented. However, calibration is crucial for all retrievals and especially so for variational methods, as inconsistent calibration between channels will lead to biases or non-convergence. One side effect of the GPM mission's reliance on a constellation of radiometers is that relative and absolute calibration of microwave imagers have received increased attention from researchers (Draper et al. 2015; Berg et al. 2016). That work on sensor calibration

underpins the work presented throughout this document, in that no tuning of Level 1 data was required to provide either good validation statistics or Gaussian T_B residuals (Figure 2.2). As intercalibration of GPM constellation radiometers improves, robust climate trend analysis from GPM- and TRMM-era sensors should be increasingly possible, meaning that long term analyses of parameters derived by the CSU 1DVAR should be stable.

In a reciprocal manner, a variational retrieval such as the CSU 1DVAR can be used in studies to improve both relative and absolute sensor calibration due to its physical forward model. While not included in this document, such work has been ongoing and should inform future intercalibration efforts. In addition, scene-dependent biases such as those from an emissivity model can be mitigated via iterative analyses of T_B residuals from the retrieval. This has already been done in the form of an emissivity model correction based on wind speed, and applied to a recent research version of the CSU 1DVAR. The potential thus exists to use the work herein to improve the calibration of GPM constellation radiometers, or assess and mitigate forward model biases in future versions of the algorithm.

Crosstalk errors are also related to discussion of the viability of climate trend analysis from the passive microwave data record, as any aliasing of radiometric signals from one parameter to another could affect such an analysis. Chapter 2 investigated crosstalk errors and concluded that a fuller representation of observation errors and *a priori* covariances could improve this long-standing problem in satellite retrievals. In particular, the description of posterior errors afforded by a variational retrieval, and these errors' covariance, allows for quantitative assessment of crosstalk. While crosstalk errors can be addressed as a post hoc correction (Elsaesser et al. in review), better treatment of crosstalk in the retrieval itself should lead to better downstream products.

Treatment of correlated errors for satellite radiances has shown promise in operational DA schemes (Weston et al. 2014; Bormann et al. 2015). It is argued in this work that assimilation of passive microwave radiances in cloudy or precipitating scenes, known as all-sky conditions, could benefit from the type of error covariance analysis presented in Chapters 2 and 3. This is consistent with the literature and the shared mathematics of OE and DA as discussed in Section 1.2, but was outside the scope of the work presented here and is therefore an unexplored implication of this work.

5.3 Future work

Chapters 2 and 3 used data from the GMI sensor while Chapter 4 focused on operational products from the US AMSR2 team. A few elements of future work are related to these two missions.

As mentioned in Section 4.7.3, a unified AMSR product called AMSRU is forthcoming that will provide a consistent, intercalibrated record from the AMSR-E and AMSR2 sensors. This provides an opportunity to run the CSU 1DVAR for the 15+ year record and analyze any trends over this time period. The integrated approach outlined in Chapter 4 can also be implemented in an extended form, including the land algorithms, for a data product that can be applied in global water cycle studies. Any such analysis could leverage the CSU 1DVAR's posterior error estimates and changes in T_B residuals to assess the significance of any trends that are present.

As with the integrated algorithms presented in Chapter 4, the development process that led to Chapter 3's warm rain retrieval included a phase in which GPROF and the CSU 1DVAR were run side by side. The blended Bayesian/variational approach to precipitation retrieval

(Section 3.6) could thus potentially improve the GPM constellation's representation of drizzle at high latitudes (Table 3.2), a known problem despite the advances realized from DPR (Skofronick-Jackson in press). While it would be difficult to apply operationally to all constellation radiometers due to their varying channel suites, it may provide a way to maximize the information content gleaned from GMI-like sensors.

REFERENCES

Alishouse, J. C., S. A. Snyder, J. Vongsathorn, and R. R. Ferraro (1990), Determination of oceanic total precipitable water from the SSM/I, *IEEE Trans. Geosci. Rem. Sens., 28*, 811–816.

Andersson, E., J. Pailleux, J.-N. Thépaut, J. R. Eyre, A. P. McNally, G. A. Kelly, and P. Courtier (1994), Use of cloud-cleared radiances in three/four-dimensional variational data assimilation, *Q. J. R. Meteorol. Soc.*, *120*, 627-653.

Aonashi, K., and Coauthors (2009), GSMaP passive microwave precipitation retrieval algorithm: algorithm description and validation, *J. Meteorol. Soc. Japan*, *87A*, 119-136.

Auligné, T., McNally, A. P. and Dee, D. P. (2007), Adaptive bias correction for satellite data in a numerical weather prediction system, *Q. J. R. Meteorol. Soc.*, *133*, 631–642, doi:10.1002/qj.56.

Aumann, H. H., and Coauthors (2003), AIRS/AMSU/HSB on the Aqua mission: Design, science objectives, data products, and processing systems, *IEEE Trans. Geosci. Rem. Sens., 41*(2), 253–264.

Backus, G. and F. Gilbert (1970), Uniqueness in the inversion of inaccurate gross Earth data, *Philos. Trans. R. Soc. London A, Math. Phys. Sci.*, *266*, 1173, 123–192.

Bauer, P., P. Lopez, A. Benedetti, S. Saarinen, E. Moreau (2006), Implementation of 1D+4D-Var assimilation of precipitation-affected microwave radiances at ECMWF, II: 4D-Var, *Q. J. R. Meteorol. Soc., 132*, 2307-2332.

Bauer. P., A. Thorpe, and G. Brunet (2015), The quiet revolution of numerical weather prediction, *Nature*, *525*, 47–55, doi:10.1038/nature14956.

Bayes, T., and R. Price (1763), An essay towards solving a problem in the doctrine of chance. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S., *Phil. Trans. Royal Soc. London*, *53*, 370-418, doi: 10.1098/rstl.1763.0053.

Behrangi, A., M. Lebsock, S. Wong, and B. Lambrigtsen (2012), On the quantification of oceanic rainfall using spaceborne sensors, *J. Geophys. Res.*, *117*, D20105, doi:10.1029/2012JD017979.

Behrangi, A., and Coauthors (2016), Status of high-latitude precipitation estimates from observations and reanalyses, *J. Geophys. Res. Atmos., 121*, 4468-4486.

Bennartz, R. and G. W. Petty (2001), The sensitivity of microwave remote sensing observations of precipitation to ice particle size distributions, *J. Appl. Meteorol.*, *40*, 345-364.

Bennartz, R. and P. Bauer (2003), Sensitivity of microwave radiances at 85-183 GHz to precipitating ice particles, *Radio Sci.*, *38*, 4, doi: 10.1029:2002RS002626.

Bennartz, R. (2007), Global assessment of marine boundary layer cloud droplet number concentration from satellite, *J. Geophys. Res., 112*, D02201, doi:10.1029/2006JD007547.

Bennartz, R. and T. Greenwald (2011), Current problems in scattering radiative transfer modeling for data assimilation, *Q. J. R. Meteorol. Soc., 137*, 1952-1962.

Berg, W., M. Sapiano, J. Horsman, and C. Kummerow (2013), Improved geolocation and Earth incidence angle information for a fundamental climate data record of the SSM/I sensors, *IEEE Trans. Geosci. Rem. Sens., 51*, 1504-1513.

Berg, W., and Coauthors (2016), Intercalibration of the GPM microwave radiometer constellation, *J. Atmos. Ocean. Tech.*, *33*, 2639-2654, doi: 10.1175/JTECH-D-16-0100.1.

Bessho, K. and coauthors (2016), An Introduction to Himawari-8/9 – Japan's new-generation geostationary meteorological satellies, *J. Meteorol. Soc. Japan*, *94*, 2, 151-183.

Bettenhausen, M. H., C. K. Smith, R. M. Bevilacqua, N. Wang, P. W. Gaiser, and S. Cox (2006), A nonlinear optimization algorithm for WindSat wind vector retrievals, *IEEE Trans. Geosci. Rem. Sens., 44*(3), 597-610.

Birman, C., J.-F. Mahfouf, M. Milz, J. Mendrok, S. A. Buehler, and M. Brath (2017), Information content on hydrometeors from millimeter and sub-millimeter wavelengths, *Tellus A: Dyn. Meteorol. Oceanog., 69:1*, 1271562, doi: 16000870.2016.1271562.

Bormann, N., A. J. Geer, and P. Bauer (2011), Estimates of observation-error characteristics in clear and cloudy regions for microwave imager radiances from numerical weather prediction, *Q. J. R. Meteorol. Soc., 137*, 2014–2023, doi:10.1002/qj.833.

Bormann, N., M. Bonavita, R. Dragani, R. Eresmaa, M. Matricardi, and T. McNally (2015), Enhancing the impact of IASI observations through an updated observation error covariance matrix, *ECMWF Tech. Memo., 756*, available from http://www.ecmwf.int/

Boucher, O., and Coauthors (2013), Clouds and Aerosols, In: Climate Change 2013: The Physical Science Basis, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Boukabara, S.-A., and Coauthors (2011), MiRS: An all-weather 1DVAR satellite data assimilation and retrieval system, *IEEE Trans. Geosci. Rem. Sens., 49*(9), 3249– 3272, doi:10.1109/TGRS.2011.2158438.

Boukabara, S.-A., K. Garrett, and W. Chen (2010), Global coverage of total precipitable water using a microwave variational algorithm, *IEEE Trans. Geosci. Rem. Sens.*, *48*(10), 3608-3621, doi: 10.1109/TGRS.2010.2048035.

Bringi, V. N., V. Chandrasekar, J. Hubbert, E. Gorgucci, W. L. Randeu, and M. Schoenhuber (2003), Raindrop size distribution in different climatic regimes from disdrometer and dual-polarized radar analysis, *J. Atmos. Sci., 60*, 354–365.

Bringi, V. N., T. Tang, and V. Chandrasekar (2004), Evaluation of a new polarimetrically based Z-R relation, *J. Atmos. Ocean. Tech.*, *23*, 612-623.

Cavalieri, D. J., P. Gloersen, and W. J. Campbell (1984), Determination of sea ice parameters with the NIMBUS 7 SMMR, *J. Geophys. Res.*, *89*, D4, 5355-5369.

Chen, R., Z. Li, R. Kuligowski, R. Ferraro, and F. Wang (2011), A study of warm rain detection using A-Train satellite data, *Geophys. Res. Lett., 38*, L04804, doi:10.1029/2010GL046217.

Chernokulsky, A. and I. I. Mokhov (2012), Climatology of total cloudiness in the Arctic: An intercomparison of observations and reanalyses, *Advan. Meteorol.*, *2012*.

Comiso, J. C., D. J. Cavalieri, and T. Markus (2003), Sea ice concentration, ice temperature, and snow depth using AMSR-E data, *IEEE Trans. Geosci. Remote Sensing*, *41*, 243–252.

Comiso, J. C. and F. Nishio (2008), Trends in the sea ice cover using enhanced and compatible AMSR-E, SSM/I, and SMMR data, *J. Geophys. Res.*, *113*, C02S07, doi: 10.1029/2007JC004257.

Comiso, J. C. (2012), Large decadal decline of the Arctic multiyear ice, *J. Clim.*, *25*, 1176-1193.

Comstock, K. K., R. Wood, S. E. Yuter, and C. S. Bretherton (2004), Reflectivity and rain rate in and below drizzling stratocumulus, *Q. J. R. Meteorol. Soc., 130*, 2891–2918, doi:10.1256/qj.03.187.

Dang, H. V. T, Lambrigtsen, B., Manning, E., Eds. (2012), AIRS/AMSU/HSB Version 6 Level 2 Performance and Test Report, NASA Jet Propulsion Laboratory, Pasadena, CA, 197 pp.

Deblonde, G. And S. English (2003), One-Dimensional variation retrievals from SSMIS-simulated observations, *J. Appl. Meteorol., 42*, 1406-1420, doi:10.1175/1520-0450(2003)042<1406:OVRFSO>2.0.CO;2.

Dee, D.P., and Coauthors (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc., 137*, 553–597. doi:10.1002/qj.828.

Draper, D., D. Newell, F. Wentz, S. Krimchansky, and G. Skofronick-Jackson (2015), The global precipitation measurement (GPM) microwave imager (GMI): Instrument overview and early on-orbit performance, *IEEE J. Sel. Topics Appl. Earth Obs. Rem. Sens., 8*(7), 3452-3462, doi:10.1109/JSTARS.2015.2403303.

Duncan, D. I. and C. D. Kummerow (2016), A 1DVAR retrieval applied to GMI: Algorithm description, validation, and sensitivities, *J. Geophys. Res. Atmos.*, *121*, 7415-7429.

Duncan, D. I., C. D. Kummerow, and W. M. Meier, An integrated examination of AMSR2 products over ocean, *IEEE J. Sel. Topics Appl. Earth Obs. Rem. Sens*, in review.

Eliasson, S., G. Holl, S. A. Buehler, T. Kuhn, M. Stengel, F. Iturbide-Sanchez, and M. Johnston (2013), Systematic and random errors between collocated satellite ice water path observations, *J. Geophys. Res. Atmos., 118*, 2629–2642, doi:10.1029/2012JD018381.

Elsaesser, G. S., and C. D. Kummerow (2008) Toward a fully parametric retrieval of the nonraining parameters over the global oceans, *J. Appl. Climatol., 47*, 1599-1618.

Elsaesser, G. S., C. W. O'Dell, M. D. Lebsock, R. Bennartz, T. J. Greenwald, and F. J. Wentz, The Multi-Sensor Advanced Climatology of Liquid Water Path (MAC-LWP), *J. Clim.*, in review.

English, S. J., R. J. Renshaw, P. C. Dibben, A. J. Smith, P. J. Rayer, C. Poulsen, F. W Saunders, and J. R. Eyre (2000), A comparison of the impact of TOVS and ATOVS satellite sounding data on the accuracy of numerical weather forecasts, *Q. J. R. Meteorol. Soc.*, *126*, 2911-2931.

Ferraro. R. R., and Coauthors (2013), An evaluation of microwave land surface emissivities over the continental United States to benefit GPM-era precipitation algorithms, *IEEE Trans. Geosci. Remote. Sens.*, *51*, 1, 378-398.

Flanner, M. G., K. M. Shell, M. Barlage, D. K. Perovich, and M. A. Tschudi (2011), Radiative forcing and albedo feedback from the Northern Hemisphere cryosphere between 1979 and 2008, *Nature Geosci.*, *4*, 151-155.

Geer, A. J., P. Bauer, and P. Lopez (2010), Direct 4D-Var assimilation of all-sky radiances. Part II: Assessment, *Q. J. R. Meteorol. Soc.*, *136*, 652, 1886-1905.

Geer, A. J., and P. Bauer (2011), Observation errors in all-sky data assimilation, *Q. J. R. Meteorol. Soc.*, *137*, 2024-2037, doi: 10.1002/qj.830.

Gong, J. and D. L. Wu (2017), Microphysical properties of frozen particles inferred from Global Precipitation Measurement (GPM) Microwave Imager (GMI) polarimetric measurements, *Atmos. Chem. Phys., 17*, 2741-2757.

Greenwald, T. J., G. L. Stephens, T. H. Vonder Haar, and D. L. Jackson (1993), A physical retrieval of cloud liquid water over the global oceans using special sensor microwave/imager (SSM/I) observations, *J. Geophys. Res., 98*, 18471–18488.

GPM Science Team (2016), GPM GMI Level 1C common calibrated brightness temperatures collocated, version 04, Greenbelt, MD, USA, NASA Earth Science Data and Information Services Center (GES DISC), Accessed 6 April 2016.

Hartmann, D.L., and Coauthors (2013), Observations: Atmosphere and Surface, In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth

Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Hilburn, K. A., and F. J. Wentz (2008), Intercalibrated passive microwave rain products from the unified microwave ocean retrieval algorithm (UMORA), *J. Appl. Meteorol. Climatol., 47*, 778–794.

Hollinger, J. P., J. L Peirce, and G. A. Poe (1990), SSM/I instrument evaluation, *IEEE Trans. Geosci. Remote Sens.*, *28*, 781-790.

Hou, A. Y., and Coauthors (2014), The Global Precipitation Measurement mission, *Bull. Amer. Meteorol. Soc., 95*, 701–722.

Houze, R., L. McMurdie, W. Petersen, M. Schwaller, W. Baccus, J. Lundquist, C. Mass, B. Nijssen, S. Rutledge, D. Hudak, S. Tanelli, J. Mace, M. Poellot, D.Lettenmaier, J. Zagrodnik, A. Rowe, J. DeHart, L. Madaus, and H. Barnes (2017): The Olympic Mountains Experiment (OLYMPEX). *Bull. Amer. Meteorol. Soc.*, doi:10.1175/BAMS-D-16-0182.1, in press.

Hsu, S. A., E. A. Meindl, and D. B. Gilhousen (1994), Determining the power-law wind-profile exponent under near-neutral stability conditions at sea, *J. Appl. Meteorol., 33*, 757–765.

Imaoka, K., and Coauthors (2010), Global Change Observation Mission (GCOM) for monitoring carbon, water cycles, and climate change, *Proc. IEEE*, *98*, 5, 717-734.

Iturbide-Sanchez, F., S.-A. Boukabara, R. Chen, K. Garrett, C. Grassotti, W. Chen, and F. Weng (2011), Assessment of a variational inversion system for rainfall rate over land and water surfaces, *IEEE Trans. Geosci. Rem. Sens., 49*, 9, 3311-3333.

Kay, J. E., T. L'Ecuyer, A. Gettelman, G. Stephens, and C. O'Dell (2008), The contribution of cloud and radiation anomalies to the 2007 Arctic sea ice extent minimum, *Geophys. Res. Lett.*, *35*, L08503.

Kazumori, M., T. Egawa and K. Yoshimoto (2012), A retrieval algorithm of atmospheric water vapor and cloud liquid water for AMSR-E, *Euro. J. Rem. Sens., 45*, 63-74, doi: 10.5721/EuJRS20124507.

Kazumori, M. and S. J. English (2015), Use of the ocean surface wind direction signal in microwave radiance assimilation, *Q. J. R. Meteorol. Soc., 141*, 1354–1375, doi:10.1002/qj.2445.

Kelly, R. E., A. T. Chang, L. Tsang, and J. L. Foster (2003), A prototype AMSR-E global snow area and snow depth algorithm, *IEEE Trans. Geosci. Remote Sensing*, *41*, 230–242.

Kida, S., S. Shige, T. Manabe, T. L'Ecuyer, and G. Liu (2010), Cloud liquid water path for the rain/no-rain classification method over ocean in the GSMaP algorithm, *Trans. JSASS Aero. Tech. Japan, 8*, 27, 19-23.

King, J. M., C. D. Kummerow, S. C. van den Heever, and M. R. Igel (2015), Observed and modeled warm rainfall occurrence and its relationships with cloud macrophysical properties, *J. Atmos. Sci., 72*, 4075-4090, doi:10.1175/JAS-D-14-0368.1.

Kongoli, C., S.-A. Boukabara, B. Yan, F. Weng, and R. Ferraro (2011), A new sea-ice concentration algorithm based on microwave surface emissivities—application to AMSU measurements, *IEEE Trans. Geosci. Remote Sens.*, *49*, 1, 175-189.

Kulie, M. S. and R. Bennartz (2009), Utilizing spaceborne radars to retrieve dry snowfall, *J. Appl. Meteorol. Climatol.*, *48*, 12, 2564–2580.

Kummerow, C., W. S. Olson, and L. Giglio (1996), A simplified scheme for obtaining precipitation and vertical hydrometeor profiles from passive microwave sensors, *IEEE Trans. Geosci. Remote Sens.*, *34*, 5, 1213-1232.

Kummerow, C., and Coauthors (2000), The status of the tropical rainfall measuring mission (TRMM) after two years in orbit, *J. Appl. Meteorol., 39*, 12, 1965–1982.

Kummerow, C. D., S. Ringerud, J. Crook, D. Randel, and W. Berg (2011), An observationally generated a priori database for microwave rainfall retrievals, *J. Atmos. Oceanic Technol.*, 28, 113–130.

Kummerow, C. D., D. Randel, M. Kulie, N.-Y. Wang, R. Ferraro, S. J. Munchak, and V. Petkovic (2015), The evolution of the Goddard profiling algorithm to a fully parametric scheme, *J. Atmos. Oceanic Technol., 32*, 2265-2280, doi:10.1175/JTECH-D-15-0039.1.

Lean, P., A. Geer, and K. Lonitz (2017), Assimilation of Global Precipitation Mission (GPM) Microwave Imager (GMI) in all-sky conditions, *ECMWF Tech. Memo., 799*, available at http://www.ecmwf.int/.

Lebsock, M. D., G. L. Stephens, and C. Kummerow (2008), Multisensor satellite observations of aerosol effects on warm clouds, *J. Geophys. Res., 113*, D15205, doi:10.1029/2008JD009876.

Lebsock, M. D. and T. S. L'Ecuyer (2011), The retrieval of warm rain from CloudSat, *J. Geophys. Res., 116*, D20209, doi: 10.1029/2011JD016076.

Lebsock, M., and H. Su (2014), Application of active spaceborne remote sensing for understanding biases between passive cloud water path retrievals, *J. Geophys. Res. Atmos., 119*, 8962–8979, doi:10.1002/ 2014JD021568.

L'Ecuyer, T. S., W. Berg, J. Haynes, M. Lebsock, and T. Takemura (2009), Global observations of aerosol impacts on precipitation occurrence in warm maritime clouds, *J. Geophys. Res., 114*, D09211, doi:10.109/2008JD011273.

L'Ecuyer, T. S. and J. H. Jiang (2010), Touring the atmosphere aboard the A-Train, *Physics Today*, July 2010, 36-41.

Le Vine, D. M., E. P. Dinnat, T. Meissner, S. H. Yueh, F. J. Wentz, S. E. Torrusio, and G. Lagerloef (2015), Status of Aquarius/SAC-D and Aquarius salinity retrievals, *IEEE J. Sel. Topics Appl. Earth Obs. Rem. Sens., 8*(12), 5401-5415, doi:10.1109/JSTARS.2015.2427159.

Lee, S.-M., and B.-J. Sohn (2015), Retrieving the refractive index, emissivity, and surface temperature of polar sea ice from 6.9 GHz microwave measurements: A theoretical development, *J. Geophys. Res. Atmos.*, *120*, 2293–2305, doi:10.1002/ 2014JD022481.

Lin, B., and W. B. Rossow (1994), Observations of cloud liquid water path over oceans: Optical and microwave remote sensing methods, *J. Geophys. Res., 99*, 20907–20927, doi:10.1029/94JD01831.

Liu, C., and E. J. Zipser (2009), ‘‘Warm rain’’ in the tropics: Seasonal and regional distribution based on 9 years of TRMM data. *J. Climate*, *22*, 767–779, doi:10.1175/2008JCLI2641.1.

Liu, Q., and F. Weng (2013), Using advanced matrix operator (AMOM) in community radiative transfer model, *IEEE J. Sel. Topics Appl. Earth Obs. Rem. Sens., 6*, 1211-1218.

Liu, Q. and S. Boukabara (2014), Community radiative transfer model (CRTM) applications in supporting the Suomi National Polar-orbiting Partnership (SNPP) mission validation and verification, *Remote Sens. Env., 140*, 744-754, doi:10.1016/j.rse.2013.10.011.

Liu, W. T., W. Tang, and F. J. Wentz (1992), Precipitable water and surface humidity over global oceans from special sensor microwave imager and European center for medium range weather forecasts, *J. Geophys. Res.*, *97*(C2), 2251-2264.

Maeda, T., Y. Taniguchi, and K. Imaoka (2016), GCOM-W1 AMSR2 level 1R product: Dataset of brightness temperature modified using the antenna pattern matching technique, *IEEE Trans. Geosci. Remote Sens.*, *54*, 2, 770-782.

Manaster, A. C. O’Dell, and G. Elsaesser (2017), Evaluation of cloud liquid water path trends using a multidecadal record of passive microwave observations, *J. Clim.*, doi:10.1175/JCLI-D-16-0399.1, in press.

Markus, T. M. and D.J. Cavalieri (2009), The AMSR-E NT2 sea ice concentration algorithm: Its basis and implementation, *J. Rem. Sens. Soc. Japan*, *29*(1), 216-225.

Mätzler, C., P. W. Rosenkranz, and J. Cermak (2010), Microwave absorption of supercooled clouds and implications for the dielectric properties of water, *J. Geophys. Res., 115*, D23208, doi:10.1029/2010JD014283.

Mears, C. A., M. C. Schabel, and F. J. Wentz (2003), A Reanalysis of the MSU channel 2 tropospheric temperature record, *J. Clim.*, *16*, 3650-3664.

Meier, W. N., F. Fetterer, J. S. Stewart, and S. Helfrich (2015), How do sea-ice concentrations from operational data compare with passive microwave estimates? Implications for improved model evaluations and forecasting, *Annals Glaciol.*, *56*, 69, 332-340.

Meier, W. N. and A. Ivanoff, Intercalibration of AMSR2 NASA Team 2 algorithm sea ice concentrations with AMSR-E slow rotation data, *IEEE J. Sel. Topics Appl. Earth Obs. Rem. Sens.,* in review.

Meissner, T. and F. J. Wentz (2002), An updated analysis of the ocean surface wind direction signal in passive microwave brightness temperatures, *IEEE Trans. Geosci. Remote Sens.*, *40*(6), 1230–1240.

Meissner, T., and F. Wentz (2012), The emissivity of the ocean surface between 6 and 90 GHz over a large range of wind speeds and earth incidence angles, *IEEE Trans. Geosci Remote Sens.*, *50*(8), 3004–3026, doi:10.1109/TGRS.2011.2179662.

Munchak, S. J., and C. D. Kummerow (2011), A modular optimal estimation method for combined radar-radiometer precipitation profiling, *J. Appl. Meteorol. Clim.*, *50*, 434-483, doi: 10.1175/2010JAMC2535.1.

Munchak, S. J., and G. Skofronick-Jackson (2013), Evaluation of precipitation detection over various surfaces from passive microwave imagers and sounders, *Atmos. Res.*, *131*, 81-94.

Munchak, S. J., R. Meneghini, M. Grecu, and W. Olson (2016), A consistent treatment of microwave emissivity and radar backscatter for retrieval of precipitation over water surfaces, *J. Atmos. Ocean. Tech.*, *33*, 215-229, doi:10.1175/JTECH-D-15-0069.1.

Njoku, E. (1982), Passive microwave remote sensing of the Earth from space: a review, *Proc. IEEE*, *70*, 7, 728-750.

Njoku, E. G., T. J. Jackson, V. Lakshmi, T. K. Chan, and S. V. Nghiem (2003), Soil moisture retrieval from AMSR-E, *IEEE Trans. Geosci. Remote Sensing*, *41*, 215–229.

O'Dell, C. W., F. J. Wentz, and R. Bennartz (2008), Cloud liquid water path from satellite-based passive microwave observations: A new climatology over the global oceans, *J. Clim., 21*, 1721–1739.

O'Gorman, P. A., and C. J. Muller (2010), How closely do changes in surface and column water vapor follow Clausius-Clapeyron scaling in climate change simulations?, *Environ. Res. Lett., 5*, 025207, doi:10.1088/1748-9326/5/2/025207.

Parkinson, C. L. (2003), Aqua: An Earth-observing satellite mission to examine water and other climate variables, *IEEE Trans. Geosci. Rem. Sens.*, *41*(2), 173-183, doi:10.1109/TGRS.2002.808319.

Parkinson, C. L. and J. C. Comiso (2013), On the 2012 record low Arctic sea ice cover: Combined impact of preconditioning and an August storm, *Geophys. Res. Lett.*, *40*, 1356-1361.

Petty, G. W., and K. Li (2013), Improved passive microwave retrievals of rain rate over land and ocean. Part I: Algorithm description, *J. Atmos. Ocean. Tech.*, *30*, 2493-2508, doi: 10.1175/JTECH-D-12-00144.1.

Rapp, A. D., M. Lebsock, and C. Kummerow (2009), On the consequences of resampling microwave radiometer observations for use in retrieval algorithms, *J. Appl. Meteorol. Climatol., 48*, 1981-1993, doi:10.1175/2009JAMC2155.1.

Rienecker, M. M., and Coauthors (2011), MERRA: NASA's modern-era retrospective analysis for research and applications, *J. Clim., 24*, 3624–3648.

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, *J. Clim., 20*, 5473-5496.

Robertson, F.R., M.G. Bosilovich, J.B. Roberts, R.H. Reichle, R. Adler, L. Ricciardulli, W. Berg, and G.J. Huffman (2014), Consistency of estimated global water cycle variations over the satellite era, *J. Climate, 27*, 6135–6154, doi:10.1175/JCLI-D-13-00384.1.

Rodgers, C. D. (1976), Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation, *Reviews Geophys. Space Phys.*, *14*, 4, 609-624.

Rodgers, C. D. (2000), Inverse Methods for Atmospheric Sounding: Theory and Practice, World Scientific Publishing Co., Singapore.

Seethala, C., and Á. Horváth (2010), Global assessment of AMSR-E and MODIS cloud liquid water path retrievals in warm oceanic clouds, *J. Geophys. Res.*, *115*, D13202, doi:10.1029/2009JD012662.

Sherwood, S. C., R. Roca, T. M. Weckwerth, and N. G. Andronova (2010), Tropospheric water vapor, convection, and climate, *Rev. Geophys., 48*, RG2001, doi:10.1029/2009RG000301.

Skofronick-Jackson, G. M., B. T. Johnson, and S. J. Munchak (2013), Detection thresholds of falling snow from satellite-borne active and passive sensors, *IEEE Trans. Geosci. Remote Sens., 51*, 7, 4177-4189.

Skofronick-Jackson, G., and Coauthors (2017), The Global Precipitation Measurement (GPM) mission for science and society, *Bull. Amer. Meteorol. Soc.*, in press, doi: 10.1175/BAMS-D-15-00306.1.

Spencer, R. W., H. M. Goodman, and R. E. Hood (1989), Precipitation retrieval over land and ocean with the SSM/I: Identification and characteristics of the scattering signal, *J. Atmos. Ocean. Tech.*, *6*, 254-273.

Staelin, D. H. (1969), Passive remote sensing at microwave wavelengths, *Proc. IEEE*, *57*, 4, 427-439.

Staelin, D. H., A. L. Cassel, K. F. Kunzi, R. L. Pettyjohn, R. K. L. Poon, P. W. Rosenkranz, and J. W. Waters (1975), Microwave atmospheric temperature sounding: effects of clouds on the Nimbus-5 satellite data, *J. Atmos. Sci.*, *32*, 1970-1976.

Staelin, D. H. (1977), Inversion of passive microwave remote sensing data from satellites, in: Inversion Methods in Atmospheric Remote Sounding, edited by: Deepak, A., Academic, New York, 361–394.

Stephens, G. L., and Coauthors (2002), The CloudSat mission and the A-Train: A new dimension of space-based observations of clouds and precipitation, *Bull. Amer. Meteorol. Soc., 83*, 12, 1771-1790.

Stephens, G. L. and J. M. Haynes (2007), Near global observations of the warm rain coalescence process, *Geophys. Res. Lett., 34*, L20805, doi:10.1029/2007GL030259.

Stephens, G. L., and C. D. Kummerow (2007), The remote sensing of clouds and precipitation from space: A review, *J. Atmos. Sci., 64*, 3742-3765.

Stephens, G. L., T. L'Ecuyer, R. Forbes, A. Gettlemen, J.-C. Golaz, A. Bodas-Salcedo, K. Suzuki, P. Gabriel, and J. Haynes (2010), Dreary state of precipitation in global models, *J. Geophys. Res., 115*, D24211, doi:10.1029/2010JD014532.

Trenberth, K. E., J. Fasullo, and L. Smith (2005), Trends and variability in column-integrated atmospheric water vapor, *Clim. Dyn., 24*, 741–758, doi:10.1007/s00382-005-0017-4.

Turk, J. (2016), CloudSat-GPM Coincidence Dataset version 1C, NASA Technical Report, Pasadena, CA.

Turner, D. D., and Coauthors (2007), Thin liquid water clouds: Their importance and our challenge, *Bull. Am. Meteorol. Soc., 88*(2), 177–190, doi:10.1175/BAMS-88-2-177.

Ulbrich, C. W. (1983), Natural variations in the analytical form of the raindrop size distribution, *J. Climate Appl. Meteorol.*, *22***,** 1764–1775.

Weng, F. (2007), Advances in radiative transfer modeling in support of satellite data assimilation, *J. Atmos. Sci., 64*, 3799-3807, doi:10.1175/2007JAS2112.1.

Wentz, F. J. (1992), Measurement of oceanic wind vector using satellite microwave radiometers, *IEEE Trans. Geosci. Rem. Sens.*, *30*(5), 960–972, doi:10.1109/36.175331.

Wentz, F. J. (1997), A well-calibrated ocean algorithm for special sensor microwave/imager, *J. Geophys. Res., 102*, 8703–8718, doi:10.1029/ 96JC01751.

Wentz, F. J. and R. W. Spencer (1998), SSM/I rain retrievals within a unified all-weather ocean algorithm, *J. Atmos. Sci.*, *55*, 1613-1627.

Wentz, F. J. and T. Meissner (2000), AMSR ocean algorithm, version 2, Remote Sens. Syst., Santa Rosa, CA, USA, Tech. Rep. 121599A-1.

Wentz, F. J., L. Ricciardulli, K. Hilburn, and C. Mears (2007), How much more rain will global warming bring?, *Science*, *317*, 233-235.

Weston, P. P., W. Bell, and J. R. Eyre (2014), Accounting for correlated error in the assimilation of high resolution sounder data, *Q. J. R. Meteorol. Soc.*, *140*, 2420–2429, doi:10.1002/qj.2306.

Wilheit, T. T., A. T. C. Chang, M. S. V. Rao, E. B. Rodgers, and J. S. Theon (1977), A satellite technique for quantitatively mapping rainfall rates over the oceans, *J. Appl. Meteorol.*, *16*, pp. 551-560.

Wilheit, T. T., and A. T. C. Chang (1980), An algorithm for retrieval of ocean surface and atmospheric parameters from the observations of the scanning multichannel microwave radiometer, *Radio Sci.*, *15*(3), 525–544, doi:10.1029/RS015i003p00525.

Wilheit, T., C.D. Kummerow, and R. Ferraro (2003), Rainfall algorithms for AMSR-E, *IEEE Trans. Geosci. Remote Sensing*, *41*, pp. 204–214.

Woodruff, S. D., and Coauthors (2011), ICOADS Release 2.5: Extensions and enhancements to the surface marine meteorological archive, *Intl. J. Climatol.*, *31*(7), 951–967, doi:10.1002/joc.2103.

Yu, H., P. E. Ciesielski, J. Wang, H.-C. Kuo, H. Vömel, and R. Dirksen (2015), Evaluation of humidity correction methods for Vaisala RS92 tropical sounding data, *J. Atmos. Ocean. Tech.*, *32*, 397-411, doi:10.1175/JTECH-D-14-00166.1.

Zabolotskikh, E., L. M. Mitnik, and B. Chapron (2015), Radio-frequency interference identification over oceans for C- and X-band AMSR2 channels, *IEEE Geosci. Rem. Sens. Lett.*, *12*, 8, 1705-1709.

APPENDIX

Whether or not the gamma DSD parameters (Section 3.3.3) are retrievable from a passive microwave platform is central to the algorithmic assumptions made in Chapter 3 for variational warm rain retrieval. Fig. 3.4 demonstrates that RWC is more strongly tied to observed radiance variations. The argument is thus that, with limited information and a limited number of retrievable parameters, there simply is not enough information content to retrieve the rain DSD in addition to other significant parameters. The binary stratiform/convective DSD switch in the algorithm described within Chapter 3 is built on this idea.

To test retrieving the DSD, an experiment was performed in which the gamma parameters μ and D_0 were added to the retrieved state vector. This was done for the case shown in Figs. 3.7 and 3.10, shallow rainfall near France on March 30th 2016. RWC is already a retrieved parameter via the first principal component of RWP, and so the gamma DSD is fully described by retrieval of μ and D_0, as N_w can be solved for given the other three parameters (Eq. 4 in Bringi et al. 2003). In the iteration μ and D_0 are allowed to vary from -2 to 10 and 0.4 mm to 2.5 mm, respectively. The *a priori* standard deviations were derived from the same in-situ disdrometer data used in Chapter 3, with *a priori* values of 1 for μ and 1.5 mm for D_0. The number of retrieved parameters was increased to 6 in the warm rain retrieval, simply adding these two DSD parameters to the state vector.

Fig. A.1 shows the resultant rain rates and χ^2 for the original retrieval (with fixed DSD) and the DSD-varying version. The main characteristics of the scene do not change from one retrieval to the other despite adding two retrieved parameters. The number of pixels that reach convergence barely changes, and the resultant rain rates are similar if not almost identical. From

this set of plots one might conclude that adding DSD parameters to the state vector is neither harmful nor particularly helpful as the fit to observed radiances barely shifts for pixels where the warm rain retrieval is run and has little effect on rain rates.

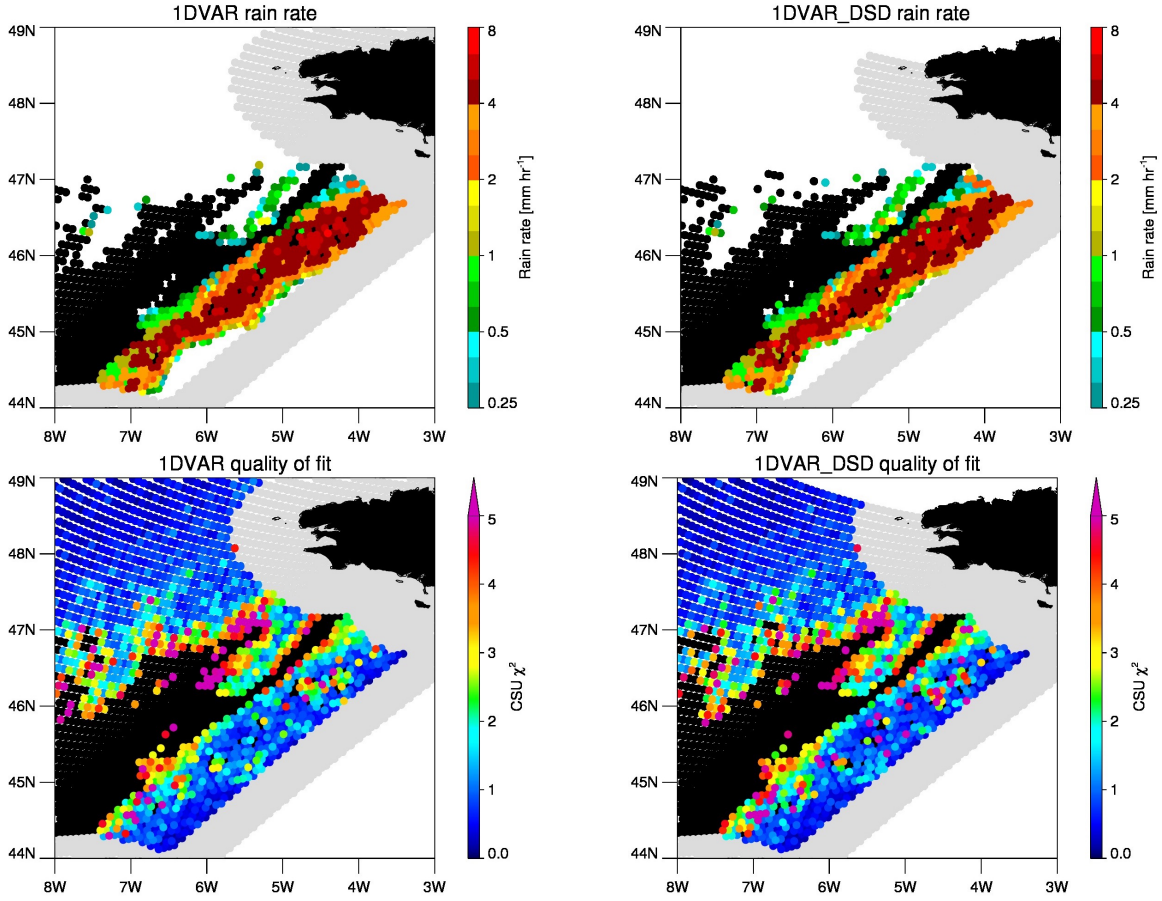

Figure A.1 Rain rates (top) and χ^2 (bottom) from the warm rain retrieval, standard version (left) and with DSD parameters retrieved (right), for the same case given in Figs. 3.7 and 3.10.

Fig A.2 plots the trace of the A-matrix, degrees of freedom for signal (DFS), for the original and DSD-varying retrievals. For a more just comparison, the standard retrieval in this case was run with four retrieved variables in the non-scattering case (i.e. first PC of water vapor only); this is why pixels in the upper left of the first panel have lower DFS in a non-raining region. Pixels for which the scattering retrieval was not run—a majority of the raining points—exhibit relative homogeneity with DFS varying around 3.0 to 3.6. Pixels with the scattering

retrieval turned on, however, exhibit large DFS variations, from around 2.5 in the standard retrieval to nearly 4.5 in the DSD-varying retrieval.

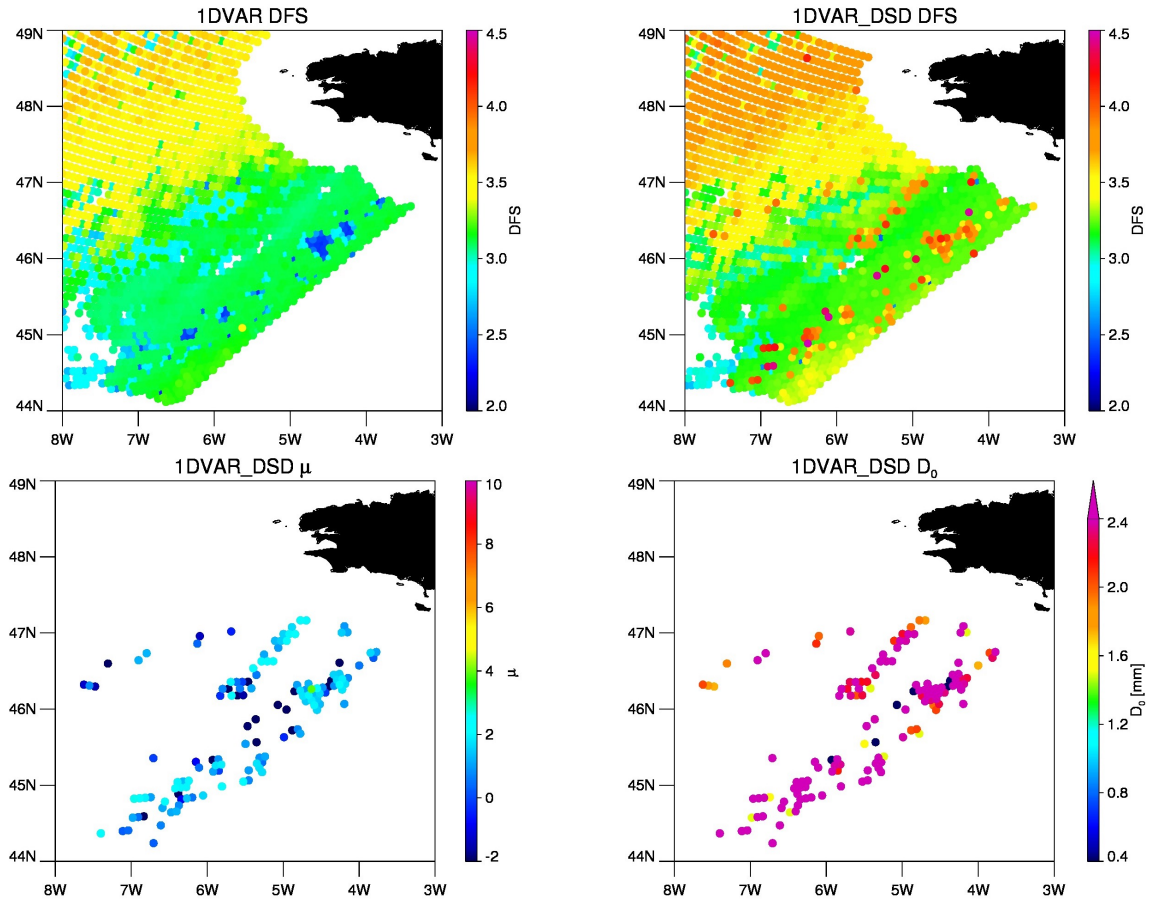

Figure A.2 Degrees of freedom for signal (DFS) from the warm rain retrieval, standard version (left) and with DSD parameters retrieved (right), for the same case given in Figs. 3.7 and 3.10. On bottom, retrieved DSD parameters for pixels where warm rain retrieval was run and converged.

What causes this stark difference in DFS seen in Fig. A.2 for pixels run through the scattering retrieval, and are the retrieved DSD parameters physically reasonable? Examination of the lower panels of Fig. A.2 shows μ values with a mean of 0.8 and fairly small variability; for D_0 the mean is 2.2 mm, also with little variability. Both parameters' standard deviations are less than those prescribed in S_a while μ is near its prior and D_0 is larger than its prior on average. These results do not seem especially physical. For one thing, values of D_0 above 2mm constitute

a small fraction of the more convective DSDs from in-situ observations (Fig. 3.3). Secondly, while the DSD of a raining volume can change on small spatial scales, it is unlikely for the shallow drizzling areas that converged acceptably with the non-scattering algorithm to be speckled with pixels containing convective-looking rain DSDs. Thus while it appears that the addition of DSD parameters to the state vector is allowing more DFS to be utilized in some cases, caution should be exercised as these results deserve some skepticism.

To conclude this experiment with retrieving the DSD in raining areas, it is determined that DSD retrieval requires further constraint but some signal to retrieve it does indeed exist within the observation vector. For very few pixels does the A-matrix yield values near 1 regarding μ and D_0, and the sum of the two is typically less than 1 as well. This indicates that while some signal exists regarding the distribution of drops within a rain volume, at most one parameter should be retrieved that can describe the distribution. Even then, further constraint seems necessary to get physically reasonable values from a passive-only retrieval.