

**Potential Reduction of Uncertainty in Passive Microwave
Precipitation Retrieval by the Inclusion of Dynamical and
Thermodynamical Constraints as the Cloud Dynamics Radiation
Database Approach**

by

Wing Yee Hester Leung

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

Master of Science
(Atmospheric and Oceanic Sciences)

at the

UNIVERSITY OF WISCONSIN-MADISON

2011

Abstract

In order to achieve better understanding of the hydrological cycle and the distribution of global precipitation, various microwave satellite platforms have been launched in the past to allow significant advance in precipitation measurement from directly measuring microwave radiances and reflectivity from space. Nevertheless, ambiguities in precipitation estimation from the only use of sets of brightness temperature measurements could lead to significant error. These ambiguities can be reduced with the addition of complementary data sets that until this point have not been employed in retrieval algorithms. In this paper, the potential improvements to estimating precipitation that are possible by combining observed brightness temperature measurements with other available sources of information will be investigated.

One way of passive microwave precipitation retrieval for the satellite-borne microwave radiometers is to be accomplished by the use of physical inversion-based algorithms, which uses Cloud Radiation Databases (CRDs). CRDs are composed of a large amount of vertical microphysical profiles, which are produced by various cloud resolving model simulations, and their corresponding brightness temperatures are calculated by radiative transfer model using the microphysical profiles as input.

Unfortunately, the relationship between the simulated microphysical profiles and the simulated multi-spectral brightness temperatures is not strictly unique. Therefore during precipitation retrieval, given a set of observed brightness temperatures, one can often match sets of microphysical profiles with strongly differing precipitation outcomes. To improve precipitation estimation, additional constraints are needed.

Fortunately, such constraints are virtually always available in the form of recent or short-term projections of the synoptic situation, which dramatically reduces the number of applicable profiles in the database, when the profiles include the synoptic situation in effect when the profiles were simulated. The Cloud Dynamics and Radiation Database (CDRD) is an attempt to include this additional information in the CRD to increase the available constraints in selecting applicable database entries used in the estimation procedure. This additional information includes the dynamical and thermodynamical structure of the atmosphere, which are stored as dynamical and thermodynamical tags in the CDRD. By using a Bayesian-based statistical estimation method, it is expected that more appropriate microphysical profiles can be chosen and thus precipitation retrieval uncertainties can be reduced.

In this study, the degree to which uncertainty in precipitation estimation can be reduced through the addition of these dynamic and thermodynamic constraints will be estimated quantitatively. This will be accomplished through a procedure whereby a CDRD of 120 cloud resolving model simulations will be statistically analyzed to determine the impact which several of the strongest dynamic and thermodynamic constraints have on the variance in the predicted columnar liquid water paths, ice water paths, and surface rain rates associated with simulated multichannel brightness temperatures.

Acknowledgements

I would first like to thank Dr. Gregory Tripoli, my advisor, for the opportunity to work on this project under his guidance. I greatly appreciate Greg's patience in teaching me in how to be a better scientist in conducting scientific research. I am also thankful for his encouragement and trust in me to work independently and all his guidance and ideas when I needed them. I have also enjoyed doing fieldwork research with Greg. When I was in undergraduate, Greg accompanied us to Storm Peak Laboratory at Steamboat Springs, CO, where I did a field experiment for the first time. I have also enjoyed going storm chasing with him in the summer.

I would also like to thank Dr. Eric Smith for all his assistance on this project. I have learnt a lot in analyzing scatter plots and doing statistical analyses from Eric. I am very thankful for his patience in teaching me. I would like to thank Dr. Alberto Mugnai and other members of his research group in Italy, Daniele Casella, Marco Formenton, and Paolo Sanò for their help on the radiative transfer model. It has been a great pleasure to work with them all. I would also like to thank Dr. Daniel Vimont, Dr. Ralf Bennartz, and Dr. Grant Petty for their advices and excellent suggestions on this project.

I would also like to thank Pete Pokrandt for all the assistance in solving computer problems and also for allowing me to use the computers in Room 1411 to complete all 120 cloud resolving model and radiative transfer simulations. Without Pete's help, this project would still not be completed. In addition, I would like to thank Angel Skram for helping me in the process of submitting this thesis.

I also appreciate all the friendships I have made over the past few years in the Atmospheric and Oceanic Sciences department. I would like to thank specifically Mark

Kulie, Tempei Hashino, John Rausch, and everyone else who have helped me throughout the years. I would also want to thank Emily Niebuhr and Agnes Lim for all the fun times hanging out together and baking after work.

Lastly I would like to thank my parents who have always given me constant support and understanding throughout the years. They have always trusted me in choosing what I want to work on in life. I am very thankful to have them as parents.

This work is funded by NASA grant NNX07AD43G.

TABLE OF CONTENTS

Abstract.....	i
Acknowledgements.....	iii
Table of Contents.....	v
1. Introduction.....	1
2. Scientific Background.....	5
2.1 Remote Sensing in the Microwave Region.....	5
2.1.1 Emission Method.....	6
2.1.2 Scattering Method.....	7
2.2 History of Passive Microwave Remote Sensing.....	7
2.3 Microwave Precipitation Retrieval Algorithms.....	10
2.4 Data Mining.....	16
3. Methodology.....	19
3.1 Cloud Dynamics and Radiation Database (CDRD) Modeling Systems.....	19
3.1.1 The Concept of CDRD.....	19
3.1.2 Bayes' Theory.....	19
3.2 Description of Models.....	20
3.2.1 The Cloud Resolving Model: University of Wisconsin – Nonhydrostatic Modeling System (UW-NMS).....	20
3.2.2 The Radiative Transfer Model.....	22
3.2.2.1 Radiometer Model.....	22
3.2.2.2 Surface Emissivity Models.....	23
3.2.2.3 Scattering Models.....	23
3.2.2.4 Radiative Transfer Models.....	24
3.3 Generation of Cloud Dynamics and Radiation Database (CDRD).....	25
3.3.1 Selection of Simulations.....	25
3.3.2 Generation of Microphysical Profiles.....	27
3.3.3 Dynamical Variables.....	29
3.3.4 Brightness Temperatures.....	39
4. Analysis.....	39
4.1 Database Statistics.....	39
5. Conclusions and Future Work.....	66
5.1 Conclusions.....	66
5.2 Future Work.....	69
Appendix A.....	70
References.....	73

1. Introduction

Passive microwave remote sensing started about 30 years ago and has provided us tremendous amount of precipitation data. This helps us to gain valuable knowledge about precipitation systems, regional and global hydrologic cycle and to improve upon weather and climate forecasts. Smith et al. (2007) states that “globally distributed, continuous, and high-quality” precipitation “intensity, accumulation, and temporal evolution” measurements are important for a wide range of research and applications, such as short term weather forecasting and rainfall data assimilation for numerical weather prediction models, prediction of regional and global scale hydrologic cycles, monitoring global climate trends, and development of rain rate retrieval products and verification techniques for rain gauges.

Satellite rain estimate products are a valuable supplement to land-based rain gauges and radar data because they can continuously monitor the variable and spatially heterogeneous rainfall pattern over space and time domain. Moreover, there is a lack of rain gauge networks over ocean and remote land areas as well as insufficient good quality precipitation data from high precision precipitation sensors over land where they are measured. More accurate global coverage of precipitation is made possible with passive microwave remote sensing from space. This data provides important inputs for hydrological models for regional and global analyses to allow for drought and flood monitoring.

Microwave radiances sensed by remote sensing platforms over satellite footprints are mostly converted to brightness temperatures (BTs) through Rayleigh-Jeans approximation. A spectral array of observed BTs are then used to estimate precipitation

through rainfall retrieval algorithms. The Goddard Profiling Algorithm (GPROF; Kummerow et al. 2001) is a commonly used algorithm and is applied to datasets from the Special Sensor Microwave/Imager (SSM/I), Tropical Rainfall Measuring Mission TRMM Microwave Imager (TMI), and the Advanced Microwave Scanning Radiometer – Earth Observing System (AMSR-E).

BTs are matched with similar microphysical precipitation structures to estimate precipitation rates through a “retrieval algorithm”. The algorithm makes use of an a-priori database that is composed of microphysical profiles that are simulated by cloud resolving model to represent a few different types of precipitation systems’ vertical microphysical structures and properties consisting of information about the hydrometeor sizes, shapes, and distributions. These profiles are then related to microwave BTs and surface precipitation rates.

For a given set of multispectral microwave observations at a given location, there is no single unique hydrometeor profile that could match the observations. Instead, various configurations of hydrometeors could be radiometrically consistent with a set of BTs observation such that iterative methods in finding a unique solution would not essentially result in a better estimate (Smith et al., 1994). In addition, Panegrossi et al. (1998) points out that it is crucial to identify the typology of the observed precipitation event and associated it with appropriate hydrometeor profiles that are generated by simulations that share similar microphysics and environmental features in order to improve retrieval precision and accuracy of profiles.

Hoch (2006) suggests that the microphysical profiles retrieved from a priori Cloud Radiation Database (CRD) are all mixed from simulations of various

environmental features. Although for a CRD to be statistically significant for retrieval in a particular region, it has to include sufficient cloud resolving model simulations that can represent various regimes of precipitation that happen under different atmospheric environments and seasons for a given location. The accuracy of the retrieved microphysical profiles are improved by matching the multispectral set of observed brightness temperatures to the simulated BTs in the CRD. Hoch (2006) has proposed a new approach which uses a Cloud Dynamics and Radiation Database (CDRD), which is an extension to the CRD by including of dynamical and thermodynamical information in the form of “dynamical tags” for each individual profiles in the database. This potential information of the synoptic states of the atmosphere is relatively new to be explored in retrieval process that uses Bayesian estimation methods. This additional knowledge of a precipitation event’s synoptic situation, geographical, and temporal location will be embedded in the “dynamical tags” and to be exploited in a tag-based Bayesian data mining technique so to be used as environmental constraints during the Bayesian retrieval process to select a more atmospheric dynamically relevant subset of microphysical profiles that are more consistent with the atmospheric environment in which the precipitation event occurs. Several studies (e.g., Hoch, 2006; Casella et al., 2009; Sanò et al., 2010) have shown that there is potential to reduce variability of retrieved microphysical profiles by including the dynamical tags in the retrieval process and thus increase the accuracy of the retrieved microphysical profiles.

This study will focus on determining quantitatively how much particular dynamical tags used in conjunction with short term model predictions and satellite derived brightness temperatures observations can potentially reduce uncertainty in the

diagnosis of microphysical properties and associated precipitation rates. This thesis is organized as follows. Chapter 2 reviews the scientific background of this research. Basic concepts of passive microwave radiation transfer through precipitating clouds over land and ocean will be described. Then the history of passive microwave remote sensing of precipitation will be presented, with discussions of both the currently available and future passive microwave remote sensing missions. Microwave imagers that are on board as part of the mission include SSM/I, AMSR, TRMM, and the upcoming Global Precipitation Measurement (GPM). Furthermore the different types of precipitation retrieval algorithms will be summarized. Commonly used data mining techniques will be discussed.

Chapter 3 explains the CDRD concept in more detail. Microphysical profile, dynamic and thermodynamic tag variables are presented. Database tags are selected based on their ability to distinguish differing atmospheric environments. In addition, the tools that are used in this study to construct the CDRD system are described in this chapter. They are: 1) Bayesian theorem, 2) A cloud resolving model: University of Wisconsin – Nonhydrostatic Modeling System (UW-NMS), and 3) A radiative transfer model.

Chapter 4 displays database statistics through scatter plots and correlation coefficients between the dynamic tags and the targeted microphysical variables (TMVs). Multiple linear regression models are used to determine the additional variances of each of the TMVs that a dynamic tag or a combination of tags could explain in addition to what the multichannel BTs could explain. Results will be presented and discussed in this chapter. Chapter 5 offers conclusions and suggests future work.

2. Scientific Background

2.1 Remote Sensing in the Microwave Region

Passive microwave instruments sense terrestrial radiation from about 5- to 200-GHz. They can observe clouds, precipitation, and water vapor, monitor land and sea surfaces, and get temperature and humidity profiling of the atmosphere. Microwave remote sensing of precipitation is achieved by sensing microwave radiances observed at the top of atmosphere, which majorly come from two sources: the Earth's surface and atmospheric constituents. The radiances emitted from the surface differ depending mainly on the type of surfaces (ocean or land) and the temperature of the surface. Atmospheric constituents such as water vapor, liquid and frozen hydrometeors can absorb, emit, and scatter radiation to contribute to changes in radiances observed at the top of the atmosphere. In other words, precipitation estimates by microwave remote sensing involves sensing the cloud water droplets below the freezing level, rain water droplets below the cloud, and ice particles in the cloud above the freezing level. Lower frequency (below 50-GHz) Microwave channels are more sensitive to thermal emission from liquid water droplets, so this emission effect dominates the atmospheric effects while higher frequency (above 50-GHz) channels are affected more by the scattering of ice particles. Various cloud and precipitation particle properties such as size, shape, and vertical distribution can affect the emission and scattering signatures. Through the distinctive differences between dominate effects in the atmosphere between liquid and ice particles reveal by BTs signatures of both lower and higher frequency channels, algorithms were first developed based on the emission method that utilizes emission information from the

lower frequency channels, and the scattering method that gets scattering information from the higher frequency channels.

2.1.1 Emission Method

The emission method is employed over the ocean at frequencies less than 50-GHz. Radiances that are emitted by the ocean can be represented as ϵT , where ϵ , the ocean surface emissivity is relatively low ($\epsilon \leq 0.5$), and T , the sea surface temperature is around 300K. Therefore the ocean has a uniform radiometrically cold background and the atmosphere is highly transparent under most circumstances. Any raindrops and cloud droplets from a precipitating cloud, water vapor, and oxygen over the ocean would absorb and emit radiation at their own thermodynamic temperature and thus increase the observed BTs at the top of the atmosphere can form a significant contrast to the cold ocean background. The absorption and emission is proportional to the droplets' masses. By holding the ocean surface temperature and emissivity constant, increase rain rate or increasing cloud thickness would result an increase in BTs seen at the top of the atmosphere. This method does not apply very well over land because of the more highly variable land surface emissivity ($\epsilon \sim 0.9$). It is more difficult to discriminate the radiances emitted by cloud water and rain droplets from the radiometrically warm surface background due to the lack of contrast.

There are studies that show how each frequency's BTs behave with increasing rain rate (Wu and Weinman, 1984 and Adler et al., 1991). At 10-GHz, the BTs are more strongly sensitive to liquid droplets but not ice particles. At 19-GHz, the BTs start to decrease sooner than those at 10-GHz with increasing rain rates because of the stronger effect of ice due to shorter wavelength. At 36-GHz, the effect of ice is even more

dominant and the BTs start to be depressed sooner than those at 19-GHz. The BTs mainly respond to stronger rain rates at this frequency than in lower frequencies. At higher rain rates, the sensitivity to rain rates decrease until it reaches zero as the column is saturated for low frequency channels.

2.1.2 Scattering Method

This method is mainly used over land using frequencies above 50-GHz. In higher frequencies, the emission effects on upwelling BTs are no longer dominant, instead scattering from ice particles above the freezing level of clouds play a larger role in the contribution of lowering BTs seen at the top of the atmosphere. The presence of ice particles help to scatter the upwelling radiation from the surface and liquid hydrometeors and depress the observed BTs. The rain estimates that use this method are not as direct as those using the emission method because it senses the amount of ice in a column but not the amount of rain itself.

2.2 History of Passive Microwave Remote Sensing

Passive microwave remote sensing of precipitation first started in late 1970s after Nimbus-5 Electrically Scanning Microwave Radiometer (ESMR-5) was launched. In 1978, the first multispectral passive microwave radiometer, the Nimbus 7 Scanning Multichannel Microwave Radiometer (SMMR) was launched. Almost a decade later, the Special Sensor Microwave/Imager (SSM/I; Hollinger et al., 1987), a sun-synchronous polar orbiting satellite as part of the Defense Meteorological Satellite Program (DMSP) was launched in 1987, which not only has increased the quality of data but also further nourished the development of rain rate retrieval algorithms for use in operational passive microwave satellite sensors (Hollinger, 1989; 1991).

The launch of Tropical Rainfall Measuring Mission (TRMM; Kummerow et al., 1998) on 27th of November 1997 has marked another important point in passive microwave remote sensing. As the name hints, the main purpose of the mission was to provide data over the tropical regions of the globe. A lot of understanding in tropical rainfall has been accomplished through this mission. TRMM has the TRMM Microwave Imager (TMI) and the Precipitation Radar (PR) onboard. The frequencies on TMI is similar to those on SSM/I, but TMI has the extra 10.7-GHz channel, which is designed to give a more linear response for high rain rates associated with tropical precipitation systems. The higher spatial resolution and wider swath width of TMI make it better than SSM/I. The TMI featured coverage of the tropics about 1 to 2 times per day depending on the latitude. PR is incredibly helpful in improving retrievals from algorithms. One of the key features of the PR is its functionality in providing three-dimensional maps of storm structures. Beside the TMI and the PR, TRMM also includes a Lightning Imaging Sensor and a Clouds and The Earth's Radiant Energy System.

The Advanced Scanning Microwave Radiometer – Earth Observing System onboard Aqua (AMSRE; Kawanishi et al., 2003) was launched on the 4th of May 2002 and it has a higher spatial resolution that could improve precipitation retrieval in comparison to older devices. Table 1 shows how the characteristics of all sensors with microwave remote sensing capability and the microwave instruments evolved since their start of the new precipitation-measuring era. There is an improvement in spectral coverage, swath width, and spatial resolution due to improvements in the reflector and the addition of 6.295-GHz channels. The swath is 1445 km. The AMSRE measures horizontally and vertically polarized BTs at six different frequencies and its function is to

retrieve data consisting of variables related to the precipitation but also things like sea surface winds, temperature and ice concentrations.

Microwave Imager	Operating Period	Type of Scan	Central Frequencies (GHz) & Polarization	FOV (km x km)	Swath Width (km)
SSM/I	1987-present	Conical	19.35 V + H	69 x 43	1400
			22.235 V	60 x 40	
			37.0 V + H	37 x 29	
			85.50 V + H	15 x 13	
TMI	1997-present	Conical	10.65 V + H	37 x 63	780
			19.35 V+ H	18 x 30	
			21.3 V	18 x 23	
			37.0 V+ H	9 x 16	
			85.5 V + H	5 x 7	
AMSR-E	1998-2020	Conical	6.925 V + H	43 x 75	1600
			10.65 V + H	29 x 51	
			18.7 V + H	16 x 27	
			23.8 V	14 x 21	
			36.5 V+H	9 x 14	
			89.0 V + H	4 x 6	
GPM	Expect launch date: July 2013	Conical	10.65 V + H	19 x 32	850
			18.70 V + H	11 x 18	
			23.80 V	9 x 15	
			36.5 V + H	9 x 14	
			89.0 V+H	4 x 7	
			166		
			183		

Table 1. Current and future satellite platforms information.

Global Precipitation Measurement (GPM; Smith et al., 1994; Smith et al., 2007) is the forthcoming satellite mission that is expected to be launched in 2013 and will bring improvements in precipitation monitoring and to also improve understanding of the

precipitation physics globally. One of the goals set for the mission is also to try to provide freshwater availability indicators. It involves international collaborations between space agencies, research and hydro meteorological forecast services, various US, Japanese, and European research teams, and individual scientists (Smith et al, 2007). The GPM center constellation will include a core satellite, with a dual-frequency precipitation radar (DPR) and a multichannel microwave imager (GMI), which is similar to the TRMM design but only with better radar capabilities, so to have greater measurement sensitivity to light rain and cold-season solid precipitation. Moreover GPM will have an orbit that will cover not only the tropics, but to higher latitudes of 65-70°. GPM will use the constellation of operational radiometers to provide global, three hourly precipitation products. The communication between the satellites will be through a transfer standard for inter-calibration of constellation radiometers.

2.3 Microwave Precipitation Retrieval Algorithms

The development of microwave precipitation retrieval algorithms for operational use on microwave sensors has flourished since the launch of DMSP SMM/I in 1987 and has been ongoing research for the last 25 years. Wilheit et al. (1977) is one of the earliest algorithms designed by using a single spectral measurement, 19-GHz or 37-GHz channels, to estimate a single rainfall parameter through a BT-rain rate relationship.

Since late 1980s, there emerge a few main categories of algorithms to estimate surface rain rates: 1. Statistically-derived algorithms, 2. Quasi-physical algorithms and 3. Physical inversion-based algorithms. Many statistically-derived algorithms are based on each channel's response to precipitation-sized particle in its effects on upwelling

radiation (Kidd et al., 1998). If the size of the precipitation particle is small compared to the wavelength of the radiation, emission effect that alters the upwelling radiation dominates; but once the size of the precipitation particle is more comparable to the wavelength of the radiation, the scattering effects dominate in causing extinction of upwelling radiation (Fowler et al., 1979). Statistical regressions between measured single channel or multichannel BTs dataset and rainfall amounts from rain gauges or radar measurements are derived and used in this type of algorithm (Smith et al., 1998). Berg and Chase (1992) is an example of this type of algorithm that uses the lower frequencies channels, 19-, 22-, and 37-GHz BTs as independent variables to capture the emission effects on upwelling radiation caused by the liquid precipitation particles. Todd and Bailey (1995) is another example that utilizes a single channel, 85-GHz for its dominated scattering signals, to estimate rainfall in the mid latitudes. A polarization corrected temperature has been formulated to eliminate the radiation variability contributed by surface emissions. Kidd et al. (1998) discusses the advantages and disadvantages of this type of algorithm and they can be summarized as follows. One major disadvantage is that there is a predominance of light rain rates than heavy rain rates in the observations, which in turn would make the statistical relationships for heavy rain rate insignificant. Moreover, statistically-derived empirically calibrated algorithms are not stable with regard to retrieval accuracy because of variations in BTs, therefore the rain rate relationship can vary depending on the physical mechanisms that cause the precipitation in the situation (Mugnai et al., 1993). The main advantage of this type of algorithm is that it uses simple formulations and does not require heavy computing resources.

Quasi-physical algorithms estimate rain rate through theoretically-derived functions of rain rates and BTs by the use of radiative transfer calculations and cloud models (Smith et al., 1998). Alder et al (1993; 1994) uses cloud model generated rain rates and radiative transfer calculated 85-GHz BTs to generate a linear regression to be used in the retrieval. Alder's algorithm is also an example of scattering algorithm, since only 85-GHz is being considered. Liu and Curry (1992) present an algorithm that is derived from the results of a radiative transfer model of plane-parallel clouds and both emission and scattering signatures to determine the amount and the nature of the precipitation. Horizontally polarized brightness temperatures at 19- and 85-GHz are used to form a linear function, which is used as a parameter to relate to rain rates. Spencer et al. (1989) introduces an algorithm that uses scattering information taken from the polarized corrected temperature, which is derived from radiative transfer calculation considering the dual-polarization 85-GHz brightness temperatures. Petty (1994b) notes that this type of algorithms only requires simple algebraic and logical operations thus they do not require heavy computing resources. However, these algorithms have not included any processes that can differentiate and alter BTs-rain rate relationships dynamically that are caused by varying precipitation microphysics and spatial variability of precipitation in different precipitating environments and background BTs differences, which are associated with the varying surface background types (Petty, 1994b).

Several studies (e.g., Smith and Mugnai, 1988, 1989; Smith et al., 1992a) have shown that multichannel microwave BTs have a more direct relationship with the vertical distribution and amount of various hydrometeors than surface rain rate. These studies gave rise to the development of multichannel physical inversion-based algorithms (e.g.

Olson, 1989; Mugnai et al, 1993; Kummerow and Giglio, 1994; Petty, 1994a; Kummerow et al., 2001), which use different frequencies to detect microphysical quantities and distributions at different levels. Physical inversion-based algorithms retrieve the rain rate and/or vertical distribution of various hydrometeor categories via multichannel BTs inversion. Some algorithms in this category might retrieve vertical profiles of various hydrometeors first, before rain rate is being diagnosed from the retrieved profiles. They use an a-priori database that includes detailed hydrometeor profiles that are part of cloud resolving model's simulations, coupled with explicit radiative transfer calculations for each of the profiles to yield the associated multichannel BTs. During the retrieval, probabilistic methods such as the Bayesian method are used to estimate rain rate. This method does not only provide one single solution, instead it will be able to provide a probability distribution of solutions that are most likely to be applicable to the atmospheric state at the time at which the rain rate is being retrieved (Stephens and Kummerow, 2007).

The GPROF algorithm, which is the operational retrieval algorithm for TMI, also uses a physical Bayesian approach (Kummerow et al., 1996, 2001). It uses a-priori large database of hydrometeor profiles that are generated by simulations by cloud resolving models and each hydrometeor profile's associated upwelling microwave BTs are calculated through radiative transfer calculations. During retrieval, the whole database of hydrometeor profiles is scanned to match a given set of observed multichannel microwave BTs to the profiles in the database that correspond more consistently with those observed BTs. Olson et al. (2007) describes that surface precipitation rates, latent heat profiles, scattering indices and polarization indices by Petty (1994a), and a term to

differentiate area fraction of convective and stratiform rain, which depends on the size of the satellite footprint, and the freezing level are all GPROF estimated profile parameters. Those precipitation and latent heat profiles that have an associated set of BTs that are radiatively consistent with the observed BTs contribute more strongly in the final estimation of rain rates. This type of algorithm is more complex, requires heavy computational resources, and a lot of microphysical assumptions have to be made in the microphysical profiles simulating process by cloud resolving models and also in the forward radiative transfer calculation (Petty, 1994b). However, this type of algorithm is able to consider the BTs changing relationship with rain rates that is due to the different dynamics involved in various types of precipitation systems.

Petty (1994b) describes another type of physical inversion-based algorithm for retrieving rain rate over the ocean with SSM/I that does not require the use of microphysical assumptions. Instead of directly inverting raw BTs, it inverts the normalized polarizations for 19.35-, 37-, and 85.5-GHz together with an 85.5-GHz scattering index, which is sensitive to the ice particles. The normalized polarizations have more direct relationships to the amount of column optical depth that is affected by the amount of liquid water present, because the normalized polarization indices can help to factor out the radiation extinction only due to liquid water from the part that is due to polarized ocean background and the presence of water vapor and ice aloft. Background variability is caused by differences in surface wind speeds and roughness, as explained in Petty (1994a). It mainly uses 19-GHz and 37-GHz for rain rates retrieval, and 85.5-GHz to provide more information in heavy rain conditions.

Tassa et al. (2003) presents another physical inversion-based algorithm that uses the Bayesian method. The retrieval scheme is trained by outputs from simulations created by a cloud-resolving model, together with associated output multichannel microwave BTs from radiative transfer models. The output vertical hydrometeor structures and simulated BTs are stored in a cloud radiation database. The Database Matching Index (DMI) is used to evaluate the representativeness of the cloud model simulations by checking how close the match is for the observations to the simulated BTs. The DMI also calculates the percentage of observed brightness temperature pixels that would have at least one simulated point that have its Euclidean distance to the observed measurement to be minimized to a given percentage error. Model errors are quantified through the use of a minimum mean square criterion.

Smith et al. (1998) presents and discusses the results of the second WetNet Precipitation Intercomparison Project, which is a project that evaluates the performance of 20 satellite precipitation retrieval algorithms, and concluded that the bias uncertainty of many passive microwave algorithms is about $\pm 30\%$. This value of uncertainty is below than the radar and rain gauge data's uncertainty that is used in the project, therefore it is not possible to pick the best algorithm from the approach of using ground validation data (Smith et al., 1998).

The technique being investigated in this study is most applicable to physical inversion-based algorithms that employ a cloud radiation database that is generated by cloud resolving model simulations. It is because the accuracy of the results from those algorithms depends mostly on the hydrometeor profiles retrieved from the a-priori database. Updated dynamical information that could potentially be used to differentiate

different typologies of precipitation events could be readily obtained from global forecasting model every 6 hours, these dynamical information could be used as extra independent information during the selection of most appropriate hydrometeor profiles through a Bayesian method during retrieval. A major goal of this study is to investigate the usefulness of dynamical information in explaining additional variances in the retrieval rain rates, liquid and ice columnar amounts. Next, there is an attempt to determine the best combination and numbers of dynamical variables to be used that could potentially be applied in global retrieval of columnar ice and liquid amounts and rain rates.

2.4 Data Mining

Manual data mining has been around for many centuries but its just in the last couple of decades with the improvement of computer technology that scientists have been able to use really big and complex data sets for data mining. With data mining, large amounts of data are captured in databases, data that often contains large numbers of variables and relationships. Data mining is the process of analyzing data from various perspectives and to summarize the results to get useful information, including the patterns, associations, or relationships among all data points. This information can often be transformed to knowledge of historical patterns and future trends. The data mining process is mostly being done by data mining software nowadays.

Although this method is relatively new to meteorology and atmospheric science there have already been several studies that have been done with the help of data mining. Diner (2004) was able to analyze large datasets of atmospheric aerosol during the Exploratory Data Analysis and Management (EDAM) project. This study was part of the

Progressive Aerosol Retrieval and Assimilation Global Observing Network (PARAGON), which aimed for a systematic, integrated approach to aerosol observation and modeling. In another study, Li (2008) uses data mining as part of a method for real time storm detection and weather forecast activation. With the help of algorithms, it proposes a way to carry out in a continuous basis in real time over large volumes of observational data. A few most commonly used data mining techniques are listed below.

An Artificial Neural Network (ANN) is based on the biological neural network that is a component in all Eumetazoa (all animals excluding a few very simple ones). ANN uses the brain's function of learning as a model for its analytical technique. Just like a human brain, an ANN can use processed information to construct new predictions. A study including data mining and ANN is Hong (2004) who uses it to construct a neural network cloud classification system to estimate precipitation from remotely sensed measurements.

Genetic algorithms are based on the concepts of natural evolution. This method uses processes such as genetic combination, mutation, and natural selection, to retrieve the desired quantity based on an optimal set of criteria or combinations. Another method of data mining that has been around since the 1960s is a Decision Tree. Quinlan (1993) gains acknowledge for its contribution to the development of automated decision trees. The name derives from the fact that it sometimes looks like an upside down tree. It is a good and pedagogical way to display an algorithm. The decision tree model is used to analyze data and have the tree built up of rules that divides the data. There are different algorithms that can be used, for instance the Classification And Regression Trees (CART) model.

Nearest neighbor method is a technique that classifies each record in a dataset based on how similar they are in metric spaces to other points within the data warehouse. It is sometimes called the k-nearest neighbor technique. The “k “ represents the number of nearest neighbors. This method is more of a searching technique than to be used to learn about the dataset. Data visualization provides graphic tools to illustrate data relationships. It is good for visual interpretation of complex relationships in multidimensional data.

Finally, rule induction is a technique to extract useful if-then rules from data based on statistical significance. It is the best choice in mining data from CDRD based on the relationships found from CDRD, if-then rules using the dynamical variables could be deduced to retrieve a subset of microphysical profiles. In the next section, the CDRD modelling system will be described in detail.

3. Methodology

3.1 Cloud Dynamics Radiation Database (CDRD) Modeling Systems

3.1.1 The Concept of CDRD

CDRD is developed based on the Bayes' theorem (Hoch, 2006). It is an extension of the CRD, which for each realization includes dynamical and thermodynamical information of the profile in addition to vertical distributions of ice and liquid hydrometeors and associated multispectral BTs, which are already part of a CRD.

3.1.2 Bayes' Theory

Bayes' theorem for rain rate retrieval can be expressed as the following:

$$P(R|Tb) = \frac{P(R) \times P(Tb|R)}{P(Tb)} \quad (1)$$

where R represents the vertical hydrometeor profiles and Tb are the multispectral BTs. The first term at the top on the right hand side, P(R), is the probability that a certain hydrometeor profile is observed and it is being computed by the cloud resolving model. The second term, P(Tb|R), is the probability that a set of BTs is observed under the condition of also having the certain hydrometeor profile R. This probability can be computed by radiative transfer model. Bayesian retrieval algorithms employ this idea to find the term on the left hand side, P(R|Tb), which is the probability of a particular hydrometeor profile given a certain set of BTs.

During retrieval, dynamical and thermodynamical information are readily available from large scale forecasting models such as ECMWF and GFS together with the BTs measurements from a satellite overpass. Therefore, the additional dynamical and thermodynamical information can be utilized as further constraints during the data

mining process. CDRD is built upon this idea for it has included 24 dynamical tags. The Bayes' theorem equation for CDRD becomes

$$P(R|Tb,Tag) = \frac{P(R) \times P(Tb|R) \times P(Tag|R,Tb)}{P(Tb) \times P(Tag|Tb)} \quad (2)$$

in which the dynamical tag is to be used to help to classify the atmospheric state at the time during retrieval and hence a more accurate subset of profiles could be chosen to compute R. It could decrease the variance of the resulting retrieval profile.

A cloud-resolving model is used to generate the hydrometeor profiles and dynamical and thermodynamical information. The associated BTs are calculated by a radiative transfer model, which needs vertical profile of hydrometeors as input from the cloud-resolving model. Descriptions of the models used are given in the following section.

3.2 Description of Models

3.2.1 Cloud Resolving Model: University of Wisconsin – Nonhydrostatic Modeling System (UW-NMS)

The cloud model used in this study is the University of Wisconsin – Nonhydrostatic Modeling System (UW-NMS) described in Tripoli (1992) that can simulate convection and its interaction with atmospheric phenomena with horizontal scales ranging from mesoscale to synoptic-scale. Through simulating the weather events listed in Appendix A, microphysical profiles of various precipitation events are generated as part of the completion of the CDRD. This model is chosen because of its ability to achieve accuracy in simulating scale-interaction processes majorly through enstrophy and kinetic energy conservation that is imposed in the model.

This nonhydrostatic regional mesoscale model is formulated on Arakawa “C” grid with multiple two-way nesting that is being put on a local rotated spherical grid. The two-way interactive nesting scheme allows increased resolution in focused areas. The initial data for the outer grid can be interpolated from another model such as the European Centre for Medium-Range Weather Forecasts (ECMWF) model and the National Centers for Environmental Prediction (NCEP) Global Forecasting System (GFS) or from a horizontally homogenous state (Tripoli, 1992). The model employs non-Boussinesq, quasi-compressible dynamical equations. The variable ice-liquid water potential temperature is used as a predictive thermodynamics variable in the model (Tripoli and Cotton, 1981). There is an advantage to using the ice-liquid water potential temperature because it is conserved in all phase changes. Potential temperature, water vapor, and cloud water are all diagnostic variables. One unique feature of this model is that it has a terrain-following vertical coordinate with variable stepped topography. It is competent in capturing steep as well as subtle topographical features and slopes therefore can also accurately simulate the dynamics of terrain-induced flows (Tripoli, 1992; Tripoli and Smith, 2010).

UW-NMS uses a bulk microphysics scheme by Flatau et al. (1989) and Cotton et al. (1986) in each of the grids in order to predict the 3-D mixing ratios of six different hydrometeors. The six categories of hydrometeors include: 1) cloud droplets, 2) rain droplets, 3) pristine ice crystals, 4) ice aggregates, 5) low density graupel and 6) high density graupel. All particles are assumed to be spherical.

3.2.2 The Radiative Transfer Model

In order to compute the upwelling BTs as part of the CDRD, a radiative transfer model is utilized and vertical microphysical profiles, surface skin temperature, and the wind, temperature, moisture profiles of 120 simulations that are generated by the aforementioned cloud resolving model are used as inputs. The radiative transfer model used for the study is a three-dimensional (3-D) adjusted plane parallel radiative transfer scheme.

To simulate the BTs, it uses:

- 1) A radiometer model, which specifies all the characteristics of a radiometer selected.
- 2) Various surface emissivity models that include emissivity properties at different frequencies of a few land types such as land, ocean, and snow.
- 3) Various scattering models for the liquid and ice hydrometeors to calculate the optical parameters of the simulated column.
- 4) Radiative transfer model is used to compute the monochromatic upwelling radiances that a specified radiometer would observe from the top of the atmosphere at its viewing angle at full cloud resolving model resolution, considering the microphysical profiles and the selected microwave frequencies and polarizations. Then the upwelling BTs will be computed and adjusted to the resolution of the selected radiometer's channels and also take into consideration of the radiometer characteristics.

3.2.2.1 Radiometer Model

The radiometer model specifies all characteristics of a radiometer including the chosen frequencies, polarization and width of the channels, the viewing angle of the radiometer, field of view and antenna pattern of various channels, and their radiometric

noise. This is the process in defining an instrumental transfer function for each channel so to calculate the upwelling BTs from the monochromatic radiances. The channel characteristics for AMSR-E and TMI are chosen for this study.

3.2.2.2 Surface Emissivity Models

The surface emissivity models are used to best represent the different surface types of all the selected cloud resolving model simulations. The surface emissivity has a significant impact on the upwelling BTs particularly in the lower window frequencies. Frequency and polarization, observation geometry, and other surface characteristics such as land types, surface roughness, soil types, soil moisture content, etc. The three surface emissivity models that are employed in the calculation are:

- For land surfaces, a model that calculates the forest and agricultural land surface emissivity by Hewison (2001) is used;
- For ocean surfaces, a fast and accurate ocean emissivity model of English and Hewison (1998), Hewison and English (2000) and Schluessel and Luthardt (1998) is used;
- For snow cover surfaces, a snow emissivity model by Hewison and English (1999) is used.

3.2.2.3 Scattering Models

The computation of the single scattering properties of various hydrometeors is accomplished by utilizing scattering models. To compute real natural ice hydrometeors have been a primary challenge since they occurs in a wide range of sizes, densities, and

shapes. As all the particles from UW-NMS are assumed to be spherical, several assumptions are made for the single scattering computations. Liquid particles including cloud and rain droplets are assumed to be spherical and homogeneous and thus their scattering properties can be calculated by Mie theory (Bohren and Huffman, 1983).

Gaupel particles are assumed to be spherical with densities close to pure ice (0.9 g cm^{-3}). They are assumed to be equivalent homogeneous spheres that have an effective dielectric function attained from a combination of the dielectric functions of ice and air, or water in the case of melting by the effective medium Maxwell-Garnett mixing theory that is applicable to a two-component mixture of air / water in ice (Bohren and Huffman, 1983). Therefore Mie theory can be applied in this case also. Mie theory cannot be applied for the pristine crystals, as they are highly non-spherical. Neither Mie theory can be applied to snow and aggregates because they are low density particles.

3.2.2.4 Radiative Transfer Models

A radiative transfer (RT) model is used to simulate BTs that would be observed by a microwave radiometer. RT code is being applied to the microphysical outputs of the simulated precipitation events from the cloud resolving model to simulate the BTs. Since fully 3-D RT schemes are computationally expensive, a 3-D adjusted plane parallel RT scheme that is developed by Roberti et al. (1994) is used. Plane parallel cloud structures are generated from the cloud model paths in the direction along the sight of the radiometer, but not in the vertical from cloud model columns. Therefore, the RT as well is performed along a slanted profile and monochromatic upwelling radiances are being computed at the same resolution (2km) as the inner grid of UW-NMS set up. After that,

instrument transfer functions are used to compute the BTs for each channel. It is done by first integrating the monochromatic upwelling radiances over the channel width, considering also the channel's spectral response. Second is to integrate the channel upwelling radiances over the field of view, which contain all the pixels of the cloud resolving model that were included in a field of view, with the consideration of radiometer antenna pattern and radiometric noise. Vertical profiles of liquid and ice water contents, together with surface skin temperature, and vertical temperature and humidity profiles are needed in the RT process. Other inputs include information from the radiometer model, the surface emissivity model, and the single scattering model.

3.3 Generation of Cloud Dynamics and Radiation Database (CDRD)

3.3.1 Selection of Simulations

North America CDRD consists of 120 simulations from a one-year period, November 2007 to October 2008, with 10 simulations selected for each month. In order for the CDRD to be robust and useful in retrieving rain rates under various atmospheric phenomenon, it has to include all types of meteorological events that have various mesoscale and synoptic environments and dynamical forcing, which occur in diverse locations and happen at different times of the year. Precipitation systems, that are caused by large-scale dynamical forcings like mid latitude cyclones, are included in the database. Mesoscale convective systems such as squall lines, mesoscale convective complexes, convection along fronts, lake effect snow, and tropical cyclone are also included. A few orographic events are also selected. The simulations are also picked to spread over both land and ocean from the tropical latitudes to higher latitudes to eliminate land / ocean

biases. Appendix A gives more information about each simulation. Each simulation can consist of precipitation that can be classified as various precipitation regimes. For example, near the center of a mid latitude cyclone and along the warm front, it is common to see convective cloud structures embedded within stratiform cloud structures. Another example would be simulations of a passage of a cold front, both convection in the warm sector of the cyclone and the slantwise convection along the cold front would be included in the simulation. Thus, all simulations would obtain stratiform and convective cloud structures at some time as the weather system being simulated is going through its lifecycle. Fig. 1 shows the location of the simulations across North America. All the boxes shown in Fig. 1 represent the center location of the inner grid of the model.

This is just a start to build a more “complete” database. It could never be perfect because that would mean to having all precipitation events over North America for a long period of time with different seasons and years included in the database. Many more simulations are necessary to capture the enormous spatial and temporal variability of all precipitation events. This dataset only provides an initial baseline representation of the natural variability.

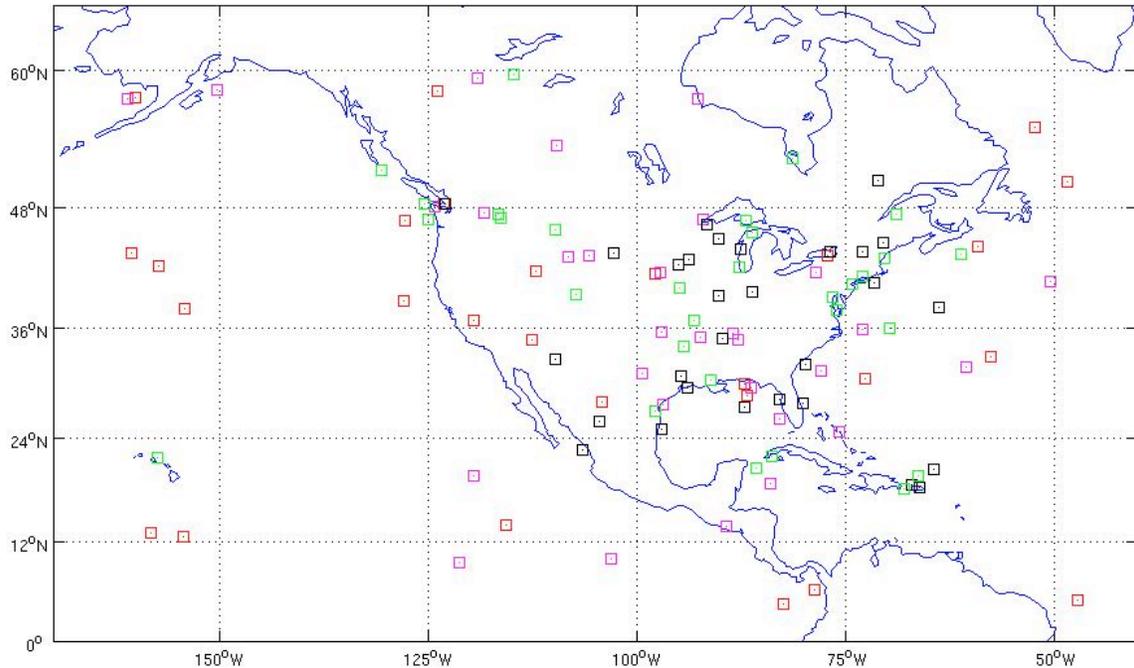


Figure 1. The location of simulations selected over North America divided by seasons. Simulations in winter, spring, summer, and autumn are in red, pink, black, and green, respectively.

3.3.2 Generation of Microphysical Profiles

After the events are being selected, the cloud-resolving model, UW-NMS is used to generate microphysical profiles and the associated atmospheric dynamical and thermodynamical variables. The model is run with three nested grids and the horizontal resolutions of the outer grid, intermediate grid, and inner grid are 50km, 10km, and 2 km, respectively. Table 2 summarizes the grid properties used in all the simulations.

NCEP GFS gridded analysis data is used to set up initial conditions for the model and to determine the outer boundary of the outer grid of the model every 6 hours of simulation time throughout the whole simulation. The simulation is set for 18 to 36 hours with a 12-hour spin-up time depending on the developing and dissipating speed of that

particular weather system to be simulated. The 12-hour spin up time is needed to allow local forcing to develop.

Grid Number	Horizontal Points	Vertical Points	Horizontal Resolution (km)	Horizontal Size (km)
1	92x92	35	50	4550x4550
2	92x92	35	10	910x910
3	252x252	35	2	502x502

Table 2. UW-NMS grid properties for all the simulations.

Hydrometeor Variables (Rain, Snow, Graupel, Aggregate, Pristine Crystals)	Other Variables	
Mixing Ratio (g kg^{-1})	Water Vapor Mixing Ratio (g kg^{-1})	Total Water Path (kg m^{-2})
Terminal Velocity (cm s^{-1})	Cloud Water Mixing Ratio (g kg^{-1})	Liquid Water Path (kg m^{-2})
Diameter (micrometer)	Zonal Wind (m s^{-1})	Ice Water Path (kg m^{-2})
Concentration ($\# \text{ cm}^{-1}$)	Meridional Wind (m s^{-1})	Height (m)
Density (g cm^{-1})	Vertical Velocity (m s^{-1})	Temperature (K)
Surface Rate (mm hr^{-1})	Surface Skin Temperature (K)	

Table 3. UW-NMS variables included in a microphysical profile.

After that, microphysical profiles, dynamical and thermodynamical variables over all grid points in the domain are to be saved hourly whenever there is one single point in the domain that has a surface rain rate of 0.01 mm hr^{-1} or greater or a surface frozen (snow, graupel, aggregates, and pristine crystals) precipitation of 0.1 mm hr^{-1} or greater.

By saving it hourly in simulation time, microphysical profiles of precipitation systems at different development stages can all be included in the database. Table 3 presents what a microphysical profile contains. This data is saved at all 36 vertical levels for each grid point except for the water paths which have just one value per profile.

3.3.3 Dynamical Variables

A total of 24 dynamic and thermodynamic variables are chosen based on their ability and potential to provide more information in helping to differentiate atmospheric states that could initiate and support various types of precipitation events. These parameters attempt to provide information on the stability of the atmosphere, the amount of mesoscale and large scale dynamical forcing and low-level moisture available, and topography influences that has an effect on the potential to promote convection. Table 4 provides a list of all the dynamical variables chosen. All the variables on Table 4 are generated by UW-NMS and are saved in 50 km grid spacing so to make them comparable to global operational forecasting model (such as ECMWF, NAM, and GFS) resolutions.

Over the last 30 years with higher model resolution available and better physical parameterization and data assimilation techniques, the initial condition error of the prediction has reduced by a significant amount and thus the predictability of global forecasting models has greatly improved. The predictability of large-scale phenomena is good in a 6-hour time frame. However, forecast models might not be able to resolve small-scale processes, such as turbulence, convection, and cloud processes. They heavily rely on model parameterizations to represent those processes.

The forecasts for large-scale synoptic forcing might be better than those for smaller scales in a situation where the large-scale synoptic forcing were dominant over the small-scale forcing effects. It is because the large-scale synoptic forecast attempt to predict at about the same resolution as in real, but the small-scale convection processes has to depend on the use of convective parameterization, which is a method to try to estimate much smaller scale convection in the much larger model grid size. In addition, the initial condition set up for the models can miss important fine scale details for convection. If small-scale effects are more important in a situation, the predictability of the large-scale synoptic forcing then might be similar to that of the smaller scale forcing because of the assumed convective parameterization that would in turn affect the forecast for large-scale humidity, temperature, and wind fields. Since the models' surface physics packages formulations to diagnose surface variables might not be applicable in all situations, it could cause errors in some situations and thus affect the predicted temperature, humidity, and wind fields.

Brunt Väisälä Frequency (s^{-1})	Convective Inhibition ($J kg^{-1}$)	CAPE ($J kg^{-1}$)	Divergence 700 hPa ($s^{-1} * 10^{-5}$)
Equivalent Potential Temperature at the surface (K)	Freezing Level (m)	Latent Heating Rate ($K day^{-1} * 10^{-5}$)	Lifted Index (K)
Lifting Condensation Level (m)	Mid-level Lapse Rate ($K km^{-1}$)	Omega 500hPa ($hPa s^{-1} * 10^{-2}$)	Omega 700 hPa ($hPa s^{-1} * 10^{-2}$)
PBL Height (m)	Potential Vorticity Advection at 250 hPa ($s^{-2} * 10^{-4}$)	Potential Vorticity Advection at 700 hPa ($s^{-2} * 10^{-4}$)	Richardson Number (unitless)
Surface Divergence ($s^{-1} * 10^{-5}$)	Surface Froude Number (unitless)	Surface Temperature (deg C)	Vertical Heat Flux ($W m^{-2}$)
Thickness 500-1000hPa (m)	Thickness 700-1000 hPa (m)	Vertical Moisture Flux 50mb above ground ($W m^{-2}$)	Vertical Wind Shear 0-6km ($m s^{-1} km^{-1}$)

Table 4. Dynamical variables selected for CDRD at 50km grid spacing.

With constant improvements in physical parameterizations used in the models, large-scale averaged dynamical variables forecasts would be accurate to be used to further categorize various precipitation systems atmospheric conditions in providing additional information than what a set of multispectral BTs could provide. Below are descriptions of each of the selected dynamical variable:

1. Brunt–Väisälä frequency (N)

$$N^2 = \left(\frac{g}{\theta_o} \right) \frac{\partial \theta}{\partial z} \quad (3)$$

N^2 is a measure of buoyant instability. N^2 is negative when it is in an unstable

equilibrium and when $\frac{\partial \theta}{\partial z}$ is negative. Convection and overturning will result.

N^2 is positive under a statically stable condition. The greater N is, the more stratified the atmosphere is, thus its more stable. N is the frequency at which the parcel will oscillate when being vertical displaced in a statically stable environment. Latent heat release during condensation can change the sign of $\frac{\partial\theta}{\partial z}$ and thus N^2 also.

It is useful to determine how likely the gravity waves are to grow or damp. Gravity waves can alter the vertical moisture gradient and trigger convective instability for initiating convection. N^2 is predicted based on the temperature structure of the atmosphere in the model.

2. Convective Inhibition (CIN)

CIN measures the amount of energy needed to lift an air parcel vertically from its original position to its level of free convection to initiate convection. The larger it is, the stronger the capping inversion is, which suppresses the development of thunderstorms. The cap is important in severe weather events because it can separate the warm, moist air below from the cool, drier air above. So potential instability can built up to a larger amount with continue surface moistening and heating by the sun to support severe weather development later in the afternoon of the day. CIN is sensitive to the thermal and moisture structure of the atmosphere predicted in the model.

3. Convective Available Potential Energy (CAPE)

CAPE measures the amount of energy available for convection. It is related to the maximum potential updraft speed and thus the larger CAPE is, the stronger potential there is for severe weather to occur. CAPE can be used to

differentiate convective precipitation from stratiform precipitation. CAPE is sensitive to the thermal and moisture structure of the atmosphere predicted in the model.

4. Divergence at 700 hPa

Divergence at 700 hPa can enhance upward vertical motion from below. It is an important large scale forcing for promoting upward vertical motion. It is sensitive to the predicted wind field in the model. It is useful to determine how strong the updrafts are in both synoptic and mesoscale environment.

5. Equivalent Potential Temperature (Θ_e)

Θ_e is the temperature of a parcel of air after all the latent heat has been released and then brought back to the reference level 1000 hPa. Therefore, it is a measure of moisture content of the parcel. Increase in dew point and temperature result in a higher theta-e and thus increase instability. Regions with high theta-e represent that there are warm moist air, which is good for severe weather development. It can also help to indicate warm moist tropical air masses from drier air masses.

6. Freezing Level (FL)

Freezing level marks the pressure level at which the temperature reaches 0 deg C. Ice formation can occur above the freezing level. It can be used to determine the amount of ice and the amount of liquid in a column. If the FL is low, the cloud contains more ice; if FL is high, the cloud contains more liquid. Since FL depends on the temperature profile of the atmosphere, it changes quickly with the change in thermal structure of the atmosphere as well by

low-level temperature advection, convection, and evaporation cooling from precipitation. It can be used to differentiate tropical from higher latitudes environment.

7. Latent Heat

Latent heat release is important for the growth and development of both synoptic and mesoscale circulations. However, mechanisms that are involved in them, including evaporation, condensation, and cloud droplet growth, are on scales that are too small to allow explicit calculations in models. They are often times approximated by some formulations that use parameters that are in resolvable scale.

8. Lifted Index (LI)

Lifted Index is the difference between the temperature of a parcel that has been lifted to 500 hPa from the surface to the environmental temperature at 500 hPa. It measures the stability of the troposphere with respect to convection that is originated near the surface. If LI is negative, it indicates instability. It is a good severe weather parameter to assess the atmospheric stability for deep convection originated from the surface. It is useful in severe weather environment.

9. Lifting Condensation Level (LCL)

LCL is the pressure level in which air reaches saturation by lifting from a pressure level below. It is used to determine the height of the cloud base from surface lifting. However, it is only useful in cases where there is surface

lifting from low-level convergence, but not for cases where the lifting is purely caused by positive buoyancy.

10. Mid Level Lapse Rate 850-500 hPa

Mid Level Lapse Rate 850-500 hPa can be altered by differential horizontal temperature advection and differential vertical diabatic heating. The change in mid level lapse rate has an important effect on the CAPE and CIN. It is used in assessing the atmospheric stability. When it is less than 5.5 to 6 degrees Celsius per km, it is stable; when it is near 9.5 degrees Celsius per km, it is absolutely unstable. In between these two values, it is considered to be conditionally unstable.

11. Omega at 500 hPa

Omega at 500 hPa is the vertical motion in p-coordinates at 500 hPa. The greater the vertical motion there is, the stronger the cloud will be and thus more intense precipitation.

12. Omega at 700 hPa

Omega at 700 hPa is the vertical motion in p-coordinates at 500 hPa. The greater the vertical motion there is, the stronger the cloud will be and thus more intense precipitation.

13. Planetary Boundary Layer (PBL) Height

PBL is the bottom layer of the troposphere that is in contact with the surface. The height of the PBL varies during the course of a day. During daytime, rising thermals and turbulent eddies in the mixed layer can deepen the depth of the PBL by the process of entrainment. Around sunset, turbulence decays

and the mixed layer is being transformed to a residual layer, which still contains the residual moisture and heat from the mixed layer that were formed from the previous day. During nighttime, the surface cools by outgoing longwave radiation and a statically stable boundary layer forms. With greater PBL depths, moisture from the surface can be mixed to higher levels. Then it allows more moisture to be evaporated into the layer, thus it provides more moisture for convective activities.

14. Positive Vorticity Advection (PVA) at 250 hPa

If there is a differential PVA in the vertical together with low-level warm air advection, then it is associated to rising motion and precipitation. It is a good parameter to suggest large-scale vertical motion.

15. Positive Vorticity Advection (PVA) at 700 hPa

If there is a differential PVA in the vertical together with low-level warm air advection, then it is associated to rising motion and precipitation. It is a good parameter to suggest large-scale vertical motion.

16. Richardson Number (Ri)

$$Ri = \frac{N^2}{\left(\frac{\partial u}{\partial z}\right)^2 + \left(\frac{\partial v}{\partial z}\right)^2} \quad (4)$$

It is a dimensionless ratio of the buoyant production of turbulence to the shear production of turbulence. It is used to indicate dynamic stability of the atmosphere. The critical value of Ri is 0.25 and that is the point in which the flow will become unstable and turbulent if Ri gets below it. It is a measure if

the wind shear is strong enough to overpower static stability in the atmosphere to produce turbulence.

17. Surface Froude Number (Fr)

Froude Number is a nondimensional parameter that is used to characterize the flow over mountain and to predict what kind of flow pattern that might occur depending on the wind speed of the flow, Brunt–Väisälä frequency, and the width of the mountain. Flows that are forced to go around the mountain have smaller values of Fr . Flows with larger values of Fr are associated with flow going on top and over the mountain. It is a good parameter to capture the topography effects on flow.

18. Surface Divergence

If it is negative, there is surface convergence and it is one of the primary lifting mechanisms for rising air to form precipitating clouds. So the stronger the convergence, the stronger the updraft will be and thus the more intense precipitation will result.

19. Surface Temperature

Surface temperature changes over various surface types. It can be used to roughly distinguish tropical latitudes from higher latitudes. It can also be used to determine the degree of surface heating.

20. Surface Vertical Heat Flux

Surface vertical heat flux measures the amount of heat transported upward directly from surface by small eddies. Stronger surface heat flux destabilized the lower levels.

21. Thickness 1000 – 500 hPa

Thickness 1000 – 500 hPa is the height between 1000 and 500 hPa pressure levels of the atmosphere. It is proportional to the mean temperature of that layer. It can be used to identify distinguish tropical latitudes from higher latitudes. The warmer the environment is the larger the thickness.

22. Thickness 1000 – 700 hPa

Thickness 1000 – 700 hPa is the height between 1000 and 700 hPa pressure levels of the atmosphere. It is proportional to the mean temperature of that layer. It can be used to identify distinguish tropical latitudes from higher latitudes.

23. Vertical Moisture Flux near surface to 700 hPa (VMF)

Vertical moisture flux combines the moisture and the vertical motion effects into one variable. It should be one of the more powerful variables because it contains information of both the amount of moisture and lifting, which are both needed for the formation of clouds and precipitation. With a great VMF, there should be more intense precipitation.

24. Wind Shear 0 – 6 km

Wind shear can affect the organization, type, longevity, and severe level of convection. With an increase of wind shear, it can help to tilt the updraft and allow precipitation to fall away from the updraft region. However if the wind shear is too large, it can be harmful to weak updrafts. It can also encourage downdrafts by the increase of mid-level entrainment. It is usually considered to be weak if it has values less than 10 ms^{-1} , and is considered to be moderate

if it has values around 10 to 20 ms^{-1} , and strong if it has values greater than 20 ms^{-1} .

3.3.4 Brightness Temperatures (BTs)

Microwave frequencies of 10.65-, 19.35-, 22.23-, 23.8-, 36.5-, 85.5-, 89.0-, and 150-GHz are included in the database. All of the channels are dual polarization except 22.23-GHz and 23.8-GHz. Microphysical profiles are needed as an input for RTM to simulate BTs in these channels. These channels are selected because they already exist in current satellite platforms.

Dynamic and thermodynamic variables can provide extra information about the synoptic situation of an event. Through a Bayesian approach, it is hypothesized that information could help to pick a more relevant subset of profiles during the process of retrieval. In order to know which variables are more correlated with microphysical variables and as a result to be more powerful in being able to minimize the variance in the retrieved microphysical profiles, statistic analyses of CDRD are performed and the results will be presented in the following section.

4. Analysis

4.1 Database Statistics

In this analysis, only realizations over water are selected because water surface has low and almost constant emissivity, allowing the increase of BTs related to the changes of the amount of liquid and ice hydrometeors in the column above the water surface to have a good contrast over the cold ocean background. Over land, emissivity is

highly variable depending on land surface types and the amount of land moisture present. Detecting changes in BTs by cloud water and raindrop emissions in the lower frequencies is more difficult as the signals are blend in with the highly emissive background.

There are a total of 2141304 (about 2.1 million) realizations over water in the database. Table 5 presents the number distribution of realization by seasons. There are already well-studied relationships between BTs in various frequencies and in various amounts of liquid columnar content, ice columnar content and rain rate (e.g. Petty, 1994a; Panegrossi et al., 1998), so liquid and ice columnar contents are important variables in the retrieval of rain rate. With a goal to determine which dynamical and thermodynamical variables have more potential to explain variances of columnar Ice Water Path (IWP), Liquid Water Path (LWP) and Rain Rate (RR) (hereafter named as Targeted Microphysical Variables (TMVs)) beyond what a set of BTs could explain, correlation coefficients between the dynamical tags and the TMVs are calculated and listed in Tables 6a, 7a, 8a, and 9a. Then the best 6 tags with the highest correlations can be identified for each of the TMVs per season as listed under Tables 6b, 7b, 8b, and 9b.

Season	Number of Realizations
Winter	837466
Spring	630640
Summer	299235
Autumn	373963

Table 5. Number of realizations distributed by season.

Correlation coefficient r represents the normalized measure of the strength of linear relationship between two variables (x and y). The value of r can vary from 1 to -1,

with 1 meaning that x and y have a strong positive linear correlation, and -1 meaning that x and y have a strong negative correlation.

	Winter	LWP	IWP	RR
1	Brunt-Väisälä frequency (N)	-0.38	-0.54	-0.36
2	Convective Inhibition (CIN)	0.17	0.30	0.14
3	Convective Available Potential Energy	0.25	0.30	0.21
4	Divergence $_{700 \text{ hPa}}$	0.25	0.03	0.21
5	Equivalent Potential Temperature Θ_e	0.31	0.40	0.25
6	Freezing Level (FL)	0.28	0.36	0.22
7	Latent Heat	0.24	0.32	0.19
8	Lifted Index (Li)	-0.04	-0.18	-0.04
9	Lifting Condensation Level (LCL)	0.09	0.29	0.10
10	Mid Level Lapse Rate	0.03	0.02	0.03
11	Omega ω $_{500 \text{ hPa}}$	-0.41	-0.67	-0.39
12	Omega ω $_{700 \text{ hPa}}$	-0.42	-0.40	-0.40
13	PBL Height	0.47	0.55	0.44
14	PVA 250 hPa	0.06	0.01	0.05
15	PVA 700 hPa	0.08	0.14	0.09
16	Richardson Number (Ri)	-0.23	-0.41	-0.23
17	Surface Divergence	0.25	0.03	0.21
18	Surface Froude Number (Fr)	0.07	0.16	0.08
19	Surface Temperature	0.17	0.32	0.14
20	Surface Vertical Heat Flux $\rho\theta'w'$	0.15	0.00	0.11
21	Thickness $_{1000-500 \text{ hPa}}$	0.18	0.28	0.15
22	Thickness $_{1000-700 \text{ hPa}}$	0.17	0.28	0.14
23	Vertical Moisture Flux (VMF) $_{\text{surface to } 700\text{hPa}}$	0.37	0.18	0.33
24	Wind Shear 0-6km	-0.16	-0.18	-0.14

Table 6a. This table shows the correlation coefficients of all the dynamic variables and LWP, IWP, and RR for the winter season.

For the winter season:

Table 6a shows the correlation coefficients between all the selected dynamical variables and LWP, IWP, and RR in the winter season. Tags that have stronger correlation values of 0.50 to 0.65 include: vertical motion (omega ω) at 500 hPa and 700 hPa and planetary boundary layer (PBL) height. Omega at 500 hPa correlates much better

with IWP than with LWP because the larger omega is at 500 hPa the stronger the updraft present to supply more moisture up higher and longer in the cloud allowing for more ice crystals to form in the cloud. PBL height also correlates better with IWP than LWP. Weaker correlations values of around 0.3 to 0.4 can be found between the TMVs and the following dynamical variables: Brunt–Väisälä frequency (N), Richardson number (Ri), vertical moisture flux from the surface to 700 hPa ($VMF_{sfc-700hPa}$), surface equivalent potential temperature (θ_e), and freezing level (FL). Out of these 5 variables, N , Ri , θ_e , and FL have a better correlation with IWP than with LWP or RR. On the other hand, $VMF_{sfc-700hPa}$ is opposite and has a better correlation with LWP or RR than with IWP. The calculation of Ri depends on N^2 so they should have the similar relationship to the TMVs. FL seems to have a better relationship with IWP than LWP or RR because FL represents the level that ice formation to be possible, but it gives not much information on how much LWP there is below that level. A relatively thicker and warmer layer must be present in the lower troposphere to have a higher FL. The dynamical variables are in general correlate just slightly better to LWP than RR, this might due to the fact that LWP calculations include the number of cloud droplets and rain droplets and the dynamical variables have a more direct relationship to cloud formation than precipitation formation, while RR only considers the surface rain. All the rest of the dynamical tags are weakly correlate with the TMVs.

Table 6b contains the best 6 tags that are linearly correlated with TMVs. They are then used to proceed to the next step in the analysis as predictor variables in a multiple linear regression model. They are chosen based not only for their large correlation coefficients, but also the independency of the variables among other chosen variables to

minimize redundant information to be contained in the selected combination of tags. To predict LWP, PBL height, omega at 700hPa, N , VMF, θ_e , and FL are picked. Although omega 500 hPa has almost the same correlation coefficient value as omega 700 hPa, but it is not picked because it is highly possible for it to have redundant information as omega 700 hPa. To predict IWP, omega at 500 hPa, PBL, N , Ri , θ_e , and FL are picked. Omega at 500 hPa correlates significantly better with IWP than omega at 700 hPa, therefore it is chosen instead of omega at 700 hPa. Finally, to predict RR, PBL, omega at 700 hPa, N , VMF, θ_e , and Ri are picked.

Winter	1	2	3	4	5	6
LWP	PBL Height	ω 700 hPa	N	VMF _{surface to 700hPa}	Θ_e	FL
IWP	ω 500 hPa	PBL Height	N	Ri	Θ_e	FL
RR	PBL Height	ω 700 hPa	N	VMF _{surface to 700hPa}	Θ_e	Ri

Table 6b. The best 6 dynamical tags that linearly correlated to LWP, IWP, and RR for the winter season are listed.

For the spring season:

Table 7a shows the correlation coefficients between all the selected dynamical variables and LWP, IWP, and RR in the spring season. Tags that have stronger correlation values of about 0.50 to 0.63 include: omega ω at 500 hPa and 700 hPa, N , and PBL height. Omega at 500 hPa correlates better with IWP than with LWP and RR, but omega at 700 hPa has the same correlation coefficient for all three TMVs. N and PBL height also have similar relationships with all three TMVs, with the correlation coefficients associate with IWP just slightly higher than that with LWP and RR. Weaker correlations values of around 0.3 to 0.4 can be found between the TMVs and the

following dynamical variables: vertical moisture flux from the surface to 700 hPa ($VMF_{sfc-700hPa}$), surface equivalent potential temperature (θ_e), freezing level (FL), and latent heat. Out of these 4 variables, latent heat, θ_e , and FL have a better correlation with IWP than with LWP or RR. They all have a correlation coefficient value of around 0.35 with LWP, and 0.33 with RR, and 0.41 with IWP. Conversely, $VMF_{sfc-700hPa}$ has a better correlation with LWP and RR than with IWP. All other dynamical variables are only weakly correlated with the TMVs.

	Spring	LWP	IWP	RR
1	Brunt–Väisälä frequency (N)	-0.47	-0.51	-0.48
2	Convective Inhibition (CIN)	0.27	0.34	0.28
3	Convective Available Potential Energy	0.25	0.27	0.23
4	Divergence $_{700\text{ hPa}}$	0.24	0.15	0.23
5	Equivalent Potential Temperature Θ_e	0.36	0.43	0.35
6	Freezing Level (FL)	0.34	0.42	0.33
7	Latent Heat	0.34	0.40	0.33
8	Lifted Index (Li)	-0.26	-0.36	-0.28
9	Lifting Condensation Level (LCL)	0.02	0.09	0.02
10	Mid Level Lapse Rate	0.17	0.24	0.20
11	Omega $\omega_{500\text{ hPa}}$	-0.51	-0.63	-0.51
12	Omega $\omega_{700\text{ hPa}}$	-0.49	-0.49	-0.49
13	PBL Height	0.50	0.56	0.50
14	PVA 250 hPa	0.18	0.17	0.17
15	PVA 700 hPa	-0.11	-0.08	-0.10
16	Richardson Number (Ri)	-0.30	-0.32	-0.29
17	Surface Divergence	0.24	0.15	0.23
18	Surface Froude Number (Fr)	0.08	0.02	0.06
19	Surface Temperature	0.26	0.33	0.26
20	Surface Vertical Heat Flux $\rho\theta'w'$	0.31	0.27	0.30
21	Thickness $_{1000-500\text{ hPa}}$	0.22	0.24	0.20
22	Thickness $_{1000-700\text{ hPa}}$	0.20	0.22	0.18
23	Vertical Moisture Flux $_{\text{surface to }700\text{hPa}}$	0.43	0.36	0.43
24	Wind Shear 0-6km	0.08	0.14	0.09

Table 7a. This table shows the correlation coefficients of all the dynamic variables and LWP, IWP, and RR for the spring season.

The best 6 tags that are linearly correlated with TMVs are shown in Table 7b. Same as in the winter season, these tags are selected based on not only their best linear relationship with the TMVs but also the independency of the variables among other chosen variables to minimize redundant information to be contained in the selected combination of tags. To predict LWP, omega at 500 hPa, PBL height, N , VMF, θ_e , and FL are picked. Although omega 700 hPa has almost the same correlation coefficient value as omega 500 hPa, but it is not picked because it is highly possible for it to have redundant information as omega 500 hPa. To predict IWP, omega at 500 hPa, PBL, N , θ_e , FL, and latent heat are chosen. Omega at 500 hPa correlates again significantly better with IWP than omega at 700 hPa, therefore it is chosen instead of omega at 700 hPa. Finally, to predict RR, PBL height, omega at 500 hPa, N , VMF, θ_e , and FL are picked.

Spring	1	2	3	4	5	6
LWP	ω 500 hPa	PBL Height	N	VMF _{surface to 700hPa}	Θ_e	FL
IWP	ω 500 hPa	PBL Height	N	Θ_e	FL	Latent Heat
RR	ω 500 hPa	PBL Height	N	VMF _{surface to 700hPa}	Θ_e	FL

Table 7b. The best 6 dynamical tags that linearly correlated to LWP, IWP, and RR for the spring season are listed.

For the summer season:

Table 8a shows the correlation coefficients between all the selected dynamical variables and LWP, IWP, and RR in the summer season. Tags in general have weaker correlations with TMVs in summer than in winter or spring. Higher correlation coefficient values of around 0.35-0.4 involve the following variables: omega ω at 500-

hPa and 700 hPa, PBL height, N , and VMF. Among these variables, N and omega at 500-hPa has about the same correlation coefficients in all TMVs. Weaker correlations values of around 0.25 can be found between the TMVs and the following dynamical variables: Divergence at 700 hPa and Ri . All the rest of the dynamical tags are more poorly correlated with the TMVs.

	Summer	LWP	IWP	RR
1	Brunt–Väisälä frequency (N)	-0.36	-0.37	-0.37
2	Convective Inhibition (CIN)	-0.04	0.00	-0.02
3	Convective Available Potential Energy	0.07	0.12	0.05
4	Divergence _{700 hPa}	0.28	0.10	0.29
5	Equivalent Potential Temperature Θ_e	0.11	0.18	0.10
6	Freezing Level (FL)	0.10	0.16	0.10
7	Latent Heat	0.00	-0.03	0.01
8	Lifted Index (Li)	-0.04	-0.02	-0.04
9	Lifting Condensation Level (LCL)	-0.06	-0.04	-0.06
10	Mid Level Lapse Rate	-0.04	-0.05	-0.04
11	Omega ω _{500 hPa}	-0.37	-0.35	-0.34
12	Omega ω _{700 hPa}	-0.36	-0.14	-0.35
13	PBL Height	0.36	0.39	0.34
14	PVA 250 hPa	0.00	0.01	0.00
15	PVA 700 hPa	-0.01	0.07	-0.01
16	Richardson Number (Ri)	-0.22	-0.23	-0.24
17	Surface Divergence	0.28	0.10	0.29
18	Surface Froude Number (Fr)	0.09	0.12	0.08
19	Surface Temperature	0.05	0.05	0.05
20	Surface Vertical Heat Flux $\rho\theta'w'$	0.16	0.13	0.18
21	Thickness _{1000-500 hPa}	0.09	0.08	0.08
22	Thickness _{1000-700 hPa}	0.09	0.05	0.07
23	Vertical Moisture Flux _{surface to 700hPa}	0.33	0.15	0.34
24	Wind Shear 0-6km	-0.03	0.06	-0.06

Table 8a. This table shows the correlation coefficients of all the dynamic variables and LWP, IWP, and RR for the summer season.

Table 8b contains the best 6 tags that are linearly correlated with TMVs. To predict LWP, omega at 700 hPa, N , PBL height, VMF, surface divergence and Ri are

selected. Although omega 700 hPa has almost the same correlation coefficient value as omega 500 hPa, but it is not picked because it is highly possible for it to have redundant information as omega 500 hPa. To predict IWP, PBL height, N , omega at 500 hPa, Ri , θ_e , and FL are picked. Omega at 500 hPa correlates significantly better with IWP than omega at 700 hPa, therefore it is chosen instead of omega at 700 hPa. Finally, to predict RR, N , omega at 700 hPa, PBL height, VMF, surface divergence, and Ri are chosen.

Summer	1	2	3	4	5	6
LWP	$\omega_{500 \text{ hPa}}$	N	PBL Height	VMF _{surface to 700hPa}	DIV _{surface}	Ri
IWP	PBL Height	N	$\omega_{500 \text{ hPa}}$	Ri	Θ_e	FL
RR	N	$\omega_{700 \text{ hPa}}$	PBL Height	VMF _{surface to 700hPa}	DIV _{surface}	Ri

Table 8b. The best 6 dynamical tags that linearly correlated to LWP, IWP, and RR for the summer season are listed.

For the autumn season:

Table 9a shows the correlation coefficients between all the selected dynamical variables and LWP, IWP, and RR in the autumn season. The higher correlation coefficient values are around 0.35-0.52 and are associated with the following variables: N , Ri , omega at 500 hPa and 700 hPa, VMF, and PBL height. N , Ri , and PBL height all have about the same correlations with all the TMVs. Same as results from other seasons, omega 500 hPa shows a better correlation with IWP than LWP and RR, while omega 700-hPa shows the opposite. VMF has much strong correlation with LWP and RR than IWP because it measures directly the amount of moisture flux in the lower levels. Other dynamical tags are weakly correlated with the TMVs.

	Autumn	LWP	IWP	RR
1	Brunt–Väisälä frequency (N)	-0.42	-0.41	-0.42
2	Convective Inhibition (CIN)	0.08	0.00	0.11
3	Convective Available Potential Energy	0.29	0.19	0.28
4	Divergence $_{700 \text{ hPa}}$	0.23	0.01	0.27
5	Equivalent Potential Temperature Θ_e	0.27	0.18	0.26
6	Freezing Level (FL)	0.24	0.15	0.23
7	Latent Heat	0.17	0.07	0.16
8	Lifted Index (Li)	-0.10	-0.14	-0.10
9	Lifting Condensation Level (LCL)	0.02	0.11	0.01
10	Mid Level Lapse Rate	0.09	0.18	0.08
11	Omega $\omega_{500 \text{ hPa}}$	-0.39	-0.52	-0.37
12	Omega $\omega_{700 \text{ hPa}}$	-0.38	-0.23	-0.38
13	PBL Height	0.42	0.40	0.41
14	PVA 250 hPa	0.11	0.00	0.10
15	PVA 700 hPa	-0.01	-0.02	-0.01
16	Richardson Number (Ri)	-0.33	-0.37	-0.33
17	Surface Divergence	0.23	0.01	0.27
18	Surface Froude Number (Fr)	0.09	0.12	0.09
19	Surface Temperature	0.09	0.12	0.09
20	Surface Vertical Heat Flux $\rho\theta'w'$	0.18	0.04	0.20
21	Thickness $_{1000-500 \text{ hPa}}$	0.17	0.12	0.15
22	Thickness $_{1000-700 \text{ hPa}}$	0.17	0.13	0.16
23	Vertical Moisture Flux $_{\text{surface to } 700\text{hPa}}$	0.34	0.09	0.36
24	Wind Shear 0-6km	-0.14	-0.06	-0.13

Table 9a. This table shows the correlation coefficients of all the dynamic variables and LWP, IWP, and RR for the autumn season.

Table 9b contains the best 6 tags that are linearly correlated with TMVs. To predict LWP, omega at 500 hPa, N , PBL height, VMF, Ri , and CAPE are selected. Omega at 700 hPa is not selected to prevent repeated information. Omega at 500 hPa, N , PBL height, Ri , CAPE, and θ_e , are chosen to predict IWP. Finally, to predict RR, N , PBL height, omega at 700 hPa, VMF, Ri , and CAPE are chosen.

Autumn	1	2	3	4	5	6
LWP	PBL Height	N	$\omega_{500 \text{ hPa}}$	VMF _{surface to 700hPa}	Ri	CAPE
IWP	$\omega_{500 \text{ hPa}}$	N	PBL Height	Ri	CAPE	Θ_e
RR	N	PBL Height	$\omega_{700 \text{ hPa}}$	VMF _{surface to 700hPa}	Ri	CAPE

Table 9b. The best 6 dynamical tags that linearly correlated to LWP, IWP, and RR for the autumn season are listed.

The following dynamical variables have better linear correlations with TMVs in all 4 seasons: Omega at 500 hPa, Omega at 700 hPa, PBL height, Ri , VMF, and N . Omega 500 hPa has a stronger linear correlation with IWP than LWP and RR, and VMF is the opposite and have a stronger linear correlation with LWP and RR than IWP in all seasons. Summer is the only season that omega at 500 hPa does not have a significant better correlation with IWP than LWP and RR. This suggests that omega 500 hPa is a more important dynamical factor in affecting IWP in other three seasons. FL has a stronger linear relationship with the TMVs in winter and spring than in summer and autumn. Other dynamical tags are weakly linearly correlate with the TMVs.

These results verify that dynamical tags are precipitation-regimes or situation dependent. In other words, since clouds and different kinds of precipitation formation does not always rely on just one single environmental factor, as it generally has to have sources of moisture at the lower levels, atmospheric instability, and some lifting mechanisms to initiate and enhance cloud development and precipitation formation, a single dynamical tag cannot be always helpful in promoting cloud formation and precipitation in all situations thus it would be more suitable to use a combination of dynamical tags to specifically explain a particular situation.

The best 6 tags that correlated with the TMVs linearly are being selected and used as predictor variables in a multiple linear regression model. Since winter season contains the highest number of realizations, thus it has higher probability to be closer to represent a more full database than other seasons and thus it is being chosen for a full analysis to be performed on and only the results from winter will be provided in this study. Data analysis results in predicting LWP will be presented first, followed by analysis results in predicting IWP; lastly the results in predicting RR will be given.

Since LWP, IWP, and RR have strong positive skew towards light precipitation cases, \log_{10} transformation is applied on the TMVs to make the distributions more symmetric.

The six tags chosen to predict LWP are:

1. PBL height
2. $\omega_{700 \text{ hPa}}$ - Omega at 700 hPa
3. N - Brunt–Väisälä frequency
4. $\text{VMF}_{\text{sfc-700hPa}}$ – Vertical Moisture Flux from surface to 700 hPa
5. Θ_e – Equivalent Potential Temperature
6. FL – Freezing Level

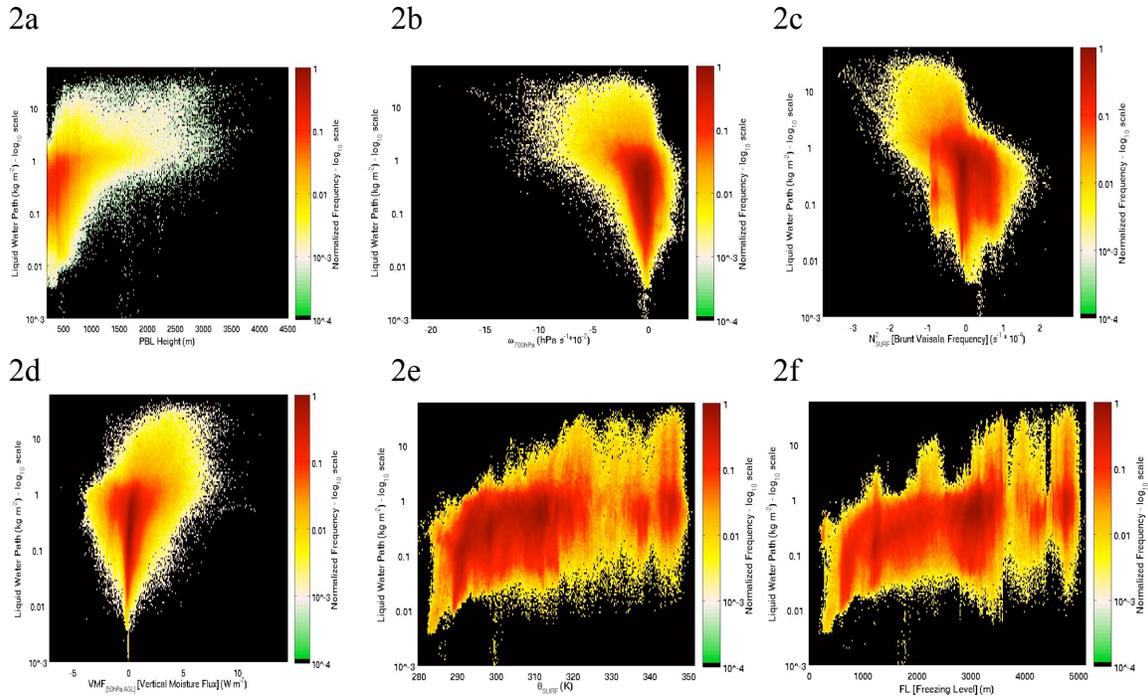
As part of the verification process in the selection of the best tags to continue the investigation, scatter plots of the tags and TMV in the database are produced to allow good visualization of the relationship between the variables. Scatter plots in Figures 2a to 2f indicate how much \log_{10} LWP is affected by a dynamical variable. The color bar expressed in common logarithmic scale represents a normalized frequency of occurrences, which is calculated by the absolute frequency divided by the maximum

value of the absolute frequency. A bin's value denotes how close one is to the bin that contains the highest number of points that happened to be in it. The color bar goes from dark red color, which represents a crowded bin, to yellow, white, and green colors, which stand for an almost empty bin.

In Fig. 2a, the bins with more data points are indicated by the red color show a more linear relationship while the bins with fewer observations show two splits in the tail structure with increasing \log_{10} LWP and PBL height. The relationship is positively correlated. Fig. 2b illustrates that $\omega_{700 \text{ hPa}}$ and \log_{10} LWP appears to have a curved relationship. Thus linear regression model will not create a good fit to predict the relationship. Another example of curvilinear regression, polynomial regression, which tries to find a curve to better fit the data points, is probably better than a linear regression. A polynomial equation has x raised to integer powers. In the case of a parabola, it can be expressed in quadratic equation that has the form of

$$y = c + b_1x + b_2x^2 \quad (5)$$

where y is the dependent variable (\log_{10} LWP), x is the independent predictor variable ($\omega_{700 \text{ hPa}}$) in this case, c is the y -intercept, b_1 and b_2 are the coefficient constants. N in Fig. 2c shows a more complex relationship. It appears that there are majorly 2 linear relationships being stacked upon one another. $\text{VMF}_{\text{sfc-700hPa}}$, surface Θ_e , and FL are shown to have a more linear relationship with \log_{10} LWP in Figures 2d, 2e, and 2f, respectively. Hence $\omega_{700 \text{ hPa}}$ is the only variable viable for a quadratic regression.



Figures 2a to 2f show scatter plots of the best six dynamical tags (a. PBL height, b. vertical motion in p-coordinates at 700 hPa ($\omega_{700 \text{ hPa}}$), c. Brunt–Väisälä frequency (N), d. Vertical Moisture Flux from surface to 700 hPa ($\text{VMF}_{\text{sfc-700hPa}}$), e. surface equivalent potential temperature (Θ_e), and f. Freezing Level (FL)) chosen as they have the highest correlation with logarithmic LWP for the winter dataset.

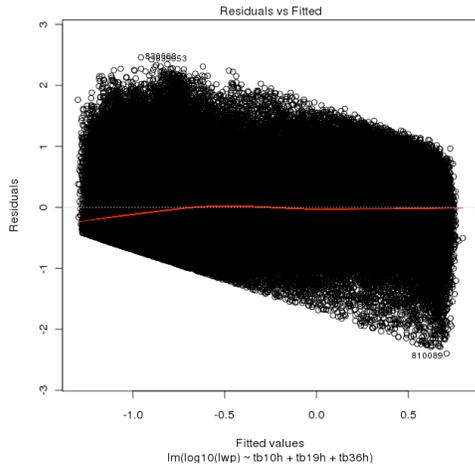
There are 2 goals in this analysis. First is to determine which independent explanatory variables (the dynamical tags) are important predictors of the dependent variable (\log_{10} LWP), and the amount of variances of the predicted \log_{10} LWP can be explained by the tags in addition to the use of the BTs. Second is to determine whether a combination of tags exists that could be more universally applicable to most precipitating situations and to find out the usefulness of individual tags.

At the beginning, horizontally polarized BTs are used as individual predictor variables in the multiple linear regression models. Key assumptions for a multiple linear regression model include: 1) All the x variables and y has a linear relationship, 2)

Residuals are independent. 3) Residuals are normally distributed with zero mean and a constant variance.

A quantile-quantile (q-q) plot, also named a normal probability plot, which plots the quantiles of one dataset against the quantiles of another dataset. It is a good way to check if the residuals of a model can be fitted to a normal distribution. The residuals are plotted against the fitted \log_{10} LWP in Fig. 3a to check if the variability of the residuals is constant throughout the range of fitted values of y . The residuals are seen to be mostly constant. The q-q plot shown in Fig. 3b shows the residuals from a linear regression model of the 10-, 19-, and 36-GHz BTs as predictor parameters with \log_{10} LWP to be the response variable plotted against the normal distribution. It shows that from -2 to 2 quantiles of the normal distribution, the residuals are also fitted quite well with the normal distribution and this means most of the data is fitted in a normal distribution. But the overall S shaped line indicates that the residuals distribution is more skewed and has longer tails than the normal distribution. Therefore it can be concluded that using a linear regression model may not be optimal with the predictor variables in use.

3a



3b

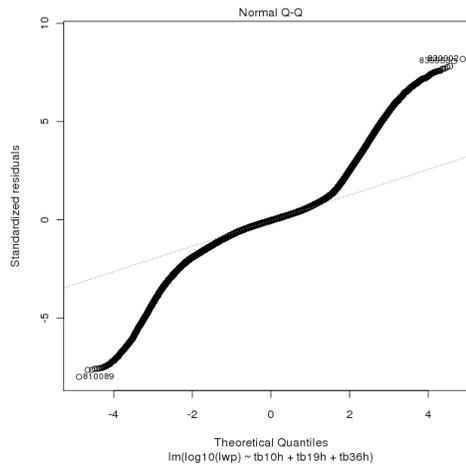


Figure 3a. Residuals versus fitted values plot. The residuals are computed from fitting \log_{10} LWP to a multiple linear regression model that uses the brightness temperatures at 10, 19, and 36-GHz as predictor variables. Residuals plotted against the fitted values show the variance is almost constant. The red line represents the trend of the residuals. However, the q-q plot in Figure 3b strongly suggests that the relationships between the model parameters are not linear.

Consequently there is a need to transform the dataset in order to fit a multiple linear regression model. The nonlinearity of BTs and LWP can be lowered by the usage of a normalized polarization difference (P ; Petty, 1994a), which is defined as

$$P \equiv \frac{T_V - T_H}{T_{V,O} - T_{H,O}} \quad (6)$$

where T_V and T_H are the observed vertically and horizontally polarized BTs at one frequency, and $T_{V,O}$ and $T_{H,O}$ are the hypothetical BTs for the same scene under clear sky condition. When viewed at an oblique angle over the ocean, the observed difference between T_V and T_H is dominated by the polarized emissivity of the ocean surface. P is 0 represents an opaque situation while P is 1 represents a cloud-free condition. Through normalizing the observed polarization difference with the cloud free polarization difference in P , the BTs' sensitivity to water vapor in a column can be

isolated out from its sensitivity to the ocean surface emissivity so that P can be directly related to the amount of column transmittance (Petty and Katsaros, 1990, 1992). Petty (1994a) also points out that P has a monotonic relationship with increasing optical depth due to the increase of rain and cloud droplets and is weakly sensitive to the scattering effects of the cloud. These advantages make P to be a better predictor variable than BTs themselves as P corresponds to the columnar LWP more strongly. P is then computed for 10-, 19-, 36-, and 85-GHz channels.

R^2 , the percentage of variance explained, is a good indicator to assess the goodness-of-fit of a model. By comparing various models as recorded in Table 10, it is found out that by using P10, P19, and P36 as predictive parameters results in the largest R^2 . Therefore P10, P19, and P36 are selected to be the base predictive parameters and \log_{10} LWP is the response variable in the multiple linear regression model.

	P10	P19	P36	R^2
\log_{10} LWP	✓			0.004
\log_{10} LWP		✓		0.023
\log_{10} LWP			✓	0.026
\log_{10} LWP	✓	✓		0.050
\log_{10} LWP	✓	✓	✓	0.051
\log_{10} LWP		✓	✓	0.026

Table 10. Multiple linear regression model comparisons with varying predictor variables and the percentage of variance explained R^2 .

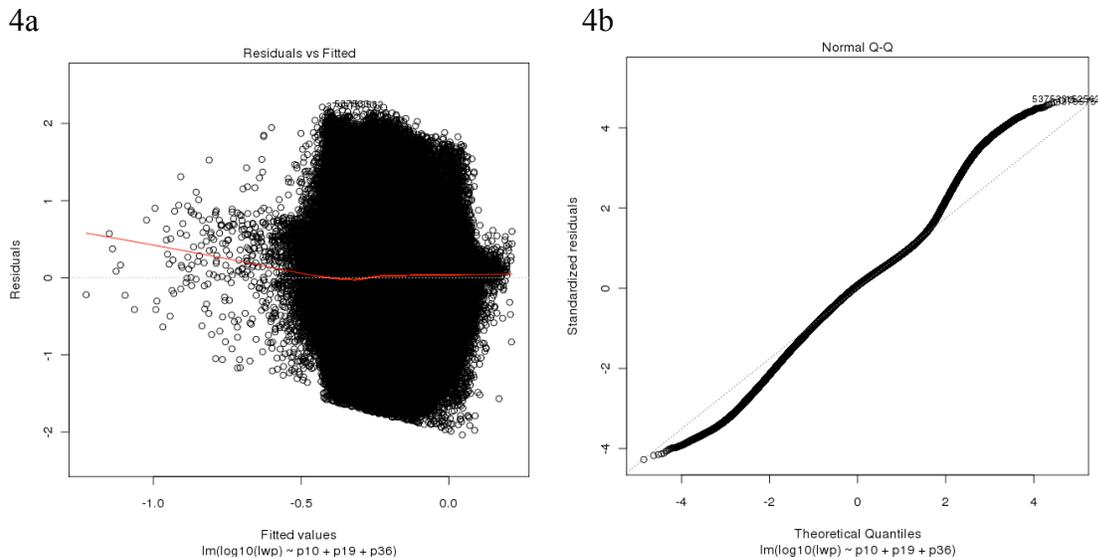


Figure 4a. Residuals versus fitted values plot. The residuals are computed from fitting \log_{10} LWP to a multiple linear regression model that uses the normalized polarization indices of 10, 19, and 36-GHz channels as predictor variables. The red line represents the trend of the residuals. Residuals plotted against the fitted values show the variance of the residuals is almost constant. However, the q-q plot in Figure 4b strongly suggests that the relationships between the model parameters are more linear after the use of normalized polarization indices.

The residuals are plotted against the fitted values in Fig. 4a and it shows that the variability of the residuals is almost constant for most part of the data. The q-q plot (shown in Fig. 4b) is again used to assess the normality of the residuals as it compares the residuals to an ideal normal distribution. In compare to Fig. 3b, Fig. 4b shows the residuals fitted significantly better onto the reference line. Since the points are a lot closer to the reference line than in Fig. 3b, the results suggest that it is more optimal to fit P linearly to \log_{10} LWP than to fit the raw BTs.

Before building a model, the data are partitioned into two mutually exclusive datasets: a training dataset and a testing dataset. The training dataset is used to fit the multiple linear regression models so to compute the regression coefficients. In

order to find out the accuracy of the model on unseen data, the testing dataset is used to compute the error in prediction, which is the discrepancy between the actual value and the predicted value of the response variable. Mean squared error (MSE) of average percentage error measures the overall accuracy of the model. The training dataset was not to be used to compute the accuracy of the model fit because that would result an excessively optimistic estimate of the accuracy, as the training process makes sure the accuracy of the model for the training dataset is as high as possible. Hence by employing a separate dataset that is unseen to the model to calculate the accuracy of the model can give a more realistic estimate.

Next, Bayesian Information Criteria (BIC; Schwarz, 1978) attempts to determine a model that best explains the data with a minimum combination of tag variables. It is a criterion that can be used as a tool for regression variable selection to form a best-fitted model, a model that has the most optimal combination of predictor parameters that result in maximal precision. However, overfitting may result because it is possible to increase the likelihood by adding parameters when the selection of model parameters is done through maximum likelihood estimation. Maximum likelihood refers to the probability of the observed results to be as large as possible after a model that has gone through parameters estimation in order to pick a few better parameters to produce the model. This probability always has values in between 0 and 1, and it is common to evaluate likelihoods on a logarithmic scale multiplied by -2. BIC not only awards the goodness of fit, but also includes a penalty that is an increasing function of the number of estimated parameters to discourage overfitting. Log likelihood for BIC can be expressed as:

$$-2 \times (n + n \log 2\pi + n \log(RRS/n) + (\log(n))(p + 1)) \quad (7)$$

where n is the number of observations, p is the number of parameters used in the model, and RSS is the residual sum of squares. RSS/n is the maximum likelihood estimates and the last term is a penalty term. The preferred model is the one with the lowest BIC value. BIC is calculated using the training data set through statistics package R with a forward selection procedure that starts with the model with only P10, P19, and P36 as its based model and add dynamical tags one at a time until no further addition significantly improves the fit. In each step, it considers all models obtained by adding one more dynamical tag that has not been included to the current model, and then computes its extra sum-of-squares, and add the variable with the largest extra sum-of-squares. Then the process starts over again until all the dynamical tags have been considered. From the BIC output results, all 7 dynamical parameters chosen to be included in the fitted model are in the following order: 1. Freezing level, 2. Vertical motion at 700 hPa, 3. Vertical motion at 700 hPa squared, 4. Brunt–Väisälä frequency, 5. Surface equivalent potential temperature, 6. Planetary boundary layer height, and 7. Vertical moisture flux from surface to 700 hPa.

Model comparisons between the based model and models with one additional dynamical tag added one at a time following the order suggested by BIC are performed using the testing dataset. All the models are given in Table 11a and their statistics are calculated and listed in Table 11b. R^2 represents the percentage of variance of the predicted \log_{10} LWP explained and it indicates that if only P10, P19, and P36 are to be used as explanatory variables, the fitted model will be able to explain 5% of the variance. By just adding one dynamical tag, the freezing level, R^2 increases to 26%. By adding just

one more dynamical tag, omega at 700 hPa, R^2 increases to 37%. From this point on, adding more dynamical variables will still increase R^2 , but the increase is much less significant than when adding the first two.

Fit	P10+P19+P36	FL	$\omega_{700\text{hPa}}$	$\omega_{700\text{hPa}}^2$	N	Θ_e	PBL	VMF
1	✓							
2	✓	✓						
3	✓	✓	✓					
4	✓	✓	✓	✓				
5	✓	✓	✓	✓	✓			
6	✓	✓	✓	✓	✓	✓		
7	✓	✓	✓	✓	✓	✓	✓	
8	✓	✓	✓	✓	✓	✓	✓	✓

Table 11a. The predicted variables chosen for the fitted models 1 to 8.

Fit	R^2	Increased R^2 (%)
1	0.0513	5.13
2	0.2643	21.3
3	0.3732	10.89
4	0.3783	0.51
5	0.3837	0.54
6	0.3897	0.6
7	0.3912	0.15
8	0.3914	0.02

Table 11b. The statistics for fitted models 1 to 8.

This section starts the statistical analysis part for predicting IWP. By comparing various models as recorded in Table 12, it is found out that by using P36 and P85 to be predictive parameters result in the largest R^2 . Therefore P36 and P85 are selected to be

the base predictive parameters and \log_{10} IWP is the response variable in the multiple linear regression model.

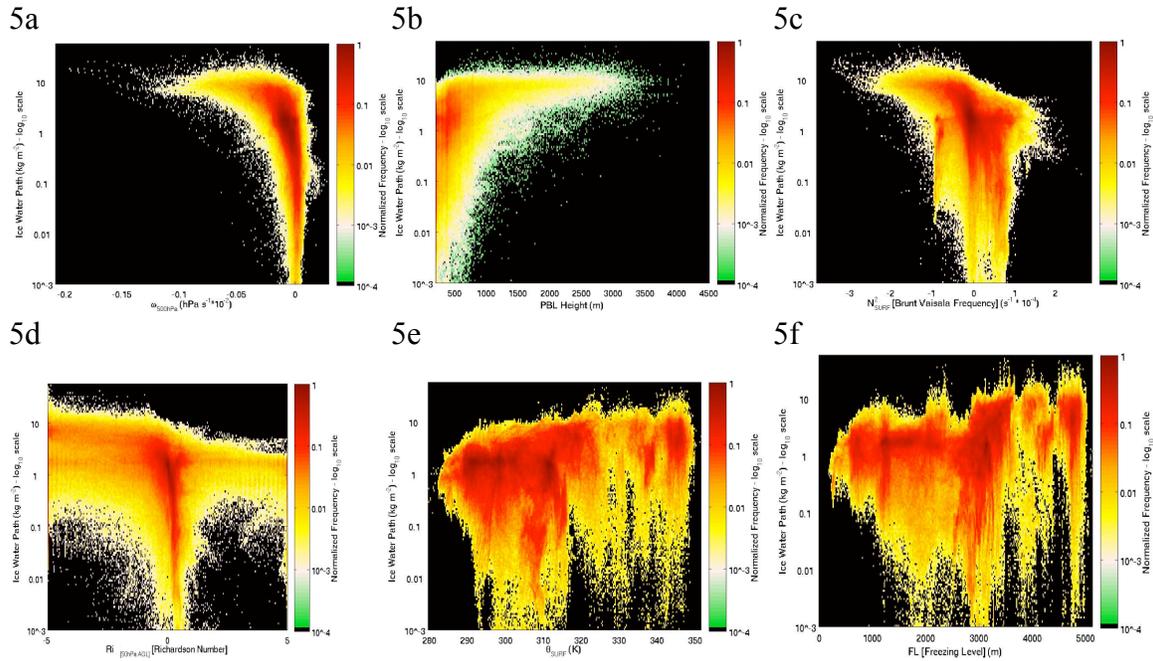
	P36	P85	R ²
Log ₁₀ IWP	✓		0.0072
Log ₁₀ IWP		✓	0.0061
Log ₁₀ IWP	✓	✓	0.0073

Table 12. Multiple linear regression model comparisons with varying predictor variables and the percentage of variance explained R².

The six tags chosen to predict IWP are:

1. $\omega_{500 \text{ hPa}}$ - Omega at 500 hPa
2. PBL Height
3. N - Brunt-Väisälä frequency
4. Ri - Richardson Number
5. Θ_e - Equivalent potential temperature
6. FL - Freezing level

Figures 5a to f display the scatter plots of the \log_{10} IWP plotted against the dynamical tags. Omega at 500 hPa and \log_{10} IWP seem to have a curved relationship as plotted in Fig. 5a. By looking at the orange to red portion (more densely populated bins), it indicates that PBL height has a more linear relationship to \log_{10} IWP in Fig. 5b. The same applies to N , Ri , Θ_e , and FL in Figures 5c, d, e, f, respectively.



Figures 5a to 5f show scatter plots of the best six dynamical tags (a. vertical motion in p-coordinates at 500 hPa ($\omega_{500 \text{ hPa}}$), b. Planetary Boundary Layer (PBL) height, c. Brunt–Väisälä frequency (N), d. Richardson number (Ri), e. surface equivalent potential temperature (Θ_e), and f. Freezing Level (FL)) chosen as they have the highest correlation with logarithmic IWP for the winter dataset.

By calculating the BIC using the training data set, it is concluded that all dynamical variables are selected to fit into a multiple linear regression model. All the models are shown in Table 13a and their statistics are calculated and listed in Table 13b. R^2 represents the percentage of variance of the predicted \log_{10} IWP explained and it shows that if only P36 and P85 are to be used as explanatory variables, the fitted model will be able to explain 0.76% of the variance. By just adding one dynamical tag, omega at 500 hPa, R^2 increases to 25%. By adding one more dynamical tag, omega at 500 hPa squared, R^2 increases to 32%. From this point on, adding more dynamical variables will still increase R^2 , but the increase is much less significant than when adding the first two. The result here is the same as how predicting LWP is shown in previous session.

Fit	P36+P85	$\omega_{500\text{hPa}}$	$\omega_{500\text{hPa}}^2$	N	PBL	FL	Ri	Θ_e
1	✓							
2	✓	✓						
3	✓	✓	✓					
4	✓	✓	✓	✓				
5	✓	✓	✓	✓	✓			
6	✓	✓	✓	✓	✓	✓		
7	✓	✓	✓	✓	✓	✓	✓	
8	✓	✓	✓	✓	✓	✓	✓	✓

Table 13a. The predicted variables chosen for the fitted models 1 to 8.

Fit	R^2	Increased R^2 (%)
1	0.007636	0.764
2	0.2562624	24.863
3	0.3254547	6.919
4	0.3296333	0.418
5	0.3316978	0.206
6	0.3332696	0.157
7	0.3338743	0.06
8	0.3338985	0.002

Table 13b. The statistics for fitted models 1 to 8.

This section starts the statistical analysis part for predicting RR. By comparing various models as recorded in Table 14, it is found out that by using P10, P19, and P36 to be predictive parameters result in largest R^2 . Therefore P10, P19, and P36 are chosen to be the base predictive parameters and $\log_{10}RR$ is the response variable in the multiple linear regression model.

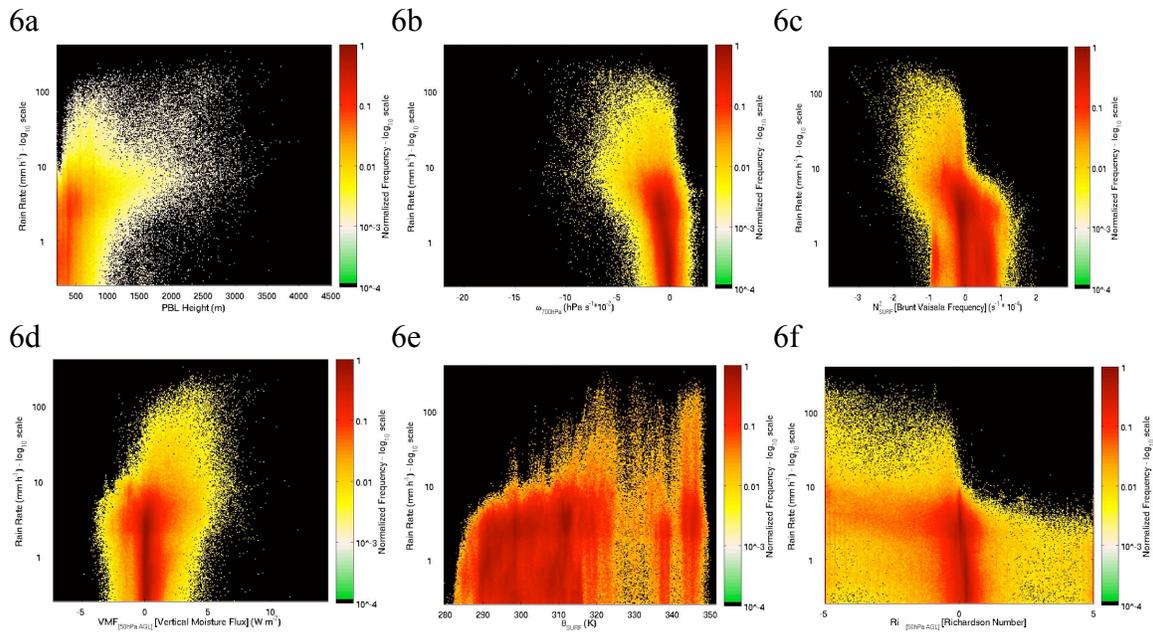
	P10	P19	P36	P85	R ²
Log ₁₀ RR	✓				0.00172
Log ₁₀ RR		✓			0.01661
Log ₁₀ RR			✓		0.01729
Log ₁₀ RR				✓	0.00663
Log ₁₀ RR	✓	✓			0.04742
Log ₁₀ RR	✓		✓		0.02763
Log ₁₀ RR	✓			✓	0.00671
Log ₁₀ RR		✓	✓		0.01745
Log ₁₀ RR		✓		✓	0.01709
Log ₁₀ RR			✓	✓	0.02181
Log ₁₀ RR	✓	✓	✓		0.05282
Log ₁₀ RR		✓	✓	✓	0.02261
Log ₁₀ RR	✓		✓	✓	0.03816
Log ₁₀ RR	✓	✓		✓	0.05255
Log ₁₀ RR	✓	✓	✓	✓	0.05333

Table 14. Multiple linear regression model comparisons with varying predictor variables and the percentage of variance explained R².

The six tags chosen to predict RR are:

1. PBL Height
2. $\omega_{700 \text{ hPa}}$ - Omega at 700 hPa
3. N - Brunt–Väisälä frequency
4. $\text{VMF}_{\text{sfc-700hPa}}$ - Vertical Moisture Flux from surface to 700 hPa
5. Θ_e - Equivalent potential temperature
6. Ri - Richardson Number

Figures 6a to f show the scatter plots of the log₁₀RR plotted against the dynamical tags. PBL height, omega at 700 hPa, N , vertical moisture flux, equivalent potential temperature, and Richardson number all has linear relationship with precipitation rate.



Figures 6a to 6f show scatter plots of the best six dynamical tags (a. Planetary Boundary Layer (PBL) height, b. vertical motion in p-coordinates at 700 hPa, c. Brunt–Väisälä frequency (N), d. vertical moisture flux from surface to 700 hPa ($VMF_{sfc-700hPa}$), e. surface equivalent potential temperature (Θ_e), and f. Richardson number (Ri)) chosen as they have the highest correlation with rain rate for the winter dataset.

Fit	P10+P19+P36	ω_{700hPa}	Θ_e	Ri	PBL	VMF	P85
1	✓						
2	✓	✓					
3	✓	✓	✓				
4	✓	✓	✓	✓			
5	✓	✓	✓	✓	✓		
6	✓	✓	✓	✓	✓	✓	
7	✓	✓	✓	✓	✓	✓	✓

Table 15a. The predicted variables chosen for the fitted models 1 to 8.

Fit	R ²	Increased R ² (%)
1	0.0525432	5.254317
2	0.2179523	16.540913
3	0.2540491	3.60968
4	0.2675856	1.35365
5	0.271627	0.40414
6	0.274456	0.2829
7	0.2763142	0.18582

Table 15b. The statistics for fitted models 1 to 8.

By calculating the BIC, it is concluded that all dynamical variables are selected to fit into a multiple linear regression model. All the models are shown in Table 15a and their statistics are calculated and listed in Table 15b. R² represents the percentage of variance of the predicted log₁₀ RR explained and it shows that if only P10, P19, and P36 are to be used as explanatory variables, the fitted model will be able to explain 5.25% of the variance. Then R² increases to 22% by just adding one dynamical variable, omega 700hPa. To add one more, equivalent potential temperature, R² increases to 25%. From this point on, adding more dynamical variables will still increase R², but the raise is much less significant than when adding the first two, which is the same results obtained from analyzing the statistics to predict LWP and IWP.

5. Conclusions and future work

5.1 Conclusions

Precipitation retrieval algorithms have been evolving and improve precipitation estimates since the 1980s. The Bayesian based algorithms depend heavily on the a-priori Cloud Radiation Database (CRD) to match observed BTs to the simulated BTs with associated microphysical profiles. However, microphysical profiles from different types of precipitation systems, such as isolated convection, extra-tropical cyclones, and tropical convections, could potentially be mixed together into the retrieval outcome because a set of multispectral BTs can match with many microphysical profiles. This can contribute to in accurate estimations of microphysical quantities and thus leads to imprecise estimation of precipitation amounts.

The Cloud Dynamics and Radiation Database (CDRD) concept makes use of the idea that short-term projection of the atmospheric environments that describes the synoptic situation being retrieved can be used to categorize and to help to select microphysical profiles that are more applicable by introducing a dependence on the weather situation. This additional information is readily available through various operational short-term (<6 hour) global forecasting model forecasts. In the CDRD, atmospheric dynamical and thermodynamical information, averaged to a global model grid scale, that is deemed to likely have a high degree of short-term predictability in these global forecasts, is derived from convection resolving model experiments. This data is linked with the cloud resolving model simulated microphysical profiles, and derived multispectral BTs that are consistent with the cloud resolving model simulation.

In this study, a North America CDRD consisted of 120 simulations of various precipitation events over North America over a time period of a year was being constructed. Through statistical analysis, this study demonstrated that by adding just two dynamical variables could increase explanation of the variation of the predicted columnar liquid water path, ice water path, and surface rain rate by a significant amount. Among all the dynamical variables chosen, freezing level and omega at 700 hPa and 500 hPa appear to be the variables that contribute most additional information relative to BTs toward explaining variance of surface precipitation rate and liquid water path. Therefore, the results suggest that the dynamical variables can bring additional information that is helpful to improve precipitation estimates.

Quantitative results for the winter season include:

1. By adding just freezing level as one of the explanatory variables can increase R^2 the explained variance of predicted LWP by $\sim 21\%$. By adding omega at 700-hPa can increase R^2 by another $\sim 11\%$.
2. By adding just omega at 500 hPa and its squared as explanatory variables can increase R^2 the explained variance of predicted IWP by $\sim 32\%$.
3. By adding just omega at 700 hPa as explanatory variables can increase R^2 the explained variance of predicted RR by $\sim 17\%$.

Although the calculations of BIC suggest including all dynamical variables to form the best fitted multiple linear regression models for the prediction of TMVs, these results might happen just because the amount of training dataset (around 0.4 million realizations) is so big that every time one extra dynamical variable is included in the model, it would still be able to improve the fit and reduce the residual sum of squares without increasing

as much on the penalty term in the calculation of BIC. However, through this analysis of comparing multiple linear regression models that use different combination of dynamical variables, the results demonstrate that by using only 1-2 dynamic variables in additional to the based model predictive parameters, the normalized polarization indices of various channels, can already significantly improve LWP, IWP, and surface rain rates estimates. The TMVs estimates would continue to be improved by including more and more dynamic variables in the multiple linear regression models, but the improvements are not as significant beyond adding just 2 dynamic variables. Therefore, two might be the optimal number of dynamic variables to be included in the models to be useful in categorizing hydrometeor profiles during the retrieval process of Bayesian physical inversion-based algorithms.

There are uncertainties in the assumption of linear relationships in between some dynamic variables and the TMVs in this analysis. In some of the scatter plots (e.g. Figures 2a, 2d, 5a, 5c, and 6a), although the more densely populated bins shown in dark red color illustrate a more linear relationship between the dynamic variable plotted with the TMV, not all the bins on the scatter plot are consistent with the linear relationship. Some less populated bins in yellow, orange, and bright red colors suggest a nonlinear relationship between the dynamic variables and the TMV plotted. Furthermore, other scatter plots as shown in Figures 2c, 5d, and 5f appear to have multiple linear relationship structures embedded in one scatter plot. Therefore, these scatter plots suggest that linear regression might not be the most appropriate technique for the analysis. To solve this problem, other transformations on the dynamic variables or the TMV might be needed to achieve a more linear relationship between the two for linear regression to be more valid.

The results from this study might change with different transformations to be applied on the variables.

5.2 Future Work

Since the database does not include all the possible precipitation systems in all seasons, improvements in making the database to be more completed can be done by adding in more simulations to widen the breadth of storm types in all seasons. There is a need to extend the data analysis to all other seasons and compare their results with those from the winter season. Moreover, there is a need to study the impact of model error on these results. It is because in simulating the BTs, cloud model microphysics including the hydrometeor sizes, shapes, composition, and distribution has to be assumed in the model. Secondly, surface skin temperature and the temperature and moisture profiles with correct representation of the environment are needed as input to the radiative transfer model. Then, there are also other calculations of the emissivity properties of the surface, and hydrometeor optical properties, and in the radiative transfer. Errors in any of these calculations can cause a bias in the simulated BTs. To reduce model errors, it is essential to develop more competent simulations of microphysical profiles and BTs. In addition, future work is needed to develop ways to implement the use of the dynamical variables into the retrieval process in the future algorithms. Since this method of the inclusion of dynamic variables into the CDRD also depends on the accuracy of the global forecasting model (GFS or ECMWF)'s forecasts, it is important to develop a checking system in the retrieval process to make sure that the forecasts from the forecasting models are accurate to be used.

Appendix A

Simulation Number	Month	Day	Year	Latitude	Longitude	Event
1	11	4	2007	21.70	-157.50	frontal
2	11	9	2007	19.64	-66.36	convective
3	11	9	2007	46.74	-86.97	lake effect snow
4	11	12	2007	46.98	-116.37	orographic
5	11	13	2007	20.63	-85.78	convective
6	11	14	2007	51.45	-130.61	frontal
7	11	20	2007	43.58	-61.17	frontal
8	11	25	2007	39.23	-76.55	frontal
9	11	27	2007	47.40	-116.72	frontal
10	11	30	2007	39.44	-107.31	frontal
11	12	1	2007	41.57	-97.82	frontal
12	12	3	2007	48.40	-123.14	frontal
13	12	7	2007	44.34	-59.15	frontal
14	12	11	2007	34.74	-112.59	frontal
15	12	14	2007	27.99	-104.24	frontal
16	12	17	2007	58.31	-124.01	frontal
17	12	20	2007	38.00	-154.25	frontal
18	12	20	2007	28.77	-86.84	frontal
19	12	24	2007	13.92	-115.66	convective
20	12	26	2007	57.80	-160.14	convective
21	1	4	2008	36.81	-119.66	frontal
22	1	8	2008	13.07	-158.20	convective
23	1	9	2008	46.80	-127.79	frontal
24	1	10	2008	55.28	-52.38	frontal
25	1	16	2008	45.71	-25.14	frontal
26	1	18	2008	30.03	-87.19	frontal
27	1	21	2008	43.41	-77.21	lake effect snow
28	1	25	2008	38.82	-127.97	convective
29	1	28	2008	32.99	-57.66	frontal
30	1	29	2008	50.40	-48.52	frontal
31	2	3	2008	41.90	-112.15	frontal
32	2	6	2008	12.55	-154.34	convective
33	2	12	2008	43.71	-160.66	frontal
34	2	13	2008	30.60	-72.77	frontal
35	2	15	2008	33.58	-28.65	frontal
36	2	16	2008	4.57	-82.44	convective
37	2	21	2008	5.09	-47.29	convective
38	2	25	2008	6.32	-78.75	convective
39	2	26	2008	42.42	-157.32	frontal
40	2	29	2008	53.12	-172.62	frontal
41	3	3	2008	40.85	-50.45	frontal
42	3	7	2008	29.61	-86.40	frontal
43	3	9	2008	58.45	-150.29	frontal

44	3	10	2008	48.11	-124.28	frontal
45	3	15	2008	35.88	-73.04	frontal
46	3	18	2008	31.13	-99.45	frontal
47	3	19	2008	9.97	-103.18	convective
48	3	23	2008	43.45	-105.82	orographic
49	3	25	2008	24.77	-75.85	frontal
50	3	28	2008	53.65	-109.69	frontal
51	4	1	2008	35.10	-92.46	convective
52	4	5	2008	26.23	-82.97	convective
53	4	6	2008	46.92	-92.11	frontal
54	4	9	2008	13.84	-89.39	convective
55	4	11	2008	41.71	-78.57	frontal
56	4	14	2008	47.52	-118.30	frontal
57	4	17	2008	41.77	-97.21	frontal
58	4	20	2008	31.50	-78.05	frontal
59	4	23	2008	57.66	-92.73	frontal
60	4	26	2008	27.68	-96.86	convective
61	5	1	2008	9.45	-121.29	convective
62	5	5	2008	35.60	-97.03	convective
63	5	6	2008	57.70	-161.02	frontal
64	5	7	2008	19.64	-119.53	convective
65	5	11	2008	35.46	-88.59	frontal
66	5	13	2008	31.80	-60.65	frontal
67	5	21	2008	43.33	-108.19	frontal
68	5	22	2008	59.36	-119.18	frontal
69	5	26	2008	34.74	-87.85	convective
70	5	30	2008	18.81	-84.02	convective
71	6	1	2008	43.60	-102.87	convective
72	6	8	2008	18.40	-66.10	convective
73	6	10	2008	48.34	-123.00	frontal
74	6	12	2008	34.95	-89.85	convective
75	6	14	2008	20.47	-64.51	convective
76	6	19	2008	4.66	-32.17	convective
77	6	21	2008	22.59	-106.52	convective
78	6	25	2008	28.30	-82.88	sea breeze
79	6	27	2008	38.14	-63.81	frontal
80	6	29	2008	50.51	-71.19	frontal
81	7	3	2008	39.73	-86.27	frontal
82	7	6	2008	43.07	-93.87	convective
83	7	12	2008	46.44	-91.58	frontal
84	7	16	2008	43.99	-87.58	frontal
85	7	16	2008	45.09	-90.18	convective
86	7	17	2008	42.55	-95.10	frontal
87	7	20	2008	32.66	-109.82	convective
88	7	23	2008	43.74	-76.82	convective
89	7	23	2008	24.93	-97.12	convective
90	7	24	2008	44.67	-70.47	convective

91	8	2	2008	43.77	-73.04	convective
92	8	5	2008	29.54	-94.04	convective
93	8	6	2008	39.37	-90.18	convective
94	8	11	2008	40.65	-71.54	frontal
95	8	14	2008	25.80	-104.59	orographic
96	8	18	2008	30.90	-94.75	convective
97	8	19	2008	27.96	-80.16	convective
98	8	21	2008	32.21	-79.81	convective
99	8	27	2008	18.56	-67.06	convective
100	8	31	2008	27.53	-87.19	convective
101	9	1	2008	30.45	-91.23	convective
102	9	6	2008	41.38	-72.95	convective
103	9	9	2008	40.58	-74.18	convective
104	9	12	2008	42.29	-87.71	frontal
105	9	13	2008	36.88	-93.16	convective
106	9	14	2008	27.06	-97.91	frontal
107	9	17	2008	59.62	-114.79	frontal
108	9	21	2008	18.23	-68.03	frontal
109	9	25	2008	36.03	-69.79	frontal
110	9	30	2008	37.86	-76.11	frontal
111	10	2	2008	47.40	-68.91	frontal
112	10	4	2008	48.34	-125.51	frontal
113	10	7	2008	34.16	-94.39	convective
114	10	10	2008	45.95	-109.86	frontal
115	10	12	2008	52.48	-81.39	frontal
116	10	15	2008	40.15	-94.97	frontal
117	10	19	2008	45.71	-86.13	frontal
118	10	21	2008	46.92	-125.07	frontal
119	10	26	2008	43.13	-70.40	frontal
120	10	27	2008	21.86	-83.94	convective

Appendix A. The date, latitude and longitude point at the center of the grid, and the type of precipitation event for all simulations are shown.

References

- Adler, R. F., H.-Y. M. Yeh, N. Prasad, W.-K. Tao, and J. Simpson, 1991: Microwave simulations of a tropical rainfall system with a three-dimensional cloud model. *J. Appl. Meteor.*, **30**, 924-953.
- Adler, R. F., A. J. Negri, P. R. Keehn, and I. M. Hakkarinen, 1993: Estimation of monthly rainfall over Japan and surrounding waters from a combination of low-orbit microwave and geosynchronous IR data. *J. Appl. Meteor.*, **32**, 335-356.
- Adler, R. F., G. J. Huffman, and P. R. Keehn, 1994: Global tropical rain estimates from microwave-adjusted geosynchronous IR data, *Rem. Sens. Rev.*, **11**, 125-152.
- Berg, W. and R. Chase, 1992: Determination of mean rainfall from the special sensor microwave/imager (SSM/I) using a mixed lognormal distribution. *J. Atmos. Oceanic Technol.*, **9**, 129-141.
- Bohren, C.F., and D.R. Huffman, 1983: *Absorption and Scattering of Light by Small particles*. John Wiley & Sons, 530 pp.
- Casella, D, M. Formenton, W.-Y. Leung, A. Mugnai, P. Sanò, E.A. Smith, and G.J. Tripoli, 2009: Statistical analysis of a new European Cloud Dynamics and Radiation Database. EGU General Assembly. 2009.
- Cotton, W. R., G. J. Tripoli, R. M. Rauber, and E. A. Mulvihill, 1986: Numerical simulation of the effects of varying ice crystal nucleation rate and aggregation processes on orographic snowfall. *J. Clim. Appl. Meteor.*, **25**, 1658-1680.
- Diner, D. J, R. A. Kahn, A. J. Braverman, R. Davies, J. V. Martonchik, R. T. Menzies, T. P. Ackerman, J. H. Seinfeld, T. L. Anderson, R. J. Charlson, J. Bösenberg, W. D. Collins, P. J. Rasch, B. N. Holben, C. A. Hostetler, B. A. Wielicki, M. A. Miller, S. E. Schwartz, J. A. Ogren, J. E. Penner, G. L. Stephens, O. Torres, L. D. Travis, and B. Yu, 2004: Paragon: a systemic, integrated approach to aerosol observation and modeling. AIAA Space Meeting, San Diego, California, September 28, 2004.

English, S.J., and T.J. Hewison, 1998: A fast generic millimetre wave emissivity model. *Proc. SPIE on Microwave Remote Sensing of the Atmosphere and Environment*, 22-30.

Flatau, P., G. J. Tripoli, J. Berlinde, and W. Cotton, 1989: The CSU RAMS Cloud Microphysics Module: General Theory and Code Documentation. *Technical Report 451*, Colorado State University, 88 pp.

Fowler, M. G., H. K. Burke, K. H. Hardy, and N. K. Tripp, 1979: The estimation of rain rate over land from spaceborne passive microwave sensors. *Satellite Hydrology*, American Waters Resources Association, 101-108.

Hewison, T.J., and S.J. English, 1999: Airborne retrievals of snow and ice surface emissivity at millimetre wavelengths. *IEEE Trans. Geosci. Remote Sens.*, **37**, 1871-1879.

Hewison, T.J., and S.J. English, 2000: Fast models for land surface emissivity. *Radiative Transfer Models for Microwave Radiometry* (C. Matzler, Ed.), COST Action 712, Directorate- General for Research, European Commission, Brussels, Belgium, 117-127.

Hewison, T.J., 2001: Airborne measurements of forest and agricultural land surface emissivity at millimeter wavelengths. *IEEE Trans. Geosci. Remote Sens.*, **39**, 393-400.

Hoch, J. A., 2006: The cloud dynamics and radiation database a focus on orographic precipitation. *Master thesis*, UW-Madison.

Hollinger, J. DMSP Special Sensor Microwave/Imager Calibration/Validation, Naval Research Laboratory, Washington, D.C., Vol 1, July 20, 1989.

Hollinger, J. DMSP Special Sensor Microwave/Imager Calibration/Validation, Naval Research Laboratory, Washington, D.C., Vol 2, May 20, 1991.

Hollinger, J. P., R. C. Lo, G. A. Poe, R. Savage, and J. L. Peirce, 1987. "Special Sensor Microwave/Imager User's Guide," Naval Research Laboratory, Washington D.C., Sep. 14, 1987.

Hong, Y.-H., K. L. Sorooshian, and X. G. Gao, 2004: Precipitation Estimation from Remotely Sensed Imagery using an Artificial Neural Network Cloud Classification System. *Journal of Applied Meteorology*, **43**, 1834 - 1852.

Kawanishi, T., T. Sezai, Y. Ito, K. Imaoka, T. Takeshima, Y. Ishido, A. Shibata, M. Miura, H. Inahata, and R. W. Spencer, 2003: The Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), NASDA's contribution to the EOS for global energy and water cycle studies. *Geosci. Remote Sens.*, **41**, 184-194.

Kidd, C., and E. C. Barrett, 1990: The use of passive microwave imagery in rainfall monitoring. *Rem. Sens. Rev.*, **4**, 415-450.

Kidd, C., D. Kniveton, and E. C. Barrett, 1998: The advantages and disadvantages of statistically derived-empirically calibrated passive microwave algorithms for rainfall estimation. *J. Atmos. Sci.*, **55**, 1572-1582.

Kummerow, C. and L. Giglio, 1994: A passive microwave technique for estimating rainfall and vertical structure information from space, Part 1: Algorithm description. *J. Appl. Meteor.*, **33**, 3-18.

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**, 1213-1232.

Kummerow, C., W. Barnes, T. Kozu, J. Shiue, and J. Simpson, 1998: The Tropical Rainfall Measuring Mission (TRMM) sensor package. *J. Atmos. Oceanic Technol.*, **15**, 809-816.

Kummerow, C., Y. Hong, W. S. Olson, S. Yang, R. F. Alder, J. McCollum, R. Ferraro, G. Petty, D.-B. Shin, and T. T. Wilheit, 2001: The evolution of the Goddard Profiling Algorithm (GPROF) for the rainfall estimation from passive microwave sensors. *J. Appl. Meteor.*, **40**, 1801-1820.

Li, X., B. Plale, N. Vijayakumar, R. Ramachandran, S. Graves, and H. Conover, 2008: Real-time storm detection and weather forecast activation through data mining and events processing. *Earth Science Informatics.*, **1**, 49-57.

Liu, G. and J. A. Curry, 1992: Retrieval of precipitation from satellite microwave measurement using both emission and scattering. *J. Geophys. Res.*, **97**, 9959-9974.

Mugnai, A., E. A. Smith, and G. J. Tripoli, 1993: Foundation of physical-statistical precipitation retrieval from passive microwave satellite measurements. Part II: Emission source and generalized weighting function properties of a time dependent cloud-radiation model. *J. Appl. Meteor.*, **32**, 17-39.

Olson, W. S., 1989: Physical retrieval of rainfall rates over the ocean by multispectral microwave radiometry: Application to tropical cyclones. *J. Geophys. Res.*, **94**, 2267-2280.

Olson, W.S., S. Yong, J.E. Stout, and M. Grecu, 2007: The Goddard Profiling algorithm (GPROF): description and current applications. V. Levizzani et al. (eds.), *Measuring Precipitation from Space: EURAINSAT and the Future*, 179-188.

Panegrossi, G., S. Dietrich, F. S. Marzano, A. Mugnai, E. A. Smith, X. Xiang, G. J. Tripoli, P. K. Wang and J. P. V. Poyares Baptista, 1998: Use of cloud model microphysics for passive microwave-based precipitation retrieval: Significance of consistency between model and measurement manifolds. *J. Atmos. Sci.*, **55**, 1644-1673.

Petty, G.W., 1994a: Physical retrievals of over-ocean rain rate from multichannel microwave imagers. Part I: Theoretical characteristics of normalized polarization and scattering indices. *Meteor. Atmos. Phys.*, **54**, 79-99.

Petty, G. W., 1994b: Physical retrievals of over-ocean rain rate from multichannel microwave imagery. Part II: algorithm implementation. *Meteor. Atmos. Phys.*, **54**, 101-121.

Petty, G.W. and K.B. Katsaros, 1990: Precipitation observed over the South China Sea by the Nimbus-7 Scanning Multichannel Microwave Imager during WMONEX. *J. Appl. Meteor.*, 29, 273-287.

Petty, G.W. and K.B. Katsaros, 1992: Nimbus 7 SMMR precipitation observations calibrated against surface radar during TAMEX. *J. Appl. Meteor.*, 31, 489-505.

Quinlan, J.R., 1993: C4.5: Programs for machine learning, Morgan Kaufmann, San Mateo, California.

Roberti, L., J. Haferman, and C. Kummerow, 1994: Microwave radiative transfer through horizontally inhomogeneous precipitating clouds. *J. Geophys. Res.*, **99**, 707-716.

Sanò, P., D. Casella, S. Dietrich, F. Di Paola, M. Formenton, W.-Y. Leung, A. Mehta, A. Mugnai, E. A. Smith and G. J. Tripoli, 2010: Bayesian estimation of precipitation from space using the Cloud Dynamics and Radiation Database (CDRD) approach: Application to case studies of FLASH and H-SAF Projects. *Nat. Hazards Earth Syst. Sci.*, to be submitted.

Schluessel, P. and H. Luthardt, 1998: Surface wind speeds over the North Sea from Special Sensor Microwave/Imager observations. *J. Geophys. Res.*, **96**, 4845-4853.

Schwarz, Gideon E., 1978: Estimating the dimension of a model, *Annals of Statistics* **6** (2): 461-464.

Smith, E. A., C. Kummerow, and A. Mugnai, 1994: The emergence of inversion-type profile algorithms for estimation of precipitation from satellite passive microwave measurements. *Remote Sens. Reviews*, **11**, 211-242.

Smith, E. A., and A. Mugnai, 1988: Radiative transfer to space through a precipitation cloud at multiple microwave frequencies. Part II: Results and analysis. *J. Appl. Meteor.*, **27**, 1074-1091.

Smith, E. A., and A. Mugnai, 1989: Radiative transfer to space through a precipitation cloud at multiple frequencies. Part III: Influence of large ice particles. *J. Meteor. Soc. Japan*, **67**, 739-755.

Smith, E. A., A. Mugnai, H. J. Cooper, G. J. Tripoli and X. Xiang, 1992a: Foundations for statistical-physical precipitation retrieval from passive microwave satellite measurements. Part I: Brightness-temperature properties of a time- dependent cloud-radiation model. *J. Appl. Meteor.*, **31**, 506-531.

Smith, E. A., J. E. Lamm, R. Adler, J. Alishouse, K. Aonashi, E. Barrett, P. Bauer, W. Berg, A. Chang, R. Ferraro, J. Ferriday, S. Goodman, N. Grody, C. Kidd, D. Kniveton, C. Kummerow, G. Liu, F. Marzano, A. Mugnai, W. Olson, G. Petty, A. Shibata, R. Spencer, F. Wentz, T. Wilheit, and E. Zipser, 1998: Results of WetNet PIP-2 Project. *J. Atmos. Sci.*, **55**, 1483-1536.

Smith, E. A., G. Asrar, Y. Furuhashi, A. Ginati, A. Mugnai, K. Nakamura, R. F. Alder, M. Chou, M. Desbois, J. F. Durning, J. K. Entin, F. Einaudi, R. R. Ferraro, R. Guzzi, P. R. Houser, P. H. Hwang, T. Iguchi, P. Joe, R. Kakar, J. A. Kaye, M. Kojima, C. Kummerow, K. Kuo, D. P. Lettenmaier, V. Levizzani, N. Lu, A. V. Mehta, C. Morales, P. Morel, T. Nakazawa, S. P. Neeck, K. Okamoto, R. Oki, G. Raju, J. M. Shepherd, J. Simpson, B. Sohn, E.F. Stocker, W. Tao, J. Testud, G. J. Tripoli, E. F. Wood, S. Yang, and W. Zhang, 2007: International global precipitation measurement (GPM) program and mission: an overview. V. Levizzani et al. (eds.), *Measuring Precipitation from Space: EURAINSAT and the Future*, 611-653.

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.

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

Tassa, A., S. Di Michele, A. Mugnai, F.S. Marzano and J.P.V. Póiares Baptista, 2003: Cloud-model based Bayesian technique for precipitation profile retrieval from the Tropical Rainfall Measuring Mission Microwave Imager. *Radio Sci.*, **38**, 8074-8086.

Todd, M. C. and J. O. Bailey, 1995: Estimates of rainfall over the United Kingdom and surrounding seas from the SSM/I using the polarization corrected temperature algorithm. *J. Appl. Meteor.*, **34**, 1254-1265.

Tripoli, G. J. and E. A. Smith, 2010: Scalable nonhydrostatic cloud/mesoscale model featuring variable-stepped topography coordinates: Formulation and performance on classic obstacle flow problems. *Mon. Wea. Rev.*, submitted.

Tripoli, G. J., and W. R. Cotton, 1981: The use of ice-liquid water potential temperature as a thermodynamic variable in deep atmospheric models. *Mon. Wea. Rev.*, **109**, 1094-1102.

Tripoli, G. J., 1992: A nonhydrostatic model designed to simulate scale interaction, *Mon. Wea. Rev.*, **120**, 1342-1359.

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. Meteor.*, **16**, 551-560.

Wu, R. and J. A. Weinman, 1984: Microwave radiances from precipitating clouds containing aspherical ice, combined phase, and liquid hydrometeors. *J. Geophys. Res.*, **89**, 7170-7178.