THESIS

LEVERAGING MACHINE LEARNING FOR WEATHER RADAR QUALITY CONTROL AND MICROPHYSICAL RETRIEVALS

Submitted by

Isaac R. Schluesche

Department of Atmospheric Science

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2025

Master's Committee:

Advisor: Michael M. Bell

Russ Schumacher
Haonan Chen

ABSTRACT

LEVERAGING MACHINE LEARNING FOR WEATHER RADAR QUALITY CONTROL AND MICROPHYSICAL RETRIEVALS

Data from meteorological radars must undergo an extensive quality control process in order to become useful in research and analysis. A primary part of some quality control procedures is the removal of measurements returned by non-meteorological features such as the earth's surface and biological targets (birds, insects, etc.). Existing methods struggle to achieve acceptable levels of non-meteorological data removal in a timely fashion, resulting in the development of RONIN.jl (Random forest Optimized Nonmeteorological IdentificatioN) - an open-source Julia package for the end-to-end tuning, training, and testing of random forest (RF) models to remove non-meteorological features from radar data. It is shown that Ronin is able to achieve performance that meets or exceeds current operational products while operating at a speed orders of magnitude faster than a prototype experimental machine learning based method.

Hydrometeor size distributions (HSDs) are quantities of great interest in a range of meteorological disciplines including cloud microphysics and numerical modeling. Prior literature has shown that these distributions can be succinctly represented through normalization by two or three integral moments of the distribution itself, with effective normalizations drastically reducing variability toward a distribution that does not vary across climactic regimes or precipitation habits. In this study, a novel three-moment normalization is employed to generate an extensive library of simulated HSDs, broadly conditioned on observations, in service of training a retrieval algorithm for the full distribution. Simulated radar variables are also computed for the distributions contained in the synthetic dataset. Subsequently, several Artificial Neural Networks (ANNs) are trained to use the radar variables as input to retrieve the full HSD through a quartet of different mathematical techniques. The four techniques are then evaluated on several real-world datasets. It is shown that

all methods examined are at least somewhat effective at retrieving moments, even of lower order, of observed distributions, and produce distributions that match well with both radar and disdrometer observations.

ACKNOWLEDGEMENTS

The work contained herein would not have been possible without the generous funding support of the National Oceanic and Atmospheric Administration, specifically award NA23OAR4590408-T1-01, and the National Science Foundation, specifically award AGS-2153337. I am very grateful to these institutions for allowing me the opportunity to conduct this research.

This work has also benefited from the support and guidance of a number of brilliant scientists. Dr. Paul Reasor was instrumental in the development of RONIN, and I thank him for his tireless work hand-editing radar sweeps for its training. Dr. Alex DesRosiers, Dr. John Gamache, Dr. Brad Klotz, and Kelley Neighbour also provided valuable feedback on the project. I further wish to thank my committee, Drs. Russ Schumacher and Haonan Chen. Of course, none of this work would be possible without my advisor Dr. Michael Bell, and I am thankful for his brilliance and insight.

I have benefited from being a part of an incredible community at Colorado State University and in Fort Collins. At the department, I'd like to send my gratitude to Sarah, Nate, my professors, the Weather Challenge team, the Bell group, and the incoming class of 2023. To Nick and Delián, I couldn't have done it without your friendship, support, and brilliance. To Nya, your love, kindness and support mean the world and this thesis would not have been possible without.

I'd like to thank Kaylan, Leila, Bella, Megan, Dr. Michael Morgan, Dr. Jonathan Martin, and Dr. Alex Goldstein for their friendship, support, and mentoring during my time at the University of Wisconsin-Madison.

Finally, I'd like to thank Mark, Tara, and Jonathan, my amazing and loving family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Meteorological radars are a cornerstone of observational science in the field Fabry (2018), and have enabled myriad breakthroughs in all facets of the discipline, ranging from high impact weather to numerical modeling (Wurman et al. (1996), Marks (2003), Ryzhkov et al. (2022)). The primary reason for the radar's indispensable status is their unique ability to derive information about the three-dimensional kinematic, cloud, and precipitation state of the atmosphere at relatively high spatiotemporal resolutions - valuable measurements that would be nearly impossible to obtain with in-situ instrumentation. The above succinct description may obscure many of the complexities associated with creating a three-dimensional evaluation of the atmospheric state. In general, performing such an analysis is a multiple step process involving complex mathematical interpolations of observed data to create a coherent result (Gao et al., 2004). One of the most important tasks in obtaining any high quality product occurs prior to analysis - the removal of non-meteorological or extraneous data from the input fields is essential to sensible downstream findings. At the time of writing, the most effective and accurate way of removing undesirable features, especially for airborne radars, is to do so by hand, requiring an experienced meteorologist to parse the data one gate at a time. The arduous nature of hand-editing has spurred the invention of numerous automated techniques aimed at addressing the problem, with many methods removing data not meeting thresholds of certain variables, frequently quantities akin to normalized coherent power. Unfortunately, automated methods generally result in the removal of undesirable quantities of so-called "good" data that could be used in analysis Gamache et al. (2008), Bell et al. (2013). Recently, DesRosiers and Bell (2024) developed an algorithm that used a Random Forest model to achieve increased amounts of meteorological data retention relative to prior methods. This work presents an extension to the DesRosiers and Bell technique (DesRosiers and Bell, 2024) that improves both model performance and speed.

At their core, all radar moments are intimately tied to the concentration, orientation, and velocity of atmospheric targets inside a given pulse volume Fabry (2018). In the case of a cloud, the targets dictating the measurements returned to the radar are primarily hydrometeors. As a result, much insight may be gleaned as to the evolution of the Drop Size Distribution (DSD) in space and time. The linkage has long been known - Sekhon and Srivastava began attempting to measure DSDs in thunderstorms using Doppler spectra as early as 1970 (Sekhon and Srivastava, 1970). Recent literature has begun to further capitalize on the ability of radar to inform HSDs, with Raupach and Berne (2017) and Bringi et al. (2020) both endeavoring to use the information provided by dual-polarimetric radar to retrieve lower-order moments of the drop size distribution to mixed results. Ho et al. (2023a) demonstrates through an improvement in radar-based rainfall estimation that the complex nonlinear decision making capabilities of Machine Learning (ML) also show promise in this arena. This work applies ML techniques in concert with a novel triple-moment normalization to develop a retrieval model for the full distribution from dual-polarimetric radar data. This technique is shown to be capable of producing rainfall rates and smaller order moments that are well correlated with disdrometer observations, using both data from radiative scattering models and two research radars as input.

Chapter 2

Improvements to Airborne Radar Quality Control Using Machine Learning

2.1 Introduction

The quality of the science that may be achieved with radar data is highly dependent on the efficacy of the quality control (QC) process to remove Non-Meteorological Data (NMD) and retain Meteorological Data (MD) (Lakshmanan et al., 2007). Removal has traditionally been done by hand in order to obtain the highest-fidelity analyses, requiring a meteorologist to carefully parse through each individual sweep. Processing a single sweep can take experts on the order of five or more minutes, requiring many hours of effort to clean the quantity of data required for analysis (Bell et al., 2013).

Several automated techniques have been developed for stationary ground-based radars (Steiner and Smith (2002), Lakshmanan et al. (2007)). These algorithms generally function by computing statistical properties of gates and those surrounding and providing them as input to a neural network or some other technique. Maps of areas of echo consistently provided by targets on the ground ("clutter maps") may also be developed in order to remove persistent nonmeteorological features. More recently, algorithms using Deep Neural Networks (DNNs) in concert with similar input features and polarimetric data have been proven to be successful on the United States NEXRAD operational network (Lakshmanan et al., 2014).

Mobile platforms, especially airborne ones, present an altogether more challenging scenario, generally featuring fundamentally different scanning geometries and NMD characteristics. A characteristic further distinguishing the current generation of airborne QC from the ground based case is the lack of dual-polarimetric information. Dual-pol values are often essential in the QC

pipeline, yielding valuable information about the size and shape of scatters for their classification as MD/NMD.

Fig 2.1(a) demonstrates some features common to airborne sweeps that must be removed. Elements immediately conspicuous include the strip of high (>40 dbZ) reflectivity about 4 km below the platform (earth surface), a circular band about 5 km from the radar (side lobe), and a second-trip echo along several radials to the right of the platform. Though all three appear with a high frequency in the airborne case, they vary dynamically and so prove difficult to remove using traditional techniques.

Airborne radars are of particular import in operational Tropical Cyclone (TCs) forecasting. Even when only considering the United States, TCs are responsible for over two-thirds of worldwide catastrophe monetary losses (Weinkle et al., 2018), and so accurate forecasts are a top priority for the National Weather Service (NWS) and other agencies. Oftentimes TCs develop and intensify in the deep ocean far from in-situ instrumentation, and so remote sensing techniques, including airborne radar, are integral to accurate forecasts. Toward this end, the NWS employs 3 airborne radars - two deployed aboard P3-Orion aircraft, and one aboard a Gulfstream IV jet - to fly missions through both developing and mature TCs. These radars operate at X-band (3-cm wavelength) and are known as Tail Doppler Radars (TDRs) (NOAA Hurricane Research Division, 2025). In addition to providing valuable information for human forecasters, the inclusion of TDR data in hurricane forecasting models can reduce track and intensity error (Gall et al. (2013), Christophersen et al. (2022)). As such, having high-quality QC'ed sweeps from the TDR in near real time is of great interest to forecasters.

In an effort to increase radar data availability for operational meteorology as well as decrease the effort required to obtain research-quality datasets, a couple of methods have been developed aimed at automating NMD removal for the airborne case. Gamache et al. (2008) created a multi-step process that removed several different types of non-meteorological artifacts using data thresholds specified apriori that is currently employed in operations at the National Hurricane Center (NHC). Bell et al. (2012) generalized the rules-based method to remove these features, and was

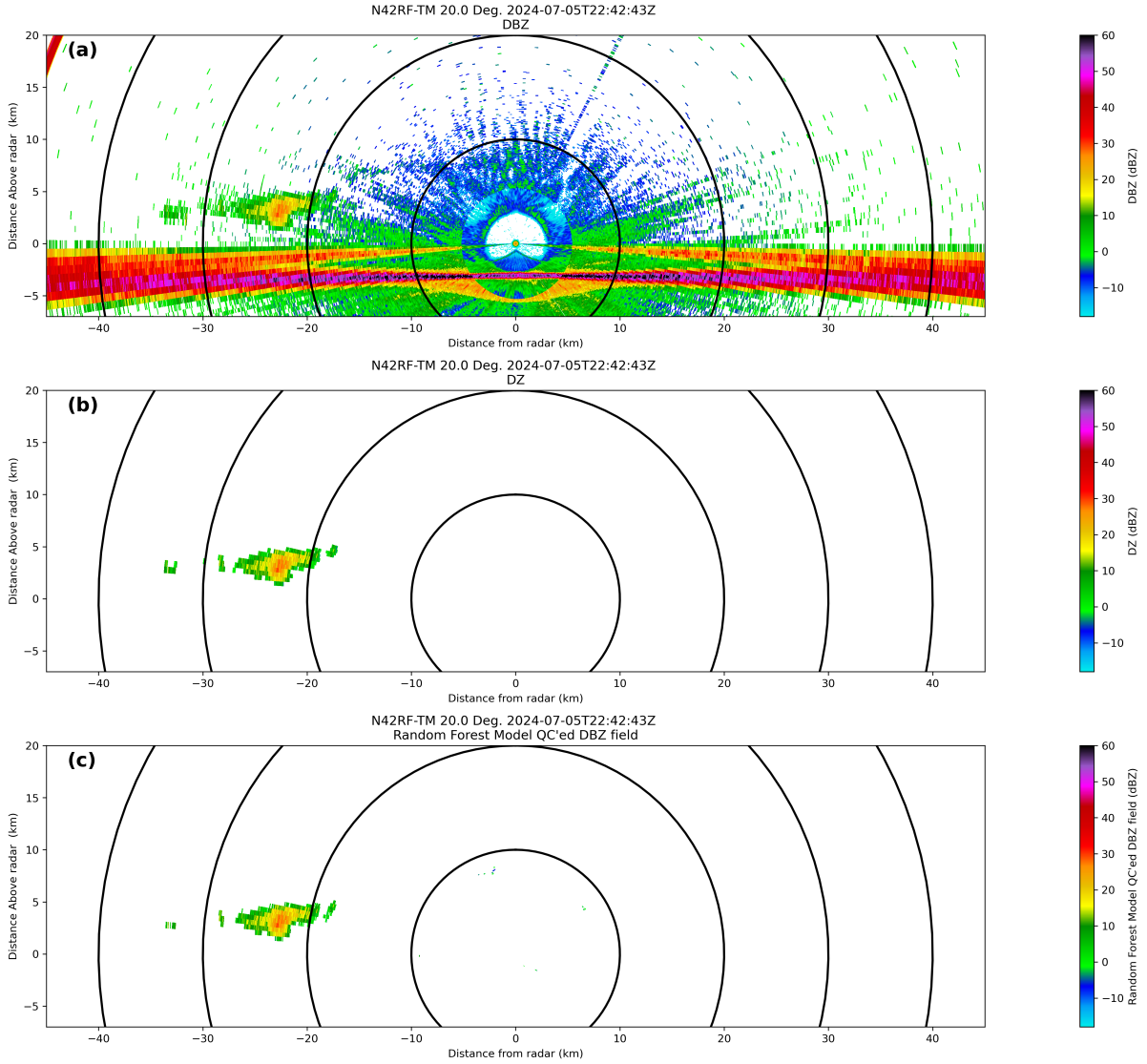

Figure 2.1: Raw Radar Refelctivity (a), hand QC reflectivity (b), and machine learning QC reflectivity (c) from the NOAA Tail Doppler Radar aboard NOAA-42 at 22:42 UTC September 5th, 2024

able to achieve removal of up to 95% nonweather echo on an airborne dataset comprising a variety of environments, but at the expense of removal of about 15% of desirable MD. In an effort to improve the amount of MD retained, DesRosiers and Bell (2024; hereafter DB24) invented a technique using Random Forest (RF) models that achieved about 96% NMD removal while still retaining 93% or more of MD.

This work aims to improve both the performance and execution speed of DB24 through a rewrite of the code base into the Julia programming language as well as modification of the RF model. This goal resulted in the development of an open source software package dubbed Random-forecast Optimized Nonmeteorological IdentificatioN (RONIN; henceforth Ronin).

The remainder of the chapter is organized as follows. Section 2.2 discusses the key points of the DB24 methodology, as well as the technical implementation of Ronin. Section 2.3 gives a discussion of the performance of Ronin on several new datasets, and Section 2.4 concludes with an overview of the results and directions for future work.

2.2 Data and Method

Ronin is a direct extension of DB24. In an effort to avoid restating their endeavors, only the points most critical to understanding of the algorithm are outlined at a high level below. Much work typically associated with ML model creation and design such as hyperparameter tuning has already been conducted and is further deferred to that manuscript. Ronin does deviate from the DB24 methodology in several key areas, however. The algorithm described in DB24 was built and tuned to be used with the Electra Doppler Radar (ELDORA), a predecessor to the TDR. Different platforms feature different NMD characteristics, so it's possible this configuration would not be optimal for other radars. Ronin was structured to be flexible enough to be used with any radar system, present or future. Additionally, Ronin is evaluated on a dataset comprising exclusively of tropical cyclones in order to show its utility in supplementing or supplanting current operational tools, in contrast to the broader range of impactful weather used by DB24. Further design details are discussed in the following sections.

2.2.1 Training, Testing, and Validation Dataset

This work utilizes a dataset comprised of TDR scans from both the P3-Orion and GIV aircraft spanning a variety of TC cases. The data are spread across a total of five cases - legs of flights from Hurricanes Earl (2022), Fiona (2022), Lee (2023), Beryl (2024), and Milton (2024). In an effort to include both land and ocean surface NMD for a more generalizable model, the Beryl case is from late in the day on July 5th, when Beryl was traversing the Yucatán Peninsula of Mexico. An example 360 degree "sweep" from the Beryl flight is shown in Figure 2.1. These sweeps were hand-edited by Dr. Paul Reasor, a radar meteorologist at the Atlantic Oceanic and Meteorological Laboratory (AOML) Hurricane Research Division (HRD), who worked through each individual sweep to remove gates considered NMD.

The dataset is selected in lieu of the dataset of DB24 both in an effort to evaluate the technique on a different set of data and radar characteristics as well as in support of Ronin's development as an option for real time QC for the NWS. The cases vary in intensity from a mid range Tropical Storm (Beryl) to an extremely intense Category 5 Hurricane (Milton), and, as mentioned above, feature overland and overwater data. These data span a range of scenarios an algorithm would be expected to encounter in operations.

The training/testing/validation split is performed largely in accordance with DB24, with an 72%/20%/8% division and similar care being taken to select consecutive periods of data for each in order to minimize temporal autocorrelation between partitions. One difference is that the division is performed on a sweep-by-sweep basis, instead of on the gate level, a change applied due to the necessity of retaining adjacent gates for the calculation of spatial features (discussed below) in an iterative methodology. The five cases in total comprise 1303 sweeps. In order to avoid model bias toward one case or another in evaluation, the same number of sweeps from each case are placed into the testing and validation set, with the remainder being assigned to the training dataset. In total, this leads to a validation dataset of 80 sweeps (16 per case), a training- dataset of 963 sweeps, and a testing dataset of 260 sweeps (52 per case).

2.2.2 Method Details

At a high level, the method described in DB24 functions as follows. For a given radar sweep, an initial QC process is undertaken where gates that are obviously non-meteorological are immediately classified as NMD in order to only apply the RF model to gates that are more difficult to classify. DB24 defines "obviously non-meteorological" as a gate having a low signal quality index (SQI) or high probability of intersecting the ground. Subsequently, input features are computed on a gate-by-gate basis, and are then used as input to a RF model, which classifies each gate in the sweep as MD or NMD, with NMD gates then removed during the QC process. More details on each step in the process are provided below.

Random Forests

Random Forests are a type of ML technique that utilize ensembles of decision trees for classification or regression problems. Though simpler than Deep Learning (DL) models, they are still extremely powerful and have been used widely in the atmospheric sciences. For classification, decision trees operate by learning functions of input features that divide data in ways that maximize similarity among resultant groups. In the case of a Random Forest, the technique is extended such that each tree in the ensemble only sees a subset of the input features, reducing possibilities for overfitting and increasing generalization abilities. In the following, RF models are trained using the hyperparameters derived in DB24 (21 trees, maximum node depth of 14). For training, gates that are removed/marked NMD in the hand-edited QC described above are assigned a label of 0/false, and those retained/marked MD are assigned a label of 1/true.

Input Feature Calculation

Perhaps the most critical portion of the QC process is the computation of input features that are fed to any downstream model for gate classification. The simplest are basic radar moments such as reflectivity (DBZ) and Doppler velocity. Following Lakshmanan et al. (2007), statistical quantities such as the average value and standard deviation of neighboring gates are also frequently used. DB24 also defines several parameters especially useful in the airborne case, such as the probability

a gate intersects the ground (PGG), and the height of a gate above mean sea level (AHT). A listing of all features currently implemented in Ronin is provided in Table 2.1 along with a brief description.

Table 2.1: Summary of Ronin model derived input features. Parenthesis following a feature name indicates the feature acts upon a radar moment such as reflectivity (DBZ) or Doppler velocity.

Feature Name	Description
AVG()	Average value of gates surrounding current gate
STD()	Standard deviation of value of gates surrounding current gate
ISO()	Number of gates surrounding current gate without data
AHT	Elevation of gate above Mean Sea Level
PGG	Probability gate intersects with ground
SIG	Signal Quality (also known as SQI or NCP)
RNG/NRG	Range/Range Normalized by height to gate

2.2.3 Technical Implementation

Probabilistic Capabilities

Inspired by the "low", "medium", and "high" levels of removal in Bell et al. (2012) as well as a desire to leverage the naturally probabilistic nature of RF models, Ronin provides an advancement over DB24 through allowing users to specify confidence levels to remove data at. As mentioned above, RFs are composed of multiple different decision trees, each providing a classification for an input gate. As such, a "pseudo-confidence" in a given class may be easily obtained by calculating the fraction of trees in agreement on a designation. It is important to note that these pseudo-confidence levels are uncalibrated - that is, enforcing that 90% of the trees in an ensemble must agree on the classification of a gate does not necessarily ensure that this classification will be correct in 90% of realizations.

Regardless, this capability allows a user to choose between prioritizing NMD removal, MD retention, or some balance. Figure 2.2 provides a demonstration of this technique. Panel (a) displays the hand QC analysis/verification. Panel (b) displays the Ronin quality control, but with the added caveat that at all decision trees in the ensemble must agree on a classification of MD for

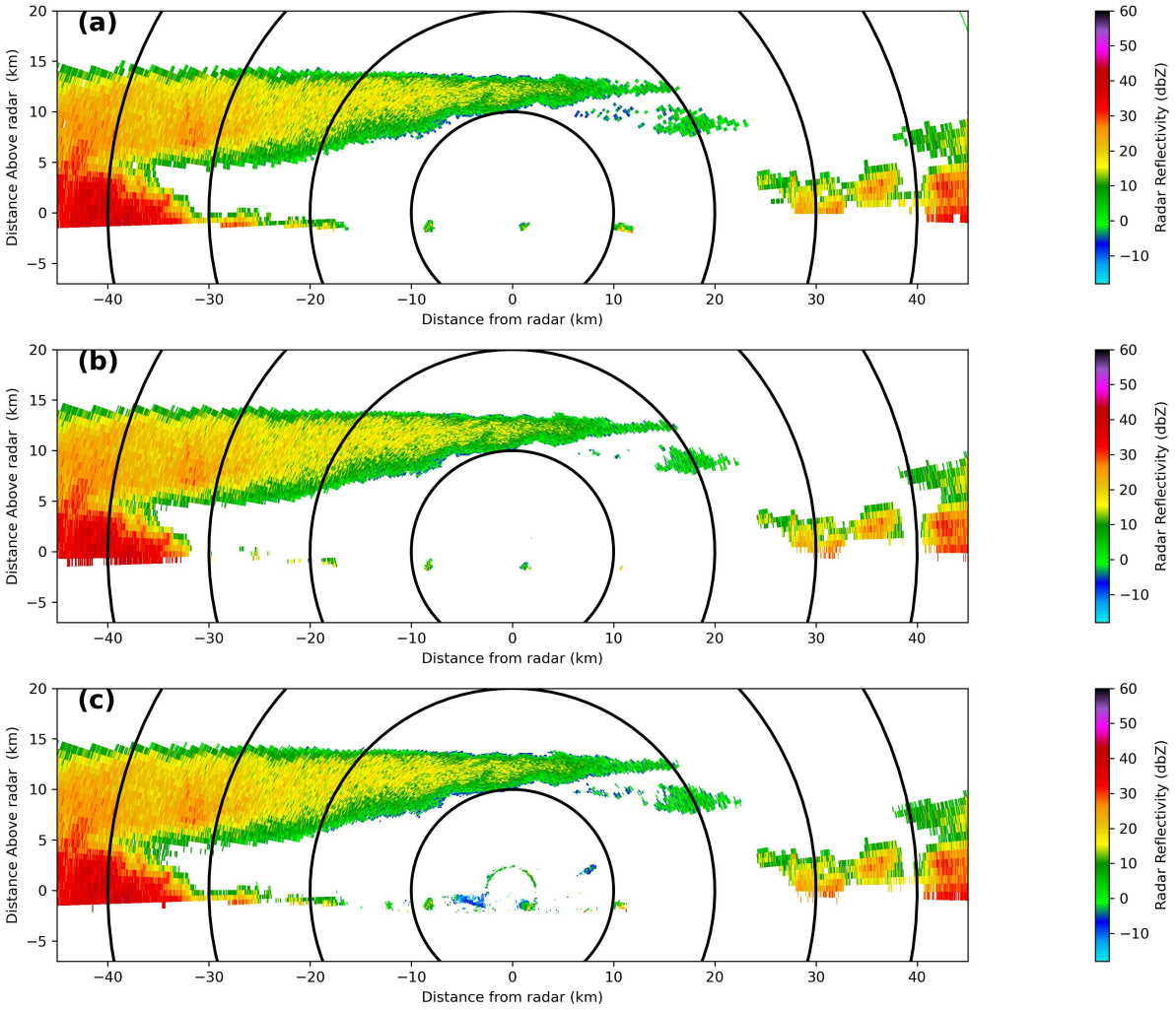

Figure 2.2: Hand QC Reflectivity (a), machine learning QC with high meteorological data threshold (b), and machine learning QC reflectivity with low meteorological data threshold (c) from the NOAA Tail Doppler Radar aboard NOAA-42 during a flight through Hurricane Fiona at 12:31 UTC September 18th, 2022

a final MD class. Panel (c) is as in (b), but now a gate is classified as meteorological if any one tree labels it MD. Panel (c) clearly retains more data than (b), and some of the retained echo is clearly meteorological, especially in the boundary layer between the -30 and -20 km range rings. However, more clutter around the radar is also readily apparent in (c) when compared to (b), which could degrade downstream analyses.

Multiple Passes

The probabilistic capability of Random Forests described above also guided the development of an iterative QC procedure, whereby gates that featured high agreement on a given classification were removed from consideration, and remaining gates subsequently used to train a new model that then acted upon them. It was hypothesized that a model trained only on so-called "difficult" gates would have better performance on them than one trained on all gates. This was further motivated by the fact that spatial statistical characteristics such as window mean and standard deviation would change for remaining gates after a first-pass QC was performed. Early experimentation has revealed only modest success with this technique, and further evaluation needs to be undertaken for a full accounting of the capabilities of an iterative method.

Julia Programming Language

One of the primary motivations for extending the DB24 technique was to improve the execution speed of the QC process. Through the author's experimentation, the Python implementation in DB24 takes on the order of several minutes to apply QC to a radar sweep. This is generally not quick enough for operational purposes. For example, in the case the NOAA TDR, new sweeps may be produced as rapidly as every two seconds on an operational mission. For cases such as flights through tropical cyclones, applying an automated QC in real time could result in dramatic improvements in realtime data availability. Ronin's implementation addresses this problem through its utilization of the Julia programming language (Bezanson et al., 2012). Julia is an open-source, dynamically typed, high-level language that is complied "just in time" (JIT). JIT compilation combines the benefits of an interpreter and a compiler allowing for an approachable, interactive syntax

while still providing performance comparable to "ahead-of-time" compiled languages such as C. This enables the entire scan cleaning process to be completed in as little as one to two seconds per sweep, meeting operational requirements.

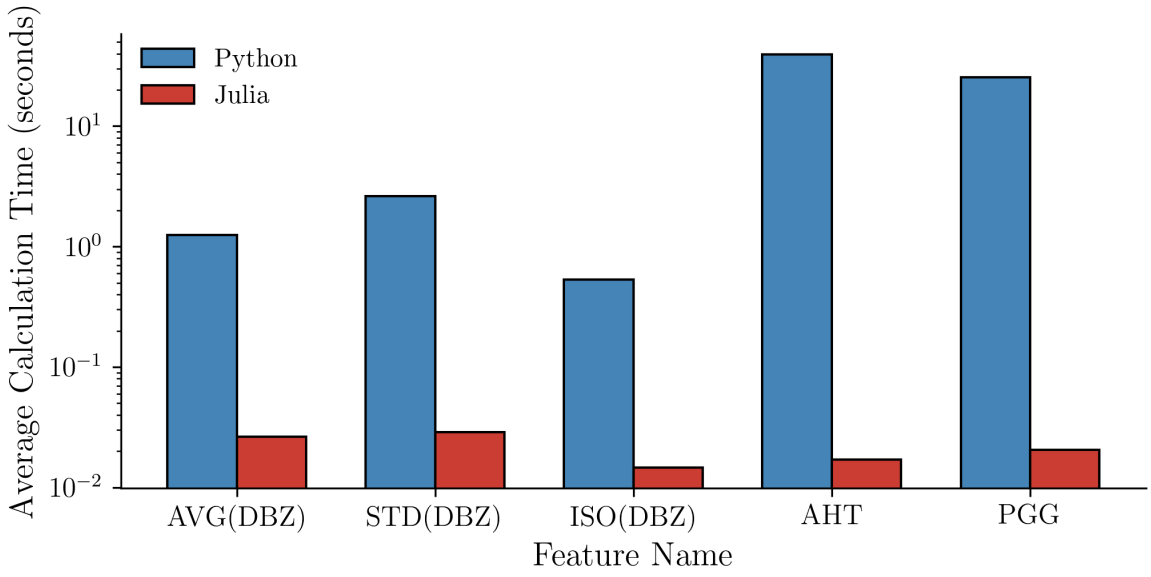

Figure 2.3: Comparison of average time in seconds to calculate window average DBZ (AVG(DBZ)), window standard deviation DBZ (STD(DBZ)), window average isolation DBZ (ISO(DBZ)), gate height (AHT), and probability gate is intersecting the ground (PGG)

Generally, the most time-consuming portion of the QC process is the calculation of input features - classification using features with an already-trained model is relatively quick. The native speed benefits of Julia over Python area are illustrated in figure 2.3. Even for simple features, such as the window-average DBZ, the Julia implementation is more than an order of magnitude quicker than the original, a benefit that increases for more complex operations. For the calculation of the gate height (AHT) and ground gate probability parameter (PGG), speedups nearing three orders of magnitude. Though it is likely that both code bases could be optimized significantly, the above provides a good example of the efficiency that can be obtained with even a relatively naive implementation in Julia.

2.2.4 Model Configuration

Details are now provided for the configuration of the RF model used henceforth. As in DB24, feature selection is performed using logistic regression with a varying penalty value λ/regularization strength in order to discern which input variables are most useful in discriminating between weather and nonweather echo. Results from this process are presented in 2.4, with panel $a)$ showing the weighted F_1 score of the model (DesRosiers and Bell, 2024). In this case, a larger λ value increases penalties on nonzero regression coefficients. We elect the features retained with a λ of $8.88 * 10^{-3}$, as the weighted F_1 score increases by less than half a percent across all other values of λ after this mark. This value is thought to strike a balance between an effective model and an appropriate number of input features. Panel $b)$ contains the resultant regression coefficients for the selected regularization strength. Features with nonzero weights are the isolation of radar reflectivity and Doppler velocity (ISO(DBZ), ISO(VEL)), radar reflectivity (DBZ), gate height AGL (AHT), range to gate (RNG), range to gate normalized by height AGL (AHT), and Probability Gate intersects ground(PGG). We elect to only utilize 5 of these features: radar reflectivity (DBZ), isolation of radar reflectivity (ISO(DBZ)), gate height AGL (AHT), range to gate normalized by altitude (NRG), and probability gate intersects ground (PGG). Isolation of Doppler velocity is removed since it is the same as isolation of reflectivity, and range is neglected since it provides information that is already contained within the combination of the AHT and NRG parameters.

2.3 Results

In order to finalize model configuration for future evaluation, a Receiver Operating Characteristic (ROC) curve was also calculated in order to determine an optimal threshold for the classification of data as meteorological. A ROC curve displays the relationship between True Positive Rate (TPR) and False Positive Rate (FPR) for a classification model. TPR is defined as the ratio between the number of successful model classifications of meteorological data and the sum of the number of successful weather data classifications and misclassifications of weather data as nonweather. FPR is further defined as the ratio between the number of misclassifications of nonmeteorological

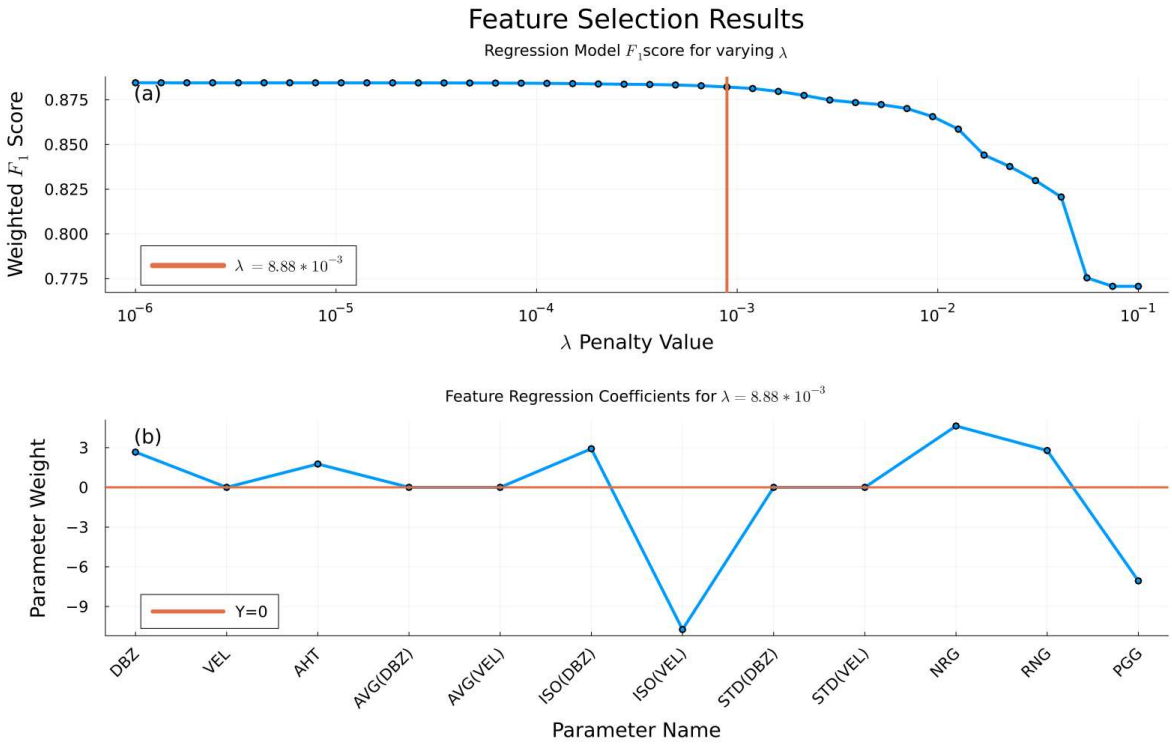

Figure 2.4: Results of Feature Selection process. a) F_1 score for logistic regression model with varying λ penalty values on X axis. Plot is annotated with selected final value of λ. b) Regression coefficients for input features with selected λ of $8.88 * 10^{-3}$.

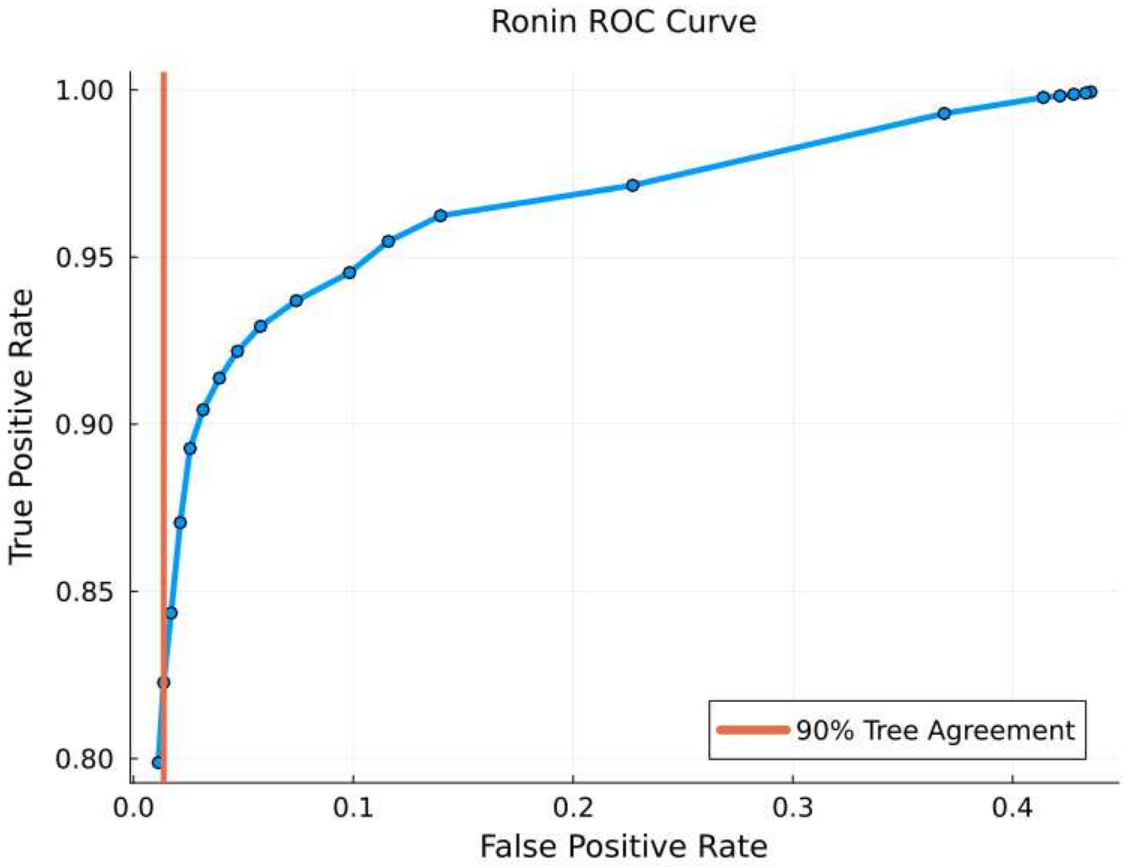

Figure 2.5: Ronin Receiver Operating Characteristic (ROC) curve evaluated on testing dataset. False Positive Rate (FPR) is displayed on the x-axis against True Positive Rate (TPR) on the y -axis. The selected threshold of 90% agreement is denoted with a vertical line

data as weather echo and the sum of these misclassifications with the number of successful classifications of NMD. Succinctly, TPR is the fraction of the total number of weather gates that are successfully classified as MD, and FPR is the fraction of the total number of nonweather gates that are misclassified as MD. To calculate a ROC curve, these rates are calculated at each different probability threshold capable for a model. In this manner, it is possible to visualize the aforementioned tradeoff between MD retention and NMD removal and make a decision according to one's priorities. Figure 2.5 displays the results of these calculations for the Ronin model, with FPR on the x axis and TPR on the y axis. The goal in this case is to maximize TPR while minimizing FPR. It is also important to note that an increasing FPR is associated with lower confidence thresholds in the underlying model. The steep positive slope with increasing FPR near small FPR values suggests that even modest decreases in confidence threshold result in large improvements in meteorological data retention. In the case of operational radar QC, even a small amount of nonweather echo may be detrimental to downstream analysis, and so we elect to use a threshold of 90% tree agreement (denoted with yellow vertical line) to prioritize NMD removal. Results for this threshold are now presented.

Ronin is currently undergoing evaluation by NOAA's Hurricane Research Division (HRD) as an option for a new realtime operational quality control. The current generation QC, dubbed NOAA-QC, is described in Gamache et al. (2008), and uses specific thresholds of signal-to-noise ratio and spectrum width to remove nonweather echo, along with other rules aimed at address side lobe and ocean surface levels. It is generally effective at removing NMD, but also suffers from an undesirable amount of MD removal. In order to compare Ronin's performance directly to NOAA-QC, both models are applied to the testing dataset. Table 2.2 contains two contingency matrices summarizing the results of the evaluation. When prioritized for NMD removal, Ronin is able to remove 98.6% of the NMD in the testing dataset, while retaining over 82% of the MD. This is an increase of about 10% over the operational dataset in meteorological data retention, a significant improvement, while still maintaining the same level of NMD removal. Testing by the author has shown that successful classification rates of weather data may be further increased by

using more input features, but a more detailed evaluation must be completed in order to obtain an understanding of the optimal parameters.

Table 2.2: Contingency Matrices for RONIN (a) and NOAA-QC (b) for the testing set. Human MD denotes data that was classified as meteorological by a human during the QC process, and NMD non-meteorological. Ronin/NOAA-QC MD represents data classified by each model as meteorological, and NMD as non-meteorological. NOAA-QC data courtesy of Paul Reasor.

(a) Ronin Testing Set

	RONIN MD	**RONIN NMD**
Human MD	82.3%	17.7%
Human NMD	1.4%	98.6%

(b) NOAA-QC Testing Set

	NOAA-QC MD	**NOAA-QC NMD**
Human MD	72.3%	27.7%
Human NMD	1.2%	98.8%

One of the end goals of radar QC is to provide data for 2D or 3D syntheses of the atmospheric state. Through we have shown that Ronin is effective in a bulk sense, it is also important to understand its impact on downstream products. Though a comprehensive analysis has yet to be completed, early results are promising. Figure 2.6 contains a pair of wind syntheses for a pass of Hurricane Beryl with data QC'ed by NOAA-QC (a) and Ronin (b). The increase in meteorological data is immediately bourne out in much greater coverage of resultant vector wind data, especially in the eyewall. An increased wind maxima is seemingly resolved that is not apparent in the NOAA-QC'ed version of the data. It is possible that increased resolution and coverage in 3D wind analyses could have beneficial impacts on NWP in the future, as TDR data has been shown to be an important source of information for tropical cyclone modeling (Christophersen et al., 2022).

2.4 Conclusions and Future Work

An extension to the ML technique for airborne radar quality control described in DB24, coined Ronin, is presented. Ronin is able to maintain state-of-the art performance in both non-meteorological data removal and meteorological data retention, while operating at speeds orders of magnitude quicker than the prior iteration. On a testing dataset comprising a diverse array of airborne tropical

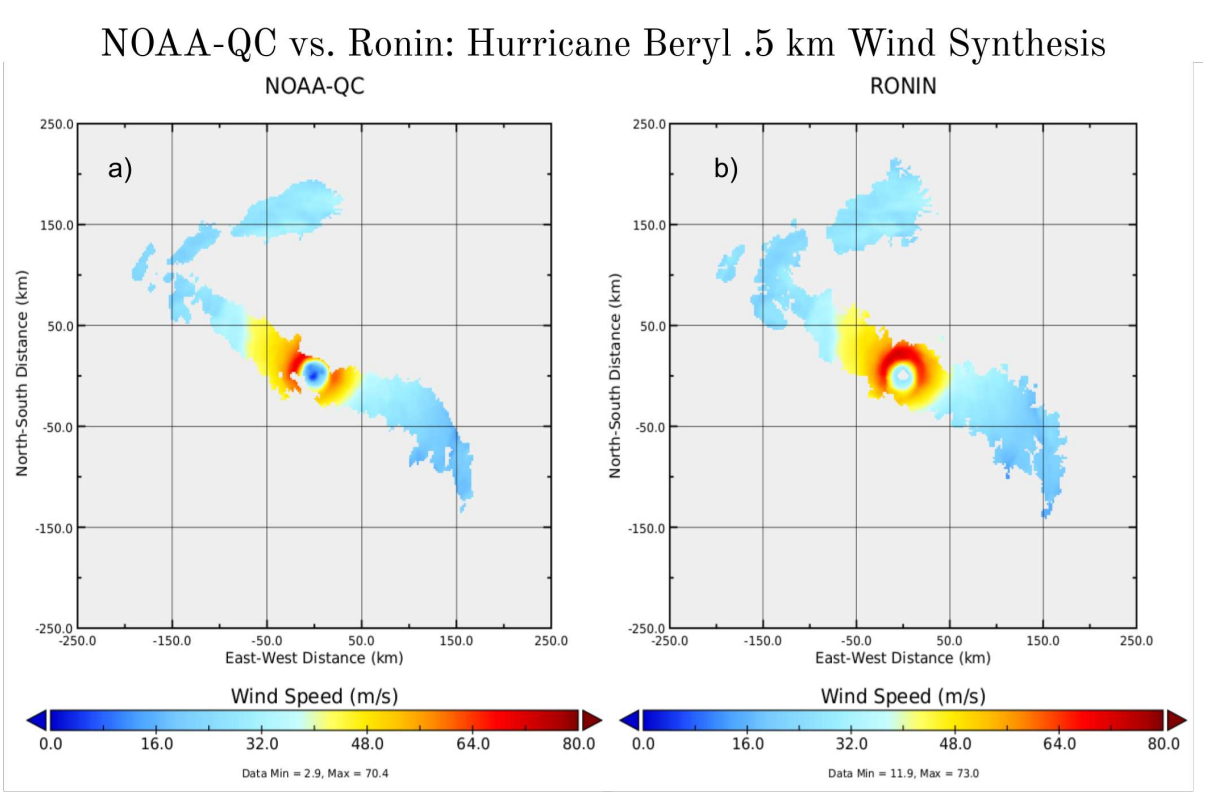

Figure 2.6: .5 km elevation 2 Dimensional Wind Syntheses with data produced from (a) NOAA-QC and (b) Ronin. Raw data from NOAA TDR during a flight through Hurricane Beryl on July 2, 2024. Figure courtesy Paul Reasor

cyclone cases, Ronin is shown to achieve the removal of 98.6 % of nonweather data while still retaining 83.3% of weather echo, in comparison to 98.8% and 73.3% for the current operational algorithm, suggesting it may be an effective upgrade for contemporary quality control methods.

Several directions for future work remain. It would be beneficial to understand how RF models directly compare to other ML techniques for the QC problem - though powerful, it is possible that more complex techniques would increase successful classifications. A more detailed evaluation of the relative importance of input features in a multiple-pass paradigm would also be beneficial, as it is unknown how the results for the feature selection would be changed on a second or third pass of an iterative QC procedure. Finally, one facet of radar quality control largely neglected in the above discussion is the unfolding of Doppler velocities exceeding the Nyquist frequency of a given platform through dealiasing. This is a common challenge, especially in proximity to the high-impact weather that airborne radar observations are often associated with. To complete a fully automated QC pipeline, Ronin must also be able to accomplish this task. Veillette et al. (2023) found success in applying a Deep Neural Network (DNN) to dealias data from a variety of radar platforms (though none were airborne), suggesting a promising direction for next steps.

Chapter 3

Machine Learning Based Retrievals of Full Hydrometeor Size Distributions in an Energetics Framework

3.1 Introduction

Hydrometeor distributions are useful microphysical quantities that describe the concentration of particles at a given value of a particular characteristic in a unit volume. This characteristic can, in theory, be any arbitrary descriptive feature of a hydrometeor, but is frequently defined in practice across particle diameters, with the distributions then being referred to as Hydrometeor Size Distributions (HSDs) or Drop Size Distributions (DSDs) when the particles are liquid. Physically, HSDs are directly tied to impactful variables such as precipitation rate and hydrometeor kinetic energy (HKE) - indeed, a portion of the early literature on the topic was motivated by understanding the relationship between HSDs, HKE, and soil erosion. Different shapes of HSD have also been shown to be associated with separate microphysical regimes in precipitation processes (Yu et al., 2012). HSDs are of great importance in atmospheric modeling and simulation - many Numerical Weather Prediction (NWP) schemes explicitly forecast moments of these distributions, and so additionally benefit from increased HSD understanding. One particularly intriguing characteristic of observed HSDs is the phenomenon whereby normalization by one or more moments of the distribution itself collapses variability in such a manner that HSDs take on an "S" shape that is common across climatic regimes and precipitation types. The characteristic shape then facilitates a complete and succinct description of the HSD by only two or three variables. Fig. 3.1 gives a visual example of the idea of normalization. Panel (a) displays the raw number concentrations for nearly 3000 DSDs obtained by Bringi et al. (2020) using disdrometer observations from a variety of rainfall and climactic conditions. Immediately conspicuous is the spread in concentrations,

sometimes nearing 10 orders of magnitude. Panel (b) displays the same 3000 DSDs, except the raw number concentrations on the left are divided by some number concentration characteristic of each individual distribution, and the particle diameter also normalized by some characteristic diameter, leading to a significantly more clustered distribution. The form of these parameters is the subject of much debate, and is further treated later in the manuscript.

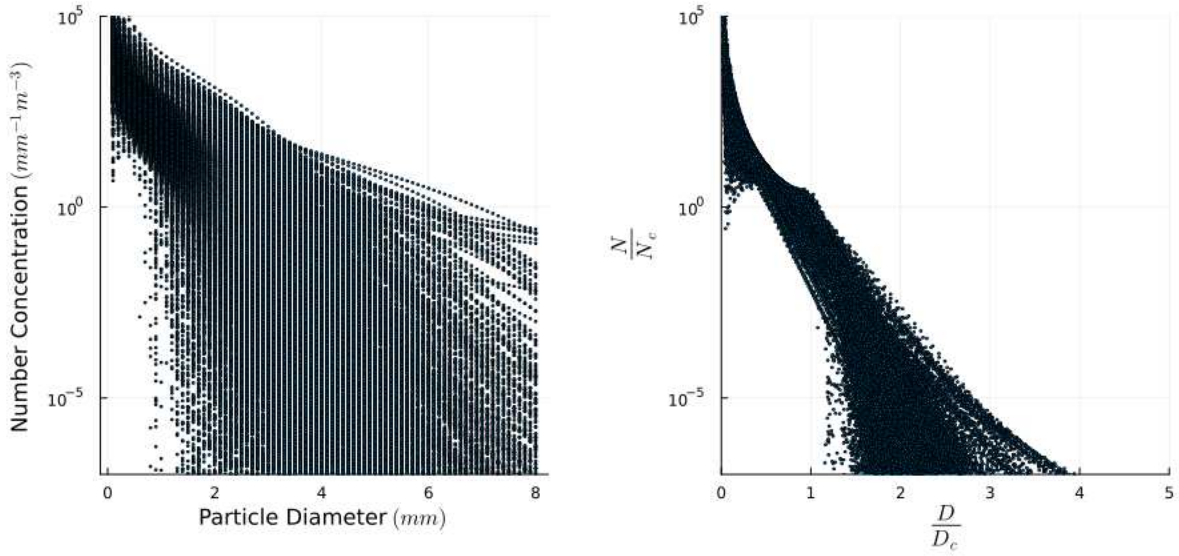

Figure 3.1: Left: scatterplot of number concentration of Temprian DSDs fit to DSDs from Bringi et al. (2020) Right: As in left panel, except D has now been normalized by some characteristic diameter D_c and N has been normalized by N_c, the forms of which is described herein.

HSD literature blossomed in the late 1800s, with the first filter-paper measurements being obtained by Weisner in Germany near the turn of the century. More focused investigation began around the time of the second world war, with Laws and Parsons (1943) taking a series of detailed measurements using the filter paper method in an effort to help design rainfall simulations aimed at investigating impacts on soil erosive and absorptive properties. Shortly following this, Marshall and Palmer (1948) (hereafter MP48), in their seminal work, fit size distributions as a function of an integral moment of the distribution (in this case, rainfall rate) for the first time. Even at this early stage, consistent similarities were identified between spectra. Gunn and Marshall (1958) further noted the applicability of the relationships from MP48 to snow size distributions. The

concept of "normalizing" HSDs was introduced by Sekhon and Srivastava in a pair of papers from 1970 and 1971 (Sekhon and Srivastava (1970), Sekhon and Srivastava (1971)). In a more detailed treatment of the MP48 fit, they suggest normalizing the diameter by the median volume diameter, and the number concentration by a factor related to the liquid water content. They further note the "remarkable fact... that the form of the size distribution is not changed" even when impacted by measurement errors. Torres et al. (1994) richly developed the theory of moment normalizations by showing that the Sekhon and Srivastava (1970) form as well as others (Marshall and Palmer (1948), Ulbrich (1983)) could be represented by a function of the particle diameter and some general integral rainfall variable while further noting the somewhat convenient microphysical interpretations this form could provide. Morrison et al. (2019) developed an approach that can use n moments, but also noted that 3 moments were generally fully sufficient to constrain variability in the non dimensional function. However, the physical meaning of the normalizing parameters in cases of 3 or more moments can become difficult to interpret. The studies discussed above suggest that HSDs observed in nature may be mathematically represented as some non-dimensional shape function and a combination of one or more integral moments of the dimensional distribution. The precise form of the shape function as well as the combination of moments that optimally reduce variability is still an open question. In general, increasing the number of moments utilized in the normalization helps further capture variability in the normalizing parameters, but can frequently result in more obfuscated physical interpretations.

Given their aforementioned utility in diagnosing microphysical processes and especially rainfall rates, much attention has also been given to the task of retrieving HSDs using observations from dual-polarimetric radar. Most modern methods such as those described in Raupach and Berne (2017) (hereafter RB17), and Bringi et al. (2020) (hereafter B20), attempt to retrieve a pair of moments, frequently M3 and M6, using power law relations or other fitting techniques from observed radar data. These moments are generally selected due to their strong relationships with radar variables - as mentioned above, M6 is directly proportional to the radar reflectivity Z in the Rayleigh scattering regime, and M3 is well-constrained by dual-polarimetric parameters such as Z_{DR}. Mo-

ments are subsequently utilized in concert with a theorized universal underlying shape function, henceforth referred to as h(x), to further derive the complete HSD and moments of other orders. These techniques have been met with some success, with B20 showing good ability to retrieve unbiased lower order moments from radar data. The reader is referred to their manuscript for a more extensive review of retrieval history along with an excellent discussion of the minutiae of validating ground-based observations relative to radar gates.

This manuscript aims to build upon prior work through the construction of a retrieval algorithm for the full size distribution from dual-polarimetric radar measurements using a pair of new approaches. Firstly, a novel 3 moment normalization described in Bell et al. (2025) (henceforth B25) is applied. The B25 paradigm is posited to provide a both effective and physically interpretable framework. Secondly, Machine Learning (ML) techniques are employed to retrieve the aforementioned normalizing parameters instead of the empirical relationships frequently applied in prior literature

The remainder of the chapter is organized as follows. Section 2 describes the energetics framework and its application in the creation of a synthetic training dataset, as well as the evaluation of the training dataset against select observations. Section 3 describes the utilization of the synthetic data data to train a machine learning model to retrieve the full hydrometeor size distribution. Section 4 presents a detailed evaluation of the trained model on a real-world datasets, and section 5 concludes with a discussion of the results and possible directions for future work.

3.2 Data and Methods

3.2.1 Dataset and Retrieval Setup in an Energetics Framework

An enduring challenges in the study of HSDs is the paucity of in-situ datasets describing the full distribution in varied conditions. Though surface based disdrometer observations of rain HSDs are common, measurements in mixed-phase or ice regions of clouds are much rarer, primarily due to the difficulty of placing observing platforms in their midst. Furthermore, it is often necessary to employ two instruments (commonly a Meteorological Particle Spectrometer/MPS and 2-

Dimensional Video Disdrometer/2DVD) in a calculated manner to obtain accurate measurements at extrema of the diameters. Droplet sizes less than about 1 mm are generally especially difficult to accurately retrieve with typical instrumentation (Bringi et al., 2020), an issue compounded by the fact that many HSD characteristics, especially their shape, are somewhat sensitive to lower order moments most impacted by the small end of the distribution. High variance on the small end of the distribution is in keeping with the literature, with much early works often highlighting the difficulty of fitting droplet sizes below 1 mm to the theorized exponential or gamma distributions (Sekhon and Srivastava (1970), Marshall and Palmer (1948)). As such, it is a very difficult task to obtain a large body of measurements that could be proven to be truly accurate representations of size distributions.

In an effort to reduce the dependence of any retrieval algorithm on a specific set of measurements, all with some of the challenges described above, it was decided to produce a large synthetic library of HSDs that were constrained by observational datasets. A synthetic approach ensures that errors in HSD quantification introduced by measurement limitations do not have an outsized impact on the actual retrieval, as the library is aimed at encompassing a large portion of all physically conceivable HSDs. Manufactured data is also posited to produce a more generalizable model applicable in a broader array of situations, as a model trained on such a dataset will have been exposed to a wider collection of possible scenarios then any single collection of in-situ measurements could be expected to represent.

The first step in creating such a library is defining a mathematical representation of the full size distribution. This study elects to represent HSDs using the energetics framework defined in B25. The reader is referred to that manuscript for the full details, but the more salient points are outlined below. B25 applies a three moment normalization, defining a characteristic number concentration N_c and characteristic diameter D_c as below:

$$N_c = \frac{M_n}{D_c^{n+1}} = \frac{M_0}{D_c}\langle \hat{D}^n \rangle \tag{3.1}$$

$$D_c = \frac{M^3 M^6}{M_4^2} \tag{3.2}$$

$$\mathcal{T} = T_H \frac{M_4^2}{M_3} = T_H \frac{M_6}{D_c} \tag{3.3}$$

$$N(D) = \frac{N_c}{\langle \hat{D}^n \rangle} (\hat{\tau} \hat{D}^{\alpha\beta-1} exp[-\eta \hat{D}^{\beta}]) \tag{3.4}$$

Equations 3.1-3.4 Definition of Parameters of B25 form of HSD. M_n is the n-th moment of the HSD, D_c is the characteristic diameter, N_c the characteristic number concentration, $\hat{D}$ the normalized diameter $\frac{D}{D_c}$, $\langle \hat{D}^n \rangle$ is the non-dimensional moment of order n, $\mathcal{T}$ is the "Temprian", and $\hat{\tau}$ an integral constraint on the non-dimensional shape function to ensure it meets the definition of a PDF. α, β, and η are parameters that define the form of the underlying shape function.

Moments three, four, and six are employed as a combination that attempts to strike a balance between selecting orders that well constrain the HSD and the amount of information available from dual-polarimetric radar data. Though using lower order moments in the normalization would likely result in further reduction in normalized HSD spread, it is well-known that radar observations are much more weakly related compared to those of higher order.

The total number concentration $N(D)$ in the B25 formulation may be obtained as in 3.4. As such, any retrieval algorithm must eventually produce values for D_c and N_c, but there are multiple avenues to this final destination. Broadly, in all scenarios, one or more integral moments or some combination of integral moments must be directly retrieved, in this case using machine learning instead of the numerical relationships applied in prior work. Then, these integral moments may be combined with knowledge of the non-dimensional distribution to analytically obtain other moments/characteristic parameters. These properties yield a variety of possible methods for retrieving a full HSD from radar data. We focus on four of these methods, further described below.

Retrieval of $\mathcal{T}$

In addition to D_c and N_c, B25 introduces a parameter coined the “Temprian” ($\mathcal{T}$) - the ratio $\frac{M_4^2}{M_3}$ - equivalently, the radar reflectivity M_6 divided by D_c (see 3.3). The reader is once again recommended to B25 for the finer points, but here it suffices to say that the $\mathcal{T}$ is directly related to

a number of different energetic characterizations of a volume of hydrometeors, and so is thought to be a useful representation of the extensive energy of any such population. It follows trivially that knowledge of $\mathcal{T}$ and sixth moment (radar reflectivity) results in D_c. Subsequently, using D_c in conjunction with a hydrometeor equation of state (derivation omitted), M_3 and thus N_c may be deduced. One advantage of the $\mathcal{T}$ retrieval approach is the obvious direct measurement of M_6 from radar, thus eliminating one source of variability.

Retrieval of M_0 or M_3

A second implication of the above prescriptions of characteristic parameters is that a moment of order n may be retrieved using the known non dimensional shape function and one of either M_0 or M_3. In the case that M_3 is known, M_0 is obtained in this fashion, and then all other moments may be derived. Reversing the technique, it is possible to use the zero order moment / total number concentration along with the non-dimensional shape function to retrieve M_3, which can in turn be used to define all other moments in the sequence and thus characteristic parameters. Prior discussion suggests that utilizing M_0 would likely aid in the reduction of variability in the normalized form since it is a low-order moment. However, work has also shown that accurate observations of M_0 are exceedingly difficult to obtain (Bringi et al., 2020) resulting in inaccuracies in downstream normalizations, indicating possible challenges with this approach.

Direct Retrieval of Characteristic Parameters

Of course, it is also a possibility to attempt to simply retrieve both of D_c and N_c from an ANN directly. Direct retrieval would have the added benefit of increased simplicity, but does not necessarily have to adhere to the mathematically consistent relationships prescribed by the B25 formulation, and so it is possible a D_c could be obtained that is inconsistent with the retrieved N_c

All of the 4 above methods are associated individual challenges and strengths - it is not readily apparent as to which will prove the most effective, or indeed if one will even prove markedly better than others. In an attempt to understand which is most accurate, this work will proceed by

creating models for all four of the techniques and endeavor to determine real-world advantages and drawbacks of each when evaluating against observed data.

3.2.2 Generation of Synthetic Training Dataset

Now that a mathematical form has been defined, it remains to determine the range that the component variables may take on in order to create a synthetic dataset. Though the use of a synthetic dataset is aimed specifically at removing direct dependence on measurements, it remains of interest to bound realizations to some range of parameters - lest retrievals be trained on nonphysical distributions and produce erroneous results. Observed parameters, such as M_0 and the characteristic diameter D_c have direct physical meaning, and so it is possible some constraints could be applied ad hoc. However, given the aforementioned differences in normalized HSD formulations, there exists little guidance in the literature as to the ranges the shape parameters α and β should take on. In order to determine the values, then, it is beneficial to examine some high-quality observational datasets and derive the values in the B25 form that must be bounded. The bounding process was undertaken as follows.

For each HSD in the dataset, the parameters of the B20 form of the HSD were "inverted" to yield the original size distribution. Subsequently, the characteristic parameters of B25 (D_c, N_c, $\mathcal{T}$) along with moments of orders 0-6 were calculated for the original HSD. It then remained to determine the optimal values of the fitting parameters α, β, and τ for each distribution. In the case that the theoretical normalized distribution is truly universal, all HSD variability would be represented in D_c and N_c, and the shape function parameters would take on only one value. In practice, some variance is observed in α and β, and though it is possible fluctuations in the shape parameters could be a result of sampling error, mathematical deficiencies, or some combination of the pair, it nonetheless holds that any retrieval algorithm should be exposed to a range of values. In a practical sense, values of both α and β will have to be selected when an actual retrieval is performed, but for now it is desirable to continue a range of possible parameters to improve generalization capabilities of the model. A brief aside - cursory observations during early fits revealed

that the η parameter was a strong function of beta, and so a secondary fitting was undertaken in order to retain a power relationship parameterizing η as a function of beta. From henceforth, we proceed only with discussion of α and β as the shape parameters.

Given a non dimensional shape function specified by D_c, α, and β a unique N_c is obtained - meaning it cannot be specified apriori. The characteristic diameter is thus not enough to solely identify the dimensional part of a DSD - further knowledge is needed. It can be shown that, given our definition of N_c using the 3rd moment, a prescribed moment of order 0 or 3 can be leveraged to produce a consistent N_c (details in appendix). We further elect to use M_3 due to its closer proximity to moments radar is sensitive to. Thus, for each B20 HSD, D_c and M_3 are passed to a Nelder-Mead solver that attempted to minimize the squared difference between the logarithm of the concentrations at each "bin" size in the full distribution by varying α and β. In this scenario, the full range of bins was defined as .1 mm bins from .1mm to 8.0 mm - aimed at identifying the full spectrum of rain hydrometeors. As in prior literature (Bringi et al., 2020), much variance was noted on the small end of the size distribution (<1mm droplets), and so a penalty term was added that scaled the loss by $\frac{1}{D}$ in order to encourage better fits here. The solver required initial values of both the alpha and beta parameter, and experimentation revealed fits were somewhat sensitive to these values. To lessen the impact of initial parameters on results, each fit was undertaken 50 times with a new random set of initial conditions, and the result with the lowest value of the loss function was retained as the "final" fit.

Once completed, however, it was determined that the methodology of simply fitting α and β to a prescribed D_c/M_3 produced fits that were not "self-consistent" - that is, the D_c returned by actually computing the moments of the distribution specified by the fit (prescribed D_c, M_3, fitted α, β) was not the same as the specified D_c. Non-self consistent HSDs would lead to inconsistencies with downstream retrievals, and so a secondary fit was undertaken that re-fit α and β constrained to a self-consistent DSD, using the original fit for the HSD as initial conditions. Qualitatively, the secondary fit worked well, generally resulting in only slight modifications to α and β. Figure 3.2

shows the results of the fitting process from 3 DSDs selected from the input dataset - one at a low rain rate ($.34\ \frac{mm}{hr}$), one at a medium rain rate ($5.42\ \frac{mm}{hr}$), and one at a high rain rate ($33\ \frac{mm}{hr}$).

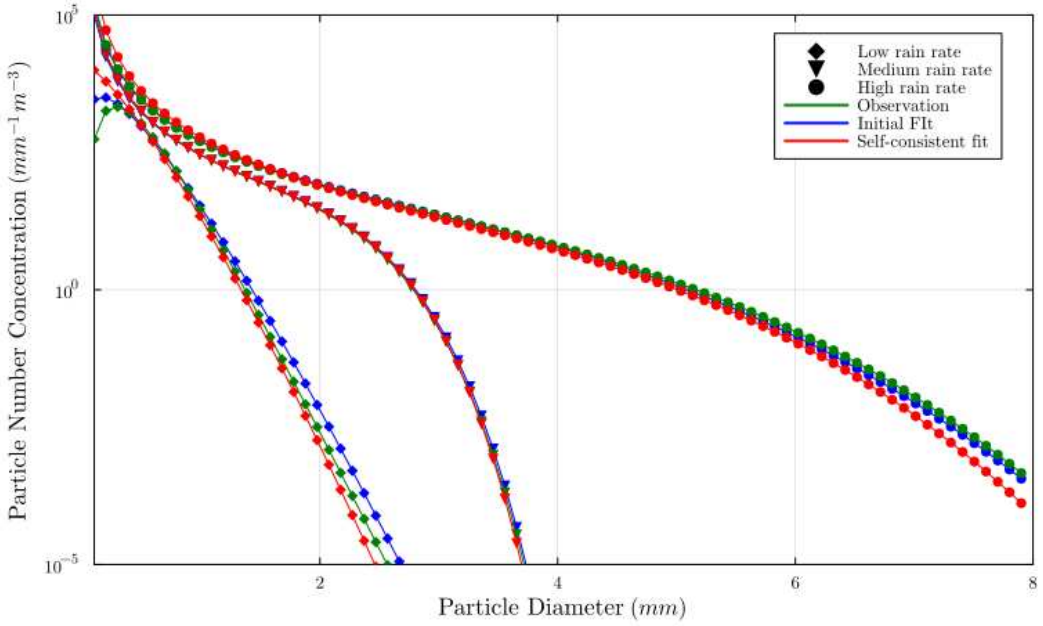

Figure 3.2: Resampled B20 DSDs (green), Fitted DSDs (blue), and self-consistent fitted DSDs (red) for a low, medium, and high rain-rate case from the B20 dataset

It is important to note that the B20 distributions contain their own set of assumptions, one primary one being their fit to a different, but still prescribed form of HSD, and as such resampling does not necessarily exactly represent the observed concentrations for a given scenario. Therefore, is not the goal of our fits to match precisely, but rather to represent the full range of parameters observed in nature.

To evaluate whether or not we were successful in replicating the span of natural parameters, we turn to the distribution of HSD moments across the dataset. Figure 3.7 displays the distribution of values of the first five moments (M_0 through M_4) and M_6 (included due to its obvious import in radar meteorology) of the fitted distributions. In a general sense, the moments match less closely with increasing order distance from the prescribed moment (M_3), with M_3 matching exactly. All distributions appear to be approximately lognormal, with peaks of increasing magnitude as order

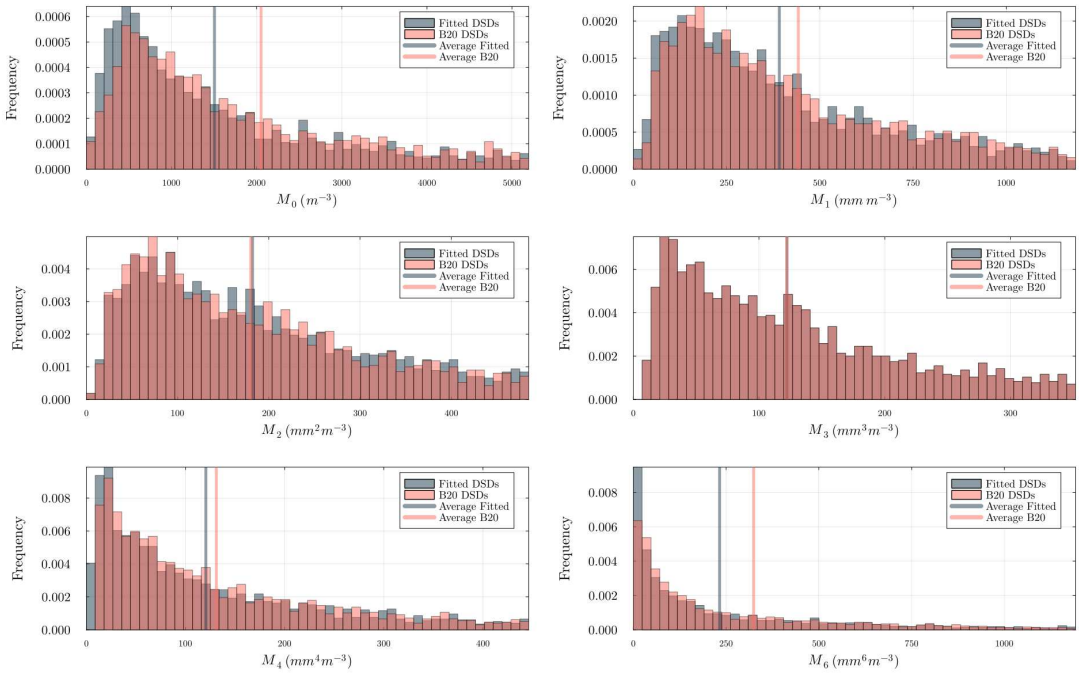

Figure 3.3: Distribution of moments for HSDS sampled by B20 (red) versus those fit to B20 using the mathematical form described in B25 (gray). Distributions truncated below 75th percentile in order to eliminate extreme tails.

increases. The lognormal distribution is commonly employed in precipitation studies (Cho et al., 2004), and so its presence here appears to be sensible at face value. The close adherence of the fits to the "observed" distributions provide increased confidence in both the validity of the mathematical representation employed in this work as well as a starting point for informing a fully synthetic dataset.

Another important consideration is that since all of D_c, M_3, α, β, are continuous variables that can take on an infinite number of real values, some discretization must be performed in order to create a tenable scenario for dataset generation. We elect to inform this process using observations as follows. A two-dimensional histogram is created for both the D_c/M_3 and the α/β space from the fitted HSDs. The edges of each bin in the histogram are obtained, and n random values of the parameter between the edges are obtained, where n is the frequency of occurrence in the specified bin. With 2928 fitted HSDs, this leads to the same number of each unique D_c/M_3 pairings and unique alpha/beta pairings. The supposition that D_c/M_3 and α/β are uncorrelated means that any

of the D_c/M_3 combinations could conceivably occur with any of the α/β pairings. As such, the synthetic dataset will contain 2928*2928 = 8573184 unique HSDs.

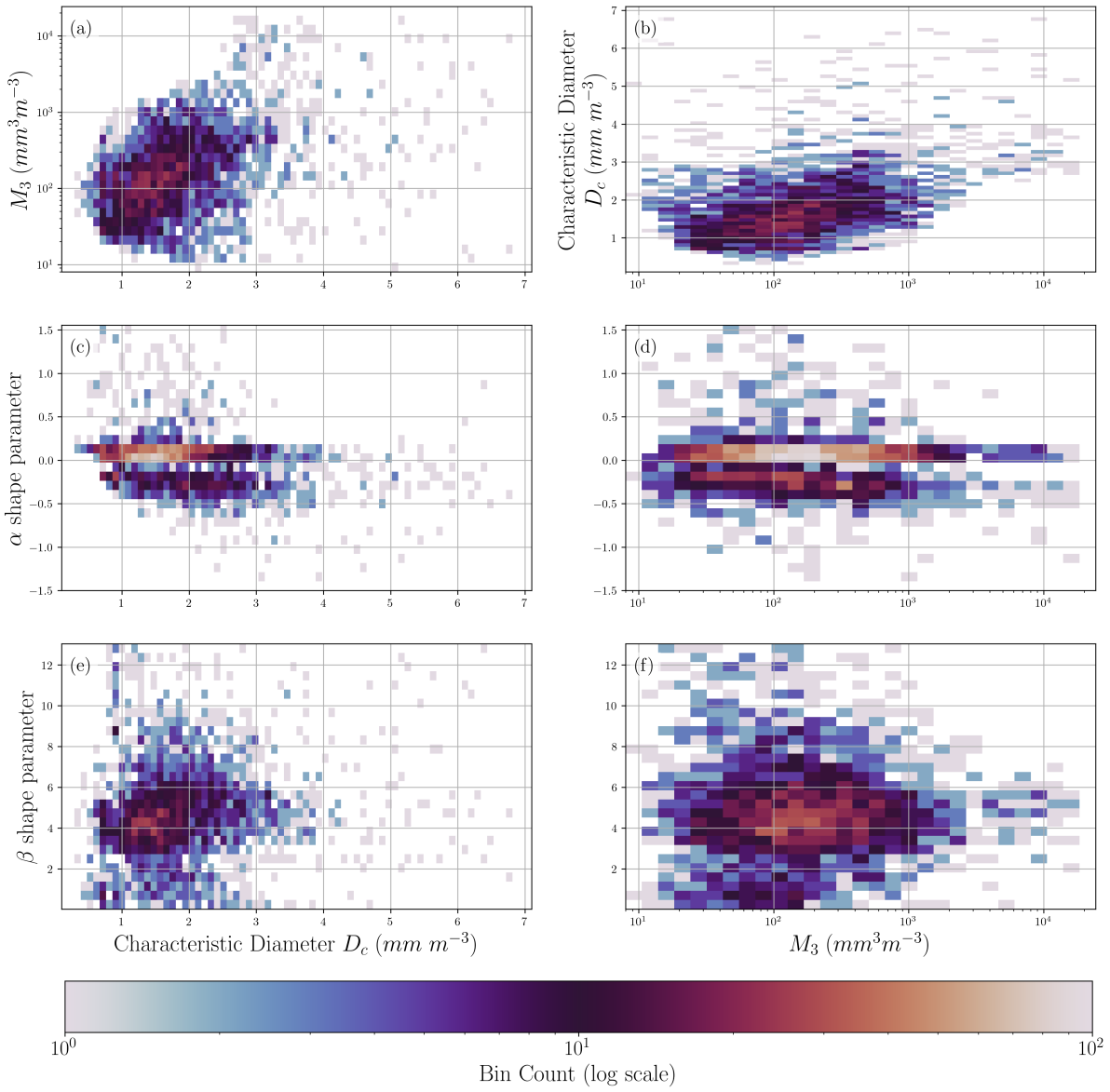

Figure 3.4: Distribution of Characteristic Diameter D_c versus a) M_3, c) α, e) β and Prescribed M_3 versus b) D_c, d) α, and f) β across the fitted dataset.

Figure 3.4 shows the distribution of Dc and M_3 versus α, β, and each other for the "self consistent" versions of the DSDs fitted to the B20 dataset. The resampled version of these histograms (not shown) appears the exact same, except for a factor of about 2500 more HSDs in each bin.

The self-consistent fitted α has a median value of $\alpha = .0728$ and the associated self consistent β of $\beta = 4.655$. These values will be used as the so-called "climatological" values for the remainder of the work. The distribution in 3.4 is also highly peaked around these values, lending credence to notion of the presence of an underlying universal normalized distribution.

Now that the unique size distributions have been selected, the necessity remains to calculate the radar variables associated with each. The T-Matrix code (Mishchenko et al., 1996), a widely utilized program for computing backscattered radiation for a particle of an arbitrary diameter, was selected for this endeavor. The wavelength of interest for this study was selected to be 5.31 cm, that of the NSF SEA-POL radar (Rutledge et al., 2019). Both backscatter and forward scatter for particles of each diameter in the discretized HSD (0.1 -> 8mm, .1mm increments) were obtained for both horizontal and vertically polarized incident radiation and subsequently multiplied by the concentrations defined by the HSD, giving the horizontal and vertical reflectivities. Knowledge of the reflectivity in both polarizations then allowed for the calculation of differential reflectivity (ZDR) and specific differential phase (KDP) following Jung et al. (2010). Finally, the radar reflectivity (Z) is calculated as the sixth moment of the HSD, yielding a full suite of radar moments for each DSD.

Another key advantage provided by using ML to retrieve reference moments becomes apparent here - transitioning to other radar frequencies is much simpler than in other approaches as it does not require a direct derivation of relationships between radar variables at different wavelengths and the reference moments. Scattering code may simply be used to understand radiative properties, with ML forming mappings between these properties and reference moments.

Figure 3.5 shows the resultant distribution of radar variables for the synthetic dataset. A cursory examination reveals that D_c is positively correlated with dBZ and K_{DP}, and especially strongly correlated with Z_{DR}. D_c in the B25 framework, though not exactly representative of the mean volume diameter D_m or other common forms, is still related to the average diameter of hydrometeors in a volume. Increases in K_{DP}, dBZ, and especially Z_{DR}, then, would be expected with increasing D_c. Z_{DR} begins to decrease above a $D_c \approx 4.5\ mm$, likely due to a transition to the Mie

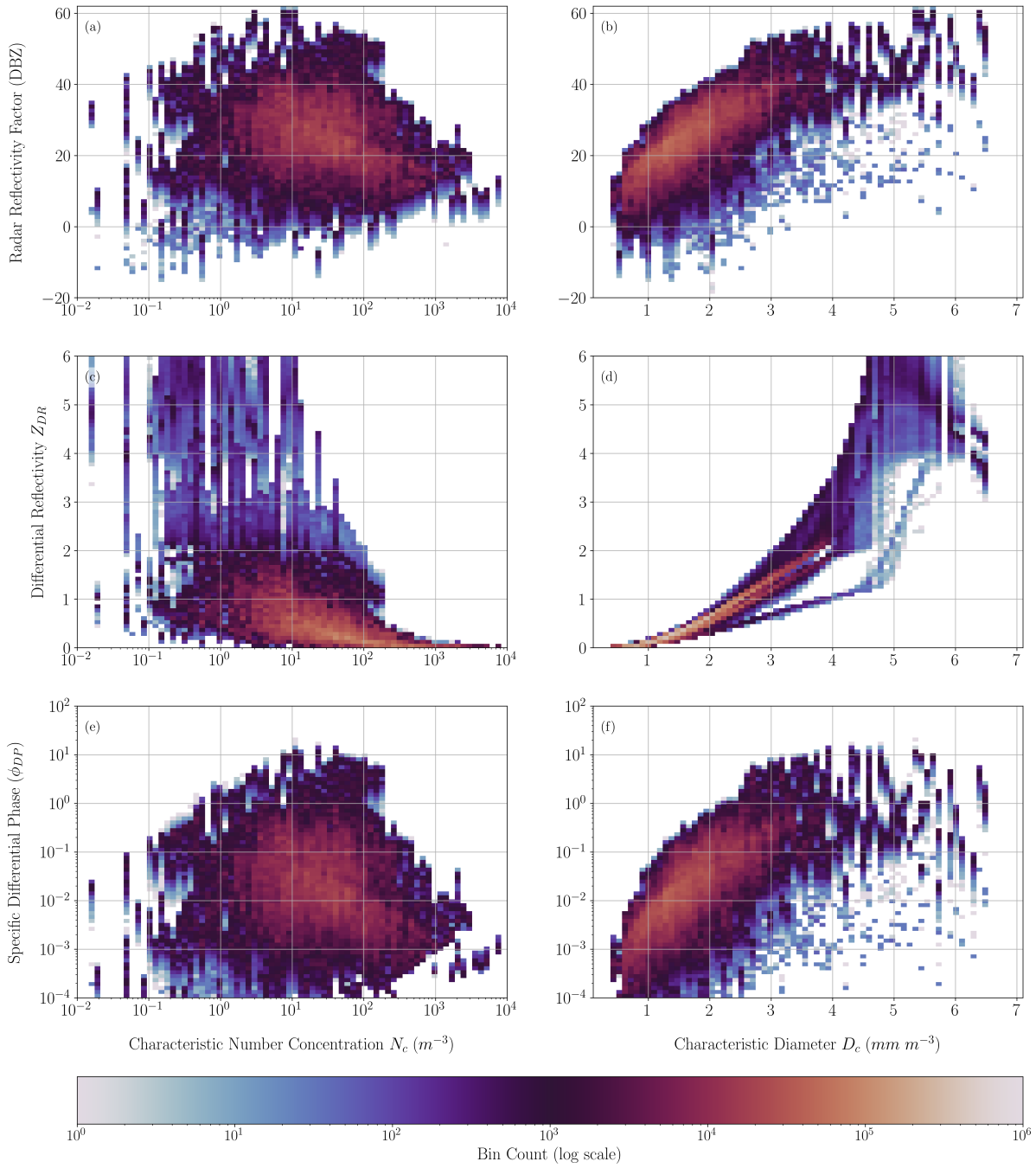

Figure 3.5: Distribution of radar variables versus characteristic parameters D_c and N_c in the synthetic dataset. For computational purposes, every $1000th$ point was selected to be plotted. dBZ and Z_{DR} presented in logarithmic units, K_{DP} and N_c on a logarithmic axis.

scattering regime at C band. N_c appears to be more weakly related to K_{DP} and dBZ, though still positively correlated. In general, the radar observations appear to be consistent with the expected physical interpretations of the characteristic parameters.

3.3 Training of Artificial Neural Net Retrieval

Artificial Neural Networks (ANNs) are a type of ML architecture that function by propagating a set of input features through a series of so-called "neurons" - nodes that apply a weight, bias, and activation function to values, eventually arriving at one or more predictands. They are a powerful technique that has been widely adopted in a multitude of scientific fields. In meteorology, some of the many tasks ANNs have been applied to include wildfire forecasting (Liang et al., 2019), reconstructing sub-grid scale tornado vortices (Zhang et al., 2023), tropical cyclone wind fields (DesRosiers et al., 2025), and even global-scale numerical modeling (Keisler, 2022). They are ideal for the retrieval problem due to their relative flexibility and simplicity, especially given the low dimensionality (mapping 3 to 5 input radar variables to 1 to 2 output characteristic parameters).

In order to evaluate the differing retrieval techniques, four models are trained as per 3.2.1 - 3.2.1. Details are contained in 3.1 All have input features of DBZ, Z_{DR}, and K_{DP}, and have basic configurations of 2 hidden layers composed of 64 nodes each. The Leaky RELU (Maas et al., 2013) function is selected for activation in an attempt to better treat negative inputs in DBZ and avoid the vanishing gradient problem (Hochreiter, 1998). Generally it is beneficial in the training of an ANN to apply a standardization of some sort to input features such that features of large absolute magnitude do not spuriously dominate the network, but some experimentation revealed it not to be necessary in this case. However, in the case of several output variables, it was beneficial to apply logarithms to decrease the variance of the retrieval.

For model validation, training, and testing, the input synthetic dataset is randomly split according to the canonical 72%/8%/20% division. With this split applied, the respective datasets then comprise of approximately 5.7 million, 630 thousand, and 1.6 million HSDS. For verification purposes, plots as in 3.5 were created for each dataset to ensure similar distributions of characteristic

and radar variables (not shown). Hyperparameter tuning was performed on the validation dataset (not shown), and a model architecture with two hidden layers and Batch Normalization on input was selected. In order to combat overfitting given the expansive input dataset, a L1 Regularization penalty was applied to a mean squared error loss function with a λ of $.001$. Table 3.1 contains details about the predictors, predictands, loss functions, and learning rates of the four models trained.

Table 3.1: Summary of ANN Configurations

Model Name	Input Features	Output Features	Activation Function	Learning Rate
Temprian Retrieval	(DBZ, KDP, ZDR)	$log_{10}(\mathcal{T})$	ReLU	.001
N_c Direct Retrieval	(DBZ, KDP, ZDR)	D_c, $log_{10}N_c$	ReLU	.001
M_0 Retrieval	(DBZ, KDP, ZDR)	D_c, $log_{10}(M_0)$	ReLU	001
M_3 Retrieval	(DBZ, KDP, ZDR)	D_c, $log_{10}(M_3)$	ReLU	.001

Figure 3.6 displays the results of applying the 4 trained models to the testing dataset and retrieving N_c (and in the $\mathcal{T}$ model's case, D_c) through the various mathematical relationships. D_c is successfully retrieved through an inverse operation with the Temprian ($\mathcal{T}$ model), retrieval of moments of other orders (M_0 model, M_3 model) as well as through direct prediction by a NN (Direct model), with all four models distributing tight clustering along the 1:1 line and Pearson Correlation Coefficient values exceeding .98.

As was expected, obtaining N_c proves to be a somewhat more challenging endeavor, likely due both to decreased radar variable sensitivity to M_3 and the factor of D_c^4 in its denominator. The reasoning for the dramatic under prediction of N_c in the M_0 retrieval is not entirely clear. The performance of the M_3 model suggests validity in the mathematical formulation of this approach, so it is possible that there is a deficiency in training data for the M_0 version. Given the challenges already present with retrieving M_0 from radar data, we will proceed by analyzing the 3 other models.

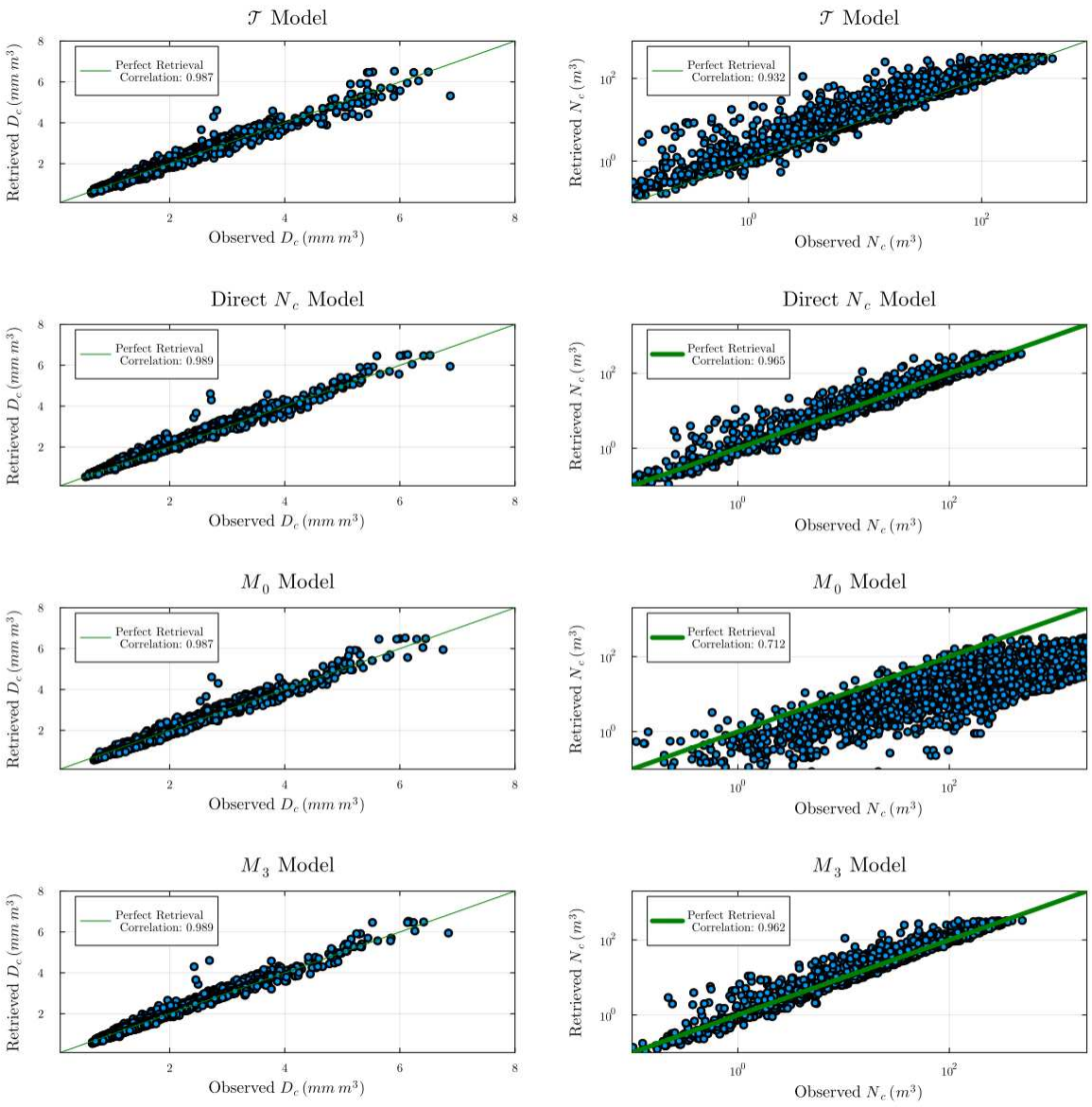

Figure 3.6: Evaluation of retrieval of D_c and N_c from NNs on testing dataset. (a) and (b) show the resultant D_c and N_c from the $\mathcal{T}$ model. (c) and (d) are as in (a)/(b), but for the direct D_c/N_c retrieval model. (e) and (f) as in (a)/(b), but for D_c/M_0 retrieval model. (g) and (h) as in (a)/(b), but for D_c/M_3 retrieval model.

3.4 Evaluation of Retrieval Techniques

3.4.1 Prediction of Rainfall Extremes Campaign In the Pacific (PRECIP)

The PRECIP Field Campaign took place in Taiwan from late May to early August 2022, coincident with both the beginning of tropical cyclone season and heightened Mei-Yu front activity. Aimed at diagnosing microphysical, thermodynamic, and kinematic processes associated with high-impact extreme precipitation events, the program involved the deployment of multiple mobile radar installations, LIDARs, ground-based disdrometers, and rawinsondes. The collocation of research-quality radars and disdrometer observations in heavy precipitation makes it an ideal candidate for evaluation of the above retrieval technique. Extended details on the instrumentation are provided in following sections.

The primary hydrometeor-measuring instrument deployed during PRECIP was a laser disdrometer known as a Parsivel (an apt contraction of "particle size and velocity"). This class of disdrometer derives particle diameter and fall velocity by measuring the size and length of shadows cast on the laser by incident hydrometeors. An important and widely known caveat of this type of instrument is difficulty obtaining accurate droplet concentrations on the extreme ends of the diameter spectrum. Raupach and Berne (2015) described a method to correct Parsivel observations using a 2-Dimensional Video Disdrometer (2DVD) as a reference instrument, and were able to greatly reduce bias in the measurements. Unfortunately, no 2DVDs were immediately available in proximity for calibration of the PRECIP disdrometers. As such, analysis will proceed with the forewarning that the Parsivel observations may primarily only be useful in determining values of the larger (M3, M4, M5, M6) order moments, which are not as impacted by concentrations on the tails of the distribution as those of lower order. In the future, it would be of interest to compare performance of this technique with other types of disdrometers, especially collocated ones, but these datasets are unfortunately sparse at the time of writing.

3.4.2 Evaluation on PARSIVEL data

Parsivel observations were obtained for the duration of the project, and time frames of continuous precipitation were identified. By far the longest consecutive period of rain took place between about 1:00 UTC on June 8, 2022 and 8:00 on the same date, and so this timeframe is selected for deeper analysis.

The disdrometer sampling period is every 30 seconds, but in an effort to combat noise and again following B20, 3-minute average spectra are obtained for all times between 1:00 UTC and 8:00 UTC on the 8th. Bulk variables provided with the DSD observations (moments, radar parameters) are also merged into a 3 minute average.

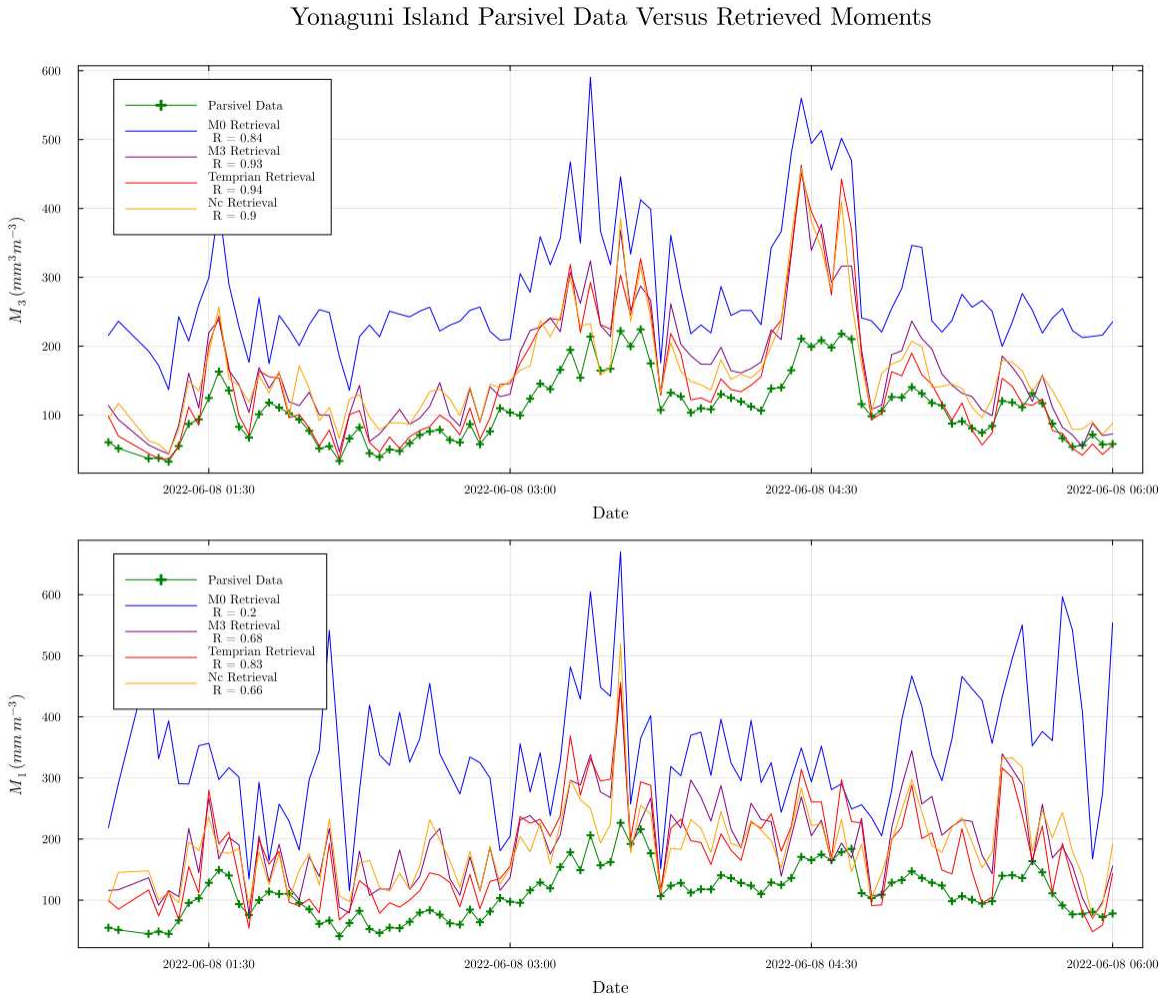

Figure 3.7: Time series of M_1 and M_3 for Parsivel observations and various flavors of neural network retrievals from disdrometer-derived dual-polarimetric radar data during June 8th, 2022 precipitation event.

Fig 3.7 contains the results of applying all 4 trained models on the aforementioned extended heavy precipitation event on June 8th. DBZ, K_{DP}, and Z_{DR} were obtained from the pre-calculated values provided in the Parsivel data for each 3 minute average period, and input into the models. The associated retrieval techniques described in 3.2.1 - 3.2.1 were then employed to obtain N_c and D_c, and finally a HSD. 3.7 contains the $1st$ and $3rd$ moments derived from numerical integration of the resultant HSD. In order to combat the effects of the somewhat truncated distributions provided by Parsivel, only the diameters with nonzero number concentrations in the correspondent observed HSD were considered for use - that is, if the Parsivel featured non-zero concentrations from $.5\ mm$ to $5\ mm$, only the portion of the retrieved HSDs between these bounds would be considered. It was decided to display moments 1 and 3 due to good performance on higher order moments.

The most immediately conspicuous aspect of 3.7 is the tendency for all retrieval algorithms to overestimate both moment orders - indeed, there are very few time steps where any retrieval algorithm has a value less than the observations. Given this as a universal tendency, it's possible that underestimation by the Parsivel is contributing to this trend. However, it's unlikely all error here may be attributed simply to measurements. The retrieval based off of a predicted M_0 is particularly afflicted, at all points being the highest prediction by a wide margin. Further investigation will need to be conducted in order to better ascertain the causes of this phenomena.

In a broad sense, the vast majority of the models are able to capture trends in the lower order moments, with the M_0 model's retrieval of M_1 the only data to feature a correlation below $.65$. As expected, M_3 proves to be a much easier retrieval for all models, with correlation values in the $.85 - 95$ range. That all models are able to achieve sensible retrievals of all moment orders throughout the course of greatly varied hydrometeor conditions further suggests validity in the normalization and synthetic dataset.

The evolution of the full HSD in time is also of microphysical importance for the reasons described above and, as such, also warrants a qualitative investigation. From about 4:39 UTC to 4:55 UTC, a convective core traverses the radar site with reflectivity values peaking near 35 dBZ at approximately 4:39, then falling to near 20 dBZ by 4:55. This provides an opportunity to examine

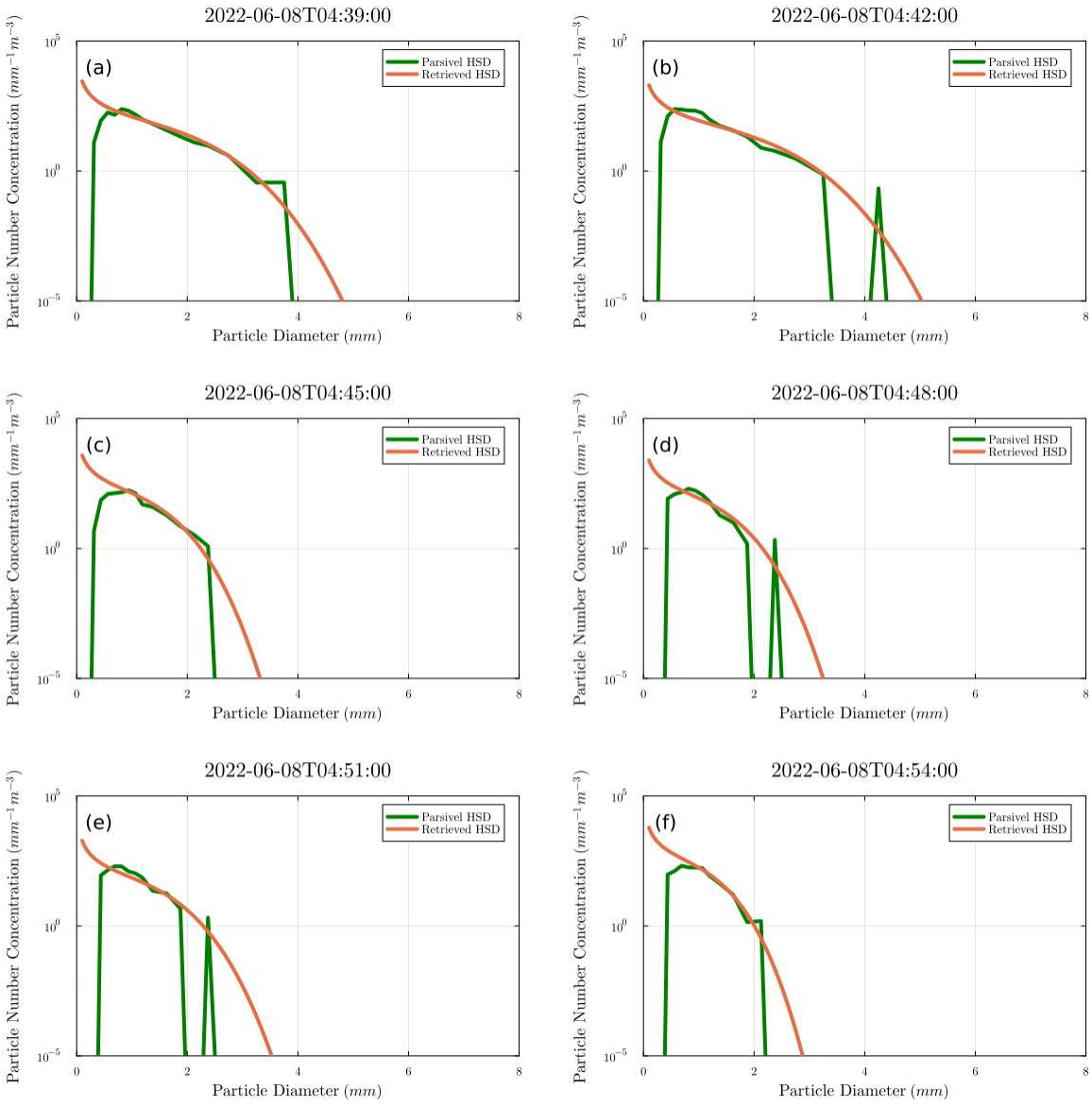

Figure 3.8: Evolution of observed 3-minute average and retrieved (using $\mathcal{T}$ model) DSDs for a period of heavy rain (32 dbZ) transitioning to light rain (20 dbZ) at Yonaguni Island between 4:39:00 and 4:54:00 on June 8th. Pearson Correlations included in annotation

how the neural net responds to a presumably wide variety of precipitation conditions in a short time. Figure 3.8 shows the evolution of the 3-minute Parsivel DSDs throughout the time period (green) compared to the retrieved (red). The Temprian retrieval model is selected for use here given its superior performance relative to the other methods throughout the period and its physical consistency.

The observed distributions begin by displaying a nearly perfectly exponential distribution, a straight line on the linear-log plot. As the intensity of precipitation decreases, the larger end of the HSD begins to taper off more and more quickly, eventually becoming a near vertical line at approximately $D = 2.5mm$. Remarkably, the concentration of droplets at or around $D = 1mm$ remains nearly steady throughout the period. Qualitatively, the neural net retrievals do an excellent job of showcasing this same structure. The small end of the distribution remains approximately constant in time, while the larger half becomes increasingly vertical. One deficiency exists in the seeming difficulty in emulating a pure exponential curve toward the beginning of the period - this could possibly be due to the prescribed β being nonzero. As a whole, the relative adherence to the observed shape, even given the aforementioned limitations associated with Parsivels, suggests that the algorithm is able to successfully give information about the evolution of the HSD through time.

3.4.3 Evaluation on Radar Data

As presented thus far, retrievals have only been evaluated using idealized radar variables derived from radiative scattering models for input. In an effort to understand the efficacy of the ML technique in the real world, we now discuss performance when using data directly from precipitation radars.

The Sea-Going Polarimetric (SEA-POL) radar (Rutledge et al., 2019) is a mobile C-band weather radar designed for deployment at both land and sea. It generally operates at a frequency of 5.65 GHz, facilitating a balance between maintaining a manageable physical size (4.3 meter antenna) and high resolution. SEA-POL features an excellent 1-degree beamwidth with a side-

lobe of about -27 dB, enabling very high quality observations. C-band radars are also of interest in precipitation microphysics studies, as some dual-pol parameters are especially sensitive at the wavelength (Rutledge et al., 2019). During PRECIP, SEA-POL was installed on the west side of Yonaguni Island, approximately 6.8 kilometers west of the Parsivel used in the study.

Radar observations from SEA-POL were employed to provide input to the retrieval model as follows. First, 81 degree azimuth (correspondent to the azimuth directly over the Yonaguni Island disdrometer) SEA-POL RHIs for the project were gridded to 250 meter resolution, with 471 points in the range dimension and 71 in the elevation. Subsequently, average values of reflectivity, differential reflectivity, and K_{DP} were calculated for a 5x3 grid point box (1250m range x 750m elevation) centered over the disdrometer (reflectivity and differential reflectivity were first converted to linear space for the average). Sweeps that took place a time where rain was occuring at the disdrometer (defined by the disdrometer measuring a nonzero M_6) were retained, and all others discarded. Additionally, sweeps that were missing values of ZDR or KDP at all gates in the aforementioned box were discarded. With these restrictions applied, 71 SEA-POL sweeps met the criteria for inclusion in the input dataset. Data from the sweeps are presented in panels a through c of Fig. 3.9. Overall, the box-average reflectivities (a) match well values observed by the disdrometer. Differential reflectivity also displays a clustering along the $1:1$ line, but some limitations of real-world data become apparent, with several data points exhibiting negative ZDR, something that cannot occur in rain. At large values, K_{DP} appears qualitatively similar between radar and disdrometer, but the Parsivel consistently returns values orders of magnitude below SEA-POL when its estimate is below about $10^{-1}\frac{deg}{km}$. The underestimation is likely a result of previously discussed disdrometer measurement limitations.

Despite these difficulties, it is still of interest to apply the ML retrieval to understand if it is a successful proof of concept. To do so, further truncation must be applied, removing data points with negative ZDR or K_{DP}, since these are not part of the training dataset. Afterwards, 39 datapoints meet all the above criteria. Each is used as input to the direct retrieval model (now selected due primarily to its increased accuracy) to retrieve the full HSD. Rain rate, M_3, and M_1

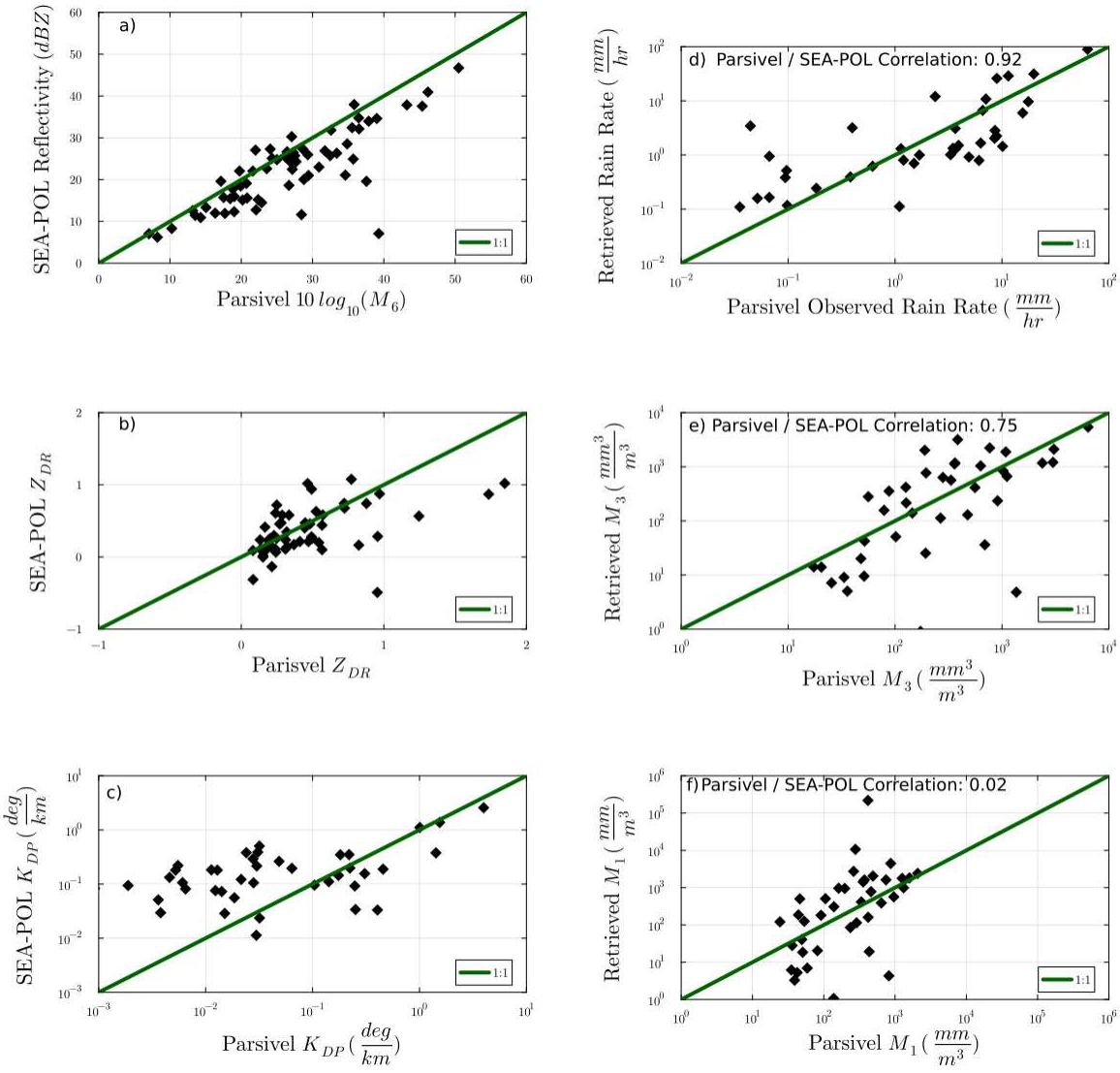

Figure 3.9: SEA-POL box-average radar observations and ML retrieved DSD moments versus parameters from Yonaguni Parsivel disdrometer $a)$ Parsivel reflectivity (obtained using Parsivel observed M_6) versus SEA-POL observed box-average radar reflectivity $b)$ Parsivel calculated differential reflectivity versus SEA-POL observed box-average differential reflectivity $c)$ Parsivel calculated K_{DP} versus SEA-POL observed box-average K_{DP} $d)$ Parsivel observed rain rate versus ML model retrieved Rain Rate ($\frac{mm}{hr}e)$ Parsivel observed M_3 versus ML model retrieved M_3 $f)$ Parsivel observed M_1 versus ML model retrieved M_1. ML Retrievals are annotated with Pearson correlation with Parsivel observations

are calculated from the HSD retrieval and compared with Parsivel observations in panels $d - f$ of 3.9.

Broadly, the ANN is successful in its retrieval of all three parameters. Rainfall rate, arguably the most important, is also the most accurate of the trio. The modeled values display a high Pearson correlation of $.92$ with the Parsivel observations, and many data points lie nearly exactly along the 1:1 "perfect retrieval" line, even across a wide range of rainfall rate magnitudes. More scatter is observed in the retrieval of M_3, but a Pearson correlation value of .75 suggests that the model is still informative. M_1 correlation is very low, with a Pearson value of $.02$, but it is likely that this is affected by outlier points. Points are still generally clustered around the 1:1 line, with no readily apparent bias or tendency for over/underestimation.

In an effort to evaluate the retrievals on a broader dataset, a similar process was undertaken for another high-quality research radar - the National Center for Atmospheric Research S-POL platform (NCAR S-POL) (Keeler et al., 2000). During PRECIP, S-POL was deployed on the west side of mainland Taiwan, near the city of Hsinchu. Another Parsivel disdrometer was located approximately 10.75 kilometers to its east at an azimuth of 85 degrees.

Scattering variables were recalculated for S-POL's wavelength as in 3.5, and a new model was trained to directly retrieve D_c and N_c as in 3.1. Gridded S-POL data was then obtained for a 48 hour period between midnight local time on May 26th, 2022 and midnight on May 28th, dates selected due to the near constant precipitation throughout. Averaging was performed in the same fashion as above, using a 5 point in range by 3 point in azimuth box in linear space for reflectivity and Z_{DR}.

Figure 3.10 gives the resultant time series of observed and Parsivel-derived radar variables for the period. Radar reflectivity once again matches the most closely of the three variables, with a correlation between S-POL box average values and Parsivel observations of .77 and a mean absolute error of around $5\ dBZ$. The Z_{DR} series is somewhat less successful, with the Parsivel continuously returning lower values than the radar. However, periods of agreement do exist, especially from 06:00-12:00 on the 26th, and from 18:00-midnight on the same date. K_{DP} once again proves

challenging, with a number of time periods featuring missing or negative values. However, when values are obtained and are positive, they match closely with the Parsivel.

Though the matchup of the radar variables with the Parsivel still leave something to be desired, the trained model is again applied to times when nonzero DBZ/ZDR/KDP is obtained, with the results presented in 3.11. The same three variables as Fig 3.9 are shown. Again, the retrievals are broadly accurate, generally capturing temporal trends displayed by the Parsivel. One new feature made apparent here is the tendency for consistent underestimation by the retrieval relative to the disdrometer. This could be attributed to a couple of different factors - it is possible that collision/coalescence processes are occurring below the view of the radar that are leading to intensification of the precipitation. Underestimation also appears to be most severe in areas of low K_{DP}, suggesting sensitivity to dual-polarimetric parameters.

At a high level, the capability of the retrieval technique to provide well-correlated results at multiple wavelengths are very promising, especially given that many of the assumptions associated with the radiative scattering code used to produce the synthetic dataset are removed when using observed data as input. A larger number of datapoints would need to be considered to draw stronger conclusions, and so in the future, it would be desirable to calculate statistics for all periods of rain during the S-POL deployment. Nonetheless, the case studies presented here are an encouraging sign.

Gridded SEA-POL data was also used to examine how the retrieval performed on RHIs for an intense squall line on July 29th. The direct M_3 retrieval model is employed for this retrieval due to its increased performance relative to the Temprian model and the desire to obtain a physically consistent solution. D_c, M_3 are retrieved directly, with N_c and $\mathcal{T}$ calculated as a result. Rainfall rate is finally calculated through a numerical integration of the obtained full HSD. Figure 3.12 displays the results of this process, with dual-pol variables in panels (a)-(c), blended rainfall rate in panel (d), N_c, D_c, and $\mathcal{T}$ in panels (e)-(g), and the retrieved rain rate in panel (h). Retrieved variables are truncated above a freezing level estimated at about 5 kilometers in height since ANNs were only trained on liquid precipitation.

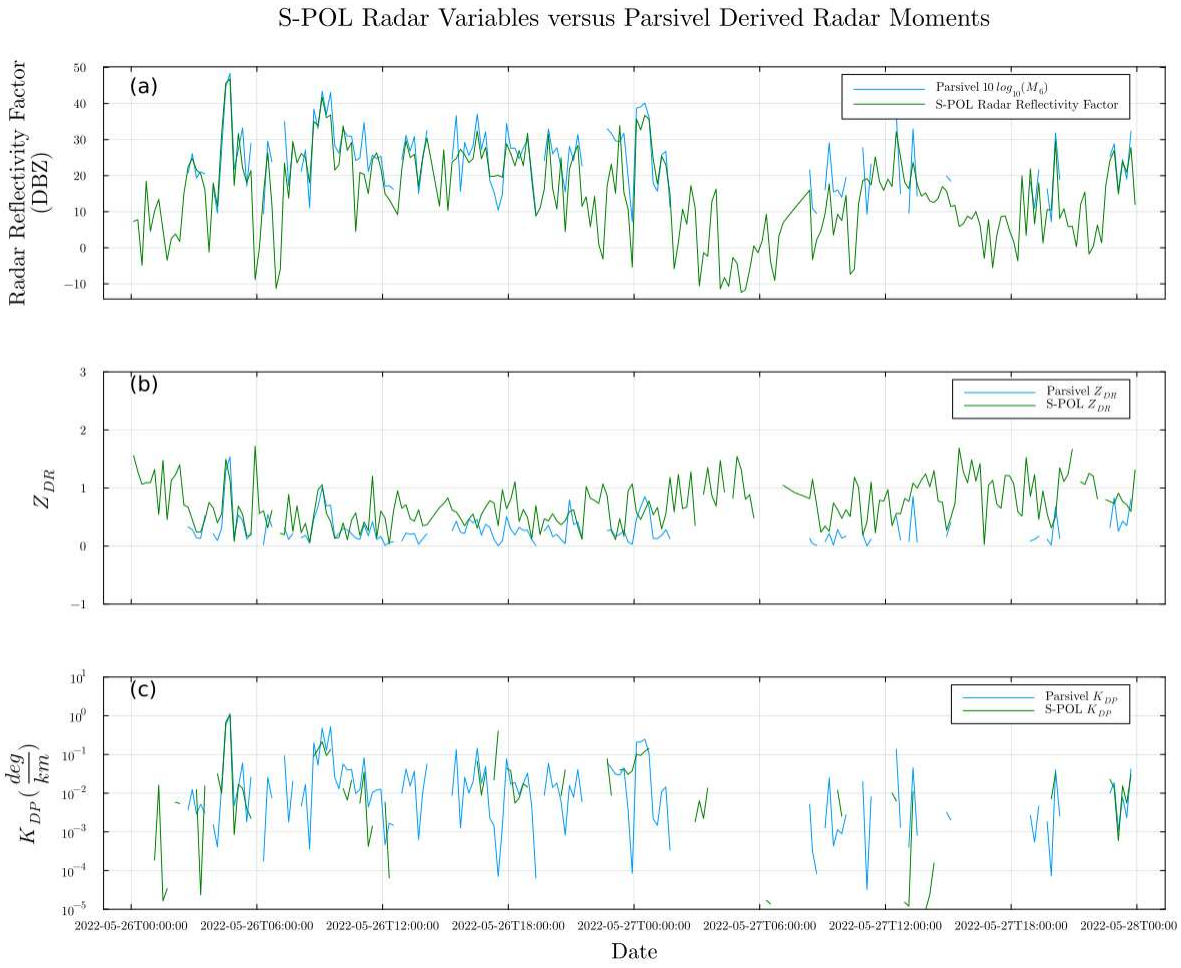

Figure 3.10: S-POL box-average radar observations versus parameters from Hsinchu Parsivel disdrometer from midnight on May 26th, 2022 to midnight on May 28th, 2022. $a)$ Parsivel reflectivity (obtained using Parsivel observed M_6) versus S-POL observed box-average radar reflectivity $b)$ Parsivel calculated differential reflectivity versus S-POL observed box-average differential reflectivity $c)$ Parsivel calculated K_{DP} versus S-POL observed box-average K_{DP}

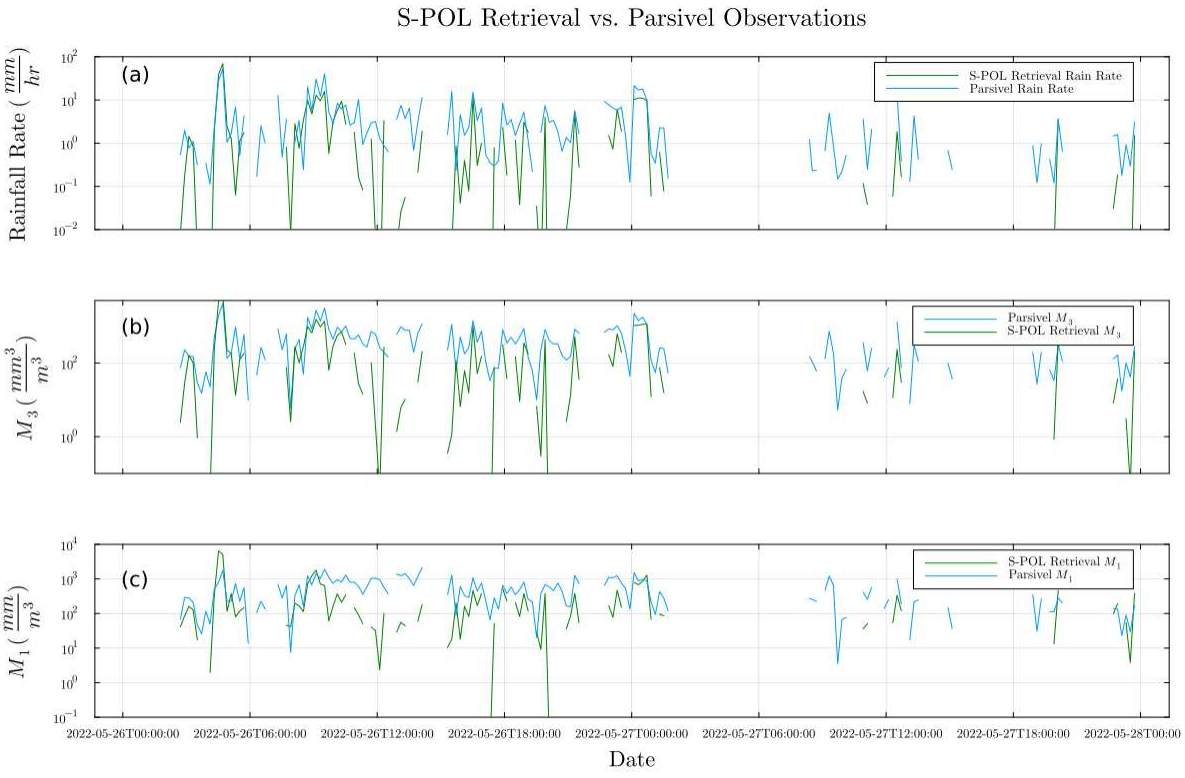

Figure 3.11: S-POL ML retrieved rainfall rate and HSD moments versus Hsinchu Parsivel disdrometer observations from midnight on May 26th, 2022 to midnight on May 28th, 2022 $a)$ Parsivel observed rain rate versus ML model retrieved rain rate ($\frac{mm}{hr}$) $b)$ Parsivel observed M_3 versus ML model retrieved M_3 ($\frac{mm^3}{m^3}$) $c)$ Parsivel observed M_1 versus ML model retrieved M_1 ($\frac{mm}{m^3}$).

Beginning by considering panel (h) suggests that the NN is producing physically consistent results. The rainfall rate structure and magnitude is almost identical to the blended product in panel (d), with a column of highest rain rate tilting down shear toward the radar. Also notable is the presence of a lobe of increased rainfall extending to the left and down from the main conglomerate, resolved in both the blended product and the Temprian-derived version. Panel (h) seems to resolve finer scale columnar structures when compared to panel (d), especially near the freezing level, but it is impossible to say if this is signal or noise at this juncture. As a whole, 3.12 suggests that the model is effective on real radar data, but more analysis is needed to determine characteristics of the retrieval.

3.5 Summary and Conclusions

A technique for developing models for retrieving the full hydrometeor size distributions from dual-pol radar data is presented. The retrievals are based off of a large library of synthetic HSDs which are generated according to a novel three-moment normalization framework and constrained by observations. Radar variables for these synthetic HSDs are calculated using a scattering model and applied to train several ANNs to retrieve the complete HSD. It is found that all methods are at least somewhat effective at retrieving lower-order moments of the size distribution when considered against ground-based disdrometer measurements. Furthermore, the most accurate approaches display several qualitative similarities in terms of the observed HSD shape and its evolution throughout the duration of a convective event. Application to a gridded radar dataset produces physically consistent results and suggests this work to be a sufficient proof-of-concept for the retrieval of full HSDs from radar. Experimentation at C band and S band suggest that successful retrievals are possible for multiple different wavelengths of radar in this framework.

In the future, the utility of other dual-polarimetric variables should also be explored for the retrieval problem. Contemporary literature has suggested that specific attenuation and derivatives thereof could be a viable alternative for K_{DP} while alleviating some of the challenges associated with calculating differential attenuation (Bringi et al., 2020). An additional limitation in the above

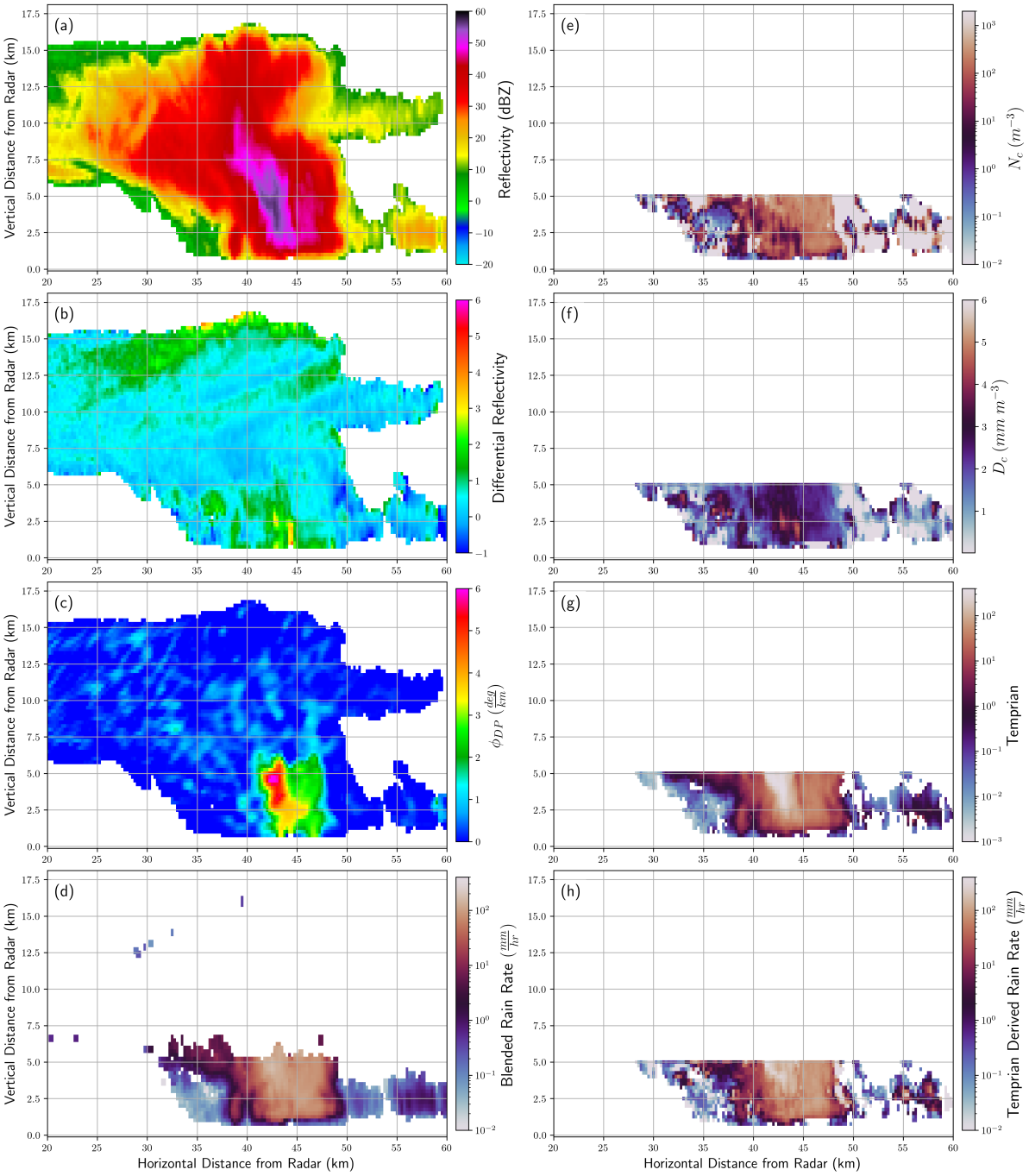

Figure 3.12: Gridded dual-polarimetric and NN retrieved variables from SEA-POL radar on July 29th, 2022. (a) Radar Reflectivity (b) Differential Reflectivity (c) ϕ_{DP}, (d) blended rainfall rate, (e) Neural Net Retrieved N_c, (f) Neural Net Retrieved D_c, (g) Neural Net Retrieved $\mathcal{T}$, (h) Rainfall rate retrieved from $M_{3.67}$ of the drop size distribution specified by retrieved D_c, N_c. Temprian-derived parameters truncated above freezing level of 5 km

approach is the lack of error in the radar measurements used to train the retrieval models. Since these moments are derived from a radiative scattering code, they represent idealized, laboratory level conditions that are generally not obtained when using observations from even research radars, let alone operational platforms. In order to use such a model with these platforms consistently, it would likely be beneficial to introduce noise of some form such that artifacts like negative Z_{DR} or K_{DP} would be within the scope of training data, but the best avenue to do so is as of yet unclear. This drawback will be addressed in future work on the problem.

Chapter 4

Conclusions and Future Work

Meteorological radars are remarkably versatile instruments, providing information across a broad range of spatial scales. The above work demonstrates not only their use as crucial sources of data for forecasting tropical cyclones, some of the largest and most impactful weather events on the planet, but also their capability for obtaining microscale observations on hydrometeor sizes, phenomenon far smaller than the human eye can distinguish.

In this work, we reach several key conclusions regarding the capabilities of radar platforms. We show

- Airborne radar may be used in concert with ML to identify and remove non meteorological features.
- Modern algorithms and programming languages facilitate for quicker and more effective removal than current methods.
- Dual-polarimetric radar information can be employed to accurately derive the full Hydrometeor Size Distribution.
- Machine learning based retrievals of the full HSD contain useful information about physically important quantities such as rainfall rate.

Several threads connect all of the above conclusions. The importance of dual-polarimetric parameters is highlighted throughout the work. It is shown that many existing quality control algorithms rely on dual-pol moments, and these same moments are fundamental to the hydrometeor retrieval problem. Without dual-pol information, it would be nearly impossible to discern anything about the orientation or composition of targets, meteorological or otherwise.

Machine learning methodologies are also integral in all facets. On the QC problem, ML is shown to be able to achieve performance exceeding operational systems, and in a performant fash-

ion. Literature also shows promise in the capabilities of ML to de-alias folded Doppler velocity, a problem that has proved challenging for traditional algorithms. Deep learning further facilitates the retrieval of the full HSD with minimal underlying assumptions, in contrast to previous work where direct relationships between radar parameters and HSD moments had to be derived by hand.

The future holds significant promise for further development of radar science, especially in the airborne realm. ML techniques continue to mature, and it is also expected that new measurement platforms will come online, with phased array technology one such example. Though still in its infancy, it has shown tremendous potential for groundbreaking observations, drastically increasing temporal and spatial resolutions of radar sweeps. Undoubtedly, radars will persist as indispensable for meteorological research and operations for many years to come.

Bibliography

Bell, M. M., W.-C. Lee, C. A. Wolff, and H. Cai, 2013: A solo-based automated quality control algorithm for airborne tail doppler radar data. *Journal of Applied Meteorology and Climatology*, **52 (11)**, 2509 – 2528, https://doi.org/10.1175/JAMC-D-12-0283.1.

Bezanson, J., S. Karpinski, V. B. Shah, and A. Edelman, 2012: Julia: A fast dynamic language for technical computing. *CoRR*, **abs/1209.5145**, 1209.5145.

Blaom, A. D., F. Kiraly, T. Lienart, Y. Simillides, D. Arenas, and S. J. Vollmer, 2020: MLJ: A julia package for composable machine learning. *Journal of Open Source Software*, **5 (55)**, 2704, https://doi.org/10.21105/joss.02704.

Bringi, V., K. V. Mishra, M. Thurai, P. C. Kennedy, and T. H. Raupach, 2020: Retrieval of lower-order moments of the drop size distribution using csu-chill x-band polarimetric radar: a case study. *Atmospheric Measurement Techniques*, **13 (9)**, 4727–4750, https://doi.org/10.5194/amt-13-4727-2020.

Cho, H.-K., K. P. Bowman, and G. R. North, 2004: A comparison of gamma and lognormal distributions for characterizing satellite rain rates from the tropical rainfall measuring mission. *Journal of Applied Meteorology*, **43 (11)**, 1586 – 1597, https://doi.org/10.1175/JAM2165.1.

Christophersen, H., J. Sippel, A. Aksoy, and N. L. Baker, 2022: Recent advancements for tropical cyclone data assimilation. *Annals of the New York Academy of Sciences*, **1517 (1)**, 25–43, https://doi.org/https://doi.org/10.1111/nyas.14873, https://nyaspubs.onlinelibrary.wiley.com/doi/pdf/10.1111/nyas.14873.

DesRosiers, A. J., and M. M. Bell, 2024: Airborne radar quality control with machine learning. *Artificial Intelligence for the Earth Systems*, **3 (1)**, e230 064, https://doi.org/10.1175/AIES-D-23-0064.1.

DesRosiers, A. J., M. M. Bell, J. C. DeHart, J. L. Vigh, C. M. Rozoff, and E. A. Hendricks, 2025: Tropical cyclone surface winds from aircraft with a neural network. *Journal of Geophysical Research: Machine Learning and Computation*, **2 (2)**, e2025JH000 584, https://doi.org/https://doi.org/10.1029/2025JH000584, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2025JH000584.

Dolan, B., B. Fuchs, S. A. Rutledge, E. A. Barnes, and E. J. Thompson, 2018: Primary modes of global drop size distributions. *Journal of the Atmospheric Sciences*, **75 (5)**, 1453 – 1476, https://doi.org/10.1175/JAS-D-17-0242.1.

Fabry, F., 2018: *Radar Meteorology: Principles and Practice*. Cambridge University Press, URL https://www.cambridge.org/core/books/radar-meteorology/9871DF1E29731166DACDBA8C29CB0B3D.

Gall, R., J. Franklin, F. Marks, E. N. Rappaport, and F. Toepfer, 2013: The hurricane forecast improvement project. *Bulletin of the American Meteorological Society*, **94 (3)**, 329 – 343, https://doi.org/10.1175/BAMS-D-12-00071.1.

Gamache, J. F., P. Dodge, and N. F. Griffin, 2008: Automatic quality control and analysis of airborne doppler data: Realtime applications, and automatically post-processed analyses for research. *28th Conference on Hurricanes and Tropical Meteorology, Preprints*.

Gao, J., M. Xue, K. Brewster, and K. K. Droegemeier, 2004: A three-dimensional variational data analysis method with recursive filter for doppler radars. *Journal of Atmospheric and Oceanic Technology*, **21 (3)**, 457 – 469, https://doi.org/10.1175/1520-0426(2004)021<0457:ATVDAM>2.0.CO;2.

Gunn, K. L. S., and J. S. Marshall, 1958: The distribution with size of aggregate snowflakes. *Journal of Atmospheric Sciences*, **15 (5)**, 452 – 461, https://doi.org/10.1175/1520-0469(1958)015<0452:TDWSOA>2.0.CO;2.

Ho, J., G. Zhang, P. Bukovcic, D. B. Parsons, F. Xu, J. Gao, J. T. Carlin, and J. C. Snyder, 2023a: Improving polarimetric radar-based drop size distribution retrieval and rain estimation using a deep neural network. *Journal of Hydrometeorology*, **24 (11)**, 2057 – 2073, https://doi.org/10.1175/JHM-D-22-0166.1.

Ho, J., G. Zhang, P. Bukovcic, D. B. Parsons, F. Xu, J. Gao, J. T. Carlin, and J. C. Snyder, 2023b: Improving polarimetric radar-based drop size distribution retrieval and rain estimation using a deep neural network. *Journal of Hydrometeorology*, **24 (11)**, 2057 – 2073, https://doi.org/10.1175/JHM-D-22-0166.1.

Hochreiter, S., 1998: The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **06 (02)**, 107–116, https://doi.org/10.1142/S0218488598000094, https://doi.org/10.1142/S0218488598000094.

Jung, Y., M. Xue, and G. Zhang, 2010: Simulations of polarimetric radar signatures of a supercell storm using a two-moment bulk microphysics scheme. *Journal of Applied Meteorology and Climatology*, **49 (1)**, 146 – 163, https://doi.org/10.1175/2009JAMC2178.1.

Keeler, R., J. Lutz, and J. Vivekanandan, 2000: S-pol: Ncar's polarimetric doppler research radar. *IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium. Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment. Proceedings (Cat. No.00CH37120)*, Vol. 4, 1570–1573 vol.4, https://doi.org/10.1109/IGARSS.2000.857275.

Keisler, R., 2022: Forecasting global weather with graph neural networks. *arXiv preprint arXiv:2202.07575*.

Lakshmanan, V., A. Fritz, T. Smith, K. Hondl, and G. Stumpf, 2007: An automated technique to quality control radar reflectivity data. *Journal of Applied Meteorology and Climatology*, **46 (3)**, 288 – 305, https://doi.org/10.1175/JAM2460.1.

Lakshmanan, V., C. Karstens, J. Krause, and L. Tang, 2014: Quality control of weather radar data using polarimetric variables. *Journal of Atmospheric and Oceanic Technology*, **31 (6)**, 1234 – 1249, https://doi.org/10.1175/JTECH-D-13-00073.1.

Laws, J. O., and D. A. Parsons, 1943: The relation of raindrop-size to intensity. *Eos, Transactions American Geophysical Union*, **24 (2)**, 452–460, https://doi.org/10.1029/tr024i002p00452.

Lee, G., V. Bringi, and M. Thurai, 2023: The retrieval of drop size distribution parameters using a dual-polarimetric radar. *Remote Sensing*, **15 (4)**, https://doi.org/10.3390/rs15041063.

Liang, H., M. Zhang, and H. Wang, 2019: A neural network model for wildfire scale prediction using meteorological factors. *IEEE Access*, **7**, 176 746–176 755, https://doi.org/10.1109/ACCESS.2019.2957837.

Maas, A. L., A. Y. Hannun, A. Y. Ng, and Coauthors, 2013: Rectifier nonlinearities improve neural network acoustic models. *Proc. icml*, Atlanta, GA, Vol. 30, 3.

Marks, F. D., 2003: State of the science: Radar view of tropical cyclones. *Meteorological Monographs*, **30 (52)**, 33 – 74, https://doi.org/10.1175/0065-9401(2003)030<0033:SOTSRV>2.0.CO;2.

Marshall, J. S., and W. M. K. Palmer, 1948: The distribution of raindrops with size. *Journal of Atmospheric Sciences*, **5 (4)**, 165 – 166, https://doi.org/10.1175/1520-0469(1948)005<0165:TDORWS>2.0.CO;2.

Mishchenko, M. I., L. D. Travis, and D. W. Mackowski, 1996: T-matrix computations of light scattering by nonspherical particles: A review. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **55 (5)**, 535–575, https://doi.org/10.1016/0022-4073(96)00002-7.

Morrison, H., M. R. Kumjian, C. P. Martinkus, O. P. Prat, and M. van Lier-Walqui, 2019: A general n-moment normalization method for deriving raindrop size distribution scaling relationships. *Journal of Applied Meteorology and Climatology*, **58 (2)**, 247 – 267, https://doi.org/10.1175/JAMC-D-18-0060.1.

NOAA Hurricane Research Division, 2025: Wp-3d radars. URL https://www.aoml.noaa.gov/hrd/about_hrd/HRD-P3_radar.html, accessed: 25 July 2025.

Raupach, T. H., and A. Berne, 2015: Correction of raindrop size distributions measured by parsivel disdrometers, using a two-dimensional video disdrometer as a reference. *Atmospheric Measurement Techniques*, **8 (1)**, 343–365, https://doi.org/10.5194/amt-8-343-2015.

Raupach, T. H., and A. Berne, 2017: Retrieval of the raindrop size distribution from polarimetric radar data using double-moment normalisation. *Atmospheric Measurement Techniques*, **10 (7)**, 2573–2594, https://doi.org/10.5194/amt-10-2573-2017.

Reasor, P. D., J. Gamache, M. M. Bell, A. J. DesRosiers, M. S. Fischer, and K. Neighbour, 2024: A new approach to real-time automated processing of hurricane reconnaissance radar data. *36th Conference on Hurricanes and Tropical Meteorology, Proceedings*.

Rutledge, S. A., V. Chandrasekar, B. Fuchs, J. George, F. Junyent, B. Dolan, P. C. Kennedy, and K. Drushka, 2019: Sea-pol goes to sea. *Bulletin of the American Meteorological Society*, **100 (11)**, 2285 – 2301, https://doi.org/10.1175/BAMS-D-18-0233.1.

Ryzhkov, A., P. Zhang, P. Bukovčić, J. Zhang, and S. Cocks, 2022: Polarimetric radar quantitative precipitation estimation. *Remote Sensing*, **14 (7)**, https://doi.org/10.3390/rs14071695.

Sadeghi, B., and Coauthors, 2022: DecisionTree.jl - A Julia implementation of the CART Decision Tree and Random Forest algorithms. Zenodo, URL https://doi.org/10.5281/zenodo.7359268, https://doi.org/10.5281/zenodo.7359268.

Sekhon, R. S., and R. C. Srivastava, 1970: Snow size spectra and radar reflectivity. *Journal of Atmospheric Sciences*, **27 (2)**, 299 – 307, https://doi.org/10.1175/1520-0469(1970)027<0299:SSSARR>2.0.CO;2.

Sekhon, R. S., and R. C. Srivastava, 1971: Doppler radar observations of drop-size distributions in a thunderstorm. *Journal of Atmospheric Sciences*, **28 (6)**, 983 – 994, https://doi.org/10.1175/1520-0469(1971)028<0983:DROODS>2.0.CO;2.

Steiner, M., and J. A. Smith, 2002: Use of three-dimensional reflectivity structure for automated detection and removal of nonprecipitating echoes in radar data. *Journal of Atmospheric and Oceanic Technology*, **19 (5)**, 673 – 686, https://doi.org/10.1175/1520-0426(2002)019<0673:UOTDRS>2.0.CO;2.

Toms, B. A., E. A. Barnes, and I. Ebert-Uphoff, 2020: Physically interpretable neural networks for the geosciences: Applications to earth system variability. *Journal of Advances in Modeling Earth Systems*, **12 (9)**, e2019MS002 002, https://doi.org/https://doi.org/10.1029/2019MS002002, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019MS002002.

Torres, D. S., J. M. Porrà, and J.-D. Creutin, 1994: A general formulation for raindrop size distribution. *Journal of Applied Meteorology (1988-2005)*, **33 (12)**, 1494–1502.

Ulbrich, C. W., 1983: Natural variations in the analytical form of the raindrop size distribution. *Journal of Applied Meteorology and Climatology*, **22 (10)**, 1764 – 1775, https://doi.org/10.1175/1520-0450(1983)022<1764:NVITAF>2.0.CO;2.

Veillette, M. S., J. M. Kurdzo, P. M. Stepanian, J. McDonald, S. Samsi, and J. Y. N. Cho, 2023: A deep learning–based velocity dealiasing algorithm derived from the wsr-88d open radar product generator. *Artificial Intelligence for the Earth Systems*, **2 (3)**, e220 084, https://doi.org/10.1175/AIES-D-22-0084.1.

Weinkle, J., C. Landsea, D. Collins, R. Musulin, R. P. Crompton, P. J. Klotzbach, and R. Pielke, 2018: Normalized hurricane damage in the continental united states 1900–2017. *Nature Sustainability*, **1 (12)**, 808–813, https://doi.org/10.1038/s41893-018-0165-2.

Wiesner, J., 1895: Beiträge zur kenntniss des tropischen regens [in german]. URL https://cir.nii.ac.jp/crid/1370564064168602643, 1397 pp.

Wurman, J., J. M. Straka, and E. N. Rasmussen, 1996: Fine-scale doppler radar observations of tornadoes. *Science*, **272 (5269)**, 1774–1777, https://doi.org/10.1126/science.272.5269.1774, https://www.science.org/doi/pdf/10.1126/science.272.5269.1774.

Yu, N., B. Boudevillain, G. Delrieu, and R. Uijlenhoet, 2012: Estimation of rain kinetic energy from radar reflectivity and/or rain rate based on a scaling formulation of the raindrop size distribution. *Water Resources Research*, **48 (4)**, https://doi.org/https://doi.org/10.1029/2011WR011437, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2011WR011437.

Zhang, H., H. Wang, Z. Xu, Z. Liu, and B. C. Khoo, 2023: A physics-informed neural network-based approach to reconstruct the tornado vortices from limited observed data. *Journal of Wind Engineering and Industrial Aerodynamics*, **241**, 105 534, https://doi.org/https://doi.org/10.1016/j.jweia.2023.105534.