UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

IDENTIFICATION OF ENHANCED RAINFALL RATES USING THE NEAR-STORM ENVIRONMENT FOR RADAR PRECIPITATION ESTIMATES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

HEATHER MARIE MOSER GRAMS
Norman, Oklahoma
2012

IDENTIFICATION OF ENHANCED RAINFALL RATES USING THE NEAR-STORM ENVIRONMENT FOR RADAR PRECIPITATION ESTIMATES

A DISSERTATION APPROVED FOR THE
SCHOOL OF METEOROLOGY

BY

_____
Dr. Phillip Chilson, Co-Chair

_____
Dr. Jian Zhang, Co-Chair

_____
Dr. Michael Richman

_____
Dr. Lance Leslie

_____
Dr. Yang Hong

ACKNOWLEDGEMENTS

I would first of all like to thank my committee co-chairs, Dr. Jian Zhang and Dr. Phillip Chilson. Their support and constructive advice made the process of research and dissertation writing much easier, and I feel that this dissertation is stronger as a result of their input. I'd also like to thank the rest of my committee, Drs. Yang Hong, Lance Leslie, and Michael Richman for their commitment of time and guidance along the way.

I'd also like to acknowledge some others who have been vital to the success of this research. Carrie Langston Killough provided invaluable assistance with the computing and coding side of the NMQ system and accessing much of the data needed for this study. Of the countless times I've stumbled over file formats and debugging C++ code, she was always available and often knew exactly how to fix the problem, which was an amazing resource to have. Dr. Kim Elmore and Dr. Valliappa Lakshmanan gave me some much-needed guidance on the statistical methods in the dissertation, which I feel strengthened the quality of the research and the results. Additionally, Drs. Lin Tang and Youcun Qi provided some helpful assistance in generating some of the dual-polarization output and the associated probabilities of enhanced rain rate for Hurricane Isaac. Dr. Suzanne Van Cooten, Kodi Monroe, and the National Sea Grant Program were the primary people who helped me get funding for my Ph.D. work, and I greatly appreciate their contributions, not just in terms of funding but also for their advice along the way.

Lastly I'd like to thank my family, friends, and my husband, Jeremy Grams, for their encouragement and unwaivering support. Graduate school is a long, tedious process, but they were always confident in my abilities as a student and researcher and

never had any doubts whatsoever that I could and would finish. When people have that kind of unconditional confidence in you, it inspires you to believe that you can accomplish anything. Thank you!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Reliable and timely flash flood warnings are critically dependent on the accuracy of real-time rainfall estimates. Precipitation is not only the most vital input for basin-scale accumulation algorithms such as the Flash Flood Monitoring and Prediction (FFMP) program used operationally by the U.S. National Weather Service, but it is the primary forcing for hydrologic models at all scales. Quantitative precipitation estimates (QPE) from radar are widely used for such a purpose due to their high spatial and temporal resolution compared to rain gauges and satellite-based algorithms. However, converting the native radar variables into an instantaneous rain rate is fraught with uncertainties.

One of those uncertainties is the varying relationship of radar observables to rain rate for different regions and storm types due to variations in drop size distributions. Many unique reflectivity-to-rain rate (Z-R) functions have been proposed in the literature over the past 70 years for single-polarization radars, and it is becoming apparent that various rain rate functions will also be needed in different environments for dual-polarization radars as well. The challenge then becomes identifying the environments in real-time such that the appropriate rain rate function can be applied. This study addresses the challenge of identifying environments conducive for tropical rain rates, or rain rates that are enhanced by highly productive warm rain processes. Rain rates in tropical environments tend to be underestimated by other operational Z-R functions and have often been associated with historic flash flooding events, so delineating them in real-time can greatly improve not only the radar-based QPE

accuracy, but the level of certainty by forecasters for issuing flash flood warnings as well.

Six consecutive months of hourly data from the 2010 warm season were used to train ensembles of statistical classification models such that probabilities of warm rain enhancement of rain rate can be derived. The predictors for the ensemble were retrieved from the 20-km Rapid Update Cycle (RUC) model analyses and were chosen to provide a general description of the thermodynamic environment from the which the rainfall developed. Those environmental predictors were trained against two different predictands: bias of rain rates for the convective Z-R function vs. collocated, quality controlled rain gauges, and the vertical gradient of radar reflectivity between the freezing level and the lowest elevation observed by the radar. The resulting probabilities from the trained ensembles were then used to delineate where tropical rain rates would be assigned in a gridded QPE product, and the resulting hourly accumulations were verified against independent rain gauges.

Overall, the probability-based precipitation type delineation scheme improved hourly rainfall accumulations for three independent cases tested when compared to both the legacy rainfall product from the National Mosaic and Multisensor Quantitative Precipitation Estimation (NMQ) project and the operational NWS rainfall product (Stage II), but neither the gauge-based nor VPR-based ensembles emerged as a clearly superior predictor than the other for all cases tested. However, spatial similarities between the two probability fields and similar results from variable importance analysis suggest that both methods are attempting to delineate the same environment. This implies that the systematic underestimation of radar-based QPE and the enhancement of

reflectivity in the warm layer from warm rain hydrometeor growth are related or at the very least are associated with the same type of environment. Initial analysis of polarimetric variables, particularly differential reflectivity, in areas of high and low probabilities also support a connection between rain rate underestimation and tropical airmasses.

CHAPTER 1: INTRODUCTION

According to the National Weather Service's natural hazards statistics for the United States, flash flooding and river flooding have killed 1030 and 338 people, respectively, since 1995. That is equivalent to about 81 fatalities per year since the NWS's national radar network was upgraded to the current operational NEXRAD doppler radar technology (i.e., pre-dual-polarization). Prior to the NEXRAD era (1940-1994), flood-related fatalities averaged 113 per year. The radar upgrade has been credited as a reason for the reduction in fatalities from flash floods, particularly in the western United States (National Research Council 2005). Other factors include improved numerical weather prediction of storms, better and more timely information from satellites and surface stations, and upgrades to the software tools forecasters used to issue their warnings (National Research Council 2011).

While NEXRAD technology has led to significant improvements in flash flood detection and warning lead time, further improvement can still be made. Two of the major scientific obstacles to improved flash flood detections and river flood predictions are 1) the accuracy of precipitation estimates and forecasts, and 2) limitations of hydrologic models. Both are very active and long-standing areas of research, but this study will focus solely on the first challenge. Specifically, this study addresses the problem of real-time precipitation estimate accuracy, particularly in heavy rain events that can produce flash floods.

The National Mosaic and Multisensor Quantitative Precipitation Estimation (NMQ) Project’s Q2 product is a fully automated, multisensor precipitation estimation tool that produces hourly rainfall accumulations on a 1-km grid over the continental

United States (Zhang et al. 2011; Vasiloff et al. 2007). The spatial distribution of rainfall is derived from a national mosaic of the level-2 data from each radar in the National Weather Service's WSR-88D network. Hourly analysis fields from the 20-km Rapid Update Cycle (RUC) model and satellite data are then used to remove non-precipitating echoes from the radar reflectivity field and segregate the different types of precipitation (stratiform and convective rain, hail, tropical rain, and snow). Different reflectivity-to-rain rate functions (Z-Rs) are then applied to the different precipitation types across the domain for each 5-minute period and are accumulated to create hourly rainfall estimates (or the liquid equivalent in areas of frozen precipitation). Z-R functions generally take the form of a power law where the coefficient A and the exponent b (Eq. 1.1) are empirically derived by fitting the function to observed data. More details on how Q2 is generated can be found in Zhang et al. (2011). While still technically an experimental product, Q2 is increasingly being used as a useful and beneficial tool by forecasters at both the NWS weather forecast offices and river forecast centers because the dynamic Z-R selection has lower bias and high accuracy than the legacy rainfall estimates which are based on single Z-Rs over the entire radar domain.

$$R = AZ^b \tag{1.1}$$

Identification of the tropical or warm rain-enhanced precipitation type for Q2 is currently based on spatially averaged vertical profiles of reflectivity (VPR) calculated for each radar location. VPRs are retrieved for each 1-km by 1-degree polar grid point located between 20 km and 80 km from the radar and are combined to produce an average profile. If the brightband identification algorithm determines that a brightband exists for that location (Zhang et al. 2008), the vertical gradient of reflectivity below the

brightband is then calculated. Otherwise, the gradient is calculated starting at the height of the 0° C isotherm. If the brightband or 0° C isotherm is at least 2.5 km above the surface and reflectivity increases with decreasing height below it or remains approximately constant, then the VPR is flagged as a "tropical" VPR (Xu et al. 2008), which basically means that enhanced growth through warm rain processes were detected in the warm layer of the storm. Otherwise, it is identified as a stratiform or convective VPR based on other criteria. If the VPR is identified as tropical, the National Weather Service's default tropical Z-R function is then applied to all locations within 200 km from the radar where the lowest available (i.e., hybrid scan) reflectivity exceeds 30 dBZ.

While the introduction of the tropical precipitation type in Q2 has greatly improved the problem of rainfall underestimation in true tropical systems like hurricanes and in environments conducive for warm rain-enhanced rainfall rates, it tends to produce large overestimates in other storm types. These types include midlatitude continental mesoscale convective systems (MCS) and extratropical cyclones.

Basing the tropical flag on VPRs alone may not be sufficient to delineate enhanced warm rain droplet growth from areas where other microphysical processes are dominating, so additional inputs are needed. Because the NMQ system already ingests the RUC analysis data for other Q2 processes, analysis fields related to humidity, instability, and temperature can be used to determine the airmass properties where tropical rainfall is being identified by the VPR algorithm. Because of the wide range of possible storm modes and environmental conditions across the United States at any

given time, however, using single thresholds in the parameters will likely not be general enough to be applicable in a real-time, year-round product like Q2. An attractive alternative is to use a statistical learning method such as support vector machines to identify different combinations of parameter values across a "training" period composed of multiple event types in different regions.

Previous studies had shown that dual-polarization rainfall algorithms can produce more accurate accumulations at high rain rates than rainfall algorithms based on horizontally-polarized reflectivity alone (Giangrande and Ryzhkov 2008; Ryzhkov et al. 2005; Chandrasekar et al. 1990). However, all these previous studies compared polarimetric rainfall algorithms to either a single Z-R function or a simplistic segregation of convective and stratiform rain rather than comparing to a dynamic Z-R product like Q2. For heavy rainfall in particular, the lack of additional Z-Rs beyond the standard convective and stratiform would have certainly given the impression of systematic, significant underestimation by the legacy rainfall algorithms. Furthermore, recent work has indicated that there exist significant variations among R(Z,Zdr) relationships for different types of precipitation systems that may require additional, environmental information for more accurate delineation (Ryzhkov, personal communication). Thus, the current study can be potentially beneficial for dual-polarization-based rainfall estimates as well, particularly at long distances from the radar where DSD changes between the radar beam's level and the ground surface can have a major impact on the accuracy of surface rainfall estimates.

Additionally, a need still exists for an accurate, high-resolution, gridded rainfall data archive for research and climatological applications. An initiative is underway to

produce a 10-year retrospective archive of not only the NMQ rainfall products, but also other radar-based and multisensor products for the continental United States. In order for the rainfall dataset to be a useful resource for future studies, the QPE should be as accurate as possible. Producing accurate QPE for high-impact, high-intensity events is of primary importance, and this study addresses that need by refining and improving the method by which high rain rates are detected for Q2.

Furthermore, this study investigates the underlying environmental ingredients that may be responsible for systematic large radar rainfall biases for heavy rain rates on the mesoscale. Most previous studies of very heavy rainfall have focused on long-term accumulations (generally daily), and the mechanisms responsible for producing large rainfall totals on longer time scales can be quite different from the mechanisms responsible for high rainfall *rates* on the hourly or sub-hourly time scale. The studies that have focused on high rain rates at short time scales (i.e., flash flood-producing rainfall), tended to either concentrate on forecasting a future flash flooding event hours in advance or examining the microphysical processes responsible for such rain rates. While understanding the microphysics of extreme rain rates is very important, particularly with the proposal of dual-polarization methods for estimating DSD model parameters, they provide little guidance on how to identify such rain rates in situations where the DSD parameters cannot be derived (i.e., Z-R-based algorithms and areas where the radar beam is too high above the surface for adequate sampling).

A predictive, probabilistic statistical model ensemble has been developed and tested in this study that can detect environmental patterns favorable for warm rain enhancement of rain rates such that localized bias correction or Z-R selection can be

applied. It also identifies environments unfavorable for such enhancement to mitigate the tendency toward false detection of tropical rain rates based on the area-averaged VPR structure alone. Training the statistical model on environment properties allows for identification of enhanced rain rates on a regular grid across the entire United States even where low-level sampling of hydrometeors from radar is not available, which can be valuable for VPR-based algorithms that attempt to reconstruct low-level reflectivity for rainfall estimation in complex terrain and areas with poor radar coverage. The environment-based approach also allows this model to applied to rainfall events where polarimetric rainfall estimates were not yet available, which is beneficial for the 10-year reanalysis initiative.

Chapter 2 provides a general background and literature review of several key aspects of this study: rainfall microphysics and growth processes, past climatological studies of heavy rainfall in the U.S., and the uncertainties associated with radar-based rainfall estimation. Chapter 3 outlines the sources of data and describes the quality control methods applied. Chapter 4 describes the methodology used in this study, including an overview of the chosen statistical tools. Chapter 5 contains the results of the analysis, and Chapter 6 provides a summary and final discussion of the results.

CHAPTER 2: BACKGROUND AND LITERATURE REVIEW

Rainfall Microphysics

Much of the pioneering field work and laboratory experiments in the area of rainfall microphysics were conducted in the 20th century, and a detailed analytic and conceptual review of this work was provided by Rogers and Yau (1993). A general summary of warm and cold rain processes are provided here as a foundation for more recent work that will be discussed in more detail.

Warm Rain

Warm rain microphysics describes the process by which rain drops form from water vapor in a supersaturated, non-freezing ambient environment. Cloud drops initially grow entirely by diffusion and/or condensation of water vapor in a supersaturated airmass where ice crystals are absent (i.e., the ambient temperature is above 0° C or latent heat release from condensation is sufficient to keep cloud droplets in the liquid phase). As the diameters of the drops increase, the rate at which they increase in size diminishes. Thus, the drop size distribution (DSD) of the population narrows over time because the smaller drops grow more rapidly than the larger ones.

Once the drops grow to a sufficient size (generally a diameter of 10-20 μm), coalescence, collision, and breakup tend to become the more dominant mechanisms controlling drop sizes. These processes act to broaden the DSD by generating more large and small drops and fewer drops in the middle range of the distribution. Bowen (1950) provided an approximate function describing the processes that control the rate of drop growth with height as it falls:

$$\frac{dR}{dz} = \frac{\bar{E}M}{4\rho_L} \frac{u(R)}{U - u(R)} \tag{2.1}$$

where $\bar{E}$ is the effective average collection efficiency, M is the cloud liquid water content per unit volume, ρ_L is the liquid density of the drop, U is the updraft velocity, and $u(R)$ is the drop's terminal fall speed. Thus, theoretically speaking, the total change in drop size before reaching the surface as rain is directly proportional to the total integrated distance the drop travels within the saturated cloud environment and inversely proportional to the strength of the updraft. For very weak updrafts, however, the drops may not be lofted as high into the cloud before their fall speed exceeds the updraft, which would limit their ultimate growth. Therefore, maximum drop growth is achieved for an updraft velocity that maximizes the total trajectory distance that the drop travels within the cloud. That growth is further enhanced by increased liquid water content, which is controlled by the water vapor content and updraft velocity of the cloud during the condensation phase (Rogers and Yau 1993). The Bowen model provides a generally accurate description of how rain begins in cumulus clouds, though there are stochastic processes occurring simultaneously that allow some drops to grow faster and fall out sooner than others (Telford 1955; Robertson 1974).

Once rain has initiated, the downdraft generated by the falling raindrops and diabatically cooled air permits smaller drops to begin raining out (Atlas and Ulbrich 2000). This process causes warm rain convection to have a transient DSD throughout the lifetime of the storm that can produce rain rates that deviate significantly from the standard Z-R functions used operationally (Carbone and Nelson 1978). This also demonstrates a potential difference on small temporal scales between the DSD of rainfall reaching the ground (i.e., what is measured by rain gauges and disdrometers)

and the DSD within the cloud (i.e., what is often measured by radar). Evaporation, time delay of drops reaching the surface, and horizontal advection of drops falling within and below the cloud base are additional factors contributing to this difference. Morin et al. (2003) estimated that raindrops observed by radar at a height of 3 km above ground could take 5-15 minutes to reach the surface and could be displaced 3-9 km from the location where they were detected. They also found that these "synchronization" errors had a greater impact on the exponent *b* in the Z-R power law than errors from hail contamination or random reflectivity errors.

Rosenfeld and Ulbrich (2003) examined each microphysical process separately with respect to its impact on DSD properties as approximated by the gamma distribution model (earlier work in this area can be found in Ulbrich and Atlas 1998, Ulbrich 1983, and Wilson and Brandes 1979):

$$N(D) = N_0 D^{\mu} \exp(-\Lambda D) \tag{2.2}$$

$$\Lambda = \frac{3.67}{D_0} \tag{2.3}$$

where D is the drop diameter (mm), $N(D)$ (m^{-3} mm^{-1}) is the number of drops in the resolution volume per drop size interval, N_0 (m^{-3} $mm^{-1-\mu}$) is an intercept parameter that varies roughly inversely to μ (Ulbrich 1983; Uijlenhoet 2001), μ is a shape parameter for the gamma function, and D_0 is the median drop volume diameter of the distribution (i.e., the diameter a spherical drop of equivalent volume would have since raindrops often adopt an oblong shape during descent due to drag).

The radar reflectivity factor Z (mm^{-6} m^{-3}) and rainfall rate R (mm hr^{-1}) can be derived from a drop size distribution as the sixth and third moments of the DSD,

respectively:

$$Z = \int_0^\infty D^6 N(D) dD \tag{2.4}$$

$$R = \frac{\pi}{6} \int_0^\infty D^3 N(D) w_t(D) dD \tag{2.5}$$

$$w_t(D) = 9.65 - 10.3 \exp(-600D) \tag{2.6}$$

where $w_t(D)$ (m s^{-1}) is the terminal fall velocity as a function of drop diameter as derived by Atlas et al. (1973), assuming no other vertical air motion. Ulbrich (1983) demonstrated that a power law function in the form $P=\alpha Q^{\beta}$ could be derived for any two moments of the gamma DSD model based on the following two equations:

$$P = a_P \frac{\Gamma(p+\mu+1)}{(3.67+\mu)^{p+\mu+1}} N_0 D_0^{p+\mu+1} \tag{2.7}$$

$$Q = a_q \frac{\Gamma(q+\mu+1)}{(3.67+\mu)^{q+\mu+1}} N_0 D_0^{q+\mu+1} \tag{2.8}$$

where p and q are the moments of the DSD (i.e., 3 and 6 for the rain rate and reflectivity, respectively), and a_p and a_q are constants that are functions of the moments (provided in a table within Ulbrich 1983). For rain rate, a_q is 33.31 mm h^{-1} m^3 $cm^{-3.67}$, and for reflectivity, a_p is 10^6 mm^6 cm^{-6}. When the above equations are solved for μ and N_0 and are cast as functions of the Z-R parameters A and b (Rosenfeld and Ulbrich 2003):

$$N_0 = \left[\frac{A \left[33.31 \Gamma \left(\frac{2.33}{b-1} \right) \right]^b}{10^6 \Gamma \left(\frac{2.33b}{b-1} \right)} \right]^{1/(1-b)} \tag{2.9}$$

$$\mu = \frac{7 - 4.67b}{b-1} \tag{2.10}$$

Based on the above relations, Rosenfeld and Ulbrich (2003) summarized the

impact that each of the major microphysical warm rain processes has on both the parameters of the DSD and the Z-R function:

1. **Coalescence** decreases the number of small drops and increases the number of larger drops. Therefore, it supports an increase in D_0, an increase in μ, and a decrease in N_0. These DSD changes result in an increase in the coefficient A and a decrease in the exponent b in the Z-R power law.
2. **Breakup** decreases D_0 and increases N_0. This results in a lower A and a small increase in b. It essentially has the opposite effect on a DSD as coalescence such that it increases the number of small drops and decreases the number of large drops.
3. **Simultaneous coalescence and breakup** greatly increases μ, which leads to a decrease in the Z-R exponent b. The change in A would correspond to the dominant process of the two, meaning A would increase if coalescence dominates or decrease if breakup dominates.
4. **Accretion** of cloud water by drops tends to raise D_0 and lower N_0. It would therefore increase A without significantly changing b.
5. **Evaporation** acts to reduce cloud water from the entire system and remove the smallest drops. Thus, N_0 would decrease, and μ and D_0 would increase. The result is a Z-R function with a higher A and a lower b.

Based on the above associations, low-level enhancement of rainfall through

warm rain processes should generally act to increase A and decrease b from the Z-R relationship representative of the resolution volume observed aloft by radar. However, estimates of N_0 and μ would be needed to derive a dynamic Z-R for which A and b could be calculated directly based on this connection to the DSD. DSD parameters cannot be retrieved from just the horizontally polarized radar parameters (reflectivity, velocity, and spectrum width), so developing a rain rate algorithm that is adaptable to warm rain enhancement for the single-polarization NEXRAD would require indirect estimation of N_0 and μ.

At high rain rates (R >= 50 mm/hr), the relationship between rainfall rate and DSD parameters (and hence, reflectivity) is less variable (Smith et al. 2009). Several studies suggested that an equilibrium DSD is achieved at the point where the largest drops are of sufficient size to experience spontaneous breakup and there is an approximate balance between coalescence, collision, breakup, condensation, and evaporation (Atlas and Ulbrich 2000; Atlas and Williams 2003; Tokay and Short 1996; List et al. 1987; List 1988; Hu and Srivastava 1995; Rosenfeld and Ulbrich 2003; Uijlenhoet et al. 2003; Bringi et al. 2003; Prat and Barros 2009). Under equilibrium conditions, the median drop diameter D_0 and DSD slope parameter remain approximately constant, and therefore the reflectivity Z is linearly dependent on the drop number concentration alone (i.e., the b parameter of the Z-R power law is 1.0 and the coefficient A varies only with N_0). Higher values of b in the power law tend to correspond to greater variability of the median drop diameter D_0 or the inclusion of multiple different precipitation regimes in the Z-R regression dataset (Atlas and Ulbrich 2000). Figure 2.1 illustrates the difference between the initial DSD within the cloud

(dashed line), the equilibrium DSD structure within the cloud (solid line), and the equilibrium DSD at the surface (dotted line). The shift to larger drops between the cloud and the surface is supported by radar studies that found the maximum observed reflectivity in tropical convection, both maritime and continental, to be at the lowest elevation (Szoke et al. 1986; Tokay et al. 1999; Atlas and Ulbrich 2000; Atlas and Williams 2003), which may introduce rainfall underestimation in locations where the lowest radar tilt is several kilometers above the surface.

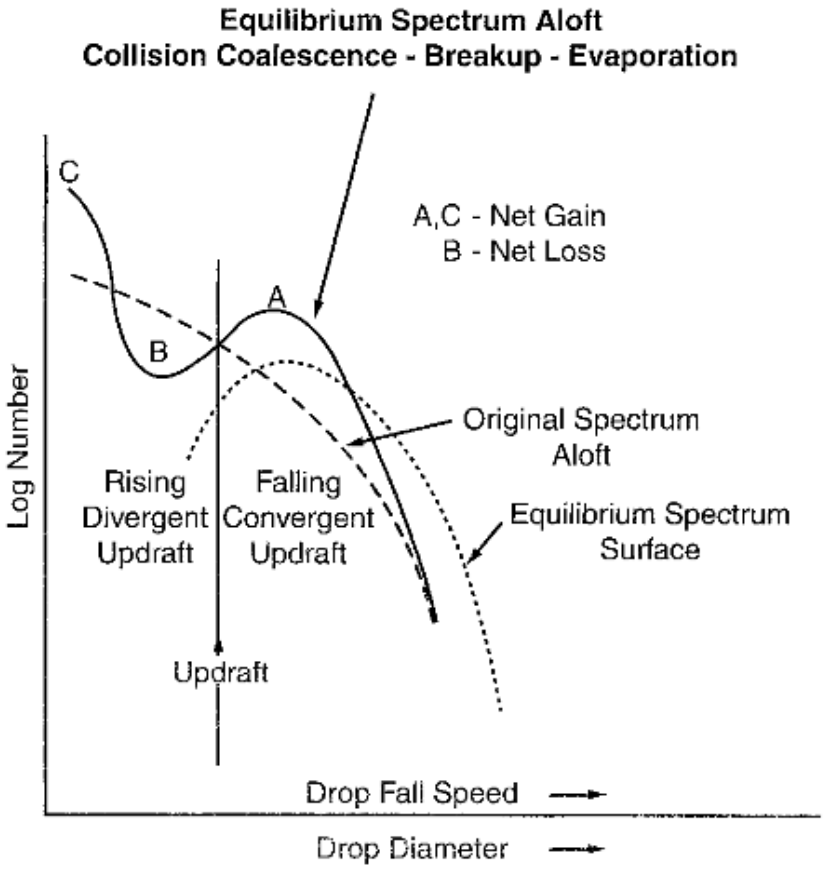


Figure 2.1. Conceptual model of the equilibrium DSD both at the surface (dotted line) and within the cloud (solid line), as well as the DSD observed initially during the growth phase of the storm (dashed line). The local maximum at A is the result of coalescence growth, the minimum at B is the result of drop breakup and loss of small drops through collection, and the maximum at C is the contribution of smaller drops from the breakup process. From Atlas and Ulbrich (2000).

Cold Rain

Cold rain processes require ice particles to be present in the cloud, though this does not represent every cloud that extends above the 0° C level of the atmosphere. Rogers and Yau (1993) noted that the cloud top temperature generally needs to be below -5° C for any ice to form and that all storms with cloud top temperatures colder than -20° C have ice processes occurring. This goes against general wisdom that water will freeze at any temperature below 0° C, which is only true when the water is in contact with an object that has a subfreezing temperature. Because clouds consist of scattered small, liquid drops, each drop must come into contact with an ice nucleus before it will begin to freeze (pure water will not homogeneously freeze until it reaches temperatures of -40° C). Additionally, latent heat release still occurs while drops and ice crystals grow by condensation and sublimation, and the updraft continues to loft liquid drops upward into subfreezing levels. Thus, even at temperatures of -20° C, liquid drops can still be present in the cloud.

Once ice crystals exist in a cloud, they will grow through similar processes as warm rain: deposition (direct transfer of water vapor to ice under supersaturated conditions) and coalescence/collision, which is generally referred to as aggregation for frozen hydrometeors. Also similarly to warm rain, the ice crystals eventually reach a critical point where their terminal velocities exceed the updraft strength and they begin to fall. There are qualities of ice crystals that make these processes more complex than what would occur for liquid drops (e.g., interaction of ice particles with both ice and liquid drops and the wide variety of shapes ice can assume through deposition), but for the purpose of a general review, the overall mechanisms are similar. The presence of ice

within a cloud does not preclude the presence of warm rain processes, however. Growth by condensation, collision, and coalescence still occurs within the warm layer and can still be the dominant growth mechanism where the "balance" between the updraft velocity and the drop terminal fall velocity exists below the freezing level (Atlas and Williams 2003). Davis (2004) attributed the dominance of warm rain processes in continental rainfall to the presence of a "warm coalescence layer" (i.e., the depth of the layer between the cloud base and the freezing level) of at least 3 km, precipitable water above 38 mm (1.5 inches), and dewpoint temperatures greater than 15 C and 25 C at the surface and 850 hPa, respectively. Additionally, moderate values of convective available potential energy or CAPE (1500-2000 J/kg) through a deep layer of the troposphere tend to be more conducive for warm rain growth than large values of CAPE because the weaker updraft extends the period of time that drops are suspended below the freezing level.

Estimation of Rainfall by Radar

Precipitation is detected by weather radars as the amount of power scattered back to the receiver by a collection of hydrometeors within the area of the radar's electromagnetic beam. The power received along the full length of each beam or radial is then segregated into bins of fixed width based on the distance of the targets, which is based on the time required for the energy to reach the targets and return to the radar at approximately the speed of light. Thus, the basic physical unit of radar-derived rainfall is a volume of hydrometeors at a known (or approximately calculated) height above ground. No direct data can be obtained regarding the sizes or number of drops in that

volume using single-polarization radars, but the intensity of returned power can be empirically associated with rainfall rate through the use of the Z-R power law relationship. As noted earlier, the idealized relationship between radar reflectivity and rain rate can be analytically described by DSD model moments and parameters, but computing that exact relationship for any single-polarization radar volume requires additional information. Because that additional information is generally not available for operational purposes, statistical fits between radar reflectivity and surface rain rate from rain gauges have been employed to approximate the Z-R relationship.

Many Z-R relationships have been derived since the earliest days of weather radar research. Battan (1973) provided several that were each derived from different rainfall events and types (Fig. 2.2), and hundreds of additional equations have been derived since. Z-Rs are most often computed using either a linear least squares regression fit on log-transformed reflectivity and rainfall rate or a nonlinear regression on the data in linear space. More recently, additional methods have been proposed, including probability matching (Rosenfeld et al. 1993; Rosenfeld et al. 1994) and artificial neural networks (Xiao and Chandrasekar 1997; Orlandini and Morlini 2000; Liu et al. 2001; Hessami et al. 2003; Xu and Chandrasekar 2005; Chiang et al. 2007; Root et al. 2010). Despite all the different methods that exist for converting Z to R, the relationship between them tends to be most appropriate for the particular rainfall event or regime from which the input data was collected, and applying them generally to other cases often introduces large biases and variance.

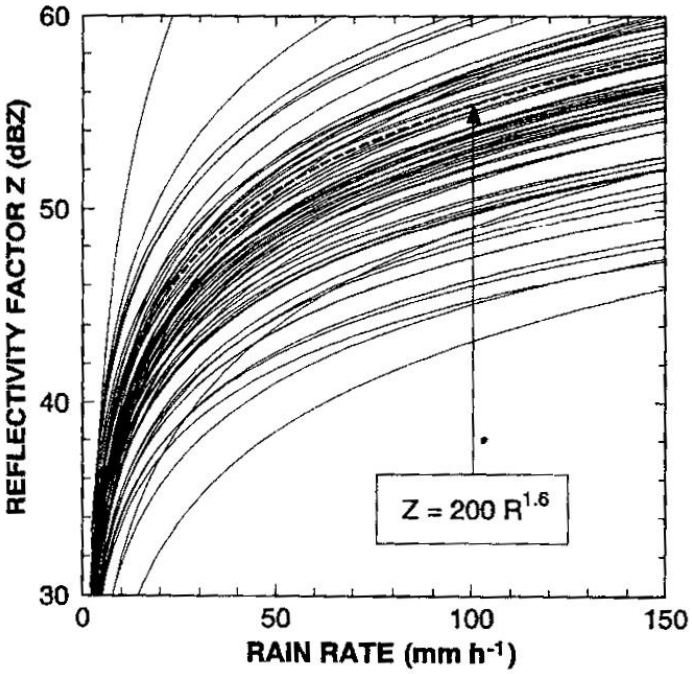


Figure 2.2. Comparison of the Marshall-Palmer Z-R function to many other empirically derived Z-R functions (from Doviak and Zrnic 1996).

One method of mitigating the large variability of the Z-R relationship with storm type is to segregate the precipitation into different rainfall types using observable three-dimensional reflectivity and environment characteristics. Multiple Z-Rs can then be used to compute rain rates appropriate for each storm type. The most common delineation used in the literature is between stratiform and convective rain. They can be easily distinguished based on their reflectivity characteristics alone (Biggerstaff and Listemaa 2000, and references therein; Amitai 2000), and the dynamics and microphysical growth differences of each tend to produce very different DSDs, and therefore different rainfall rates, from comparable values of reflectivity. While segregating into two general precipitation types can reduce errors and can easily be implemented in a real-time system with automated techniques, many situations still exist where both Z-R relationships provide a poor match to the "true" rainfall rates

observed at the surface by rain gauges and disdrometers. For example, Zawadzki (2006) showed large differences between the "stratiform" Z-R relationships derived from stratiform rain in a system undergoing extratropical transition, the trailing stratiform region of a mesoscale convective system, and broad stratiform associated with large-scale frontal uplift. In addition to the Z-Rs being different from each other, none of the Z-Rs derived by Zawadzki followed the classic Marshall-Palmer Z-R that is broadly applied as the default “stratiform” relationship. High variability in DSDs and density of rain drops has also been observed within a single stratiform event (Lee et al. 2009), though it's typically described as the most homogeneous type of rain observed in nature (Rosenfeld et al. 1995).

Deviations of observed rainfall from the prevalently used stratiform and convective Z-R relationships have led to additional precipitation types being proposed. The Precipitation Processing System (PPS) developed for operational rainfall estimation from the WSR-88D network allows custom Z-Rs to be used in regions where the default relationships have poor performance (Fulton et al. 1998). Examples include the use of a tropical Z-R in coastal areas (Davis 2004; Wood 1997) and orographic Z-Rs in areas with precipitation strongly influenced by complex terrain. Studies have also argued that a “transition” rainfall rate is needed for the region between convective and stratiform rain in mesoscale convective systems (e.g., Atlas and Ulbrich 2000). Because the transition area is characterized by low reflectivity and light rainfall accumulations, however, use of the stratiform or convective Z-R relationship does not seem to be produce large systematic errors (Uijlenhoet et al. 2003b).

The “tropical” Z-R is probably the most widely used alternative Z-R relationship

to the Marshall-Palmer and NWS default (i.e., convective) in operational and real-time Z-R-based rainfall estimation products. It was initially derived by Austin and Geotis (1979) using disdrometers on ships and DSDs collected by aircraft during the Global Atmospheric Research Program's Atlantic Tropical Experiment (GATE). A similar Z-R function was then chosen by Fulton et al. (1998) as a good alternative for WSR-88D radars located in maritime environments, and Xu et al. (2008) went on to apply Austin's Z-R in the NMQ system to rainfall with vertical profiles of reflectivity (VPR) characteristic of tropical environments. Its usefulness lies in its ability to provide more accurate rainfall estimates in heavy rainfall situations where the stratiform and convective Z-Rs exhibit large underestimation. This is particularly important in the context of issuing warnings for flash floods, because the National Weather Service's Flash Flood Monitoring and Prediction system bases flash flood likelihood on basin-accumulated rainfall from radar (or, alternatively, from River Forecast Center rainfall products with a time delay). Significant underestimation of rainfall could result in missed events or lower lead-time on warnings.

There is some uncertainty regarding how representative the tropical Z-R relationship is for midlatitude, continental heavy rainfall events. Though it seems to provide better rainfall estimates where the other Z-Rs underestimate, is it right for the wrong reason (i.e., correcting a systematic underestimation bias rather than being applied in situations with DSDs similar to the one for which it was derived)? Answering this question requires an examination of both the DSDs in maritime and continental environments, and the characteristic properties of the heavy rainfall events in midlatitudes where the tropical Z-R is often applied.

Differences between Maritime and Continental Rainfall

Convection in maritime and continental airmasses have some characteristic differences that affect the dynamics and microphysics of rainfall development. The most consistently observed difference is the updraft strengths of cells in each environment. Maritime airmasses tend to have deep warm layers with moderate or "skinny" CAPE evenly distributed throughout the entire layer (Fig. 2.3). Continental airmasses, in comparison, tend to have shallower warm layers with steeper temperature lapse rates in the lower levels (Szoke et al. 1986). This larger concentration of CAPE in a shorter layer of the troposphere is the reason for the more explosive nature of convection and much stronger updrafts. Maritime convection and tropical cyclones both tend to have weaker updrafts (Jorgenson et al. 1985; Szoke et al. 1986; Zipser and Lutz 1994; Rosenfeld et al. 1995). As a result, hail is rarely observed in maritime convection and reflectivity decreases rapidly with height above the freezing level. The weaker updrafts also extend the residence time of drops within the warm layer, which is why highly productive warm rain processes are more often found in maritime convection than in convection of continental origin.

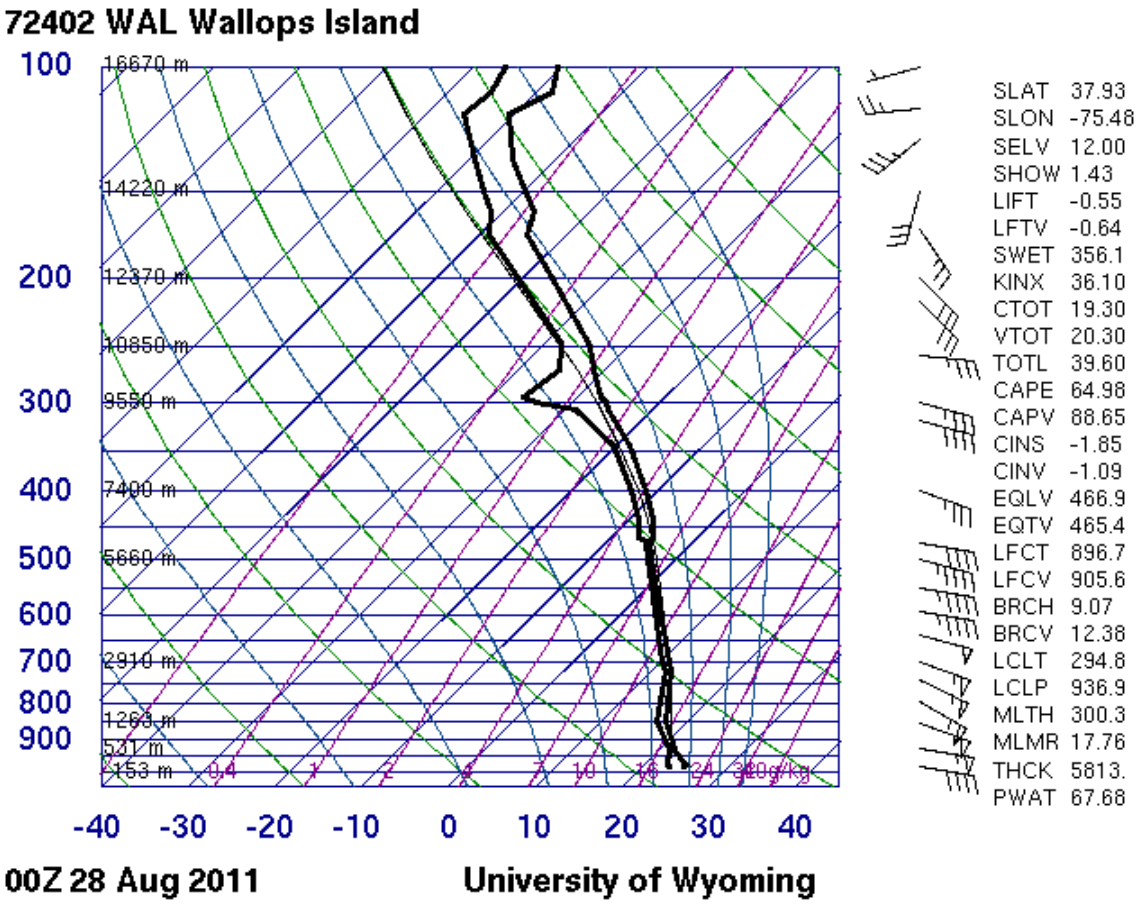


Figure 2.3. Observed sounding from Wallops Island, Virginia, during Hurricane Irene.

Continental and maritime environments produce significantly different DSD signatures due to differences in both the concentration of cloud condensation nuclei and the strength of updrafts, so reflectivity retrieved from rainfall of maritime origin may not relate to rainfall rate the same as rainfall of continental origin (Rosenfeld and Ulbrich 2003; Bringi et al. 2003; Ulbrich and Atlas 2007; Rosenfeld et al. 1995). However, the Z-Rs in the different environments tend to converge for very high rainfall rates, presumably due to the onset of equilibrium DSD processes (Uijlenhoet et al. 2003a; Bringi et al. 2003; Willis and Tattelman 1989; Blanchard and Spencer 1970). The rainfall rates do not converge to a single Z-R relationship, though. Rather, the

nonlinear power laws converge to a linear function for which the reflectivity and rain rate are only dependent on the concentration of drops in the radar sample volume, which is equivalent to setting the exponent of a Z-R function constant and only varying the coefficient (Uijlenhoet 2003a).

The structure of the vertical profile of reflectivity (VPR) has also been used to segregate continental and tropical rainfall (Xu et al. 2008). Because rain formed in a tropical environment tends to have a deeper warm cloud layer where warm rain processes dominate, the additional growth of drops falling through that cloud layer leads to a profile below the freezing level where reflectivity increases with decreasing height. Thus, the maximum reflectivity in the profile is often observed at the lowest tilt of the radar. By contrast, midlatitude continental VPRs often have a maximum reflectivity that is elevated above the lowest level, indicating either that hydrometeor growth is primarily occurring in the ice phase or that evaporation both below the cloud layer and from entrainment of dry air into the warm layer are more significant than growth through collision and coalescence. However, this alone does not appear to be a perfect discriminator between the two environments for convective rainfall, because the structure of the VPR below the freezing level is also strongly dependent on the stage of convective development in individual cells (Szoke et al. 1986; Bringi and Chandrasekar 2001). Convective cells tend to have maximum reflectivity at the lowest tilts at the end of their life cycle because they are “raining out” (i.e., the downdraft dominates to the point where the updraft is no longer lofting large hydrometeors into the mid and upper-levels of the storm), whereas cells in earlier stages often have elevated reflectivity maxima because the updraft keeps all but the largest drops suspended in the cloud.

A more consistent difference between maritime and continental convective VPRs is the rapid decrease in reflectivity with height above the freezing level (Zipser and Lutz 1994). Figure 2.4 clearly shows the difference in the reflectivity gradients in tropical maritime and midlatitude continental convection. However, the reflectivity gradients observed from the tropical continental convection seem to occupy the space of both the maritime and continental clusters. Furthermore, while this could be a useful discriminator between convection originating in the different airmasses, it requires that the mean VPRs are derived from convective rain only. Stratiform rain also has a characteristic rapid decrease of reflectivity above the freezing level, so additional criteria would be needed for the classification, such as detection of the brightband for stratiform rain (e.g., Zhang et al. 2008). This methodology could not be applied to warm rain growth in stratiform rain alone.

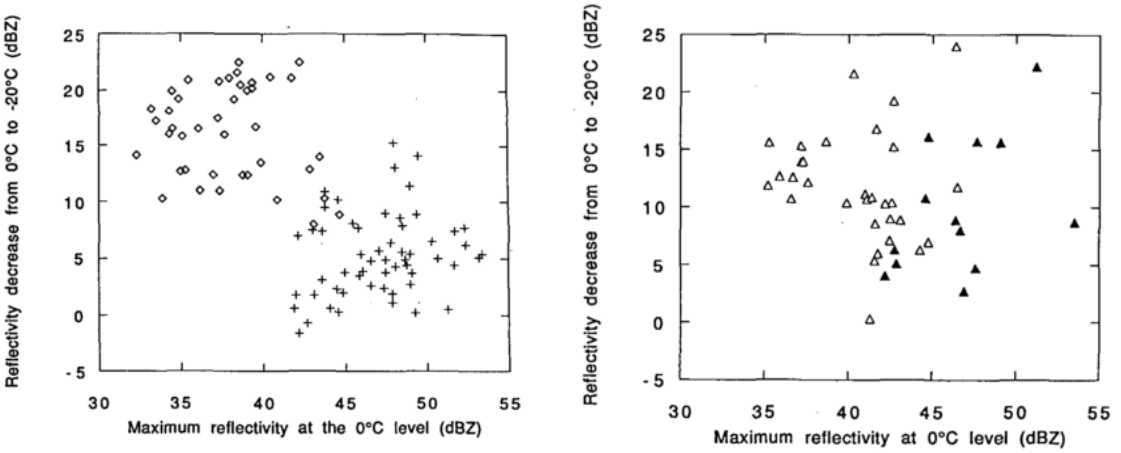


Figure 2.4. Characteristics of reflectivity above the freezing level in convection of tropical oceanic (diamonds, left), midlatitude continental (crosses, left), and tropical continental (right) origin (from Zipser and Lutz 1994).

While few studies have examined the effect of warm rain growth processes on the reflectivity structure of stratiform and convective rain in midlatitude continental environments characteristic of the United States, it is not unusual for "tropical"

environment characteristics to be present in the midlatitudes during the warm season as a result of strong moisture and temperature advection from maritime airmasses. In fact, tropical airmasses have been cited in several U.S. flash flooding case studies as a contributor to the formation and duration of intense rainfall (discussed in the next section). Therefore, it may be appropriate to apply Z-R relationships derived from maritime, tropical airmasses to midlatitude, continental regions if the environmental conditions warrant their use.

Climatology of Heavy Rainfall Events in the United States

A wealth of literature has been published on the occurrence of heavy or extreme rainfall in the United States, covering a broad range of topics. These topics include regional frequency analysis (Brooks and Stensrud 2000; Schumacher and Johnson 2006), climatology studies of favorable environments, synoptic patterns, and storm modes (Maddox et al. 1979; Funk 1991; Bradley and Smith 1994; Konrad 1997; Junker et al. 1999; Rogash 2003; Konrad and Perry 2009), ingredients-based flash flood forecasting methods (Doswell et al. 1996; Johnson and Moser 1992), and numerous flash flooding case studies (Maddox et al. 1978 is perhaps the most well-known). The motivation for most of the studies centered around either better forecasting of future flash floods or, in the non-meteorological literature, design of engineering flood control structures based on return periods of heavy rainfall events. Comparatively few studies focus on the delineation of excessive rainfall rate areas in near-real time (one notable example is a study by Noel and Dobur (2004)). In much of the literature related to flash flooding, emphasis has been placed on rainfall accumulation (often daily or longer)

rather than rainfall intensity at shorter time scales. In general it appears that this focus is primarily due to the limitations of the available data, especially of environmental variables that could only be computed from radiosondes or from numerical model analyses that up until recently were only available at 6-12 hour intervals.

In general, the key “ingredients” that need to be in place for very heavy rainfall to occur are high atmospheric water vapor content, sufficient lift or forcing for initiation, and moderate convective instability. Regardless of the specific geographic location, storm mode, or synoptic pattern, these three fundamental pieces must come together for extremely high rain rates. The resulting onset of flash flooding can also depend on other factors such as antecedent soil moisture, soil permeability, land surface topography, and storm motion, but they are all beyond the scope of this study.

Maddox et al. (1979) examined the characteristics of 151 flash flood-producing rainfall events of non-tropical origin across the United States and found them all to be associated with high surface dewpoints, high moisture content through a deep layer of the troposphere, and weak to moderate vertical wind shear. However, as shown in Fig. 2.5, the environmental characteristics of the location of heaviest rainfall may not be consistent with the airmass feeding the storms, particularly at the surface. For example, storms producing elevated heavy rainfall on the cool side of a surface warm front draw their inflow from the warm sector, so the conditions at the surface (i.e., at the rain gauge) will not provide much information about the thermodynamics needed for high rainfall rates. Analyzing environmental parameters from higher levels over a gauge may be more representative of the characteristics of the environment from which the hydrometeors are forming, though it's important to remember that parameters such as

CAPE and precipitable water (PW) may be affected by the presence of convection and precipitation. Therefore, even for automated real-time applications, attempts should be made to characterize the thermodynamics of the parcels being drawn into the updraft and contributing to the formation of precipitation rather than the thermodynamics of the environment where the rain is falling.

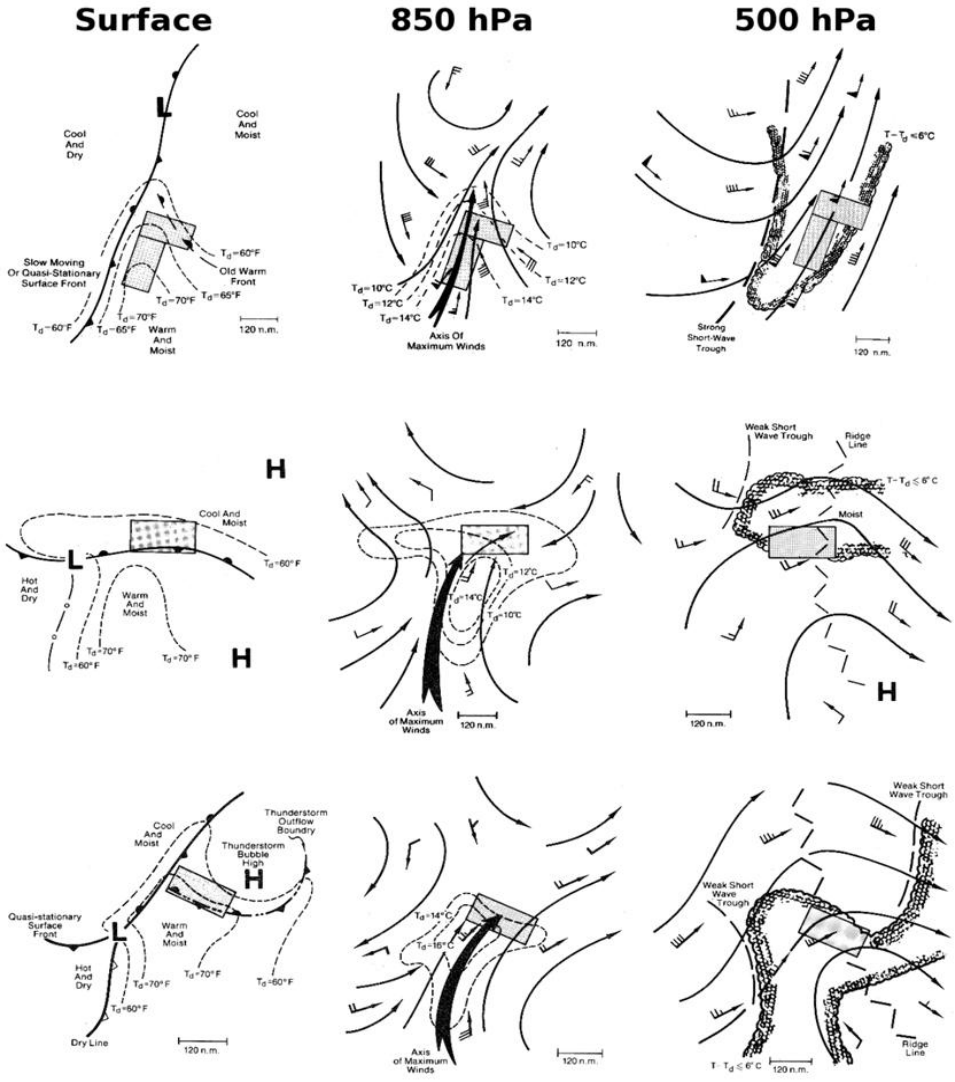


Figure 2.5. Synoptic (top), frontal (middle), and mesohigh (bottom) setups that are favorable for flash flood-producing rainfall (adapted from Maddox et al. 1979). Locations of maximum rainfall are outlined by shaded boxes.

Bradley and Smith (1994) conducted a climatological analysis of extreme rainfall events (daily accumulations greater than 125 mm) over Oklahoma by examining characteristics of radiosondes in the pre-storm environment, as well as surface maps and rain gauge data. While the CAPE values retrieved from the radiosondes indicate that some of the data may have been taken from airmasses already modified by convection (a pitfall of using such a sparsely distributed network with data available only twice daily), the PW values were consistently high for every event. Their seasonal analysis of the PW relative to the climatological mean demonstrated that the classic rule of thumb of examining PW in terms of percent of normal may not always be reliable, particularly during the summer season when average PW is sufficient to support large rainfall accumulations. The authors suggested instead that a fixed threshold of PW year-round may be more appropriate (in combination with sufficient forcing and instability), though it's possible that the threshold value may vary by region. Johnson and Moser (1992) developed a decision tree for forecasting flash floods in Louisiana using fixed thresholds of various parameters related to low and mid-level moisture, temperature advection, lift, and instability. Most notably, their thresholds for PW tended to be higher than the 25 mm value suggested by Bradley and Smith, with values below 32 mm (1.25 in.) considered to be low-risk for flash flooding. Also, no seasonal changes in the thresholds were mentioned.

Junker et al. (1999) analyzed the synoptic and dynamic features of several MCS events from the “Great Midwest Flood of 1993”, and four of the five main factors related to scale and intensity of rainfall in their study were associated with environmental moisture. Namely, the factors were 1000-500 hPa relative humidity, areal

extent of moisture advecting toward the forcing boundary (area seemed to correlate more with rainfall than magnitude), strength of low-level moisture convergence, and extent of low-level moisture convergence upstream of the heaviest rainfall (important for supporting storm propagation and redevelopment). Because their study analyzed total rainfall accumulation rather than rainfall rate, however, emphasis was placed on factors supporting sustained heavy rainfall over a single location for a long period of time rather than factors related to extreme rain rates over short periods of time.

Konrad (1997) used initialized fields of several thermodynamic parameters from a numerical model (the Limited-Area Fine Mesh model on a 55 km grid) to characterize the environments associated with heavy rainfall events across the southeast United States. To best capture the conditions within the pre-storm environment, a linear interpolation scheme was used on the 00Z and 12Z analysis fields to estimate the parameter values two hours prior to the onset of heavy rainfall for each event. Heavy rainfall events were separated by type of synoptic pattern from which they developed, because the authors felt that stronger statistical relationships between the parameters would be obtained for groups of storms with similar dynamic characteristics (much like Maddox et al.'s (1979) separation between synoptic, frontal, and mesohigh setups). Analysis of the environment associated with each synoptic pattern showed that heavy rainfall events from all five patterns were associated with high moisture content (relative to the mean) and ridging of the moisture fields in the vicinity of the heaviest rainfall. However, the authors went on to state that while these features were present in all of the heavy rainfall events, the magnitude of the moisture variables did not seem to correlate with rainfall intensity. Furthermore, relationships between excessive rainfall

and the ambient environment were strongly dependent on the synoptic pattern, and none of the parameters effectively delineated the heavy rainfall area for all events.

Konrad and Perry (2009) examined the factors that led some hurricanes to produce excessive rainfall accumulations while other hurricanes of similar intensity did not. While the correlations between the environmental parameters and the daily rainfall totals were weak (less than 0.5), statistically significant relationships with rainfall totals were nevertheless found for the areal size of the tropical cyclone, the area over which PW was greater than 2 inches (50 mm), and the area over which both PW > 2 inches and divergence at 200 hPa was present. Mean 850 hPa moisture flux, mean PW, and mean wind speeds within the tropical cyclone produced weaker, yet still statistically significant, correlations.

Funk (1991) summarized the techniques used for forecasting heavy convective rainfall by the Hydrometeorological Prediction Center (known as the Forecast Branch of the National Meteorological Center prior to the reorganization of the National Weather Service in the mid-1990s). Most of the techniques were related to pattern recognition of dynamic and thermodynamic features around baroclinic forcing boundaries, such as the location of the 850 hPa theta-e ridge axis, diffluence of 1000-500 hPa thickness, and the ridge axis of minimum 500 hPa absolute vorticity. However, Funk also listed several rules of thumb indicating that high moisture content can enhance precipitation in patterns that are otherwise not conducive for heavy rainfall.

Though pattern recognition of synoptic setups and forcing mechanisms serve a forecaster well to predict the occurrence of heavy rainfall, those features are not well-suited to an automated, real-time environment and are not as useful once the rainfall

event is already underway. Additionally, while Z-R relationships can approximate intense rainfall rates from observations of high reflectivity, the magnitude of the intensity is often underestimated. Therefore, for nowcasting of flash floods based on rainfall accumulations in basins such as in the FFMP system, there needs to be a way to delineate where systematic underestimation is likely to occur and to compensate for it with either a different reflectivity-rain rate relationship or a local bias adjustment.

Use of Rain Gauges to Validate and Adjust Radar-based Rainfall Estimates

Rain gauges have historically been the primary tool used to verify the accuracy of radar-based rainfall estimates. Many past studies have addressed the uncertainties associated with such a comparison.

Zawadzki (1975) investigated the spatiotemporal correlation between rainfall rates from an idealized radar and rain gauge (i.e., no error or bias such that differences are only a function of the sampling properties of each sensor). He found that autocorrelation of rain gauge measurements tends to decline exponentially in space and that the average or "smoothed" rainfall rates detected by a radar's beam tend to introduce randomly scattered errors when compared to instantaneous point rainfall rate measurements. The smoothing also introduced a systematic "bias" in the radar data for which locally high rainfall rates under the sampling volume would be underestimated and locally low rainfall rates would be overestimated. Integration of the gauge data over space and time tended to reduce the random errors and better match the gauge estimates to those estimated by radar.

Kitchen and Blackall (1992) conducted a similar analysis on high-resolution data

rather than using a statistical model for the rain rate distribution. They showed that even with a 2 minute accumulation period, the rain gauges could only capture about 50% of the temporal variability of the rain rate. When the gauges were sampled every 15 minutes to match the update period of the radar, representativeness errors were much larger. Coarser spatial sampling tended to average out the small-scale temporal errors, while spatial errors between the radar and point gauges increased. Longer accumulation periods reduced both spatial sampling errors and temporal sampling errors. Similar scaling relationships were found by Seo and Krajewski (2010) and Villarini and Krajewski (2008) for spatial scales of 0.5-8 km and temporal scales from 5 minutes to several days. Kitchen and Blackall also found that as much as 80% of the variance from hourly accumulations in a single convective rainfall event were the result of subpixel variability. Ciach and Krajewski (1999a) even showed that the uncertainties associated with the point-to-area comparison between gauges and radar can often be larger than the difference in rain rates between the two sensors, making objective comparison difficult. This is particularly true at short time scales (less than one hour), but continues to be an issue for accumulation periods as long as four days. Anagnostou et al. (1999) attributed up to 60% of the radar-rain gauge differences in their study to subgrid rainfall variability for hourly accumulations on a 2-km radar grid, which is comparable to the scales used in the NMQ system. The subpixel-scale uncertainty also seems to vary with season, rainfall type, and the density and location of gauges within individual radar pixels (Zhang et al. 2007; Villarini et al. 2008; Krajewski et al. 2003).

Habib et al. (2004) proposed a distribution-based method for filtering the representativeness errors out of comparisons between radar and gauge rainfall. Their

"conditional transformation" method was able to estimate the distribution of "true" areal rainfall over an area of similar scale to that sampled by radar, eliminating the need for a point-area comparison with individual gauges and the large uncertainties associated with the sampling differences. However, computation of the representativeness errors require a priori knowledge of the spatial correlation structure of rainfall at subpixel scales, which tend not to be readily available when using operational gauge datasets. Furthermore, the method is best suited for analysis over small regions or individual storm events where rain rates are statistically homogeneous.

A natural next step from the previous work showing large variability between gauges and radar is to examine what effect that variability has on the derivation of Z-R relationships, which are widely used in operational settings to obtain real-time rainfall rates and accumulations. Morin et al. (2003) examined the effect the spatial and temporal variability of rain rate has on the derivation of Z-R relationships for high-intensity convective rainfall. They found a scale dependence both in space and time of the Z-R parameters *A* and *b*, though the authors argued that *A* was inversely dependent on *b* (Fig. 2.6). As the spatial resolution was decreased from 1 km to 5 km, the exponent factor *b* decreased by approximately 15%, and as the accumulation period was increased from 5 min. to 120 min., a slightly larger decrease in *b* was observed. The largest changes were observed at the smallest temporal and spatial scales. This result highlights the importance of knowing how a particular Z-R relationship was derived before applying it deterministically to a different radar dataset. For example, the often-used Marshall-Palmer relationship (Marshall et al. 1947; Marshall and Palmer 1948) was derived using radar data with approximately a 1 km^2 spatial resolution (8 degree

beamwidth at a distance of 8.8 km from the radar) and 1-min. rain rate data from rain gauges. The reflectivity *Z* was also computed at the surface using dyed paper to capture drop sizes at comparable time intervals for derivation of the DSDs. While Morin et al.'s study did not include rainfall accumulation of less than 5 min., their results nevertheless suggest that the Z-R relationship derived from 1 min. data would be different from the Z-R relationship derived using 5 min. accumulations, and the variation of the parameters appear to be highest at short time scales.

Villarini and Krajewski (2010) attempted to estimate the systematic and random error components separately for hourly rainfall accumulations over Oklahoma using the three Z-R relationships most widely used by the National Weather Service: Marshall-Palmer (i.e., stratiform), NWS (convective), and tropical. While the systematic bias varied significantly between the different Z-R derived rainfall totals, they found that the spatial and temporal error characteristics of the random component varied independently of the rainfall rate parameterization (different Z-Rs as well as use of AP and clutter-removal algorithms).

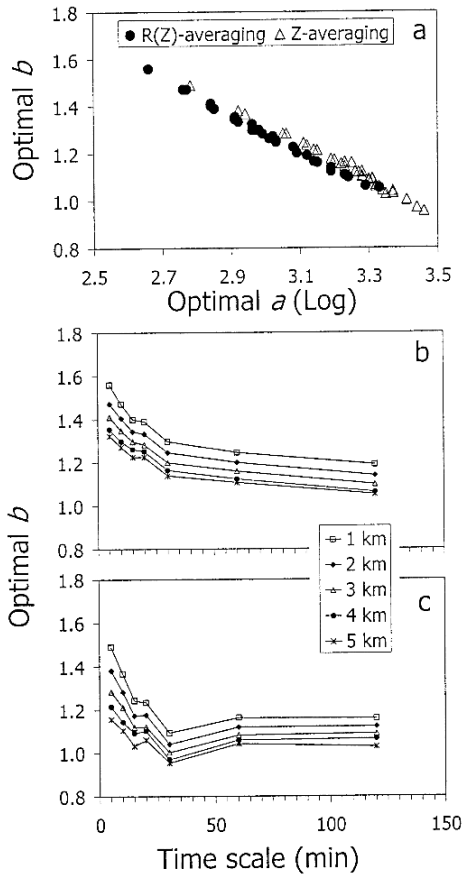


Figure 2.6. Variation of the *A* and *b* parameters of the Z-R relationship on temporal and spatial scale differences (from Morin et al. 2003)

Ciach and Krajewski (1999a) argued that the exponent *b* of a Z-R relationship (and thus, the coefficient *A* as well) may also vary depending on which method was used to derive it, even when using the same data. Using an idealized statistical model of rain rate and reflectivity variance, they analytically derived three different equations for a quantity β based on three well-established methods of deriving Z-R: direct nonlinear regression, reverse nonlinear regression, and the probability matching method (Rosenfeld et al. 1994). The parameter β is a function of the *b* exponent from the true or "physical" Z-R dependency and various error terms associated with radar and gauge measurement errors.

A confounding issue related to the comparison of radar and gauge rainfall is the quality control of the "ground truth." Numerous studies have addressed the various uncertainties of measuring rainfall with rain gauges, including low spatial density, undercatchment due to high winds and biological interference, false tips from melting ice, drizzle, birds, rodents, and insects, and general mechanical malfunctions (Ciach 2003; Vasiloff et al. 2009; Tokay et al. 2010). Krajewski et al. (2003) argued that any gauge network design that is based on the use of single gauges at each location rather than dense clusters of gauges is useless for rainfall spatial variability and decorrelation analysis, because it is impossible to discern good quality measurements from erroneous ones. Tokay et al. (2010) disagreed, however, and stated that single gauges of good quality could still be used for research applications, though they also stressed the value of having multiple, collocated gauges. While having collocated gauges may mitigate a wide range of quality control issues, it still may not be sufficient to capture the rainfall rate variability. Ciach (2003) showed that even with a network of 15 research-quality collocated gauges, large variability of instantaneous rainfall rate was still observed, particularly for light rain (Fig. 2.7). Despite all the possible sources of error and variability, however, rain gauge networks (most with single gauges only) continue to be the most prevalent tools for radar rainfall verification, because no other observing platform currently exists at the ground surface with a sufficient density to capture the spatiotemporal variability of rainfall accurately.

Good quality control is of utmost importance when using gauges, especially on short time scales with combined datasets of multiple gauge networks and sensor types,

such as the NWS's Hydrometeorological Automated Data System (HADS) that is used for gauge bias adjustment in the NMQ system. Tokay et al. (2010) compared gauges from several of the operational networks included in the HADS dataset and found data quality to vary significantly not only between the different networks, but within networks as well. This variability in data quality should be explicitly addressed when using HADS for radar rainfall verification or adjustment, and low quality sites should be identified and removed. Unfortunately it is very difficult to determine quality of individual gauges especially for large datasets, but automated, multisensor quality control methods (e.g., Kondragunta and Shrestha 2006) can aide in filtering out outliers both in space and magnitude, and past work comparing the accuracy of different observing networks (e.g., Tokay et al. 2010) can identify problems with specific sensor types, reporting methods, or maintenance practices.

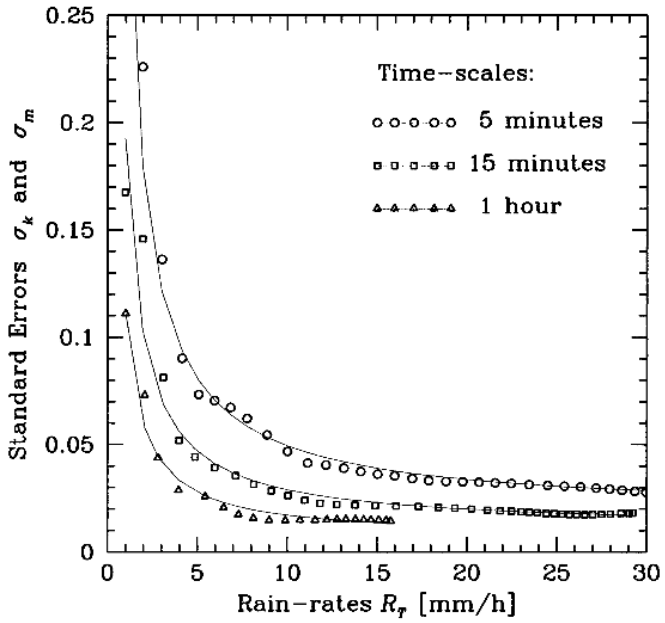


Figure 2.7. Standard deviations observed from 15 collocated, identical rain gauges in Oklahoma (from Ciach 2003).

CHAPTER 3: DATA

The three primary data sources for this study are rainfall estimates and other products derived from radar reflectivity, environmental analysis from a numerical weather model, and hourly accumulations of rainfall from rain gauges. Each dataset will be described in detail in the following subsections.

Radar-derived Fields: The NMQ System

The National Mosaic and Next Generation Quantitative Precipitation Estimate system (NMQ or Q2) is a testbed across the contiguous United States (CONUS) for real-time QPE and short-term QPF analysis and verification (Zhang et al. 2011; Vasiloff et al. 2007). The backbone of NMQ is the Multi-Radar Multi-Sensor system (MRMS), which ingests, quality controls, and mosaics together all available radar data across the CONUS in real time (updating every five minutes). It also generates a suite of single-radar and multi-radar products through additional processing of both the radar data and hourly analysis fields from the 20-km RUC model. The MRMS algorithms are a combination of code written specifically for NMQ and radar processing algorithms available in the Warning Decision Support System - Integrated Information (hereafter called WDSS-II) software environment (Lakshmanan et al. 2007a). The automated steps for generating the NMQ products are illustrated in Figure 3.1 and are described in the following steps.

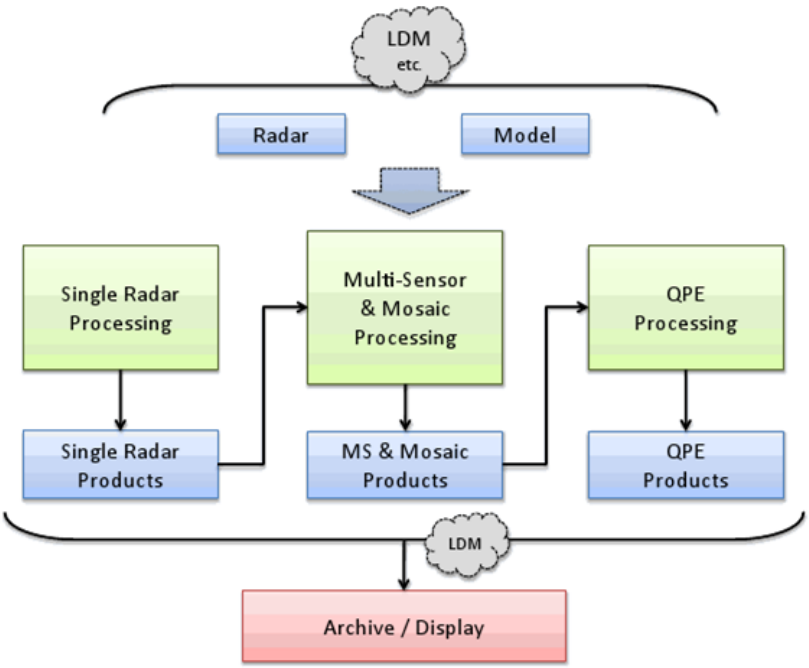


Figure 3.1. Flowchart of the NMQ rainfall product generation process.

1. Raw Data Processing. All subsequent steps require that the input data be in one of three data formats, depending on the algorithm. The formats are NetCDF, an NMQ-specific binary format, and XML. Neither the raw radar nor the RUC data are in one of those formats when retrieved via the Local Data Manager (LDM) software, so they must be converted. All of the data conversion (gribToNetcdf, ldm2netcdf, and w2radarenv) is done in WDSS-II.

2. Remapping of Model Analysis Fields. The 20-km RUC's horizontal resolution and projection are not the same as the NMQ domain, so the data must be resampled. For the national domain in the real-time system, the model data resolution is resampled to 10 km using a nearest neighbor method and are projected onto a latitude-longitude grid. New products are also derived in this step that are based on the model fields, such as

temperature profiles at each radar location and the surface wet bulb temperature. The full list of model analysis parameters used for generation of the rainfall products includes:

a. Height of all vertical pressure levels

b. Temperature at all vertical levels, including surface

c. Height of the 0° C and -10° C temperature surfaces

d. Surface dewpoint temperature

e. Surface wet bulb temperature

All fields are retrieved from the isobaric vertical coordinate analysis, so vertical levels are defined by 25 hPa pressure intervals.

3. Single-Radar Data Quality Control (w2qcnn). w2qcnn is a neural network-based WDSS-II quality control algorithm that checks for the presence of ground clutter or anomalous propagation (AP) signatures in the reflectivity and velocity fields of a single radar (see Lakshmanan et al. 2007b for a more detailed description). Bins flagged as ground clutter or AP are masked from the dataset.

4. Single-Radar Vertical Profiles of Reflectivity and Delineation of Stratiform vs. Convective Rainfall. VPRs are needed in later steps for identification and removal of brightband contamination, identification of tropical rainfall, and correction of reflectivity where range and orographic effects are present. The VPR algorithm computes three different VPR types for each radar: stratiform, convective, and a global mean. In order to separate stratiform and convective VPRs, the algorithm uses three

threshold values of reflectivity. If the VPR at any lat/lon location contains either reflectivity above 50 dBZ or a reflectivity value at or above the -10° C temperature level exceeding 30 dBZ, the VPR is flagged as convective. All other VPRs with reflectivity exceeding 15 dBZ are then flagged as stratiform. Mean VPRs for each precipitation type are then computed in an annulus region of 20-80 km in range from the radar location, and a global VPR is computed from both rainfall types together over the same area. Linear interpolation is then used between the elevation angles to derive reflectivity values at 200-meter intervals from 400 meters above ground to the highest available level, and the final three VPRs are written to a text file. New VPRs and precipitation flags are computed for every five-minute update, and an additional "average" VPR is computed using a one-hour running mean of the 5-minute VPRs.

5. Bright band Identification and Analysis. Estimation of the bright band height is based on a combination of the 20-km RUC model's analysis of the 0° C height and the stratiform VPRs derived in the previous step (see Zhang et al. 2008 for a more detailed description). The RUC analysis serves as a background field or first guess. The algorithm then searches downward through the stratiform VPR starting at 500 meters above the RUC's 0° C height level for a local maximum in the reflectivity. If a local maximum is found, the top and the bottom of the bright band layer are computed based on the level at which reflectivity decreases from the maximum by a user-defined percentage (Fig 3.2). If the vertical distance between the top and bottom layers are less than 1.5 km or if the vertical distance from each is less than 1.0 km from the maximum reflectivity level, then the VPR is flagged as containing a bright band.

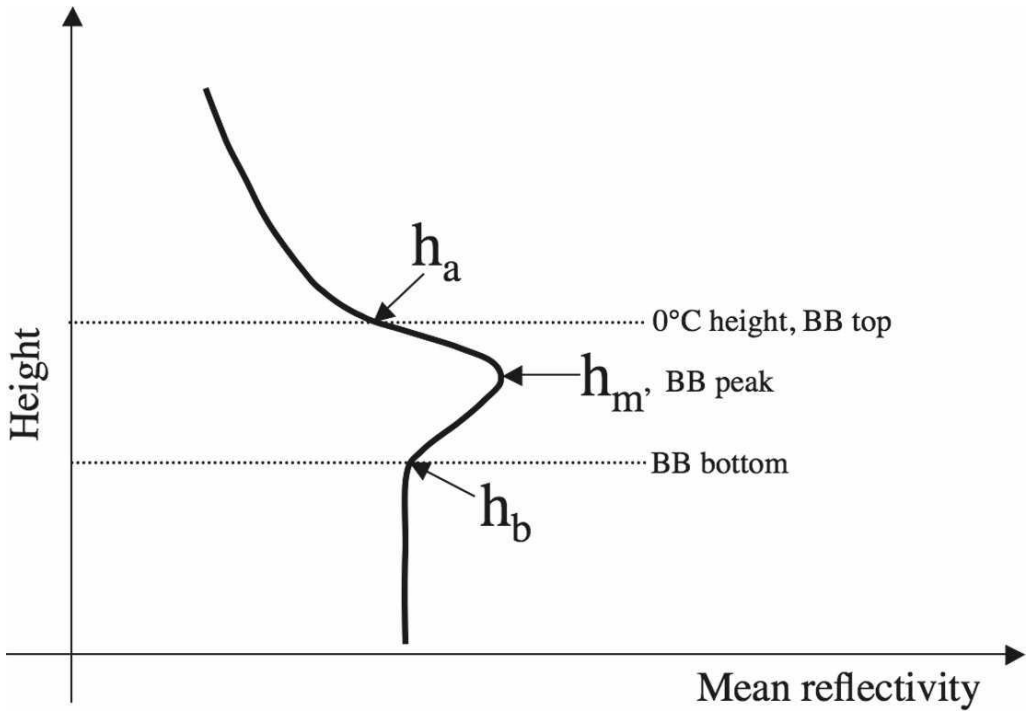


Figure 3.2. Conceptual illustration of a vertical profile of reflectivity exhibiting a brightband structure. The top, bottom, and peak of the brightband or melting layer are computed automatically by the NMQ brightband identification algorthm (from Zhang et al. 2008).

6. Tropical Rainfall Identification. After the VPRs are generated for every radar in the CONUS domain, the tropical identification algorithm checks the reflectivity profiles below the freezing level for increasing reflectivity with decreasing height toward the surface (see Xu et al. 2008 for a more detailed description). The reflectivity gradient is indicative of significant warm rain microphysical growth processes that are often found in rainfall from tropical environments. If the reflectivity slope in the warm layer is either constant or positive toward the surface, the radar is flagged as containing tropical rainfall rates (Fig. 3.3).

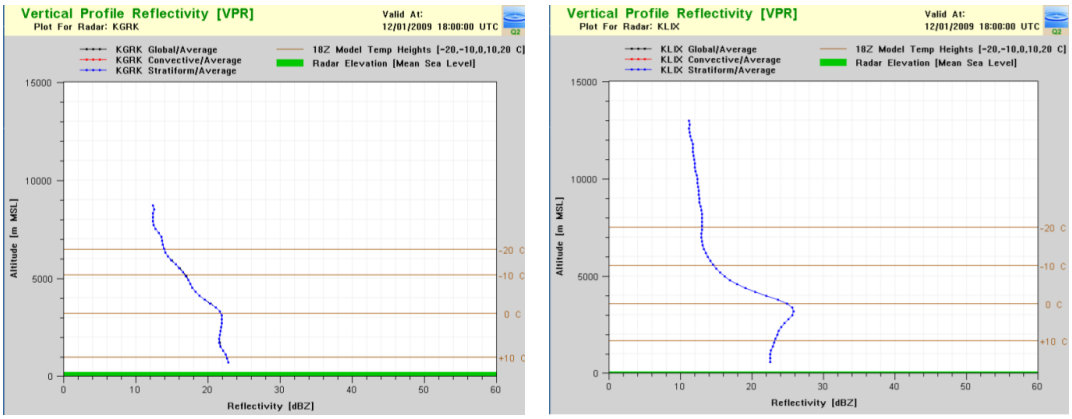


Figure 3.3. Comparison of VPR profiles characteristic of tropical (left) and non-tropical (right) environments.

7. Remapping of Radar Data (w2merger). All remaining steps in the NMQ product generation process require the data to be remapped to a common, cartesian grid. For the single radar data fields, this means resampling and interpolating the reflectivity from the radar's native polar coordinates and elevation angles to a three-dimensional, rectangular grid with a 0.01 degree horizontal resolution (in latitude/longitude coordinates) and 31 vertical levels. The remapping method is a combination of interpolation between radar tilts in vertical space and nearest neighbor resampling in azimuth and range (Zhang et al. 2005). The interpolation and resampling are needed in order to fill in gaps between radar tilts in the various volume coverage patterns and to smooth discontinuities that are introduced when different elevation angles are mapped to a single horizontal height level (Fig. 3.4).

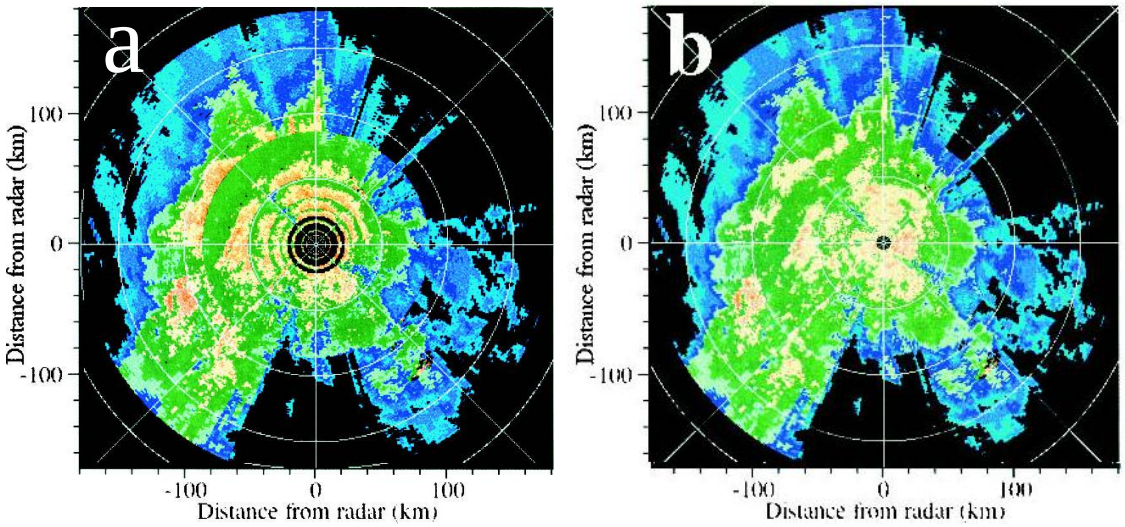


Figure 3.4. A horizontal cross-section of reflectivity from a stratiform rain event before and after the resampling and interpolation processing (from Zhang et al. 2005).

8. Mosaic of Single Radars Across CONUS. Once all the single-radar processing is completed and both the radar and model-derived fields are in the same cartesian coordinate system, the individual radars can be mosaicked into a single, CONUS-wide domain and the final products can be generated. The mosaic algorithm (see Zhang et al. 2005) uses a distance weighted mean function where multiple radars overlap the same location, which preserves the fine-scale details of the reflectivity field by emphasizing radars with the lowest amount of beam spreading with range. Additionally, VPR correction techniques are used where range from the nearest radar still prevents sampling of the lowest levels, which are most important for rainfall estimation.

The three-dimensional CONUS reflectivity product is then used to create a suite of products related to rainfall type, probability of the presence of hail, liquid water content, and heights of the lowest and highest available reflectivity levels. It is at this stage that the tropical precipitation type is delineated in the gridded rainfall type product. The

tropical Z-R is assigned anywhere that is within a user-defined radius from a radar flagged as tropical (generally between 120 and 200 km) where reflectivity exceeds 30 dBZ. The 30 dBZ threshold was chosen due to the nonlinear tendency of the Z-R function to assign very similar rainfall rates for areas of low reflectivity. The rainfall rates diverge at around the 30 dBZ level, so every pixel below that is left as stratiform (Xu et al. 2008).

9. Precipitation Accumulation. Z-R relationships are assigned to the different delineated precipitation types and 5-minute rainfall rates are computed from the hybrid-scan reflectivity field. The rainfall rates for each 5-minute period are then compiled into accumulations for 1, 3, 6, 24, and 72-hour periods. The 1-hour and 3-hour accumulations are updated every 5 minutes in the real-time system, and the longer accumulations are updated hourly. A local bias correction is applied to the precipitation totals using the Hydrometeorological Automated Data System (HADS) rain gauge dataset, which is discussed further below. Rainfall data on the cartesian grid are matched to individual rain gauges using an inverse distance weighting approach. Because this study uses the HADS gauges as a validation dataset, the Q2 rainfall product used in the analysis is the hourly accumulation of rainfall *prior* to gauge adjustment.

Environmental Variables: The Rapid Update Cycle Model

The Rapid Update Cycle, or RUC, model was an operational short-range numerical weather prediction system that is run at hourly intervals for forecasts up to 12

hours into the future. The primary advantage of using the RUC for this study is that the initial objectively analyzed data fields are updated every hour, rather than at 6-hr or 12-hr intervals typical of longer range models such as the Global Forecast System (GFS), the European Centre's Medium-Range Weather Forecast (ECMWF) model, and others. The horizontal spatial resolution of the RUC has evolved over the years from 60 km in 1994, to 40 km in 1998, to 20 km in 2002, and finally to 13 km in 2006 (Benjamin et al. 2004a). The NMQ system, however, still uses the 20 km version of the analysis for generation of its rainfall products. When the 13 km RUC became operational, additional updates were made to the data assimilation and processing scheme of the model, and the 20 km RUC was modified to take advantage of these changes such that the analysis on the 20 km grid was simply an aggregation and resampling of the 13 km product. All RUC data evaluated in this study are from cases that occurred after that change.

The RUC analyzes data on two vertical coordinate systems: isobaric and "native", which is a hydrostatic, hybrid isentropic-sigma, terrain-following coordinate system (Benjamin et al. 2004b). A 3DVAR-like data assimilation scheme is used to integrate new observations into the analysis fields for the hourly updates. The 1-hour forecast from the previous hour's run is used as the background analysis field, and it is subtracted from the current hour's observed data to obtain the forecast error. This error field is then added to the background to generate the new analysis (see Benjamin et al. 2004a for a more detailed description of the data assimilation process). The data assimilated into each analysis field are generally taken within 30 minutes of the analysis valid time. The only exception is for rawinsondes, which can be received anywhere from 45 minutes before to 50 minutes after their associated valid time (either 00Z or

12Z).

The observational data are quality controlled primarily through “buddy checks” of neighboring observations in the forecast error field. If no nearby observations are present, then parameter-dependent thresholds are in place to flag unusally large errors between the observation and the background field. When analysis values were compared to observations from rawinsonde data by Benjamin et al. (2004a), differences of 2.8-3.8 m/s were found for wind speed (error increasing with height), 0.5-0.9 degrees Celsius for temperature (generally decreasing with height), 6-10 meters for height (increasing with height), and 6.8-9 percent for relative humidity (generally decreasing with height).

Cloud cover in the RUC is based on Geostationary Operational Environmental Satellites or GOES fields and cloud observations from surface observing networks (Weygandt et al. 2006), and the assimilated cloud fields are used to adjust the cloud water and water vapor mixing ratio fields up or down based on the presence or absence of clouds that differs from the 1-hour forecast from the previous run (Benjamin et al. 2002). High temporal frequency PW measurements from global positioning systems (GPS) platforms were assimilated into the RUC starting in 2005. Prior to that model update, PW was computed by vertically integrating the specific humidity at all levels (based mainly on rawinsonde data), as well as through assimilation of PW from the GOES satellites over land (Smith et al. 2007). Radar reflectivity is also assimilated into the RUC and used to adjust the microphysics scheme in the vicinity of observed precipitation, which is an important consideration when using any of the hydrometeor or mixing ratio fields for an environmental analysis. This adjustment based on

reflectivity suggests that the microphysical and moisture parameters in the analysis data will not necessarily be an independent predictor of rainfall properties and instead are influenced by the radar observations. Because of the connection between radar data assimilation and precipitation microphysics in the model, they were not included as predictor parameters for this study.

Rain Gauges: The Hydrometeorological Automated Data System

The Hydrometeorological Automated Data System (HADS) is a real-time data collection and processing system for hourly, automated rain gauges across the United States. The gauges included in HADS come from a large number of networks managed by many entities, including but not limited to the U.S. Geological Survey, the National Oceanic and Atmospheric Administration, the U.S. Department of the Interior, the U.S. Department of Agriculture, the U.S. Army Corps of Engineers, the National Park Service, the Federal Aviation Administration, the Tennessee Valley Authority, and numerous other university, state, federal, and private groups (see http://www.weather.gov/oh/hads/dcp_operators.html for a complete list). The quality control, maintenance, siting practices, and instrumentation accuracy and precision vary significantly across the various rain gauge networks, making quality control of the full dataset a significant challenge. However, because the HADS processing is primarily focused on acquiring and disseminating the data as quickly as possible for real-time users, little post-acquisition quality control is done on the data before it is sent out (Kim et al. 2009). Automated QC algorithms developed for HADS and other real-time datasets generally treat all gauges as equal in terms of precision and accuracy and use

consistency checks both in space and time as well as other observing platforms (e.g., radars, satellites, or other surface observables such as temperature, dewpoint, visibility, or ceiling heights) to identify erroneous measurements (Tollerud et al. 2005; Kondragunta and Shrestha 2006; Kim et al. 2006; Fiebrich et al. 2010). Accuracy differences between gauge networks do exist, though, as Tokay et al. (2010) demonstrated, and the network differences should be accounted for in the QC process wherever feasible.

The quality control procedure for HADS data used in this study closely follows the methods outlined by Kondragunta and Shrestha (2006) with a few modifications. The steps of the procedure are:

1. **Single-gauge sanity check.** If the value reported by the gauge is physically unrealistic, the gauge is rejected. Physically unrealistic values include negative accumulations and extremely high 1-hour accumulations (e.g., greater than 100 mm). Missing gauges are also removed from the dataset in this first step. A list of rejected gauges is generated as a reference, and if at any later step the gauge's value is shown to possibly be valid, it is moved from the rejected list back to the valid list. Negative accumulations obviously would not be validated by additional information, but large accumulations can be if the rainfall is identified as convective in nature.

2. **Multiple-gauge spatial consistency check.** Gauges that survived the single-gauge check step are checked against other nearby gauges using statistical properties of the sample. Starting from one corner of the CONUS domain, a fixed-size rectangular window is moved first horizontally across the domain such that each new window

overlaps half of the previous one, and the same overlap is used when the window moves vertically to the next row. This means that each gauge will undergo the spatial consistency check four times with four different samples, and in order to be rejected it must fail the check all four times. If for any of those sample sets the gauge is not flagged as an outlier, it is kept on the valid list.

The statistical analysis involves identifying gauges as outliers based on an index derived from the median, 25th and 75th quartiles, and the Mean Absolute Deviation of the sample:

$$Index = \frac{|x_i - Q50|}{IQR} \quad \text{for Q75} \neq \text{Q25} \tag{3.1}$$

$$Index = \frac{|x_i - Q50|}{MAD} \quad \text{for Q75} = \text{Q25} \tag{3.2}$$

where

$$MAD = \frac{1}{N}\sum(|x_i - Q50|) \tag{3.3}$$

Q50 = Median

Q25, Q75 = 25th and 75th Quartiles, respectively

IQR = Interquartile Range (Q75-Q25)

N = number of gauges in sample

If the Index value is higher than a user-defined threshold (set to 3.0 for this study), the gauge is defined as an outlier and is moved to the rejected list. Note that both high and low outliers can be identified, so this procedure should identify under-reporting gauges in addition to the over-reporting ones. Selection of the index threshold essentially determines how much of the distribution tails are classified as outliers, with larger

values including more gauges.

3. Convective Rainfall Check. Following the single and multi-gauge checks, the locations of rejected gauges with non-negative accumulations are checked for the presence of convective rainfall. The reasoning for this check is that convective rainfall exhibits large horizontal gradients in rain rates, is associated with heavier rainfall, and tends to have discontinuous spatial distributions. These properties are not well-treated by a spatial consistency check because the rainfall accumulations can vary significantly over short distances. Kondragunta and Shrestha (2006) identified convection based on lightning observations, but there are other ways to identify it with the radar-based data already available through the NMQ system. Precipitation types are identified for assignment of Z-R relationships in NMQ, and one of those precipitation types is convective rainfall. The delineation of convection is based on reflectivity thresholds, and it is mapped to a latitude/longitude grid that can readily be compared to the location of the rejected gauges. Thus, for this study, the rejected gauges were flagged as convective and moved back to the valid list if any NMQ grid point within a 3x3 window around the gauge was flagged as convective during the gauge's accumulation period (i.e., the previous hour).

4. Multi-sensor Check. Malfunctioning rain gauges often either report zero accumulation while rainfall is occurring or non-zero accumulations when no precipitation is nearby, and checking the rain gauge's location against radar reflectivity can help expose these malfunctions. This check requires that the reflectivity data has undergone some quality control of its own to remove nonprecipitation returns such as ground clutter, anomalous propagation, and other artifacts. The hybrid-scan reflectivity

(HSR) product generated within the NMQ system can serve such a purpose, because it is a quality-controlled product. It represents the lowest available elevation angle from the radar, which mitigates the inclusion of virga and other high-level hydrometeors that are not reaching the ground as precipitation. The use of HSR in this study differs from Kondragunta and Shrestha (2006) because they compared the gauge values to radar-derived rainfall estimates and rejected gauges based on a difference threshold. The comparison between HSR and the gauge is a boolean, true/false comparison as opposed to a difference check. If the observed reflectivity within a 3x3 grid point window of the gauge exceeded a minimum, user-defined threshold (set to a high enough level to filter out clear air and drizzle returns, here defined as 10 dBZ) for any update in the gauge's accumulation period, then precipitation was occurring in the gauge's vicinity. The gauge was then either kept or rejected based on the following contingency table:

	Radar Detected Rainfall	**No Radar Detected Rainfall**
Gauge > 0.0	VALID GAUGE	REJECT GAUGE
Gauge = 0.0	REJECT GAUGE	VALID GAUGE

Table 3.1. Contingency table for multisensor quality control of HADS rain gauges.

A boolean comparison of radar reflectivity to the gauge value eliminates any ambiguity that may exist when comparing gauges to radar-derived rainfall estimates that rely on selection of a Z-R relationship.

5. Temporal Consistency Check. Kondragunta and Shrestha (2006) discussed the use of a temporal check for which radar rainfall and gauges are compared over a period of time to determine if a gauge is “stuck” or not based on the time series of the radar's rainfall estimates. However, they also noted that this particular step is much more

computationally intensive than any of the other checks. It relies on the same radar-derived rainfall estimates that the multi-sensor check uses, but because it looks at the change of the rainfall estimate over time it is not sensitive to the choice of Z-R relationship. Because this study is focused on long periods of archived gauge data for the training set rather than real-time gauge QC, a less computationally demanding method was chosen to evaluate the tendency of gauges to be stuck. A running total of the number of failures in each QC category for the previous 24 hours were maintained in a separate text file. If a particular gauge either failed one of the first three checks or was missing for more than 10% of the previous 24 hours of observations, it was added to the rejected list. It then remained on the rejected list until the failure rate improved. Additionally, gauges were set to missing and were rejected for the first hour following a missing observation. This check was added to filter out erroneously high gauge values that were interpreted to be multi-hour accumulations of the missing period but were not flagged as such within the HADS system.

This method of gauge QC is highly heuristic, with many user-defined thresholds included in the analysis. Because the gauges in this analysis are used solely for the purpose of training a statistical model on an archived dataset, a degree of subjective intervention by the human expert seems appropriate in order ensure the best possible quality for the training data. Furthermore, the original purpose of this QC method was that of an initial QC pass of obvious gauge problems before they were further analyzed manually by the forecasters at the NWS River Forecast Centers. It was never intended to be a fully objective process, and the number of uncertainties contained in a compiled dataset like HADS would be difficult to account for in a fully objective approach.

Overall Data Characteristics

The training period for the classification models was hourly observations from 1 April 2010 - 30 September 2010. Hourly data from three additional case studies that did not occur during the training period were also retrieved for independent analysis, and details of those events will be discussed in the next section.

The domain of the analysis was roughly the eastern half of the United States (east of -105 degrees west longitude), which includes 4,084 unique HADS gauge sites (post-QC) and 112 WSR-88D radars (Fig. 3.5). Gauges and radars west of -105 W were excluded for training due mainly to the added uncertainty that lack of radar coverage and beam blockages in complex terrain contribute to radar/gauge comparisons. Because this analysis is focused on heavy rain rates, restricting the domain to the eastern half of the U.S. also makes sense based on the long-term climatology of heavy rainfall occurrence. Brooks and Stensrud (2000) conducted a heavy rainfall frequency analysis for the U.S. and found that rainfall accumulations in excess of an inch (25.4 mm) per hour occurred predominantly east of the front range of the Rocky Mountains (Fig. 3.6).

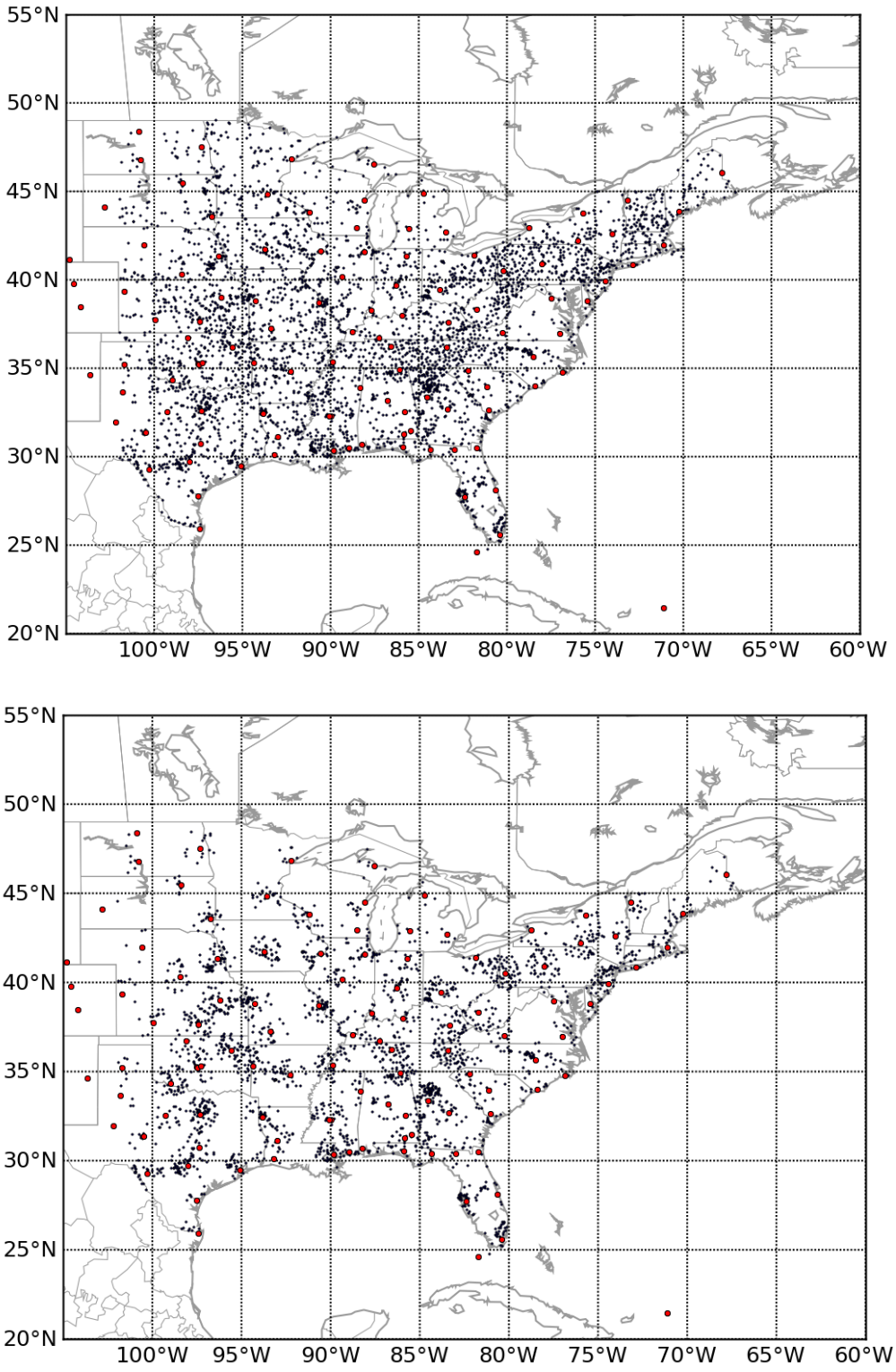


Figure 3.5. Gauges (black points) and WSR-88D radars (red points) located within the study domain. The lower panel shows the subset of gauges located within 100 km of the nearest radar.

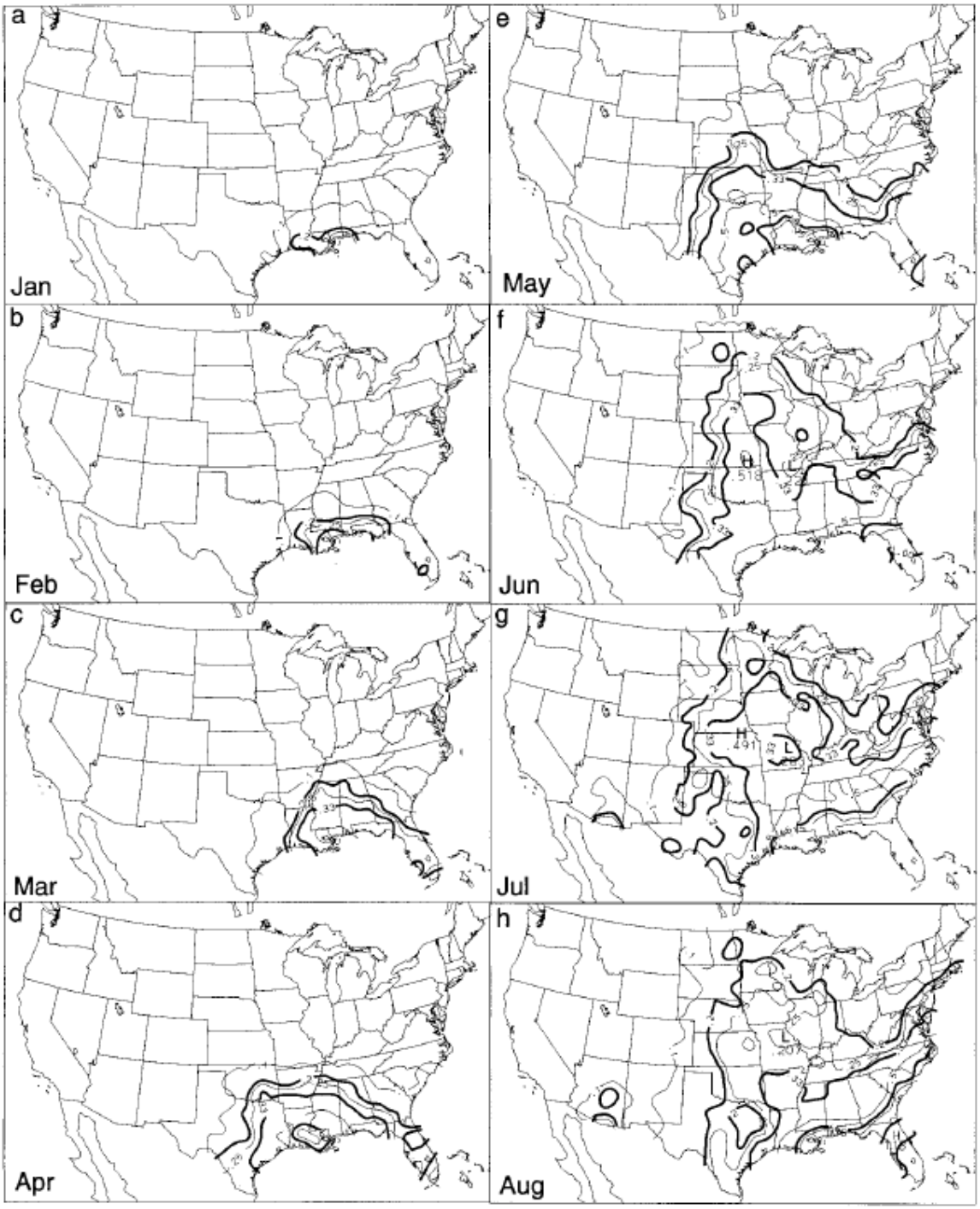


Figure 3.6. Objective analysis of the number of 1 inch per hour rainfall accumulations per year segregated by month for period of 1948-1994 (from Brooks and Stensrud 2000).

For model training purposes, only gauges within 100 km from the nearest radar were retained for analysis to mitigate added uncertainties due to beam broadening and the increased height of the radar beam with range. Of the 2,518 gauges that met the radar distance criterion and passed all QC checks (Fig. 3.5, lower panel), there were over 170,000 nonzero hourly rainfall accumulations recorded within the domain and training period, a number much too large for training of a machine learning algorithm. By restricting the analysis to gauges that received over half an inch (12.7 mm) of rainfall in an hour, the population size decreases substantially to 7,467 hourly gauge observations during the training period. This input data population size is much more manageable for machine learning purposes and appropriately places the focus of the analysis on heavier rainfall events for which flash flooding is of greatest concern.

Overview of Validation Events

Hurricane Irene

Irene made landfall in the United States near Cape Lookout, North Carolina, at 1130Z on 27 August, 2011, as a Category 1 hurricane on the Saffir-Simpson intensity scale (per National Hurricane Center public advisory statements). After crossing the Outer Banks, the center of Irene moved offshore as a weak Category 1 storm before making landfall again near Little Egg Inlet, New Jersey, at 0935Z on 28 August. The center of Irene then moved offshore once again until it made a third landfall over New York City at 1300Z on 28 August as a tropical storm with maximum sustained winds of 65 miles per hour (100 km hr^{-1}). Irene then accelerated northward through New England and became extratropical just before crossing the Canadian border from New

Hampshire.

While the high winds and several tornadoes caused damage to trees and homes in North Carolina and Virginia, the primary impacts of Irene were from flooding, both from the storm surge along the coastline and the heavy inland rainfall. Figure 3.7 shows the storm total rainfall for the entire event, with many areas receiving over 10 inches of rain over the course of just a few days. The storm had a very similar track and evolution to Hurricane Floyd in 1999, which produced catastrophic flooding in North Carolina. Figures 3.8-3.10 compare surface and upper air observations between Irene (left) and Floyd (right) as they made landfall, weakened, and underwent extratropical transition. Atallah and Bosart (2003) analyzed Hurricane Floyd's precipitation distribution and found that rain rates were enhanced on the western side of the storm by mesoscale lift from a pre-existing surface front oriented parallel and to the left of the storm track (Fig. 3.8). Furthermore, as Floyd moved north along the coast, it became embedded in an approaching midlatitude trough (Fig. 3.9), leading to quasi-geostrophic enhancement of lift north and west of the circulation center as a result of absolute vorticity advection by the thermal wind. Prior to that interaction, the maximum of absolute vorticity associated with the tropical cyclone is coincident with the center of the circulation (and the thickness ridge) and is disconnected from the maximum in absolute vorticity associated with the trough (Fig. 3.10 - top panels). Thus, the tropical cyclone has a vertically stacked, equivalent barotropic structure and no advection of the absolute vorticity takes place. As the cyclone interacts with the approaching trough, however, the absolute vorticity maximum becomes increasingly displaced from the thickness ridge and moves into a strongly baroclinic zone characterized by a strong

thickness gradient (Fig. 3.10 - middle and lower panels). The thermal wind along the thickness contours then advects the absolute vorticity, leading to enhanced upward vertical motion as demonstrated by the quasi-geostrophic omega equation (Atallah and Bosart 2003; Holton 2004).

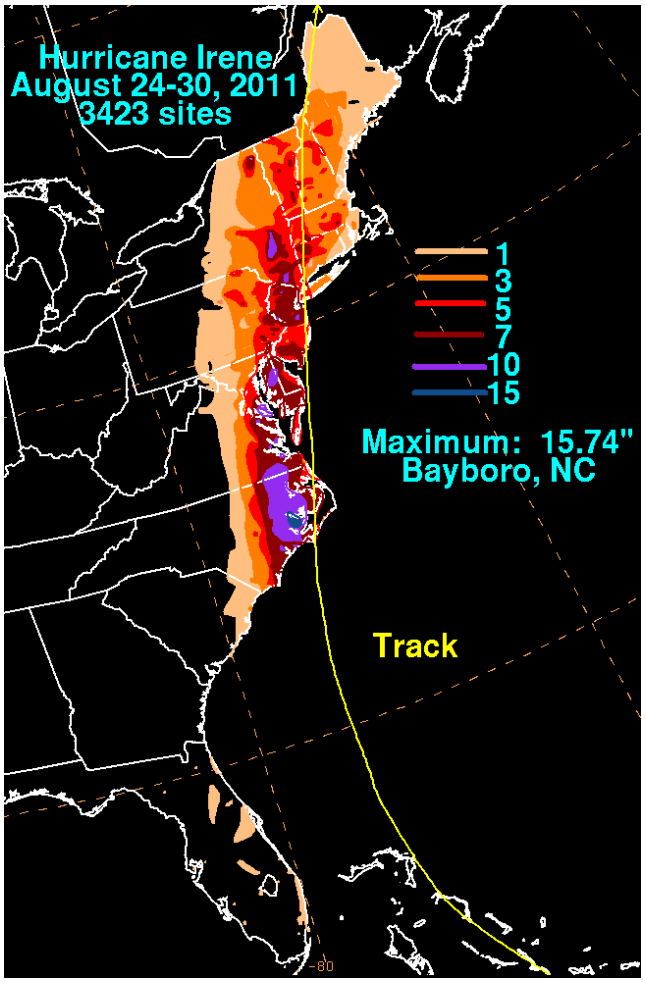


Figure 3.7. Storm total rainfall (inches) from Hurricane Irene (source: National Weather Service Hydrometeorological Prediction Center)

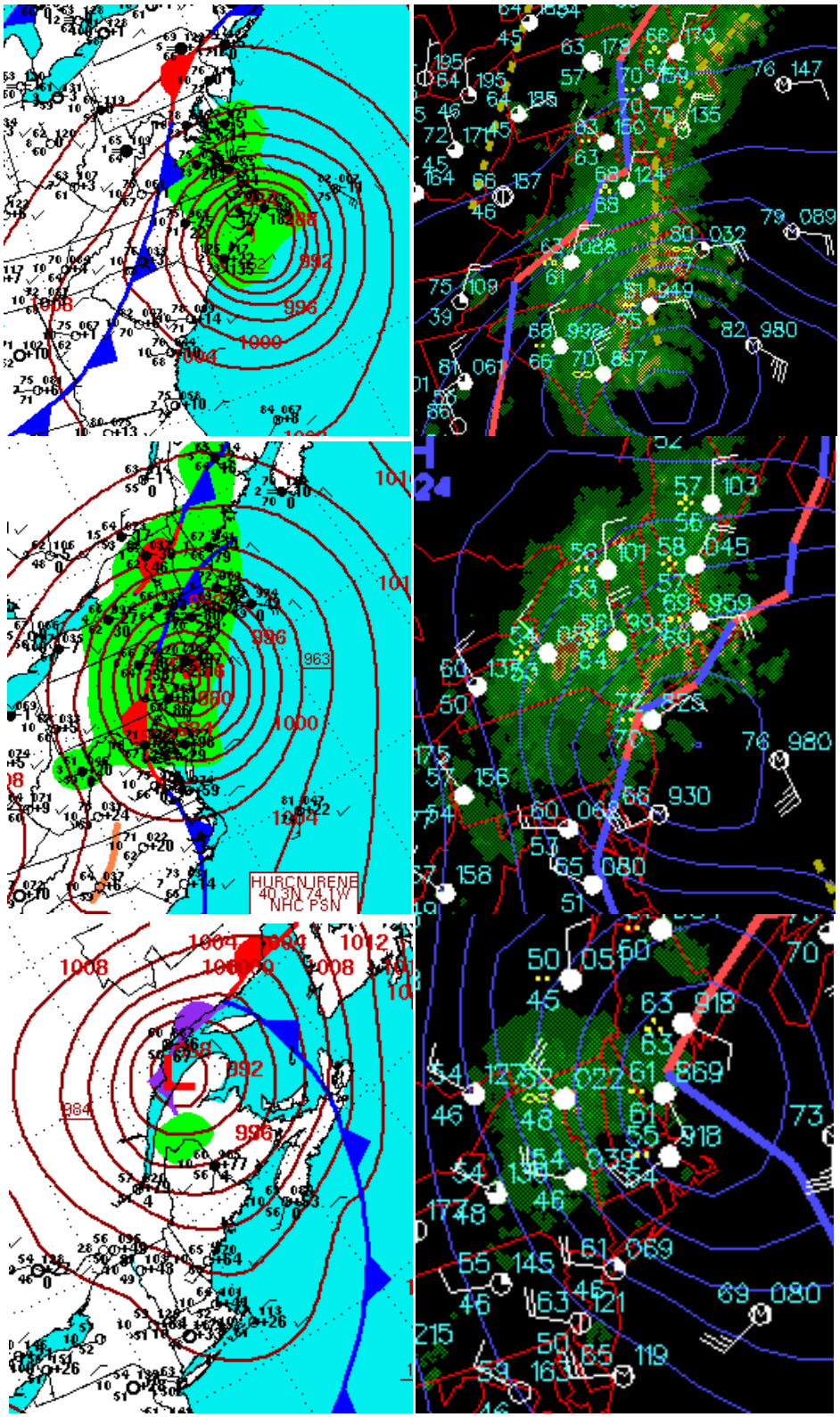


Figure 3.8. Surface station plots and analyzed boundaries for Hurricane Irene (left) and Hurricane Floyd (right).

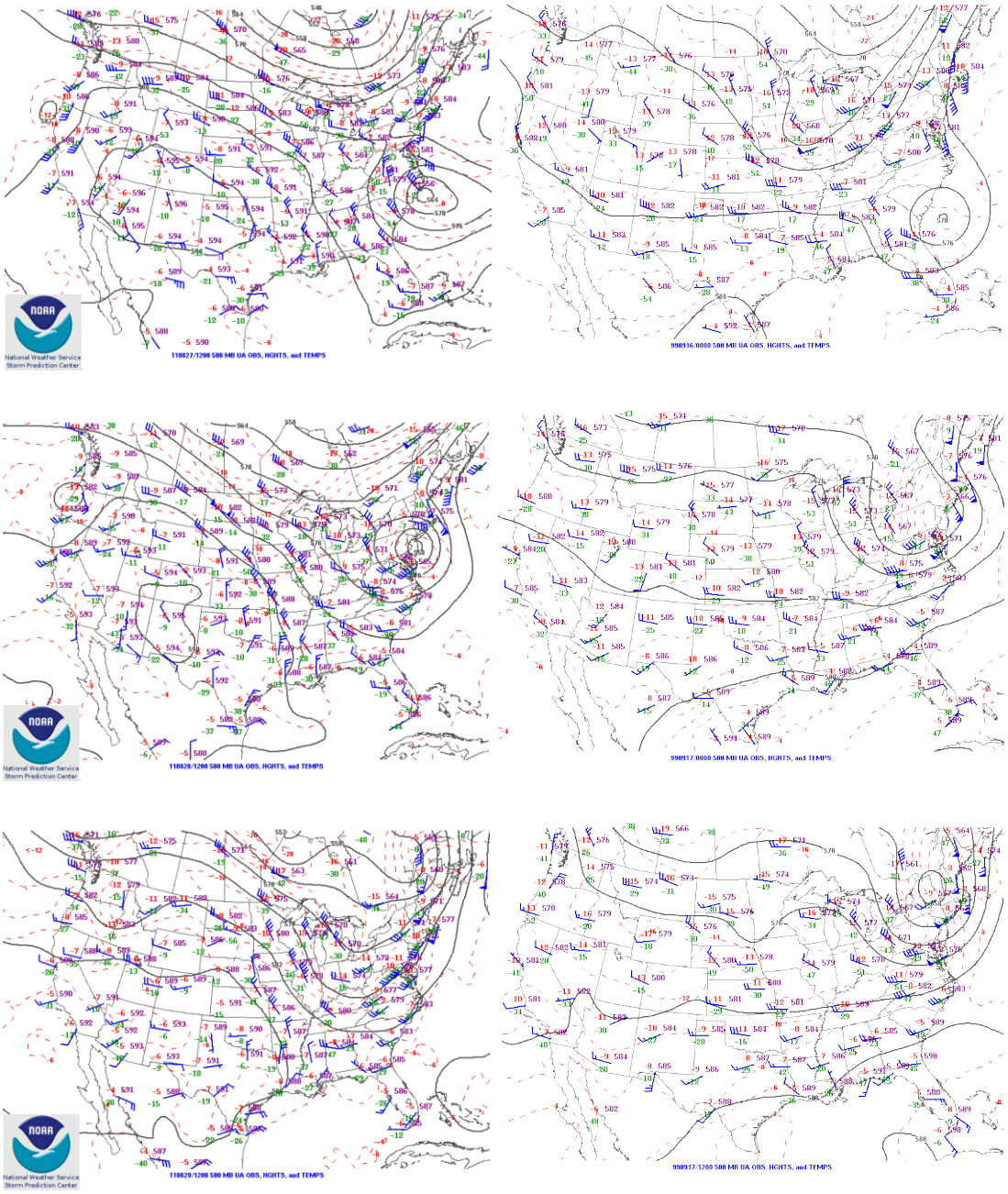


Figure 3.9. 500 hPa Radiosonde station plots and analyzed geopotential height contours for Hurricane Irene (left) and Hurricane Floyd (right).

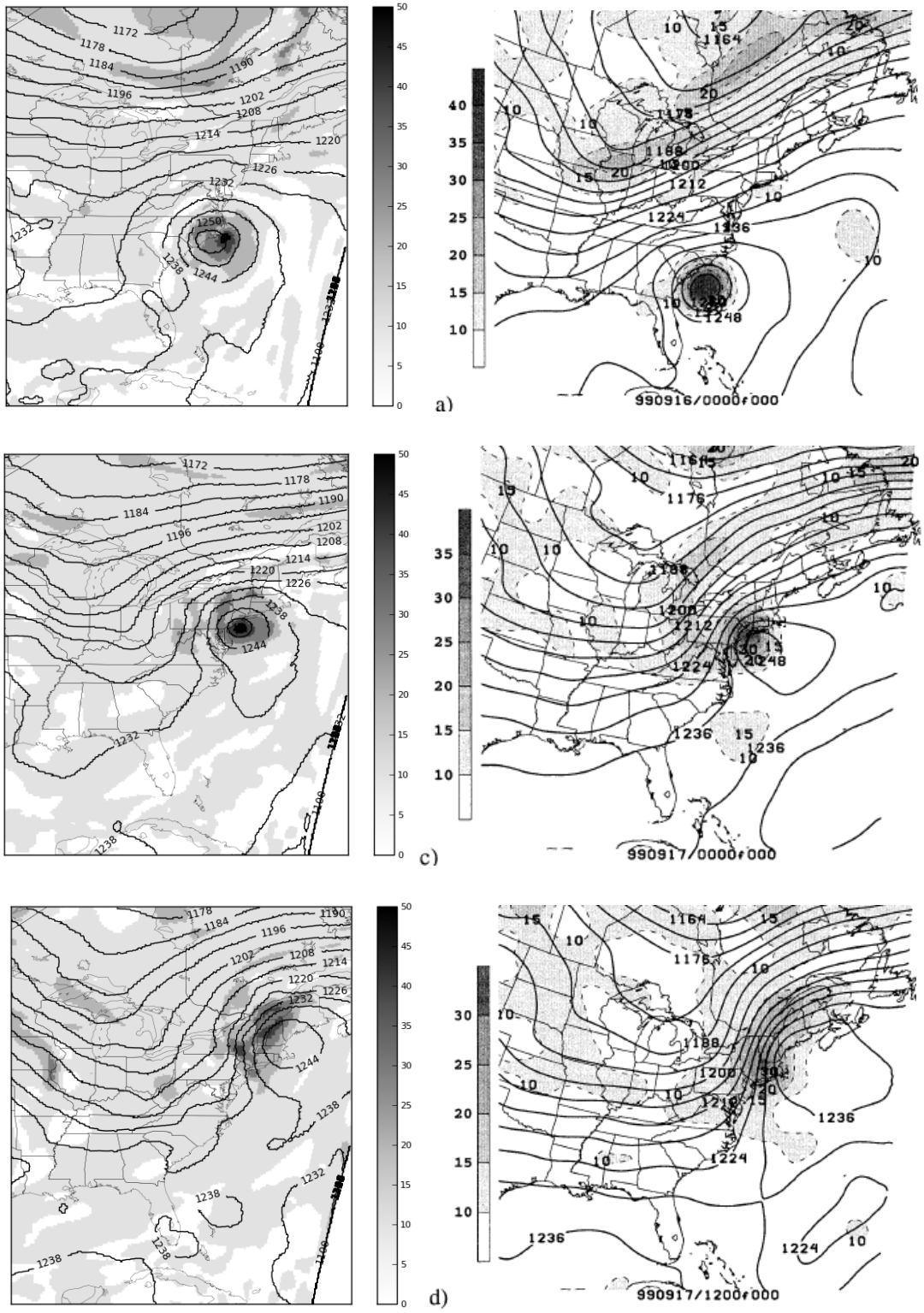


Figure 3.10. 1000-200 hPa thickness (contours, in dm) and 500 hPa absolute vorticity (shaded, in s^{-1}) for Hurricane Irene (left) and Hurricane Floyd (right). The right panels are from Atallah and Bosart (2003), and absolute vorticity is the 700-400 hPa mean.

Rainfall rates prior to the extratropical transition of Irene were not estimated well by the standard Z-R functions used operationally by the National Weather Service. Figs. 3.11 and 3.12 show the bias of the NWS Default (i.e., convective) Z-R and the tropical Z-R relative to hourly rain gauge accumulations from 00Z on 26 August 2011 to 00Z on 29 August 2011. Using the default Z-R would have led to significant underestimation, but even the tropical Z-R underestimated rainfall in many locations by 20 mm or more. The tendency for rain gauges to underestimate rain rate under high wind conditions (e.g., hurricanes) implies that the radar bias could have been even more severe than indicated. Thus, Hurricane Irene represents an ideal case for which environment-based identification of enhanced rain rates can improve remotely-sensed rainfall estimates.

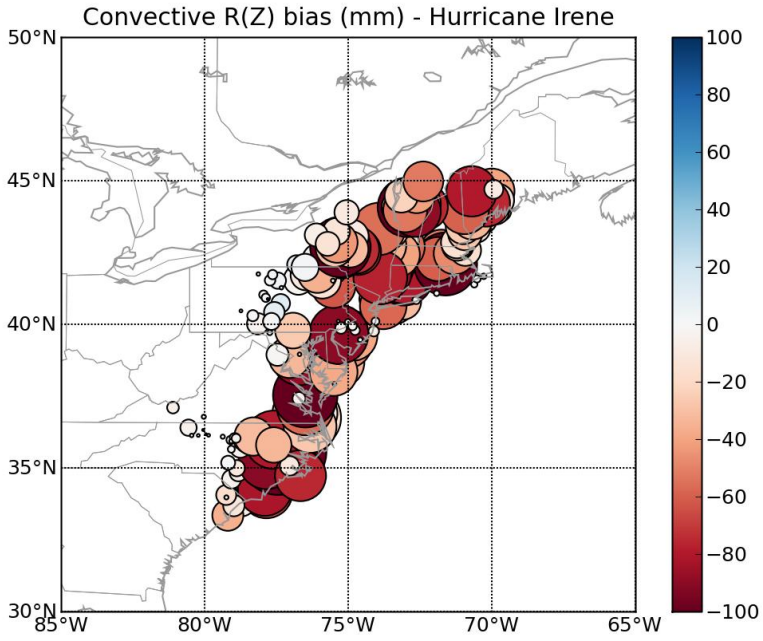


Figure 3.11. Difference between rainfall accumulations in mm from gauges and the convective Z-R function for the period 00Z 26 August 2011 - 00Z 29 August 2011. Negative (warm) values represent underestimation of radar relative to the gauge. Circle size is proportional to rainfall accumulation magnitude.

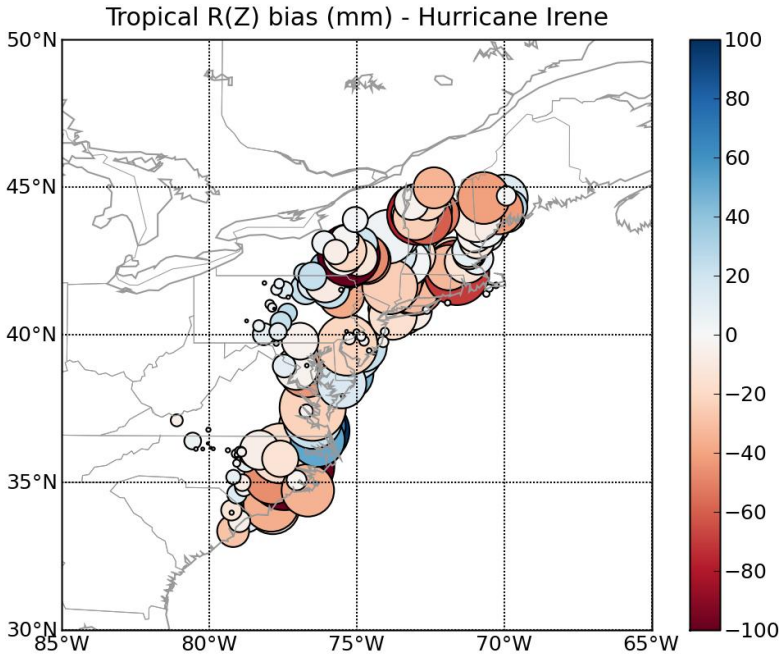


Figure 3.12. Same as Fig. 3.11, but for the tropical Z-R function.

Hurricane Ida

Hurricane Ida actually never made it to the United States as a tropical cyclone. After battling wind shear and unfavorable sea surface temperatures in the Caribbean Sea and the Gulf of Mexico, Ida weakened and underwent extratropical transition just before making landfall on the Alabama coast at 1200Z on 10 November 2009 (Avila and Cangialosi 2010). It then slowly drifted eastward and dissipated over the Florida Panhandle. The remnants of Ida then redeveloped into a new midlatitude extratropical cyclone and rapidly intensified into a strong nor-easter that impacted the mid-Atlantic states before moving offshore (Fig. 3.13). The storm total rainfall maximum occurred in Virginia and North Carolina as a result of the extratropical reintensification (Fig. 3.14).

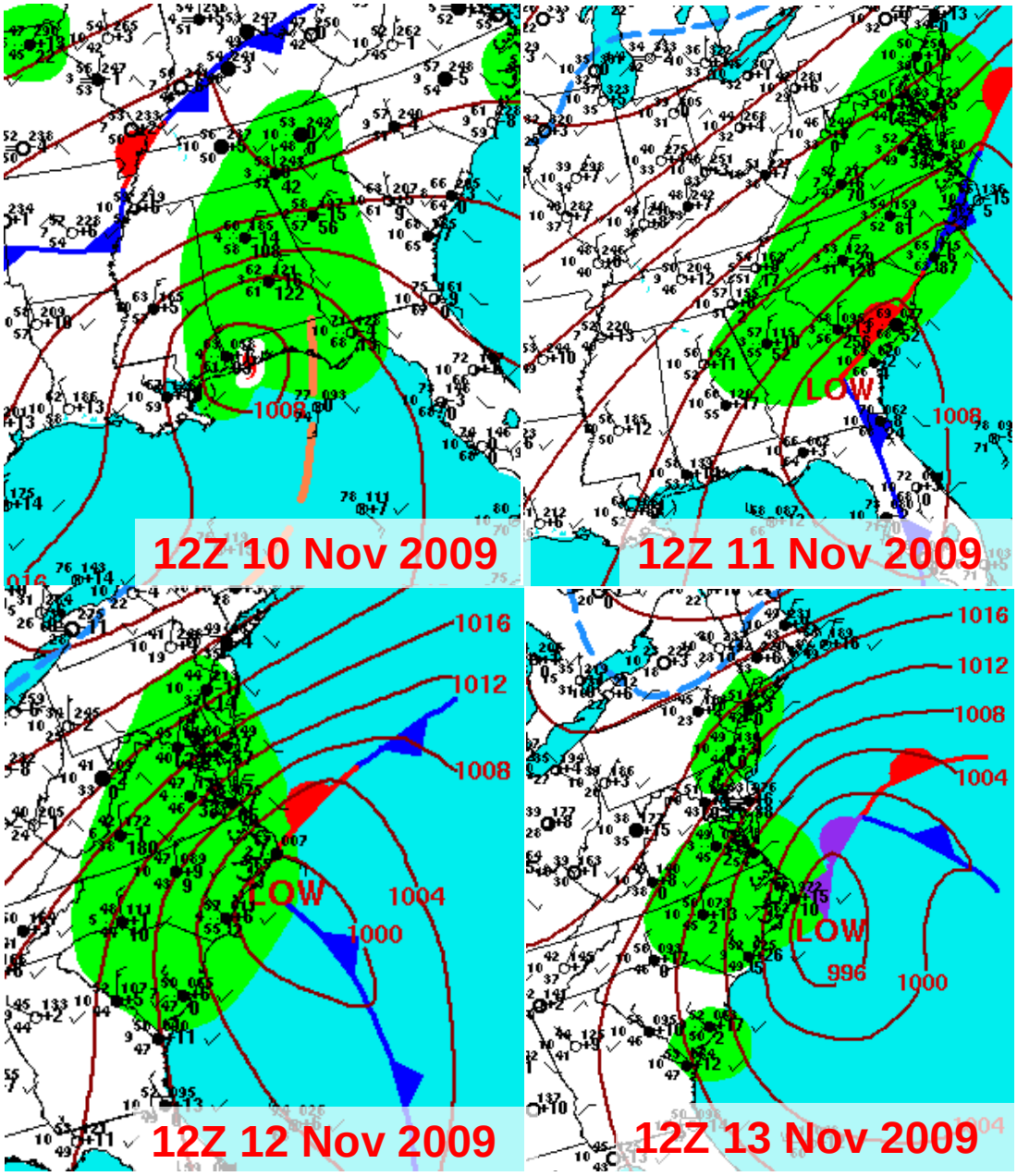


Figure 3.13. 12Z Surface observations, analyzed sea level pressure, and surface fronts for the four days following landfall of extratropical storm Ida (source: National Weather Service Hydrometeorological Prediction Center)

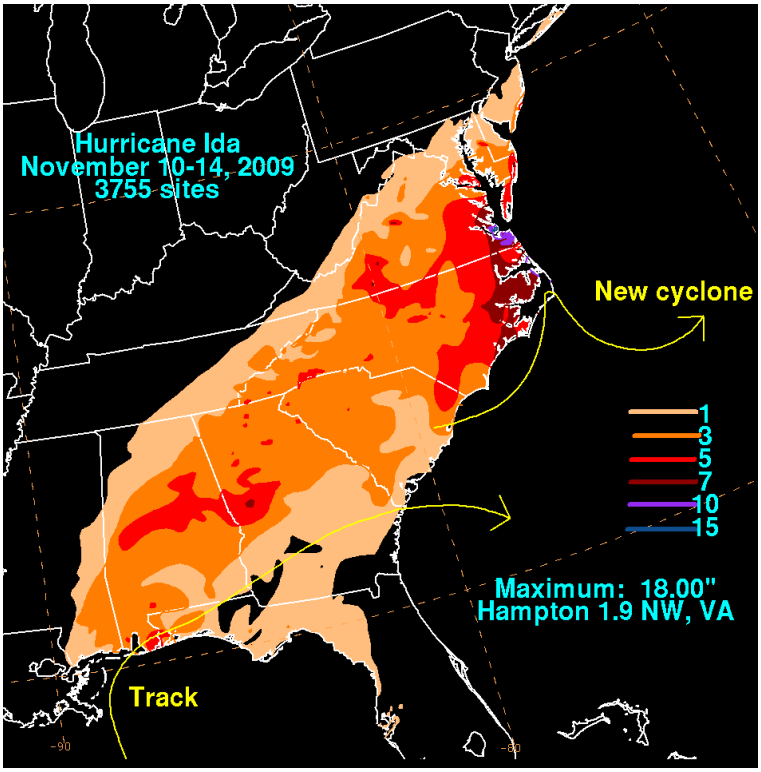


Figure 3.14. Storm total rainfall (inches) from the extratropical remnants of Hurricane Ida (source: National Weather Service Hydrometeorological Prediction Center)

Despite Ida's extratropical nature over the United States, rain rates were still largely underestimated by the standard stratiform and convective Z-R functions used for midlatitude continental rainfall (Fig. 3.15). The tropical Z-R also underestimated the rainfall accumulations in the Carolinas (Fig. 3.16). Vertical profiles of reflectivity computed near KRAX show the warm rain enhancement signature of increasing reflectivity toward the surface (Fig. 3.17). However, the VPR also contains a brightband signature near the freezing level and generally low reflectivity throughout the profile, indicating that the dominant rainfall process near the radar is stratiform in nature, likely produced by forced ascent of the warm, moist maritime airmass along the warm front (Fig. 3.13).

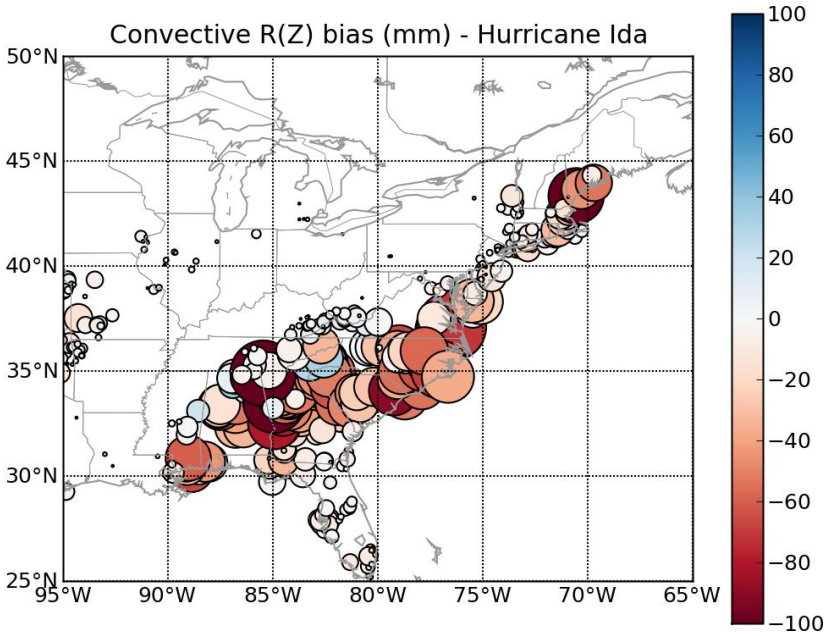


Figure 3.15. Difference between rainfall accumulations in mm from gauges and the convective Z-R function for the period 00Z 9 November 2009 - 00Z 16 November 2009. Negative (warm) values represent underestimation of radar relative to the gauge. Circle size is proportional to rainfall accumulation magnitude.

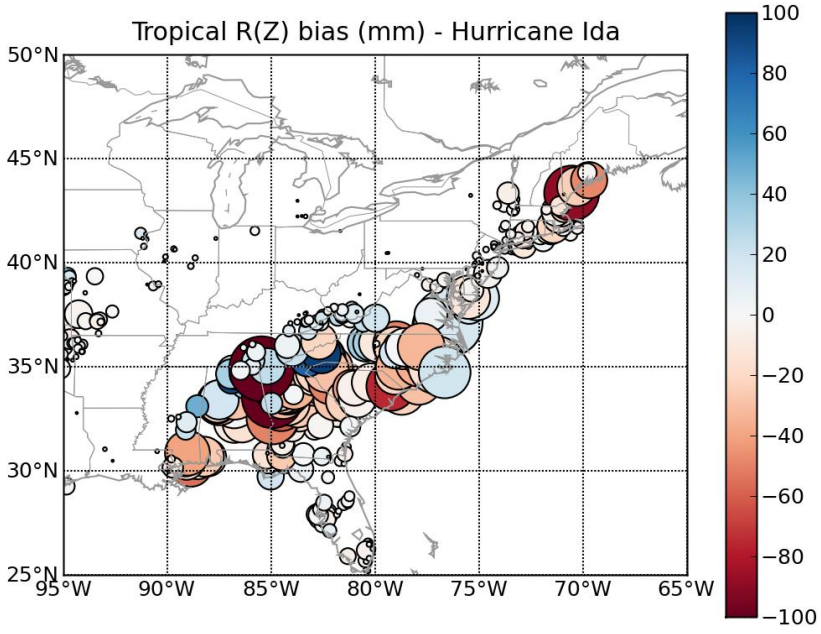


Figure 3.16. Same as Fig. 3.15, but for the tropical Z-R function.

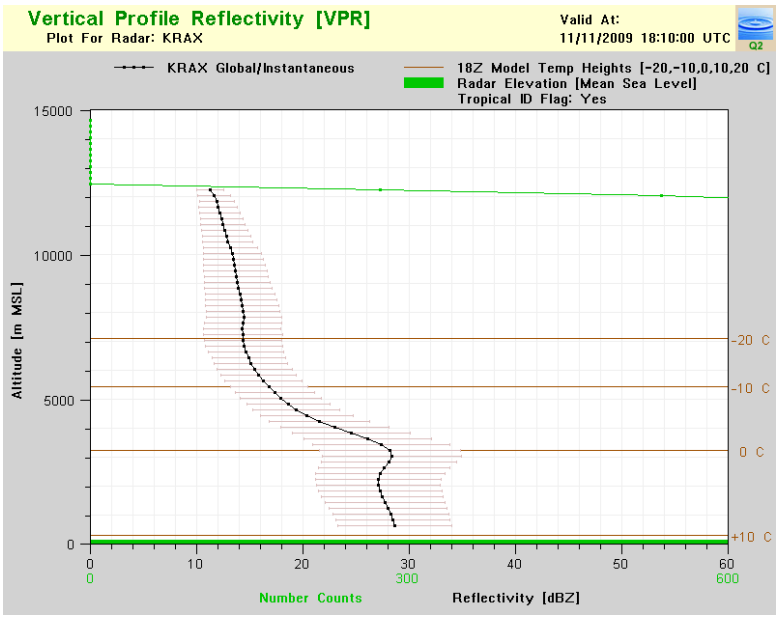


Figure 3.17. Mean and standard deviation of vertical profiles of reflectivity within 80 km of KRAX for one radar volume scan.

April 2011 Southeast U.S. MCS

On 4-5 April 2011, a large linear MCS moved across the Southeast U.S., bringing several inches of rainfall and damaging winds. Over 1300 reports of severe winds were received by the National Weather Service for the event. In terms of rainfall estimation, the case provides an important null example of when warm rain enhancement was not a significant contributor to rain rates and the convective/stratiform rainfall types provided a reasonable estimate of total accumulation (Fig. 3.20). Based on VPR structure alone, the Q2 system's warm rain identification algorithm intermittently flagged radars as "tropical" throughout the event (Fig. 3.18) and ultimately led to large overestimation of rainfall by the non-gauge-adjusted product wherever the tropical Z-R function was applied (Fig. 3.21). The sounding launched just prior to the storm in Birmingham, AL, shows a profile more typical of a continental convective environment than a maritime tropical one with steep

lapse rates and larger dewpoint depressions (Fig. 3.19).

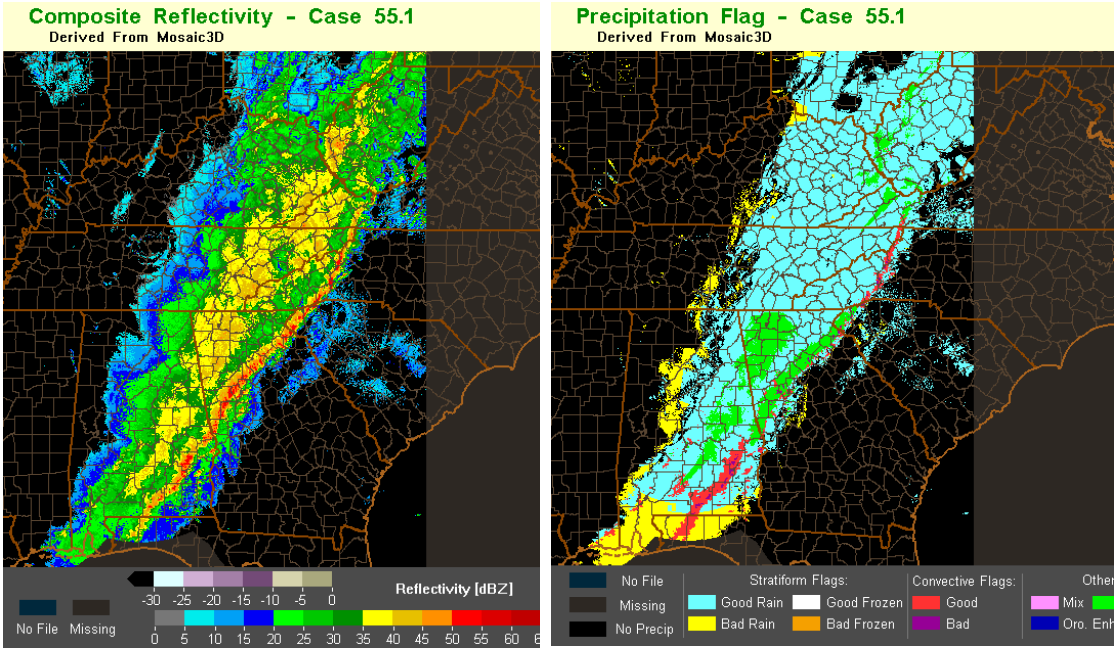


Figure 3.18. Composite reflectivity (left) and assigned Q2 precipitation types (right) for 04Z on 5 April 2011.

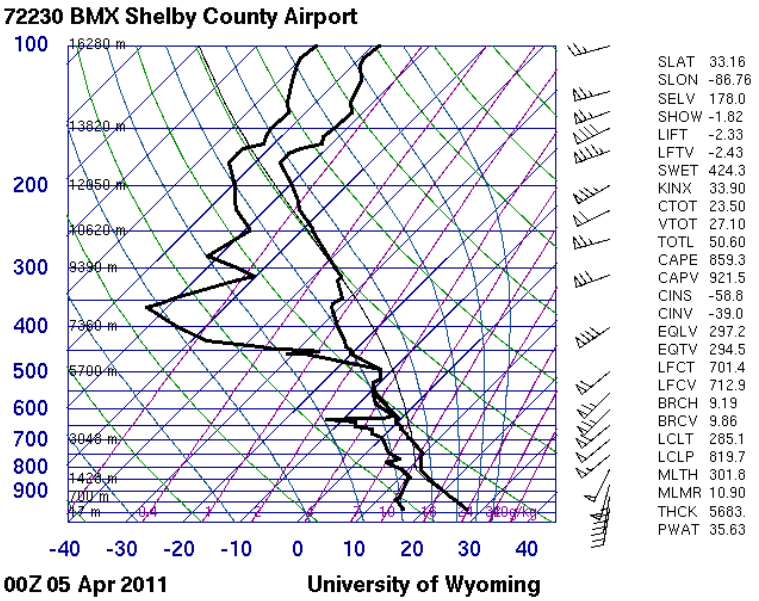


Figure 3.19. Pre-storm sounding from Birmingham, AL.

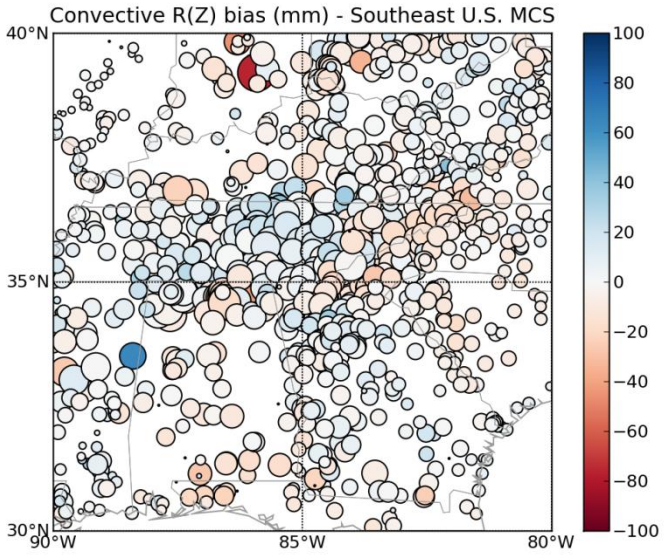


Figure 3.20. Difference between rainfall accumulations in mm from gauges and the convective Z-R function for the period 12Z 4 April 2011 - 12Z 5 April 2011. Negative (warm) values represent underestimation of radar relative to the gauge. Circle size is proportional to rainfall accumulation magnitude.

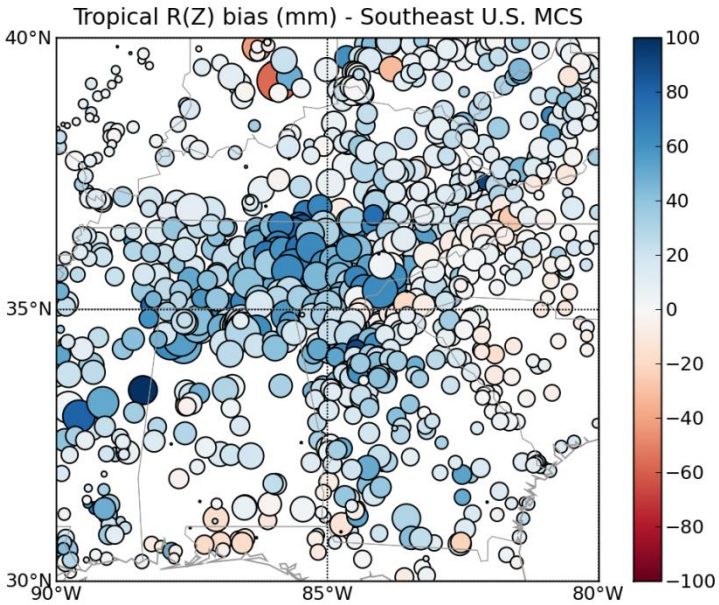


Figure 3.21. Same as Fig. 3.20, but for the tropical Z-R function.

CHAPTER 4: MODELING METHODOLOGY

Two different predictands were evaluated in this study for classification of enhanced warm rain processes based on the RUC environmental inputs:

1. The bias of hybrid scan reflectivity-based radar rain rate relative to hourly rain gauge accumulations at the nearest gauge collocated to the radar rainfall grid point on a 0.01 degree latitude-longitude cartesian grid:

$$\text{Bias} = R_{radar} - R_{gauge} \tag{4.1}$$

2. The gradient of the vertical profile of reflectivity below the brightband or freezing level for the radar nearest to the gauge (an approximation of the methodology for delineating tropical rainfall currently in the NMQ system).

The rain rate from the radars were computed using the National Weather Service default or "convective" Z-R function (Fulton et al. 1998):

$$Z=300R^{1.4} \tag{4.2}$$

where Z is the hybrid scan reflectivity and R is the hourly rain rate (mm hr^{-1}).

Binary classes for the gauge-based models were assigned to each training example based on either the sign of the bias, with negative (positive) values representing Z-R underestimation (overestimation) relative to the gauge. The decision to compute bias from a single Z-R rather than from different Z-Rs based on classification of rainfall as stratiform or convective was because of the emphasis on heavier rain rates in this study. At reflectivities greater than 35 dBZ, the convective (NWS default) Z-R will always produce a higher rain rate estimate than the Marshall-Palmer stratiform Z-R. The objective of this study is to accurately identify environments conducive for

rainfall enhancement, and the use of a stratiform-convective classification scheme could potentially introduce erroneous underestimates in areas where convective rainfall is misclassified as stratiform. Conversely, the nonlinear nature of the Z-R function tends to produce very similar rainfall rates in areas of low reflectivity regardless of which equation is used, so the impact of classifying stratiform rain as convective has a smaller impact on the overall bias.

Binary classes for the VPR-based models were assigned to each training example based on the sign of the vertical reflectivity gradient, with negative (positive) values representing an increase (decrease) of reflectivity with decreasing height (i.e., reflectivity differences were computed starting at the higher level and working downward toward the surface).

The choice of binary classes rather than using a regression-based learning algorithm on the raw bias and VPR slope values was made to mitigate some of the uncertainties inherent in all the datasets used in this analysis. Trying to fit to the exact bias or VPR slope values would allow influences unrelated to the environment to train the classification model, such as instrument calibration and measurement errors (gauges, radar), uncertainty related to spatial averaging and smoothing of the VPRs, and random sampling errors. Training on simply the sign of the bias or VPR slope provides a general analysis of the environments consistent with the rainfall and reflectivity trends. Furthermore, a weighted sampling scheme was employed to emphasize gauges and radars with more extreme deviations from the classification threshold, which will be discussed in more detail later. Thus, values very close to the threshold between the two classes had far less influence on the training than values that were clearly situated

within one or the other. Altering the sample distribution in such a fashion (creating a bimodal distribution from an approximately normal one) could impact the ability of a regression-based model to accurately classify future events for which the weighted sampling would not be performed.

Binary classification algorithms tend to be sensitive to the balance of observations assigned to each class, especially when classification accuracy is the optimization criterion. For example, consider a population for which 90% of the data values belong to class A while the remaining 10% belong to class B. Rather than trying to fit to the rare class, the algorithm will often fit to the trivial scenario for which every observation is assigned to class A to achieve a total classification accuracy of 90%. Three ways to force the algorithm to attempt to fit both classes in a more equitable manner are either to change the optimization criterion, assign a larger misclassification cost penalty to the smaller class in the optimization function, or to employ a sampling strategy that will produce a sample distribution containing equal numbers of each class (either by oversampling the smaller class or undersampling the larger class). Both the second and third approaches were evaluated in this study because of the predominance of underestimated rainfall by the radar Z-R functions at higher rainfall accumulations.

When developing a machine learning-based model for prediction or diagnostic analysis, the input data is typically split into training and testing sets such that the final model can be evaluated on data independent from that for which it was optimized. Testing on an independent set can demonstrate the model's generality and its predictive performance in new scenarios. The testing set can either originate from the same population as the training set (referred to as “out-of-box” data that were left over after

the training sample was chosen) or it can come from an entirely separate population. When dealing with spatiotemporal observations such as rainfall, using the out-of-box measurements may not be entirely independent due to the gauges or radar samples potentially originating from the same rainfall event, either at a different time or a different gauge. Thus, the only true independent test set would have to be a rainfall event or time period outside of the training period. For this study, we examined both the out-of-box observations and several independent events, both tropical and continental in nature, and compared the accuracy and variable importance results derived from each group. If the model is overfit to the training period, then the out-of-box dataset would have higher classification accuracy.

Overview of Supervised Classification Methods

To produce probabilities of enhanced rain rates, 100-member ensembles of statistical models were generated using two binary (two-class) classifiers: Support Vector Machines and Decision Tree Ensembles. Each of these classifiers are described in further detail in the following sections.

Support Vector Machines

Support vector machine (SVM) is a supervised learning method of nonlinear classification (Scholkopf and Smola 2002; Kanevski et al. 2009). The basic form of the SVM is a binary classifier that uses a number of predictor variables to find the best separation function between two classes. The two classes are defined by the user and are provided in the training dataset as the correct outcomes of the given predictor values at

each data point. Unless the user is working with data that is linearly separable in its raw input form, the SVM method first linearizes the predictor data by mapping them to a higher dimensional feature space using a kernel function of the user's choice. For this study, we have chosen to use the radial basis function kernel:

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right) \tag{4.3}$$

$$\gamma = \frac{1}{2\sigma^2} \tag{4.4}$$

where $\phi(x_i)$ and $\phi(x_j)$ represent the feature vectors at a support vector point and a testing data point, respectively, and σ controls the width of the Gaussian RBF curve. Transposing the two vectors as shown in Equation 4.3 is equivalent to taking the dot product between them. The radial basis function is then solved for that dot product to obtain the kernel function for which the SVM is trained.

Once the parameters are in linear space, the SVM algorithm then finds an optimal decision surface (i.e., a hyperplane) through the N-dimensional cloud of points that divides the classes as accurately as possible while avoiding overfitting. The decision function is the solution of the primal problem (Chang and Lin 2011):

$$\min_{\mathbf{w},b,\varepsilon} \left(\frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^{l} \varepsilon_i \right) \tag{4.5}$$

$$y_i(\mathbf{w}^T \phi(x_i) + b) \geq 1 - \varepsilon_i \tag{4.6}$$

where **w** is the vector that defines the resulting hyperplane and the margins between the two classes, and the right-hand term in Equation 4.5 ($C \sum_{i=1}^{l} \varepsilon_i$) is a misclassification loss function for data that are not perfectly separable. The loss

function introduces a "slack variable" ε_i for points that fall on the wrong side of the hyperplane, and the cost parameter *C* can be tuned by the user to increase or decrease the influence of incorrectly classified points on the final decision surface solution (Fig. 4.1). Because Equation 4.5 is an optimization function that seeks the minimum of the terms in parenthesis, however, it will attempt to fit a function that not only best separates the classes but also minimizes the sum of the slack variables needed for misclassified points.

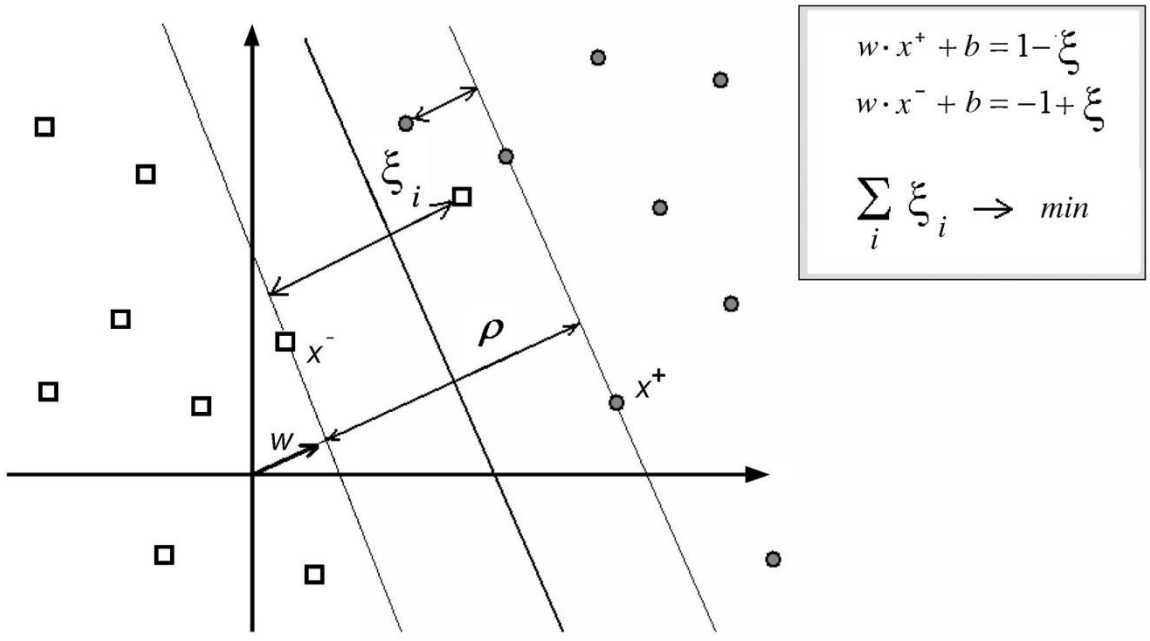


Figure 4.1. A conceptual illustration of the support vector machine two-class separation process in linear space when the classes are non-separable. The parameter ρ represents the width of the margin separating the two classes, which is maximized by the optimization function (from Kanevski et al. 2009).

Solving for the hyperplane is a quadratic programming problem requiring a dual solution that can be expressed as:

$$\min_{\alpha}\left(\frac{1}{2}\boldsymbol{\alpha}^T Q\boldsymbol{\alpha} - \mathbf{e}^T\boldsymbol{\alpha}\right) \tag{4.7}$$

$$\text{subject to} \quad \mathbf{y}^T\boldsymbol{\alpha} = 0$$

$$0 \le \alpha_i \le C$$

where **e** is a vector of non-zero values, $\boldsymbol{\alpha}$ is the vector of real numbers to be minimized, and Q is a square, positive semidefinite matrix that is a function of the training kernel:

$$Q_{ij} = y_i y_j K(x_i, x_j) \tag{4.8}$$

The resulting decision function takes the following form such that the sign of the function in parenthesis determines the predicted class of the i^{th} data point (Kanevski et al. 2009; Chang and Lin 2011):

$$sgn\left(\sum_{i=0}^{N} y_i\, \alpha_i K\left(x_i, x_j\right) + b\right) \tag{4.9}$$

Many different combinations of both the kernel function shape parameters and samples of the training set are evaluated using K-fold cross validation on the training set within a grid search of possible kernel function parameter sets. This grid search/cross validation process finds the optimal SVM model that balances accurate classification and generalization. The binary classification approach of SVM has also been adapted for multi-class datasets and regression (Kanevski et al. 2009).

SVMs have been applied to a wide range of environmental classification and regression problems. In meteorology, SVMs have been used to estimate radar-based rainfall rates in the place of Z-R functions (Trafalis et al. 2005), map climatological

rainfall, temperature, and wind speed in complex terrain (Kanevski et al. 2009), discriminate between tornadic and non-tornadic severe weather outbreaks (Mercer et al. 2009; Shafer et al. 2010), and to estimate monthly pan evapotranspiration (Eslamian et al. 2008).

The open source SVM software package LIBSVM (Chang and Lin 2011) was used for analysis in this study. The LIBSVM package includes various scripts for scaling data, separating datasets into training and testing sets, finding optimal kernel parameters through grid searches, running the K-fold cross validation and training the SVM model, and predicting classes in new data. LIBSVM also has an active community of developers who have adapted the code to several programming languages and software environments.

Decision Trees

Decision trees have been widely used for classification problems in meteorology and other fields because of their ease of interpretation and computational efficiency for multivariate datasets. In automated data mining applications, decision trees are a popular choice in meteorology because the structure of the tree is similar to the ingredients-based forecasting methodology. When making an ingredients-based prediction, the forecaster examines many different environmental or remote sensing observations and determines the likelihood of an event occurring based on the relative values of the parameters. For example, high precipitable water is an important factor for the occurrence of heavy convective rainfall, but it also needs to be coincident with an unstable airmass and sufficient forcing for initiation (Doswell et al. 1996). Decision

trees attempt to objectively quantify the thresholds above or below which the event is more likely to occur. Few events in meteorology can be based on fixed thesholds, however, and the combination of "ingredients" required for an event to occur tends not to be unique. A good example would be the relative values of CAPE and wind shear needed to produce a tornadic storm. Thus, evaluating multiple decision trees in an ensemble framework can potentially be used to derive the probability that the event will occur based on trees trained on random subsets of training events.

Decision trees can be subdivided into two types: classification trees and regression trees (Rokach and Maimon 2008). Classification trees are focused on assigning data instances to a finite number of descriptive classes, whereas regression trees attempt to replicate patterns for a predictand comprised of continuous, real numbers. Algorithms written to automatically generate decision trees use optimization criteria to find the thresholds that best segregate the classes based on training examples, and the complexity of the tree can be explicitly controlled by the user to mitigate overfitting. For example, the tree-growing algorithm can be set to stop once a minimum number of data instances has been parsed to a tree node rather than allowing it to continue until every single data instance is correctly classified, many into terminal nodes containing only one data point.

The MATLAB function ClassRegTree was used to construct classification trees for this study. ClassRegTree allows the user a wide range of options to control how the decision trees are constructed. The split criterion for each tree node was the maximization of the Gini Index (Hastie et al. 2001):

$$\sum_{k \neq k'} \hat{p}_{mk} \hat{p}_{mk'} = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk}) \quad (4.10)$$

where $\hat{p}_{mk}$ is the proportion of class k observations in the m^{th} node of the decision tree. The training and testing sets were collected following the same sampling methodology as was used for the support vector machines. The training set was then used to train two different types of decision trees:

a) Random Forest Classification

b) All-Parameter Classification Tree Ensemble

The difference between the Random Forest and All-Parameter approaches primarily lie in the features included for consideration at each node split in the tree. In a Random Forest (Breiman et al. 2001), a randomly selected subset of the predictors are considered at each node, and the split is based on the predictor that provides the optimal split on the data instances available at that node. This random subset selection introduces a large amount of variability in the structure of all the trees in the ensemble. By contrast, when an ensemble of trees is grown by considering all possible predictors at every node (i.e., the All-Parameter approach), the trees tend to be more similar to each other, particularly in the top few tiers. The variability in the deeper tiers of each tree is higher due to the differences in the random samples used to train the trees.

Association of Gauges To Their Environment

The 19 environmental parameters used as predictors for the statistical models are listed in Table 4.1. They encompass a range of isobaric levels within the lower half of the atmosphere and include both temperature and moisture-based variables for a general depiction of the thermodynamic environment.

RUC-20 Variable (abbrev.; units)
Vertically-Integrated Precipitable Water (PW; kg m^{-2})
1000-700 hPa Precipitation Efficiency (PE; kg m^{-2})
1000-700 hPa Mean Relative Humidity (RH; percent)
900 hPa Relative Humidity (RH900; percent)
850 hPa Relative Humidity (RH850; percent)
700 hPa Relative Humidity (RH700; percent)
500 hPa Relative Humidity (RH500; percent)
Surface Equivalent Potential Temperature (Theta-E; K)
Surface Temperature (TSFC; degrees C)
850 hPa Temperature (T850; K)
700 hPa Temperature (T700; K)
500 hPa Temperature (T500; K)
Height of 0 C Isotherm (HGT0C; m)
Surface-Based CAPE (CAPE; J kg^{-1})
850-700 hPa Lapse Rate (LRLOW; K km^{-1})
850-500 hPa Lapse Rate (LRDEEP; K km^{-1})
Surface-850 hPa Wind Shear (SHEAR850; m s^{-1})
Surface-700 hPa Wind Shear (SHEAR700; m s^{-1})
Surface-500 hPa Wind Shear (SHEAR500; m s^{-1})

Table 4.1. RUC analysis parameters used as inputs to the statistical models.

Once the HADS rain gauges were quality controlled, the associated environmental values for each gauge were retrieved based on the storm-relative inflow at the gauge location, which is the vector difference between the wind vector at any level and the storm motion vector (i.e., the mean surface-500 hPa wind). Rather than retrieve the parameters at the nearest RUC grid point to the gauge (a location where the environment is likely thermodynamically modified by convection and precipitation), the model's U and V wind components were first used to derive the speed and direction of the storm relative wind at all the levels corresponding to the predictor parameters (surface, 900 hPa, 850 hPa, 700 hPa, 500 hPa, and a mean of all five levels for vertically integrated parameters). The storm relative wind vector at each of those levels was then the basis for the direction and distance from a gauge that environmental predictors were retrieved (Fig. 4.2). Because RUC analyses were updated hourly, the wind velocity was converted to units of km hr^{-1}, and that value determined how many 20 km grid boxes away from the storm to retrieve the input values as an estimate of storm inflow properties. Thus, for storms with stronger inflow, the environment was sampled further away from the gauge.

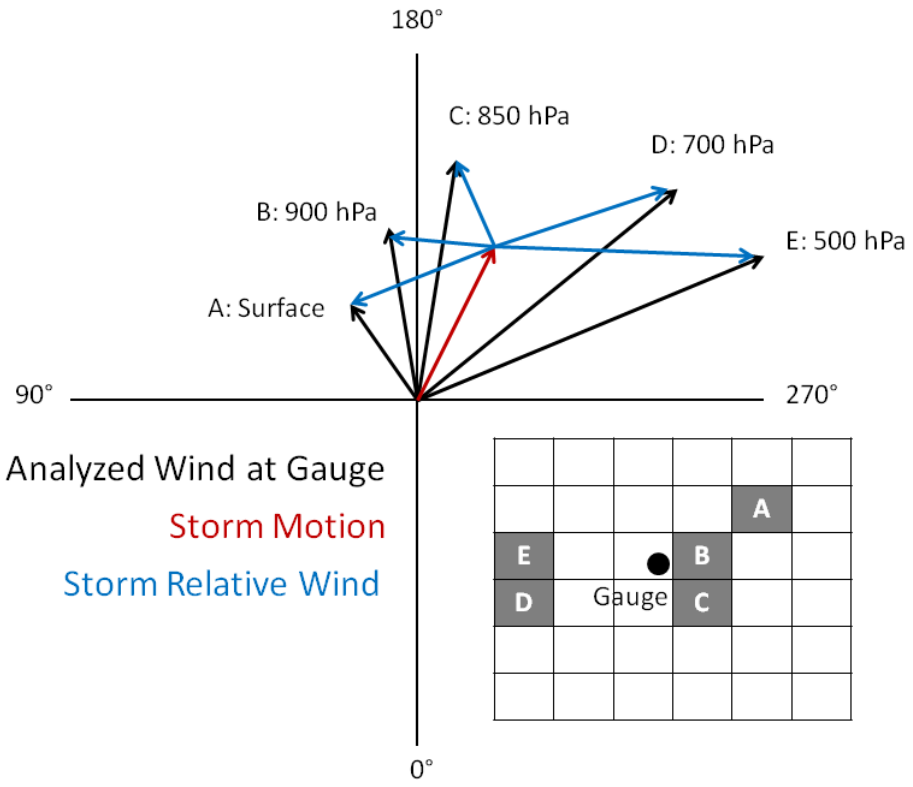


Figure 4.2. Conceptual illustration of how inflow parameters are retrieved based on the storm relative wind vector at each level. The box at the lower right shows which grid points relative to the gauge location would be chosen in horizontal RUC grid space from the wind vector example.

Once the inflow properties near each gauge were retrieved from the RUC analysis, some additional filtering on the dataset was conducted to remove potential frozen precipitation types (gauges with surface temperature less than 10° C) and gauges for which any of the RUC inputs were missing (e.g., gauges located near the edge of the RUC model's domain). The predictor variables were then normalized to a common scale for which the minimum and maximum values of each variables were assigned as -1 and 1, respectively, with values in between adjusted to scale by linear interpolation. The normalization of the input variables prevents the classification algorithms from biasing its results toward variables with larger quantities or variability over larger ranges. For

example, the height of the freezing level can vary between 2000 to 5000 meters whereas 500 hPa temperature may vary only over a few degrees, so the changes in rainfall bias would appear to be more sensitive to freezing level height just by virtue of its units.

Random sampling with replacement was performed on the full 2010 warm season dataset to create training sets of gauges for the classification algorithms, and remaining gauges were kept as the out-of-box test set. Weighted sampling was employed to emphasize the influence of gauges with large bias on the training of the classification model. An inverse of the Nuttall window filter centered on zero bias assigned sampling weights to the gauges such that bias values near zero had weights near zero while larger positive and negative biases were given weights near 1 (Fig. 4.3). The result was that gauges with larger bias magnitude were preferentially selected by the random sampling algorithm.

In order to address the tendency for unbalanced class sizes, two different approaches were tested. The first approach involved modifying the random sampling algorithm to force an equal number of each class by undersampling the majority class, which for this study tended to be gauges where the convective Z-R underestimated rainfall due to the focus on heavier rain rates (Fig. 4.3). Because this study focused on training ensembles of models to create probabilistic output, however, undersampling the gauges for one ensemble member did not preclude the left out gauges from being included for other ensemble members. The second approach retained the class imbalance in the random sampling and instead assigned a larger misclassification cost penalty to the smaller class for each classification algorithm's respective optimization function.

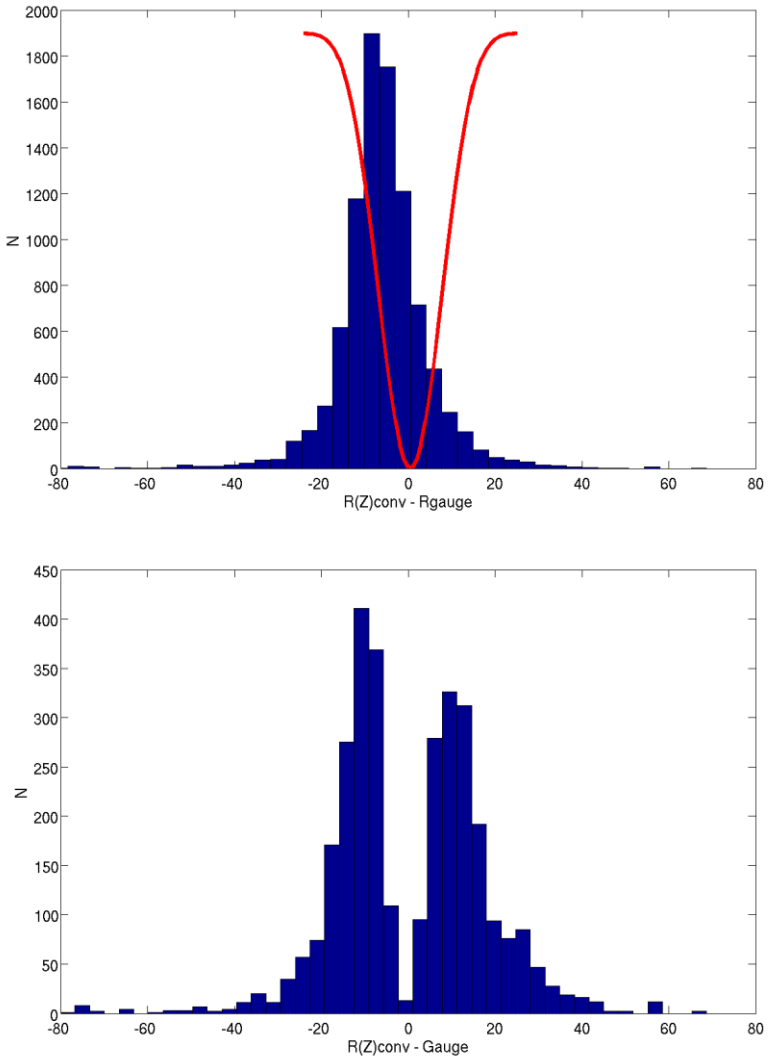


Figure 4.3. a) Distribution of rainfall bias for all 2010 warm season gauges that exceeded 12.7 mm hr^{-1} (0.5 in. hr^{-1}), and the inverse Nuttall window filter used to assign selection weights to the gauges (ranging from 0-1). b) An example distribution of gauges that resulted from the weighted random selection, which divides the gauges into two, equal-sized classes for training.

Ensemble Training and Evaluation

For the support vector machine runs, 100-member ensembles were generated for multiple kernel parameter configurations rather than running the computationally

expensive parameter grid search for every ensemble member. The selection of the radial basis function γ parameter and the cost penalty of misclassified gauges is driven by finding the best balance between model generality and accuracy. High values of both γ and C produce a model that overfits the training examples to achieve very high accuracy that may not generalize well to data outside the training set. Conversely, low values of γ and C produce a smoother, more general decision surface between the classes but can have a larger rate of misclassification. The grid search tests the entire parameter space to find the parameter set that produces the highest accuracy through cross-validation of the training data. Because the training sets for the ensemble members are samples of the same population with several gauges potentially being included more than once, however, the SVM grid algorithm tended to repeatedly select a small subset of the parameter combinations when run on many ensemble members. Thus, for greater computational efficiency in training, the SVM models were trained for the 100 ensemble members with the kernel parameters held fixed for the entire ensemble, and all the predominant combinations were evaluated.

Similar parameter adjustments can be made when training decision trees to achieve varying levels of accuracy vs. generality. For example, the user can set the minimum number of training examples or "leaves" that can be assigned to a node. Setting a higher minimum forces the training algorithm to stop sooner, which gives a shallower tree with fewer nodes (i.e., a more general result). Setting a minimum value of one training example, on the other hand, allows the algorithm to continue splitting the data until every example is correctly classified. Such a result may give high accuracy for the training set, but may not work well on other datasets. For this study, the

minimum leaf size was adjusted at various intervals to test the effect on the output, and a fixed value of 10 training examples was chosen as a balance between model accuracy and generality.

Support Vector Machines		Decision Trees	
	Undersampled Selection Sets:		Undersampled Selection Sets:
S1	$\gamma = 2$; C = 4	**D1**	Random Forest (RF)
S2	$\gamma = 2$; C = 8	**D2**	All Parameters (AP)
S3	$\gamma = 2$; C = 32		Weighted Cost Penalty Sets:
S4	$\gamma = 8$; C = 2	**D3**	RF; W(min) = 3
S5	$\gamma = 0.5$; C = 128	**D4**	AP; W(min) = 3
S6	$\gamma = 0.5$; C = 2048	**D5**	RF; W(min) = 6
S7	$\gamma = 0.125$; C = 8192	**D6**	AP; W(min) = 6
	Weighted Cost Penalty Sets:		Unbalanced Class Sizes:
S8	$\gamma = 0.125$; C = 8192; W(min) = 3	**D7**	RF; Weighted Sampling
S9	$\gamma = 0.125$; C = 8192; W(min) = 6	**D8**	AP; Weighted Sampling
S10	$\gamma = 0.125$; C = 8192; W(maj) = 0.16	**D9**	RF; Unweighted (control)
S11	$\gamma = 0.125$; C = 8192; W=1 (control)	**D10**	AP; Unweighted (control)

Table 4.2. Configuration of each ensemble. For the support vector machine ensembles, γ is the radial basis function kernel shape parameter, C is the misclassification cost penalty, and W is the weighted multiplier applied to C for the minority (min) or majority (maj) class. For the decision tree ensembles, RF represents random forests, and AP represents the non-Random Forest ensembles. W is the same as for the SVMs.

Table 4.1 lists every ensemble parameterization that was evaluated. For the classification algorithms, both undersampling and cost penalty adjustments were tested for a range of model-specific parameter sets. In addition, a control ensemble was run that was trained on a dataset that was not corrected for class imbalance (i.e., having a similar distribution to the population).

Probabilities of radar rainfall underestimation were computed as the ratio (0 to 1) of ensemble members that predicted the underestimation class for a given gauge location and its associated inflow environment. Verification of the probabilities was based on a number of performance metrics that can be either computed from the probabilities themselves or as part of a contingency table (Fig. 4.4) if a probability threshold for classification is set to segregate overestimation or underestimation predicted classes. These metrics included Brier Score (where p is the probability and o is the observed outcome with 1 representing underestimation and 0 representing overestimation), Heidke Skill Score (HSS), Hit Rate (HR), Probability of Detection (POD), False Alarm Rate (FAR), Critical Success Index (CSI), and Equitable Threat Score (ETS) (Wilks 1995; Mason 2003):

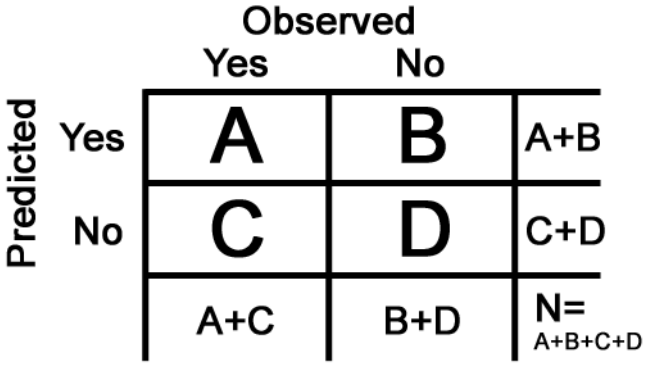


Figure 4.4. Standard contingency table for a binary outcome upon which many verification equations are based.

$$BS = \frac{1}{N}\sum_{k=1}^{N}(p_k - o_k)^2 \tag{4.11}$$

$$HSS = \frac{2(AD - BC)}{(A + C)(C + D) + (A + B)(B + D)} \tag{4.12}$$

$$HR = \frac{A + D}{N} \tag{4.13}$$

$$POD = \frac{A}{A + C} \tag{4.14}$$

$$FAR = \frac{B}{A + B} = 1 - POD \tag{4.15}$$

$$CSI = \frac{A}{A + B + C} \tag{4.16}$$

$$ETS = \frac{A - E}{A + B + C - E} \tag{4.17a}$$

$$E = \frac{(A + B)(A + C)}{N} \tag{4.17b}$$

In addition to computing the probability of radar rainfall underestimation, a variable importance analysis was conducted to determine which of the RUC environment predictors were most strongly influencing the models. Each trained model was evaluated using a testing set for which one variable at a time was permuted or randomized, essentially removing its correlation with the rainfall bias. The change in the Brier Score with the randomized variable was then computed and normalized by the standard deviation of the increased error across all ensemble members. This permutation process was bootstrapped to assess the robustness of the variable ranking to variance of the input data.

CHAPTER 5: RESULTS AND CONCLUSIONS

In terms of spatial representation of probabilities, decision trees and SVMs produced somewhat different depictions of where they predicted enhancement of rain rates through warm rain processes to occur. Figs. 5.1-5.3 provide a snapshot from each of the three independent cases, highlighting differences between the models for that time given the same exact RUC input values. The general depiction of probabilities are similar among the models, particularly with regards to the location of the highest probabilities around the storm of interest. However, differences are apparent, particularly between the A and C panels for each event, which are the two decision tree-based models for gauge bias and VPR slope.

Because the gauge bias-based classifiers were trained on the gauges themselves, they appear to align better with the bias tendencies than the VPR-based classifiers do. While this is not a surprising finding, one would guess that if the gauge bias and the VPR slope were varying due to the same phenomena (i.e., warm rain enhancement of rainfall), their probabilities should look more spatially consistent and the VPR-based probabilities would align better with the gauge biases observed. Furthermore, none of the models appear to adequately capture the smaller scale variability of the gauge bias, particularly for the Southeastern U.S. MCS event (Figure 5.3). The high wind, fast storm motion, and convective nature of the MCS likely contributed to a large amount of variability in the gauge accumulations over short space and time scales, which would be difficult for either the radar-averaged VPR or the hourly RUC analysis to predict.

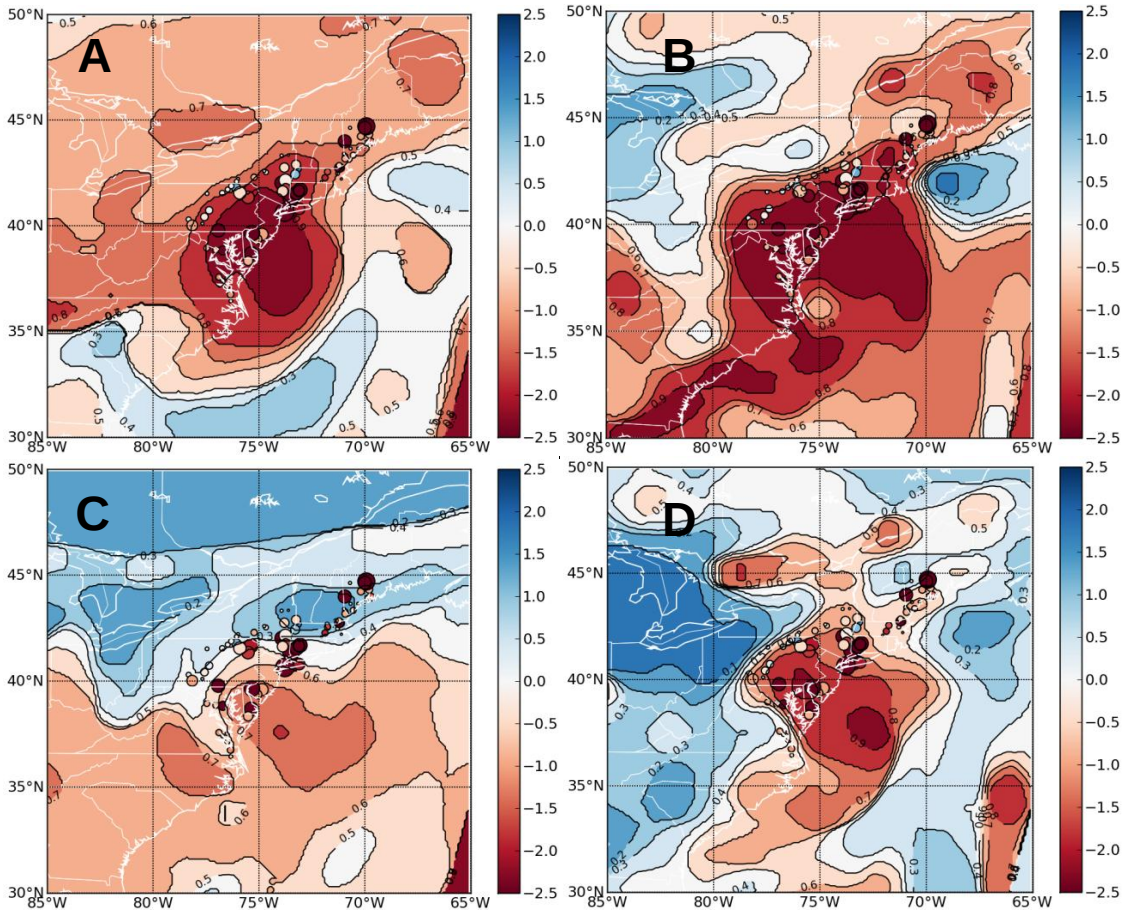


Fig. 5.1. Warm rain probabilities generated from the 20-km RUC analysis for 06Z 28 Aug 2011 (Hurricane Irene) using A) ensemble D2 with gauge bias as the classification criterion, B) ensemble S7 with gauge bias as the criterion, C) D2 with VPR slope as the criterion, and D) S7 with VPR slope as the criterion. Size of colored circles represent hourly gauge accumulations, and the circle colors represent bias to radar rainfall estimates in mm (magnitude shown by scale on right). Colors of contoured areas range from dark blue for low warm rain probabilities to dark red for high warm rain probabilities.

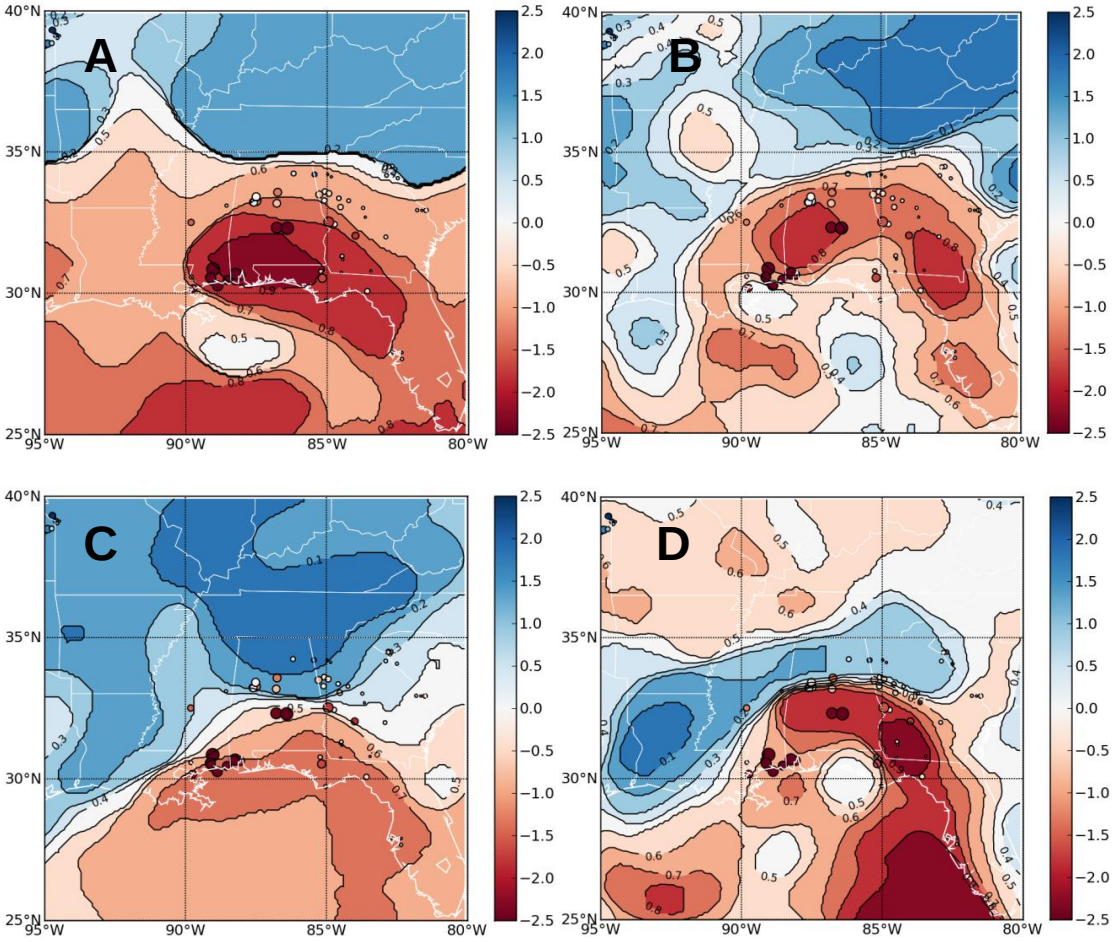


Fig. 5.2. Warm rain probabilities generated from the 20-km RUC analysis for 06Z 10 Nov 2009 (extratropical storm Ida) using A) ensemble D2 with gauge bias as the classification criterion, B) ensemble S7 with gauge bias as the criterion, C) D2 with VPR slope as the criterion, and D) S7 with VPR slope as the criterion. Size of colored circles represent hourly gauge accumulations, and the circle colors represent bias to radar rainfall estimates in mm (magnitude shown by scale on right). Colors of contoured areas range from dark blue for low warm rain probabilities to dark red for high warm rain probabilities.

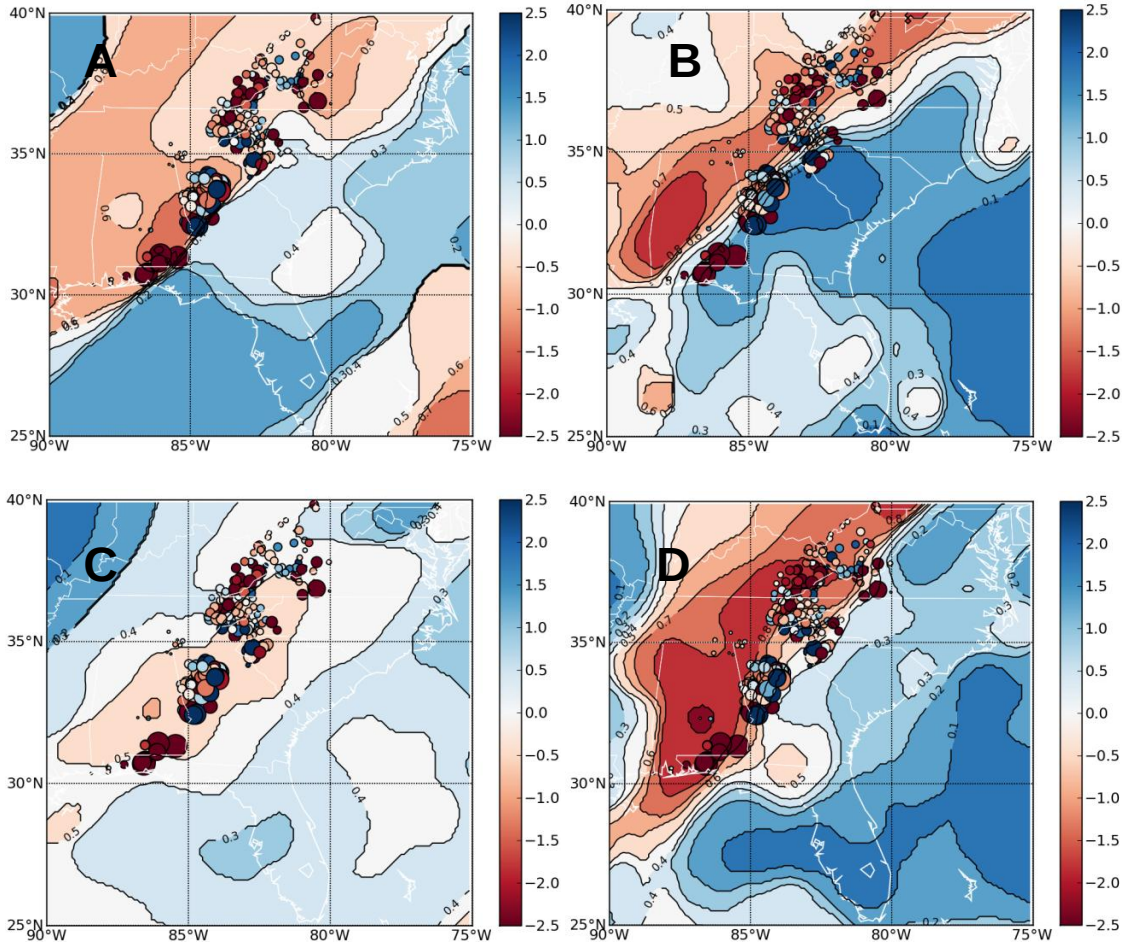


Fig. 5.3. Warm rain probabilities generated from the 20-km RUC analysis for 05Z 5 Apr 2011 (southeast U.S. MCS) using A) ensemble D2 with gauge bias as the classification criterion, B) ensemble S7 with gauge bias as the criterion, C) D2 with VPR slope as the criterion, and D) S7 with VPR slope as the criterion. Size of colored circles represent hourly gauge accumulations, and the circle colors represent bias to radar rainfall estimates in mm (magnitude shown by scale on right). Colors of contoured areas range from dark blue for low warm rain probabilities to dark red for high warm rain probabilities.

In terms of computational feasibility for implementation as part of a national, real-time system (NMQ), the decision tree ensembles (both all-parameter and random forest) held a significant advantage over the SVM ensembles. Both training and prediction ran much faster using the decision tree-based classifiers. Predicting enhanced rain rate probabilities from an ensemble of 100 members for a single hour on the full

RUC domain (246,051 10 km grid points after remapping to the NMQ cartesian grid) required an approximate runtime of 42 seconds using decision trees vs. 251 minutes using the SVMs. Obviously the amount of time required for the SVM prediction would not be feasible in a real-time framework, but fortunately (as will be shown later) the skill of SVM is not significantly better than the decision tree method. Due to the huge difference in computational expense between the two classifiers and the similarity in skill, all of the variable importance and ensemble structure analysis was conducted on the decision tree ensembles only. They would also be the classifiers of choice when moving forward with implementation of the statistical model for enhanced rain rate probability on the NMQ system.

Radar Rainfall Bias Ensembles

Variable Importance

The decision tree variable importance testing on the predictor variables revealed that the gauge-based classification accuracy was most sensitive to the absence of three of the variables in particular (Fig. 5.4), thus implying that the classification models found an important association between them and the bias of radar rainfall relative to gauges. Those three variables were the 850 - 500 hPa lapse rate, freezing level height above ground, and the mean 1000 - 700 hPa relative humidity. The bootstrapped analysis of the variable importance metric in Fig. 5.4 showed that the relative rankings of the variables, particularly the most significant ones, were stable with respect to variability in the training data. The spread of each box plot represents the variable importance computed from 100 iterations of the 100 member ensemble (i.e., 10000 total

members), and the notches represent the bootstrapped confidence interval for the median value.

Table 5.1 contains the bootstrapped mean and standard deviations of the RUC predictor variables as sampled from the balanced sample distribution (i.e., values were drawn with replacement from the full population of data, and the majority bias class was undersampled to produce equal sized classes). The balanced set was selected using the weighting function described in Chapter 4 to emphasize data points with larger bias. Bootstrapping the predictors from the two classes after undersampling and weighted selection gives values more representative of the dataset used to train the classification models rather than examining the full data population. Table 5.1 also shows the achieved significance level (ASL) or p-value resulting from testing the null hypothesis that the two rainfall bias classes were pulled randomly from the same distribution and thus have equal means (following the two-sample permutation test methodology outlined by Efron and Tibshirani 1993). Low values of the ASL indicate that the null hypothesis can be rejected, thus suggesting that there is a significant difference between the RUC parameters values associated with areas of positive and negative radar rainfall bias relative to gauges. The means provided in Table 5.1 show what those relative differences are and support the argument that the underestimation is observed in environments conducive for enhanced warm rain processes (e.g., higher moisture content, higher relative humidity, and weaker lapse rates through the mid-levels of the atmosphere). However, the standard deviation columns show that there is some amount of overlap between the two distributions, which likely contributes to the uncertainty of the classification models.

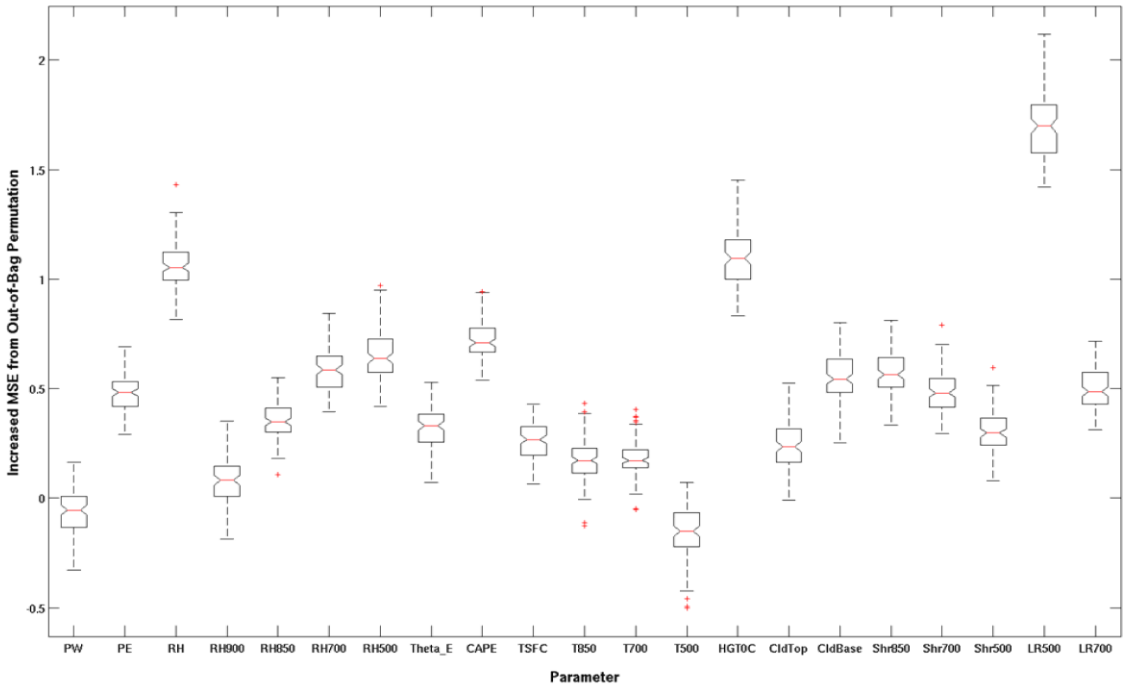


Figure 5.4. Increase in decision tree ensemble mean squared error (normalized by standard deviation) when individual parameters are permuted or randomized. The relative values of normalized MSE can be useful as a test of variable importance to the classification model.

	Bias < 0 (Underestimated)		Bias > 0 (Overestimated)		Two-Sample Test ASL
Parameter (units)	**Mean**	**St. Dev.**	**Mean**	**St. Dev.**	
Precipitable Water (kg m^{-2})	55.5 (0.4)	8.6 (0.3)	50.4 (0.5)	10.1 (0.3)	<0.001
Precipitation Efficiency (kg m^{-2})	48.2 (0.5)	10.3 (0.3)	38.8 (0.5)	10.6 (0.3)	<0.001
1000-700 hPa Mean RH (%)	85.6 (0.5)	10.5 (0.3)	78.0 (0.5)	11.5 (0.4)	<0.001
900 hPa RH (%)	86.8 (0.6)	11.9 (0.4)	79.6 (0.6)	14.1 (0.4)	<0.001
850 hPa RH (%)	89.2 (0.5)	10.4 (0.5)	82.5 (0.6)	13.5 (0.5)	<0.001
700 hPa RH (%)	82.4 (0.7)	16.2 (0.6)	74.4 (0.8)	17.8 (0.6)	<0.001
500 hPa RH (%)	84.4 (0.9)	19.2 (0.9)	79.1 (0.9)	21.1 (0.7)	<0.001
Equivalent Pot. Temperature (K)	348.3 (0.3)	11.2 (0.2)	346.3 (0.3)	12.5 (0.2)	<0.001
Surface Temperature (C)	24.7 (0.1)	4.3 (0.1)	25.0 (0.1)	4.9 (0.1)	0.04
500 hPa Temperature (K)	266.6 (0.1)	2.3 (0.1)	264.8 (0.1)	3.1 (0.1)	<0.001
Freezing Level Height (km agl)	4.6 (0.03)	0.6 (0.07)	4.3 (0.03)	0.7 (0.07)	<0.001
CAPE (J kg^{-1})	1110 (54)	1218 (57)	1489 (63)	1415 (62)	<0.001
850 hPa Temperature (K)	290.9 (0.1)	2.5 (0.1)	291.2 (0.1)	3.0 (0.1)	0.002
700 hPa Temperature (K)	281.7 (0.1)	2.3 (0.05)	281.3 (0.1)	2.9 (0.04)	<0.001
Sfc-500 hPa Shear (m s^{-1})	13.3 (0.3)	7.7 (0.2)	15.3 (0.4)	8.5 (0.2)	<0.001
Sfc-700 hPa Shear (m s^{-1})	10.9 (0.3)	7.0 (0.2)	10.9 (0.3)	6.3 (0.2)	0.883
Sfc-850 hPa Shear (m s^{-1})	10.0 (0.1)	6.6 (0.1)	8.9 (0.1)	5.6 (0.1)	<0.001
850-500 hPa Lapse Rate (K km^{-1})	-4.9 (0.02)	0.5 (0.02)	-5.4 (0.02)	0.5 (0.02)	<0.001
850-700 hPa Lapse Rate (K km^{-1})	-4.6 (0.03)	0.8 (0.02)	-5.0 (0.03)	0.8 (0.02)	<0.001

Table 5.1. Mean values of the bootstrapped mean and standard deviation for all RUC parameters used to train the gauge bias-based classification model (drawn from a balanced and weighted sample of the 2010 warm season population). Numbers in parenthesis represent the standard deviation of the target statistic among the bootstrap samples. The rightmost column contains the achieved significance level or p-value of a two-sample permutation test that the RUC parameter distributions of the two bias classes have equal means.

Verification on Independent Cases

While the support vector machine ensembles performed better in terms of the Heidke Skill Score for the training period, they had much lower skill than decision trees for the three independent cases (Figs. 5.5-5.12). Varying the sampling methods and RBF parameters had little impact on the skill of the SVM ensembles, as evidenced by the very minimal spread between the 11 parameter sets tested for several of the verification metrics computed.

For the decision tree ensembles, varying the parameters and sampling strategies had a large impact on the resulting model's behavior, particularly with respect to which probability threshold maximized the model's segregation skill. The best probability threshold also varied from event to event among the three independent cases tested, but the relative optimal probabilities among the parameter sets were consistent between the 2010 training period and the independent cases. There were differences, however, among the different verification metrics in selecting the "best" ensemble. For scores that relied on overall accuracy (POD, HR, FAR, and CSI), ensembles tended to perform better that correctly predicted the majority class, even if it meant that it performed very poorly for the minority class.

It should be noted that no resampling was done on the independent cases to ensure equal-sized classes or to emphasize gauges with larger radar bias. For example, Hurricane Irene was largely an event that was underestimated by radar. Out of 474 hourly rain gauge measurements exceeding 12.7 mm hr^{-1}, only 4 of those were overestimated by the convective Z-R function. Thus, the ensembles that produced very high probabilities of underestimation everywhere in the domain had very good POD,

HR, and FAR and poor HSS. The HSS was designed to assess the skill of a prediction for the rare class, so the successful identification of overestimation for those four gauges has a major impact on the HSS for the event.

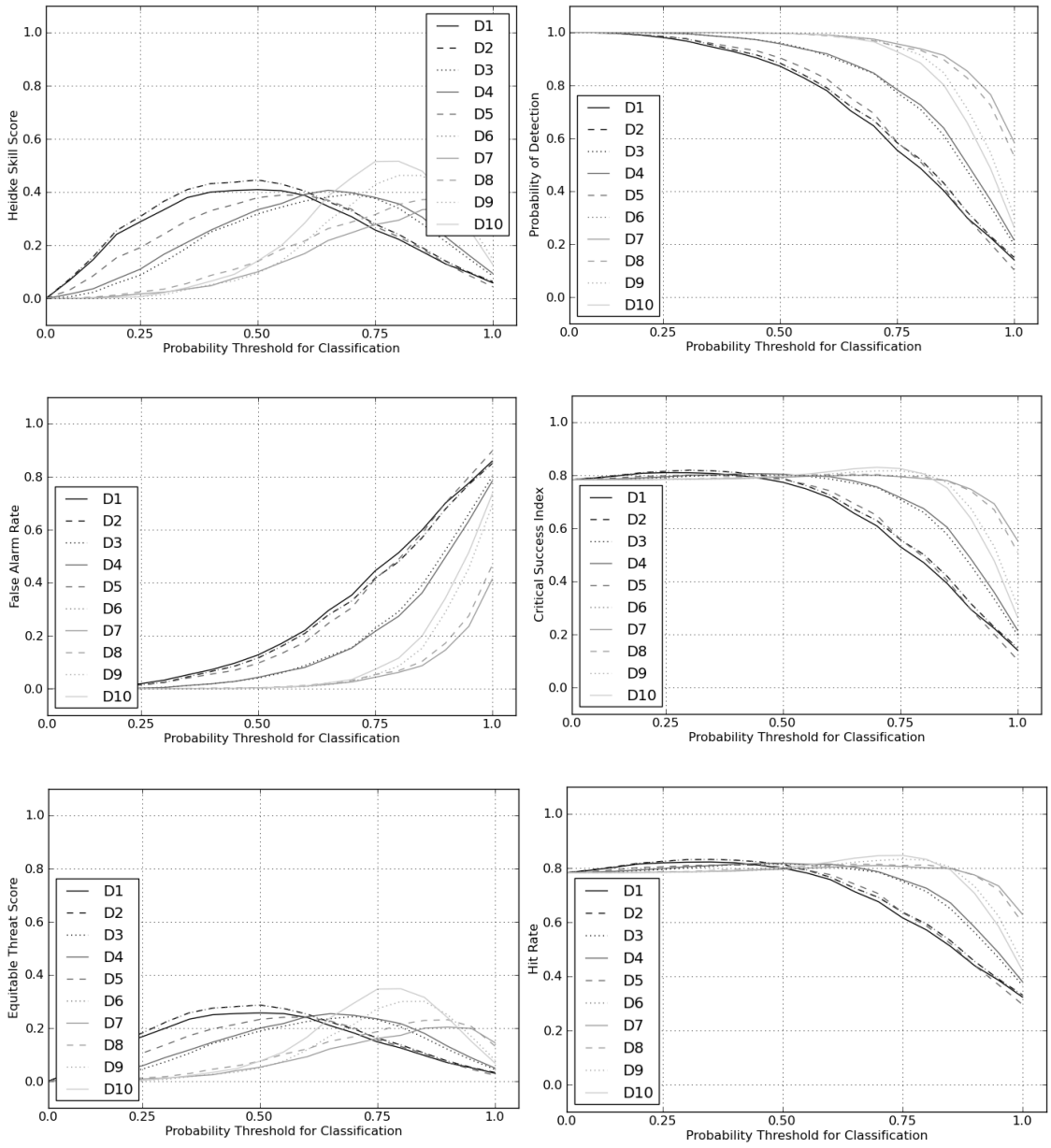


Figure 5.5. Decision tree ensemble verification threshold plots for the 2010 warm season. Each line represents the median score for the 100 ensemble members.

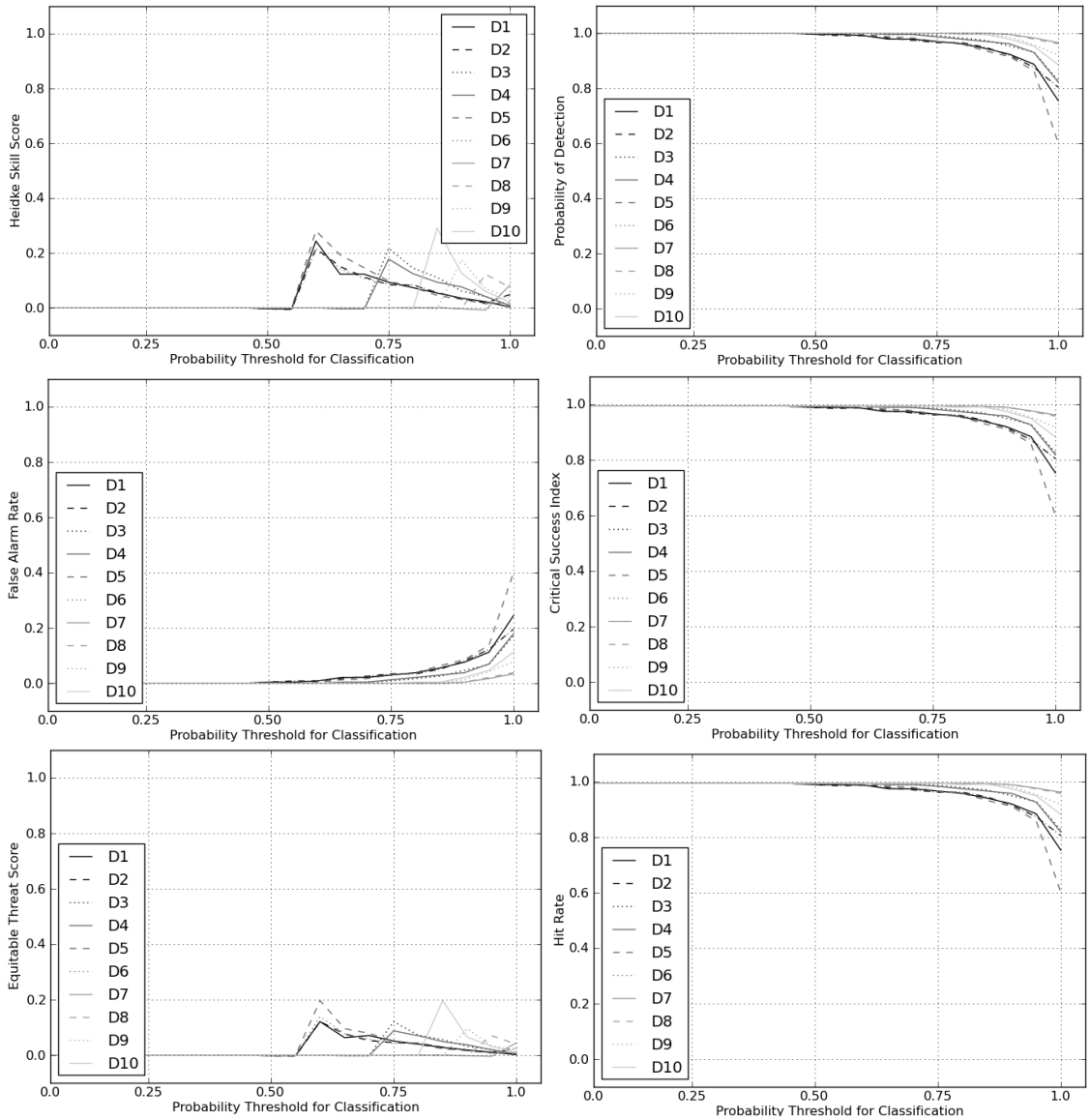


Figure 5.6. Decision tree ensemble verification threshold plots for Hurricane Irene. Each line represents the median score for the 100 ensemble members.

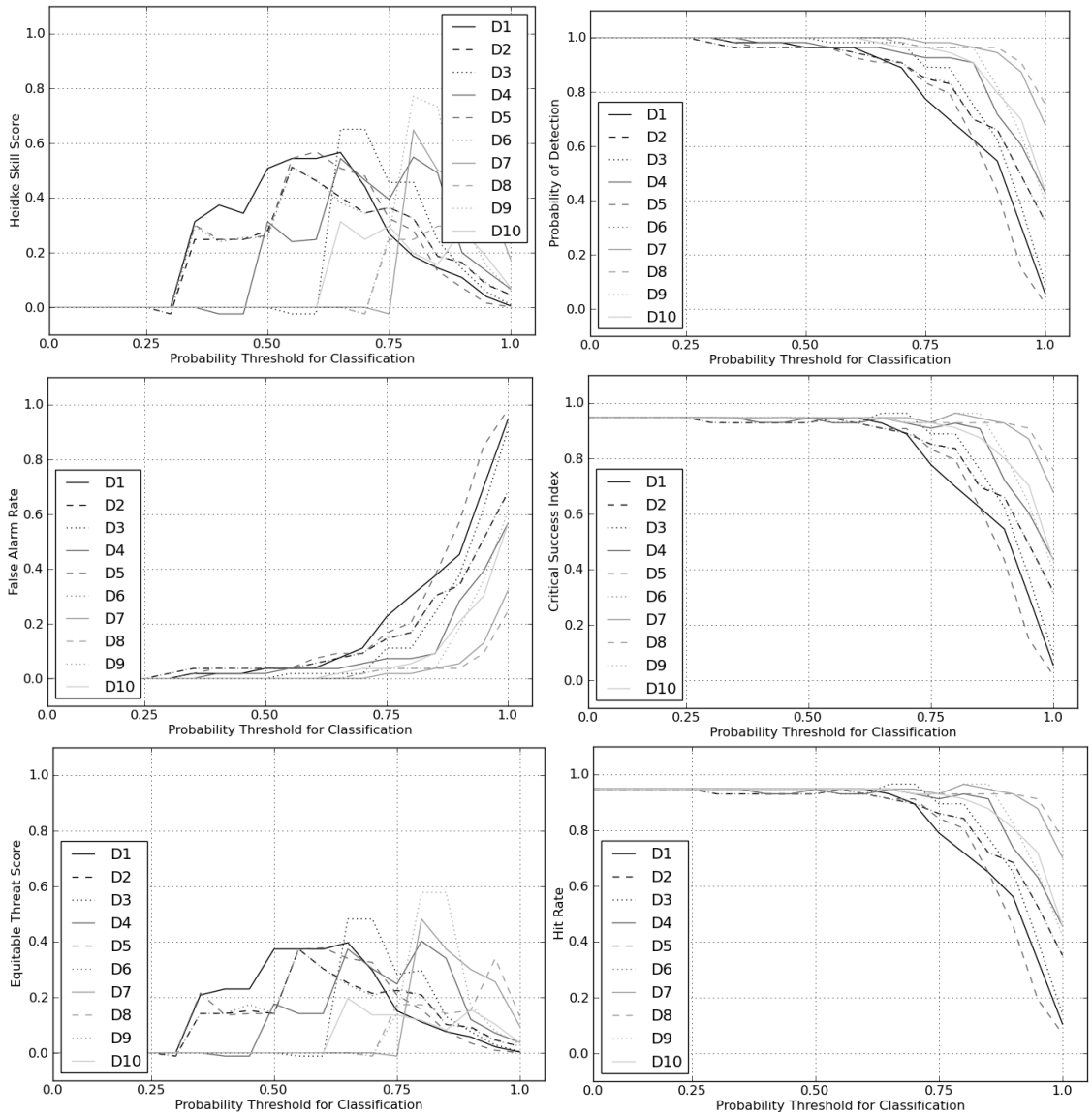


Figure 5.7. Decision tree ensemble verification threshold plots for extratropical storm Ida. Each line represents the median score for the 100 ensemble members.

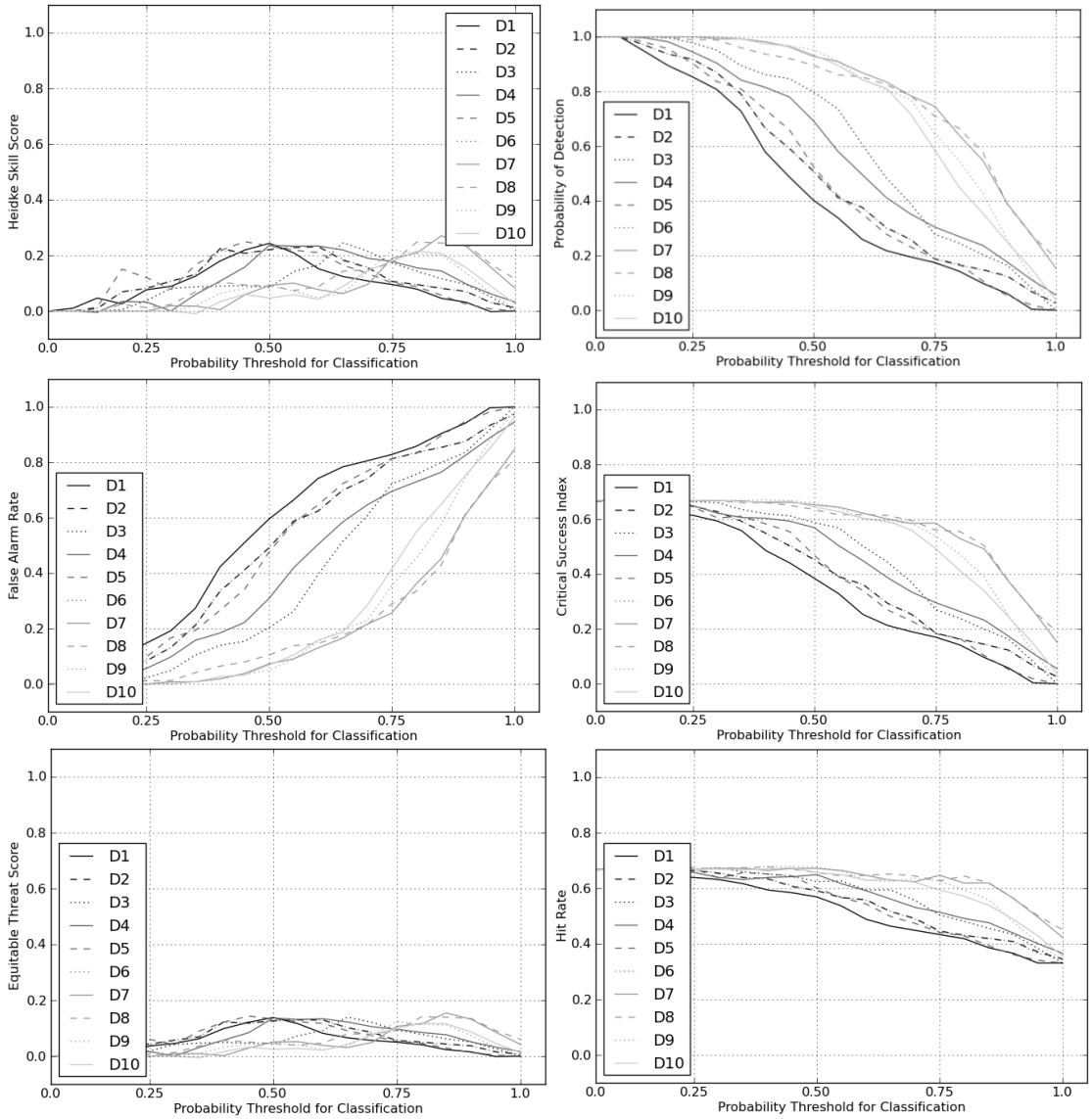


Figure 5.8. Decision tree ensemble verification threshold plots for the 2011 southeast U.S. MCS event. Each line represents the median score for the 100 ensemble members.

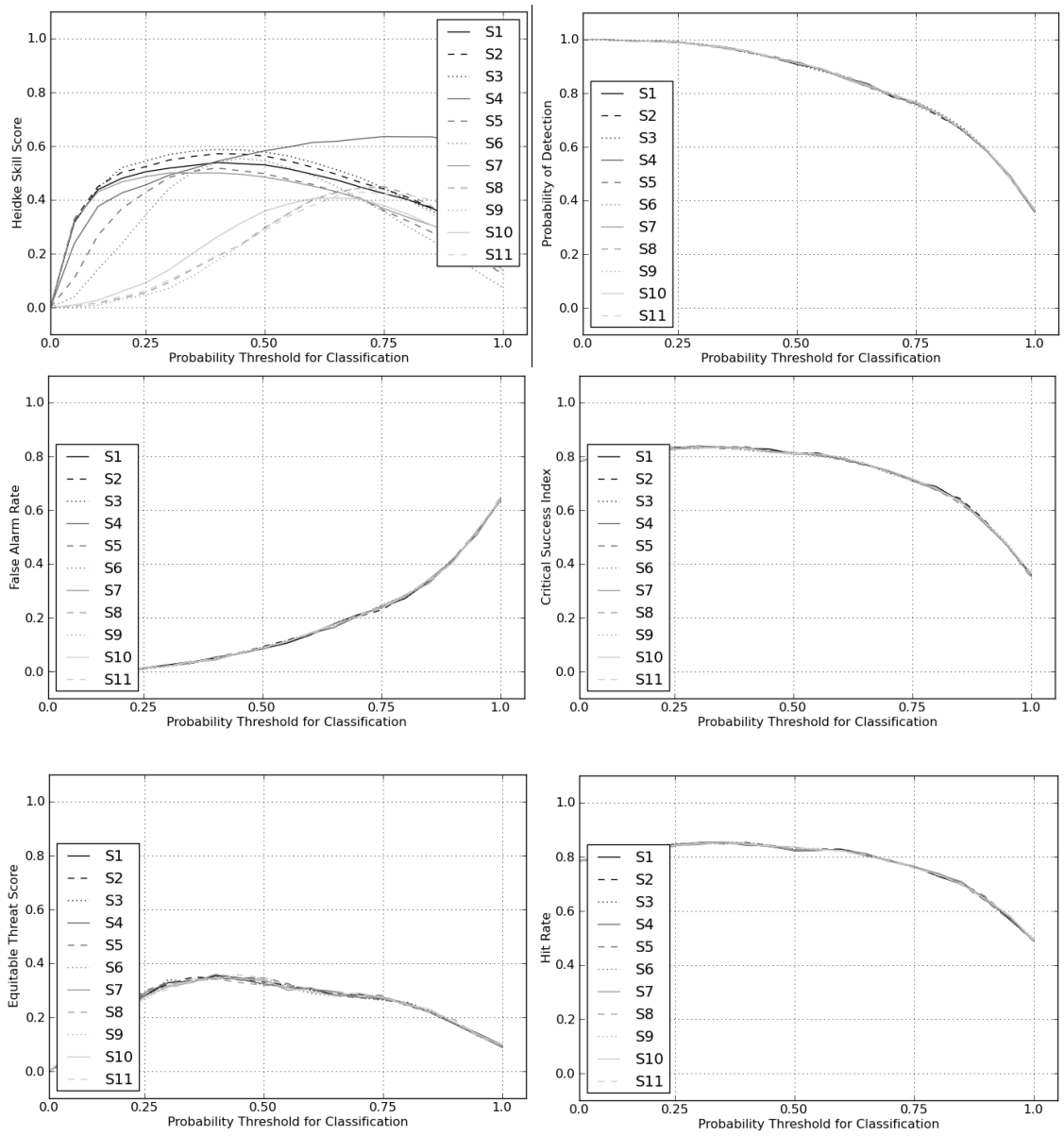


Figure 5.9. Support vector machine ensemble verification threshold plots for the 2010 warm season. Each line represents the median score for the 100 ensemble members.

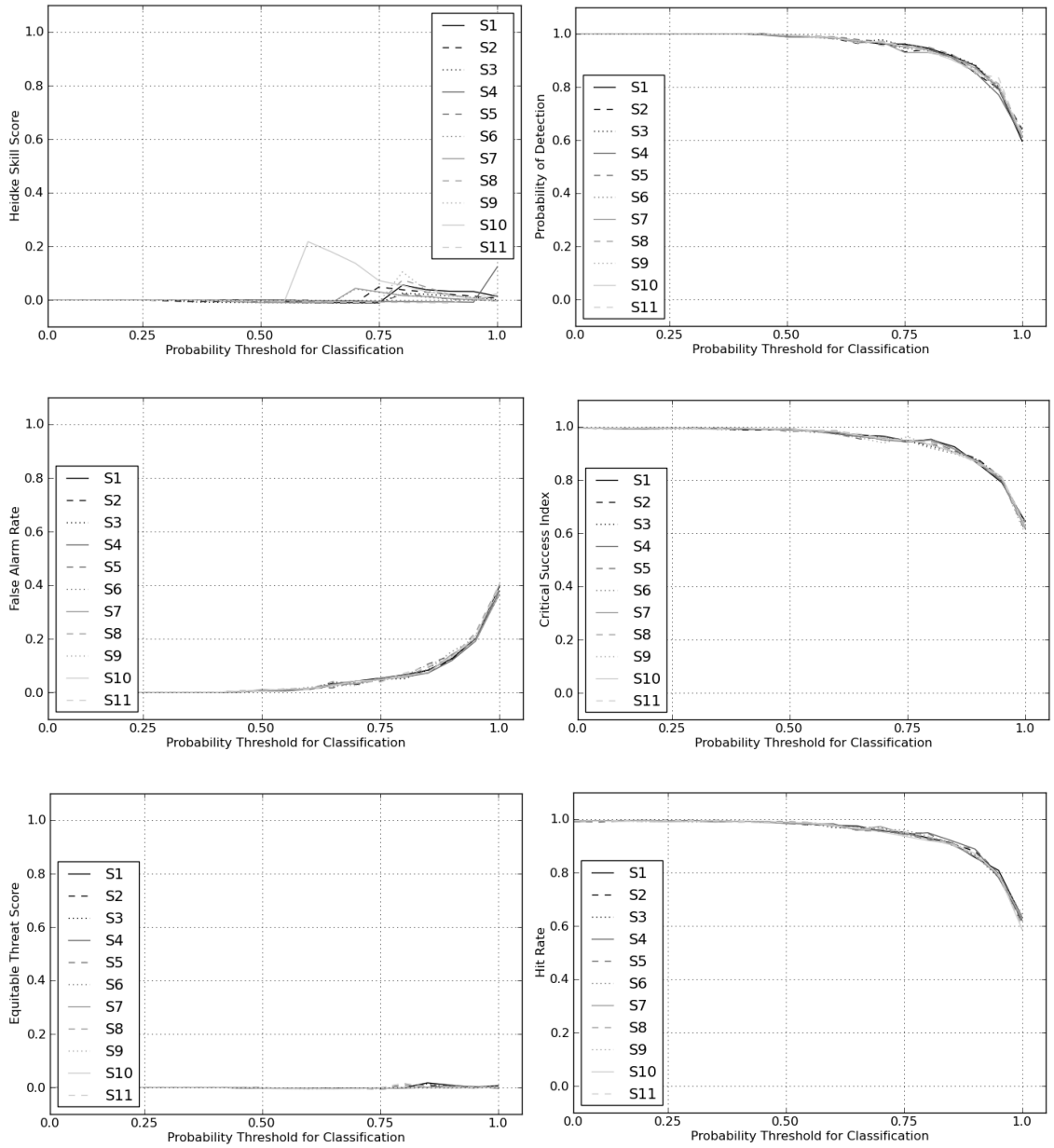


Figure 5.10. Support vector machine ensemble verification threshold plots for Hurricane Irene. Each line represents the median score for the 100 ensemble members.

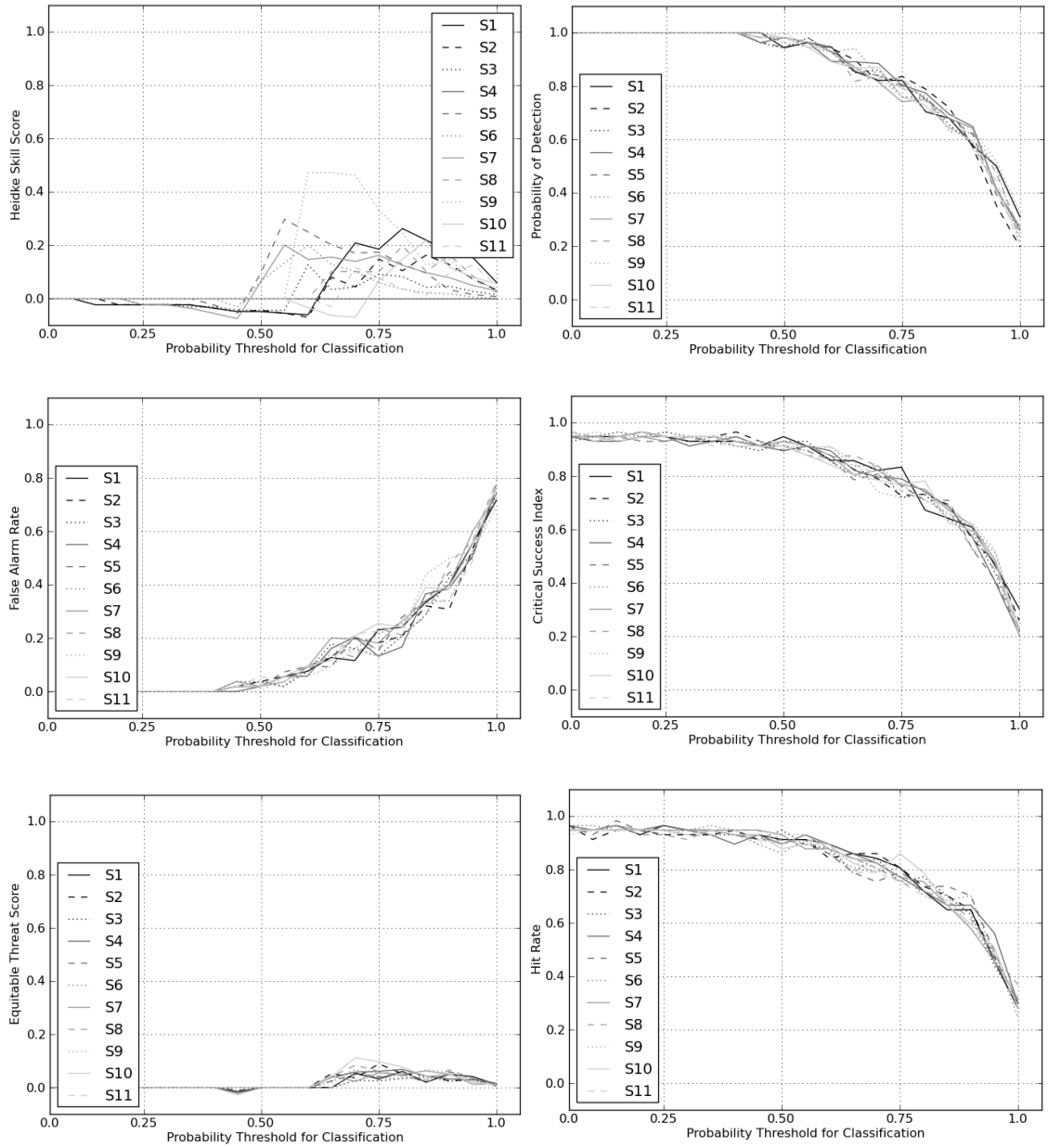


Figure 5.11. Support vector machine ensemble verification threshold plots for extratropical storm Ida. Each line represents the median score for the 100 ensemble members.

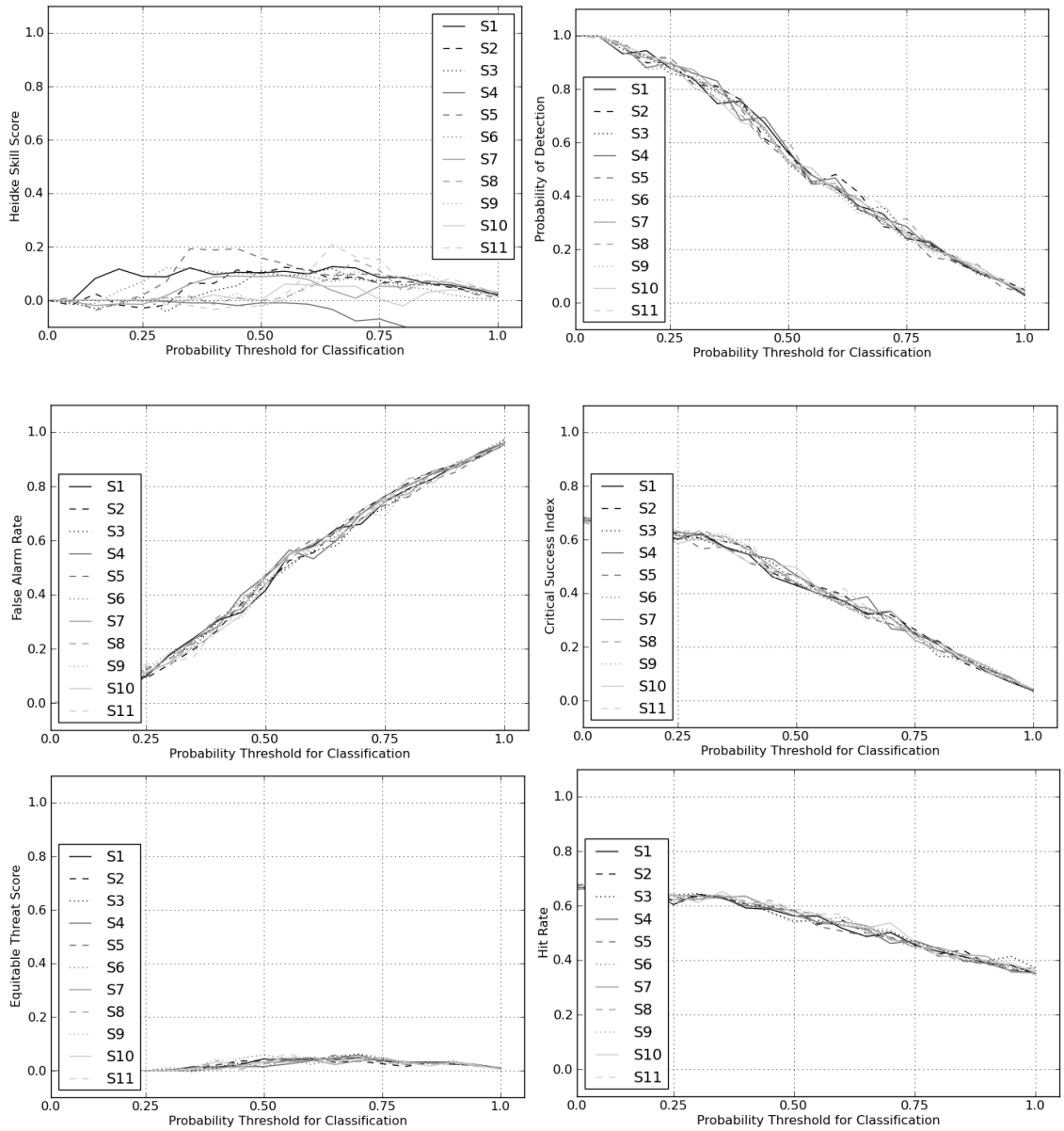


Figure 5.12. Support vector machine ensemble verification threshold plots for the 2011 southeast U.S. MCS event. Each line represents the median score for the 100 ensemble members.

Ensemble Characteristics

For the "all-parameter" decision tree ensembles, the 100 member trees had very similar structures in the top few tiers where the first optimal splits were found for the training samples. The first split or node in every tree was based on the 850-500 hPa lapse rate, and the split threshold was the same (-5.1 deg km^{-1}). The second tier of the trees was where some differences emerged (Fig. 5.13). For data points with 850-500 hPa lapse rate less than -5.1 deg km^{-1} (i.e., more negative), 70% of the next node decision was based on precipitable water, 20% was based on 500 hPa temperature, and smaller portions were based on precipitation efficiency (5%), a second threshold of 850-500 hPa lapse rate (2%), and freezing level height (1%). Where data points were sampled from areas with lapse rate greater than -5.1 deg km^{-1} (i.e., less negative), 67% of the next split was based on 1000-700 hPa mean relative humidity, 20% were based on a second threshold of lapse rate, 8% were based on surface-based CAPE, and 2% looked at 850 hPa relative humidity. Thus, in general, where mid-level lapse rates were steep or strongly negative, the decision trees evaluated total moisture content and mid-level temperature first; where lapse rates were closer to moist adiabatic, decision trees examined low-level inflow humidity and instability.

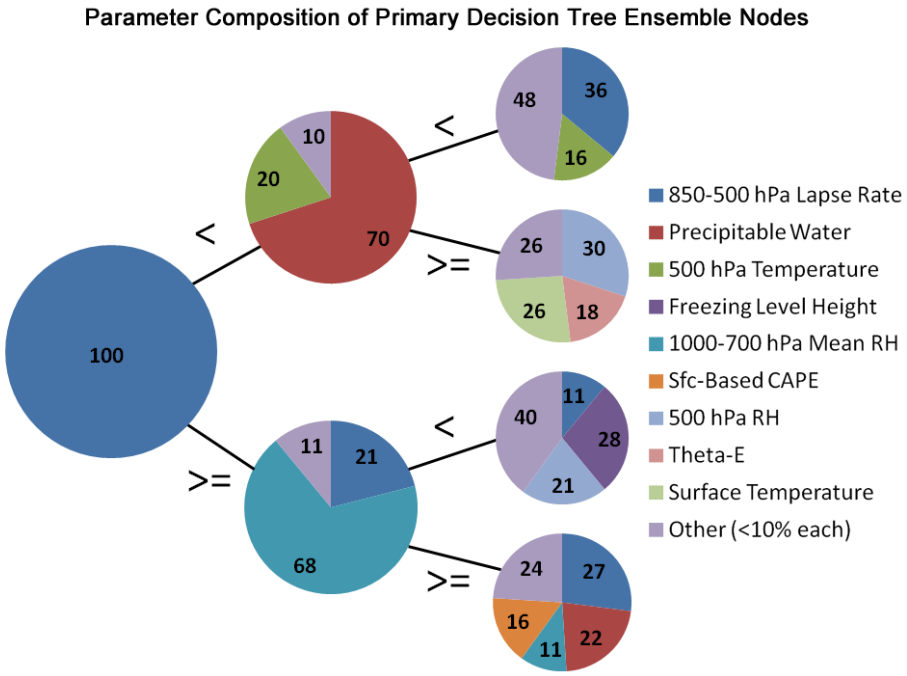


Figure 5.13. Parameter composition (in percent) of the first three tiers of each decision tree in the "all-parameter" ensemble (D2).

While similarity among ensemble members aids the physical interpretation of the classification, diversity among the members is preferred to uniformity as long as all members perform with reasonable accuracy (Brown et al. 2005). The aggregation of a diverse ensemble of accurate classifiers will tend to be more accurate than any individual member, because the members all have different failure modes. By contrast, an ensemble of highly correlated members will not provide a much better classification than any one member. Increasing ensemble diversity to achieve higher accuracy is the motivation behind techniques such as bagging, boosting, and randomized feature subset selection (e.g., random forests). Ideally, each member of the ensemble will have high accuracy on its own for a test dataset and will be statistically independent from all other members, but in practice perfect statistical independence can be difficult to achieve.

Event	Diversity (Q)	Hit Rate	HSS
2010 Warm Season	0.04 (AP)	0.77 (AP)	0.40 (AP)
	0.03 (RF)	0.76 (RF)	0.39 (RF)
	0.28 (SVM)	0.79 (SVM)	0.44 (SVM)
Irene	0.74 (AP)	0.99 (AP)	0.22 (AP)
	0.73 (RF)	0.99 (RF)	0.24 (RF)
	0.66 (SVM)	0.96 (SVM)	-0.01 (SVM)
Ida	0.36 (AP)	0.93 (AP)	0.46 (AP)
	0.24 (RF)	0.95 (RF)	0.54 (RF)
	0.43 (SVM)	0.81 (SVM)	0.14 (SVM)
SE MCS	0.19 (AP)	0.56 (AP)	0.23 (AP)
	0.15 (RF)	0.49 (RF)	0.15 (RF)
	0.40 (SVM)	0.48 (SVM)	0.10 (SVM)

Table 5.2. Comparison of ensemble diversity and accuracy for decision tree ensembles D1 (Random Forest - RF) and D2 (All Parameters - AP), and SVM ensemble S9. The verification statistics were based on the threshold probability with the best mean HSS of the three independent events for each ensemble, which was 55% for the decision trees and 65% for the SVM.

Table 5.2 compares the ensemble diversity between the all-parameter decision tree method, the random forest method, and the SVM, as well as the hit rate and Heidke Skill Score of each for the three independent test cases and the 2010 training period. As described by Brown et al. (2005), ensemble diversity can either be quantified in a pairwise fashion (comparing each member to all the other members and computing a mean value) or a non-pairwise fashion (comparing each member to the ensemble aggregated classification and computing a mean across the ensemble). The Q statistic

was selected for the pairwise comparisons, which is derived from a 2x2 contingency table of ensemble member classifications between two data points (similar to Fig 4.4 except comparing the predicted class between two ensemble members):

$$Q = \frac{AD - BC}{AD + BC} \tag{5.1}$$

Q has similar properties to the Pearson's correlation coefficient, with a range of -1 to +1 and a value of 0 representing statistical independence. Somewhat surprisingly, very little difference in diversity is apparent between the AP and RF ensembles, despite the added randomness of the subset feature selection in the RF. While the RF ensemble has slightly lower Q than the AP ensemble, it does not seem to translate to an improvement in the accuracy or skill of the majority vote classification. With the exception of Hurricane Irene, the SVM ensembles had consistently lower diversity (i.e., higher Q) and verification scores than the decision tree ensembles.

Vertical Profile of Reflectivity Ensemble

When the classification was based on the slope of the VPR below the freezing level rather than the bias of the radar rainfall relative to gauges, the most significant predictors for classification accuracy changed (Fig. 5.14). Rather than being based largely on the 850-500 hPa lapse rate of temperature, accuracy of the VPR-based model appeared to rely more heavily on the height of the freezing level and the 500 hPa relative humidity. Table 5.3 shows that the VPRs exhibiting increasing reflectivity with decreasing height toward the surface (negative VPR slope values) were observed in environments with much higher freezing levels and lower mid-level humidity. Humidity at lower levels in each category were similar, however, and the precipitable water

trended higher in the negative VPR slope environment. The lower humidity at 500-700 hPa may be a reflection of the warmer temperatures at those levels (assuming the storm relative winds are sampling an unsaturated environment), which would be commensurate with the higher freezing level. Alternatively, they could be an indicator of evaporation in the mid-levels of the storm as drier air is advected or mixed into the cloudy air, which could also lead to a reflectivity profile which is increasing with decreasing height as evaporation acts to diminish drop sizes across the DSD spectrum aloft.

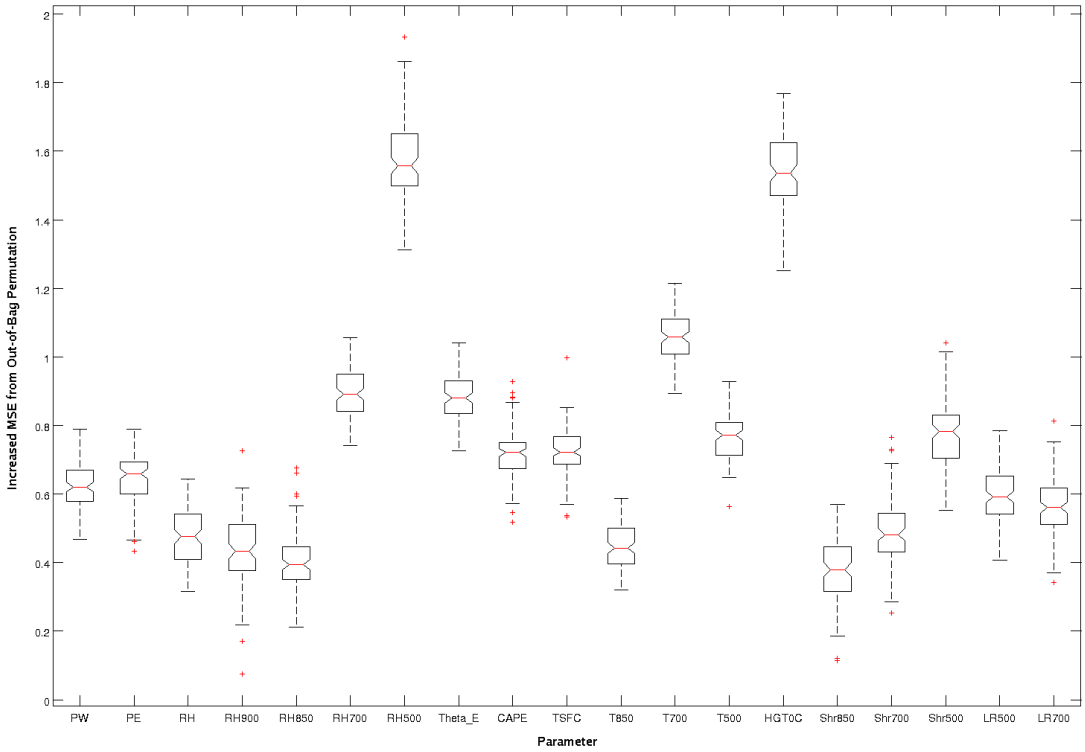


Figure 5.14. Increase in decision tree ensemble mean squared error (normalized by standard deviation) when individual parameters are permuted or randomized.

	VPR Slope < 0 (Warm Rain)		VPR Slope >= 0 (Non-Warm Rain)		Two-Sample
Parameter (units)	**Mean**	**St. Dev.**	**Mean**	**St. Dev.**	**Test ASL**
Precipitable Water (kg m^{-2})	51.7 (0.4)	9.9 (0.3)	42.3 (0.5)	10.4 (0.3)	<0.001
Precipitation Efficiency (kg m^{-2})	43.8 (0.5)	11.1 (0.3)	36.8 (0.5)	11.1 (0.3)	<0.001
1000-700 hPa Mean RH (%)	84.0 (0.4)	10.0 (0.3)	86.2 (0.5)	11.1 (0.4)	<0.001
900 hPa RH (%)	86.1 (0.5)	12.3 (0.6)	86.0 (0.6)	13.5 (0.6)	0.677
850 hPa RH (%)	87.5 (0.5)	11.1 (0.5)	88.4 (0.5)	11.8 (0.6)	0.038
700 hPa RH (%)	77.8 (0.7)	16.7 (0.5)	88.7 (0.6)	13.8 (0.7)	<0.001
500 hPa RH (%)	81.5 (0.7)	17.4 (0.7)	89.6 (0.7)	15.8 (0.9)	<0.001
Equivalent Pot. Temperature (K)	345.3 (0.6)	13.5 (0.4)	330.2 (0.6)	13.6 (0.4)	<0.001
Surface Temperature (C)	24.0 (0.2)	4.8 (0.2)	19.1 (0.2)	5.0 (0.1)	<0.001
500 hPa Temperature (K)	265.8 (0.1)	2.9 (0.1)	262.2 (0.2)	3.6 (0.1)	<0.001
Freezing Level Height (km agl)	4.6 (0.02)	0.5 (0.02)	3.9 (0.03)	0.7 (0.02)	<0.001
CAPE (J kg^{-1})	1105 (55)	1188 (56)	386 (33)	734 (50)	<0.001
850 hPa Temperature (K)	290.4 (0.1)	3.1 (0.1)	287.2 (0.2)	4.1 (0.1)	<0.001
700 hPa Temperature (K)	281.7 (0.1)	2.8 (0.1)	277.5 (0.2)	3.6 (0.1)	<0.001
Sfc-500 hPa Shear (m s^{-1})	15.3 (0.4)	8.6 (0.3)	16.8 (0.4)	8.3 (0.2)	<0.001
Sfc-700 hPa Shear (m s^{-1})	11.9 (0.3)	6.1 (0.2)	12.7 (0.3)	6.6 (0.2)	0.055
Sfc-850 hPa Shear (m s^{-1})	10.6 (0.3)	5.8 (0.2)	10.7 (0.3)	5.8 (0.2)	0.758
850-500 hPa Lapse Rate (K km^{-1})	-4.9 (0.02)	0.5 (0.02)	-4.9 (0.02)	0.5 (0.02)	0.989
850-700 hPa Lapse Rate (K km^{-1})	-4.4 (0.04)	0.9 (0.04)	-4.6 (0.04)	1.0 (0.04)	0.604

Table 5.3. Mean values of the bootstrapped mean and standard deviation for all RUC parameters used to train the VPR-based classification model (drawn from a balanced and weighted sample of the 2010 warm season population). Numbers in parenthesis represent the standard deviation of the target statistic among the bootstrap samples. The rightmost column contains the achieved significance level or p-value of a two-sample permutation test that the RUC parameter distributions of the two bias classes have equal means.

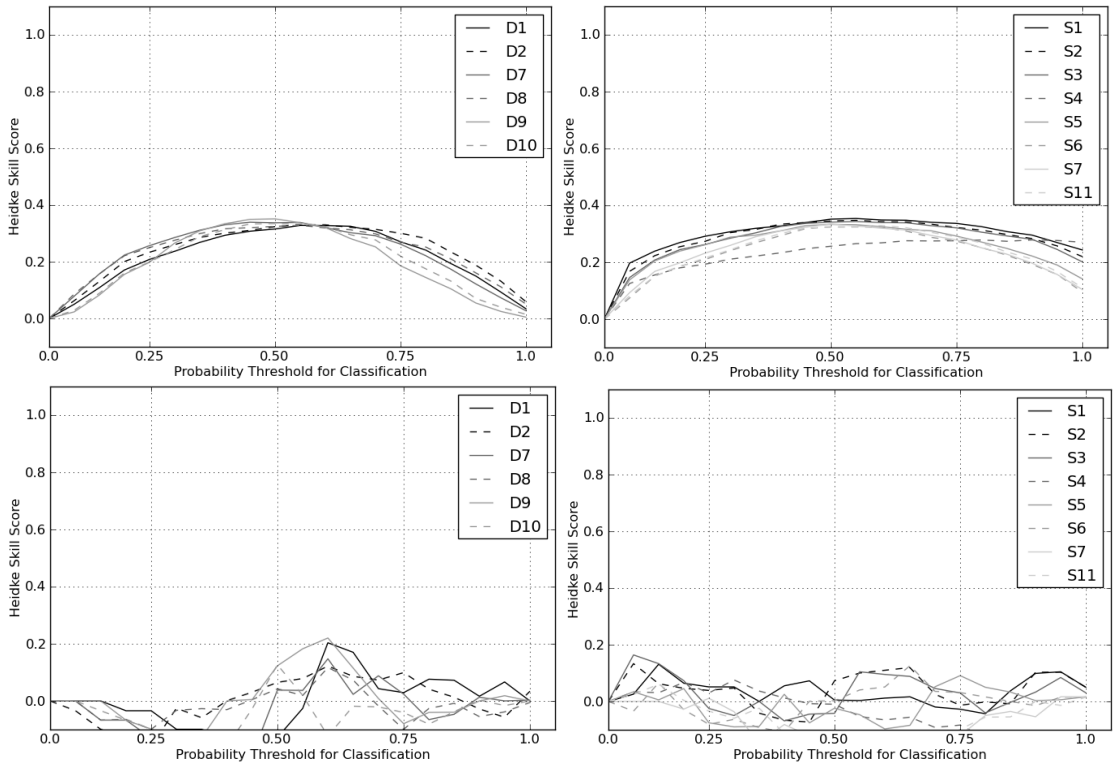


Figure 5.15. Verification threshold plots for the 2010 training period (top) and extratropical storm Ida (bottom) based on the VPR classification. Decision tree ensembles are shown on the left, and SVM ensembles are on the right. Each line represents the median score for the 100 ensemble members.

Overall, classification accuracy for the VPR-based models were not as high as the gauge bias-based models, including for the training period. Furthermore, the models trained using the 2010 warm season did not appear to generalize well to the independent cases, as is shown in Fig. 5.15 in a comparison between the Heidke Skill Scores from the training period and extratropical storm Ida (the other events performed similarly to Ida). The independent case datasets were constructed somewhat differently from the training set, however. For the training set, the mean of the RUC grid boxes within 80 km of the radar was computed as the predictors of the radar's VPR slope, whereas for the independent cases the storm relative inflow method was used to retrieve the environmental predictors at the locations of gauges within 100 km of the radar. Thus,

it's possible that the environmental predictors that the VPR classification was trained on was more reflective of the thermodynamics of the atmosphere modified by existing precipitation whereas the environments of the independent cases represented more of the unmodified inflow to the storms. The sampling differences may explain why the variable importance of the VPR set emphasized freezing level height over the mid-level lapse rates, which would tend to be more uniformly moist adiabatic across the training set when sampled within the precipitation. The lapse rates in Table 5.3 suggest that this is the case, with both categories showing values that are less negative (closer to moist adiabatic) than the median lapse rates computed for the gauge bias-based categories (Table 5.1).

When the VPRs were trained on environments derived from the storm relative inflow rather than the mean value around the radar, lapse rates and wind shear replaced the 500 hPa relative humidity as significant parameters and freezing level height remained significant (Fig. 5.16). In other words, the significant predictors fell more in line with those identified for the gauge bias-based classification. However, the spatial distribution of probabilities indicate that even when using storm-relative inflow for selection of the environment, the VPR-based probabilities are still more similar to each other than to the gauge bias-based classification (Fig. 5.17).

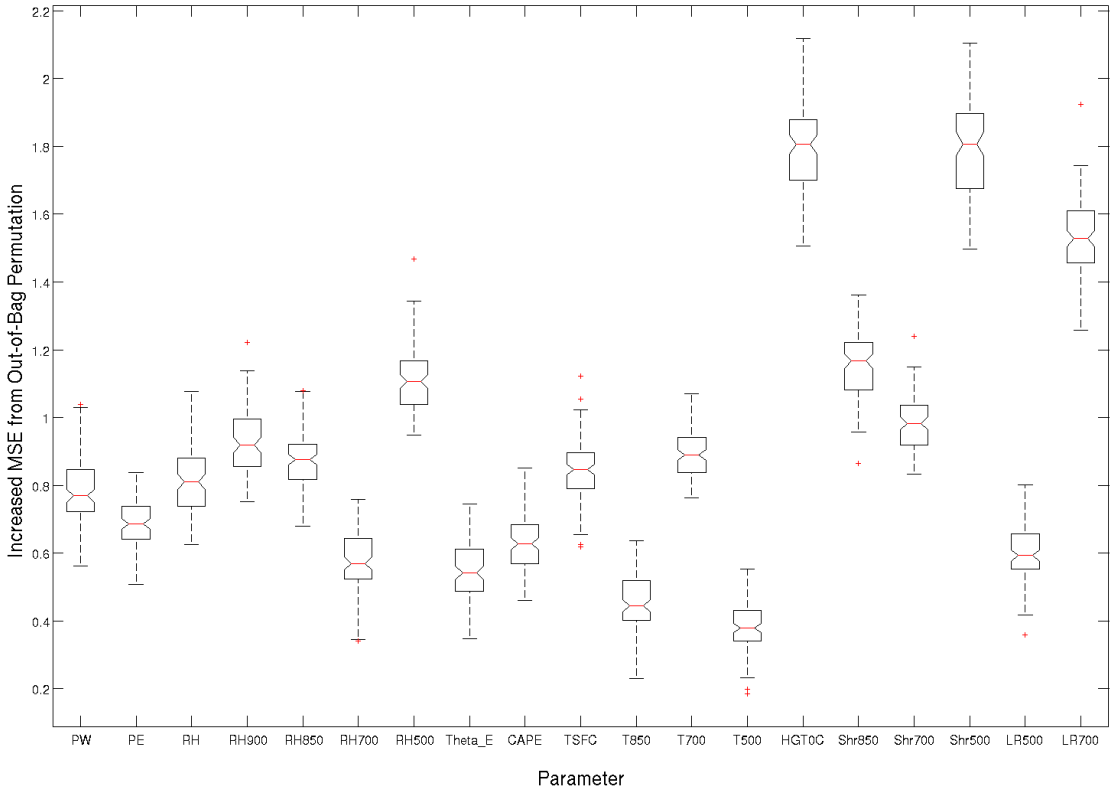


Figure 5.16. Increase in decision tree ensemble mean squared error (normalized by standard deviation) when individual parameters are permuted or randomized. Classification criteria were the VPR slope of radars based on the storm relative inflow environment around nearby rain gauges.

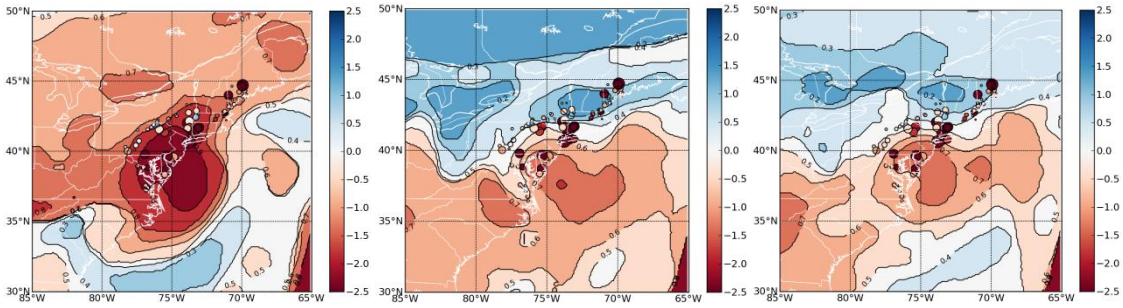


Figure 5.17. Comparison of decision tree based probabilities (ensemble D2) valid 06Z 28 Aug 2011 for the gauge bias-based criterion (left), the VPR slope criterion using mean radar environments (middle), and the VPR slope criterion using storm-relative inflow environments (right).

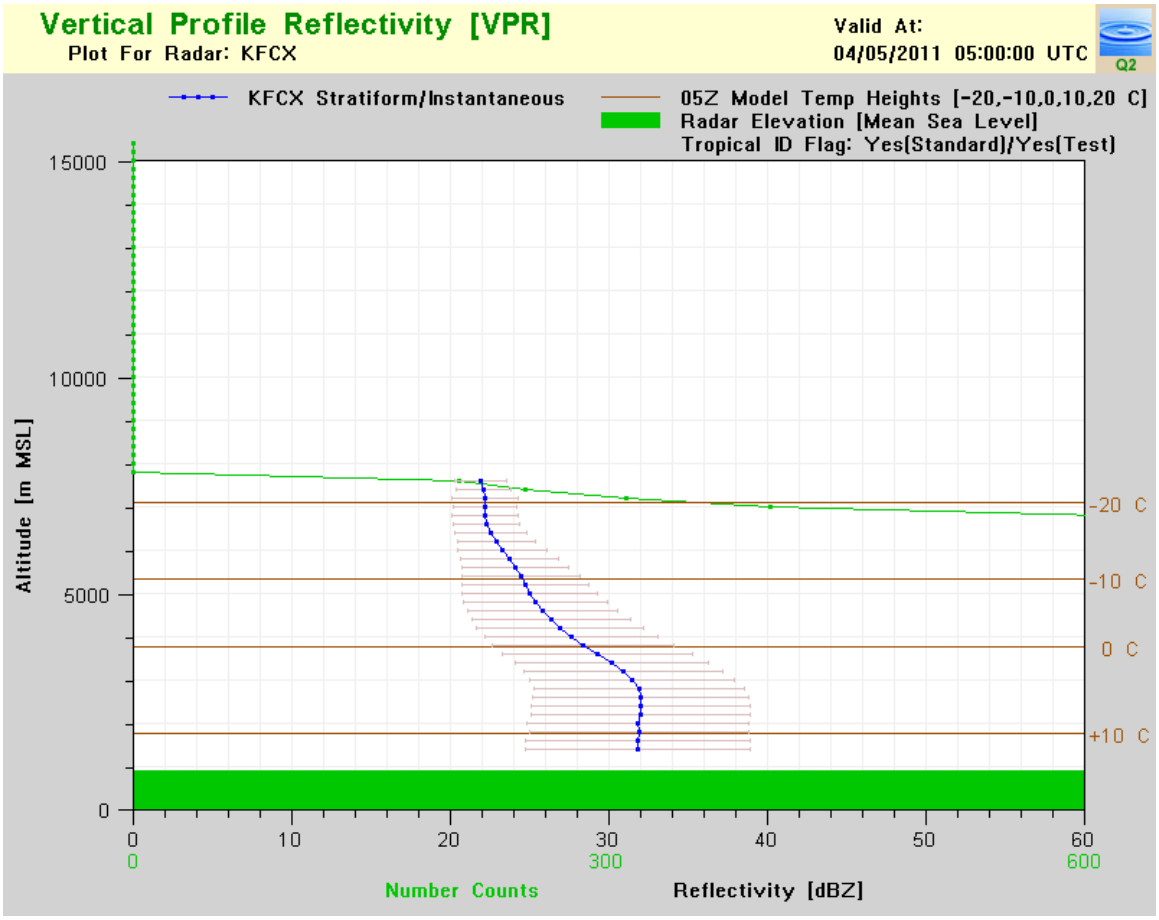


Figure 5.18. Instantaneous (i.e., from one volume scan) mean vertical profile of reflectivity observed during the Southeast U.S. MCS event at KFCX with standard deviation shown.

According to this analysis, environments conducive for VPR structure exhibiting increasing reflectivity with decreasing height are not solely useful for predicting where enhanced rain rates will occur within the next hour. However, this result may well be due to limitations in how the VPR slope is measured and computed in the NMQ framework rather than a physical disconnect between vertical reflectivity structure and changes in rain rate. The VPRs used to train the ensemble represent averages of reflectivity only observed within 80 km of the radar and also undergo interpolation and smoothing. The averaging and smoothing, especially on heterogeneous rainfall types

within the radar's radius of influence, tend to center the lower part of the VPR to a slope of near zero. Figure 5.18 shows an averaged VPR with the standard deviation of reflectivity at each vertical level (denoted with error bars). The large spread of the error bars even for just the stratiform rain type shows how much the VPR can vary even for a 5-minute snapshot within a short distance of the radar, so the mean profile may not be representative of all the point measurements of rainfall at the gauges within the radar's umbrella. The same could also be said for hourly RUC analyses on a 20-km grid resolution, though a 20-km grid would have better spatial resolution than the 160-km diameter circles around radars that are as much as 300 km apart.

The advantage that the VPRs have over the RUC-based rain rate identification is the temporal resolution. New VPRs are computed for every volume scan of the radar whereas the RUC analyses are only available hourly. For fast moving systems or for rapidly changing conditions, the probability of enhanced rain rate computed at the start of the accumulation hour will not reflect the changes that occur during the following hour. In order for the RUC-based probabilities to be most advantageous on the time scale of 5 minute rain rates, the probabilities would either need to be interpolated between the 00-hour analysis and the 01-hour forecast or an object tracking algorithm would be needed to identify and follow storm elements as they move away from the location where the probability was calculated. These are certainly potential areas of future study and improvement, and NMQ's experimental development environment easily facilitates such testing in a real-time mode.

As the WSR-88D network is upgraded to dual-polarization capability nationwide, changes are being made to the algorithms that compute rain rate

operationally. Rather than assigning a single Z-R at each radar based on the forecaster's interpretation of the predominant environment, an automated algorithm assigns different dual-polarization-based rain rate functions for each precipitation type delineated by the Hydrometeor Classification Algorithm or HCA (Ryzhkov et al. 2005; Giangrande and Ryzhkov 2008). The rain rates were derived based on rainfall events in Oklahoma, however, and employing them in other climates and regions around the country are revealing that dual-polarization-based rainfall rates may not be universally applicable in all places. Thus, environment-based analyses and algorithms like the one within this study still hold potential utility in the dual-pol era, particularly where radars continue to struggle with adequately sampling the lowest elevations that are most similar to rain rates measured at the surface.

VPRs vs. Gauge Bias: Effects on Derived QPE

Due to the spatial differences apparent between the VPR-based and gauge bias-based model ensembles, it is important to assess which ensembles provide the most improvement over the current NMQ system in terms of reducing overall QPE error through better delineation and assignment of the enhanced rain rates. For this comparison, the same Z-Rs were applied in the new precipitation typing algorithm that are used currently for Q2. Thus, any change in Q2 QPE error should be the result of the spatial delineation of precipitation types alone.

Ensemble D2 (the All-Parameter ensemble with weighted sampling on classes balanced through undersampling the majority class) was chosen for both the gauge bias and VPR ensembles. The threshold probability for delineating enhanced rain rates was

selected based on the probability that maximized the Heidke Skill Score for the three independent events for ensemble D2 (Figs. 5.5-5.8, 5.15), which was consistently around 0.6 on a scale of 0-1. Thus, for both the VPR and gauge bias ensembles, areas with probabilities equal to or higher than 0.6 were flagged as conducive for enhanced warm rain processes. Within those areas, Q2 grid points with hybrid scan reflectivity greater than or equal to 30 dBZ were assigned the tropical Z-R. Areas that did not meet these criteria were assigned either convective or stratiform Z-Rs based on Q2's existing precipitation typing methodology, described in Chapter 3. Once hourly rainfall accumulations were computed from the 5-minute rain rates, they were compared against the QCed HADS gauge set for verification using mean difference bias, root mean squared error (RMSE), a bias ratio (the ratio between the sums of total accumulations within the domain for the two QPE products), and coefficient of determination (R^2).

The new enhanced rain rate delineation schemes performed with mixed results for the three independent cases (Figs. 5.19-5.21; Tables 5.4-5.6). Relative to the legacy Q2 tropical identification algorithm based solely on VPR slope derived from the radar data, the probability-based algorithm provided major improvements in areas where the tropical Z-R was either inadequately used when needed or overused where not appropriate. Additionally, the Hurricane Irene event (Fig. 5.19) demonstrates the advantage of basing the precipitation type on model input rather than radar input when radar data is missing. KDIX, the NEXRAD radar serving the Philadelphia area, was offline for several hours while the event was underway. The mosaic of adjacent radars allows for precipitation rates to be calculated in the vicinity of the missing radar (albeit using higher elevation tilts). VPRs cannot be computed in Q2 for that radar, however, so

the precipitation type is left as stratiform or convective. The probability-based algorithms both generate high probabilities near KDIX even while the radar is missing, so the tropical Z-R is delineated as it would be if KDIX was available. The use of the tropical Z-R around KDIX led to large reductions in QPE mean absolute error over New Jersey and southern New York. Between the gauge-based and the VPR-based probabilities for Irene, the gauge-based algorithm had lower overall bias and lower RMSE (Table 5.4).

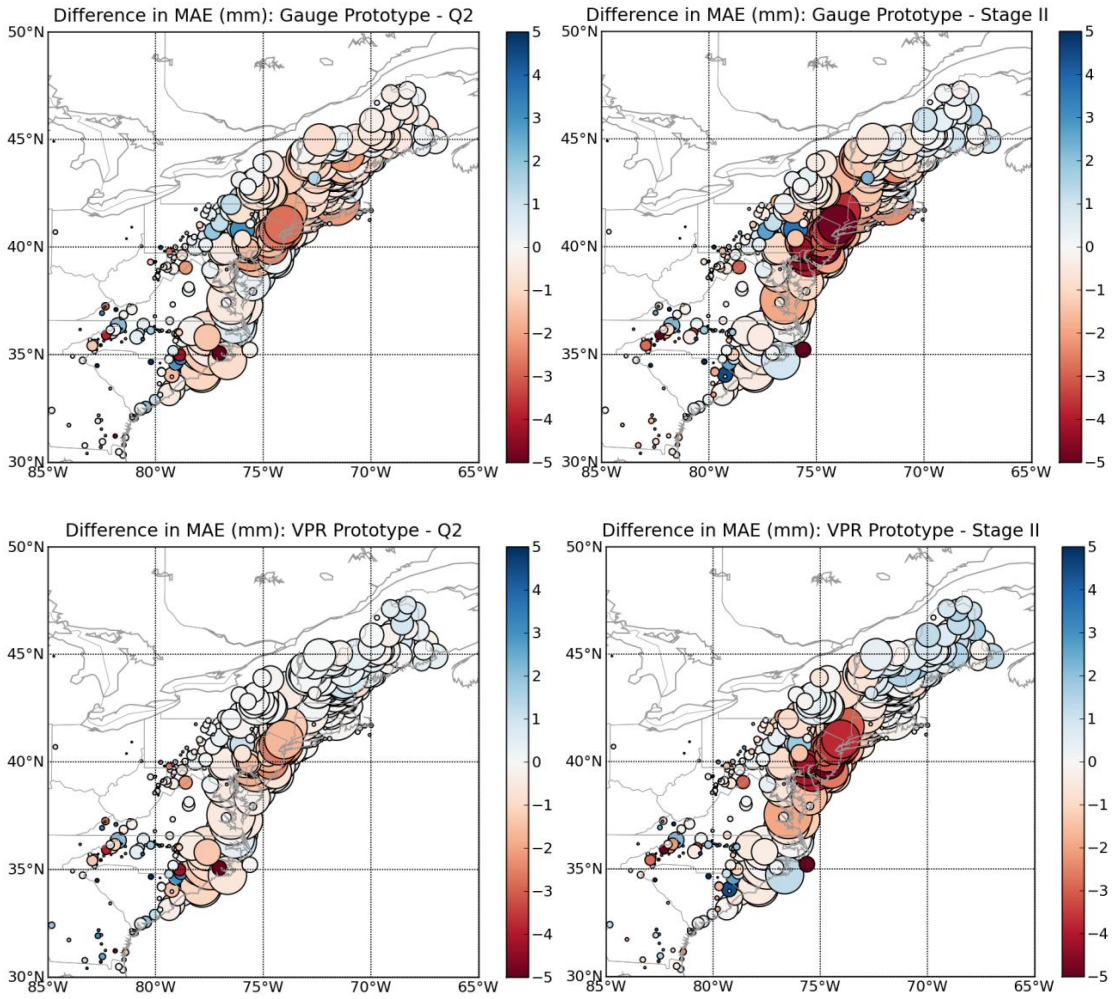


Figure 5.19. Comparison of mean absolute error (MAE) between the new prototype rainfall rates and the legacy products (Q2 and NWS Stage II, respectively) for Hurricane Irene. Negative (warm) colors denote improvement from the prototype QPE product.

QPE Product	Mean Bias	RMSE	Ratio Bias	R^2
Stage II	-2.29	5.77	0.58	0.24
Q2	-2.06	4.44	0.62	**0.60**
Gauge-Based	**-0.80**	**3.96**	**0.85**	0.57
VPR-Based	-1.62	4.24	0.70	0.55

Table 5.4. Verification scores for Hurricane Irene between the three QPE products tested.

For the April MCS event, however, the VPR-based algorithm outperformed the gauge-based algorithm, but the perhaps more interesting result was that Stage II was more accurate than any of the Q2 products (Table 5.5). Q2 and both of the prototype algorithms overestimated rainfall amounts for the event as shown in the bias values, while Stage II underestimated. The probable explanation is that the Weather Forecast Offices impacted by the event used only the convective Z-R function for the storm duration, whereas Q2 and the two prototypes assigned some of the heavier rain areas as tropical. RMSE is sensitive to outliers with large errors, so while the mean difference bias was improved with the new algorithms, it's possible that the tropical Z-R led to major overestimates of rain rates for a few gauges that impacted the RMSE. The comparisons of the new algorithms to the Stage II mean bias in Figure 5.20 (right panels) do indeed show that precipitation estimates for a few gauges were much worse with the new algorithms than they were with Stage II. When compared to the legacy Q2 method for delineating enhanced rain rates, however, both prototype algorithms provided substantial improvement in QPE accuracy (left panels).

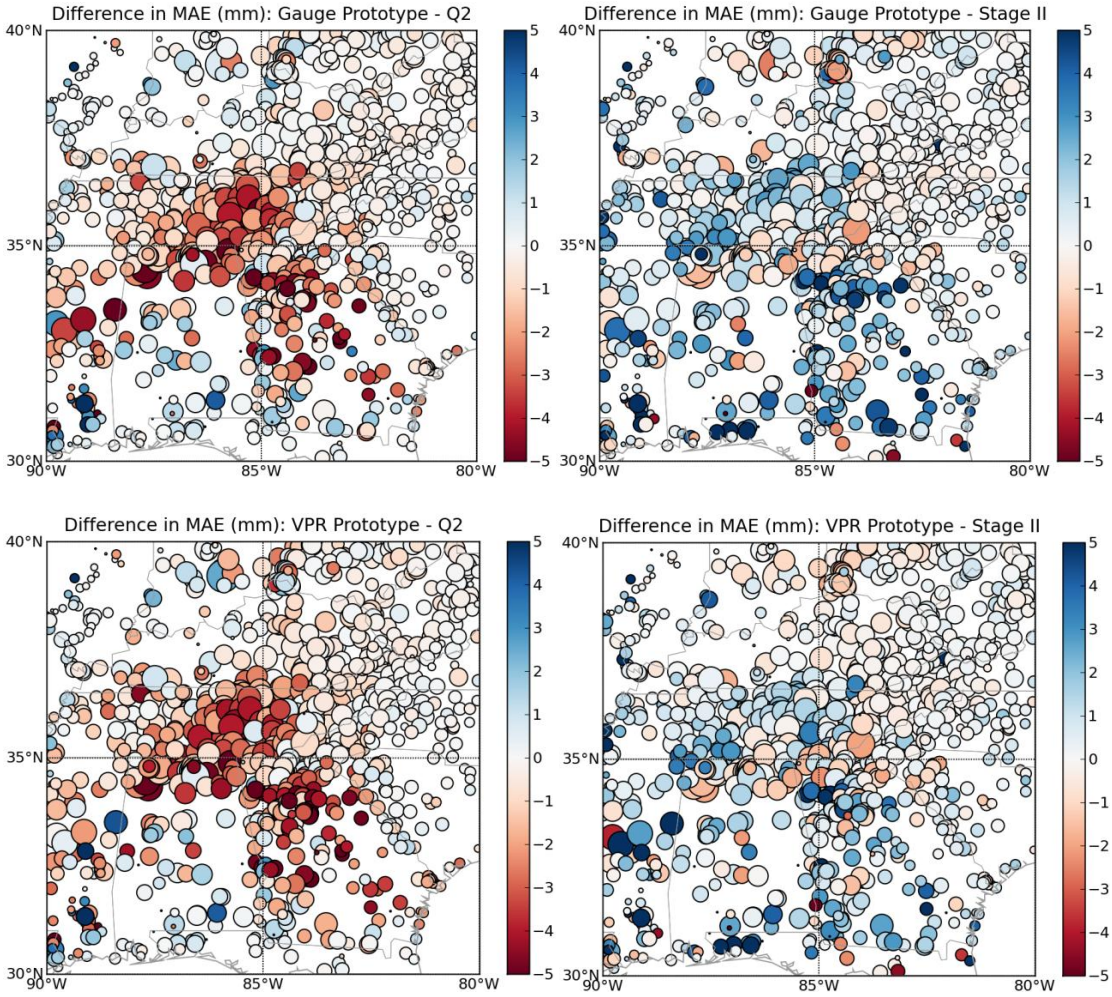


Figure 5.20. Comparison of mean absolute error (MAE) between the new prototype rainfall rates and the legacy products (Q2 and NWS Stage II, respectively) for the April 2011 Southeast U.S. MCS event. Negative (warm) colors denote improvement using the prototype QPE product.

QPE Product	Mean Bias	RMSE	Ratio Bias	R^2
Stage II	-0.63	**2.78**	0.85	**0.66**
Q2	1.66	5.04	1.41	0.60
Gauge-Based	0.63	3.88	1.16	0.54
VPR-Based	**0.39**	3.70	**1.09**	0.54

Table 5.5. Verification scores for the April 2011 Southeast U.S. MCS event between the three QPE products tested.

For the extratropical storm event (Ida), Table 5.6 shows that all three Q2 products outperformed Stage II across all verification metrics. Among the Q2 suite, the legacy product performed only slightly better than the gauge-based prototype, while the VPR-based prototype had the best RMSE. In this instance, it appears that in an overall sense relative to Stage II, the legacy Q2 VPR-based tropical Z-R performed well on its own without additional environmental inputs. All four QPE products underestimated hourly rainfall accumulations similarly to Hurricane Irene even when using the tropical Z-R, which suggests that either the Z-R was not appropriate for the event or that the warm rain processes were so efficient in these tropical cases that the reflectivity observed aloft is an underestimation of what's occurring near the surface. Because neither the gauge-based nor VPR-based ensembles emerged as clearly superior for every independent case tested in terms of QPE accuracy, further analysis will be needed to determine if there are systematic storm type-specific advantages of using one over the other. Three cases is clearly too small a sample to reach a definitive conclusion to justify using one exclusively over the other within the real-time NMQ environment.

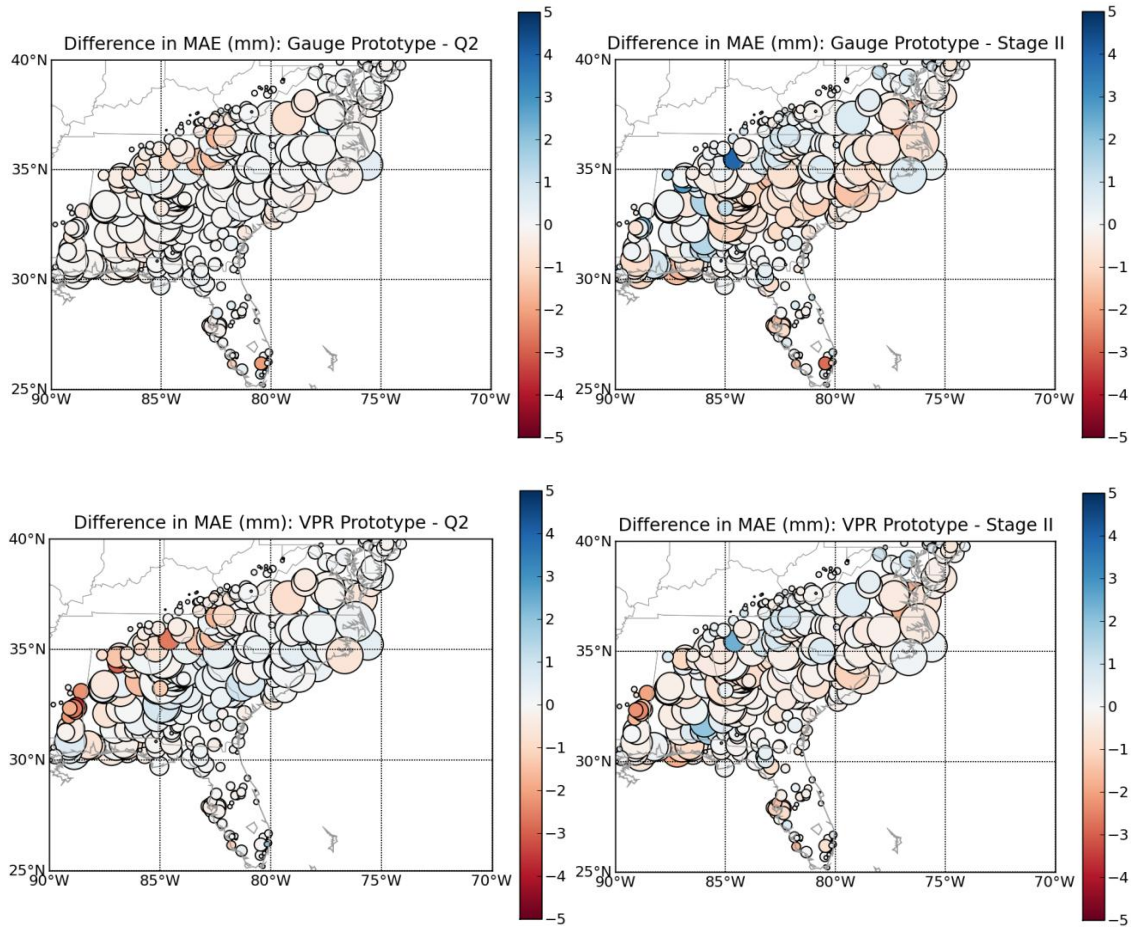


Figure 5.21. Comparison of mean absolute error (MAE) between the new prototype rainfall rates and the legacy products (Q2 and NWS Stage II, respectively) for extratropical storm Ida. Negative (warm) colors denote improvement using the prototype QPE product.

QPE Product	Mean Bias	RMSE	Ratio Bias	R^2
Stage II	-1.18	3.64	0.60	0.22
Q2	**-0.23**	3.62	**0.92**	0.25
Gauge-Based	-0.33	3.52	0.89	**0.26**
VPR-Based	-0.71	**3.45**	0.76	**0.26**

Table 5.6. Verification scores for extratropical storm Ida between the three QPE products tested.

Impact of Model Resolution: 20-km RUC vs. 13-km RUC

As models are updated with higher-resolution grids and/or data assimilation methods, it is important to understand what impact the changes have on algorithms that rely on them for input. When the 13-km RUC model was released in 2006, a 20-km analysis continued to be produced which was a simple resampling of the 13-km fields. To examine what impact changing the model resolution has on the probability of enhanced rain rate and its most important input variables, the 13-km analysis was processed for the 2010 warm season in the same manner that the 20-km analysis was for training the machine learning models. Table 5.7 shows the results of a two-sample permutation test between the variables from each model resolution using the storm-relative wind inflow retrieval method.

It is clear in Table 5.7 that parameters varying significantly over short distances in the vicinity of precipitation are impacted by the change in model resolution, even when the 20-km grid is directly derived from the 13-km fields. Moisture variables such as precipitable water and relative humidity appear to have the largest difference, as well as surface-based CAPE. The 20-km resample of the 13-km grid likely extends the influence of rain-cooled air to further distances from where the precipitation is occurring such that the "inflow" variables retrieved from the 20-km analysis may not be truly characteristic of the near-storm environment. This is most clearly apparent when looking at the difference in CAPE between the two analyses because of the fact that CAPE is converted to other forms of energy during the formation and evolution of convection. Thus, the near-storm environment would generally be expected to have higher values of CAPE.

	13-km RUC Analysis		20-km RUC Analysis		Two-Sample Test ASL
Parameter (units)	**Mean**	**St. Dev.**	**Mean**	**St. Dev.**	
Precipitable Water (kg m^{-2})	50.0 (0.2)	10.8 (0.2)	55.3 (0.2)	9.1 (0.1)	<0.001
Precipitation Efficiency (kg m^{-2})	39.8 (0.3)	12.8 (0.2)	46.3 (0.2)	11.1 (0.1)	<0.001
1000-700 hPa Mean RH (%)	78.0 (0.3)	13.0 (0.2)	83.0 (0.2)	10.8 (0.2)	<0.001
900 hPa RH (%)	80.0 (0.3)	15.2 (0.2)	84.5 (0.3)	12.6 (0.2)	<0.001
850 hPa RH (%)	81.9 (0.3)	15.0 (0.3)	87.9 (0.2)	11.0 (0.3)	<0.001
700 hPa RH (%)	74.7 (0.4)	19.0 (0.3)	79.6 (0.4)	16.9 (0.3)	<0.001
500 hPa RH (%)	72.8 (0.6)	25.8 (0.3)	83.5 (0.4)	19.7 (0.4)	<0.001
Equivalent Pot. Temperature (K)	348.3 (0.3)	11.9 (0.2)	348.9 (0.2)	11.2 (0.2)	0.086
Surface Temperature (C)	25.6 (0.1)	4.6 (0.1)	25.3 (0.1)	4.4 (0.1)	0.105
500 hPa Temperature (K)	266.2 (0.1)	2.8 (0.1)	266.2 (0.1)	2.6 (0.05)	0.481
Freezing Level Height (km agl)	4.6 (0.01)	0.5 (0.01)	4.6 (0.01)	0.5 (0.02)	0.763
CAPE (J kg^{-1})	677 (27)	1316 (28)	1297 (27)	1257 (28)	<0.001
850 hPa Temperature (K)	291.1 (0.1)	2.8 (0.1)	291.2 (0.1)	2.5 (0.05)	0.395
700 hPa Temperature (K)	281.6 (0.1)	2.6 (0.1)	281.8 (0.1)	2.3 (0.04)	0.021
Sfc-500 hPa Shear (m s^{-1})	12.0 (0.2)	7.6 (0.1)	12.3 (0.2)	7.5 (0.2)	0.323
Sfc-700 hPa Shear (m s^{-1})	9.5 (0.1)	5.6 (0.1)	10.2 (0.1)	6.1 (0.1)	<0.001
Sfc-850 hPa Shear (m s^{-1})	8.3 (0.1)	5.3 (0.1)	9.4 (0.1)	6.1 (0.1)	<0.001
850-500 hPa Lapse Rate (K km^{-1})	-5.0 (0.01)	0.5 (0.01)	-5.0 (0.01)	0.5 (0.01)	0.329
850-700 hPa Lapse Rate (K km^{-1})	-4.8 (0.02)	0.9 (0.01)	-4.8 (0.02)	0.8 (0.01)	0.183

Table 5.7. Mean values of the bootstrapped mean and standard deviation for RUC parameters from both the 13-km and 20-km analyses (drawn from the 2010 warm season population). Numbers in parenthesis represent the standard deviation of the target statistic among the bootstrap samples. The rightmost column contains the achieved significance level or p-value of a two-sample permutation test that the RUC parameter distributions of the two model resolution classes have equal means.

Figure 5.22 shows the variable importance computed from the gauge-based decision tree ensemble training when the 13-km analysis was used as input in place of the 20-km analysis (shown in Fig. 5.4). While freezing level is still a significant predictor, the 850-500 hPa lapse rate and mean RH are no longer the top variables. The height of the freezing level is not likely to change drastically in the vicinity of rainfall except for the possible case of diabatic heating aloft within the center of hurricanes and MCSs. Mid-level temperature lapse rates and low-level relative humidity, however, can be modified by precipitation.

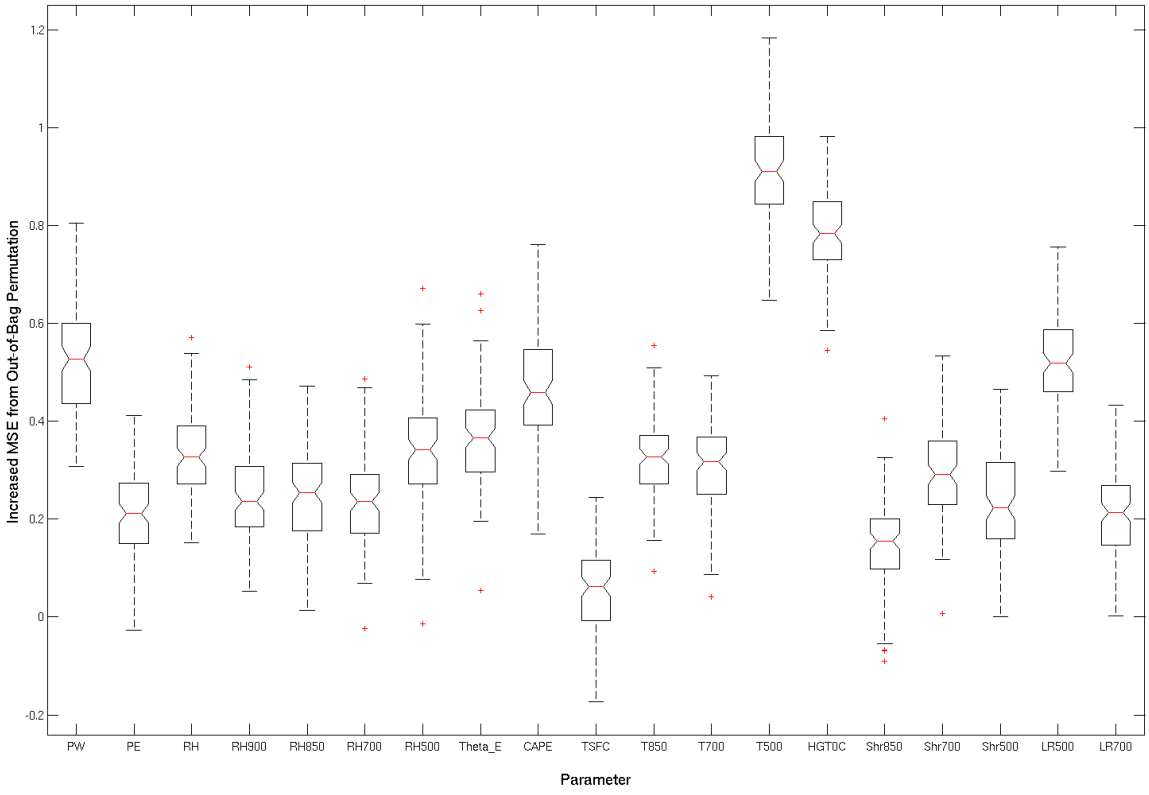


Figure 5.22. Increase in decision tree ensemble mean squared error (normalized by standard deviation) when individual parameters are permuted or randomized. Classification criteria was bias of the convective Z-R-derived rain rate trained on the storm relative inflow environment around nearby rain gauges.

The comparison between the 13-km and 20-km RUC analyses demonstrate that the training of the statistical model ensembles is sensitive to the resolution of the NWP model due to the impact of grid spacing on parameters that have large gradients in the vicinity of precipitation. Thus, the ensemble would need to be retrained whenever an upgrade to a higher resolution model is made.

Comparison of Probabilities to Dual-Polarized Radar Variables

Because the probabilities of warm rain enhancement of rainfall were not trained on actual DSD or microphysical quantities, it is important to compare them to observed DSD properties within storms to determine whether there is a physical connection between the statistical models' output and the impact of warm rain growth processes on rain rate. The challenge to such a comparison is the same reason the models were not trained in such a fashion: the available data for DSD parameters is sparse and generally not continuously available, and they're typically only available at the surface in the form of disdrometer measurements. This was especially true prior to 2011 when the NEXRAD network was still entirely operating in single-polarization mode. However, the ongoing dual-polarization upgrade allows for some additional analysis opportunities that were not previously available on a large scale.

Differential reflectivity quantifies the mean aspect ratio of all drops within the radar's resolution volume by dividing the reflectivity from the vertically polarized beam (Z_v) from the reflectivity of the horizontally polarized beam (Z_h):

$$Z_{dr} = 10log_{10}\frac{Z_h}{Z_v} \tag{5.2}$$

Pruppacher and Klett (1978) showed that raindrops become more oblate as they grow in size due to the effects of air resistance. Thus, a resolution volume with a high Z_{dr} implies that it contains relatively large drops because the power returned to the radar is greater from the horizontally-polarized beam (Wakimoto and Bringi 1988). For the purposes of this study, the directly proportional relationship between Z_{dr} and drop size is useful because any increase in mean drop size due to enhanced collision and coalescence processes should be evident in the vertical profile of Z_{dr}. Specifically, Bringi and Chandrasekar (2001) explained that in warm convective rain, the maximum in Z_{dr} (and potentially Z_h as well) occurs near the surface due to size sorting of the largest drops formed by collision and coalescence. The larger, heavier drops preferentially fall out of cloud because the updraft can no longer suspend them aloft, so resolution volumes observed near the surface contain the largest proportion of large drops.

Specific differential phase or K_{dp} is a polarimetric variable that quantifies the shift in radar beam phase between the horizontal and vertical channels and is an approximate indicator of the bulk liquid water content of the resolution volume that the beam is propagating through (Bringi and Chandrasekar 2001):

$$K_{dp} = \frac{\Phi_{dp}(r_2) - \Phi_{dp}(r_1)}{2(r_2 - r_1)} \approx \left(\frac{180}{\lambda}\right) 10^{-3} CW(1 - \bar{r}_m) \qquad (5.3)$$

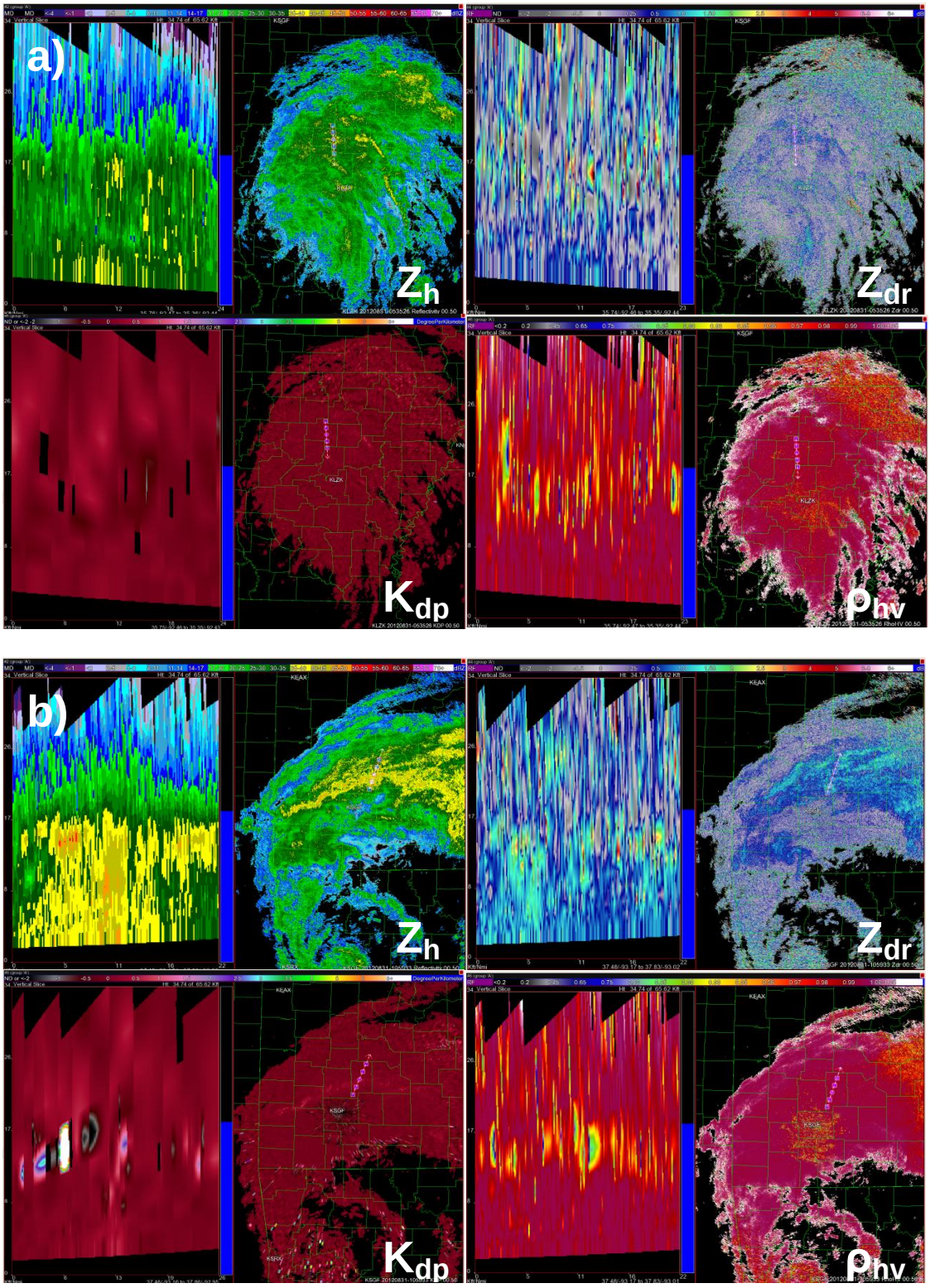
where Φ_{dp} is the cumulative differential phase along the beam path at two different distances from the radar (r_1 and r_2), λ is the radar's wavelength, C ~ 3.75, W is the rainwater content, and $\bar{r}_m$ is the mass-weighted mean drop axis ratio. Therefore, an increase in both K_{dp} and Z_{dr} with decreasing height below the freezing level would be a

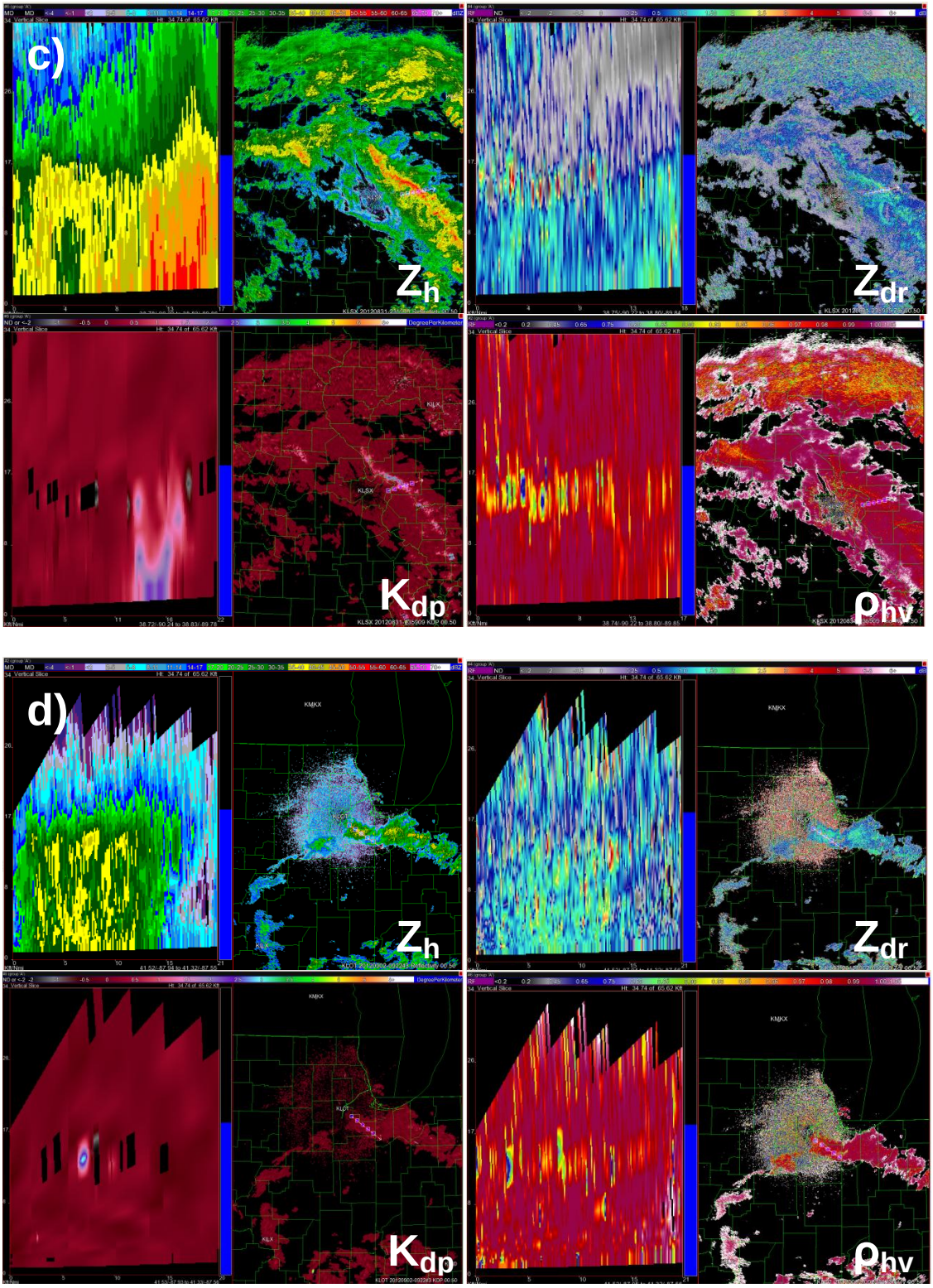
fairly strong indicator of productive warm rain growth within the storm.

Correlation coefficient, or ρ_{hv}, estimates the linearity of the relationship between the horizontally polarized and vertically polarized returned power and phase shifts over the collection of pulses comprising one resolution volume:

$$\rho_{hv} = \frac{\langle S_{hh} S_{vv}^{*} \rangle}{\sqrt{\langle |S_{hh}|^2 \rangle \langle |S_{vv}|^2 \rangle}} \tag{5.4}$$

The correlation between the two polarizations provides insight into the type and homogeneity of targets that are contained in the resolution volume. Thus, pure rain and dry snow tend to have a characteristically high correlation coefficient, whereas mixed precipitation and non-meteorological targets (e.g., insects and birds) have much lower values. For estimating rainfall, the primary benefits of ρ_{hv} are 1) estimation of the height of the melting layer, and 2) quality control of ground clutter and other non-precipitation signals.


a)
Z_h
Z_{dr}
K_{dp}
ρ_{hv}
b)
Z_h
Z_{dr}
K_{dp}
ρ_{hv}


c)
Z_h
Z_{dr}
K_{dp}
ρ_{hv}
d)
Z_h
Z_{dr}
K_{dp}
ρ_{hv}

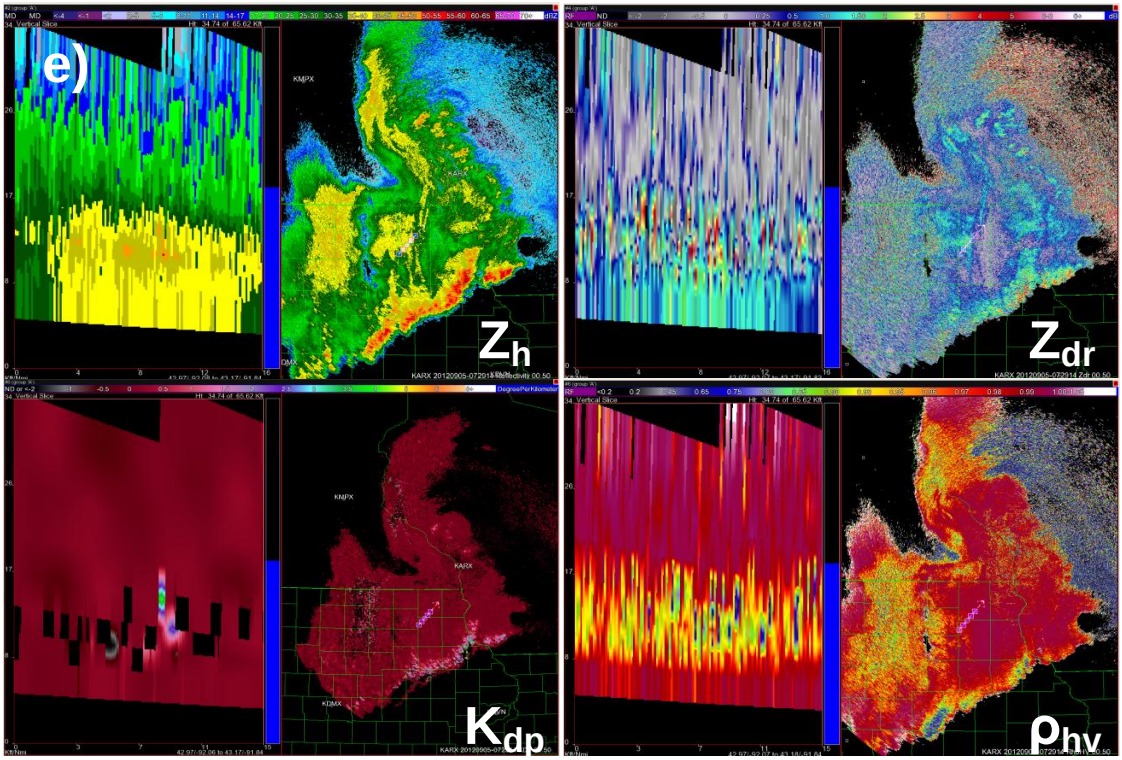


Figure 5.23. Vertical cross-sections of polarimetric variables for five different radars. Panels (a), (b), (c), and (d) were retrieved from Hurricane Isaac (2012) as it weakened and moved inland from a) KLZK (Little Rock, AR) to d) KLOT (Chicago, IL). Panel (e) was retrieved from an MCS event that occurred in Iowa on 5 Sep 2012.

The panels in Figure 5.23 illustrate some of the differences in observed polarimetric variables between a case transitioning from tropical to extratropical (panels a-d) and a purely continental MCS case (panel e). Unfortunately KLZK was the nearest radar to the landfall of Hurricane Isaac that had polarimetric capability at the time, and by the date and time of panel (a) Isaac had already been downgraded to post-tropical status. However, a transition to more continental characteristics is still apparent as the storm moved further inland from KLZK (panel a) to KSGF (panel b) to KLSX (panel c) to KLOT (panel d). The freezing level depicted in the ρ_{hv} field gradually lowered over time, and the magnitude of the Z_{dr} observed below the freezing level tended to increase, implying a transition from a DSD composed of many small, spherical drops to a DSD with larger drops. In contrast, the event shown in panel (e) was characterized by a much

lower freezing level and higher Z_{dr} values in the warm layer despite how shallow it is.

What's missing from the four Isaac cross-sections is an obvious vertical gradient in the Z_{dr} below the freezing level except for perhaps a slight indication in panels (a) and (c), even in areas where Z_h appears to be increasing toward the surface. However, the reason for this is likely due to the fact that the cross-sections are evaluating rainfall predominantly of the stratiform type, for which size sorting would generally not be observed due to the very weak lift's inability to suspend small drops aloft. The size sorting signature is more apparent in panel (c) where the cross-section intersects a region of convective rainfall, and a small increase in Z_{dr} with decreasing height is indeed observed there.

The more apparent trend from panels (a) to (e) is an overall increase in Z_{dr} below the freezing level. The gradual increase in overall Z_{dr} in time may be due to the height of the bright band within the stratiform rain. VPRs retrieved from KLZK and KMPX corresponding to the times of panels (a) and (e), respectively, depict the bright bands at very different heights (Fig. 5.24). The higher bright band at KLZK is associated with a much shallower ice region above it than was observed at KMPX. The shallower ice region may be limiting the growth of ice by deposition as well as the extent that aggregation can occur before the hydrometeors begin to melt as they fall, which would limit the drop sizes of rain in the warm layer of the profile. Thus, a lowering of the bright band over time as Isaac moved north and became increasingly non-tropical could explain the gradual increase in Z_{dr} below the freezing level in areas of stratiform rain. Another explanation for the higher Z_{dr} may be the higher variability of vertical motion within the trailing stratiform region of MCSs as well as the possible

presence of melting graupel. Small upward vertical velocity is possible, which may be enough to loft some small liquid drops above the freezing level, allowing for additional growth of frozen hydrometeors through accretion (Houze 1993).

Another notable property of the VPR retrieved from KLZK in Figure 5.24 is the increase in reflectivity with decreasing height within the warm layer. While drop growth through warm rain processes is often associated with convective rain due to the ability of the updraft to suspend drops within the cloud, it is apparent that substantial drop growth can also occur within stratiform rain if the warm layer is deep enough. The primary mechanism responsible for warm rain growth in stratiform rain is probably collection of small drops and cloud drops by larger drops as they fall faster due to their size (Houze 1993). Therefore, while the dominant growth mechanisms may be different for warm rain in convective and stratiform modes in a tropical environment, the thermodynamic structure of the environment in both is similar (i.e., higher freezing levels and weaker vertical velocities relative to convective environments, as well as potentially higher moisture content).

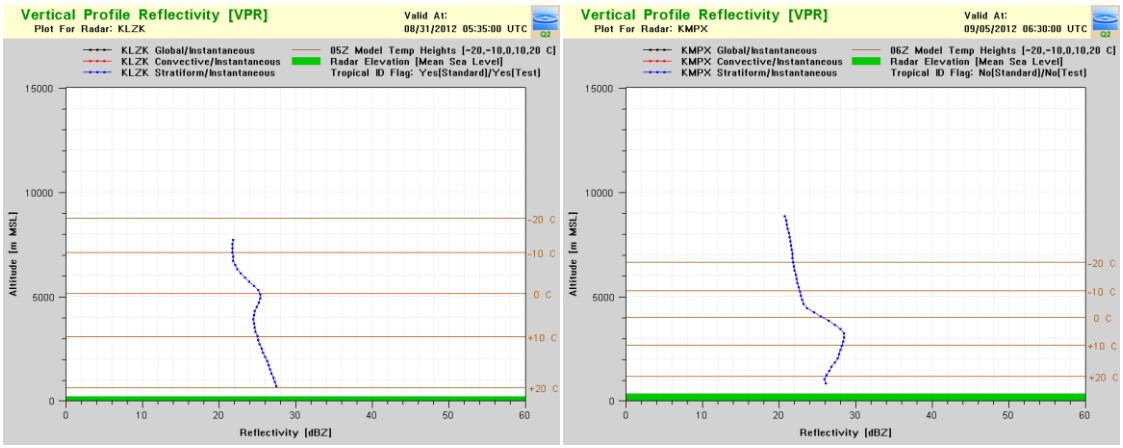


Figure 5.24. Vertical profiles of reflectivity corresponding to stratiform rain in the vicinity of the cross-sections in panels (a) and (e) of Figure 5.23, respectively.

In terms of the probabilities of enhanced rain rate for each case, the difference is striking (Fig. 5.25). White stars denote the location for where the cross-sections were retrieved for panel (c) (left) and panel (e) (right). Probabilities were of similar values as the left image for the other three cross-sections retrieved from Isaac. The probabilities computed for the continental MCS event are much lower than for the weakening post-tropical Isaac event.

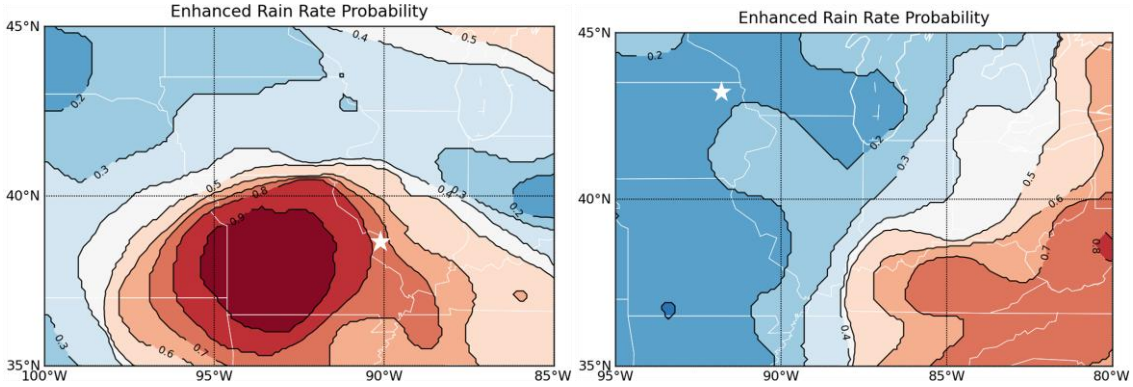


Figure 5.25. Probabilities of enhanced rain rate for 00Z 1 Sep 2012 (left) and 07Z 5 Sep 2012 (right). White stars mark the locations of vertical cross-sections in Figure 5.17's panels (c) and (e), respectively.

The microphysical reasoning for enhanced warm rain growth in tropical airmasses would seem to give credence to the results of the statistical modeling in this study, particularly based on the analysis of variable importance among the predictors. This is just a cursory evaluation, however, and deeper analysis will be needed once a longer coincident record between the new probabilities and dual-polarized radar data is established.

CHAPTER 6: SUMMARY

Estimation of rainfall in real-time from radar and rain gauges is a challenging problem with many uncertainties and potential sources of error. Whether using single-polarized or dual-polarization radar technology, one of the key steps in producing rain rates is to determine the rainfall type. For single-polarization, the rainfall type determines which reflectivity to rain rate (Z-R) function to apply, and different Z-Rs can produce vastly different rain rates. Thus, correctly associating a rainfall type with its most appropriate Z-R function is paramount to the accuracy of the resulting rainfall accumulations.

This study has examined the specific challenge of identifying where warm rain drop growth processes can enhance rainfall beyond what the default operational Z-Rs would provide. This enhancement of rainfall can occur in maritime and tropical airmasses where updraft velocity tends to be weaker than in more continental airmasses because the weak updraft extends the residence time of drops within the non-freezing levels of the cloud, allowing for additional drop growth and interaction with the saturated environment. It was postulated in this study that environments favorable for enhanced warm rain growth could be identified in near real-time using numerical weather prediction model analyses and aid in the delineation of an enhanced rainfall rate for improved rainfall accuracy. This automated enhancement of rain rate would be particularly beneficial in heavy rainfall situations that would be conducive for flooding, which are often underestimated by radar. More accurate rainfall estimates would improve not only flash flood warning and prediction, but also the accuracy of longer term river flooding forecasts.

The major contributions of this work are as follows:

- The newly developed probabilistic model for enhanced rainfall rates advances the state of the operational science by improving real-time QPE in environments conducive for flash flood-producing rainfall. Q2 is increasingly being looked to by forecasters both at the U.S. National Weather Service and abroad as a beneficial tool for the timely and accurate issuance of flash flood and flood warnings due to its dynamic and automated, multisensor-based Z-R selection, but improvements in the delineation of specific precipitation types have been very much needed. The environment-based probabilities reduce both the overestimation and underestimation bias caused by the incorrect identification of tropical rain using the VPR method alone. Because operational flash flood guidance is still solely based on basin accumulation of QPE rather than on hydrologic routing of runoff by numerical models, QPE accuracy is of critical importance to the NWS's mission to protect life and property.
- Scientifically, the probabilities provide an opportunity to advance the current "state of the art" in radar QPE, which are the polarimetric-based rainfall products. As radars in the U.S. continue to acquire dual-polarization capability, new challenges are presenting themselves in terms of the applicability of the previously derived QPE equations to untested meteorological environments and phenomena, including tropical rain. Furthermore, dual-polarization does not solve the long-standing problem of the NEXRAD network's inability in many places to adequately sample hydrometeors near the surface where they are most closely related to what is measured by a rain gauge. The probabilities can

provide much-needed guidance on the characteristics of the radar's environment and the resulting rainfall DSD properties near the surface such that rain rates can be dynamically adjusted in real-time.

- The new probability product will enhance the accuracy and quality of the 10-year reanalysis archive of gridded Q2 rainfall estimates that are currently being compiled for future scientific analysis and research. The lack of dual-polarization capability over those 10 years necessitates the use of Q2's current single-polarization-based algorithms in generating the archive, so it is critical that the challenge of reliably delineating tropical rain is addressed.

Six months of hourly data from rain gauges, the RUC model, and mosaicked radar reflectivity and rainfall fields from the NMQ Q2 system were retrieved for the 2010 warm season as a training set. The gauges and radars were assigned environment characteristics using a near-storm environment approach, and the analysis was restricted to gauges that met predefined criteria related to hourly rainfall accumulation, distance from the nearest radar, and filtering out potential areas of frozen precipitation. Three independent events - one tropical cyclone, one extratropical storm, and one continental MCS - were also selected to test how well the models performed on data outside of their training period.

Machine learning models were then trained using two possible classifications: one based on the structure of the vertical profile of reflectivity (VPR) below the freezing level, and one based on the bias of radar-derived rainfall when compared to hourly gauges. In general, areas undergoing enhancement through warm rain processes

below the freezing level are assumed to a) have increasing reflectivity with decreasing height toward the surface, and b) be systematically underestimated by the NWS default or convective Z-R function, which was derived from thunderstorms in continental-type airmasses. These assumptions formed the basis for the classifications, and the near-storm environment parameters derived from the RUC analyses were used as predictors.

To generate probabilities of warm rain enhancement, 100-member ensembles of the machine learning models (either support vector machines or automatically-generated decision trees) were trained using weighted bootstrap samples from the training period, and probabilities were derived as the ratio of the 100 members that predicted a favorable environment for the classification criterion tested. Contingency table-based performance metrics were then used to determine both what the optimal probability threshold was between the warm rain and non-warm rain classes, and which parameter settings in the model training were most appropriate for the data provided.

Overall, the decision tree-based ensembles tended to produce better skill scores and more diverse ensembles than the support vector machine-based ensembles and also produced probability fields that were spatially more physically consistent with the storm systems being analyzed. Furthermore, the decision trees proved to be far more computationally efficient, which is a major consideration for an algorithm intended for use in real-time.

When comparing the gauge bias classification to the VPR classification, the gauge bias ensembles were more skillful when verified against gauges than the VPRs (which were verified against VPRs). Not surprisingly, the gauge bias-based models aligned much better with where the radar-derived rain rates were underestimating,

which is what this algorithm ultimately seeks to improve. However, performance was more mixed when the two classification methods were evaluated in terms of their ultimate effect on QPE after precipitation types were identified.

Bootstrapped variable importance analysis of the individual predictor variables indicated that the radar-to-gauge bias during training was generally most sensitive to 850-500 hPa lapse rates of temperature, mean 1000-700 hPa relative humidity, and freezing level height. Two-sample permutation testing of the variables belonging to each class showed that areas of gauge underestimation corresponded to weaker (i.e., less negative) lapse rates, higher low to mid-level relative humidity, and higher freezing level, which would be consistent with a tropical airmass favorable for rainfall modification by warm rain processes. The variable importance analysis for the VPR-based models yielded similar results, indicating that the segregation between positive and negative VPR slope below the freezing level was most sensitive to freezing level height, 850-700 hPa lapse rate, and surface to 500 hPa wind shear. The similarity of the variable importance analyses between the two classification methods support the idea that the effect of enhanced warm rain growth on the DSD and ultimately the radar-based rain rate is a substantial contributor to QPE underestimation when compared to rain gauges, particularly in tropical environments.

Future Research Directions

The major disadvantage to the RUC-based delineation of enhanced rain rate probabilities is the temporal resolution. While VPRs are updated with every new volume scan from the radar, RUC and other short-term forecast model analyses are only

currently available hourly. While this is a considerable improvement over longer-range models that produce analyses on a 6 or 12-hour update cycle, it is still problematic for fast-moving storms or in environments that are rapidly evolving. In order for the probabilities to have maximum benefit on the time scale of 5-minute rain rates such as in Q2, efforts will need to be made to bridge the gap between the hours whether through interpolation to the 1-hour forecast or through object-based tracking of storms as they move in time. This may be in part why the probabilities appear to perform much better on tropical cyclones due to their slow movement relative to storms that form in environments of much stronger flow.

Further investigation is also needed regarding the effect of varying model initialization analysis methodologies and model grid resolution on both the probability skill and the selection of most significant predictors by the machine learning algorithms. As was shown in the previous section, simply aggregating the 13-km RUC analysis up to 20-km with no change in the data assimilation methods had an impact on several of the predictor variables related to moisture content and thermodynamics of the environment near precipitating systems. We have yet to explore how different models affect the probabilities, such as the latest operational update to the RUC (known as Rapid Refresh or RAP) or the 3-km analysis from the High-Resolution Rapid Refresh. At the present time the most pressing issue is to evaluate the differences between the RUC analysis and the RAP analysis, because the real-time NMQ system very recently switched to the RAP when it replaced the RUC as the short-term operational forecast model.

Development of a new precipitation typing scheme for Q2 is currently underway

that takes advantage of the new probability of enhanced rainfall. Rather than simply assigning the tropical Z-R function where probabilities are favorable, the algorithm uses fuzzy logic based on the probabilities and other inputs to determine where either the tropical Z-R or even higher rain rates are needed (e.g., rain rates that may be associated with equilibrium DSDs in very intense rain). The algorithm also separates convective from stratiform rain types for the tropical rain rates.

Once the upgrade to dual-polarization is complete across the national WSR-88D network, further studies can be conducted on the nature of the relationship between the enhanced rain rate probabilities and the new dual-polarization variables. Because differential reflectivity, specific differential phase, and correlation coefficient tend to vary with the microphysical composition of precipitation in a more direct way than horizontally polarized reflectivity, they allow for a more physically based analysis of environments conducive for enhanced warm rain processes than was previously possible. A potential avenue of study would be using the dual-polarization parameters as predictors in the training of the statistical classification models as a complement to the environmental parameters, but uncertainties in the radar observables would have to be addressed first (i.e., range-height dependence of variables, precision and calibration of Z_{dr} and K_{dp}, and handling of these values where radar coverage is poor or non-existent). Despite the added uncertainties, however, the improved temporal resolution of the radar variables along with their microphysical association with precipitation offer potentially significant benefits.

REFERENCES

Amitai, E., 2000: Systematic variations of observed radar reflectivity-rainfall rate relations in the tropics. *J. Appl. Meteor.*, **39**, 2198-2208.

Anagnastou, E. N., W. F. Krajewski, and J. Smith, 1999: Uncertainty quantification of mean-areal radar-rainfall estimates. *J. Atmos. Ocean. Tech.*, **16**, 206-215.

Atallah, E. H., and L. F. Bosart, 2003: The extratropical transition and precipitation distribution of Hurricane Floyd (1999). *Mon. Wea. Rev.*, **131**, 1063-1081.

Atlas, D., and C. R. Williams, 2003: The anatomy of a continental tropical convective storm. *J. Atmos. Sci.*, **60**, 3-15.

Atlas, D., and C. W. Ulbrich, 2000: An observationally based conceptual model of warm oceanic convective rain in the Tropics. *J. Appl. Meteor.*, **39**, 2165-2181.

Atlas, D., R. C. Srivastava, and R. S. Sekhon, 1973: Doppler radar characteristics of precipitation at vertical incidence. *Rev. Geophys. Space Phys.*, **11**, 1-35.

Austin, P. M., and S. G. Geotis, 1979: Raindrop sizes and related parameters for GATE. *J. Appl. Meteor.*, **18**, 569-578.

Avila, L. A., and J. Cangialosi, 2010: Hurricane Ida (AL112009). National Hurricane Center Tropical Cyclone Report. 19 pp.

Battan, L. J., 1973: *Radar Observation of the Atmosphere*. Univ. of Chicago Press, Chicago, Ill., 324 pp.

Benjamin, S. G., D. Devenyi, S. S. Weygandt, K. J. Brundage, J. M. Brown, and G. A. Grell, 2004a: An hourly assimilation - forecast cycle: The RUC. *Mon. Wea. Rev.*, **132**, 495-518.

Benjamin, S. G., G. A. Grell, J. M. Brown, T. G. Smirnova, and R. Bleck, 2004b: Mesoscale weather prediction with the RUC hybrid isentropic terrain-following coordinate model. *Mon. Wea. Rev.*, **132**, 473-494.

Benjamin, S. G., D. Kim, and J. M. Brown, 2002: Cloud/Hydrometeor initialization in the 20-km RUC cusing GOES and radar data. *10th Conf. on Aviation Range and Aerospace Meteor.*, Am. Meteor. Soc., Portland, OR.

Biggerstaff, M. I., and S. A. Listemaa, 2000: An improved scheme for convective/stratiform echo classification using radar reflectivity. *J. Appl. Meteor.*, **39**, 2129-2150.

Blanchard, D. C., and T. A. Spencer, 1970: Experiments on the generation of raindrop-

size distributions by drop breakup. *J. Atmos. Sci.*, **27**, 101-108.

Bowen, E. G., 1950: The formation of rain by coalescence. *Australian J. Sci. Res.*, **3**, 192-213.

Bradley, A. A., and J. A. Smith, 1994: The hydrometeorological environment of extreme rainstorms in the southern plains of the United States. *J. Appl. Meteor.*, **33**, 1418-1431.

Breiman, L., 2001: Random forests. *Machine Learning*, **45**, 5-32.

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

Bringi, V. N., and V. Chandrasekar, 2001: *Polarimetric Doppler Weather Radar: Principles and Applications*. Cambridge University Press. 636 pp.

Brooks, H. E., and D. J. Stensrud, 2000: Climatology of heavy rain events in the United States from hourly precipitation observations. *Mon. Wea. Rev.*, **128**, 1194-1201.

Brown, G., J. Wyatt, R. Harris, and X. Yao, 2005: Diversity creation methods: A survey and categorisation. *Information Fusion*, **6**, 5-20.

Carbone, R. E., and L. D. Nelson, 1978: The evolution of raindrop spectra in warm-based convective storms as observed and numerically modeled. *J. Atmos. Sci.*, **35**, 2302-2314.

Chandrasekar, V., V. N. Bringi, N. Balakrishnan, and D. S. Zrnic, 1990: Error structure of multiparameter radar and surface measurements of rainfall. Part III: Specific differential phase. *J. Atmos. Ocean. Tech.*, **7**, 621-629.

Chang, C.-C., and C.-J. Lin, 2001: LIBSVM: A Library for Support Vector Machines. *ACM Trans. on Intelligent Systems and Tech.*, **2**, 27:1-27.

Chiang, Y.-M., F.-J. Chang, B. J.-D. Jou, and P.-F. Lin, 2007: Dynamic ANN for precipitation estimation and forecasting from radar observations. *J. Hydrol.*, **334**, 250-261.

Ciach, G. J., 2003: Local random errors in tipping-bucket rain gauge measurements. *J. Atmos. Ocean. Tech.*, **20**, 752-759.

Ciach, G. J., and W. F. Krajewski, 1999a: Radar-rain gauge comparisons under observational uncertainties. *J. Appl. Meteor.*, **38**, 1519-1525.

Davis, R. S., 2004: The impact of tropical rainfall rates on flash flood detection. 22nd Conf. on Severe Local Storms, Hyannis, MA., Am. Meteor. Soc., 11B.5 [Available

online at http://ams.confex.com/ams/pdfpapers/81521.pdf]

Doswell, C. A., H. E. Brooks, and R. A. Maddox, 1996: Flash flood forecasting: An ingredients-based methodology. *Wea. Forecast.*, **11**, 560-581.

Doviak, R. J., and D. S. Zrnic, 1993: *Doppler Radar and Weather Observations*. Dover Publications, 562 pp.

Efron, B., and R. J. Tibshirani, 1993: *An Introduction to the Bootstrap*. Monographs on Statistics and Applied Probability, 57, Chapman and Hall/CRC, 436 pp.

Eslamian, S. S., S. A. Gohari, M. Biabanaki, and R. Malekian, 2008: Estimation of monthly pan evaporation using artificial neural networks and support vector machines. *J. Appl. Sci.*, **8**, 3497-3502.

Fiebrich, C. A., C. R. Morgan, A. G. McCombs, P. K. Hall, and R. A. McPherson, 2010: Quality assurance procedures for mesoscale meteorological data. *J. Atmos. Ocean. Tech.*, **27**, 1565-1582.

Fulton, R. A., J. P. Breidenbach, D.-J. Seo, D. A. Miller, and T. O'Bannon, 1998: The WSR-88D rainfall algorithm. *Wea. Forecast.*, **13**, 377-395.

Funk, T., 1991: Forecasting techniques utilized by the Forecast Branch of the National Meteorological Center during a major convective rainfall event. *Wea. Forecast.*, **6**, 548-564.

Giangrande, S. E., and A. V. Ryzhkov, 2008: Estimation of rainfall based on the results of polarimetric echo classification. *J. Appl. Meteor.*, **47**, 2445-2462.

Habib, E., G. J. Ciach, and W. F. Krajewski, 2004: A method for filtering out raingauge representativeness errors from the verification distributions of radar and raingauge rainfall. *Advances in Water Resources*, **27**, 967-980.

Hastie, T., R. Tibshirani, and J. Friedman, 2001: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics, Springer, 533 pp.

Hessami, M., and A. A. Viau, 2003: An adaptive neuro-fuzzy inference system for the post-calibration of weather radar rainfall estimation. *J. Hydroinformatics*, **1**, 63-70.

Holton, J. R., 2004: *An Introduction to Dynamic Meteorology*. International Geophysics Series, 88, Elsevier Academic Press, 535 pp.

Houze, R. A., 1993: *Cloud Dynamics*. International Geophysics Series, 53, Academic Press, 573 pp.

Hu, Z., and R. C. Srivastava, 1995: Evolution of raindrop size distribution by

coalescence, breakup, and evaporation: Theory and observations. *J. Atmos. Sci.*, **52**, 1761-1783.

Johnson, G. A., and J. Moser, 1992: A decision tree for forecasting heavy rains from mid-latitude synoptic patterns in Louisiana generally from late fall through spring. NOAA Tech. Memo. NWS ER-87, 24 pp.

Jorgensen, D. P., E. J. Zipser, and M. A. LeMone, 1985: Vertical motions in intense hurricanes. *J. Atmos. Sci.*, **42**, 839-856.

Junker, N. W., R. S. Schneider, and S. L. Fauver, 1999: A study of heavy rainfall events during the great midwest flood of 1993. *Wea. Forecast.*, **14**, 701-712.

Kanevski, M., A. Pozdnoukhov, and V. Timonin, 2009: *Machine Learning for Spatial Environmental Data*. CRC Press, Boca Raton, FL. 377 pp.

Kim, D., E. Tollerud, S. Vasiloff, and J. Caldwell, 2009: Comparison of manual and automated quality control of operational hourly precipitation data at the National Weather Service. *23rd Conf. on Hydrol.*, Am. Meteor., Soc., Phoenix, AZ.

Kim, D., B. Nelson, and L. Cedrone, 2006: Reprocessing of historic Hydrometeorological Automated Data System (HADS) precipitation data. *10th Symp. on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface*, Am. Meteor. Soc., Atlanta, GA.

Kitchen, M., and R. M. Blackall, 1992: Representativeness errors in comparisons between radar and gauge measurements of rainfall. *J. Hydrol.*, **134**, 13-33.

Kondragunta, C. R., and K. Shrestha, 2006: Automated real-time operational rain gauge quality control tools in NWS hydrologic operations. *20th Conf. on Hydrol.*, Am. Meteor., Soc., Atlanta, GA.

Konrad, C. E., and L. Perry, 2009: Relationships between tropical cyclones and heavy rainfall in the Carolina region of the USA. *Int. J. Clim.*, **30**, 522-534.

Konrad, C. E., 1997: Synoptic-scale features associated with warm season heavy rainfall over the interior southeastern United States. *Wea. Forecast.*, **12**, 557-571.

Krajewski, W. F., G. J. Ciach, 2003: An analysis of small-scale rainfall variability in different climatic regimes. *Hydrological Sciences J.*, **48**, 151-162.

Lakshmanan, V., T. Smith, G. J. Stumpf, and K. Hondl, 2007a: The Warning Decision Support System - Integrated Information (WDSS-II). *Wea. Forecast.*, **22**, 596-612.

Lakshmanan, V., A. Fritz, T. Smith, K. Hondl, and G. J. Stumpf, 2007b: An automated technique to quality control radar reflectivity data. *J. Appl. Meteor. Clim.*, **46**, 288-305.

Lee, C. K., G. W. Lee, I. Zawadzki, and K.-E. Kim, 2009: A preliminary analysis of spatial variability of raindrop size distributions during stratiform rain events. *J. Appl. Meteor. Clim.*, **48**, 270-283.

List, R., 1988: A linear radar reflectivity-rainrate relationship for steady tropical rain. *J. Atmos. Sci.*, **45**, 3564-3572.

List, R., N. R. Donaldson, and R. E. Stewart, 1987: Temporal evolution of drop spectra to collisional equilibrium in steady and pulsating rain. *J. Atmos. Sci.*, **44**, 362-372.

Liu, H., V. Chandresekar, and G. Xu, 2001: An adaptive neural network scheme for radar rainfall estimation from WSR-88D observations. *J. Appl. Meteor.*, **40**, 2038-2050.

Maddox, R. A., C. F. Chappell, and L. R. Hoxit, 1979: Synoptic and meso-α aspects of flash flood events. Bull. Am. Meteor. Soc., 60, 115-123.

Maddox, R. A., L. R. Hoxit, C. F. Chappell, and F. Caracena, 1978: Comparison of meteorological aspects of the Big Thompson and Rapid City flash floods. *Mon. Wea. Rev.*, **106**, 375-389.

Marshall, J. S., and W. M. Palmer, 1948: The distribution of raindrops with size. J. Meteor., 5, 165-166.

Marshall, J. S., R. C. Langille, and W. M. Palmer, 1947: Measurement of rainfall by radar. J. Meteor., 4, 186-192.

Mason, I. B., 2003: Binary events. *Forecast Verification - A Practitioner's Guide in Atmospheric Science*. I. T. Jolliffe and D. B. Stephenson, Eds., Wiley, 37-76.

Mercer, A. E., C. M. Shafer, C. A. Doswell, L. M. Leslie, and M. B. Richman, 2009: Objective classification of tornadic and nontornadic severe weather outbreaks. *Mon. Wea. Rev.*, **137**, 4355-4368.

Morin, E., W. F. Krajewski, D. C. Goodrich, X. Gao, and S. Sorooshian, 2003: Estimating rainfall intensities from weather radar data: The scale-dependency problem. *J. Hydrometeor.*, **4**, 782-797.

National Research Council, 2011: *The National Weather Service Modernization and Associated Restructuring*. National Academy Press, 150 pp.

National Research Council, 2005: *Flash Flood Forecasting over Complex Terrain: With an Assessment of the Sulphur Mountain NEXRAD in Southern California*. National Academy Press, 148 pp.

Noel, J., and J. Dobur, 2002: A pilot study examining model-derived precipitation

efficiency for use in precipitation forecasting in the eastern United States. *National. Wea. Digest*, 3-8.

Orlandini, S., I. Morlini, 2000: Artificial neural network estimation of rainfall intensity from radar observations. *J. Geophys. Res.*, **105**, 24849-24861.

Prat, O. P., and A. P. Barros, 2009: Exploring the transient behavior of Z-R relationships: Implications for radar rainfall estimation. *J. Appl. Meteor. Clim.*, **48**, 2127-2143.

Pruppacher, H. R., and J. D. Klett, 1978: *Microphysics of clouds and precipitation*. D. Reidel, 714 pp.

Robertson, D., 1974: Monte Carlo simulations of drop growth by accretion. *J. Atmos. Sci.*, **31**, 1344-1350.

Rogash, J., 2003: Meteorological aspects of south-central and southwestern New Mexico and far western Texas flash floods. *National Wea. Digest*, **27**, 1-27.

Rogers, R. R., and M. K. Yau, 1996: *A Short Course in Cloud Physics*. International Series in Natural Philosophy, 113, Butterworth-Heinmann, 290 pp.

Rokach, L., and O. Maimon, 2008: *Data Mining with Decision Trees: Theory and Applications*. Series in Machine Perception and Artificial Intelligence, 69, World Scientific, 244 pp.

Root, B., T.-Y. Yu, M. Yeary, and M. B. Richman, 2010: The added value of surface data to radar-derived rainfall-rate estimation using an artificial neural network. *J. Atmos. Ocean. Tech.*, **27**, 1547-1554.

Rosenfeld, D., and C. W. Ulbrich, 2003: Cloud microphysical properties, processes, and rainfall estimation opportunities. *Radar and Atmospheric Science: A Collection of Essays in Honor of David Atlas*. R. M. Wakimoto and R. Srivastava, Eds., Am. Meteor. Soc., 237-258.

Rosenfeld, D., E. Amitai, and D. B. Wolff, 1995: Classification of rain regimes by the three-dimensional properties of reflectivity fields. *J. Appl. Meteor.*, **34**, 198-211.

Rosenfeld, D., D. B. Wolff, and E. Amitai, 1994: The window probability matching method for rainfall measurements with radar. *J. Appl. Meteor.*, **33**, 682-693.

Rosenfeld, D., D. B. Wolff, and D. Atlas, 1993: General probability-matched relations between radar reflectivity and rain rate. *J. Appl. Meteor.*, **32**, 50-72.

Ryzhkov, A. V., S. E. Giangrande, T. Schuur, 2005: Rainfall estimation with a polarimetric prototype of WSR-88D. *J. Appl. Meteor.*, **44**, 502-515.

Schumacher, R. S., and R. H. Johnson, 2006: Characteristics of U.S. extreme rain events during 1999-2003. *Wea. Forecast.*, **21**, 69-85.

Scholkopf, B., and A. J. Smola, 2002: *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press, Cambridge, MA. 626 pp.

Seo, B.-C., and W. F. Krajewski, 2010: Scale dependence of radar rainfall uncertainty: Initial evaluation of NEXRAD's new super-resolution data for hydrologic applications. *J. Hydrometeor.*, **11**, 1191-1198.

Shafer, C. M., A. E. Mercer, L. M. Leslie, M. B. Richman, and C. A. Doswell, 2010: Evaluation of WRF model simulations of tornadic and nontornadic outbreaks occurring in the spring and fall. *Mon. Wea. Rev.*, **138**, 4098-4119.

Smith, J. A., E. Hui, M. Steiner, M. L. Baeck, W. F. Krajewski, and A. A. Ntelekos, 2009: Variability of rainfall rate and raindrop size distributions in heavy rain. *Water Resources Res.*, **45**, 1-12.

Smith, T. L., S. G. Benjamin, S. I Gutman, and S. Sahm, 2007: Short-range forecast impact from assimilation of GPS-IPW observations into the Rapid Update Cycle. *Mon. Wea. Rev.*, **135**, 2914-2930.

Szoke, E. J., E. J. Zipser, and D. P. Jorgensen, 1986: A radar study of convective cells in mesoscale systems in GATE. Part I: Vertical profile statistics and comparison with hurricanes. *J. Atmos. Sci.*, **43**, 182-197.

Telford, J. W., 1955: A new aspect of coalescence theory. *J. Meteor.*, **12**, 436-444.

Tokay, A., P. G. Bashor, and V. L. McDowell, 2010: Comparison of rain gauge measurements in the Mid-Atlantic Region. *J. Hydrometeor.*, **11**, 553-565.

Tokay, A., D. A. Short, C. R. Williams, W. L. Ecklund, and K. S. Gage, 1999: Tropical rainfall associated with convective and stratiform clouds: Intercomparison of disdrometer and profiler measurements. *J. Appl. Meteor.*, **38**, 302-320.

Tokay, A., and D. A. Short, 1996: Evidence from tropical raindrop spectra of the origin of rain from stratiform versus convective clouds. *J. Appl. Meteor.*, **35**, 355-371.

Tollerud, E., R. S. Collander, Y. Lin, and A. Loughe, 2005: On the performance, impact, and liabilities of automated precipitation gauge screening algorithms. *21st Conf. on Weather Analysis and Forecasting*, Am. Meteor. Soc., Washington, D.C.

Trafalis, T. B., B. Santosa, and M. B. Richman, 2005: Learning networks in rainfall estimation. *Computational Management Sci.*, **2**, 229-251.

Uijlenhoet, R., J. A. Smith, and M. Steiner, 2003a: The microphysical structure of extreme precipitation as inferred from ground-based raindrop spectra. *J. Atmos. Sci.*, **60**, 1220-1238.

Uijlenhoet, R., M. Steiner, and J. A. Smith, 2003b: Variability of raindrop size distributions in a squall line and implications for radar rainfall estimation. *J. Hydromet.*, **4**, 43-61.

Uijlenhoet, R., 2001: Raindrop size distributions and radar reflectivity - rain-rate relationships for radar hydrology. *Hydrol. and Earth System Sci.*, **5**, 615-627.

Ulbrich, C. W., and D. Atlas, 2007: Microphysics of raindrop size spectra: Tropical continental and maritime storms. *J. Appl. Meteor. Clim.*, **46**, 1777-1791.

Ulbrich, C. W., and D. Atlas, 1998: Rainfall microphysics and radar properties: Analysis methods for drop size spectra. *J. Appl. Meteor.*, **37**, 912-923.

Ulbrich, C. W., 1983: Natural variations in the analytical form of the raindrop size distribution. *J. Clim. Appl. Meteor.*, **22**, 1764-1775.

Vasiloff, S. V., K. W. Howard, and J. Zhang, 2009: Difficulties with correcting radar rainfall estimates based on rain gauge data: A case study of severe weather in Montana on 16-17 June 2007. *Wea. Forecast.*, **24**, 1334-1344.

Vasiloff, S. V., D.-J. Seo, K. W. Howard, J. Zhang, D. H. Kitzmiller, M. G. Mullusky, W. F. Krajewski, E. A. Brandes, R. M. Rabin, D. S. Berkowitz, H. E. Brooks, J. A. McGinley, R. J. Kuligowski, and B. G. Brown, 2007: Improving QPE and very short term QPF: An initiative for a community-wide approach. *Bull. Am. Meteor. Soc.*, **88**, 1899-1911.

Villarini, G., and W. F. Krajewski, 2010: Sensitivity studies of the models of radar-rainfall uncertainties. *J. Appl. Meteor. Clim.*, **49**, 288-309.

Villarini, G., P. V. Mandapaka, W. F. Krajewski, and R. J. Moore, 2008: Rainfall and sampling uncertainties: A rain gauge perspective. *J. Geophys. Res.*, **113**, 1-12.

Villarini, G., and W. F. Krajewski, 2008: Empirically-based modeling of spatial sampling uncertainties associated with rainfall measurements by rain gauges. *Advances in Water Resources*, **31**, 1015-1023.

Wakimoto, R. M., and V. N. Bringi, 1988: Dual-polarization observations of microbursts associated with intense convection: The 20 July storm during the MIST Project. *Mon. Wea. Rev.*, **116**, 1521-1539.

Weygandt, S., S. G. Benjamin, D. Devenyi, J. M. Brown, and P. Minnis, 2006: Cloud and hydrometeor analysis using METAR, radar, and satellite data within the RUC/Rapid

Refresh model. *12th Conf. on Aviation Range and Aerospace Meteor.*, Am. Meteor. Soc., Atlanta, GA.

Whitaker, C. J., and L. I. Kuncheva, 2000: Examining the relationship between majority vote accuracy and diversity in bagging and boosting. Tech. Report, Bangor University School of Computer Sci. 19 pp.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. International Geophysics Series, 59, Academic Press, 467 pp.

Willis, P. T., and P. Tattelman, 1989: Drop-size distributions associated with intense rainfall. *J. Appl. Meteor.*, **28**, 3-15.

Wilson, J. W., and E. A. Brandes, 1979: Radar measurement of rainfall: A summary. *Bull. Am. Meteor. Soc.*, **60**, 1048-1058.

Wood, L. T., 1997: Using the tropical Z-R relationship to improve precipitation estimates during a heavy rain event in southeast Texas. 28th Conf. on Radar Meteor., Austin, TX, Am. Meteor. Soc.

Xiao, R., and V. Chandresekar, 1997: Development of a neural network based algorithm for rainfall estimation from radar observations. *IEEE Trans. on Geoscience and Remote Sensing*, **35**, 160-171.

Xu, G., and V. Chandresekar, 2005: Operational feasibility of neural-network-based radar rainfall estimation. *IEEE Geosci. and Remote Sensing Letters*, **2**, 13-17.

Xu, X., K. W. Howard, and J. Zhang, 2008: An automated radar technique for the identification of tropical precipitation. *J. Hydromet.*, **9**, 885-902.

Zawadzki, I., 2006: Sense and nonsense in radar QPE. *Proc. European Radar Conf.*, Barcelona, Spain.

Zhang, J., K. Howard, C. Langston, S. Vasiloff, B. Kaney, A. Arthur, S. Van Cooten, K. Kelleher, D. Kitzmiller, F. Ding, D.-J. Seo, E. Wells, and C. Dempsey, 2011: National Mosaic and Multi-Sensor QPE (NMQ) system: Description, results, and future plans. *Bull. Am. Meteor. Soc.*, **92**, 1321-1338.

Zhang, J., C. Langston, and K. W. Howard, 2008: Brightband identification based on vertical profiles of reflectivity from the WSR-88D. *J. Atmos. Ocean. Tech.*, **25**, 1859-1872.

Zhang, J., K. W. Howard, and J. J. Gourley, 2005: Constructing three-dimensional multiple radar reflectivity mosaics: Examples of convective storms and stratiform rain echoes. *J. Atmos. Ocean. Tech.*, **22**, 30-42.

Zhang, Y., T. Adams, and J. V. Bonta, 2007: Subpixel-scale rainfall variability and the effects on separation of radar and gauge rainfall errors. J. Hydrometeor., 8, 1348-1363.

Zipser, E. J., and K. R. Lutz, 1994: The vertical profile of radar reflectivity of convective cells: A strong indicator of storm intensity and lightning probability? *Mon. Wea. Rev.*, **122**, 1751-1759.