VALIDATION AND DEVELOPMENT OF EXISTING AND NEW RAOB-BASED WARM-SEASON CONVECTIVE WIND FORECASTING TOOLS FOR CAPE CANAVERAL AIR FORCE STATION AND KENNEDY SPACE CENTER

by

Mitchell Hollis McCue

B.S., Plymouth State University, 2007

THESIS

Submitted to Plymouth State University

in Partial Fulfillment of

the Requirements for the Degree of

Master of Science

in

Applied Meteorology

April, 2010

This thesis has been examined and approved.

Thesis Director, James P. Koermer
Professor of Meteorology
Department of Atmospheric Sciences and Chemistry

Thomas R. Boucher
Associate Professor of Statistics
Department of Mathematics

William P. Roeder
Meteorologist, Systems Staff Meteorology
45th Weather Squadron, Patrick AFB, Florida

Date

ACKNOWLEDGEMENTS

This research would not have been possible if it were not for the valuable contributions of many people. To name all of them here is a difficult task. First off, I would like to extend a sincere thanks to the NASA Space Grant Program under University of New Hampshire Subcontract #PZ05008 for funding my research at Cape Canaveral Air Force Station in the summer of 2009. I would also like to thank NOAA for providing additional financial support while undertaking my graduate studies at Plymouth State University. In addition, I offer a heartfelt thanks to Dr. James Koermer of the Judd Gregg Meteorology Institute at Plymouth State University for selecting me to be a part of a terrific research opportunity, for serving on my thesis committee, and for providing invaluable guidance during the course of completing this thesis. I would also like to thank Dr. Thomas Boucher and Mr. William Roeder for serving on my thesis committee, as well as for providing considerable support and guidance while undertaking this research project. Furthermore, I would like to thank the NCDC, the 14th Weather Squadron in Asheville, North Carolina, the Applied Meteorology Unit (AMU), and Computer Sciences Raytheon (CSR) for supplying valuable archived data for my work. I would additionally like to thank all personnel from the AMU and the 45th Weather Squadron for contributing invaluable advice while working at CCAFS during the summer of 2009.

Finally, I am indebted to my parents for giving me the inspiration to pursue my education through graduate school and to become the meteorologist I have become. I have come a long way since the days of turning on the floodlight in the middle of the night to check the snow and stepping outside during a thunderstorm. It has been an exciting journey through good times and bad, and I would never have gotten this far if it

were not for your invaluable and everlasting support, kindness, and generosity. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

VALIDATION AND DEVELOPMENT OF EXISTING AND NEW RAOB-BASED WARM-SEASON CONVECTIVE WIND FORECASTING TOOLS FOR CAPE CANAVERAL AIR FORCE STATION AND KENNEDY SPACE CENTER

by

Mitchell H. McCue

Plymouth State University, April, 2010

Using a 15-year (1995 to 2009) climatology of 1500 UTC warm-season (May through September) rawinsonde observation (RAOB) data from the Cape Canaveral Air Force Station (CCAFS) Skid Strip (KXMR) and 5 minute wind data from 36 wind towers on CCAFS and Kennedy Space Center (KSC), several convective wind forecasting techniques currently employed by the 45th Weather Squadron (45 WS) were evaluated. Present forecasting methods under evaluation include examining the vertical equivalent potential temperature (θ_e) profile, vertical profiles of wind spend and direction, and several wet downburst forecasting indices. Although previous research found that currently used wet downburst forecasting methods showed little promise for forecasting convective winds, it was carried out with a very small sample, limiting the reliability of the results. Evaluation versus a larger 15-year dataset was performed to truly assess the forecasting utility of these methods in the central Florida warm-season convective environment. In addition, several new predictive analytic based forecast methods for

predicting the occurrence of warm-season convection and its associated wind gusts were developed and validated. This research was performed in order to help the 45 WS better forecast not only which days are more likely to produce convective wind gusts, but also to better predict which days are more likely to yield warning criteria wind events of 35 knots or greater, should convection be forecasted. Convective wind forecasting is a very challenging problem that requires new statistically based modeling techniques since conventional meteorologically based methods do not perform well.

New predictive analytic based forecasting methods were constructed using R statistical software and incorporate several techniques including multiple linear regression, logistic regression, multinomial logistic regression, classification and regression trees (CART), and ensemble CART using bootstrapping. All of these techniques except the ensemble CART methods were built with data from the 1995 to 2007 warm-seasons and validated with a separate independent dataset from the 2008 and 2009 warm-seasons. Ensemble CART models were built using randomly selected data from the 1995 to 2009 RAOB dataset and validated with data not used in constructing the models. Three different ensemble CART algorithms including the random forests, bagging, and boosting algorithms were tested to find the best performing model.

Quantitative verification results suggest that the presently used convection and wet downburst forecasting techniques do not show much operational promise. As such, it is not recommended that the 45 WS use vertical profiles of θ_e, wind speed, or wind direction to make specific predictions for which days are likely to produce convection or warning threshold wind gusts. None of the wet downburst indices used displayed much potential either. Although, the linear regression based predictive analytic models do not

perform too well, CART based models perform better, especially those that utilize a binary response variable. Of the new techniques, the ensemble CART models displayed the most promise with the boosting algorithm showing nearly perfect results for predicting which days would produce convection and which days would produce warning threshold winds should convection be predicted.

CHAPTER 1

1. Introduction and background

Strong winds produced by warm-season convection in and around the Kennedy Space Center (KSC), Cape Canaveral Air Force Station (CCAFS), and Patrick Air Force Base (AFB) pose a significant operational hazard to many weather sensitive aviation, pre-launch, launch, and post-launch activities. Consequently, the forecasters at the 45th Weather Squadron (45 WS) located at CCAFS and Patrick AFB have the responsibility of disseminating accurate wind warnings for the entire KSC/CCAFS complex and Patrick AFB in order to minimize adverse impacts to costly equipment and assure human safety (Harms *et al.* 1999). Warnings are based on the intensity of the winds with warning thresholds at 35 knots and 50 knots for winds up to 300 feet above ground level (AGL). Desired lead times are 30 minutes for the 35 knot wind threshold and 60 minutes for the 50 knot threshold. Until the summer of 2009, the 45 WS also issued a convective wind warning for 60 knot winds with a desired lead time of 60 minutes.

Convective wind warnings, which are issued when winds are likely to exceed the warning thresholds, are the second most common type of weather advisory issued by the 45 WS behind lightning (Wheeler and Roeder 1996) with more than 175 convective warnings per year (Roeder 2009). As such, it is essential that forecasters have as thorough an understanding as possible about the various atmospheric conditions that may lead to the occurrence of strong convective winds in order to provide as much warning as possible to the appropriate parties. As the downburst funnel forecasting conceptual model developed by the 45 WS shows in Fig. 1, predicting convective winds begins with a general knowledge of the central Florida convective wind climatology.

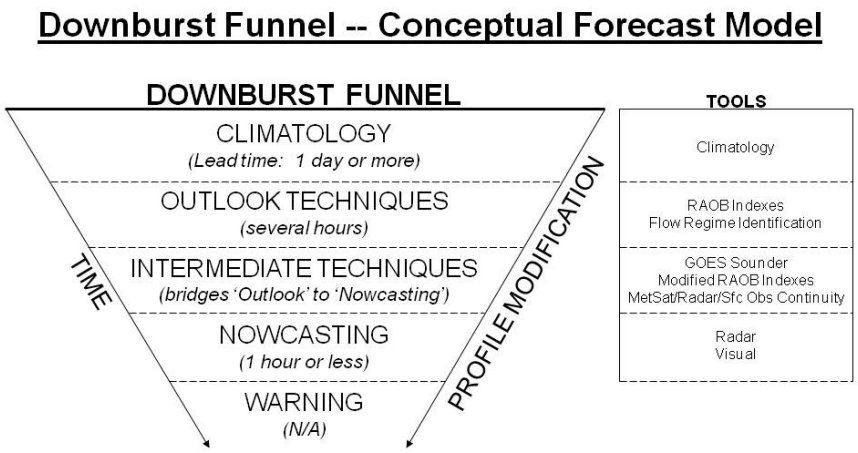

Figure 1. Downburst funnel forecasting conceptual model (Roeder 2009).

Unlike the winter months in Florida and in the higher latitudes at all times of year, the local weather throughout the warm-season (May through September) is dominated by numerous weak interacting low-level boundaries as opposed to stronger synoptic scale features. Examples of local low-level boundaries responsible for initiating convection in the area include, but are not limited to: east and west coast sea-breeze fronts from the Atlantic Ocean and Gulf of Mexico, respectively, Indian River and Banana River breeze fronts, lake breeze fronts, thunderstorm outflow boundaries, and the interaction between these and other boundaries.

Cummings *et al.* (2007) studied the number of convective wind events from May through September in and around CCAFS/KSC and found that the greatest quantity of convective events occurred during the afternoon and evening hours from approximately 1600 UTC to 0000 UTC (Fig. 2) and in the month of August, which follows the general pattern of thunderstorm frequency. A convective wind event is defined as an event that produces a convectively generated wind gust of at least 1 knot recorded by one or more

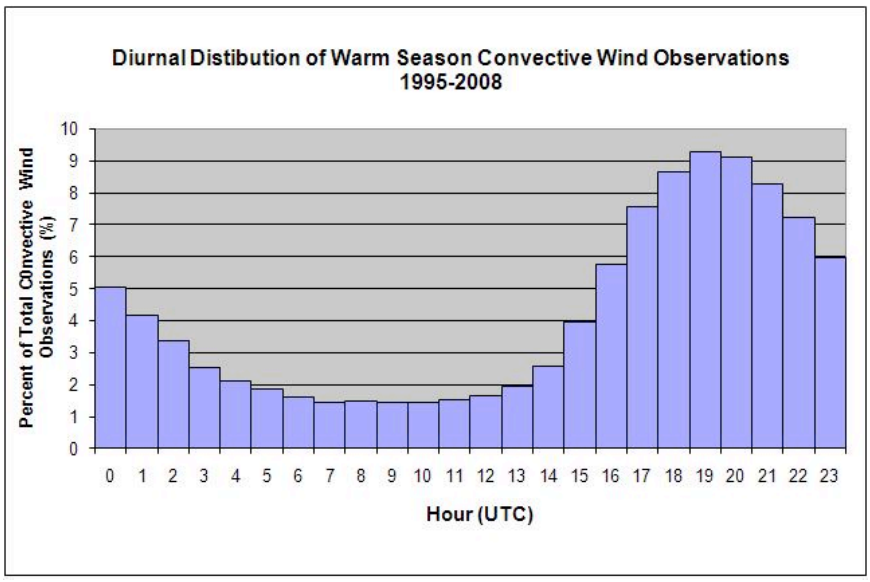

Figure 2. Diurnal distribution of warm-season convective wind observations from 1995 to 2008 for May-September for KSC/CCAFS (Cummings *et al.* 2007; Koermer 2009).

of 36 weather towers around the KSC/CCAFS complex. Cummings *et al.* (2007) also examined the distribution of the maximum winds with respect to the overall synoptic flow regime and found that, in general, the average peak convective wind speed was greater when the flow regime had a significant westerly component and weaker when it had more of an easterly component. This finding is consistent with the pattern that more and stronger thunderstorms form under westerly flow regimes. Lastly, this study discovered that the highest frequency of convective winds fell in the 20 to 24 knot interval and declined steadily above the 35 knot warning threshold. Fig. 3 displays a frequency distribution of the maximum observed peak wind gusts in knot increments and the associated Gumbel probability curve fit to the observed data for the 924 convective wind events for the warm-season months in the 1995 to 2008 study period. The median and mode peak wind speeds of the best-fit Gumbel curve are 32 knots and 36 knots, respectively.

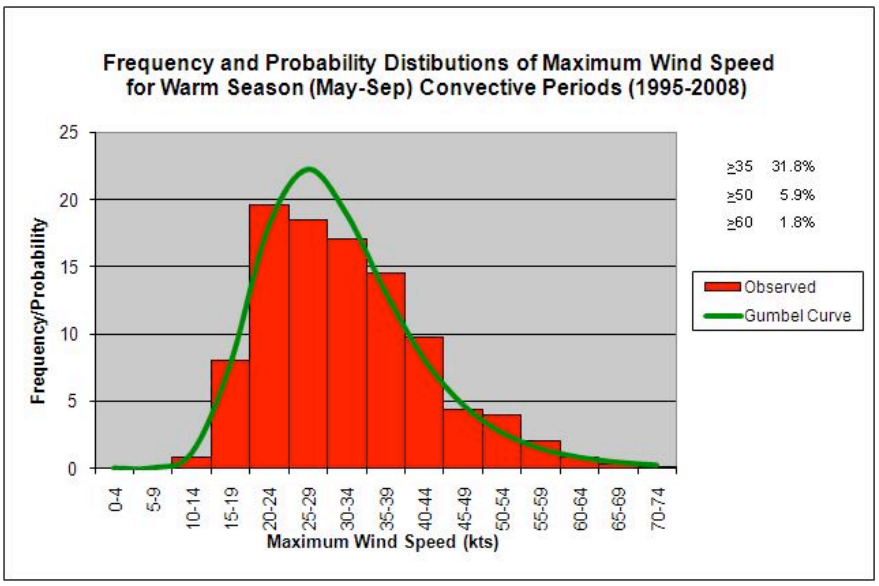

Figure 3. Frequency and probability distributions of maximum observed wind speeds from 1995-2008 (Cummings *et al.* 2007; Koermer 2009).

After the first step of "climatology" in the downburst funnel (Fig. 1), the next step in the convective wind forecasting process is the "outlook" for the next few hours. These outlook techniques are the tools used in the morning to forecast the likelihood and intensity of downbursts that afternoon. These techniques include skew-T analysis, flow regime identification, and identifying preexisting boundaries. More specifically, this thesis will investigate the 1500 UTC CCAFS Skid Strip (KXMR) rawinsonde observation (RAOB) data to see whether various atmospheric parameters suggest that convection and strong winds will occur. Note that the 45 WS uses asynoptic release times of 1000, 1500, and 2300 UTC for the local KXMR RAOB during the summer. The late morning RAOB (1500 UTC) is used to monitor how the planetary boundary layer has changed since sunrise to better predict the likelihood and intensity of afternoon thunderstorms and their associated hazards, including downbursts. Intermediate

forecasting techniques bridge the gap between the morning outlook techniques and the nowcasting warning techniques just before the downbursts occur.

Finally, nowcasting primarily involves examining local radar data from the 45 WS WSR-74C at Patrick AFB and WSR-88D at Melbourne to evaluate when and where downbursts will occur, and how strong the convectively generated winds will be. Other research at Plymouth State University is examining how to better use the WSR-88D to predict downbursts at KSC/CCAFS (Rennie *et al.* 2010).

Strong localized downbursts of convectively induced winds, which are also commonly known as microbursts or macrobursts depending on the horizontal extent and duration of the outflow, are a subject of much research since they have been known to produce winds of up to F3 tornado intensity (Fujita and Wakimoto 1981), have been a factor in many aviation accidents (Wolfson *et al.* 1994), have caused considerable property damage, and have placed many lives at risk (Kuchera and Parker 2006). As a result, much research has been done in an attempt to better forecast their occurrence and provide more warning of when and where they may take place.

A microburst can be defined as a strong downburst of damaging outflow winds with a horizontal diameter of less than 4 km wide (Fujita 1981) and a temporal duration of between 2 and 5 minutes whereas a macroburst has a horizontal diameter greater than 4 km and a duration of 5 to 20 minutes (Wakimoto 1985). Based on this, Wakimoto (1985) goes on to extend the general microburst definition to include both wet and dry microbursts. According to his definition, a dry microburst contains little or no rain between the onset and end of the high winds because it normally occurs in a low relative humidity environment where much of the precipitation evaporates before reaching the

ground. He defines a wet microburst to be accompanied by heavy rain between the beginning and end of the high wind period since they are usually situated in humid environments where the thunderstorm precipitation shaft does not have much chance to evaporate prior to reaching the ground. Due to the presence of warm and humid airmasses over central Florida during the warm season, the wet microburst is far more likely to be observed in and around the CCAFS/KSC complex.

The physical origins of a microburst have been a subject of considerable research and debate for quite some time. Research, performed by Srivastava (1985) studying evaporatively driven downdrafts by building a simple numerical model, indicated that the presence of a temperature lapse rate close to that of the dry adiabatic lapse rate below the cloud base and a high rainwater mixing ratio near the cloud base were significant contributors to spawning intense downdrafts. Although high rainwater mixing ratios act to increase negative buoyancy primarily by enhancing the amount of water available for evaporation, they can also act to add more mass to the parcel via a process commonly known as precipitation loading, which can additionally aid downward acceleration (Proctor 1989). An alternative way to look at the impact of precipitation loading is while the thunderstorm is in its formative stages; the updraft will support the weight of the hydrometeors and keep them suspended aloft. However as the storm reaches the mature stage, the updraft will weaken and sometimes collapse completely, allowing the hydrometeors to fall and aerodynamic loading (air friction) to result in a downdraft that is about as strong as the updraft. Steep low-level lapse rates can also further augment the negative buoyancy of a parcel by reducing the temperature surplus of the parcel relative to the atmosphere (reduced positive buoyancy) as the parcel descends and warms

(Kuchera and Parker 2006). The results from Srivastava's model were confirmed in an observational study of dry microbursts near Denver, Colorado in the Joint Airport Weather Study (JAWS) project during the early 1980s, when it was determined that evaporative cooling and the resultant negative buoyancy of a parcel were the major forcing mechanisms behind the dry downbursts observed in and around Denver in the JAWS project (Srivastava 1985). Subsequent research done by Srivastava (1987), which included precipitation melting processes into the simple model, revealed that the low and mid-level relative humidity profile had a strong effect on the melting of precipitation particles. This is due to the fact that frozen precipitation particles melt completely in a shorter fall distance when relative humidity values are high due to the greater wet-bulb temperatures that normally accompany elevated values of relative humidity. Meanwhile, he also explains that liquefied precipitation particles cannot evaporate completely over a similar fall distance under comparable moisture and temperature conditions implying that, despite a higher latent heat of vaporization, the amount of latent heat absorbed by melting is significantly greater than by evaporation. Consequently, melting precipitation particles such as hail can be an even larger contributor to strong downbursts, especially in high relative humidity environs conducive to wet downbursts (Srivastava 1987).

Building on Srivastava's research, Proctor (1989) discovered through numerical modeling that dry downbursts occurred from melting snow and graupel inside of convective cells situated in low relative humidity surroundings. Since snow and graupel can rapidly sublimate into water vapor in dry environments, extreme cooling of the air, due to the latent heat of sublimation being larger than that of evaporation or melting, can result. His studies also revealed that wet downbursts occurred from a combination of

melting hail and precipitation loading associated with cells located in high relative humidity environments. From his studies, Proctor (1989) drew two main conclusions: first, wet downbursts originate near the freezing level since this is where cooling of the ambient air due to phase change occurs; second, wet downbursts are more likely to occur in environments with high freezing levels, steep lapse rates, high relative humidity values below the freezing level, and lower humidity values near the freezing level. An observational study performed by Atlas *et al.* (2004) that examined a wet microburst in the Amazonian rainforest using Doppler and polarimetric radar confirmed some of the modeling studies performed by Srivastava and Proctor and found that melting hail was the main driving force behind the microburst observed.

Even though prior research has shown some potential with regard to wet downburst forecasting, current methods do not work well for predicting downbursts at KSC/CCAFS. More specifically, the presently used RAOB based downburst forecasting techniques perform poorly since Pearson correlation coefficients (to be discussed in more detail in chapter 2) between many of the RAOB derived predictors and observed downburst wind speeds are all quite close to 0 (Fig. 4). The highest Pearson correlation coefficient was only 0.3, which was for the computed storm motion wind speed (SMspd) variable. The central scientific question that this thesis will attempt to answer is: how can

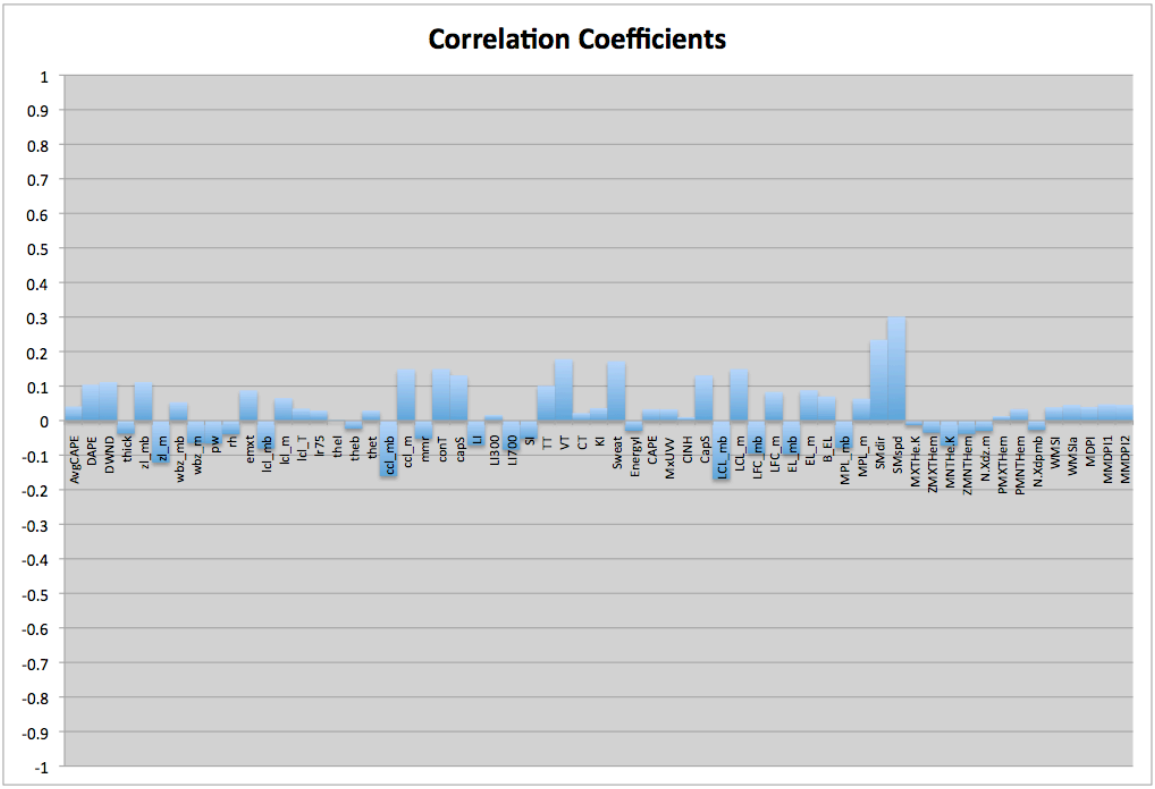

Figure 4. Pearson correlation coefficients between peak wind speed and each of the predictor variables used in this study. Please consult Table 1 in chapter 2 for acronym definitions.

wet downburst forecasts for KSC/CCAFS be improved given that many of the RAOB derived predictor variables do not correlate well with the peak wind speed response variable? It is hoped that by expanding the verification of previous RAOB indices and introducing some new outlook forecasting techniques through a rigorous statistical analysis of a 15-year climatology of RAOB data, RAOB derived thermodynamic variables, and RAOB based downburst forecasting indices that wet downburst prediction accuracy can be enhanced. More concisely, new predictive analytic procedures, which use historical data to come up with ways of predicting future events, are introduced and tested against an independent dataset. Such procedures examined include multiple linear regression (MLR), logistic regression (LR), multinomial logistic regression (MR),

classification and regression trees (CART), and ensembles of CART using several algorithms. The primary hope is that various new intermediate techniques will enable 45 WS forecasters to better tackle what is a complex forecasting environment and provide better accuracy and lead times on convective wind warnings.

CHAPTER 2

2. Data and methodology

a. Data

Data used in this study were obtained from several different sources. Raw decoded 1500 UTC KXMR RAOB text data from 1 May to 30 September for each year from 1995 to November 2008 were obtained from Computer Sciences Raytheon (CSR) in order to remove the World Meteorological Organization's (WMO) procedure of assigning a dew point depression of 30° C when the relative humidity value was 20% or less on transmitted soundings. Fig. 5 illustrates the difference between the corrected (black) and uncorrected soundings (blue). Note how significant differences in dew

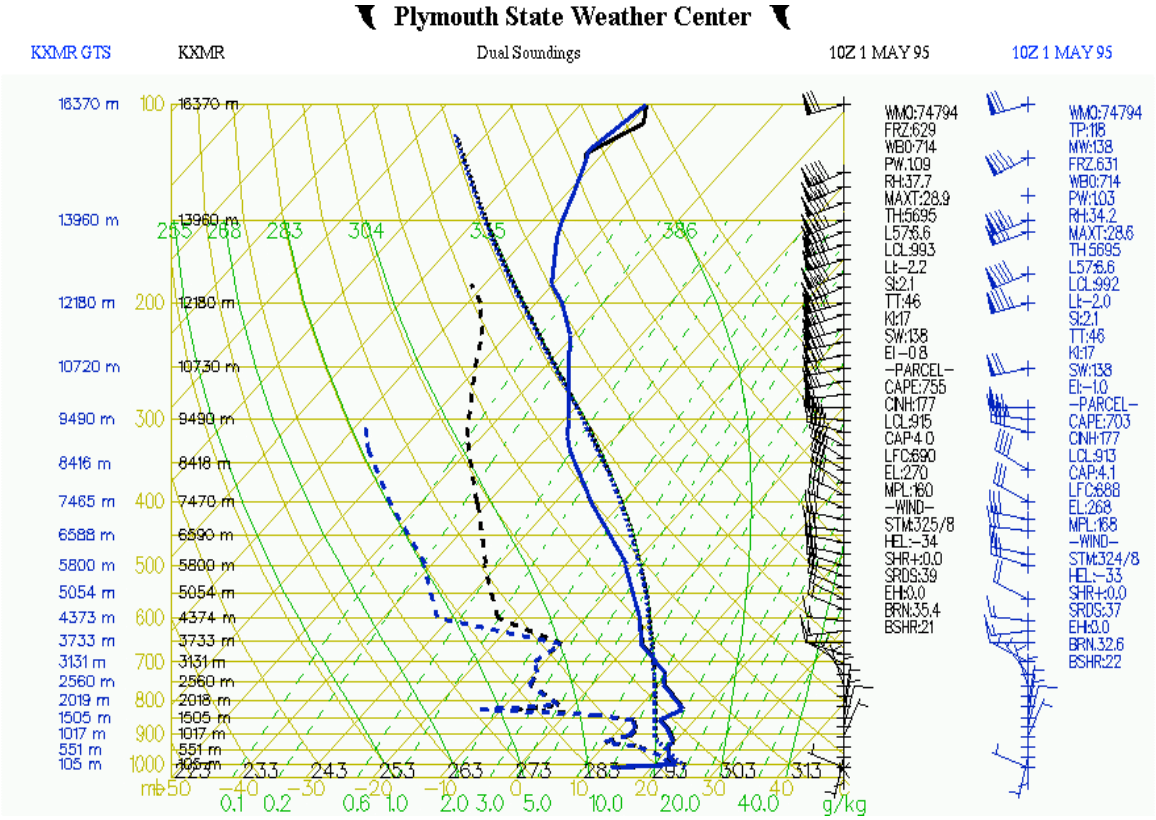

Figure 5. Dual 1000 UTC KXMR soundings for 1 May 1995 showing the difference between the WMO convention sounding (blue) and the corrected sounding (black).

point exist above the freezing level and in the 650-500 hPa layer, a layer whose equivalent potential temperature (θ_e) was found to be important in downburst formation in east central Florida (Wheeler and Roeder 1996). This correction ensures greater accuracy of computed thermodynamic variables such as θ_e and wet downburst forecasting tools that are based on atmospheric moisture content. After CSR eliminated this problem in November 2008, KXMR RAOB data came from the Global Telecommunication System (GTS). All of the KXMR RAOB data are also available online at the Plymouth State University convective wind climatology website, which can be found at http://vortex.plymouth.edu/conv_winds.

Wind speed data were obtained from the 45th Space Wing's network of 36 weather towers in and around the KSC/CCAFS spaceport. Fig. 6 displays a map of the weather tower locations and their relationship to the surrounding area. A total of 44 weather

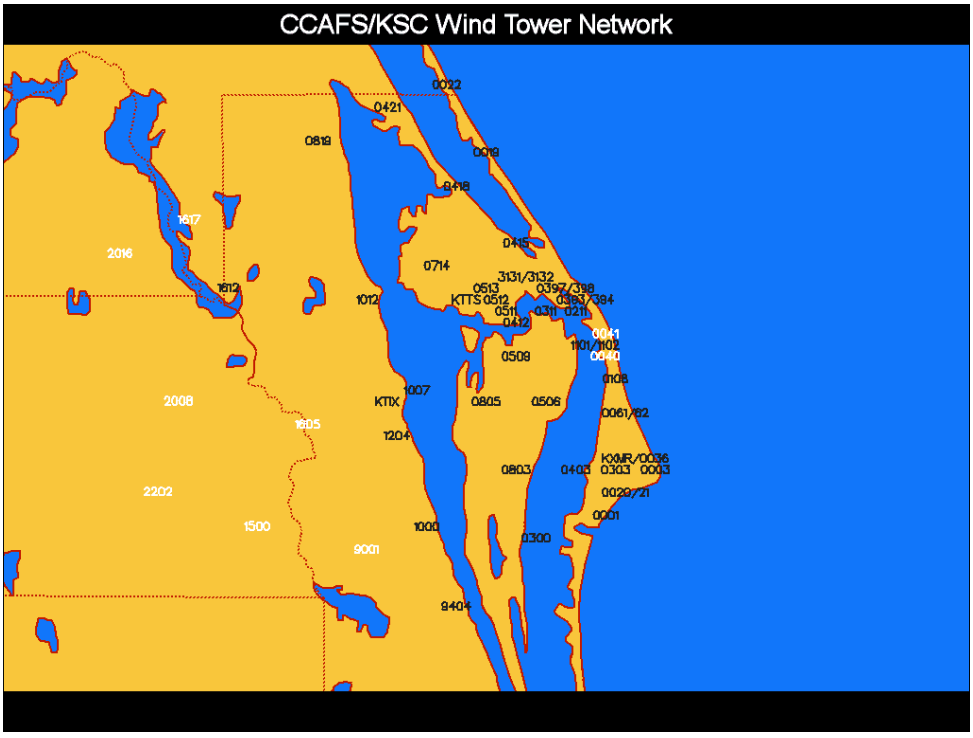

Figure 6. Map of locations of CCAFS/KSC wind towers and KTTS and KXMR. Data from black four-digit numeric tower identifiers were used in this study (Koermer 2009).

towers existed during this period, but one of the quality control requirements used in this research required 70% or more data availability. This reduced the number of towers available to 36. The weather towers measure 5 minute average peak wind speed from 12 feet to as high as 497 feet at one location at 10 different heights at various towers (Case and Bauman 2004). However, to make the results of this research match the 45 WS warning requirements as close as possible, only wind data from the ground to 300 feet were used in this study.

The 1500 UTC sounding was chosen since it is most representative of the atmospheric conditions a few hours before the time of the majority of convective wind events. Additionally, the 45 WS uses this sounding to produce a convective wind forecast for each day during the warm-season. The database also contains a consistent day-to-day record of 1500 UTC soundings, making it a more reliable data source from which to draw conclusions.

A dataset containing 61 different RAOB-based thermodynamic and wet downburst forecasting variables was computed for each day. Table 1 lists and briefly describes the predictors used in this study. These 61 predictor variables and the peak wind response variable were then read into R, a free and open-source statistical software environment (R Development Core Team 2009), in order to perform predictive analytic based forecast methods.

Table 1. List of 61 predictor variables and their associated acronyms included in the dataset.

Brief Description	Predictor Variable Acronym
Convective available potential energy (CAPE) derived from layer averages (J kg-1)	AvgCAPE
Downdraft CAPE (DCAPE) (J kg-1)	DAPE
DCAPE wind in knots	DWND
1000-500 hPa thickness (m)	thick
Freezing level (mb)	zl_mb

Freezing level (m)	zl_m
Height of the wet-bulb freezing level (mb)	wbz_mb
Height of the wet-bulb freezing level (m)	wbz_m
Precipitable water (inches)	pw
Surface to 500 hPa mean relative humidity	rh
Estimated maximum temperature (°C)	emxt
Surface Lifted Condensation Level (LCL) in mb	lcl_mb
Surface LCL (m)	lcl_m
Surface LCL temperature (°C)	lcl_T
700-500mb lapse rate (°C km-1)	lr75
Equivalent potential temperature (θ_e) index (K)	theI
Bottom layer in mb for θ_e index	theb
Top layer in mb for θ_e index	thet
Convective condensation level (CCL) in mb	ccl_mb
CCL (m)	ccl_m
Mean mixing ratio (g kg-1)	mmr
Convective temperature (°C)	conT
Cap strength	capS
Lifted index	LI
300 mb lifted index (LI)	LI300
700 mb lifted index (LI)	LI700
Showalter index	SI
Total-Totals index	TT
Vertical-Totals index	VT
Cross-Totals index	CT
K-Index	KI
SWEAT index	Sweat
Energy index	EnergyI
Parcel CAPE using 100 hPa layer (J kg-1)	CAPE
Maximum parcel upward vertical velocity (m s-1)	MxUVV
Convective inhibition (CINH) (J kg-1)	CINH
Parcel Cap Strength	CapS
Parcel LCL (mb)	LCL_mb
Parcel LCL (m)	LCL_m
Parcel level of free convection (LFC) in mb	LFC_mb
Parcel LFC (m)	LFC_m
Equivalent Level (EL) in mb	EL_mb
EL (m)	EL_m
Buoyancy at EL (J kg-1)	B_EL
Maximum parcel ascent level (mb)	MPL_mb
Maximum parcel ascent level (m)	MPL_m
Computed storm motion wind direction	SMdir
Computed storm motion wind speed	SMspd
Maximum (θ_e) in lower layer (K)	MXTHe-K
Height of the maximum θ_e (m)	ZMXTHe-K
Minimum mid-level θ_e (K)	MNTHe-K
Height of the minimum θ_e (m)	ZMNTHe-K
Difference between the maximum and minimum θ_e heights (m)	N-Xdz-m
Pressure at height of the maximum θ_e (mb)	PMXTHem
Pressure at height of the minimum θ_e (mb)	PMNTHem
Pressure difference between height of minimum and maximum θ_e (mb)	N-Xdpmb
Wet microburst severity index (WMSI)	WMSI
WMSI using average CAPE	WMSIa
Microburst downdraft potential index (MDPI)	MDPI
Modified MDPI1	MMDPI1
Modified MDPI2	MMDPI2

b. Quality control of the data

Quality control was performed on the wind data and each of the 61 RAOB derived predictor variables using an R script that performed data standardization in order to remove outliers that would otherwise contaminate the dataset and the results of this study. Data standardization involves a simple three step process: first, the arithmetic mean of each parameter is computed (Eq. 1); second, the standard deviation for each data point is calculated (Eq. 2); third, the difference between the arithmetic mean of the parameter and each observation value is divided by the standard deviation of the data point (Eq. 3). This results in a standard score for each data point. The above procedure is illustrated mathematically in Eqs. (1)-(3)

$$\mu = \frac{1}{n}\sum_{k=1}^{n} x_k \tag{1}$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(x_k - \mu\right)^2} \tag{2}$$

$$z_k = \frac{x_k - \mu}{\sigma} \tag{3}.$$

In Eqs. (1)-(3), μ is the arithmetic mean of each parameter, σ is the standard deviation of each data point, z_k is the standard score of the k^{th} data point, and x_k represents the k^{th} data point value for an arbitrary dataset of size n. This was done for each of the parameters to yield a standardized data matrix. Once this matrix was formulated, the R script removed any data point where the standard score was either less than -5 or greater than 5. A value of plus or minus 5 was chosen after some trial and error found that this value removed corrupt outliers while minimizing the removal of valid observations. This same script likewise flagged any missing data points with an NA, which stands for “not applicable”.

Since missing data points were represented with a -999 in the provided dataset, this was done so that R did not treat these values as actual data. R has several options that enable the user to specify how R handles the NAs during a particular task. Throughout the course of this study, the option was often set to have R ignore the NAs, which thereby prevented missing data from affecting the results.

An example of the benefits to quality controlling the data and removing outliers is displayed in the scatterplots in Fig. 7, which show WMSI and observed peak wind speed

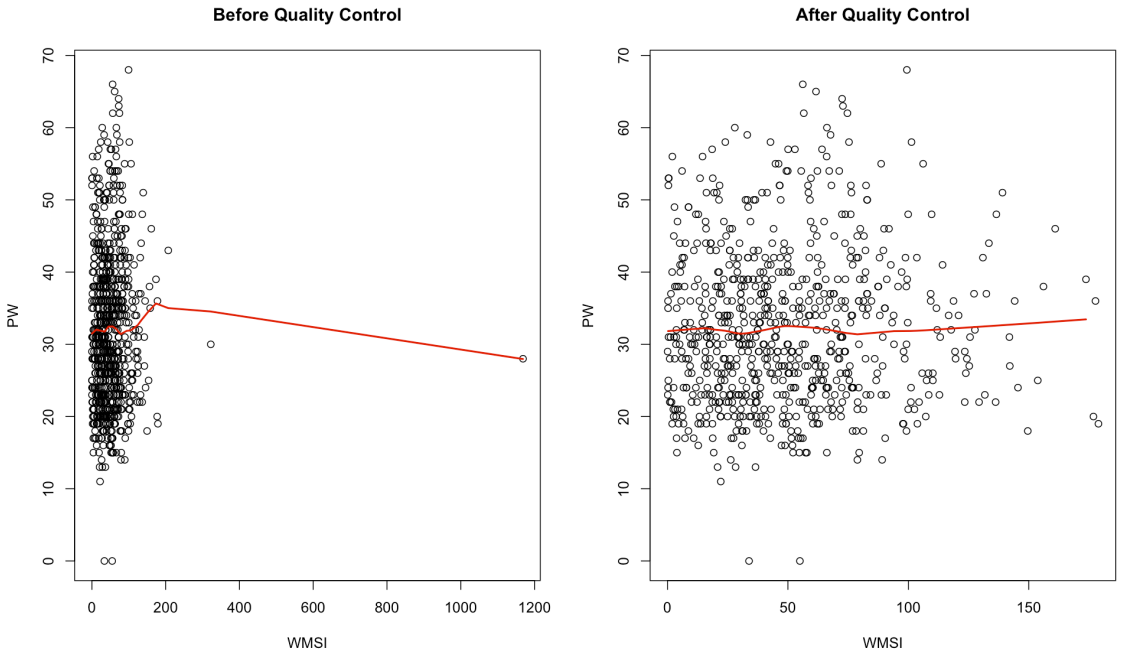

Figure 7. Scatterplots and curve fits of WMSI and peak wind data before (left) and after (right) quality control was performed.

(represented by a PW on the plots) data before (left) and after (right) it has been quality controlled. Comparing scatterplots of each of the predictors and peak wind speed before and after quality control revealed a similar pattern, illustrating the importance of "cleaning up" the data prior to using it to test current forecasting methods and build new ones.

In addition to removing outliers, all convective wind cases that occurred before 1600 UTC or after 0000 UTC were removed from the dataset. This was done because the primary purpose of this thesis is to evaluate forecasting methods that utilize the 1500 UTC KXMR sounding since this sounding is typically used to forecast convection and downbursts for that afternoon. The 1500 UTC sounding is not typically representative of atmospheric conditions during the overnight or morning hours, and consequently, should not be used to forecast convective activity during this timeframe. A late afternoon or early evening sounding is sometimes launched to provide forecasts for convection after 0000 UTC.

c. Evaluation of existing wet downburst forecasting indices

Atkins and Wakimoto (1991) developed an outlook forecasting technique used for forecasting wet downbursts in a weak synoptic wind environment that involves examining the atmospheric θ_e profile. This was done because it has been shown in modeling studies that cool, dry air in the mid-levels can aid downburst generation by being more susceptible to evaporative cooling and, therefore, greater negative buoyancy (Srivastava 1985, 1987; Proctor 1989). The Atkins and Wakimoto (1991) study on wet downburst activity over the southeastern United States found that the mean difference between the surface θ_e value and the minimum mid-level θ_e value was less than 13 K for days with thunderstorms but no downbursts and higher than 20 K for downburst days. This suggests that θ_e profiles can be used as a tool to differentiate between days with a high or low potential for wet downbursts. Loconto (2006) extended this logic to distinguish between days when winds greater than 35 knots or less than 35 knots were observed on the KSC/CCAFS complex. Using a small 66 case dataset – where one half of

the cases had winds less than 35 knots and the other half had winds equal to or exceeding 35 knots – to construct composite θ_e profiles, Loconto (2006) discovered that days where the winds reached or exceeded the 35 knot warning criterion possessed a greater θ_e difference between the surface and the mid-levels than days where the winds did not reach or exceed the 35 knot warning threshold (Fig. 8).

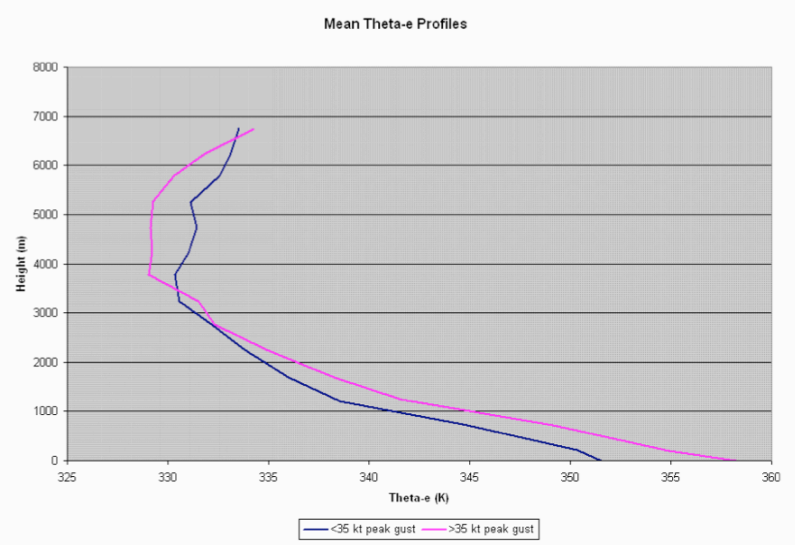

Figure 8. Composite θ_e profiles for a small 66 case dataset. Warning criteria days are in magenta, non-warning criteria days are in dark blue (Loconto 2006).

Based on the results of Atkins and Wakimoto (1991), Wheeler and Roeder (1996) derived a RAOB based microburst forecasting index, the Microburst Day Potential Index (MDPI), in an attempt to help the 45 WS forecast the likelihood of wet microbursts for any given day from the 1500 UTC KXMR RAOB data. The index they derived incorporates finding the difference between the maximum θ_e value in the low-levels of the atmosphere and the minimum θ_e value in the mid-levels in order to evaluate downdraft development potential. Mathematically, MDPI is defined as

$$MDPI = \frac{\max \theta_e - \min \theta_e}{CT}, \tag{4}$$

where max θ_e refers to the maximum θ_e value found in the lowest 150-hPa of the atmosphere, min θ_e refers to the minimum θ_e value between 650 and 500 hPa, and CT is a locally defined critical threshold that was empirically tuned to be 30 K for KSC/CCAFS. A MDPI greater than 1 implies steep θ_e lapse rates and a higher likelihood of wet downbursts with winds in excess of 35 knots should deep convection form while an MDPI of less than 1 signals a reduced risk of downbursts (Wheeler 1996).

In an attempt to include the effect of the vertical θ_e gradient on the daily potential for microbursts, the MDPI (MMDPI) was modified in two ways, each incorporating a different measure of the height of the maximum low-level θ_e and the minimum mid-level θ_e. Eq. (5) shows that MDPI was modified by assimilating the height difference in meters into the denominator while Eq. (6) illustrates that it was modified by putting in the pressure level difference (hPa) that corresponded to the height of the maximum or minimum θ_e level. Eqs. (5) and (6) are expressed as

$$MMDPI1 = \frac{100\left(\max\theta_e - \min\theta_e\right)}{z_{\min\theta_e} - z_{\max\theta_e}} \tag{5}$$

$$MMDPI2 = \frac{10\left(\max\theta_e - \min\theta_e\right)}{p_{\max\theta_e} - p_{\min\theta_e}} \tag{6}.$$

The values of 100 in (5) and 10 in (6) were used in an attempt to scale the values similar to those obtained from the MDPI. As with the MDPI, values of greater than 1 indicate steeper θ_e lapse rates and an enhanced potential for downbursts while values of less than 1 suggest a lower risk of strong convective winds.

Another convective wind forecasting index, the Wet Microburst Severity Index (WMSI), was developed by Pryor and Ellrod (2004) in order to assess both the potential

and severity of wet microbursts. Like the MDPI, it combines the difference between the low-level θ_e maximum and the mid-level θ_e minimum, but unlike the MDPI, it also incorporates Convective Available Potential Energy (CAPE), which is used to evaluate updraft potential. WMSI can be expressed mathematically as

$$WMSI = \frac{CAPE\left(\max\theta_e - \min\theta_e\right)}{1000} \quad (7).$$

A WMSI in excess of 50 was found by Pryor (2005) to produce wind speeds in excess of 35 knots.

An alternative wet microburst wind speed forecasting index derived by Proctor (1989) takes into account that wet microbursts are sensitive to the height of the melting level, the mean lapse rate below the melting level, and the ambient moisture content of the atmosphere below this level. More concisely, based on the results of his modeling studies, the index assumes that wet microburst strength increases with higher melting levels, steeper lapse rates, and as moisture content decreases from the ground to the melting level. The index may be written mathematically as

$$I = \frac{\sqrt{H_m^2\left(\gamma - \gamma_0\right) + H_m \frac{Q_l - 1.5Q_m}{3}}}{5}, \quad (8)$$

where H_m is the height of the melting level in meters AGL, γ is the lapse rate between the ground and the melting level (°C m^{-1}), γ_0 is a constant equal to $5.5*10^{-3}$ °C m^{-1}, Q_l is the ambient mixing ratio (g kg^{-1}) at 1 km AGL, and Q_m is the ambient mixing ratio at the melting level (Proctor 1989).

Building off of Proctor (1989), McCann (1994) composed the wind index (WINDEX) in order to better match observed microburst wind speeds. The parameters

are the same as in Proctor's Index except that the ambient mixing ratio is averaged over the lowest 1 km AGL to better represent the actual low-level moisture setting. Mathematically, the WINDEX can be expressed as

$$WINDEX = 5\sqrt{H_m R_Q \left(\gamma^2 - 30 + Q_l - 2Q_m\right)}, \tag{9}$$

where H_m is the height of the melting level in km AGL, Q_l is the mean mixing ratio from the surface to 1 km AGL (g kg^{-1}), Q_m is the mixing ratio at the melting level, R_Q is $Q_l/12$ but not larger than 1, and γ is the lapse rate from the surface to the melting level (°C km^{-1}) (McCann 1994).

While each of these wet microburst wind speed forecasting indices show some potential for use in the central Florida warm-season environment, they have not previously been thoroughly evaluated with a large dataset in order to assess their strengths, weaknesses, and biases. Additionally, since Proctor's index, WMSI, and WINDEX were not derived based on central Florida warm-season climatology, it is suspected that one or more of these indices may need to be tailored to better accommodate the local atmospheric conditions.

1) VERIFYING NUMERIC FORECASTING AIDS

In an attempt to evaluate the performance of Proctor's index and WINDEX, the predicted wind speeds from each of these indices were compared against the actual convective peak wind speed for days when convective winds occurred. More precisely, the observed peak wind speeds and predicted wind speeds were read into R to compute the mean error (ME), mean absolute error (MAE), root mean square error (RMSE), hit rate, and correlation coefficients for each of the RAOB derived indices versus the observed wind speed. The ME, MAE, and RMSE are defined in mathematical terms as

$$ME = \frac{1}{n}\sum_{k=1}^{n}\left(y_k - o_k\right) \tag{10}$$

$$MAE = \frac{1}{n}\sum_{k=1}^{n}\left|y_k - o_k\right| \tag{11}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(y_k - o_k\right)^2} \tag{12}.$$

In Eqs. (10)-(12), y_k is the k^{th} predicted wind speed value and o_k is the k^{th} observed wind speed value for an arbitrary dataset of size n. ME is used to diagnose whether an index has a tendency to over or under predict wind speed while MAE represents the characteristic magnitude of a forecast error for a given verification dataset (Wilks 2006). RMSE has the advantage that it retains the units of the forecasting variable and therefore is more easily interpreted as a typical error magnitude, but has the disadvantage that it can be overly high in situations where a only few of the errors are large (Wilks 2006). In an evaluation of Doppler radar data based predictive wind gust equations, Sullivan (1999) defined the hit rate as the percentage of wind speed observations that fall within plus or minus 5 knots of the predicted wind speeds. In addition to the ME, MAE, and RMSE, the hit rate method used by Sullivan (1999) was also used in the validation of the aforementioned predictive wind speed indices.

2) VERIFYING BINARY FORECASTING AIDS

However, because the MDPI, MMDPI1, MMDPI2, and WMSI assess the potential of a downburst with winds in excess of 35 knots as opposed to the actual wind speed, they were verified in a different manner. In order to verify the ability of the MDPI, MMDPI1, and MMDPI2 to forecast 35 knot or greater wind speeds, observed wind speed data were translated into a binary response (i.e. to forecast a "yes" or "no") variable. A 0

represented a day with either no convection or convection and winds less than 30 knots and a 1 represented a day where winds in excess of 30 knots occurred. The 30 knot threshold was chosen in order to provide cautious forecasting for 35 knot wind warnings. Meanwhile the MDPI, MMDPI1, and MMDPI2 data were translated into binary by setting any value of these indices that was greater than or equal to 1 to 1 and any value less than 1 to 0. In accordance with Pryor (2005), WMSI was translated into binary by equating any value of 50 or greater into a 1, and anything less into a 0. Days without an observed wind speed were excluded from the WMSI verification, since WMSI provides a wind speed related forecast, not a potential for occurrence forecast.

A common method for verifying a two class or binary forecasting aid is by means of a 2 by 2 contingency table, also known as a confusion matrix as shown in Wilks (2006).

Table 2. A schematic of a 2 by 2 confusion matrix.

Confusion Matrix		Observed	
		Yes	No
Predicted	Yes	a	b
	No	c	d

Several forecasting skill attributes can be calculated from the contingency table, including accuracy, bias, probability of detection (POD), probability of false alarm (POFA – Barnes *et al.* 2009), critical success index (CSI), Heidke's skill score (HSS), and true skill statistic (TSS) (Wilks 2006). These relationships are expressed mathematically in Eqs. (13)-(19)

$$accuracy = \frac{a+d}{a+b+c+d} \tag{13}$$

$$bias = \frac{a+b}{a+c} \tag{14}$$

$$POD = \frac{a}{a+c} \tag{15}$$

$$POFA = \frac{b}{a+b} \tag{16}$$

$$CSI = \frac{a}{a+b+c} \tag{17}$$

$$HSS = \frac{2(ad-bc)}{(a+c)(c+d)+(a+b)(b+d)} \tag{18}$$

$$TSS = \frac{ad-bc}{(a+c)(b+d)} \tag{19}.$$

Accuracy simply describes how many times a forecast was correct for an index. Its value ranges from 0 to 1, where 1 represents perfect accuracy and 0 no accuracy. Bias measures how well the forecast frequency of "yes" events compares to the observed "yes" events. It can range from 0 to infinity where values less than 1 represent an inherent under forecasting problem while values greater than 1 signify chronic over forecasting. Unbiased forecasting, meanwhile, would be characterized by a value of 1. The POD statistic merely illustrates what fraction of the observed "yes" events was correctly predicted by any given index. Its value ranges from 0 to 1, where a 1 represents perfect performance and a 0 the worst possible performance. On the other hand, the POFA defines the percentage of "yes" forecasts that failed to occur. Like POD, it too ranges from 0 to 1, but unlike POD, a 0 represents perfect performance, and a 1 the worst possible performance. Intuitively, since POD has a positive orientation and POFA a negative orientation, a desirable forecasting index is one that has a high POD and a low POFA. CSI addresses the correspondence between predicted "yes" events and observed

"yes" events including those that occurred randomly. Values close to 1 are desirable while values close to 0 are not. HSS is a statistic that evaluates the accuracy of an index with respect to random forecasting. Values between 0 and 1 represent forecasts that are better than random forecasting while values from 0 to -1 represent forecasts that are the same or worse than random forecasting, respectively. Lastly, the TSS evaluates how well the forecasting index separates observed "yes" events from observed "no" events. Its value also ranges from -1 to 1, with positive values corresponding to forecasts that more frequently match the observations and negative values representing forecasts that are more commonly the opposite of the observations. A more detailed treatment of these and other indices is provided by Wilks (2006).

d. Composite soundings

Extending Loconto's (2006) study, which constructed composite θ_e profiles for a small 66 case dataset consisting of 33 warning and 33 non-warning criteria wind events (Fig. 8), composite θ_e profiles were built for the entire 1995 to 2009 dataset in order to see if a larger dataset produced a similar outcome. In addition, a bigger dataset is more representative of the overall climatology and provides more statistically reliable results that shed light on whether or not the 45 WS can utilize vertical θ_e profiles as a forecasting tool. Composite θ_e profiles were built with a FORTRAN 95 program that first classified days based on whether or not the recorded peak wind speed reached or exceeded 35 knots before constructing separate vertical θ_e profiles for each category by computing the average θ_e in 40-hPa increments from the surface to 200 hPa. A similar process was done to see if vertical profiles of temperature, dew point, wind speed, and wind direction

showed any difference for each category and potential as a means for forecasting 35 knot or stronger convective winds.

e. Predictive analytic techniques

This section discusses the five predictive analytic techniques used to create forecast models. The five techniques are: 1) multiple linear regression, 2) logistic regression, 3) multinomial regression, 4) classification and regression trees, and 5) ensembles of classification and regression trees via the three bootstrapping algorithms of bagging, boosting, and random forests. The results of these techniques are in chapter 4.

1) MULTIPLE LINEAR REGRESSION

Multiple linear regression (MLR) models analyze the relationship between a numeric response and multiple predictor variables. The relationship is expressed as an equation that predicts the response as a linear function of each of the predictors. Models are built by minimizing the sum of the squared residuals, a process known as ordinary least squares estimation.

Numerous methods of variable selection were tested in order to come up with the most accurate and simple MLR models. The models were built using a training dataset from the 1995 to 2007 warm seasons and validated with an independent dataset from the 2008 and 2009 warm seasons. However, if the two datasets do not follow a similar distribution, using a separate dataset to validate models can present misleading verification results. As such, two simple tests were performed to assess whether or not this issue would occur. A Kolmogorov-Smirnov test (K-S test) was used to test the null hypothesis that the two separate datasets were drawn from the same continuous distribution while a Quantile-Quantile plot (Q-Q plot) was used to graphically compare

the differences in the distributions by plotting their quantiles against each other. Since the results from the K-S test gave a p-value of 0.5546 and Fig. 9 displays a linear trend that roughly follows the line y = x, the null hypothesis that the two datasets followed a similar

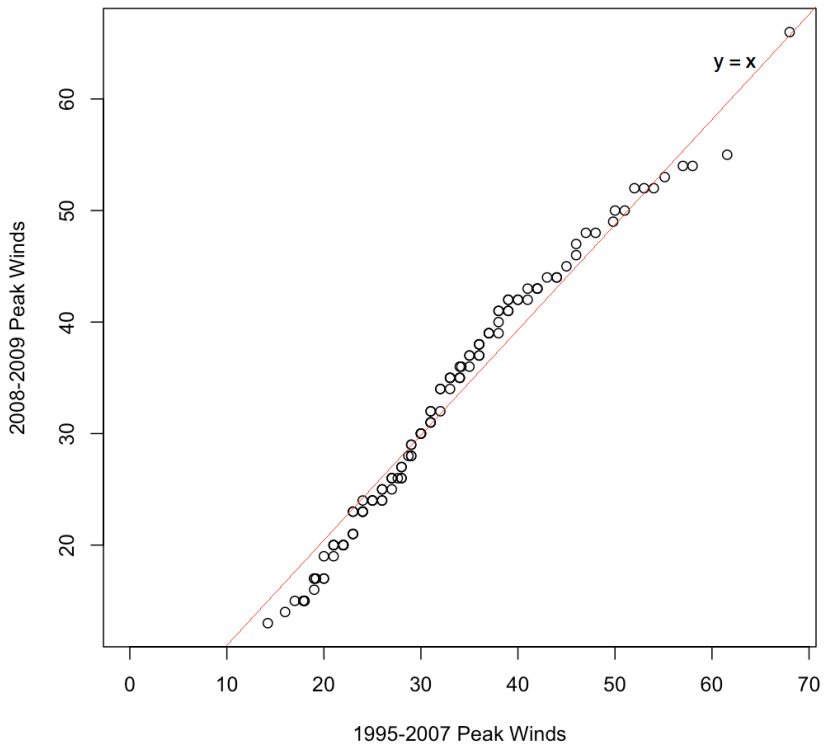

Figure 9. Q-Q plot of dependent (1995-2007) and independent (2008-2009) observed peak wind speeds.

distribution was not rejected. Consequently, it was concluded that validation of these models with this independent dataset was adequate in evaluating the model's ability to forecast in a volatile and fickle environment.

Because the MLR models are trying to predict the maximum possible wind speed for days when convective winds occur, all data from days where there was no convective wind event were excluded from model construction and validation. This is because the

intent of the MLR models is not to forecast whether or not convection will occur, but rather, to forecast the strongest expected wind gust should convection be predicted by one of the techniques discussed in the subsequent sections.

2) LOGISTIC REGRESSION

Unlike MLR, logistic regression (LR) uses a two-class categorical response as opposed to a numeric one. This was done to build a model that predicts the probability that an event will fall into a particular class. After model construction, probabilities were computed with the independent dataset. These probabilities were then translated into a binary outcome for model validation with a classification threshold that optimized model accuracy.

In order to see if LR provided more promising results, the peak wind response variable was translated into a two-class response in one of two ways. In the first method, a value of 0 was assigned to the response variable for days when no convection occurred, and a 1 for days when a convective wind gust was recorded regardless of strength. This was done in an attempt to build LR models that help 45 WS forecasters to gain a better idea whether or not the environment is conducive for convection. In the second method, the response variable was assigned a value of 0 for a peak wind speed of less than 35 knots and a 1 for wind speeds of 35 knots or greater. As with MLR, any data for non-convective days were removed since the purpose of this LR model type is to forecast whether or not warning threshold winds will occur should the first type of LR model predict convection. These LR models were considered to see if they would help 45 WS meteorologists to better diagnose days where the atmospheric conditions are more favorable for warning criteria winds in the event that thunderstorms form.

LR models for the two response variable classification methods were built with the training data using many of the same variable selection techniques utilized in constructing the MLR models. Evaluation of model performance was also done with the independent data from 2008 and 2009. As discussed previously, the 2008 and 2009 seasons were shown to be representative of the entire data sample, and thus, appropriate to use as independent test data.

3) MULTINOMIAL LOGISTIC REGRESSION

In an attempt to avoid having to use two separate LR models to make daily forecasts, multinomial logistic regression (MR) models were built and tested. MR models produce a probability that an event will fall into one of three or more classes, as opposed to just two. The class with the highest probability is then chosen as the forecasted outcome. In this case, the response variable was categorized into three classes where non-convective days were assigned a value of 0, days with convection and winds less than 35 knots a 1, and days with convection and winds greater than or equal than 35 knots a 2. As with other types of regression, several variable selection techniques were used to find the model that performed best with the independent data. Models were selected based on their ability to forecast accurately using cross-validation. Finally, the performances of the best performing LR and MR models were compared to see whether using two LR models or a single MR model provided better results.

Regression models have the advantage that the equations can be easily programmed to output a daily forecast value based on 1500 UTC RAOB data, but have the disadvantage that they require exhaustive and complex methods of variable selection that demand considerable trial and error.

Three class forecast verification

Validating three class forecasts such as those produced by MR models can be done with a 3 by 3 contingency table (confusion matrix) like the one shown below (Wilks 2006).

Table 3. A schematic of a 3 by 3 confusion matrix.

Confusion Matrix		Observed		
		o1	o2	o3
Predicted	y1	y1, o1	y1, o2	y1, o3
	y2	y2, o1	y2, o2	y2, o3
	y3	y3, o1	y3, o2	y3, o3

In the above table, y_i represent the forecasted values while o_j signify the observed values. From this, a modified version of accuracy, HSS and TSS can be computed as well as a new statistic, the Gandin-Murphy Skill Score (GMSS) (Wilks 2006; Gandin and Murphy 1992). The expressions for these performance metrics are shown in Eqs. (20)-(22)

$$HSS = \frac{\sum_{i=1}^{I} p(y_i,o_i) - \sum_{i=1}^{I} p(y_i)p(o_i)}{1 - \sum_{i=1}^{I} p(y_i)p(o_i)} \tag{20}$$

$$TSS = \frac{\sum_{i=1}^{I} p(y_i,o_i) - \sum_{i=1}^{I} p(y_i)p(o_i)}{1 - \sum_{j=1}^{J} \left[p(o_j)\right]^2} \tag{21}$$

$$GMSS = \sum_{i=1}^{I}\sum_{j=1}^{J} p(y_i,o_j)s_{i,j}, \tag{22}$$

where

$$s_{i,j} = \frac{1}{J-1}\left[\sum_{r=1}^{I-1}\frac{1}{D(r)} + \sum_{r=j}^{J-1} D(r) - (j-i)\right] \tag{23}$$

and

$$s_{j,j} = \frac{1}{J-1}\left[\sum_{r=1}^{j-1}\frac{1}{D(r)} + \sum_{r=j}^{J-1} D(r)\right] \tag{24}.$$

Finally,

$$D(j) = \frac{1 - \sum_{r=1}^{j} p(o_r)}{\sum_{r=1}^{j} p(o_r)} \tag{25}.$$

The above formulae include terms for the joint distribution of forecasts and observations, $p(y_i,o_j)$, the marginal distributions of the forecasts, $p(y_i)$, and the marginal distributions of the observations, $p(o_j)$. In calculating GMSS, scoring weights are computed where $S_{i,j}$ is the scoring weight for incorrect forecasts and $S_{j,j}$ is the scoring weight for correct forecasts. D(j) represents the odds ratios and r a dummy summation index (Wilks 2006). As with the two-class response, HSS estimates how the forecasting index compares to random forecasting while the TSS evaluates how well the index differentiates different types of forecasts from one another. GMSS is another skill score that differentiates between single and multiple-category forecast misses. Its value ranges from 0 to 1 with 1 representing perfect forecasting and 0 signifying random forecasting (Gandin and Murphy 1992). The advantage to using GMSS is that forecasts that are off by two or more categories are scored as worse forecasts than adjacent-category misses. In other words, "near miss" forecasts are penalized less by GMSS. HSS and TSS, on the other

hand, only depend on the proportion of forecasts correct (Gandin and Murphy 1992; Wilks 2006).

4) VARIABLE SELECTION TECHNIQUES

Several methods of variable selection were used in order to find which of the 61 predictor variables provided the simplest and best performing MLR, LR, and MR models. In addition, reducing the number of dimensions in a model makes it simpler to work with and interpret. More specifically, the variable selection methods used in this study included using Akaike's Information Criterion (AIC) in a stepwise regression algorithm, removing non-statistically significant predictors one by one until all remaining predictors were below a threshold p-value, determining variable importance as found by the boosting and random forests algorithms (see ensemble classification and regression tree methods in subsection 6), and finding which predictors had the most correlation with the response variable. Principle Components Analysis (PCA) was also used as a means of reducing the number of dimensions in the linear regression models, even though it does not "remove" variables in the same manner that the other methods discussed in this section do. Each of the simplified models was then compared to a model with all 61 predictors incorporated into it in order to compare how they performed in relation to the full model.

The first means for selecting which predictor variables to use in a linear regression model tested in this study involves using Akaike's Information Criterion (AIC). In its most basic form, AIC is a measure of how well the model fits the data and is used as a tool for model selection. AIC is expressed mathematically as

$$AIC = 2k - 2\ln(L), \tag{26}$$

where k is the number of predictors in the model and L is the maximized value of the likelihood function for the model (Akaike 1974). The objective of using this method is to compare multiple models with different combinations of predictors until the model with the lowest AIC value is found. In other words, the goal is to minimize AIC by striking a balance between the goodness of the model's fit, which is represented in the log-likelihood value, and a penalty term that increases with the number of parameters in the model (Wilks 2006). AIC is used in a stepwise regression algorithm by first finding the AIC value for a model with all 61 predictors included into it before using a backward elimination procedure that removes predictors one by one until a minimum AIC value is achieved.

Another variable selection technique used in this study involves eliminating non-statistically significant predictor variables via a chi-square test that computes a p-value for each predictor. The algorithm begins by computing the p-value for each predictor in the full 61-predictor model. The predictor with the highest p-value is then removed before another regression model is built. P-values for the remaining predictors were then computed via the chi-square test before the predictor with the greatest p-value was, once again, removed. This process was repeated until all of the remaining predictor variables in the regression model had a p-value of less than 0.1, a value that is generally regarded as a common threshold for an indicator of statistical significance.

The boosting and random forests ensemble classification and regression tree (CART) algorithms, which will be discussed in greater detail in subsection 6, have the ability to establish the relative importance of the variables used in ensemble CART models. This is determined by measuring the prediction accuracy of each of the

predictors in the ensemble CART models using data that was not used in constructing the models. It was thought that including those predictors deemed important by the ensemble CART models would result in better performing linear regression models. Variable importance plots produced by the random forests algorithm (Fig. 10) were then used to

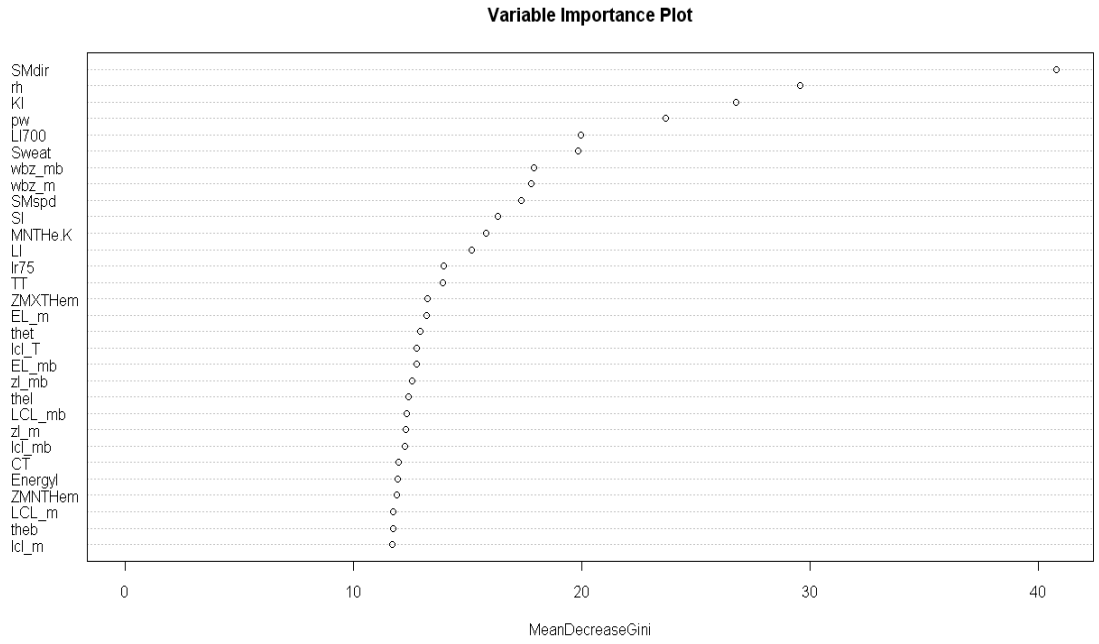

Figure 10. Example of a variable importance plot produced by the random forests algorithm. More important variables appear at the top of the plot.

determine which predictors to incorporate into MLR, LR, and MR models. Unfortunately, due to a limitation in the boosting algorithm in R that only allows it to work with a binary response, the boosting variable importance plots (Fig. 11) were only used to determine the variables to be included in the LR models. In order to minimize

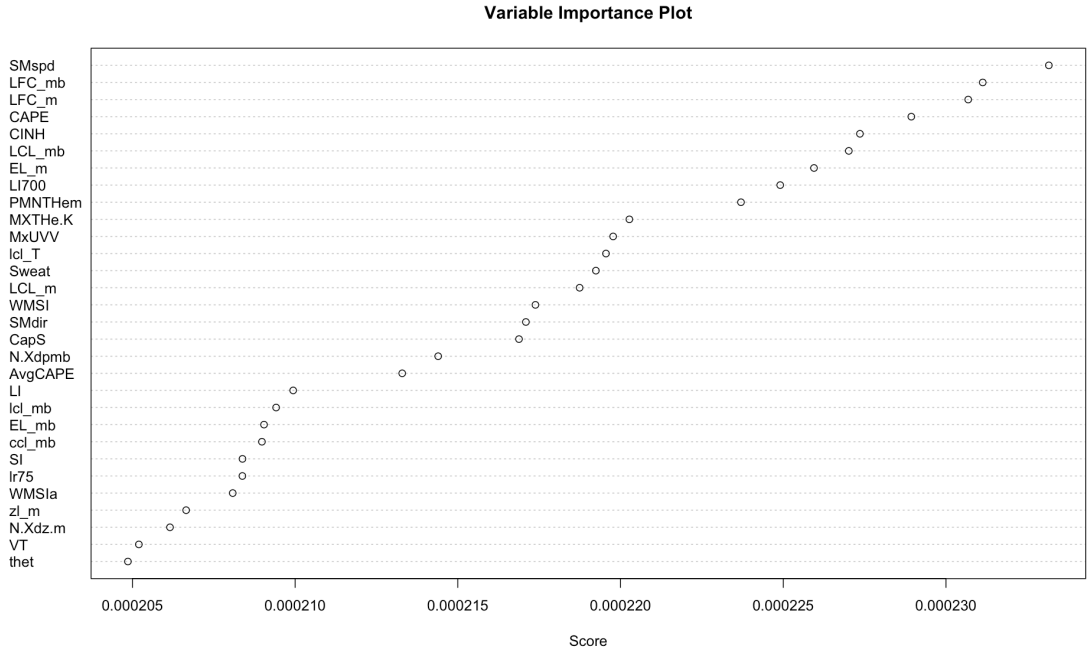

Figure 11. Example of a variable importance plot produced by the boosting algorithm. More important variables appear at the top of the plot.

subjectivity as much as possible, the ten most important variables indicated by the plots were included in the models.

Correlation coefficients between the peak wind response variable and each of the 61 predictors were calculated in order to find which ones displayed the most correlation. They were computed by taking the average of three commonly used measures of correlation: Pearson's product-moment correlation coefficient, Spearman's rank correlation coefficient, and Kendall's tau correlation coefficient. Pearson's product-moment correlation coefficient is a measure of the strength of the linear dependence between two variables. Spearman's rank correlation coefficient is simply the Pearson coefficient computed using the ranks of the data, i.e. it is a measure of how well one variable increases or decreases with respect to the increase of the other variable, regardless of whether the increase or decrease of the first variable is linear or non-linear.

Kendall's tau correlation coefficient is used to measure the degree of correspondence between a pair of rankings (Wilks 2006). It was anticipated that using a mean of the three measures of correlation would provide the best representation of the actual amount of correlation between the response and each of the predictors. Variables showing the highest degree of correlation were then selected for use in the regression models since it was thought that this might lead to a superior performing model.

PCA reduces the number of predictors in a model by linearly combining the original predictors into a smaller number of new predictors. This is done so that the first new linear combination (principle component) accounts for as much of the variability in the data as possible, while successive principle components account for as much of the remaining variability as possible (Wilks 2006). Components that contained a standard deviation in excess of one were included in the regression models examined in this study since components with lower standard deviations typically accounted for less than 10% of the variance observed in the data. Unlike the other variable selection methods described in this section, PCA does not successively remove predictors until the best model is found; rather, predictors are linearly combined to produce a model with far fewer dimensions, with each subsequent component explaining the most independent linear variance. PCA makes the resulting model far easier to work with and is especially useful in building MLR models efficiently and effectively.

5) CLASSIFICATION AND REGRESSION TREES

Classification and regression tree (CART) (Breiman *et al.* 1984) models are combinations of cluster analyses and discriminant analyses that try to optimally stratify the response variable by related groups of predictor variables. Since various assumptions

can be used in the optimization algorithm, different CART models can yield different results. Unlike linear regression methods, CART forecasting methods overcome the issue of model variable selection since all of the predictor variables are used in growing the trees They also have the advantage that they are usually easy to use and automate. Disadvantages to CART include the reasons why some variables and thresholds were chosen may not be easily understandable and the performance metrics are sometimes not as familiar as with more common techniques. Understanding the reasons why a technique works can be important in the psychological acceptance of the technique by forecasters.

Classification trees make use of a categorical response while regression trees utilize a numerical one. Both tree types were grown, validated, and tested with the same datasets and dependent variable types as the regression models. Three major tree growing algorithms were tested in the statistical program R in order to see which one produced the best performing classification and regression trees. The first algorithm – the "tree" algorithm (Ripley 2009) – evaluated uses recursive partitioning to grow decision trees by splitting the observations based on how well the predictor variables' values can separate the observations into distinct groups in terms of the homogeneity of the response. The predictor and its associated value that produce the most purely split groups is chosen for the first node of a tree. This process iterates until some suitable stopping point is reached. The second algorithm – the "rpart" algorithm (Therneau and Atkinson 2010) – works in a similar manner as the "tree" algorithm except that it uses different default control parameters that allow for more splitting to occur and, thus, larger trees to be grown. The third algorithm – the "party" algorithm (Hothorn *et al.* 2006) – uses recursive partitioning in the conditional inference framework to grow trees. The idea behind this algorithm is a

relatively straightforward process that begins by testing the null hypothesis of no association between each of the predictors and the response. If the null hypothesis cannot be rejected, tree splitting is stopped and the algorithm terminates. Otherwise, the predictor that has the strongest association (lowest p-value for the test of no association) with the response is selected to split the data into groups. This process continues until none of the predictor variables have a statistically significant relationship with the response.

However, because trees can become overly complex when using all of the predictors, a means of pruning each tree is required. The “tree” algorithm contains a function that enables the user to manually snip off individual nodes (leaves) from the tree in order to trim it to a more desirable size. An automatic version of this function snips off nodes from the tree and returns information to the user that can be used to decide which pruned version of the tree should be chosen. In the case of the “rpart” algorithm, tree pruning can normally be accomplished by specifying either the complexity parameter or the maximum depth of the tree. The primary objective is to specify the complexity parameter or the maximum depth of the tree so that both the size of the tree and the prediction misclassification error are minimized. Finding the optimal complexity parameter is a rather simple process that involves using Fig. 12. Fig. 12 displays a plot that relates the X-relative error (normalized misclassification error), size, and complexity

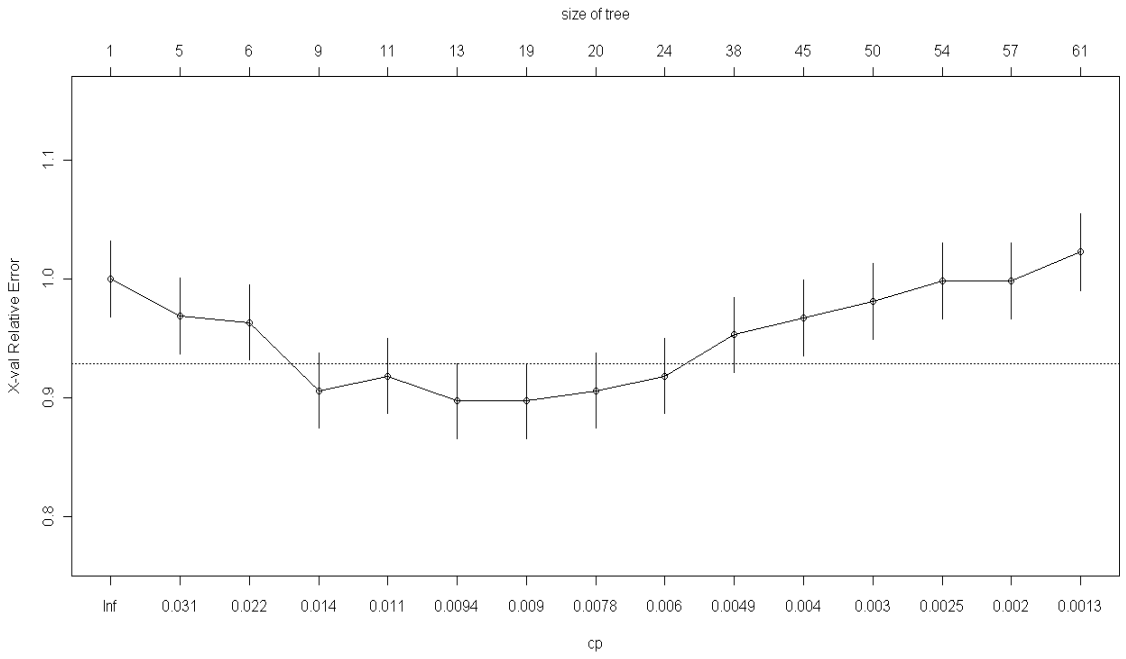

Figure 12. Plot of X-relative error, size of tree, and complexity parameter (CP) for a set of trees using all 61 predictor variables. CP is chosen so that both the size of the tree and the X-relative error are minimized.

parameter (CP) for trees using all 61 of the predictors. Typically the X-relative error will decrease to a point before "leveling off" for a while as the tree size increases. Eventually, however, the trees become so large that their misclassification errors actually begin to increase, which is analogous to over-fitting in linear regression. The intention is to choose a complexity parameter near the point where the leveling off begins so that the pruned tree will be the simplest and most accurate one possible. For example, Fig. 12 suggests that a complexity parameter of 0.0094 is a suitable one since this is a point where both the error and tree size are at their smallest. An R script automated this process by testing numerous complexity parameters to determine which tree yielded the lowest misclassification error. The simpler method of pruning the trees in the "rpart" algorithm involves specifying their maximum depth, a process that was accomplished with an R

script that tested various maximum depths until the one that yielded the tree with the lowest misclassification error was found. Pruning trees with the "party" algorithm, on the other hand, can be completed by either specifying the maximum depth of the tree, as was done with the "tree" and "rpart" algorithms, or by setting the difference between one and the p-value (1 – p-value) that must be exceeded in order for splitting to continue. Again, R scripts automated these two procedures by testing which maximum depth or 1 – p-value gave the smallest and more accurate trees.

6) ENSEMBLE CART USING BOOTSTRAPPING

Finally, bootstrapping algorithms were tried in order to see if using an ensemble of CART models provided better results. In its most simple form, bootstrapping involves the creation of multiple learning samples of the data by repeated random sampling with replacement. Bootstrapping can be applied with CART to develop three separate algorithms.

The first algorithm tested was the bagging algorithm (Breiman 1996; Peters *et al.* 2002), also known as the bootstrap aggregation algorithm. In essence, this algorithm uses bootstrapping to create multiple versions of a classifier such as classification or regression trees, each grown upon a bootstrapped sample, before aggregating these to produce a predicted result. For regression trees, the algorithm employs a simple process that begins with taking a bootstrap sample from the original dataset before fitting a tree to this data. A prediction is then made from the resulting tree. These steps are performed a large number of times (normally 50-1000) and the predictions from each tree are averaged to produce a final result. In the instance of classification trees, the process is the same except that the final prediction is chosen by a popular vote of each of the predicted

outcomes from the collection of trees instead of averaging the predictions from the all of the trees. The basic idea behind bagging is that by averaging the predictions over multiple samples, the variability of the prediction is reduced while its unbiased nature is simultaneously preserved.

The random forests (Breiman 2001; Liaw and Wiener 2009) algorithm is similar to bagging except that it chooses a random subset of predictor variables instead of using all of them. The number of randomly chosen predictors is usually fixed. By inserting randomness in this manner, the correlations between predictions generated by individual trees are reduced. This subsequently lowers the variance of the prediction error. Furthermore, by using fewer predictors in each tree, a significant computational savings is made.

Boosting (Freund and Schapire 1996; Culp *et al.* 2006) is a tool that classifies binary response variables (recall that due to a limitation in the boosting algorithm in R, it can only handle a binary response) with multiple classification trees. The basic idea behind the algorithm is to combine predictions from a group of weak classifiers in such a manner that the averaged predictions make a stronger classifier. The algorithm begins by growing a tree on a learning sample and predicting a class for an observation. If correctly classified, the observation receives less weight; if not, it gets more. Trees are repeatedly grown on the reweighted samples with incorrectly classified observations getting larger weights than the correctly identified ones. Consequently, cases that are difficult to classify receive ever-increasing weight, thereby increasing their chance of being correctly classified by the classifier. The final classification is produced by a weighted vote of the iteratively produced classifiers after the loop reaches a user specified stopping criterion.

Since the bagging and random forest algorithms can handle a non-binary categorical response, the wind speed was fed into the algorithms as either a numeric value, a two-class binary response as was done for two types of LR models, or a three-class response as it was for the MR models. Experimentation was done in order to see which method yielded the most promising result.

Validation of the bootstrapping models was done differently than it was for the linear regression or CART models. Instead, model verification was done with data not selected for any of the bootstrap samples, sometimes referred to as the out-of-bag (OOB) data. Although not a predetermined dataset like the 2008 and 2009 warm-season independent dataset used to verify the linear regression and CART models, OOB data can be interpreted as independent data since it is not used in building the bootstrapping models. An advantage to using the OOB data is that cross-validation is not necessary, resulting in a computational savings. The OOB data were then used to calculate the OOB error, which can be either a misclassification error in the case of a categorical dependent variable, or a RMSE in the case of a numeric one. Other common statistical performance metrics can be computed from the OOB data as well.

Bootstrapping methods have the advantage that they counter the problem that individual trees can be poor predictors and difficult to interpret, especially if they're large. They also mitigate the issue of excessive variance in the predicted outputs produced by single trees. Furthermore, since verification is done using data not used in any of the bootstrap samples, no independent dataset is needed, implying that an entire dataset can be used in model construction. As with CART, bootstrapping models do not need to undergo exhaustive trial and error variable selection methods. Finally, they can

be easily automated with a computer script to produce real-time forecasts. Bootstrapping methods have the primary disadvantage that the reason for the final forecast will not be known, since it has been dispersed across many variables and thresholds.

CHAPTER 3

3. Results of existing wet downburst forecasting tools

The performance results of the wet downburst forecasting tools currently used by the 45 WS are discussed here.

a. Performance of existing wet downburst forecasting indices and suggested improvements

Evaluation of Proctor's index and WINDEX with 1500 UTC KXMR RAOB data from 1995 to 2009 revealed that neither of these indices predicted peak wind speeds with much accuracy. The performance of the aforementioned wet downburst forecasting indices are summarized in Table 4. All performance metrics were discussed in chapter 2, section c.

Table 4. Performance metrics of WINDEX and Proctor's Index.

	WINDEX	Proctor
RMSE	24.92	10.75
MAE	21.94	8.644
ME	21.28	1.101
Hit Rate	0.086	0.355
Correlation	0.109	0.147

As table 4 illustrates, Proctor's index is the better performer of the two indices with an observed wind speed falling within 5 knots (hit rate) of the forecast just over 35% of the time. In addition, it also has the lower RMSE, MAE, and ME. The WINDEX over predicts wind speed since its ME value is significantly positive. Correlation coefficients for these two indices illustrate little correspondence between the predicted and actual wind speeds.

Translating Proctor's index, WMSI, WINDEX, and the observed wind speeds into binary variables (any value of Proctor's index or WINDEX greater than 35 was set to 1

and less than 35 to 0) and performing the subsequent verification yielded the results displayed in Table 5. All binary performance metrics were defined in chapter 2, section c.

Table 5. Binary forecasting verification of wet microburst forecasting indices.

	WMSI	WINDEX	Proctor
Bias	1.129	2.395	0.964
Accuracy	0.526	0.420	0.572
POD	0.473	0.982	0.455
POFA	0.581	0.590	0.528
CSI	0.286	0.407	0.302
HSS	0.034	0.015	0.108
TSS	0.034	0.019	0.108

Table 5 also indicates that Proctor's index, WMSI, and WINDEX do not have much forecasting capability since they do not display desirable performance metrics. The extreme over forecasting problem associated with the WINDEX is also evidenced here with its bias in excess of 2. Although WMSI and Proctor's index are not significantly biased, they each have a POFA that is greater than the POD. Moreover, all of these indices do not perform much better than random forecasting, as evidenced by their HSS values near 0, and have little ability to differentiate between days with or without warning criteria wind speeds, as indicated by their TSS values near 0. In general, an HSS or TSS of at least 0.3 is usually considered as a need for a forecast technique to be even marginally useful in real-world operations.

In an attempt to correct for the intrinsic over forecasting issue associated with the WINDEX, Eq. (9) was modified to better accommodate the higher mixing ratios typically found in the central Florida warm-season environment. Since the ratio $Q_l/12$, which is represented in the term R_Q, cannot be greater than 1 (recall that Q_l is the mean mixing ratio in g kg^{-1} from the surface to 1 km AGL), this implies that the mean low-level

mixing ratio must be less than 12 g kg^{-1}. Computation of this ratio for the KXMR RAOB dataset found that the ratio averaged well in excess of 1, illustrating the need for this ratio to be adjusted. After some trial and error, it was decided to increase the constant in the denominator from 12 to 18. In addition, the constant of 30 was raised to 35 to better account for the steep low-level lapse rates found in the dataset. The modified WINDEX (MWINDEX) can be written as

$$MWINDEX = 5\sqrt{H_m R_Q \left(\gamma^2 - 35 + Q_l - 2Q_m\right)}, \tag{27}$$

where all of the variables are as defined in Eq. (9) except that R_Q is $Q_l/18$ instead of $Q_l/12$.

Verification statistics of the MWINDEX are displayed in Table 6 while the binary verification statistics for the index are shown in Table 7.

Table 6. Performance metrics of the modified version of the WINDEX.

	MWINDEX
RMSE	12.38
MAE	9.878
ME	-0.462
Hit Rate	0.316
Correlation	0.130

Table 7. Binary performance metrics of the modified version of the WINDEX.

	MWINDEX
Bias	0.960
Accuracy	0.575
POD	0.457
POFA	0.523
CSI	0.304
HSS	0.114
TSS	0.113

While not overly promising, the numbers in these tables do suggest some improvement in the ability of the MWINDEX to forecast for central Florida's warm-season convective wind environment since it corrects the chronic over forecasting issue found with the WINDEX. It is speculated that using this index with an afternoon sounding to forecast wind speed in more of a nowcasting situation may yield some promise since conditions are likely to change markedly between consecutive soundings.

As Table 8 indicates, verification of the MDPI, MMDPI1, and MMDPI2 found that they performed worse by most measures than Proctor's index, WMSI, and WINDEX.

Table 8. Performance metrics of MDPI, MMDPI1, and MMDPI2.

	MDPI	MMDPI1	MMDPI2
Bias	1.755	0.541	0.766
Accuracy	0.624	0.741	0.715
POD	0.392	0.094	0.138
POFA	0.777	0.826	0.820
CSI	0.166	0.065	0.084
HSS	0.055	-0.014	-0.012
TSS	0.071	-0.012	-0.011

The rather high bias indicates that the MDPI forecasts 30 knot or greater wind speed days too frequently, while the biases of well below 1 for the MMDPI1 and MMDPI2 show that neither of these indices forecasts them with enough regularity. The relatively high accuracy values of these three indices are overshadowed by the low POD and high POFA values, suggesting that the higher accuracy may be due to chance forecasts that verified as opposed to sound forecasts that verified. Additionally, none of these indices perform better than random forecasting and have little or no ability to differentiate between days with weak or no convective winds and warning criteria winds.

Finally, a modification to Proctor's index was implemented in order to better represent the low-level moisture profile of the atmosphere. Since Eq. (8) indicates that only the 1 km mixing ratio (Q_l) is employed in calculating the index, the mean mixing ratio from the surface to 1 km AGL was used instead. This modification of Proctor's index is expressed as

$$I = \frac{\sqrt{H_m^2(\gamma - \gamma_0) + H_m \frac{Q_l - 1.5Q_m}{3.5}}}{5}, \tag{28}$$

where the terms are as defined above in Eq. (8). After some empirical tuning, a constant of 3.5 was adopted in order to account for the slightly higher values of Q_l found with mean mixing ratio profiles in the lowest 1 km of the atmosphere.

Evaluation of the modified index yielded the results shown in Table 9 while binary validation results of the index are displayed in Table 10.

Table 9. Performance metrics of the modified version of Proctor's Index.

	Modified Proctor's Index
RMSE	10.70
MAE	8.591
ME	0.737
Hit Rate	0.369
Correlation	0.155

Table 10. Binary performance metrics of the modified version of Proctor's Index.

	Modified Proctor's Index
Bias	0.946
Accuracy	0.571
POD	0.442
POFA	0.533
CSI	0.294
HSS	0.101
TSS	0.100

As such, despite a considerable effort to improve Proctor's index, the modified version of this index still does not show much potential for use as evidenced by the poor performance metrics in each of the above tables.

In sum, the overall performance of the wet microburst forecasting indices studied is not impressive when using 1500 UTC KXMR RAOB data. As such, it is advised that 45 WS forecasters exercise caution when using any of these tools to predict wet microburst wind speeds.

b. Composite soundings

1) COMPOSITE θ_e PROFILES

The composite θ_e profiles for convective days shown in Fig. 13 reveal a layer of lower θ_e values in the mid-levels for the days where winds of 35 knots or greater were

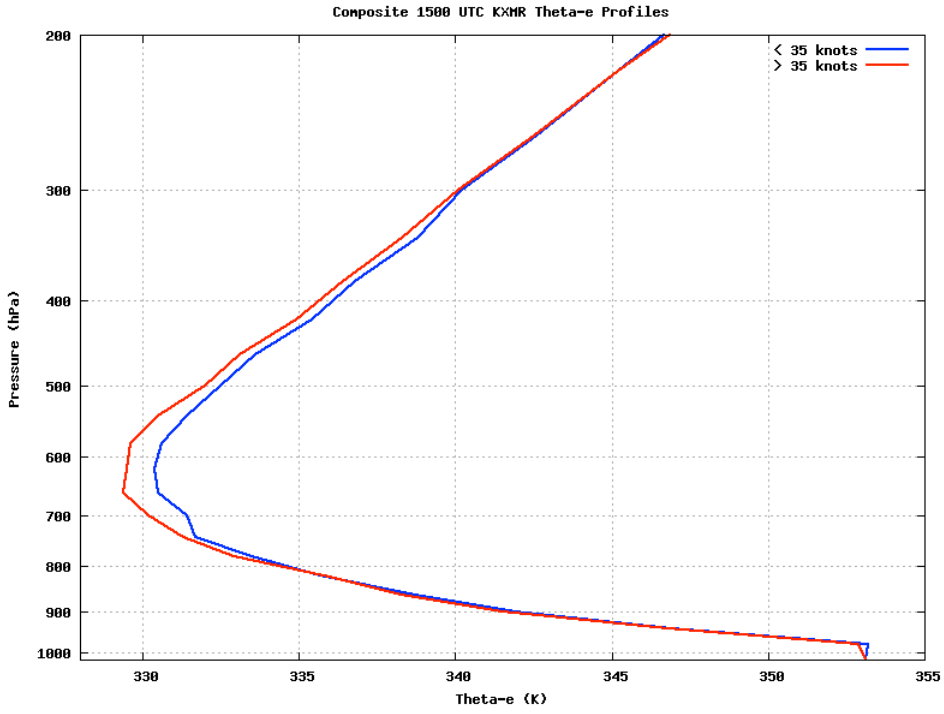

Figure 13. Composite 1500 UTC KXMR θ_e profiles for warning (red) and non-warning (blue) days. Profiles contain data from 1995 to 2009.

observed than for the days when winds of less than 35 knots were observed. More precisely, θ_e values of less than 330 K between 700 hPa and 550 hPa were typical of the stronger convective wind days.

Given the apparent difference seen in the mid-level θ_e values, a quantitative evaluation was done with the 1995 through 2009 θ_e data in order to assess whether the daily θ_e profiles have some potential in forecasting wind strength. Evaluation was done by computing the difference between the maximum low-level θ_e and the minimum mid-level θ_e ($\Delta\theta_e$) for each day and comparing these values to the average $\Delta\theta_e$ for all convective days, regardless of the observed wind speeds. This average was found to be 27.5 K. Based on the idea that greater $\Delta\theta_e$ values imply stronger convective winds, if the daily $\Delta\theta_e$ value was found to be larger than the 27.5 K threshold, a warning level wind was forecasted; if not, no warning level wind was forecasted. Table 11 illustrates the performance metrics of using a $\Delta\theta_e$ value of 27.5 K.

Table 11. Performance of using the daily $\Delta\theta_e$ value to forecast warning versus non-warning winds.

	$\Delta\theta_e$ Performance Metrics
Bias	1.194
Accuracy	0.525
POD	0.534
POFA	0.552
CSI	0.322
HSS	0.052
TSS	0.053

Contrary to the results of Loconto (2006) and Atkins and Wakimoto (1991), which found that larger $\Delta\theta_e$ values commonly coincided with stronger microbursts, the above

performance metrics indicate that using the daily $\Delta\theta_e$ value to differentiate between warning and non-warning convective wind speeds is not recommended.

2) COMPOSITE TEMPERATURE, DEW POINT, AND WIND PROFILES

The composite soundings shown in Figs. 14-16 show that there is a negligible difference between the temperature and dew point profiles for days with thunderstorms and weak winds and days with thunderstorms and strong winds. However, as anticipated, the non-convective temperature and dew point profiles appear to be slightly cooler and drier than either of the convective soundings. Instead, it is the wind barbs in each of these soundings that display the greatest amount of variance with a tendency toward a somewhat stronger and more west-southwesterly wind profile with the strong convective wind days, especially below 500 hPa. The wind speed and direction profiles displayed in Figs. 17 and 18, respectively, better illustrate this tendency. Weak convective wind days showed a somewhat fainter westerly wind component below 500 hPa, while the non-convective days actually displayed a bit of an easterly wind component in the low-levels. These results are also similar to the results of the Cummings *et al.* (2007) study.

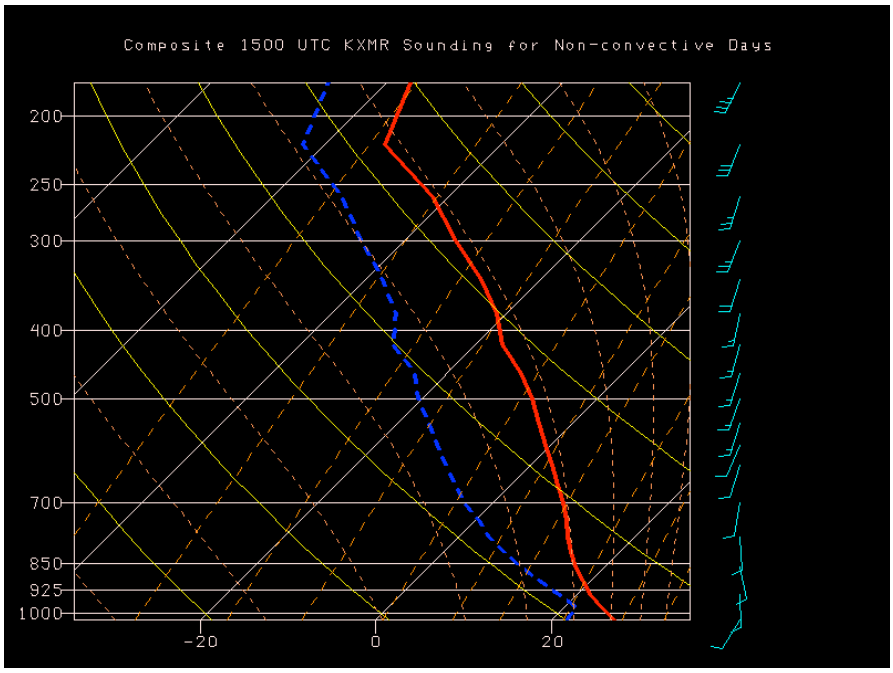

Figure 14. Composite 1500 UTC KXMR temperature (red) and dew point (dashed blue) soundings alongside with wind barbs (knots) for non-convective days. Soundings contain data from 1995 to 2009.

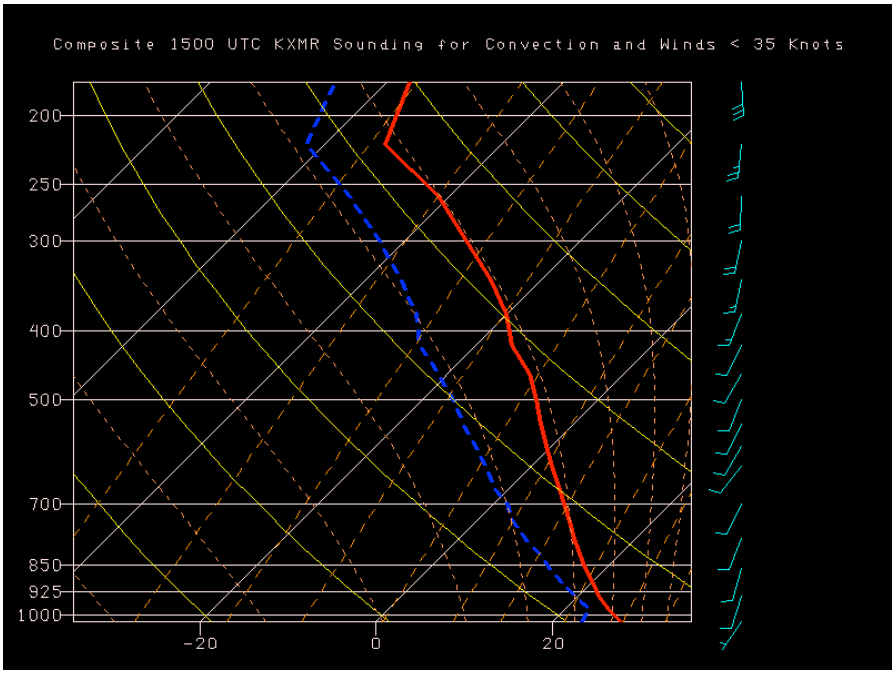

Figure 15. Composite 1500 UTC KXMR temperature (red) and dew point (dashed blue) soundings alongside with wind barbs (knots) for convective days and observed winds less than 35 knots. Soundings contain data from 1995 to 2009.

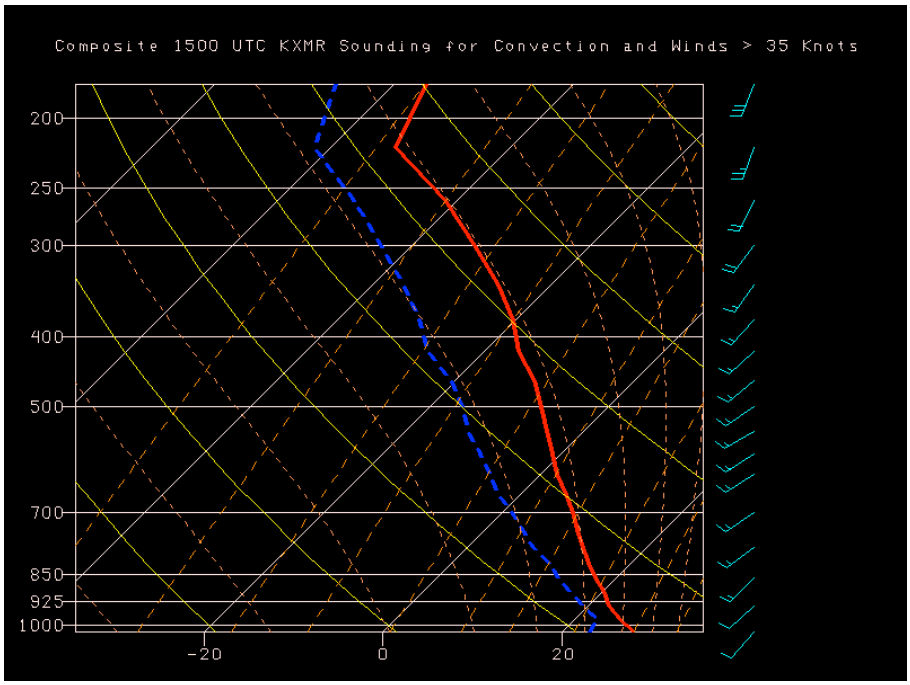

Figure 16. Composite 1500 UTC KXMR temperature (red) and dew point (dashed blue) soundings alongside with wind barbs (knots) for convective days and observed winds greater than 35 knots. Soundings contain data from 1995 to 2009.

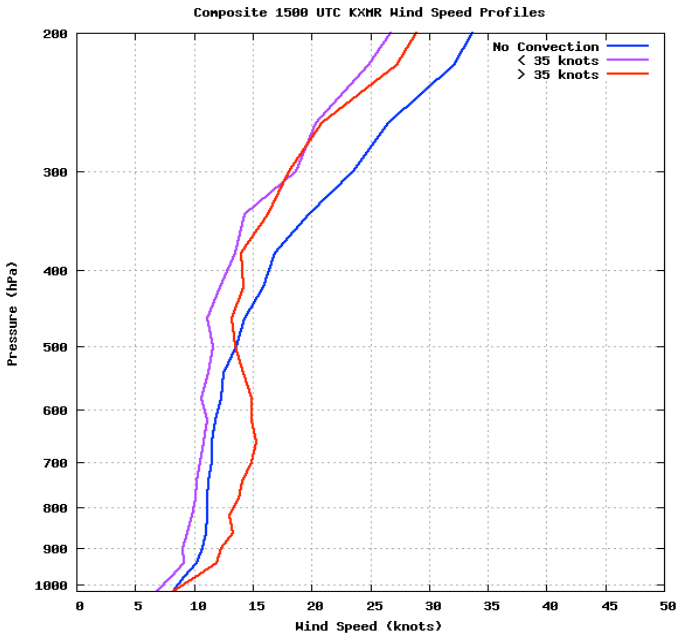

Figure 17. Composite 1500 UTC KXMR wind speed profiles for non-convective days (blue), convective days with winds less than 35 knots (purple), and convective days with winds greater than 35 knots (red).

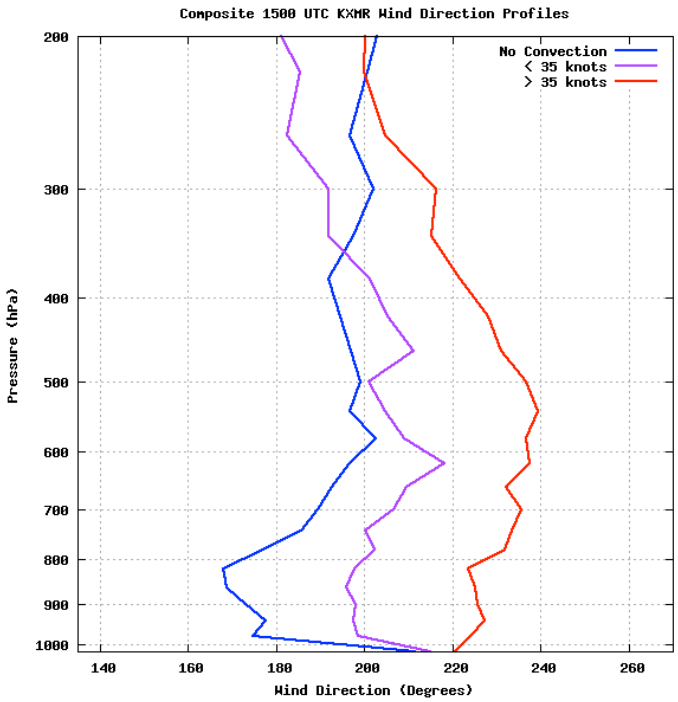

Figure 18. Composite 1500 UTC KXMR wind direction profiles for non-convective days (blue), convective days with winds less than 35 knots (purple), and convective days with winds greater than 35 knots (red).

Despite these findings, it is probably not the southwesterly flow that is causing more or stronger downbursts directly. Rather, the southwesterly flow is a flow regime that corresponds to a greater number of thunderstorms (Lambert and Roeder 2008) and possibly more intense thunderstorms as well. This is due to the southwesterly flow slowing the inland penetration of the east coast sea breeze front off of the Atlantic Ocean and increasing the convergence at the east coast sea breeze front. Greater convergence can, in turn lead to more and stronger thunderstorms over KSC/CCAFS. In addition, the southwest flow accelerates the inland motion of the west coast sea breeze front off of the Gulf of Mexico, which can result in a collision of west and east coast sea breeze fronts on the eastern shore of the Florida peninsula. Outflow boundaries from prior convection can also enhance the forward motion of the eastward moving west coast sea breeze front and

result in even greater convergence as it collides with the east coast sea breeze front. As such, the greater number of thunderstorms and strong thunderstorms can lead to a larger quantity of downbursts and more intense downbursts, especially if due to low-level boundary interactions (Ander *et al.* 2009; Dinon *et al.* 2008). However, the southwest flow regime is a potential predictor for warning level downbursts at CCAFS/KSC, regardless of the physical cause.

Since the low-level wind direction profile displays the greatest stratification between the convective classes, it was quantitatively analyzed in order to find whether or not it could be employed as a forecasting tool. Averaging the wind direction from the surface to 700 hPa for each day and comparing the averaged value with one of two thresholds achieved this. One threshold was determined by calculating the mean surface to 700 hPa wind direction for all of the days in the dataset while the other was computed by averaging the same data from the convective days only. The first wind direction threshold came out to 197.5°; the second was found to be 214.4°. Any daily mean wind direction that was found to be greater (larger westerly component) than the first threshold was predicted as a convective day while any day that had more of a westerly component than the second threshold was predicted as a day where warning level winds would occur. Verification with observed wind data yielded the results in Table 12.

Table 12. Performance metrics of two wind direction thresholds.

	Convective vs. Non-convective	Warning vs. Non-warning
Bias	1.386	1.419
Accuracy	0.621	0.586
POD	0.665	0.704
POFA	0.520	0.504
CSI	0.387	0.410
HSS	0.241	0.196
TSS	0.261	0.209

Although not impressive, Table 12 shows that using the mean low-level wind direction to differentiate between convective and non-convective days yielded better results than using it to do so for warning and non-warning wind days. As such, there does seem to be some limited potential in using it as a criterion with which to diagnose which days are more likely to produce convection, but not necessarily for forecasting which days are more likely to produce warning or non-warning winds. This is the case since convection and perhaps strong downburst winds are favored when the flow displays more of a westerly component.

From a physical standpoint, this makes sense because convection is favored on the east coast of Florida during westerly and southwesterly wind regimes. In addition to the processes mentioned previously, warmer and moister air from the interior of the peninsula will be advected toward the east coast, enhancing the instability and energy available for convection. The slightly greater ambient wind speeds may also reflect the presence of a stronger synoptic flow on warning criteria convective wind days, suggesting that some momentum transfer from the mid-levels may be contributing to increased wind velocities near the surface in these cases. Lastly, a southwesterly synoptic

flow may imply the presence of an approaching frontal boundary or upper level trough, both of which are locations favored for large-scale ascent and thunderstorm formation.

CHAPTER 4

4. Results of new wet downburst forecasting methods

The performance results of new predictive analytic wet downburst forecasting techniques are discussed in this chapter.

a. Formulation and evaluation of multiple linear regression models

After the variable selection techniques discussed in chapter 2 were employed to build simple multiple linear regression (MLR) models in R, each was evaluated to choose the one that best predicted potential wind speeds from among the 61 RAOB derived predictor variables. Recall that the objective of the variable selection methods was to include only the predictor variables that best forecasted wind speed in order to develop the most accurate, simple, and easy to use MLR model possible. In addition, predictor variable elimination was done to remove variables that did not successfully forecast peak microburst wind speeds.

Several variable selection methods were tested against a MLR model with all 61 predictor variables embedded into it in order to assess which variable selection technique performed best. Since scatterplots of each of the predictor variables versus the peak wind speed response variable yielded mostly scatter with no discernable linear or non-linear relationships, it was uncertain which predictor variables would provide the best performing MLR models. The variable selection techniques described in chapter 2, section e, subsection 4 were used in an attempt to find the simplest and best performing MLR models. Verification results for each of the MLR models generated by the variable selection techniques with the 2008 and 2009 independent data are displayed in Table 13.

Table 13. Performance of MLR wind models with all predictors and the simplified versions after variable selection was done. Each of the variable selection techniques in the table represents individual variable selection methods for separate MLR models. Variable selection techniques are explained in chapter 2, section e, subsection 4.

	All Predictors	AIC	P-value	Variable Importance	Correlation	PCA
RMSE	10.82	10.49	9.511	10.02	9.832	12.31
MAE	8.671	8.524	7.910	8.179	7.972	10.26
ME	2.180	2.120	1.524	2.196	1.707	0.526
Hit Rate	0.352	0.324	0.408	0.380	0.437	0.254
Correlation	0.312	0.338	0.403	0.348	0.348	0.222

Table 13 indicates that using several variable selection techniques to simplify MLR models does not provide any significant improvement over the model containing all 61 predictors. In fact, all of the models tested have very similar performance metrics. A weak case could, perhaps, be made that the model found by eliminating non-statistically significant predictors with a chi-square test (removing predictors one by one until all of the remaining predictors were below a threshold p-value) since this model had a slightly lower RMSE and MAE than any of the other models evaluated. This model also has a slightly higher hit rate and correlation coefficient than the other models.

Unfortunately, the MLR models do not show much on the way of improvement over some of the wet downburst forecasting indices discussed previously as indicated by their similarly low correlation coefficients and high error values. Even though their performance is poor, an advantage to MLR models over some of the wet downburst forecasting indices is that they are constructed from data that directly represents the local climatology. Furthermore, they contain many more predictors that can be used to predict convective wind speeds than the wet downburst forecasting indices. On the other hand, the primary disadvantage of MLR models is that they still cannot adequately handle the

amount of chaos involved in forecasting the strength of convectively induced winds, leading to undesirably high forecast errors.

b. Formulation and evaluation of logistic regression models

1) USING LOGISTIC REGRESSION MODELS TO DIFFERENTIATE BETWEEN NON-CONVECTIVE AND CONVECTIVE DAYS

As with MLR, several variable selection techniques were tried in order to find the best performing LR model with the fewest number of predictors in an attempt to better forecast which days are more conducive to convection (LR model type 1) and which convective days may produce a warning threshold wind gust (LR model type 2). The variable selection techniques tested here include the same ones used in evaluating the MLR models with the addition of utilizing the variable importance as found by the boosting algorithm. Again, the models produced by each of the variable selection techniques were evaluated against a model with all 61 predictors included in it to find which method provided the simplest and best performing model. Model evaluation with the independent data and the resulting performance metrics of each of the type 1 LR models found by the variable selection techniques and the LR type 1 model with all of the predictors are displayed in Table 14.

Table 14. Performance of type 1 LR models differentiating between non-convective and convective days. Variable selection techniques are discussed in chapter 2, section e, subsection 4.

	All Predictors	AIC	P-value	Variable Importance: Random Forests	Variable Importance: Boosting	Correla-tion	PCA
Bias	0.690	0.873	0.930	0.930	0.873	0.613	0.761
Accuracy	0.721	0.694	0.713	0.721	0.697	0.710	0.577
POD	0.451	0.493	0.549	0.380	0.507	0.387	0.282
POFA	0.347	0.435	0.409	0.307	0.419	0.370	0.630
CSI	0.364	0.357	0.398	0.325	0.371	0.315	0.190
HSS	0.344	0.302	0.355	0.321	0.316	0.294	0.021
TSS	0.320	0.292	0.349	0.288	0.307	0.267	0.020

Despite overall similar performance results between each of the LR models evaluated, Table 14 signifies that the model constructed by eliminating non-statistically significant predictors, once again, provided the top performing type 1 LR model as indicated by its superior performance metrics. Although the best performing type 1 LR model shows some considerable promise, the overall performance of this model suggests that employing this approach to diagnosing which days have a greater potential to produce convection must be approached with some caution. However, despite less than ideal forecasting ability, it is considerably better than both present forecasting methods and random forecasting.

2) USING LOGISTIC REGRESSION MODELS TO DIFFERENTIATE BETWEEN NON-WARNING AND WARNING CRITERIA WIND DAYS

After using the best performing type 1 LR model to forecast whether a day has potential to produce a convective event, another series of LR models were developed with the same variable selection methods as before in an attempt to find a good LR model that would forecast whether or not a day with convection is likely to produce a warning

or non-warning level wind gust (LR model type 2). The results of all of the simplified type 2 LR models found by using the variable selection techniques and the type 2 LR model with all 61 predictors are summarized in Table 15.

Table 15. Performance of type 2 LR models differentiating between non-warning and warning convective wind days.

	All Predictors	AIC	P-value	Variable Importance: Random Forests	Variable Importance: Boosting	Correla-tion	PCA
Bias	0.886	1.086	1.000	1.029	1.143	1.114	0.886
Accuracy	0.577	0.620	0.606	0.676	0.648	0.662	0.436
POD	0.514	0.657	0.600	0.686	0.714	0.714	0.371
POFA	0.419	0.395	0.400	0.333	0.375	0.359	0.581
CSI	0.375	0.460	0.429	0.511	0.500	0.510	0.245
HSS	0.153	0.240	0.211	0.352	0.297	0.325	-0.129
TSS	0.153	0.241	0.211	0.352	0.298	0.325	-0.128

After evaluating the type 2 LR models with the independent data, performance metrics found that variable selection using the random forests algorithm's variable importance function provided the best results. This is somewhat surprising since this is a different variable selection technique than the one that yielded the best performing type 1 LR model. It is not known why this is the case, and it is suspected that this may have resulted from chance. Although the performance is only marginally useful for operations, the results in Table 15 do suggest that the best performing type 2 LR model does provide some promise in helping 45 WS meteorologists to better diagnose which convective days are more likely to produce a warning level wind velocity. In general, the performance results of both LR model types indicate that translating the response variable into a binary category and removing the noise in the response provides better forecasting results than forecasting the wind speed directly as was done with the MLR models.

c. Formulation and evaluation of multinomial regression models

In order to avoid having to use two LR models, several MR models were developed using three of the above variable selection techniques discussed in chapter 2. Unfortunately, due to a deficiency in the multinomial regression function that prevented the computation of AIC or the predictor variable p-values, the only variable removal methods tested were finding the most highly correlated predictors, variable importance as found by the random forest algorithm, and PCA. The performance of the lower dimension models were, once again, compared to the model with all 61 predictors in order to choose the best performer. Validation with the independent dataset yielded the results displayed in Table 16.

Table 16. Performance of MR model.

	All Predictors	Correlation	Variable Importance	PCA
Accuracy	0.657	0.637	0.676	0.375
HSS	0.236	0.133	0.232	0.023
TSS	0.204	0.107	0.188	0.025
GMSS	0.269	0.123	0.204	0.024

Comparing the results of the models generated by variable selection techniques with a model that contained all of the parameters found that, unlike the two types of LR models that each used different subsets of predictors, the model with all 61 predictors provided the best results. In addition, the performance metrics of the simplified models were all considerably worse than that of the full model. The relatively poor performance of the MR model suggests that the 45 WS should not use this approach for convective wind forecasting. Instead, it is recommended that two LR models be used due to their higher accuracy and better performance with respect to chance forecasting. The reason

for the weak performance of this model is not well understood, but is speculated that different factors lead to determining whether convection will occur on any given day than those factors that determine the strength of a wind gust. In more succinct terms, combining the parameters that predict the outcome of two separate events into a single model is quite possibly the culprit behind the weakness of the MR models.

d. Development and validation of CART models

Since MLR, LR, and MR models displayed only mediocre results at best, CART models were built using the "tree", "rpart", and "party" algorithms. A description of these algorithms was presented in chapter 2, section e, subsection 5. Pruning the trees was accomplished by utilizing the methods discussed in chapter 2 until the best model was found. For brevity, only the results of the top performing pruned tree from each CART algorithm are included. As with the linear regression models, the CART models were constructed with the training dataset and evaluated with the independent dataset. The CART models attempted to forecast the potential downburst wind speed, whether a day will produce convection, or whether a convective day is likely to produce a warning level wind gust. An example regression tree that was grown and pruned in R is displayed in Fig. 19. Any cases that meet the condition go to the left, while cases that do not go to the

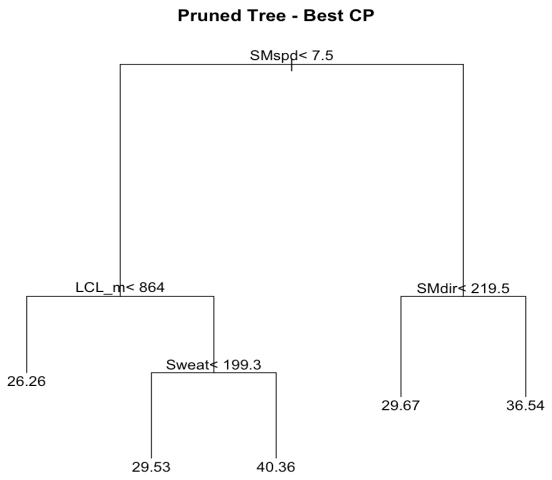

Figure 19. Example regression tree used to predict wind speed. The numbers at the end of each node represent a predicted wind speed in knots. Please consult Table 1 for the acronyms used.

right. This continues down the tree until a terminal node is reached, which, in this instance, provides the forecasted wind gust velocity. Verification statistics of the best performing pruned regression trees found by each algorithm are displayed in Table 17.

Table 17. Performance metrics of the three regression tree algorithms tested.

	"tree"	"rpart"	"party"
RMSE	14.38	10.18	10.09
MAE	11.72	7.942	8.254
ME	4.387	1.145	0.750
Hit Rate	0.280	0.413	0.360
Correlation	-0.108	0.318	0.317

Unfortunately, as these results show, the performance of the three regression tree algorithms is not any better than the best MLR model. In fact, by some measures, they are actually slightly worse. This is likely because the discrete forecasts produced by the regression trees introduce an extensive amount of variance and error. As such, it is not recommended that regression trees be used to forecast convective wind speed.

Using classification trees to make forecasts for convection and approximate wind gust strength yielded more promising results. An example of a classification tree used to forecast whether a day will produce convection (referred to as classification tree type 1) that was grown and trimmed by the "tree" algorithm is displayed in Fig. 20, while an

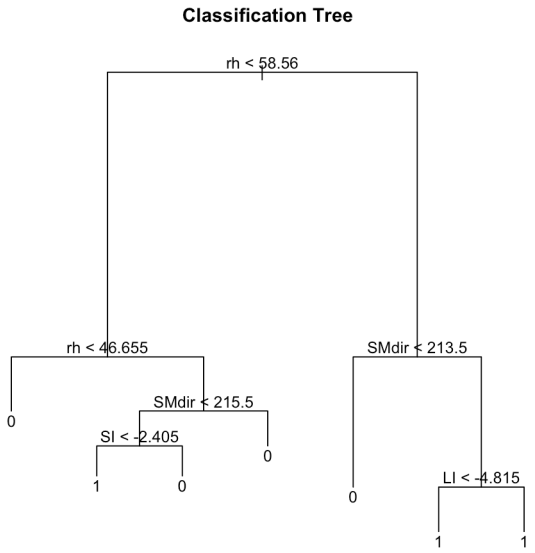

Figure 20. Classification tree used to forecast whether convection would occur on any given day. It was grown and pruned by the "rpart" algorithm. A 0 corresponds to a forecast of no convection, while a 1 corresponds to a forecast of convection. Please consult Table 1 for acronym definitions.

example of a tree used to determine if the wind strength will obtain warning threshold (referred to as classification tree type 2) is displayed in Fig. 21. The tree displayed in Fig. 21 was grown and trimmed with the "rpart" algorithm. Table 18 shows the results of the type 1 classification trees for each CART algorithm tested, while Table 19 does likewise for the type 2 classification trees.

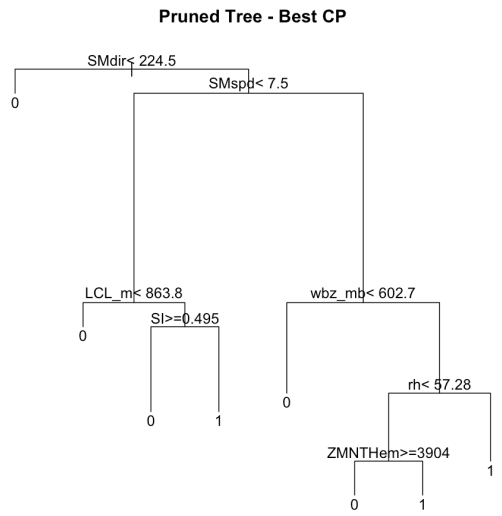

Figure 21. Classification tree used to forecast whether convective winds will reach warning threshold should convection occur on any given day. It was grown and pruned using the "rpart" algorithm. A 0 corresponds to a forecast of non-warning level winds, while a 1 corresponds to a forecast of warning level winds. Please consult Table 1 for acronym definitions.

Table 18. Performance metrics of classification tree type 1 for each CART algorithm. Type 1 predicts whether or not convection will occur.

	"tree" type 1	"rpart" type 1	"party" type 1
Bias	1.141	1.155	1.042
Accuracy	0.701	0.726	0.756
POD	0.648	0.690	0.676
POFA	0.432	0.402	0.351
CSI	0.434	0.471	0.495
HSS	0.367	0.421	0.472
TSS	0.379	0.436	0.476

Table 19. Performance metrics of classification tree type 2 for each CART algorithm. Type 2 predicts whether or not downburst winds will reach or exceed the warning threshold should type 1 forecast convection.

	"tree" type 2	"rpart" type 2	"party" type 2
Bias	0.771	0.743	1.257
Accuracy	0.493	0.676	0.648
POD	0.371	0.543	0.771
POFA	0.519	0.269	0.386
CSI	0.265	0.452	0.519
HSS	-0.018	0.350	0.298
TSS	-0.017	0.348	0.299

For the type 1 classification trees, the "party" algorithm provides the best performing model with a TSS that would actually make it marginally useful for operational forecasting endeavors. However, in the case of the type 2 classification trees, the "rpart" algorithm actually provides the best overall results. Considering the accuracy values in the vicinity of 70% and the otherwise solid performance metrics, it is recommended that both of these classification tree types be used over both linear regression models and regression trees.

Although each classification tree type can be used independently of each other to make forecasts for convection or warning threshold winds, it is recommended that both types be used in conjunction with each other. More specifically, classification tree type 1 should first be used to make a forecast for convection or no convection, and if convection is predicted, then classification tree type 2 should be used to make a forecast for whether or not the convection may produce 35 knot or greater winds. However, if the forecaster wishes to use another forecasting method to predict convection and forego using the type 1 classification tree, then he/she may do so and jump directly to using classification tree type 2 to make a forecast of whether or not the winds may reach the warning threshold.

This is because the two tree types were grown separately from one another and are intended to forecast different phenomena.

Finally, a three-category classification tree was grown in order to compare the feasibility of using this with both of the two-category classification trees and the MR model. The resulting classification tree grown and trimmed using the "rpart" algorithm is illustrated in Fig. 22. Performance metrics for all three CART algorithms are displayed in Table 20.

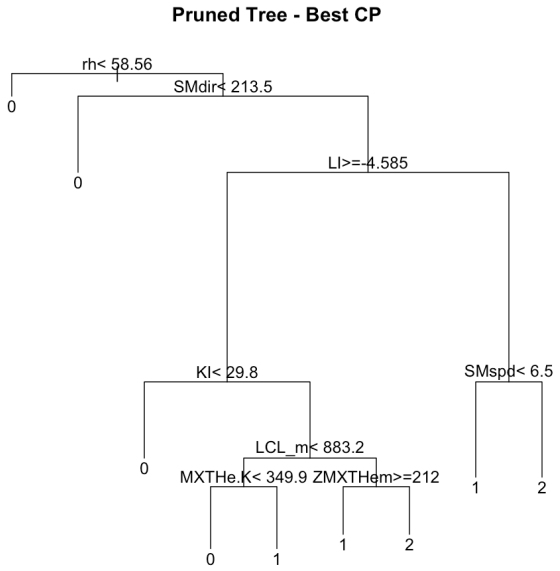

Figure 22. Three-class classification tree used to forecast whether convection will occur on any given day and, if so, whether convective winds will reach warning threshold. It was grown and pruned using the "rpart" algorithm. A 0 corresponds to a forecast of no convection, a 1 corresponds to a forecast of non-warning level winds, and a 2 corresponds to a forecast of warning level winds. Please consult Table 1 for acronym definitions.

Table 20. Performance metrics of three-category classification tree for each of the three CART algorithms.

	"tree"	"rpart"	"party"
Accuracy	0.647	0.652	0.667
HSS	0.142	0.150	0.152
TSS	0.113	0.118	0.115
GMSS	0.172	0.179	0.170

As with the MR model, using a three-class response variable decreased the predictive ability of the classification tree. All three algorithms produced similar results. It is, once again, thought that the weak performance of the three class models may have to do with merging parameters that are used to predict two separate types of events into one. Consequently, it is suggested that 45 WS forecasters use the two types of binary response classification trees to make forecasts as opposed to either the regression trees or three-class classification trees.

e. Construction and validation of ensemble CART models using bootstrapping

Several ensemble CART models using several bootstrapping algorithms were tested with the dependent variable as either a numeric, two-class, or three-class response and were compared to both regression and CART models. Both the random forests and bagging algorithms – recall that boosting can only be used with a binary response due to a limitation in the boosting algorithm – were implemented by growing 500 regression trees to predict the actual peak wind speed in knots since this number of trees provided the best results while simultaneously minimizing computing time. Keep in mind that the performance metrics for these algorithms were computed using data not selected for any of the bootstrap samples, or the out-of-bag (OOB) data. Even with this method of model

verification, the performance of both the bootstrapping models was much better than both MLR models and regression trees, as indicated by Table 21.

Table 21. Out-of-bag (OOB) performance metrics of the bagging and random forests bootstrapping algorithms with a numeric response.

	Bagging	Random Forests
RMSE	9.860	9.908
MAE	7.588	7.655
ME	0.019	0.178
Hit Rate	0.406	0.394
Correlation	0.452	0.444

Both of these algorithms produced similar results with bagging being the slightly better of the two. However, even with the improved performance, there is still too much inconsistency in the predicted wind speeds in order to accurately predict their strengths from the 1500 UTC KXMR RAOB data. As such, it is not advised that ensemble CART models be used to forecast a peak wind gust either.

Treating the response as a two-class variable in the ensemble CART models produced the most promising results by far in this study. Various tree sizes were tested in the boosting and bagging algorithms until the size that produced the highest TSS was found. Within the boosting algorithm, 100 individual 256-split trees were grown since this combination yielded the strongest results for both the non-convection versus convection forecasting model (referred to as model type 1) and the warning versus non-warning wind model (referred to as model type 2) (Fig. 23). For the bagging algorithm, 500 trees were, once again, grown. Unlike the boosting algorithm, however, varying the tree size provided negligible difference in the predictive ability of the models for bagging algorithm (Fig. 24). Since the random forests algorithm grows trees to their maximum

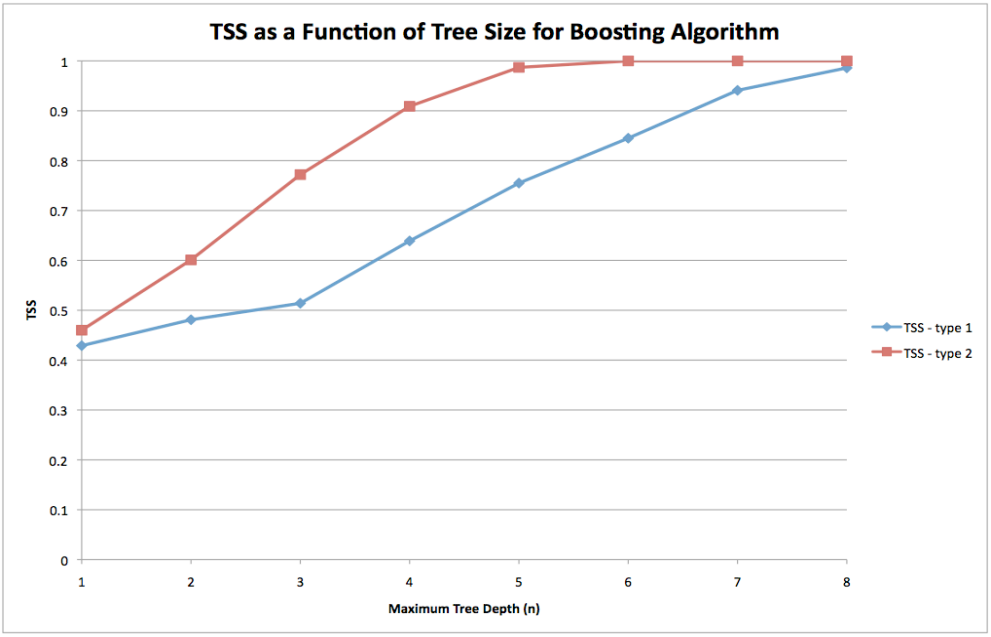

Figure 23. TSS as a function of tree size for the boosting algorithm for both types of models. The number of splits on a tree is simply 2n, where n corresponds to the maximum tree depth.

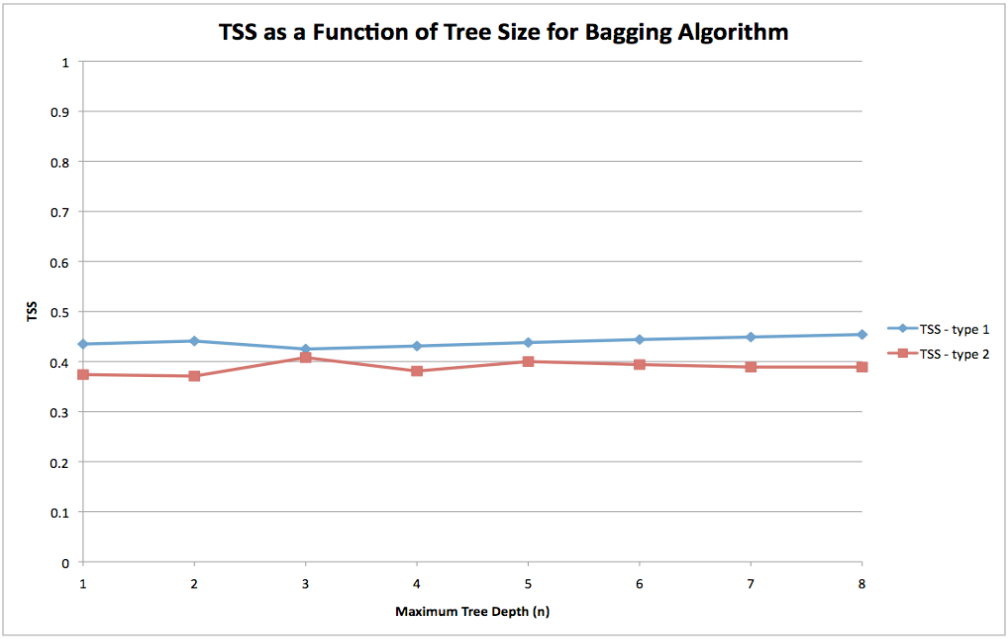

Figure 24. TSS as a function of tree size for the bagging algorithm for both types of models. The number of splits on a tree is simply 2n, where n corresponds to the maximum tree depth.

possible size and does not allow the user to specify tree size, models employing different tree sizes could not be evaluated with this algorithm. As with bagging, 500 trees were grown with this algorithm as well. Of the three algorithms evaluated, boosting had the best performance, with a nearly perfect forecasting track record for the larger tree sizes. Table 22 summarizes the OOB forecasting ability of each the bootstrapping algorithms for model type 1 while Table 23 does likewise for model type 2. In order to add some credibility to the results generated by the boosting algorithm, it was tested on a separate dataset from 2008 and 2009. These results are also displayed in Tables 22 and 23.

Table 22. OOB performance metrics of convection versus non-convection (type 1) bootstrapping models.

	Bagging	Random Forests	Boosting	Boosting 2008-2009 Data
Bias	0.940	0.765	0.995	1.000
Accuracy	0.759	0.764	0.994	1.000
POD	0.608	0.529	0.990	0.995
POFA	0.352	0.308	0.006	0.003
CSI	0.457	0.428	0.983	0.994
HSS	0.449	0.437	0.987	0.997
TSS	0.442	0.411	0.986	0.998

Table 23. OOB performance metrics of warning versus non-warning (type 2) bootstrapping models.

	Bagging	Random Forests	Boosting	Boosting 2008-2009 Data
Bias	1.086	0.818	0.996	1.000
Accuracy	0.704	0.671	0.998	1.000
POD	0.703	0.531	0.996	0.999
POFA	0.352	0.351	0.000	0.002
CSI	0.509	0.413	0.996	0.998
HSS	0.404	0.316	0.997	0.999
TSS	0.408	0.310	0.996	0.999

As Tables 22 and 23 indicate, using the boosting algorithm on a separate dataset from 2008 and 2009 yielded similar performance metrics, adding some trustworthiness to the results. It is surmised that the superior performance of the boosting model is due to the weighting scheme that corrects hard to classify observations. The nearly perfect performance of the boosting models indicates that this is the best forecasting approach to predicting whether or not convection will occur and whether the winds will reach or exceed the 35 knot threshold. This performance is, in some respects, surprisingly high and verification with a new independent dataset would help to further validate the success of this model.

Lastly, two three-class bootstrapping models were built with the bagging and random forest algorithms. Both algorithms grew 500 trees. The performance results of both algorithms are illustrated in Table 24.

Table 24. OOB performance metrics of two three-class bootstrapping models.

	Bagging	Random Forests
Accuracy	0.712	0.703
HSS	0.352	0.259
TSS	0.312	0.208
GMSS	0.376	0.244

Although better than both three-class classification trees and MR models, these results do not show much promise for these forecasting methods. To that end, it is not recommended that the 45 WS meteorologists utilize any of the three-class response models examined in this study.

CHAPTER 5

5. Summary, conclusions, and future work

The principle objective of this study was to develop ways of improving warm-season convective wind forecasting on central Florida's Space Coast using a 15 year climatology (1995-2009) of 1500 UTC RAOB data from the CCAFS Skid Strip and 5 minute averaged peak wind from the 36 weather towers selected from the network. This was done by first evaluating present wet microburst forecasting indices, two of which were then modified based on their weaknesses. Composite θ_e soundings were also constructed in order to see if the vertical θ_e gradient could be used as a convective wind forecasting tool. Likewise, composite soundings of temperature and dew point were built alongside with wind profiles in an attempt to see if any of these parameters could be used. In addition, new predictive analytic techniques such as MLR, LR, MR, CART, and ensemble CART models using bootstrapping were used to formulate new statistical forecasting models.

Evaluation of numerous existing wet microburst forecasting indices found that none of them did well in predicting peak convective wind in the central Florida warm-season environment. Although Proctor's index was the best performer among the present indices, it still did not do well. Modification of this index to include a better representation of the low-level moisture profile yielded modest improvements. Since the heavily used WINDEX seriously over predicted wind speed, it was tailored to better match the exceedingly high mixing ratios typically found in the dataset. Even though this provided better results, it is not recommended that any of these indices be used to forecast the intensity of convective winds since they all had poor skill at producing convective

wind forecasts. In addition, they are likely too simple to capture the complex near turbulent scale of dynamics of downbursts.

Construction of composite θ_e profiles found that the mid-levels were generally colder and drier on days with warning level winds indicating that a greater vertical θ_e gradient implies a higher likelihood for stronger winds. However, quantitative validation of this suggested that it should not be used as a forecasting tool. Meanwhile, it was found from the wind direction profiles that convective days had a tendency to occur on days with more of a westerly or southwesterly wind regime. Again, quantitative verification found that using low-level wind direction to make convective wind forecasts also had limited performance.

Testing and verification of several linear regression techniques found that MLR models did not do a particularly good job at forecasting peak wind gusts. On the other hand, LR models did a better job at forecasting both convective days and whether the convective days had a potential to produce warning level wind speeds. Performance declined with MR models, indicating that using two LR models is the best way to predict convective winds.

In general, CART methods yielded similar results to the linear regression models. Regression trees performed somewhat worse than MLR models while two-class classification trees did a bit better than LR models. In a manner that is consistent with the performance of the MR models, using a three-category response in a classification tree produced rather disappointing results. Once again, it is suggested that using a binary response classification tree is the best of the approaches considered in this research.

It was found that ensemble CART forecasting methods using bootstrapping algorithms yielded the best results of those studied. Random forests and bagging produced mediocre results for predicting peak wind gust, but improved considerably when using it to forecast for a two-class response. The boosting algorithm had the best performance metrics by far in this study, indicating that it is probably the best statistical model to use of those investigated in this research.

The results of the boosting algorithm answer the central scientific question posed in chapter 1 affirmatively. Recall that this question asked how could wet downbursts forecasts be improved given the poor correlation between the observed peak wind speed and each of the RAOB derived predictor variables? Using the boosting ensemble CART algorithm for wet downburst forecasting shows considerable promise and appears to make a solid attempt at tackling what appears to be a highly challenging forecasting problem. However, despite these exciting results, it is urged that caution be used in forecasting convective winds in the central Florida warm-season environment since it is far too complex of a problem to undertake with just RAOB data. This is because many factors determining the scope and intensity of thunderstorm induced winds are simply not resolvable with this type of data since downburst intensity, location, and duration depends on many microscale processes that cannot be detected with this type of data. Furthermore, other evidence suggests that local low-level boundary interactions (Ander *et al.* 2009; Dinon *et al.* 2008) play a significant role in downburst formation at KSC/CCAFS. A single location RAOB will not be able to detect such low-level boundaries and use them in local downburst prediction. However, RAOB techniques are

meant to be used as general outlook techniques (Fig. 1) and are not intended to provide detailed warnings.

Finally, the following suggestions are ways to extend, improve, and use this research in future work. Since the boosting ensemble CART model showed surprising forecast skill, it should be verified again with new independent data to verify that the performance is repeatable. In addition, the conversion of ensemble CART models to probability forecasts should be explored based on the percent of total forecasts. For example, if 375 out of 500 CART forecasts predicted that warning level downbursts would occur, the overall forecast might correspond to a 75% probability forecast. This simplest "percent of votes" should be considered first before looking for more complex conversions of number of votes to probability forecast. The performance of this "percent of votes" should be verified with traditional probability forecast techniques. In particular, a reliability diagram or an attributes diagram would indicate any systematic bias and suggest if another conversion of votes to probability would be more appropriate.

In order to implement ensemble CART models into operational forecasting, a script will first need to automatically compute values for each of the 61 parameters in Table 1 from the day's 1500 UTC KXMR RAOB data. R can then be invoked automatically to build the trees from the 15-year KXMR RAOB climatology training dataset. The data for the present day can be run through the trees as independent data to produce a forecast, which can be outputted as a text file that can be displayed either on a website or opened with any text viewer. However, it is advised that this process be tested before being placed into operational use since it is not known how well the ensemble CART algorithms will do in a real-world environment.

Also, since RAOB-based techniques tend to be broad area tools, extending the area of verification beyond KSC/CCAFS to include much of central Florida should be considered since the weather towers may not detect downbursts occurring in central Florida. Storm reports and surface observations from locations throughout central Florida could also be used to increase the verification area to be more representative of RAOB forecast tools.

In addition, a similar study of verifying current techniques and developing new techniques should be done for forecasting downbursts with peaks winds that reach or exceed 50 knots. Even though current methods do not perform well for predicting downbursts that reach or exceed the 35 knot threshold, perhaps they would work better for stronger 50 knot or greater downbursts, especially since many of those techniques were developed for predicting severe straight line winds in excess of 50 knots. Although many of the new techniques did not work well for forecasting 35 knot or greater downbursts, it is possible that they may work better for predicting the stronger downbursts. Even the new techniques that did work well for forecasting the 35 knot or greater downbursts will likely require tuning to optimize performance for predicting the stronger downbursts.

It is also thought that using an ensemble of the regression, CART, and ensemble CART models to make forecasts may provide better results than many of the results discussed. Finally, it is recommended that many of the above techniques be tested with Geostationary Operational Environmental Satellite (GOES) sounder data to see how using GOES data compares with the 1500 UTC RAOB data since, unlike the RAOB data, it is available on an hourly basis.

REFERENCES

Akaike, H., 1974: A new look at the statistical model identification. *IEEE Transaction on Automatic Control*, **19**, 716-723.

Ander, C.J., A.J. Frumkin, J.P. Koermer, and W.P. Roeder, 2009: Study of sea-breeze interactions which can produce strong warm-season convective in the Cape Canaveral area. *16th Conf. on Air-Sea Interaction/8th Conf. on Coastal Atmospheric and Oceanic Prediction and Processes*, Amer. Met. Soc., Phoenix, AZ, J8.3.

Atkins, N.T., and R.M. Wakimoto, 1991: Wet microburst activity over the southeastern United States: Implications for forecasting. *Wea. Forecasting*, **6**, 470-482.

Atlas, D., C.W. Ulbrich, and C.R. Williams, 2004: Physical origins of a wet microburst: Observations and theory. *J. Atmos. Sci.*, **61**, 1186-1196.

Barnes, L.R., D.M. Schultz, E.C. Gruntfest, M.H. Hayden, and C.C. Benight, 2009: CORRIGENDUM: False alarm rate or false alarm ratio? *Wea. Forecasting*, **24**, 1452-1454.

Breiman, L., 1996: Bagging predictors. *Machine Learning*, **24**, 123-140.

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

____, J. H. Friedman, R. A. Olshen, and C. G. Stone, 1984: Classification and Regression Trees. Chapman and Hall, 368 pp.

Case, J. L., and W. H. Bauman III, 2004: A Meso-climatology study of the high-resolution tower network over the Florida Spaceport. *11th Conference on Aviation, Range, and Aerospace Meteorology*, Amer. Met. Soc., Hyannis, MA, P7.6.

Culp, M.V., G. Michailidis, and K. Johnson, 2006: ada: An R package for stochastic boosting. *Journal of Statistical Software*, **17** (2).

Cummings, K.A., E.J. Dupont, A.N. Loconto, J.P. Koermer, and W.P. Roeder, 2007: An updated warm-season convective wind climatology for the Florida Space Coast. *16th Conf. of Applied Climatology*, Amer. Met. Soc., San Antonio, TX, J3.13.

Dinon, H.A., M.J. Morin, J.P. Koermer, and W.P. Roeder, 2008: Convective winds at the Florida Spaceport: year-3 of Plymouth State research. *13th Conf. on Aviation, Range, and Aerospace Meteorology*, Amer. Met. Soc., New Orleans, LA, J8.5.

Freund, Y., and R. E. Schapire, 1996: Experiments with a new boosting algorithm. Pages 148–156 in Machine learning: proceedings of the thirteenth international conference, San Francisco, CA.

Fujita, T.T., 1981: Tornadoes and downbursts in the context of generalized planetary scales. *J. Atmos. Sci.*, **38**, 1541-1557.

____, and R.M. Wakimoto, 1981: Five scales of airflow associated with a series of downbursts on 16 July 1980. *Mon. Wea. Rev.*, **109**, 1438-1456.

____, and ____, 1983: Microbursts in JAWS depicted by Doppler radars, PAM and aerial photographs. Preprints, *21st Conf. Radar Meteorology*, Edmonton, AB, Canada, Amer. Met. Soc., 19-23.

Gandin, L.S., and A.H. Murphy, 1992: Equitable skill scores for categorical forecasts. *Mon. Wea. Rev.*, **120**, 361-370.

Harms, D. E., A. A. Guiffrida, B. F. Boyd, L. H. Gross, G. D. Strohm, R. M. Lucci, J. W. Weems, E. D. Priselac, K. Lammers, H. C. Herring and F. J. Merceret, 1999: The many lives of a meteorologist in support of space launch, Preprints, *8th Conf. On Aviation, Range, and Aerospace Meteorology*, Dallas, TX, 5-9.

Hothorn, T., K. Hornik, and A. Zeileis, 2006: party: A laboratory for recursive partitioning. [Available online at http://cran.r-project.org/web/packages/party/vignettes/ party.pdf.]

Koermer, J.P., cited 2009: CCAFS/KSC warm-season convective wind climatology. [Available online at http://vortex.plymouth.edu/conv_winds.]

Kuchera, E. L. and M.D. Parker, 2006: Severe convective wind environments. *Wea. Forecasting*, **21**, 595-612.

Lambert, W., and W.P. Roeder, 2008: Update to the lightning probability forecast equations at Kennedy Space Center/Cape Canaveral Air Force Station, Florida. *2nd International Lightning Meteorology Conference*, 24-25 Apr. 2008, 16 pp.

Liaw, A., and M. Wiener, cited 2009: Package "randomForest". [Available online at http://cran.r-project.org/web/packages/randomForest/randomForest.pdf.]

Loconto, A.N., 2006: Improvements of warm season convective wind forecasts at the Kennedy Space Center and Cape Canaveral Air Force Station. M.S. thesis, Plymouth State University, 93 pp.

McCann, D.W., 1994: WINDEX – A new index for forecasting microburst potential. *Wea. Forecasting*, **9**, 532-541.

Peters, A., T. Hothorn, and B. Lausen, 2002: ipred: Improved predictors. R News, 2, 33-36. [Available online at http://cran.r-project.org/doc/Rnews_2002-2.pdf.]

Proctor, F.H., 1989: Numerical simulations of an isolated microburst. Part II: Sensitivity experiments. *J. Atmos. Sci.*, **46**, 2143-2165.

Pryor, K.L., 2005: Forecasting convective downburst potential over the United States Great Plains. [Available online at http://arxiv.org/html/physics/0511245v1.]

____, and G.P. Ellrod, 2004: WMSI – A new index for forecasting wet microburst severity. *National Weather Association Electronic Journal of Operational Meteorology*, 2004-EJ3. [Available online at http://www.nwas.org/ej/pdf/2004-EJ3.pdf.]

R Development Core Team, 2009: R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. [Available online at http://www.r-project.org]

Rennie, J. J., J.P. Koermer, T.R. Boucher, and W.P. Roeder, 2010: Evaluation of WSR-88D methods to predict warm-season convective wind events at Cape Canaveral Air Force Station and Kennedy Space Center. *22nd Conf. Climate Variability and Change*, Amer. Met. Soc., Atlanta, GA, P2.17.

Ripley, B., 2009: Package "tree". [Available online http://cran.r-project.org/web/packages/tree/tree.pdf.]

Roeder, W.P., 2009: Plymouth State University convective winds at CCAFS/KSC research program in brief. PowerPoint Presentation. 45th Weather Squadron, Cape Canaveral Air Force Station, Florida.

Srivastava, R.C., 1985: A simple model of evaporatively driven downdraft: Application to microburst downdraft. *J. Atmos. Sci.*, **42**, 1004-1023.

____, 1987: A model of intense downdrafts driven by melting and evaporation of precipitation. *J. Atmos. Sci.*, **44**, 1752-1773.

Sullivan, G.D., 1999: Using WSR-88D to Forecast Convective Winds at the Kennedy Space Center and Cape Canaveral Air Station. M.S. thesis, Air Force Institute of Technology. 97 pp.

Therneau, T.M. and B. Atkinson, 2010: Package "rpart". [Available online at http://cran.r-project.org/web/packages/rpart/rpart.pdf.]

Wakimoto, R.M., 1985: Forecasting dry microburst activity over the High Plains. *Mon. Wea. Rev.*, **113**, 1131-1143.

Wheeler, M., 1996: Verification and implementation of Microburst Day Potential Index (MDPI) and Wind Index (WINDEX) forecasting tools at Cape Canaveral Air Force Station. NASA Contractor Report CR-201354, 24 pp.

_____, and W.P. Roeder, 1996: Forecasting wet microbursts on the central Florida Atlantic Coast in support of the United States Space Program. Preprints, *18th Conf. on Severe Local Storms*. Amer. Met. Soc., San Francisco, CA, 654-658.

Wilks, D.S., 2006: Statistical Methods in the Atmospheric Sciences. 2nd ed. Elsevier Academic Press Publications, 627 pp.

Wolfson, M.M, R.L. Delanoy, B.E. Forman, R.G. Hallowell, M.L. Pawlak, and D.P. Smith, 1994: Automated microburst wind-shear prediction. *Lincoln Lab. J.*, **7**, 399-426.