Upgrade Summer Severe Weather Tool

Leela Watson
Applied Meteorology Unit
Kennedy Space Center, Florida

April 2011

NASA STI Program ... in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NASA Aeronautics and Space Database and its public interface, the NASA Technical Report Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include creating custom thesauri, building customized databases, and organizing and publishing research results.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at *http://www.sti.nasa.gov*

- E-mail your question via the Internet to help@sti.nasa.gov

- Fax your question to the NASA STI Help Desk at (301) 621-0134

- Phone the NASA STI Help Desk at (301) 621-0390

- Write to:
NASA STI Help Desk
NASA Center for AeroSpace Information
7121 Standard Drive
Hanover, MD 21076-1320
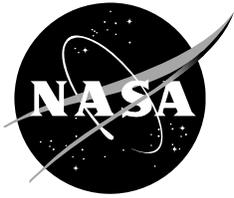
Upgrade Summer Severe Weather Tool Phase III

Leela Watson
Applied Meteorology Unit
Kennedy Space Center, Florida

April 2011

Acknowledgements

Executive Summary

The 45th Weather Squadron (45 WS) Commander's morning weather briefing includes an assessment of the likelihood of local convective severe weather for the day. This forecast is provided in order to enhance protection of personnel and material assets of the 45th Space Wing, Cape Canaveral Air Force Station (CCAFS), and Kennedy Space Center (KSC). The severe weather elements produced by thunderstorms include tornadoes, strong surface winds and/or large hail. Forecasting the occurrence and timing of these phenomena during the warm season (May-September) is challenging for 45 WS operational personnel.

In the first phase of the task, the Applied Meteorology Unit (AMU) analyzed stability parameters and synoptic patterns from east-central Florida severe weather days during the warm season in the years 1989-2003 to determine which were important to severe weather development. A HyperText Markup Language (HTML)-based tool was created that helped determine the probability of issuing severe weather watches and warnings for the day by assigning weights to the important parameters and patterns based on their threat value. A Meteorological Interactive Data Display System (MIDDS)-based Graphical User Interface (GUI) replaced the HTML tool in a follow-on task. The new tool retrieved stability parameters and other information from MIDDS automatically, minimizing the forecaster's interaction with the tool. Later, the AMU updated the severe weather database with data from the years 2004-2009, re-analyzed the data to determine the important parameters, made appropriate adjustments to the index weights depending on the results of the analysis, and updated the MIDDS GUI.

For this task, the 45 WS requested the AMU upgrade the severe weather database by adding weather observations from the 2010 warm season, update the verification dataset with results from the summer of 2010, use statistical logistic regression analysis on the database and develop a new forecast tool, and update the MIDDS GUI with the new tool if it outperforms the current tool. The added data increased the period of record (POR) from 21 to 22 years. With this update, the datasets included reported severe weather events, sounding stability parameters, and surface weather patterns and the upper jet patterns identified from surface and upper air maps.

The AMU analyzed seven stability parameters that showed the possibility of providing guidance in forecasting the occurrence of severe weather for the 2010 season, calculated verification statistics for the Total Threat Score (TTS) in the 2010 season, and calculated warm season verification statistics. Analysis of the seven stability parameters indicate that adding the 2010 data had little effect on the tool's overall severe weather predicting capability. On days that severe weather was reported, the TTS ranged from -11 to +20 compared to the 2006 summer season verification in which the TTS was never below 0 on days when severe weather was reported. Finally, the Severe Weather Worksheet TTS did not verify well in the 2010 warm season, with a high False Alarm Rate (FAR) and low values for Probability of Detection (POD), Critical Success Index (CSI), and Heidke Skill Score (HSS).

The AMU also performed statistical logistic regression analysis on the 22-year severe weather database. The candidate predictors included the flow regimes from the Florida rawinsondes, the placement of the upper-level jet, and seven stability parameters calculated from the XMR rawinsonde. The data were stratified into equation development and verification datasets and one equation for the warm season was developed. A process called screening regression was used to determine which candidate predictors to include in the equation in which an iterative technique was used to test each predictor's ability to explain the variance in the predictand individually and in combination with other predictors. Four predictors were chosen for the warm season logistic regression equation.

Four equation performance tests were conducted. The results indicated that the logistic regression equation did not show an increase in skill over the previously developed TTS. The equation showed less accuracy than the TTS at predicting severe weather, little ability to distinguish between severe and non-severe weather days, and worse standard categorical accuracy measures and skill scores over TTS. The results showed that the equation had some skill in predicting non-severe events, but no skill in predicting severe events.

Based on the findings of this study and after reviewing the results with the 45 WS, a new tool was not developed based on the performance of the logistic regression equation. The previously developed TTS and MIDDS GUI were not updated due to the inability of the 2010 severe weather season data to help improve the tool.

Table of Contents

List of Figures

List of Tables

1.   Introduction

The 45th Weather Squadron (45 WS) Commander's morning weather briefing includes an assessment of the likelihood of local convective severe weather for the day. This forecast is provided in order to enhance protection of personnel and material assets of the 45th Space Wing, Cape Canaveral Air Force Station (CCAFS), and Kennedy Space Center (KSC). The severe weather elements produced by thunderstorms include tornadoes, convective surface winds $\geq$ 50 knots, and/or hail with a diameter $\geq$ 1 inch. Forecasting the occurrence and timing of these phenomena during the warm season (May – September) is challenging for 45 WS operational personnel.

In the first phase of the task, the Applied Meteorology Unit (AMU) analyzed stability parameters and synoptic patterns from east-central Florida severe weather days during the warm season in the years 1989-2003 to determine which were important to severe weather development. The AMU then created a HyperText Markup Language (HTML)-based tool using the important parameters and patterns to help determine the probability of issuing severe weather watches and warnings for the day. The HTML tool was replaced with a Meteorological Interactive Data Display System (MIDDS)-based Graphical User Interface (GUI) in a follow-on task that retrieved stability parameters and other information from MIDDS automatically, minimizing the forecaster's interaction with the tool. Later, the AMU updated the severe weather database with data from the years 2004-2009, re-analyzed the data to determine the important parameters, made appropriate adjustments to the index weights based on the results of the analysis, and updated the MIDDS GUI.

For this task, the 45 WS requested the AMU to:

- Add severe weather reports and indices for the warm season May-September 2010 to increase the period of record to 22 years,

- Use the daily severe weather forecast threat scores from 2009 and 2010 as the verification data of the tool,

- Use logistic regression to determine the best predictors and provide a probability forecast, and then compare the performance of the logistic regression equations with the previous tool, and

- Update the MIDDS GUI implementation of the tool with the new results if the logistic regression equations are successful.

2.   Previous Work

In the initial Severe Weather Forecast Decision Aid task (Bauman et al. 2005), the AMU completed a 15-year climatological study of severe weather events and related severe weather atmospheric parameters. The period of record (POR) for the analysis was May-September, 1989-2003. Data sources included stability parameters derived from archived sounding data, local forecast rules used to set threat assessment thresholds, Cloud-to-Ground Lightning Surveillance System (CGLSS) used to differentiate between lightning and non-lightning days, surface and upper air maps, and two severe weather event databases covering east-central Florida used to identify reported severe weather. These datasets provided the foundation for analyzing stability parameters and synoptic patterns with the goal of developing an objective tool to aid in forecasting severe weather events. Based on the results from the analyses, an interactive web-based Severe Weather Decision Aid was developed to assist the duty forecaster by providing a level of objective guidance based on the stability parameters from the CCAFS morning sounding, CGLSS data, and synoptic-scale dynamics. The tool outputs the Total Threat Score (TTS), which is a measure of the level of severe weather threat. The higher (lower) the TTS, the greater (lesser) the chance of severe weather occurring.

In a follow-on study (Wheeler 2009), the functionality of the web-based tool was migrated to MIDDS. A MIDDS GUI worksheet was created using Tool Command Language and its associated Tool Kit (Tcl/Tk). The GUI retrieves and calculates most of the daily sounding stability indices needed by the worksheet when opened. The forecaster is required to answer a few more subjective questions before the TTS is calculated and displayed.

In the final follow-on study (Wheeler 2010), the AMU updated the existing severe weather tool by adding data from May-September, 2004-2009, creating a 21-year severe weather database. Data sources included local forecast rules, archived sounding data, surface and upper air maps, and two severe weather event databases covering east-central Florida. The new POR for the analysis was May-September, 1989-2009. Results from this study showed a

greater ability to predict severe weather in the added years as compared to the original study. The MIDDS GUI was also updated and mouse-over help was added to allow the forecaster to quickly compute and analyze the daily TTS.

3.   Database

For this work, the AMU updated the severe weather database with data from the 2010 warm season, increasing from a 21- to a 22-year climatological study of atmospheric stability indices and severe events from 1989-2010. To be consistent with previous work, the AMU collected the same data types and parameters used to update the severe weather database. Severe weather reports during 2010 were collected from the Storm Prediction Center (2009, SPC) and data from severe weather days in that period from the National Climatic Data Center (2010, NCDC) database. The sounding stability parameters were calculated from the 1000 UTC CCAFS soundings available from the National Oceanic and Atmospheric Administration's (NOAA)/Earth System Research Laboratory (ESRL) and from the objective portion of the daily Severe Weather Worksheet in MIDDS. Also, the 200 mb charts were analyzed to identify the placement and characteristics of the upper-level jet. With this update, the datasets included reported severe weather events, sounding stability parameters, and surface weather patterns and the upper jet patterns identified from surface and upper air maps. Each data type proved to have some relevance to forecasting the threat of convection in east-central Florida and at KSC/CCAFS.

3.1   Severe Weather Events

Severe weather events for this study included tornadoes, convective surface winds ≥ 50 knots (≥ 26 m s^{-1}), and/or hail with a diameter ≥ 1 inch (≥ 2.54 cm) that were observed in east-central Florida. The previous studies used a $\geq \frac{3}{4}$ inch diameter criterion for hail, however, the National Weather Service changed the minimum size for severe hail from ¾" to 1" in January 2010. Therefore, the POR prior to the 2010 summer season reflected the ¾" criterion and the new 1" hail criterion was used for the 2010 summer season. The 2010 database contained 16 days with reported severe weather, which included three tornadoes, six hail events, and 15 high wind events.

It is important to note that the database contains only those severe weather events reported by human observers. Severe weather events can only be recorded when observed by people in the vicinity, and then only if the proper authorities are notified. Therefore, severe weather days are more accurately described as "reported" severe weather days. To determine relationships between the data and severe weather occurrence for this and previous studies, the AMU had to assume that severe weather only occurred on reported severe weather days.

3.2   Sounding Parameters

A thorough analysis of atmospheric stability based on a local upper air sounding is needed for any convective forecast. A listing of the sounding stability indices (bold) and additional calculated parameters from MIDDS used in the TTS calculation is shown in Table 1. These sounding parameters are calculated in MIDDS from the CCAFS rawinsonde and are readily available.

3.3   Synoptic Weather Patterns

The synoptic weather patterns investigated by the AMU included the position of the upper-level jet streak if one existed and the position of the surface high pressure ridge axis over east-central Florida. It is commonly known that upper-level divergence and/or the left-exit and, to a lesser degree, right-entrance quadrant of a jet streak in the vicinity of convective systems can help produce severe weather. The 45 WS forecasters often analyze the position of the surface high pressure ridge axis protruding westward from the Bermuda high pressure center as an indicator for convection occurrence. It is generally known that if the surface ridge is south of the KSC/CCAFS area the probability for convection is increased due to the low-level convergence generated from the southwesterly flow around the ridge interacting with the east coast sea breeze off the Atlantic Ocean and the west coast sea breeze off the Gulf of Mexico.

Table 1. The eight stability parameters (in bold font) that showed the possibility of providing guidance in forecasting severe weather and the six other sounding parameters in the severe weather database and the equations used in their calculation.

Index Acronym	Definition
LI	Lifted Index = $(T_{500} - T^*)$ T^* = Temperature of a parcel characterized by the mean T_d in the lowest 3000 ft and the forecast maximum surface temperature if it were lifted dry adiabatically to saturation and then moist adiabatically to 500 mb.
KI	K Index = $(T_{850} - T_{500}) + T_{d850} - (T_{700} - T_{d700})$
TT	Total Totals = $(T_{850} - T_{500}) + (T_{d850} - T_{500})$
SSI	Showalter Stability = Index $(T_{500} - T^*)$ T^* = Temperature a parcel characterized by the T_{850} and T_{d850} would have if it were lifted dry-adiabatically to the LCL and then moist-adiabatically to 500 mb.
CT	Cross Totals = $(T_{d850} - T_{500})$
TI	Thompson Index = KI – LI
PW	Precipitable water in mm in the layer from the surface to 500 mb
CAPE FMaxT	CAPE calculated using the forecast maximum temperature (FMaxT) for the day instead of the surface temperature in the morning
10070RH	Average Relative Humidity in percent (%) from 1000-700 mb
LLJ	Low Level Jet below 5000 ft (Wind direction and speed)
INV	Height of Inversion below 8000 ft
T850	The sounding temperature at 850 mb
TDif	The difference between forecast maximum and convective temperatures
MDPI	Microburst Day Potential Index

4. Data Analysis Results

The AMU gathered severe weather reports for 2010 from SPC and data for those severe weather days from NCDC. The 200 mb charts were analyzed to identify placement and characteristics of any jet streaks overhead. The Florida flow regime patterns that identified the position of the surface high pressure ridge axis over east-central Florida were also added to the severe weather database. The datasets were integrated and compared to the severe weather reports of hail, high wind, and tornadoes to determine what the parameter values were on each of the severe weather event days.

The AMU analyzed seven of the eight stability parameters that showed the possibility of providing guidance in forecasting the occurrence of severe weather in the first phase of this task (Bauman et al. 2005). The parameter CAPE FMaxT was not calculated for the years 2004-2009 and was therefore not analyzed in this phase of the task. The parameters TT, KI, LI, TI, CT, SSI and PW were analyzed to determine if they increased the severe weather forecast capability of the tool in the 2010 data and in all 22 years (1989-2010) combined. Results indicate that adding the 2010 data had no effect on the forecast capability of the tool for TT, KI, PW, SSI, and CT. The forecast capability decreased when adding the 2010 LI and TI data. Overall, there was minimal change in the tool's overall severe weather predicting capability. The relationship between each stability parameter and threshold criteria for the severe weather threat was calculated for severe and non-severe days. The results for each of these seven parameters are detailed below.

4.1 Total Totals (TT)

The TT thresholds specify a low threat for severe weather when TT ≤ 45, a medium threat when 46 ≤ TT ≤ 48, and a high threat when TT > 48. When TT was > 48, a severe weather event was reported in 13% of the 2010 warm season days. This slightly decreased the 21-year value of 34%. The 22-year value decreased to 33%. Figure 1 displays the threat levels of Low, Medium and High with the occurrence/non-occurrence of severe weather for the 22-year POR, while Figure 2 displays the same data, but shows the individual TT values for each day. It is evident that TT poorly discerns severe from non-severe weather for our POR.
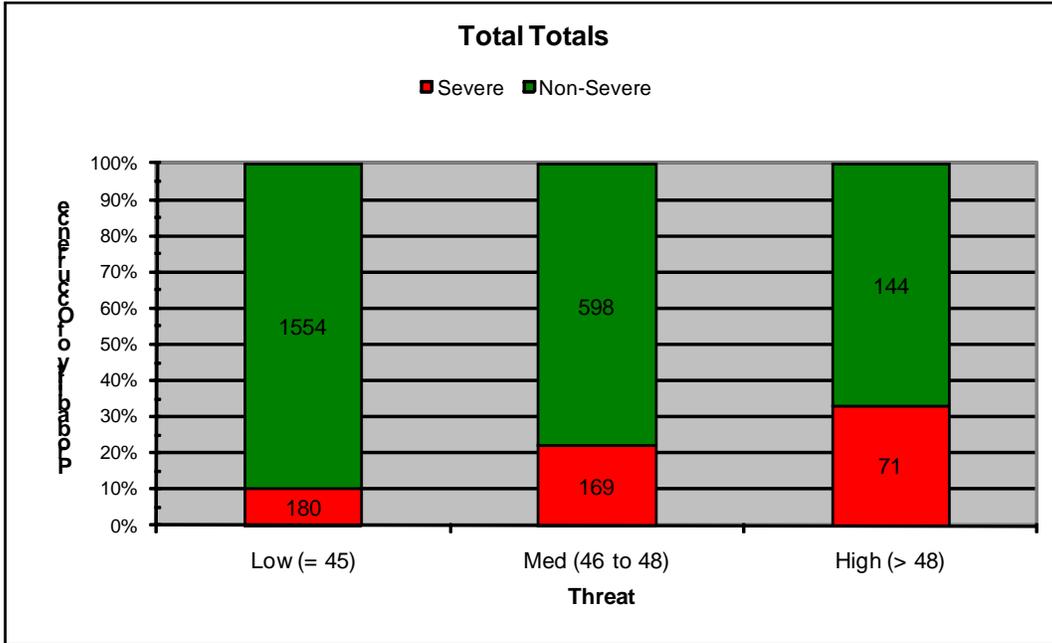
Figure 1. Stacked column graph of TT thresholds. The number of severe/non-severe occurrences for the Low, Medium and High threat thresholds for all 22 years in the severe weather database.
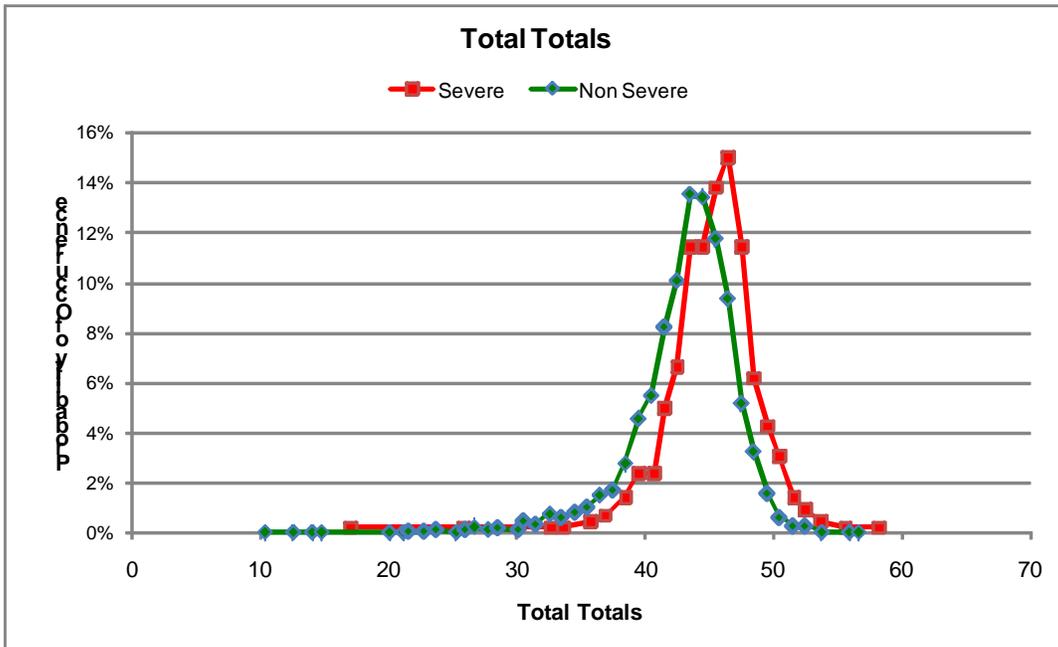
Figure 2. Scatter plot of probability of occurrence vs. TT for both the non-severe and severe days for all 22 years in the severe weather database.

10

4.2 K-Index (KI)

The KI thresholds values indicate a low threat for severe weather when KI < 26, a medium threat when $26 \geq$ KI ≥ 28, and a high threat when KI > 28. When KI was > 28, a severe weather event was reported in 16% of the 2010 days. This did not alter the 21-year value of 18%. The 22-year value remained at 18%. Figure 3 displays the threat levels of Low, Medium and High with the occurrence/non-occurrence of severe weather for the 22year POR, while Figure 4 displays the same data, but shows the individual KI values for each day. It is evident that KI poorly discerns severe from non-severe weather for our POR.
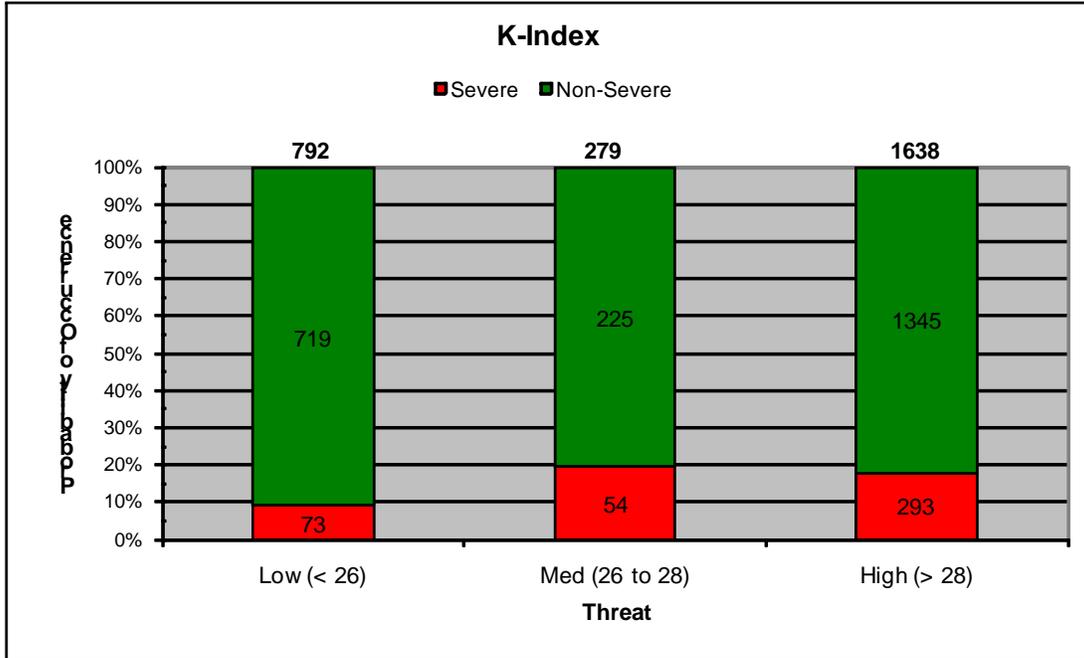

Figure 3.   Same as Figure 1 except for KI.

Figure 4. Same as Figure 2, except for KI.

11

4.3 Lifted Index (LI)

The LI thresholds values indicate a low threat for severe weather when LI ≥ -2, a medium threat when -3 ≥ LI ≥ -5, and a high threat when LI < -5. The LI was never < -5 in the 2010 dataset and, therefore, slightly decreased the percentage of severe weather in the 22-year POR from 31% to 30%. Figure 5 shows the LI Low, Medium and High threat distribution for all years in the severe weather database, while Figure 6 displays the same data, but shows the individual LI values for each day. LI also poorly discerns severe from non-severe weather for our POR.

Figure 5.   Same as Figure 1 except for LI.

Figure 6. Same as Figure 2, except for LI.

12

4.4 Thompson Index (TI)

The TI specifies a low threat when TI < 25, a medium threat when $25 \leq TI \leq 34$, a high threat when $35 \leq TI \leq 39$, and a very high threat when $TI \geq 40$. The TI value was > 40 on 4 days in the 2010 season and severe weather was not reported on any of those days. The percent occurrence decreased to 25% for the 22-year POR over the previous 21-year value of 26%. Figure 7 shows the severe weather threat distribution for all years in the severe weather database, while Figure 8 displays the same data, but shows the individual TI values for each day. As above, TI poorly discerns severe from non-severe weather in our POR.

Figure 7. Same as Figure 1 except for TI and the fourth threat category Very High.
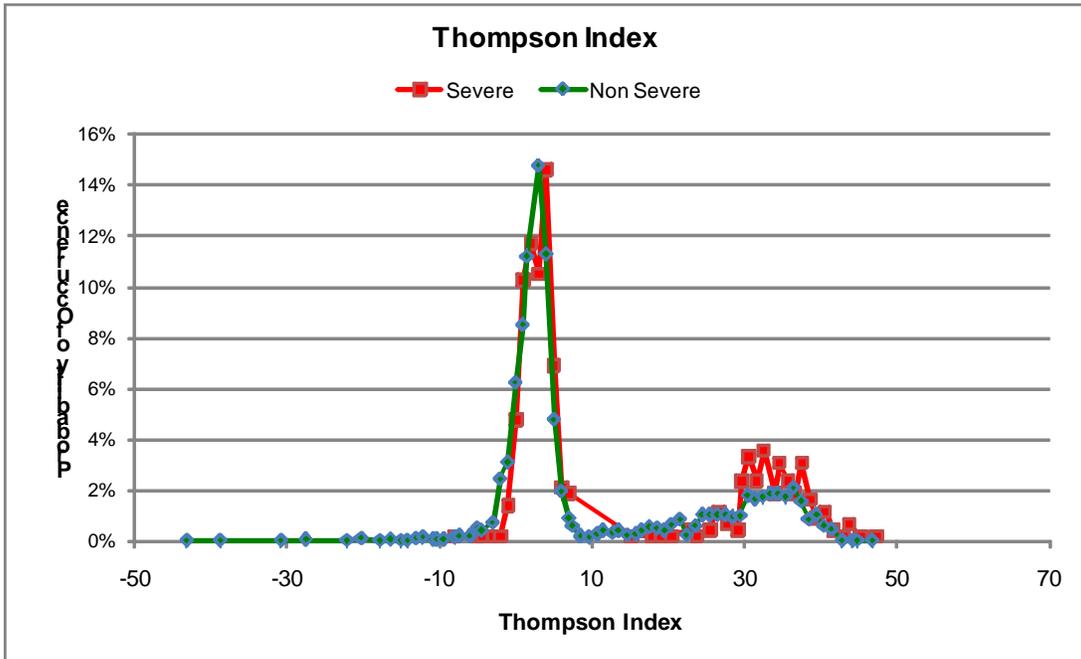
Figure 8. Same as for Figure 2, except for TI.

4.5   Cross Totals (CT)

The CT thresholds indicate a low threat when CT ≤ 19, a medium threat when 20 ≤ CT ≤ 21, a high threat when 22 ≤ CT ≤ 23, and a very high threat when CT ≥ 24. When CT was ≥ 24, a severe weather event was reported in 20% of the 2010 days. This did not alter the 21-year value of 31%. The overall 22-year value remained at 31%. Figure 9 displays the threat levels of Low, Medium, High, and Very High with the occurrence/non-occurrence of severe weather for the 22-year POR, while Figure 10 displays the same data, but shows the individual CT values for each day. It is evident that CT poorly discerns severe from non-severe weather for our POR.

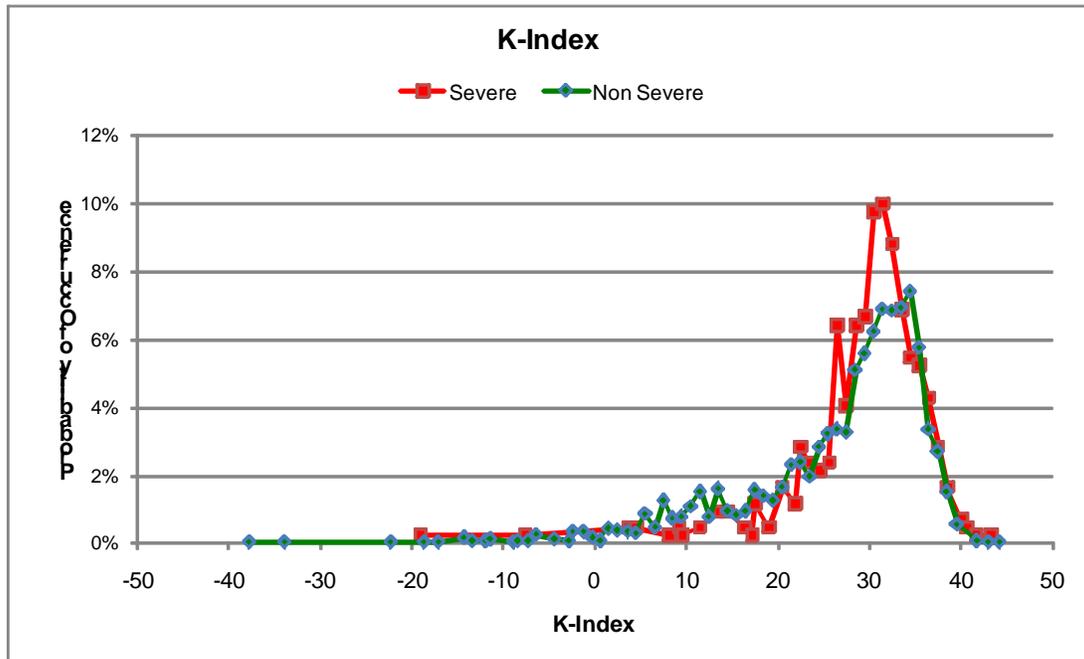Figure 9. Same as Figure 7 except for CT.
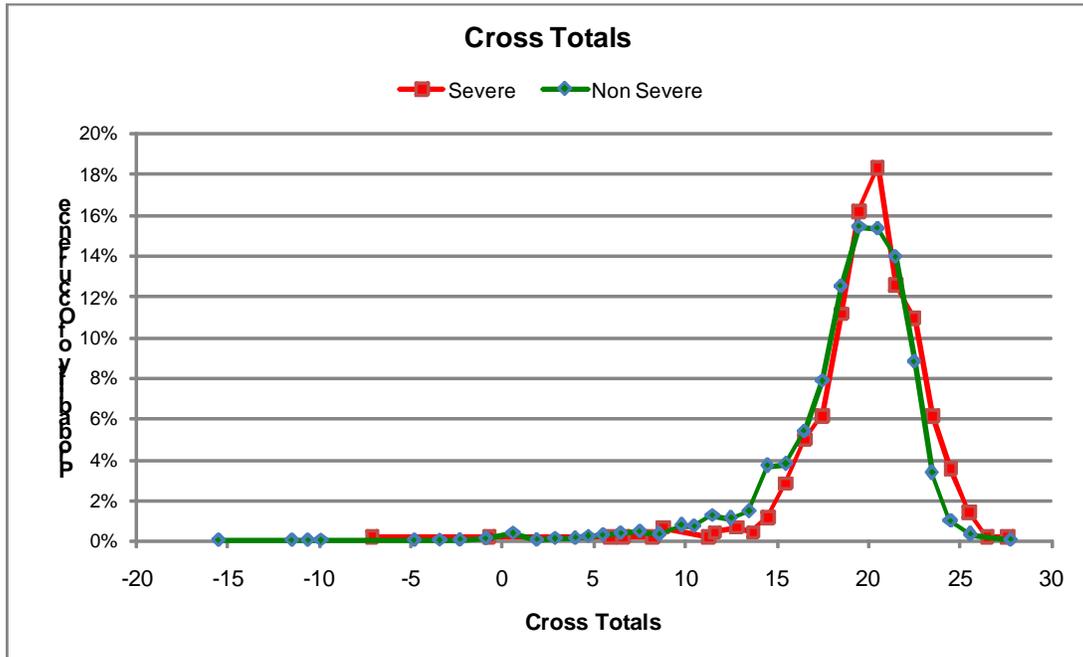
Figure 10. Same as Figure 2, except for CT.

14

4.6 Showalter Stability Index (SSI)

The SSI thresholds indicate a low threat when SSI ≥ 3, a medium threat when 2 ≥ SSI ≥ -2, and a high threat when SSI < -2. The 2010 severe weather database confirmed that SSI is a good severe weather predictor. When SSI < -2, severe weather was reported in central Florida 33% of the time. The 22-year POR value remained the same as the previous work at 37%. Figure 11 shows the SSI Low, Medium and High threat distribution for all years in the severe weather database, while Figure 12 displays the same data, but shows the individual SSI values for each. SSI poorly discerns severe from non-severe weather for our POR.
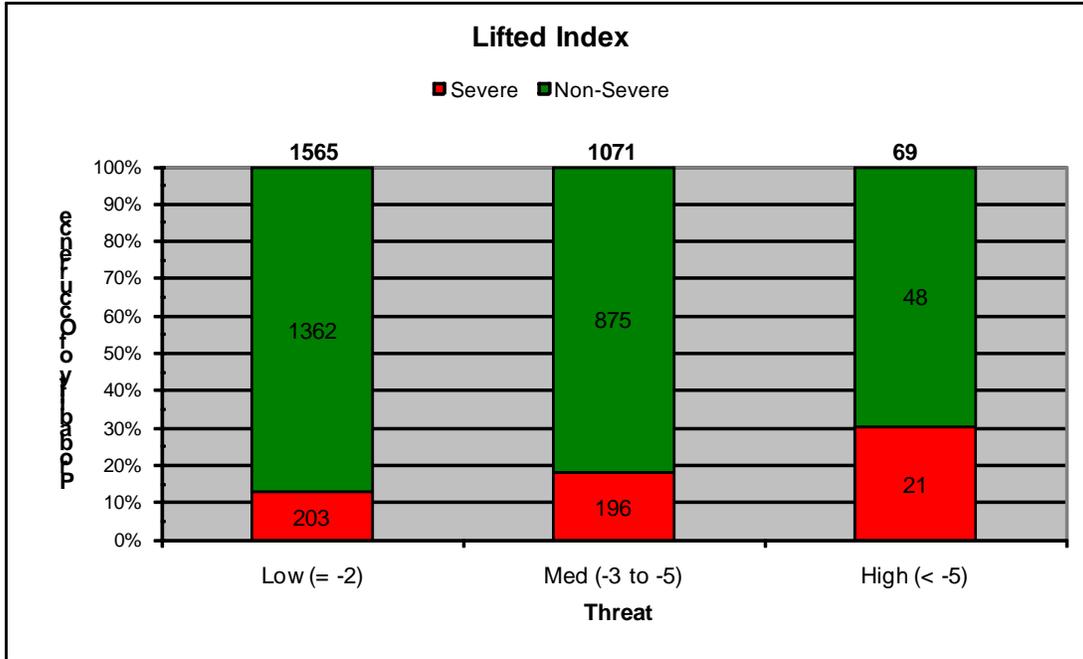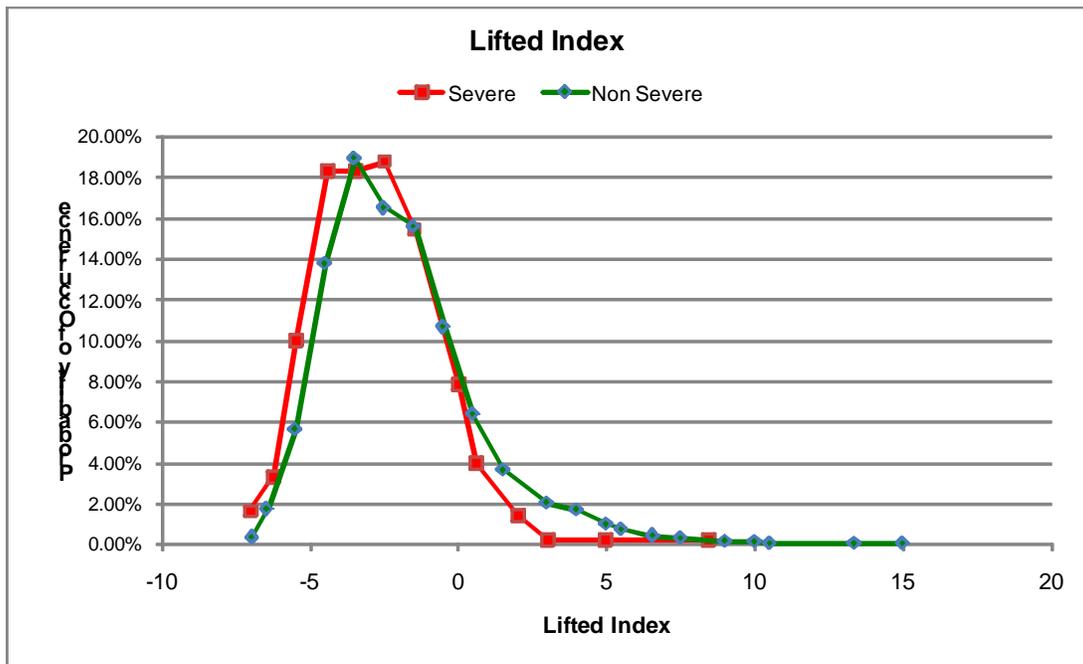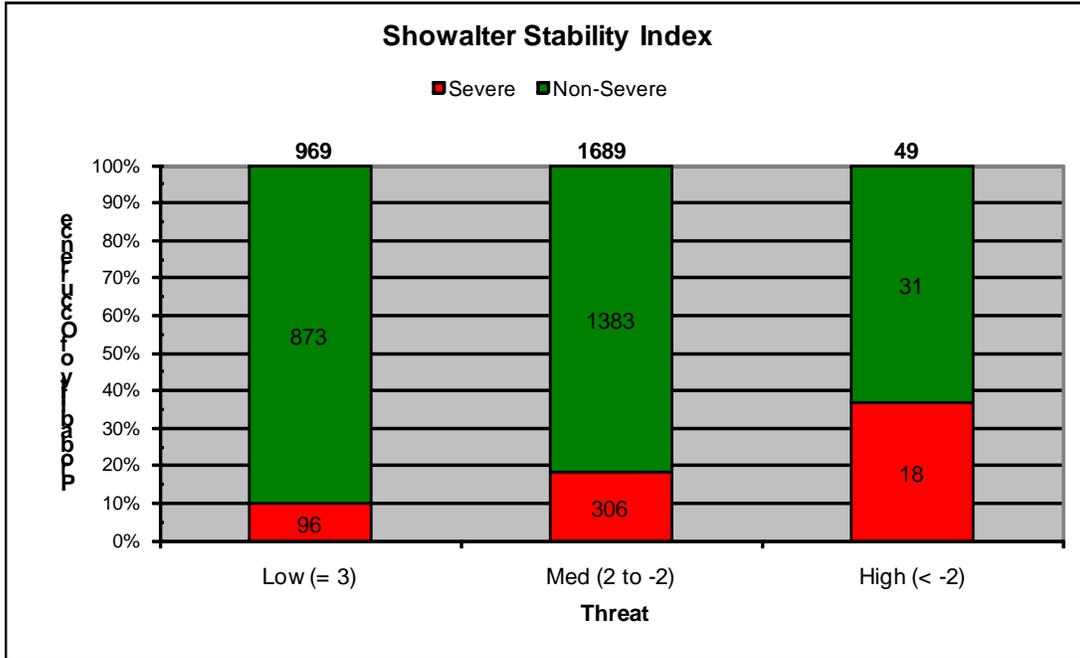

Figure 11. Same as Figure 1 except for SSI.
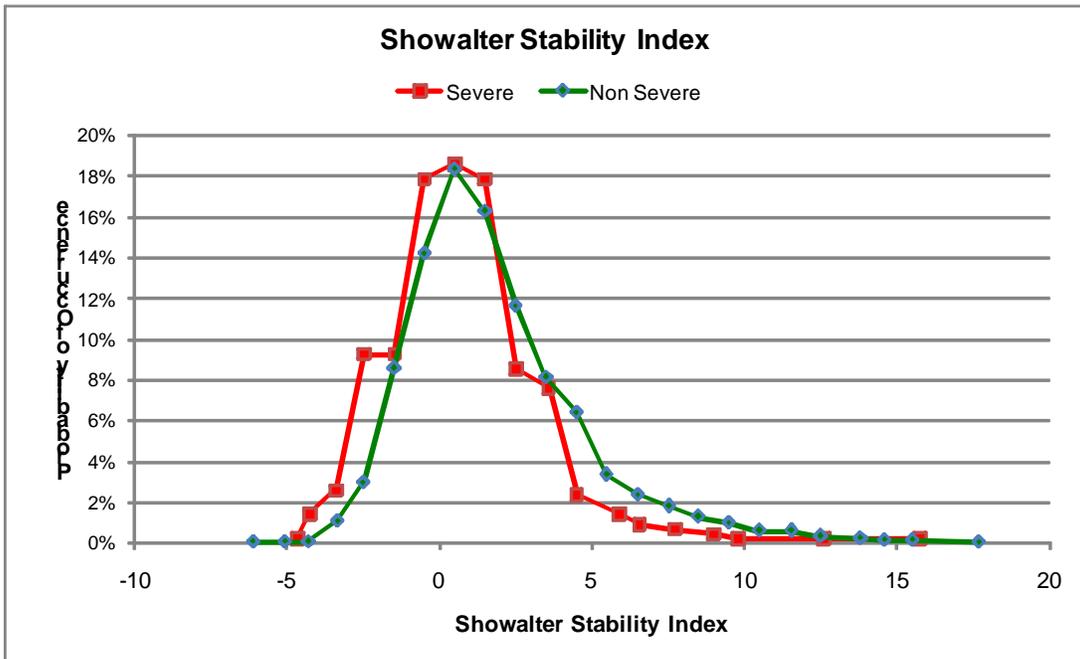
Figure 12. Same as Figure 2, except for SSI.

15

4.7 Precipitable Water (PW)

The PW thresholds indicate a low threat when PW < 1.0 in, a medium threat when 1.0 in ≤ PW ≤ 1.75 in, and a high threat when CT > 1.75 in. When PW was > 1.75, a severe weather event was reported in 13% of the 2010 days. This did not alter the 21-year value of 15%. The overall 22-year value remained at 15%. Figure 13 displays the threat levels of Low, Medium, and High with the occurrence/non-occurrence of severe weather for the 22-year POR, while Figure 14 displays the same data, but shows the individual PW values for each day. As above, PW does not discern severe from non-severe weather in our POR.
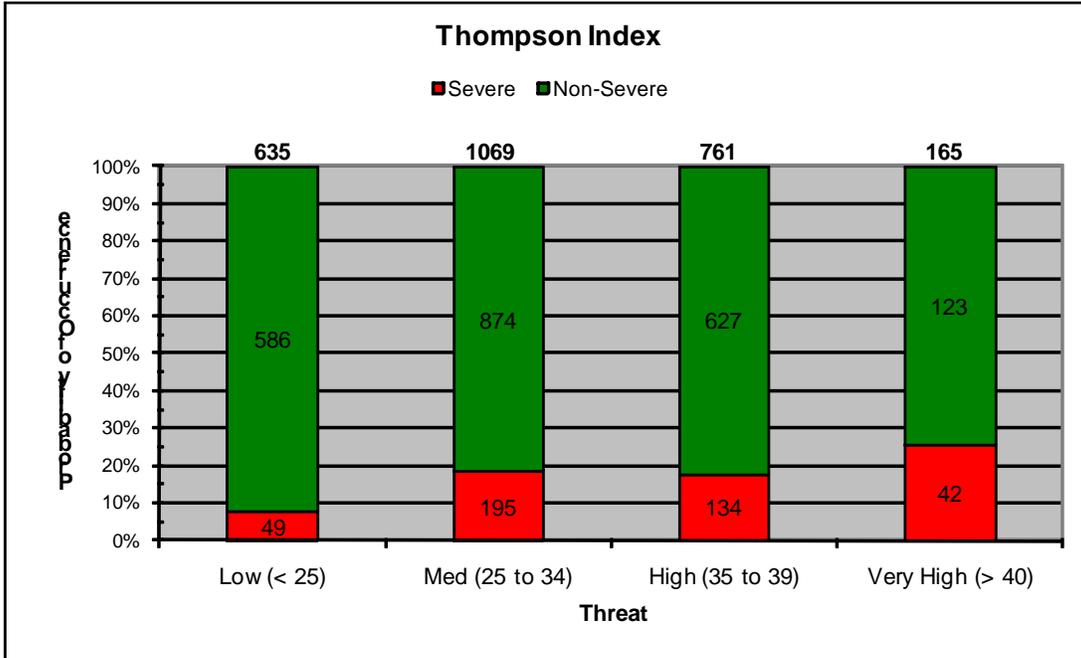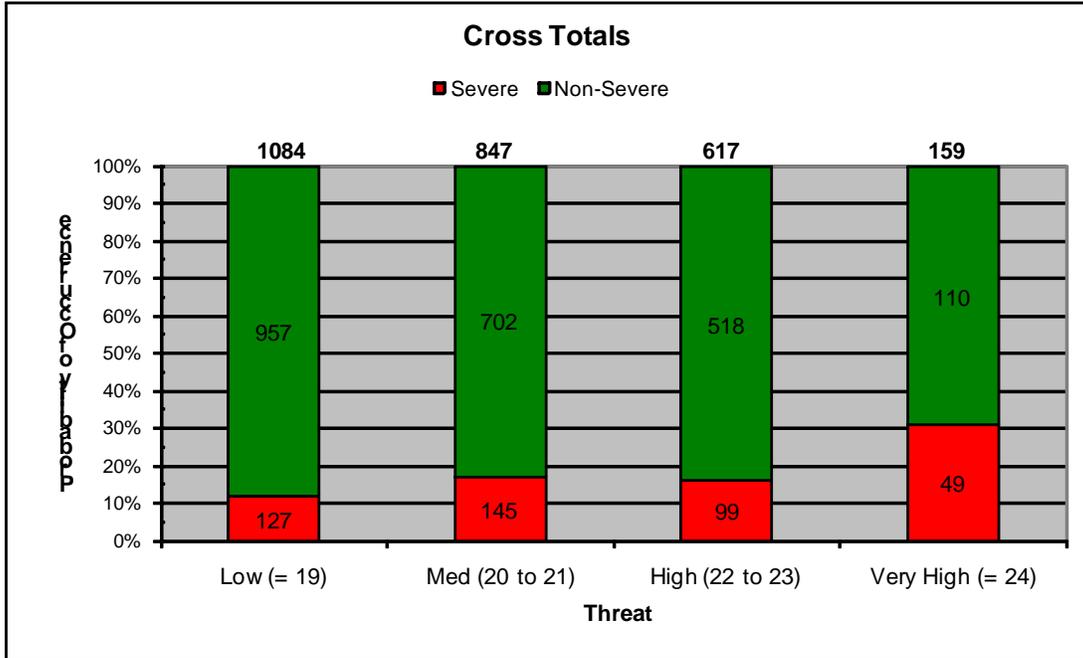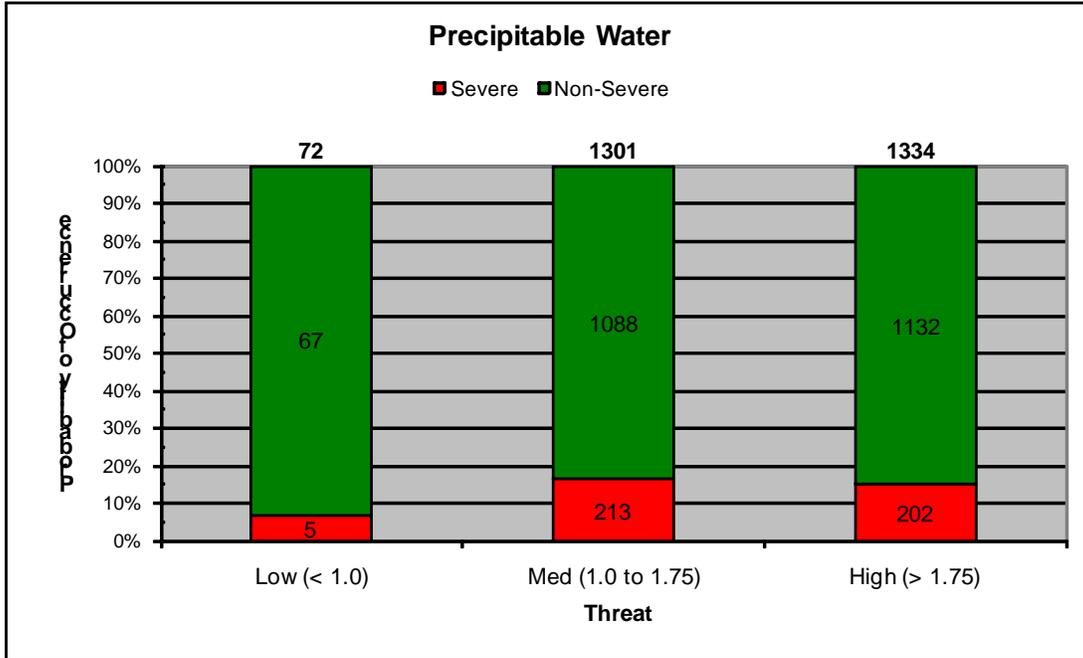

Figure 13. Same as Figure 1 except for PW.
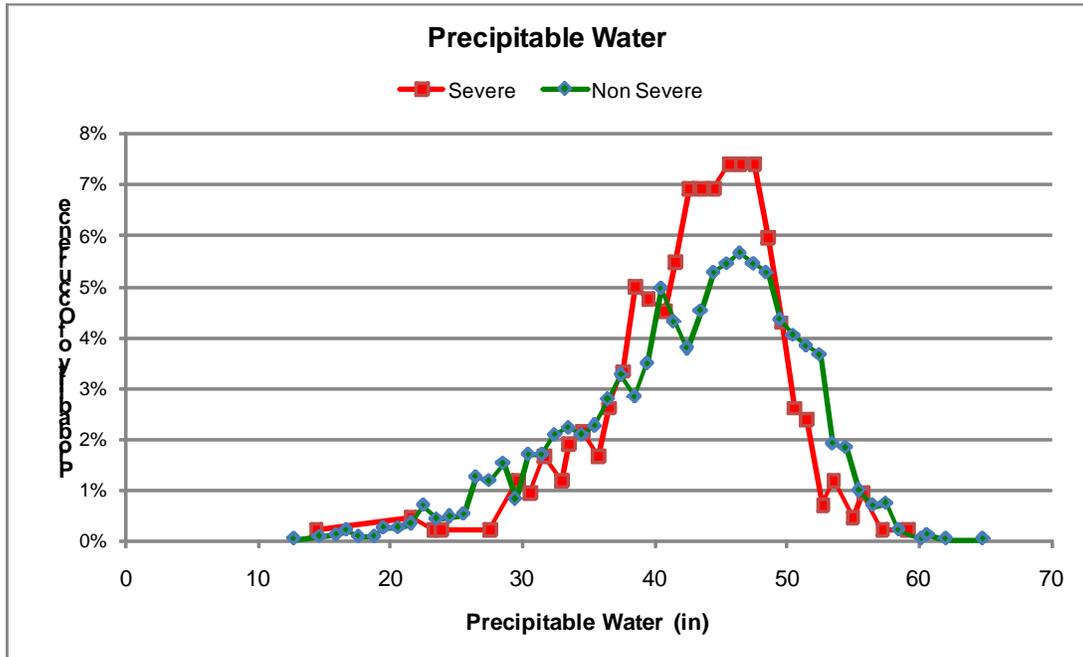
Figure 14. Same as Figure 2, except for PW.

5.   2010 Verification Results

The TTS used for verification was developed from the first severe weather study (Bauman et al. 2005). The AMU calculated verification statistics for the TTS from an independent dataset created by the 45 WS forecasters and AMU personnel during the 2010 warm season. When the 45 WS forecasters/AMU personnel completed the Severe Weather Worksheet GUI and computed the daily TTS, a text file was saved that contained their answers to the subjective questions and the sounding stability parameters for the day. This allowed a comparison of the daily TTS with reported severe weather events in 2010.

From 3 May to 30 September 2010, the AMU and 45 WS forecasters completed 132 worksheets. Total Threat Scores ranged from -23 to 20. Severe weather was reported in east-central Florida on 16 of the 153 days. Figure 15 shows the TTS values color-coded for reported severe weather. On days that severe weather occurred, the TTS ranged from -11 to 20. During the 2006 warm season verification (Bauman 2006), the TTS was never below 0 on days when severe weather was reported.

The 2010 warm season was one of the hottest and driest summers on record across east-central Florida. This was due to the placement and strength of the surface Atlantic ridge and high pressure ridge aloft. This resulted in few severe weather reports during the 2010 summer season and may account for the wide range of TTS values when severe weather was reported over the 2010 season.
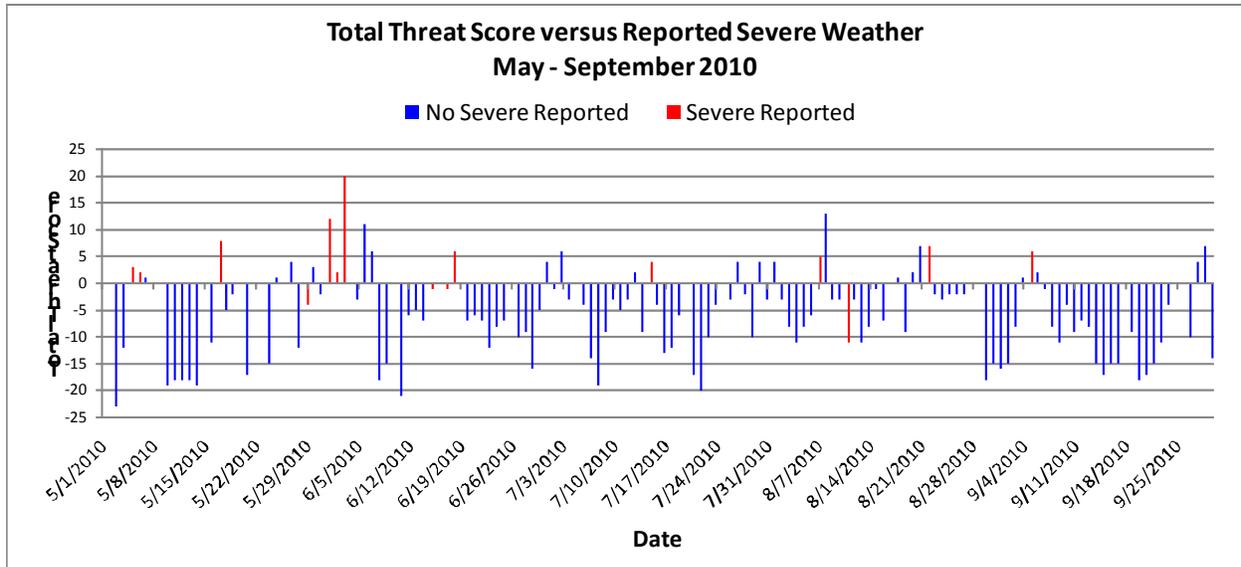
Figure 15. Total Threat Score versus date from 1 May – 30 September 2010. Red lines represent days with reported severe weather, blue lines represent days with no reported severe weather.

The AMU computed verification statistics for the 2010 warm season. The standard 2x2 contingency table shown in Table 2 was used to calculate the statistics and scores described in the last row of Table 2:

- The False Alarm Rate (FAR) is the fraction of 'yes' forecasts that are incorrect,
- Probability of Detection (POD) is the fraction of 'yes' forecasts that are correct,
- Critical Success Index (CSI) measures the fraction of observed or forecast events that were correctly predicted,
- Heidke Skill Score (HSS) is the probability of a correct 'yes' forecast by random chance, and
- True Skill Statistic (TSS) measures how well the forecast separated the 'yes' events from the 'no' events compared to random chance, but with an assumption of an unbiased forecast.

17

Table 2. The standard contingency table used for forecast verification.		Observed Event	
		Yes	No
Forecast Event	Yes	a	b
	No	c	d

N = a + b + c + d

False Alarm Rate (FAR) = b/(a+b)

Probability of Detection (POD) = a/(a+c)

Critical Success Index (CSI) = a/(a+b+c)

Heidke Skill Score (HSS) = [(a+d) - E]/(N-E)

E = [(a+c)(a+b)+(b+d)(c+d)]/N, N=a+b+c+d

True Skill Statistic (TSS) =a/(a+c) – b/(b+d)

Table 3. Warm season 2010 TTS Verification Statistics		Observed Severe			
		Yes	No	FAR	0.46
				POD	0.44
Forecast Severe	Yes	7	6	CSI	0.32
	No	9	110	HSS	0.42
				TSS	0.39
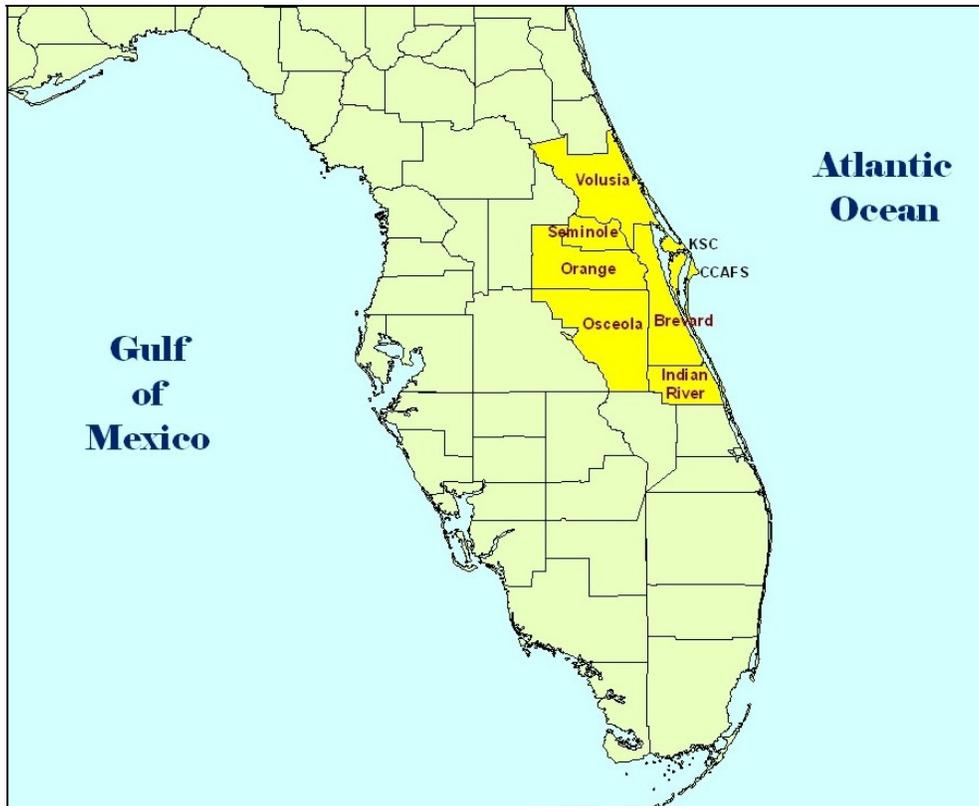
Figure 16. Map of Florida showing the six counties (shaded in yellow) included in the severe weather events database. The location of KSC and CCAFS are shown on the map.

Table 3 shows the contingency table statistics for the 2010 warm season. The TTS forecast threshold value for the contingency table was chosen based on the results of the 2006 summer season verification. If the TTS was < 5 it was considered a No forecast and if ≥ 5 it was a Yes forecast. The east-central Florida severe weather verification area (Figure 16) included three coastal counties (Brevard, Volusia, Indian River) and three inland counties (Seminole, Orange, Osceola), all of which are typically in the same large-scale air mass as KSC/CCAFS on most warm season days. If severe weather was reported in these Florida counties, that was classified as observed Yes. The Severe Weather Worksheet TTS did not verify well in the 2010 warm season, with a high FAR and low values for POD, CSI and HSS. However, it should be noted again that the 2010 warm season was atypical being much warmer and drier with much less severe weather than normal.

Figure 17 displays a scatter plot of probability of occurrence/non occurrence of severe weather vs. the TTS for both the 2009 and 2010 summer seasons. There were 36 severe weather and 186 non-severe weather days for which the TTS was calculated. The results show that TTS was a fairly good indicator of severe weather, particularly when the value was ≥5. When TTS was ≥5, the occurrence of non-severe weather was 1% or less. Based on these results and those from Section 4, it is evident that the final daily value of the TTS is driven by the subjective questions that the forecaster is required to answer as the daily stability indices poorly differentiated severe from non-severe weather.
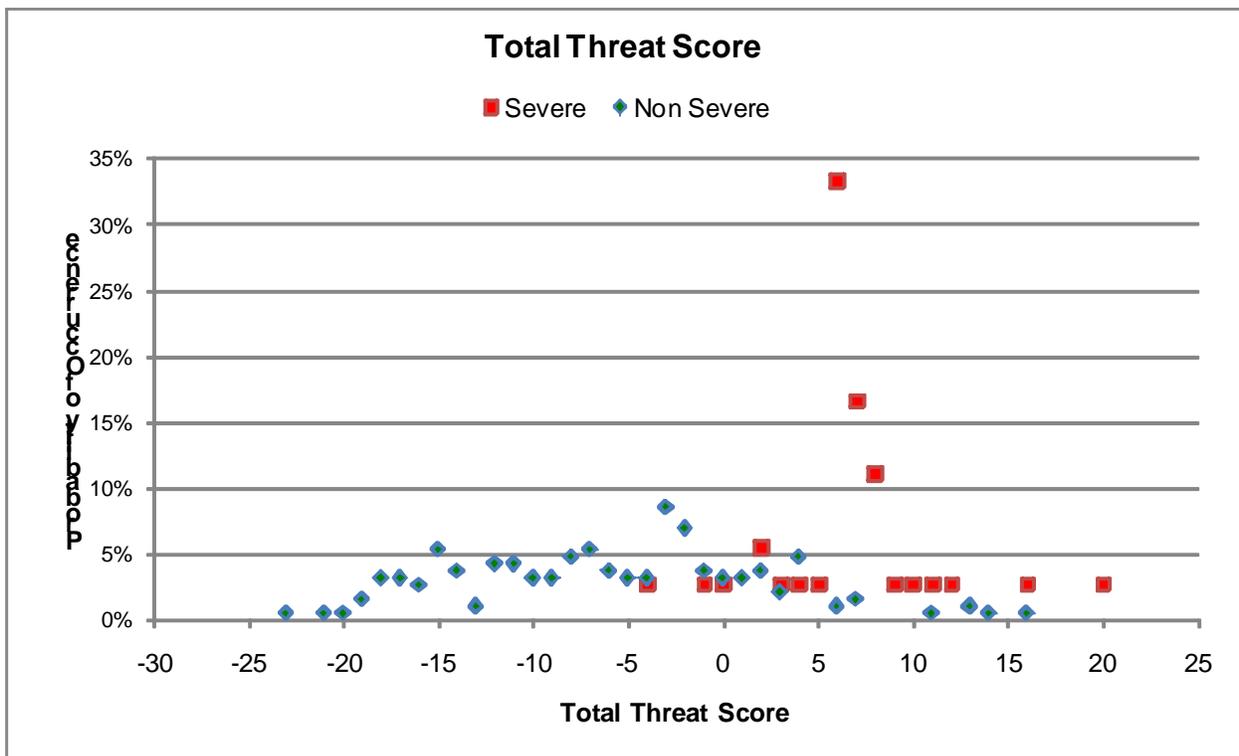
Figure 17. Scatter plot of probability of occurrence vs. TTS for both the non-severe and severe days for the 2009 and 2010 seasons in the severe weather database.

6. Logistic Regression Analysis

The AMU was tasked to perform statistical logistic regression analysis on the 22-year severe weather database and possibly develop a new forecast tool. There were three major steps in this portion of the task:

- Determine the elements of the equation,
- Develop the logistic regression equation, and
- Determine the equation performance.

To accomplish this, the AMU followed procedures outlined by Lambert and Wheeler (2005).

6.1 Elements of the Logistic Regression Equation

The necessary elements to create the logistic regression equation include a predictand and candidate predictors. The predictand is the element to be predicted. The SPC and NCDC severe weather reports provided the occurrence of severe weather in the area and were used to create the predictand. The predictand value was set to "1" if severe weather occurred within the six east-central Florida counties on a specific day and set to "0" if no severe weather occurred. As mentioned previously, an assumption had to be made that severe weather only occurred on reported severe weather days. The candidate predictors included the flow regimes from the Florida rawinsondes, the placement of the upper-level jet, and seven stability parameters calculated from the XMR rawinsonde.

6.1.1 Flow Regime Probabilities

Probabilities of severe weather occurrence based on flow regime pattern for each day were calculated using the severe weather binary predictand. The number of days each regime occurred was compared to the severe weather predictand to see how many of those days reported severe weather. The probability was calculated by dividing the number of severe weather days within a particular regime by the total number of days the regime occurred.

6.1.2 Upper-level Jet Probabilities

Probabilities of severe weather occurrence based on the placement of the upper-level jet were calculated in the same manner as the flow regime probabilities. The number of days each jet pattern occurred was compared to the severe weather predictand to see how many of those days reported severe weather. The probability was calculated by dividing the number of severe weather days with a particular jet pattern by the total number of days the particular pattern occurred.

6.1.3 Stability Index Parameters

The stability indices chosen as candidate predictors were based on the results from the previous phases of this work. All seven indices showed some skill in predicting severe weather. The stability indices were calculated for each day in the database from the 1000 UTC XMR sounding and are available to the forecasters through MIDDS. The stability index candidate predictors included the

- Total Totals (TT),
- Cross Totals (CT),
- K-Index (KI),
- Lifted Index (LI),
- Thompson Index (TI),
- Showalter Stability Index (SSI), and
- Precipitable water (PW).

6.2 Development of the Logistic Regression Equation

The amount of data available for equation development was critical to the reliability of the new equation. Data had to be stratified into equation development and verification datasets, which limited the amount of data available for equation development. Therefore, the amount of available data was determined before development began. After determining that an appropriate amount of data was available, one equation for the warm season was developed.

6.2.1 Data Availability

The World Meteorological Organization (1992, WMO) states that there should be at least 250 events in the dataset in order to derive stable statistical relationships. There are 153 days in the warm season, which equates to 3366 days over the 22-year POR. However, sounding data were not available for every day in the POR. Data were available for 3192 days or 95% of the time. Of these days, there were 422 reported severe weather days. This was sufficient to satisfy the WMO standard after stratifying the full dataset into development and verification datasets.

6.2.2 Development and Verification Datasets

The candidate predictors and predictand were stratified into development and verification datasets. The development dataset was required to contain enough samples so that the resulting logistic regression equation was stable. The verification dataset was needed for equation testing to ensure that the equation would perform sufficiently in an operational setting. If the performance was much worse with the verification data, this would indicate that the development dataset was too small or there were too many predictors and the equations were fit too strongly to the development data.

The daily TTS values were archived for the years 2009 and 2010. Therefore, these two years were chosen as the verification dataset in order to compare the accuracy of the equation vs. the TTS. This left 20 years of data (1989-2008) for the development dataset. The development dataset contained 380 severe weather events, while the verification dataset contained 42.

6.2.3 Equation Development

The development of the logistic regression equation follows the procedure outlined in Lambert and Wheeler (2005). One logistic regression equation was developed using candidate predictors determined from the previous phases of this task.

6.2.3.1 Logistic Regression

Choosing the correct statistical regression method is essential when creating a reliable probability forecast tool. Logistic regression is deemed most appropriate when using data with a predictand that is binary (Wilks 2006). Logistic regression is represented by the equation

$$y = \frac{e^{(b_0 + b_1 x_1 + \cdots + b_k x_k)}}{1 + e^{(b_0 + b_1 x_1 + \cdots + b_k x_k)}}, \tag{1}$$

where y is the predicted value, $x_1 \ldots x_k$ are the set of predictors, and $b_1 \ldots b_k$ are the coefficients for the corresponding predictors. For this task, y represents the probability of a severe weather event occurring and is bound between the values 0 and 1. The candidate predictors for $x_1 \ldots x_k$ are those listed in Sections 6.1.1, 6.1.2, and 6.1.3 and the method for determining the corresponding coefficients is outlined in Section 6.2.3.2. A detailed description of logistic regression can be found in Section 4.2.1 of Lambert and Wheeler (2005).

6.2.3.2 Predictor Selection

Following Lambert and Wheeler (2005), predictor selection was conducted using the S-PLUS[®] statistical software (Insightful Corporation 2005), which has a built-in logistic regression function. The software also determines each predictor's contribution to the reduction in variance of the predictand, called the reduction in the residual deviance. For logistic regression, the residual deviance is used to assess the fit of the overall model. The smaller (larger) the deviance is the better (worse) the fit of the model. A detailed description of residual deviance can be found in Section 4.2.2 of Lambert and Wheeler (2005).

A process called screening regression was used to determine which candidate predictors to include in the logistic regression equation. In this approach, predictors were added to the equation one at a time. At each step, the candidate predictor that created the biggest reduction in the residual deviance was chosen as the next predictor in the

equation. Selection began with the prediction equation $y = \frac{e^{(b_0)}}{1 + e^{(b_0)}}$ (NULL model), where the only term in Equation (1) is the intercept. In the next step, each of the seven candidate predictors was added as a lone predictor in Equation 1 resulting in seven single predictor equations. The predictor that caused the largest reduction in the residual deviance from the NULL model was chosen as the first predictor in the equation. At this stage, the prediction

equation is $y = \frac{e^{(b_0 + b_1 x_1)}}{1 + e^{(b_0 + b_1 x_1)}}$. Next, the other six candidate predictors were added individually to the equation creating a set of two-predictor equations. The second predictor that caused the largest reduction in residual deviance was chosen as the second predictor. This continued for all candidate predictors. It is important to note that it is generally not useful to include all potential predictors in a final equation since most predictor variables are mutually correlated so that the full set of predictors includes redundant information (Wilks 2006). This could create unrealistic results.

Figure 18 shows the percent reduction in residual deviance from the NULL model as each predictor was added. The TT reduced the residual deviance by the most (~8%) and was chosen as the first predictor in the equation. The second predictor was the flow regime probabilities, which brought the total reduction of residual deviance to ~13%. The LI and jet probabilities were the third and fourth predictors in the equation, respectively, producing the final reduction in residual deviance of 15%.

There was no sufficient fifth predictor for the equation. In other words, no other candidate predictor reduced the residual deviance by a significant amount, thereby not providing added value for predicting severe weather. The regression coefficients for each predictor, $b_1 \ldots b_k$, should maintain the correct sign during each step described above. A positive regression coefficient means that the predictor increases the probability of the outcome, while a negative coefficient means that the predictor decreases the probability of that outcome. For this study, KI, TI, CT, TI, PW, flow regime probabilities, and jet probabilities should have positive coefficients indicating that larger (smaller) values of each variable increase (decrease) the chance of severe weather. The variables LI and SSI should have negative coefficients indicating that larger (smaller) values decrease (increase) the chance of severe weather. None of the fifth candidate predictor coefficients had the correct sign, indicating that none of the predictors added value for predicting severe weather.
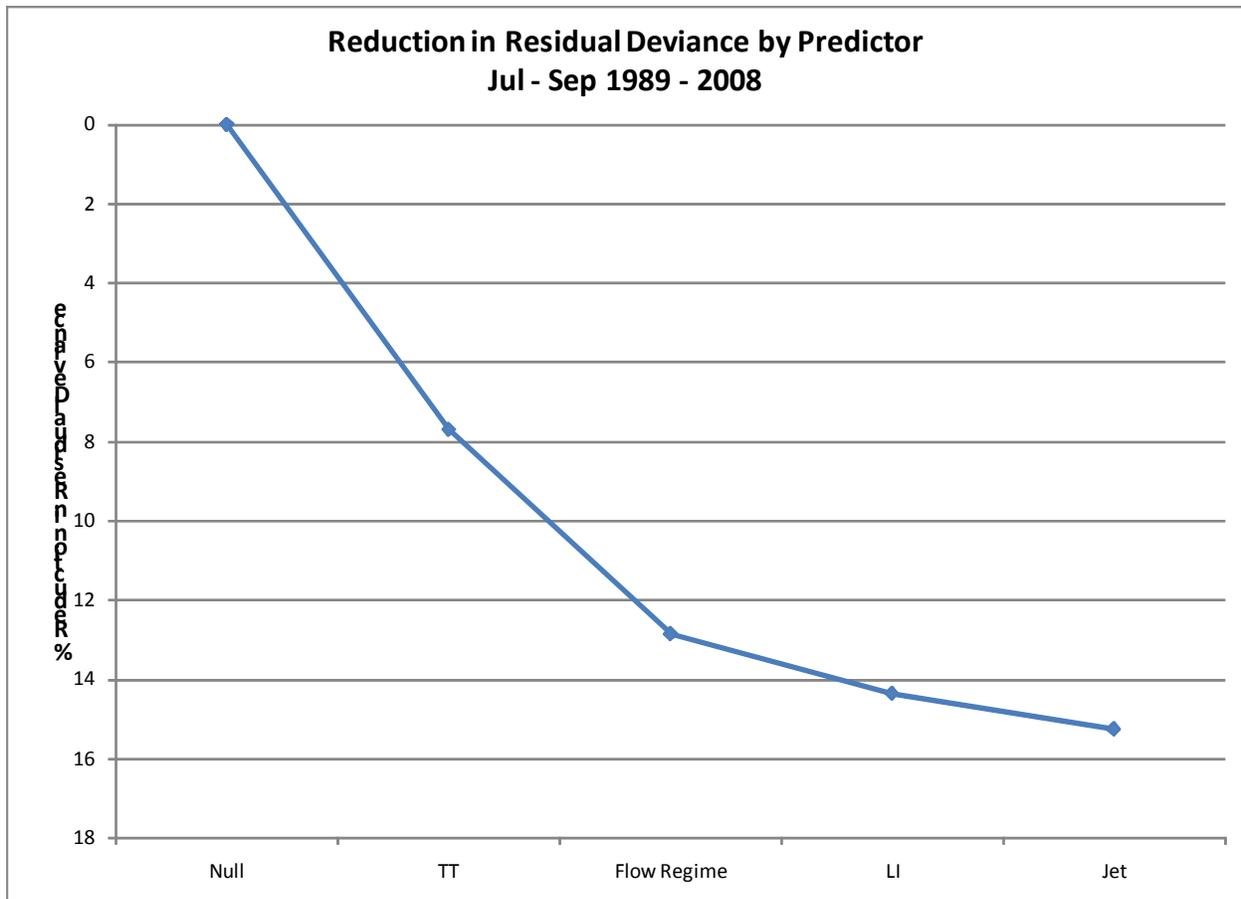

Figure 18. The total percent reduction in residual deviance from the NULL model as each predictor was added to the equation using the development dataset.

6.3   Logistic Regression Equation Performance

Forecast probabilities were produced using the four predictors from the verification dataset. These probabilities were compared with the binary severe weather observations in the verification dataset using four tests that measure forecast performance. The tests included

- Mean Squared Error, which evaluates equation performance,
- Brier Skill Score, which measures equation performance against other forecast methods,
- Distributions of the probability forecasts for days with and without severe weather, and
- Contingency table statistics.

6.3.1   Mean Squared Error

The Mean Squared Error (MSE) is the mean of the squared differences between the forecast probabilities and the observations. The MSE is given by

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2, \tag{2}$$

where n is the number of forecast/observation pairs, p_i is the probability calculated from the equation, and o_i is the corresponding binary severe weather observation (Wilks 2006). The MSE for a perfect forecast is 0, with larger MSE indicating decreasing accuracy of the forecast.

The MSE was computed for the four-predictor equation using the development and verification datasets. The MSE for the full development dataset was 0.10, which indicates skill in predicting severe weather. However, when the data was split into severe and non-severe events, the MSE was 0.61 and 0.03, respectively. Similarly, the MSE for the full verification dataset was 0.01, but 0.59 and 0.02 for severe and non-severe events, respectively. These results indicate that the equation was biased towards predicting non-events and failed to adequately predict severe weather events. The MSE was also computed for the TTS using the verification dataset. Based on previous work by Bauman (2006) and Wheeler (2010), a TTS $\geq$ 5 was used as the threshold for severe weather where a TTS < 5 was assigned a 0 (a No forecast) and a TTS $\geq$ 5 was assigned a 1 (a Yes forecast). The observation was subtracted from the TTS forecast value, the result was squared, and then the final mean was taken of all the squared differences. The MSE for the full verification dataset was 0.07 and was 0.26 and 0.03 for severe and non-severe events, respectively. Based solely on MSE, the TTS was a better predictor of severe weather events than the logistic regression equation.

6.3.2   Brier Skill Score

The Brier Skill Score (BSS) measures the improvement in skill of the logistic regression equation against a reference forecast. It is calculated using the MSE as

$$BSS = \left(\frac{MSE_{eqn} - MSE_{ref}}{MSE_{perfect} - MSE_{ref}}\right) * 100, \text{ (Wilks 2006)} \tag{3}$$

where MSE_{eqn} is the MSE of the equation being tested, MSE_{ref} is the MSE of the reference forecast method, and $MSE_{perfect}$ is the MSE of a perfect forecast, which is always 0. The BSS denotes a percent improvement (degradation) in skill of the equation over the reference forecast when it is positive (negative). The calculated TTS for the verification dataset was used for the reference forecast.

The BSS values for the verification dataset were -57% for the full dataset, -131% for the severe weather events, and 34% for non-severe weather. As with the MSE, these results indicate that the logistic regression equation is biased towards predicting non-events as the percent improvement for the non-severe weather is large. However, the percent degradation for predicting severe events is quite large, again indicating that TTS is a better tool for predicting severe weather.

6.3.3 Probability Distributions

The equation probability forecasts from the verification dataset were stratified by severe and non-severe weather days. The distribution of the probability values was calculated for each stratification. Figure 19 shows the probability distribution for severe days (red curve) and non-severe days (blue curve). If the equation performance was considered "good", the red (blue) curve should have a minimum (maximum) in the lower probability values that increase to a maximum (minimum) at the higher values.

The non-severe weather days have a peak frequency near 65% at probability values of 0.1 and then decrease to near 0 at 0.6. It shows a high percentage of low probabilities for non-severe events and a low percentage of high probabilities as expected for good performance. The severe weather days have a small peak near 30% at probability values near 0.2 followed by a dip and then another small peak near 15% at probability values at 0.4. This indicates that the equation performed poorly for severe weather days. The maximum at 0.2 and minimum at 0.6 suggests the equation is under-forecasting severe weather events. It should be noted that forecast probabilities for both severe and non-severe days were never greater than 0.7.
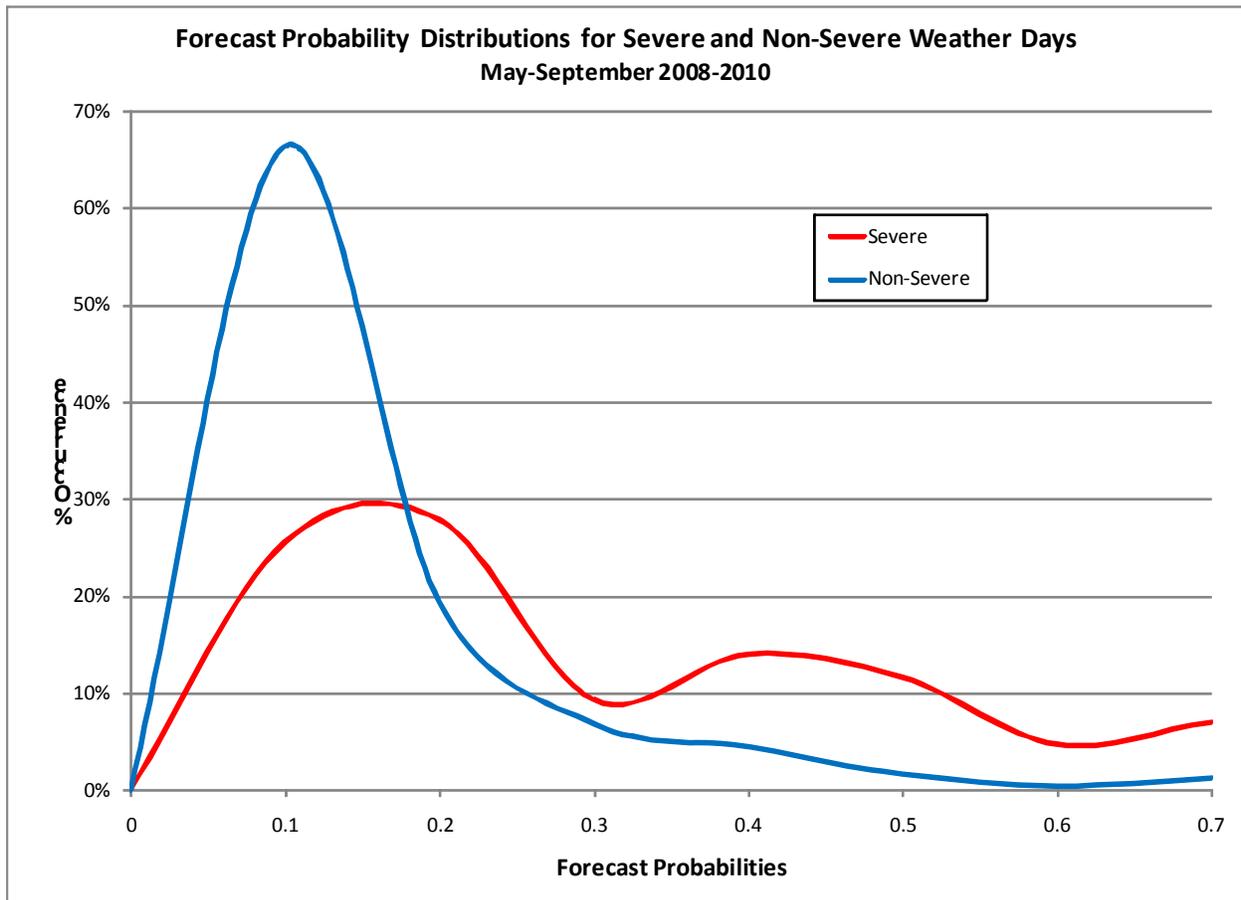
Figure 19. Forecast probability distributions for severe (red) and non-severe (blue) days in the verification data. The y-axis values are the frequency of occurrence of each probability value, and the x-axis values are the forecast probability values output by the equation.

6.3.4 Contingency Table Statistics

Contingency table statistics were computed for the verification dataset TTS and equation probabilities. As in Section 5, the standard 2x2 contingency table shown in Table 2 was used to calculate the statistics and scores. The contingency table statistics were computed using the same threshold values for TTS as in Section 5: if < 5 it was a No forecast and if ≥ 5 it was a Yes forecast. The procedure outlined by Wilks (2006) was used to choose the proper threshold values for the equation probabilities. Figure 20 shows the CSI and Bias values for the equation probabilities over a range of equation output probability values from 0.05 to 0.7 in increments of 0.05. The cutoff

value that had the maximum value of CSI and a bias value closest to 1 (no bias) was chosen. The resulting probability cutoff was 0.35 indicated by the vertical black line in Figure 20.

Table 4 and Table 5 show the contingency table statistics for the TTS and equation probabilities for the verification dataset, respectively. The POD and CSI are 1 for a perfect forecast and 0 for no skill, and vice versa for FAR. The HSS and TSS are 1 for a perfect forecast, 0 for performance equal to a random forecast, and < 0 for performance worse than that of a random forecast. It is evident that the TTS (Table 4) outperforms the equation (Table 5) in every computed statistic.
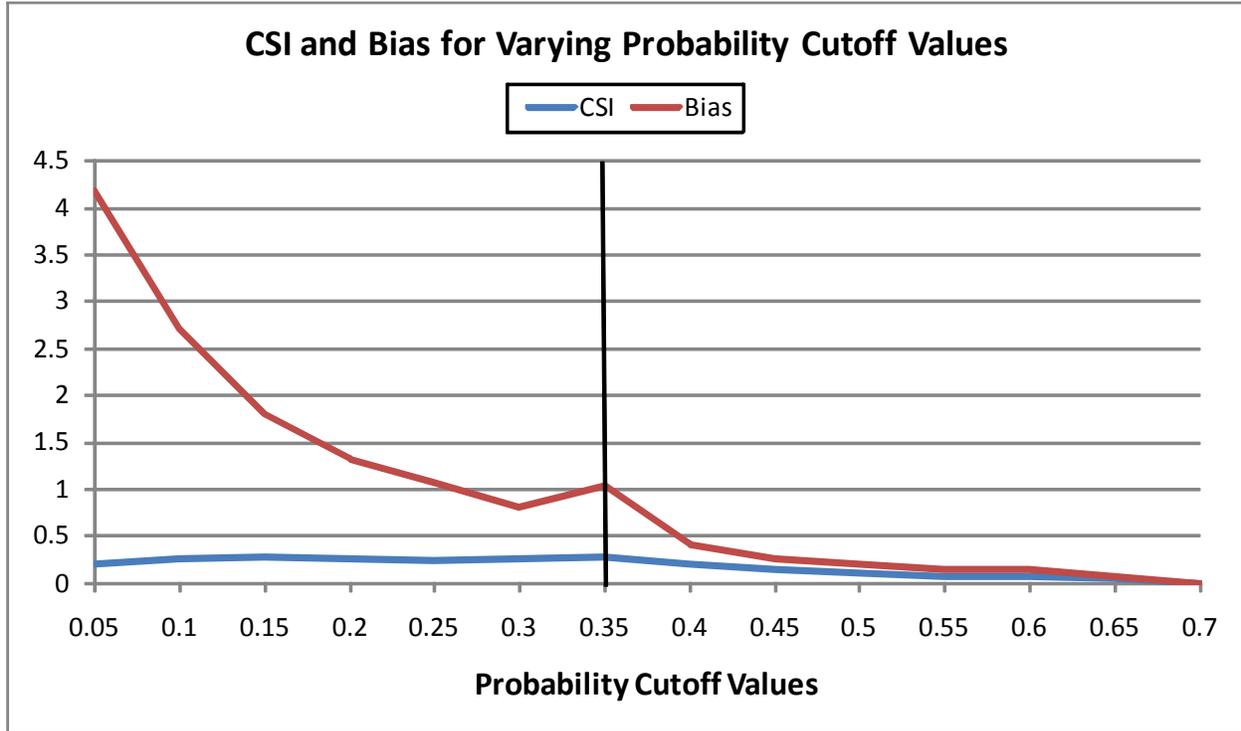
Figure 20. Graph showing the CSI (blue) and bias (red) values for the equation probabilities over a range of equation output probability values from 0.05 to 0.7 in increments of 0.05. The vertical black line shows the resulting probability cutoff that had the maximum value of CSI and a bias value closest to 1.

Table 4. TTS Verification Dataset Statistics		Observed Severe		FAR	0.23
		Yes	No	POD	0.73
Forecast Severe	Yes	**30**	**9**	CSI	0.60
	No	**11**	**176**	HSS	0.70
				TSS	0.68

Table 5. Equation Probabilities Verification Dataset Statistics		Observed Severe		FAR	0.42
		Yes	No	POD	0.35
Forecast Severe	Yes	**15**	**11**	CSI	0.28
	No	**28**	**238**	HSS	0.36
				TSS	0.30

6.3.5 Equation Performance Summary

All four equation performance measures indicated that the logistic regression equation did not show an increase in skill over the previously developed TTS. The equation showed less accuracy than the TTS at predicting severe weather, little ability to distinguish between severe and non-severe weather days, and worse standard categorical accuracy measures and skill scores over TTS. The MSE, BSS, and probability distributions show that the equation had some skill in predicting non-severe events, but no skill in predicting severe events.

The overriding difference between the logistic regression equation and the TTS is the inclusion of subjective questions and answers in computing the final value for the TTS. The logistic regression equation only takes into account the objective stability parameter values. Bauman et al. (2005) found that persistence, squall line activity, moisture boundaries, and sea breeze and boundary collisions were important for severe weather development and included questions to account for these phenomena when calculating the TTS. The results of this analysis emphasize the importance of these subjective factors.

7. Summary

This report presented a severe weather forecasting tool developed from a 22-year climatological study of severe weather events and related severe weather atmospheric parameters. Data sources included archived sounding data from the 1000 UTC XMR soundings, surface and upper air maps, and two severe weather event databases covering east-central Florida. The NCDC and SPC severe weather events databases were used to identify days with reported severe weather. These datasets provided the foundation for analyzing the stability parameters and synoptic patterns that were used to develop the original objective tool to aid in forecasting severe weather events. The severe weather database was upgraded by adding weather observations from May-September 2010. The new period of record for the analysis was May-September, 1989-2010.

Stability parameter analysis results indicate that adding the 2010 data had a minimal effect on the severe weather forecast potential of the tool. The AMU calculated verification statistics for the TTS values calculated by the 45 WS forecasters and AMU personnel in the 2010 warm season. On days that severe weather occurred, the TTS ranged from -11 to 20 compared to the 2006 summer season verification in which the TTS was never below 0 on days when severe weather was reported. Standard contingency table statistics showed a high FAR and low POD, CSI and HSS. The 2010 warm season was one of the hottest and driest summers on record across east-central Florida due to the placement and strength of the surface Atlantic ridge and high pressure ridge aloft. This resulted in few severe weather reports during the 2010 summer season and accounted for the wide range of TTS values when severe weather was reported over the 2010 season.

The AMU created a logistic regression equation that predicted the probability of severe weather occurrence for the day in east-central Florida. The equation was tested using four methods described in Section 6.3.1 - 6.3.4. The results from the tests show a degradation in skill in predicting severe weather over the TTS. The equation also showed less accuracy than the TTS at predicting severe weather, little ability to distinguish between severe and non-severe weather days, and worse standard categorical accuracy measures and skill scores over TTS.

Based on the findings of this study and after reviewing the results with the 45 WS, a new tool was not developed based on the performance of the logistic regression equation. The previously developed TTS and MIDDS GUI were not updated due to the inability of the 2010 severe weather season data to help improve the tool.

26

References

Bauman, W., M. Wheeler and D. Short, 2005: Severe Weather Forecast Decision Aid: Final Report. NASA Contractor Report CR-2005-212563, Kennedy Space Center, FL, 50 pp. [Available from ENSCO, Inc., 1980 N. Atlantic Ave., Suite 830, Cocoa Beach, FL, 32931 and online at http://science.ksc.nasa.gov/amu/final-reports/severe-tool-final.pdf.]

Bauman, W., 2006: Severe Weather Forecast Decision Aid Verification Testing. AMU Memo, 11pp. [Available from ENSCO, Inc., 1980 N. Atlantic Ave., Suite 830, Cocoa Beach, FL, 32931.]

Insightful Corporation, 2005: *S-PLUS 7 for Windows User's Guide*, Insightful Corp., Seattle, WA, 664 pp.

Lambert, W. and M. Wheeler, 2005: Objective lightning probability forecasting for Kennedy Space Center and Cape Canaveral Air Force Station. NASA Contractor Report CR-2005-212564, Kennedy Space Center, FL, 54 pp. [Available from ENSCO, Inc., 1980 N. Atlantic Ave., Suite 230, Cocoa Beach, FL, 32931, and at http://science.ksc.nasa.gov/amu/final.html.]

NCDC, 2010: U.S. Storm Events Database. [Available online at http://www4.ncdc.noaa.gov/cgi-win/wwcgi.dll?wwEvent~Storms.]

SPC, 2009: Severe Weather Geographic Information System Database. [Available online at http://www.spc.noaa.gov/gis/svrgis/.]

Wheeler, M., 2009: Severe Weather and Weak Waterspout Checklist in MIDDS. NASA Contractor Report CR-2009-214760, Kennedy Space Center, FL, 16 pp. [Available from ENSCO, Inc., 1980 N. Atlantic Ave., Suite 830, Cocoa Beach, FL, 32931 and online at http://science.ksc.nasa.gov/amu/final-reports/svr-wx-wksht-midds.pdf.]

Wheeler, M., 2010: Upgrade Summer Severe Weather Tool in MIDDS. NASA Contractor Report CR-2010-216282, Kennedy Space Center, FL, 13 pp. [Available from ENSCO, Inc., 1980 N. Atlantic Ave., Suite 830, Cocoa Beach, FL, 32931 and online at http://science.ksc.nasa.gov/amu/final-reports/severe-tool-upgrade.pdf.]

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2d ed. Academic Press, Inc., San Diego, CA, 467 pp.

World Meteorological Association (WMO), 1992: *Methods of Interpreting Numerical Weather Prediction Output for Aeronautical Meteorology.* Technical Note No. 195, ISBN 92-63-10770-X, 89 pp.

List of Acronyms

10070RH	1000 to 700 mb average Relative Humidity	MDPI	Microburst Day Potential Index
45 WS	45th Weather Squadron	MIDDS	Meteorological Interactive Data Display System
AMU	Applied Meteorology Unit	MSE	Mean Squared Error
BSS	Brier Skill Score	NOAA	National Oceanic and Atmospheric Administration
CAPE FMaxT	Cape using Maximum forecast Temperature	NCDC	National Climatic Data Center
CCAFS	Cape Canaveral Air Force Station	POD	Probability of Detection
CGLSS	Cloud-to-Ground Lightning Surveillance System	POR	Period of Record
		PW	Precipitable Water
CSI	Critical Success Index	SPC	Storm Prediction Center
CT	Cross Totals	SSI	Showalter Stability Index
ESRL	Earth System Research Laboratory	T_{850}	Temperature at 850 mb
FAR	False Alarm Rate	Tcl/Tk	Tool Command Language/Tool Kit
HTML	HyperText Markup Language	TDif	Forecast maximum temperature – convective temperature
HSS	Heidke Skill Score		
GUI	Graphical User Interface	TI	Thompson Index
KSC	Kennedy Space Center	TSS	True Skill Statistic
INV	Height of Inversion	TT	Total Totals Index
KI	K-Index	TTS	Total Threat Score
LI	Lifted Index	WMO	World Meteorological Organization
LLJ	Low Level Jet		

28

NOTICE

Mention of a copyrighted, trademarked or proprietary product, service, or document does not constitute endorsement thereof by the author, ENSCO Inc., the AMU, the National Aeronautics and Space Administration, or the United States Government. Any such mention is solely for the purpose of fully informing the reader of the resources used to conduct the work reported herein.