J2.5 **Developing a Severe Weather Alert System**

by

Ron Holmes[1]

And

Richard H. Grumm

NOAA/NWSFO, State College, PA

1. INTRODUCTION

Forecasting severe weather requires forecasters to process a significant number of parameters (Doswell et al. 1996) and to identify patterns associated with severe weather. Doswell (1986) pointed out the need for humans in the forecast process to perform both diagnostic and prognostic tasks. High speed computers and the ability to diagnose and integrate large datasets facilitate using automated algorithms to aid in performing these tasks. These algorithms can provide alerts based on patterns and past history of a potential weather event.

A severe weather pattern classifier was developed by Root et. al (2007) in which a clustering algorithm was employed to objectively identify anomaly patterns or fingerprints associated with severe weather. They showed some skill in predicting severe weather events based on these anomaly patterns. Though less sophisticated than that approach this paper describes a methodology employed at a local forecast office to accomplish a similar task.

It will be shown here, that by using data mining techniques predictors can be developed to assist forecasters in anticipating severe weather events. A serious question is whether or not integrating large databases and using AI algorithms is sufficient to adequately describe severe weather predictors based on what we already know. This paper attempts to accomplish this and will present a severe weather alert system developed by analyzing previous severe weather events and existing grid re-analysis datasets with AI algorithms.

2. DATA and METHODS

i. datasets

The National Climatic Data Center (NCDC) severe weather data base was used to identify severe weather events. The area of study was confined to the Mid-Atlantic region. The Storm Prediction Center (SPC) dataset was also used for comparative purposes. This provided a set of severe weather events based on time and geographic location. The types of severe weather included convectively produced severe winds (winds>50Kts), hail (size>0.75 inches), and tornadoes (F1-F5). Events from 1979-2007 were used in this study since this corresponded to the gridded data set period.

The gridded data used for this study was the North American Regional Reanalysis (NARR: Messinger et. al). This facilitated obtaining data variables within 32km of any severe weather report and provides a robust set of variables. The NARR variables extracted are presented in Table 1. This list represents a small subset of variables available for study. The NARR data has 3-hourly resolution which facilitated time matching within 1.5 hours of any event. Brooks et al. (1994) used proximity soundings to evaluate severe weather parameters and limited cases to where a sounding was available within 5 hours and 300 km of the observed severe weather. The NARR allows for the retrieval of closer proximity data in both space and time.

ii. Methodology

Severe weather events were extracted from the severe weather database. Initially, reports within 50 km of State College were collected and used to compare to the NARR variables. This produced too few cases. The severe weather collection method was revised to obtain a more robust dataset. Thus, the search area was expanded to include the Mid-Atlantic region, which comprised of Pennsylvania and its surrounding states.

The Mid-Atlantic severe data set produced 1443 tornadic events, 5366 hail events, and 21,772 wind events. For each event, the time, latitude, and longitude were saved. The event time was rounded to the nearest 3 hour time interval in order to match the NARR data set. NARR data was then extracted for the variables listed in Table 1 to build the training dataset.

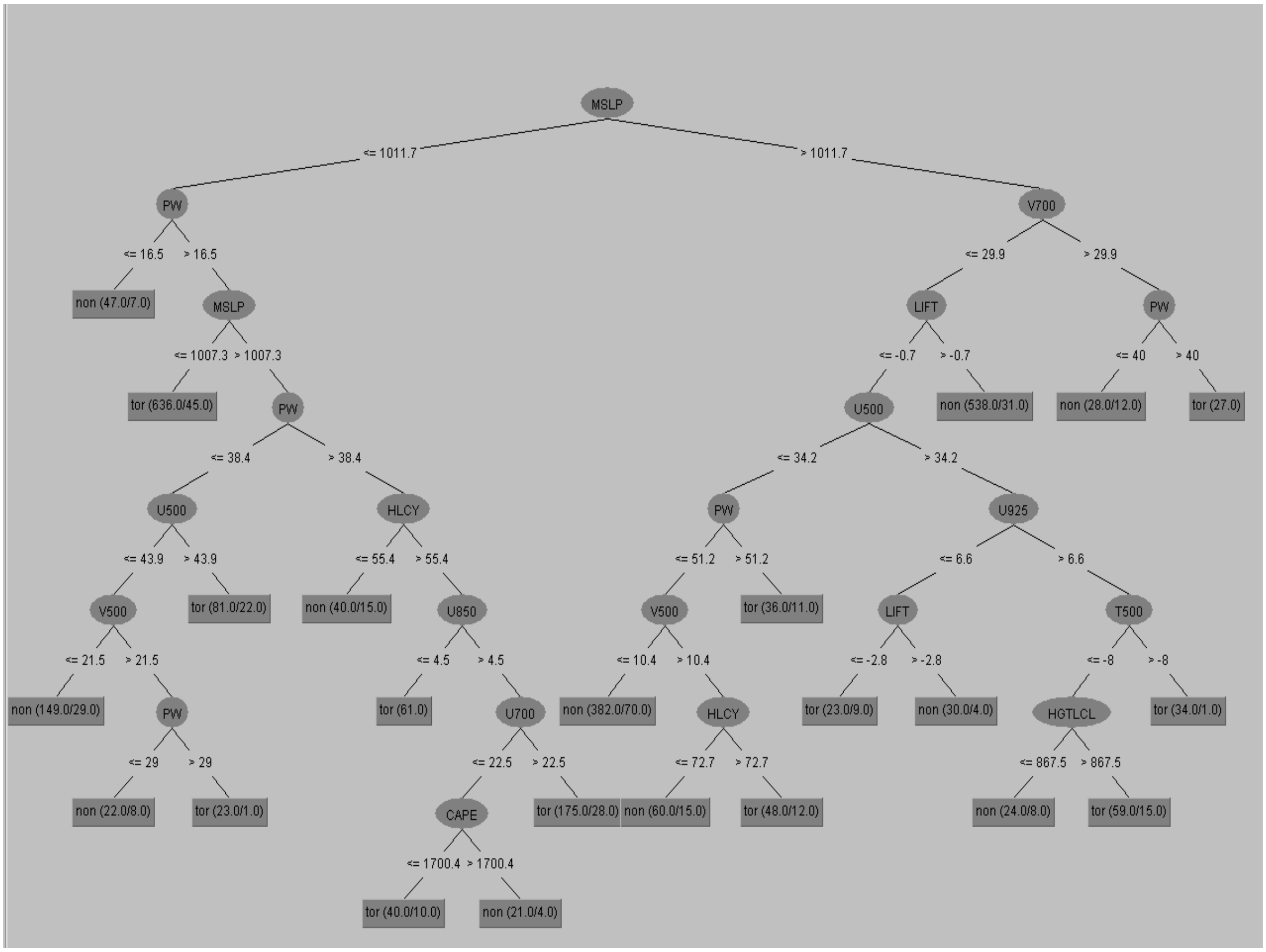

Figure 1. Decision tree for tornadic events produced by WEKA. Each node (ovals and circles) show the variable and the criteria used to produce the decision. Rectangles show terminal decisions for tornadic or non-tornadic events. For each event type forecast, the number of correctly and incorrectly forecast events is shown. For example, for MSLP less than or equal to 1011.7 hPa and PW less than or equal to 16.5 mm 47 non-tornadic events were correctly forecast and 7 events were incorrectly forecast.

The Waikato Environment for Knowledge Analysis (WEKA:Witten and Frank, 2005) was used to develop decision trees and neural networks from the tornado, hail, wind, and combined severe weather dataset WEKA employs Quinlan's C4.5 decision tree algorithm (Quinlan, 1993) and an implementation of a neural network. Several variables in WEKA can be adjusted to increase pruning and to prevent over fitting. This produces a decision tree that is more robust and able to generalize well on new unseen data. It also produces a smaller tree that can easily be turned into an operational program for alerting forecasters to the possibility of severe weather, similar to the system presented by Root et al. (2007). The neural network implementation has similar tunable variables to prevent over fitting.

In order to use WEKA to train on the data a comparable number of non-event days were required. For simplicity, non-event days were obtained by picking times 24 hours prior to each observed event. An improvement would require ensuring no severe weather was observed on that date.

3. RESULTS

Tables 2 & 3 show the skill scores and sizes of the resulting decision trees by severe weather type. The overall results are promising and imply gridded data can be used to alert forecasters to the potential for severe weather on the left side of the tree. The mean sea level

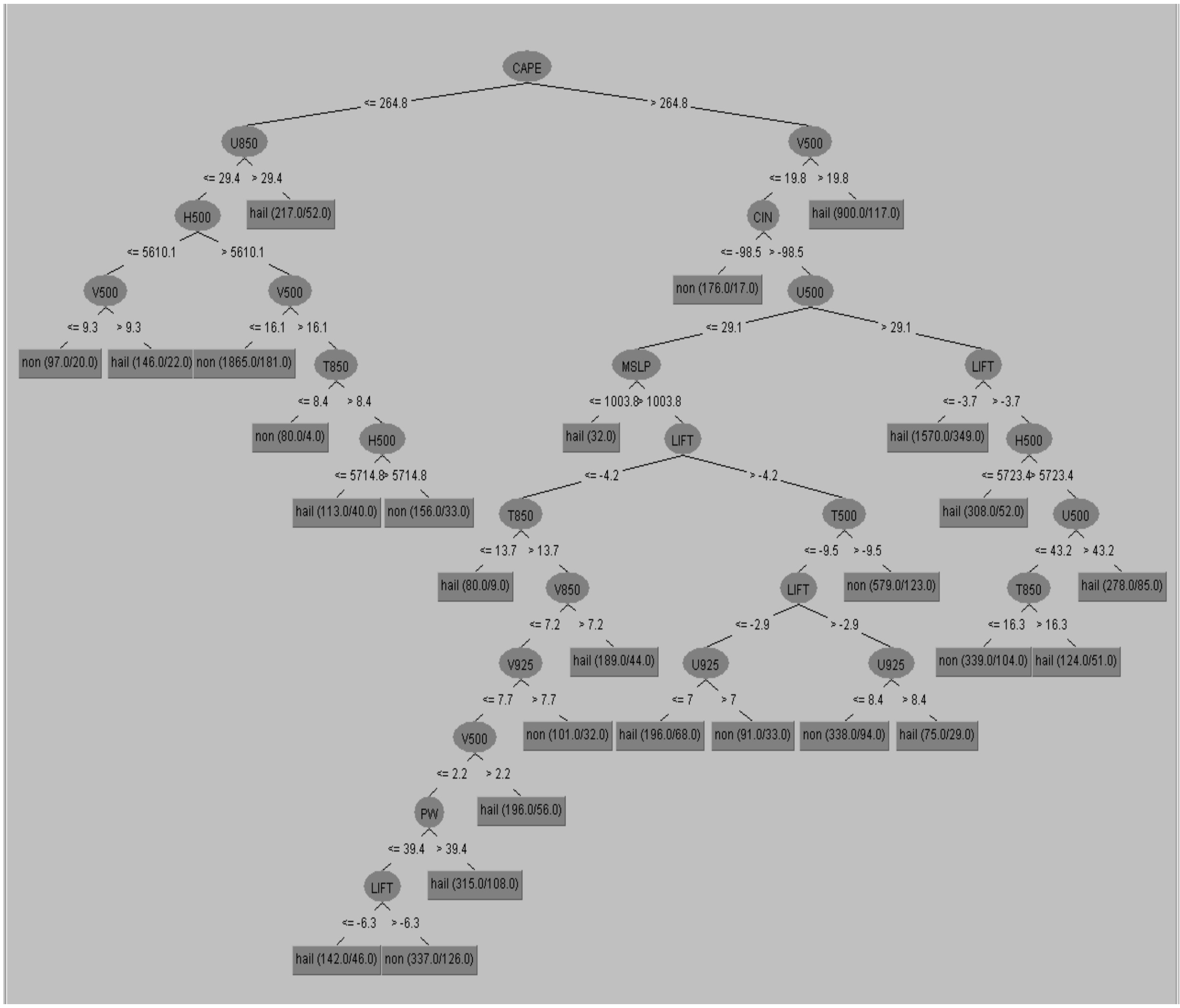

Figure 2 As in Figure 1 except for severe thunderstorms producing hail.

events approximately 80% of the time. The strongest signal was associated with tornadic events which had slightly higher correctly classified events, a higher critical success index (CSI), and the higher CSI for detecting non-events.

The rest of this section will focus on analyzing the decision tree logic for each event to ensure the results truly represent the state of the atmosphere when producing the severe weather.

i. Tornadic events

Figure 1 shows the decision tree for tornado events. The first branch of the tree is based on the value of mean sea-level pressure. The majority of tornadoes were correctly predicted pressure (MSLP) and precipitable water (PW) allowed for quick retrieval of 636 events, based on the combination of surface pressure less than 1007 hPa and PW values greater than 16.5 mm.

Similar to past studies, the 500 hPa winds (Giordano and Fritsch 1990) helped retrieve an additional 104 events with 81 northwest flow events. The helicity, convectively available convective energy (CAPE), and low and mid-level winds helped to identify over 300 additional events.

The right hand branch of the tree, with MSLP greater than 1011.7 hPa excelled at predicting non-tornadic events. Tornadic events were generally much harder to obtain in this portion of the tree. The simplest branch was associated with strong 700 hPa southerlies and high Precipitable Water (PW). This branch identified 27 tornadoes 2 levels down indicating a considerable amount of logic processing was required to capture tornadic events down this branch. Similar to the left side, PW, winds, and helicity played critical roles in describing the tornadic events deep down this tree. This represented a difficult logic progression for a human to make, though it contained many of the signals forecasters often examine when predicting severe weather. For example, the concept of the lifting condensation level (LCL) height (Markowski et al 2000) and its relationship to the strength of the rear flank downdraft (RFD) helped to identify 59 additional tornadic events at such a deep level in the tree.

ii. *Hail events*

The decision tree for Hail events is shown in Figure 2. Unlike the tornado tree, in this tree CAPE was the most influential predictor in distinguishing hail events from non-hail events. It should be noted that the threshold values were quite small (265JKg^{-1}).

The majority of the non-hail events (1865) appear on the left side of the tree with the lower values of CAPE. The hail events that appear on the left side may indicate problems with the severe weather dataset or cooler mid-level conditions under northwest flow. The weak v-wind components at 500 mb, strong u-wind component and the relatively cool 850 hPa temperatures may suggest cooler conditions associated with northwest flow severe weather events.

The right side of the tree contains the majority of hail events (4984) compared to the left side (476). This side of the tree describes an atmosphere of high instability and shear. One level down, the 500 hPa v-wind obtains 900 hail

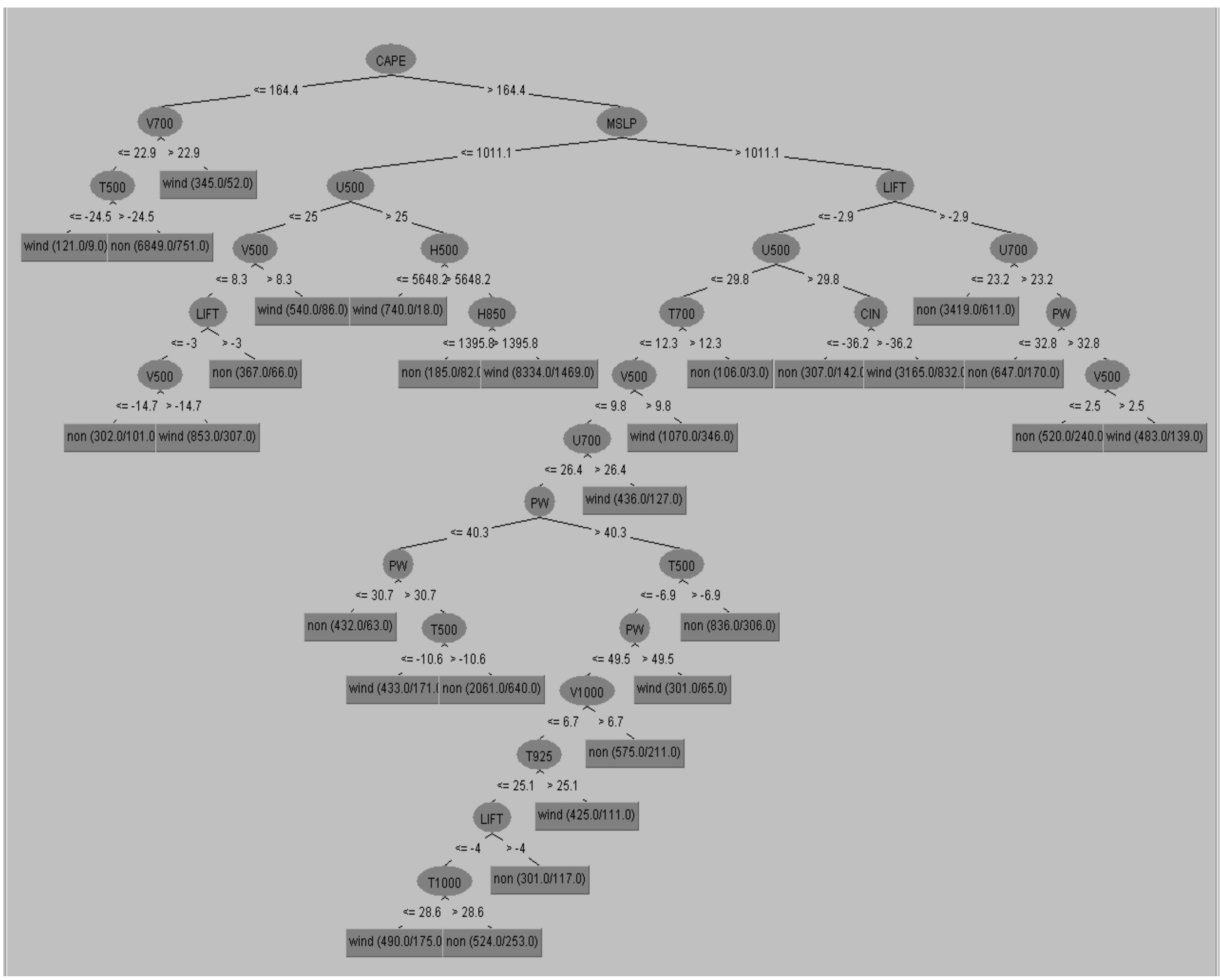

Figure 3 As in Figure 1 except for severe thunderstorm wind events.

Parameter	Level or Layer
Precipitable Water	Column
Mean Sea Level Pressure	Surface
CAPE	Column
CIN	Column
Helicity	Column? Or 0-3km?
Lifted Index	Surface based?
LCL Height	Arbitrary
Geopotential Height	1000, 925, 850, 700, 500 mb
Temperature	1000, 925, 850, 700, 500 mb
Wind U-component	1000, 925, 850, 700, 500 mb
Wind V-component	1000, 925, 850, 700, 500 mb

Table 1. Parameters extracted from the NARR data for each severe weather event and used in the training data set.

events. The majority of hail events (1570) are obtained 5 levels down and were associated with moist, unstable air and strong southwest flow as indicated by the low lifted indices and strong u,v wind components at upper levels.

With weaker winds at the u-wind 500 hPa decision point, the hail events are more difficult to discern. These conditions likely represent the weakly sheared pulse thunderstorm environment.

Type	Number of Instances	Percent Correctly Classified	Percent Incorrectly Classified	Number of Leaves	MAE	RMSE
Tornado	2584	81.9	18.1	23	0.2515	0.3737
Hail	9040	76.4	23.6	27	0.3222	0.4109
Wind	35167	77.2	22.8	29	0.3235	0.4047

Table 2. Size of decision tree and various skill scores for tornadic, hail, and severe wind events for the Mid Atlantic region. Mean Average Error (MAE) and Root Mean Square Error (RMSE) are also shown.

These events require more unstable conditions associated with higher heat and humidity. One portion of the right branch differs somewhat from the rest and leads to a specific type of severe thunderstorm. It shows hail events at the bottom of the tree associated with high instability and deep low-level warmth with LI's less than -4.2 and 850 temperatures > 13.7 C. Following the decision tree beneath these levels reveals the weakly sheared environments associated with true pulse thunderstorms. Moisture and mid-level instability are important indicators of this event type (Medlin, et. al., 2006).

For hail events, the strongly forced events are easily uncovered by WEKA. However, many of the weakly forced events required complicated decisions, ideally suited for an AI alert system.

iii. Convective wind events

Figure 3 shows the decision tree to distinguish between severe convective wind events and non-events. Overall, the key decision is based on CAPE, similar to that found for hail, and the majority of non-events are found on the left branch. This decision tree has a complex right side associated with higher CAPE and a simple left side that quickly identifies a large number of non-events (6849 successfully classified with only 51 non events misclassified) only 2 levels down from the root node.

The right side of the tree defines the majority of severe wind events where CAPE values are higher. There appear to be two types of convective wind scenarios. One branch is associated with a relatively deep low pressure system with MSLP less than 1011 hPa and quickly captures 8334 events when the 500 u-wind is strongly southerly (u-wind > 25), the 500 hPa height is greater than 5648 m and the 850 hPa height is greater than 1395 m. Further down this branch weaker mid and upper level winds combined with very unstable air (LI < -3) hint at the possibility of a pulse severe storm environment.

Type	Event Type	POD	FAR	CSI
Tornado	Tor	0.799	0.161	0.832
	Non-Tor	0.839	0.201	0.807
Hail	Hail	0.773	0.246	0.759
	Non-Hail	0.754	0.227	0.769
Wind	Wind	0.787	0.243	0.764
	Non-Wind	0.757	0.213	0.781

Table 3. Event vs. Non-Event skill scores for tornadic, hail, and severe wind events for the Mid Atlantic region. Probability of Detection (POD), False Alarm Rate (FAR), and Critical Success Index (CSI).

The rightmost branch under the MSLP node shows moist, unstable conditions, and strong winds are required to generate convective winds when surface pressures are greater than 1011 hPa. One branch quickly gets 3165 events when the LI is less than -2.9 and the 500 hPa u-winds are greater than 29.8kts and the CIN is greater than -36.2JKg^{-1}. This branch confirms the idea that most strong severe weather in the Mid-Atlantic Region events occur with strong instability and a convective cap that must be broken in order to release deep-moist convection. With weaker 500 hPa winds, another 1070 events can be obtained with cold 700 hPa temperatures. The remainder of this branch describes an environment for the classic pulse severe storm. Very high heat and moisture at low levels are indicated by the high PW values, Lifted Index, and low level temperatures. At mid and upper levels relatively weak winds and cold temperatures were present. These factors contribute to slow-moving deep, moist convection. Similar useful predictors associated with pulse severe storm environments were also shown in Figure 2.

4. CONCLUSIONS

A decision tree was used to identify atmospheric predictors for 3 types of severe weather in the Mid-Atlantic region: tornadic, hail, and severe winds. Overall, the decisions trees showed meteorologically consistent results and the initial trees suggest some AI logic could aid forecasters in searching out and identifying useful information in complex severe weather situations.

The decision tree appeared to show the most skill at identifying tornadic environments with a CSI of 0.83. This skill is most likely due to the combination of favorable atmospheric parameters that come together to produce a strong signal. Generally, tornadic events in the Mid-Atlantic region result from a strongly forced synoptic environment and thus a strong signal is present during most of these events. The signal is characterized by low pressure; high low-level heat and moisture; strong winds to produce favorable environmental shear and helicity; and a thermodynamic profile that favors deep, moist convection.

Decision tree skill is slightly lower for hail and severe wind events. The tree shows that hail events have two distinct types: strongly forced with high shear and weakly forced, where higher PW and mid-level cool temperatures are needed to identify pulse severe events. These harder to identify events are ideally suited for AI applications to aid forecasters.

The wind events are similar to hail events and are associated with two types of scenarios: a highly sheared environment and a low-shear, highly-buoyant atmosphere. Again, in the lower sheared environment the AI approach may aid forecasters in identifying these events.

Decision trees confirm the current ideas and methodologies long used by operational forecasters in identifying synoptic scale severe weather characteristics. In these cases, the alert system could lend *increased confidence to forecasters* in their identification of a potential severe weather day. What cannot be overlooked is the ability for the decision tree to easily identify non-events. *Some of the non-event logic is elegantly simplistic*. The alert system could be tuned to give forecasters more confidence in a low probability event day.

Subtle nuances for identifying severe events are shown by the more complex decision tree logic. An automated classification and alert system using decision tree logic may prove helpful in identifying severe weather events associated with weak synoptic scale conditions. In addition to the alert for such potential the system should produce some indication of what predictors, not normally used, triggered the alert so that forecasters can investigate the potential further. The decision tree could run on operational model guidance as new model data comes into the forecast office. The output of the decision tree could be used to display a pop-up window at the forecaster's workstation alerting them to the potential of severe weather for that particular forecast valid time. In addition, the output would let forecasters know which predictors and threshold values led to this decision so that they could investigate those predictors further.

5. References

Brooks,H.E., C.A. Doswell III, and J. Cooper 1994: On the environments of tornadic and non-tornadic mesocyclones. *Wea. and Fore.,***9**,606-618.

Brooks, A. R. Anderson, K. Riemann, I. Ebbers, and H. Flachs, 2005: Climatological aspects of convective parameters from the NCAR/NCEP reanalysis. Atmos. Res., submitted.

Doswell, C.A. III (1986): The Human Element in Weather Forecasting, *Nat. Wea. Dig*., **11**, 6-17.

Doswell,C.A III, Harold E. Brooks, and Robert A. Maddox (1996): Flash Flood Forecasting: An Ingredients-Based Methodology *Wea. And Fore*, **11**, 560-581.

Hall, T., H.E. Brooks, and C.A. Doswell III (1999): Precipitation forecasting using a neural network. *Wea .and Forecasting*, **14**, 338-345.

North American Regional Reanalysis. Fedor Mesinger, Geoff DiMego, Eugenia Kalnay, Kenneth Mitchell, Perry C. Shafran, Wesley Ebisuzaki, Dušan Jović, Jack Woollen, Eric Rogers, Ernesto H. Berbery, Michael B. Ek, Yun Fan, Robert Grumbine, Wayne Higgins, Hong Li, Ying Lin, Geoff Manikin, David Parrish, and Wei Shi *Bulletin of the American Meteorological Society* Volume 87, Issue 3 (March 2006) pp. 343–360
DOI: 10.1175/BAMS-87-3-343

Ian H. Witten; Eibe Frank (2005). Data Mining: Practical machine learning tools and techniques, 2nd Edition. Morgan Kaufmann, San Francisco. Retrieved on 2007-06-25.

Markowski P. M., J. M. Straka, and E. N. Rasmussen, 2000: Surface thermodynamic characteristics of rear flank downdrafts as measured by a mobile mesonet. Preprints *20th Conf. on Severe Local Storms,* Orlando, FL, Amer. Meteor. Soc., 251–254

McCaul, E.W. Jr., C. Cohen and C. Kirkpatrick (2005): The sensitivity of simulated storm structure, intensity, and precipitation efficiency to environmental temperature, *Mon. Weather Rev.* **133**, pp. 3015–3037.

Medlin, Jeffrey M., and Jack Cullen (2006): A Theromdynamic Investigation of the Early Afternoon Wet Microburst Pre-Storm Environment over Southern Alabama and the Western Florida Panhandle. *National Wea. Dig.*, **30**, 61-67.

Quinlan, R., 1993. Programs for Machine Learning. ISBN:1-55860-238-0 Morgan Kauffman, 1993

Witten Ian H. and Eibe Frank, 2005. Data Mining. Practicle Machine Learning Tools and Techniques. 2nd Ed. Morgan Kaufmann Publishers. 500 Sansome Street, Suite 400, San Fransisco, CA 94111