Fingerprinting Significant Weather Events

By

Paul Knight[1], Jeremy Ross, Ben Root, George Young,

The Pennsylvania State University

and

Richard Grumm

National Weather Service, State College, PA 16803

1. INTRODUCTION

Forecasting significant weather events continues to pose a challenge to meteorologists. While there are notable success stories of predicting important storms far in advance; such as the 12-14 March 1993 Superstorm and the 6-8 January 1996 Mid-Atlantic snowstorm; many important weather events are still only predicted with short notice. Pattern recognition offers a paradigm for this prediction challenge. The premise of pattern recognition is that significant weather events have repeatable and very specific configurations of atmospheric anomaly fields.

For example, flooding in Virginia is found to occur with specific combinations of anomalous conditions. These patterns include anomalously large values of precipitable water, a maximum of 700 hPa thermal field east of the region, and unusually strong negative values of the 850 hPa u wind component coupled with large positive values of the 200 hPa v wind component along the East Coast. (Knight and Evans 2000)

Once the "fingerprint" for a significant weather event has been identified, this information can be used to assess the risk of a future event based on the configuration of model forecasts of anomaly fields. Several steps are followed to produce this significant weather forecast system. A comprehensive relational database of significant weather events is assembled from existing events (National Climatic Data Center (NCDC), the Storm Predictions Center (SPC) and others). The following weather events are chosen for the domain of the Middle Atlantic River Forecast Center (MARFC); flash flooding, widespread flooding, snow storms, ice storms, non-convective wind storms, record heat and cold, tornadoes, convective winds, hail and forest fires. All events are within the specified region, with the exception of fires where only data from Pennsylvania was available. The events database period of record varied from 1950 to 2003 with most event types within the database covering at least a 40 year span.

This paper addresses key predictors associated with various significant weather events. The first section will detail the criteria used for the events along with their sources. The second section explains an objective methodology used to rank the importance of atmospheric data fields for each event, with a discussion describing the technique of spatially clustering maximum and minimum values of anomalies and ranking their magnitudes. The results section outlines preliminary findings for the future implementation of an early warning system of significant weather events within the MARFC domain.

2. METHODS

i) Events Database

Significant weather event occurrences are typically associated with distinct pattern signatures that can be objectively determined through pattern recognition. The underlying

[1]*Corresponding author address*: Paul Knight
503 Walker Building, University Park, PA 16802

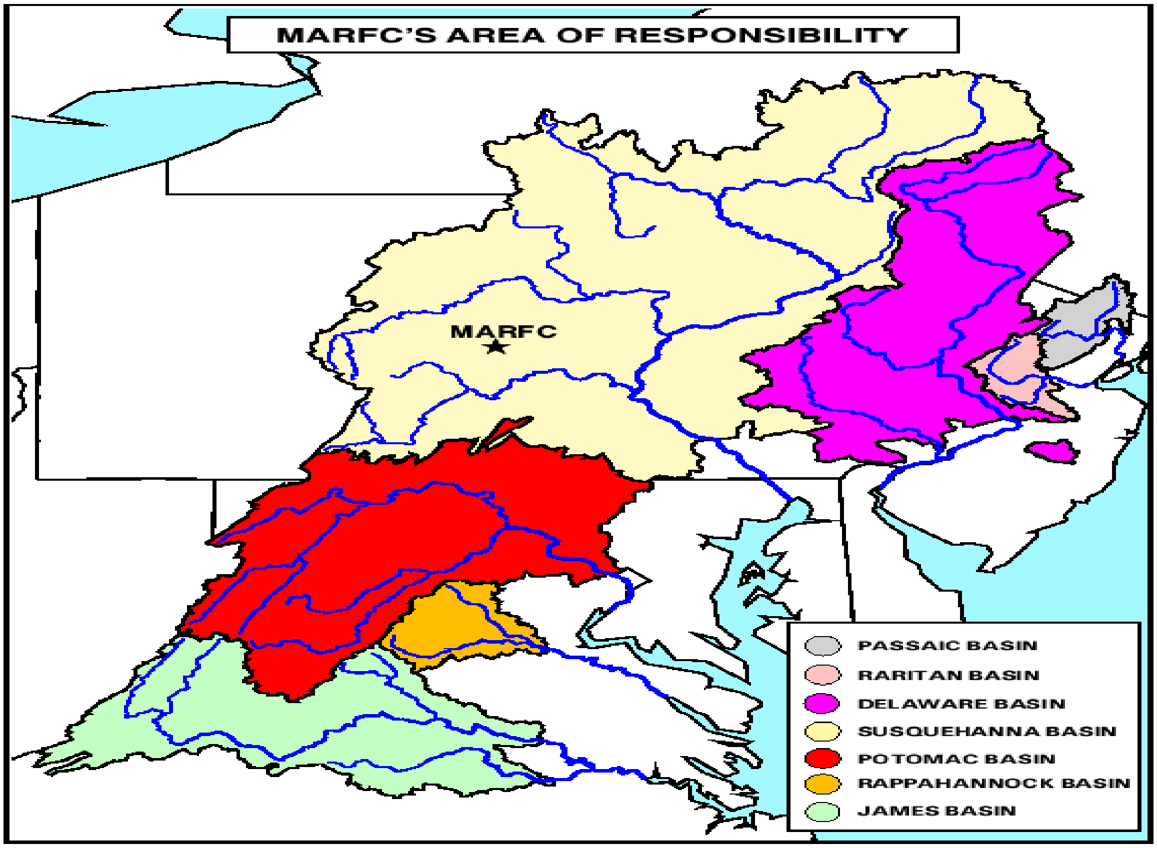

Figure 1. Domain of the Middle-Atlantic River Forecast Center (MARFC).

success of identifying these events and their connecting patterns is largely dependent on the accuracy and sufficiency of available historical datasets. For purposes of studying significant events and weather patterns, it is therefore beneficial to develop relational databases for extreme events and to use historical meteorological data compiled through coherent data assimilation techniques to analyze the associated patterns.

The significant event database is developed using 11 types of significant weather events including: snow storms, ice storms, heat waves, cold waves, tornadoes, convective and non-convective winds, hail, widespread floods, flash floods, and fires. The events are identified for the Mid Atlantic River Forecast Center (MARFC) region (Fig 1), with data obtained from several sources including the regional NOAA climate offices, the SPC severe weather database, and the NCDC storm database. For each event, thresholds of significance are developed subjectively from observational experiences and include the magnitude and area impact of the event across the MARFC domain. Columns 2 and 3 in Table 1 indicate the threshold magnitude and location used to determine each significant event. In addition, the dataset length and the number of events found within that period are given. All researched events and details such as the type, magnitude, location, and date of the event, are compiled into a relational database using the open source software MySQL. The development of relational databases allows for efficient extraction of dates of events based on chosen thresholds and can quickly be altered to accommodate different levels of event significance.

Relational databases are only a portion of the data needed to develop a statistical pattern recognition model. The patterns associated with extreme events must be analyzed from historical datasets with periods as extensive as that from which the date of the significant event occurred. Fortunately, during the decade of the 90's, the National Centers for Environmental Prediction (NCEP) and National Centers for Atmospheric Research (NCAR) cooperated in a project denoted "reanalysis" to produce a long record (from 1948-present) of global historical analyses of key atmospheric fields for the needs of the research and climate communities (Kalnay et al. 1996). Using a reduced resolution version of the NCEP global operational model from 1995, a consistent method of data assimilation and quality control is employed using most available data at time of analysis (land surface, ship, rawinsondes, pibal, aircraft, and satellite).

In spite of the project's best efforts, it is recognized that further increases in available data from 1948 onwards may impact the consistency of the dataset over oceans and other data sparse regions. However, this research focuses on the eastern U.S. where the density and quality of the observing network is sufficiently consistent throughout the reanalysis period, and therefore adequate for our purposes. A higher resolution and more detailed North American regional reanalysis (NARR) project (Mesinger et al. 2004) has been developed and will be transitioned to in future work.

Several important meteorological variables are examined and used from the NCEP/NCAR reanalysis dataset. These include geopotential heights, temperatures, specific humidity, meridional and zonal winds, sea-level pressure, precipitable water, 2m temperature, and 10m meridional and zonal winds. In addition, several other datasets are constructed from the primitive variables to produce thickness, shear, and wind magnitudes at various levels. The reanalysis dataset is global with a horizontal resolution of 2.5° x 2.5° at 17 vertical pressure levels, with a 6-hr temporal resolution beginning from 00Z on 1 January 1948. Since we are interested in the significance of the patterns, a technique similar to that developed by Hart and Grumm (2001) is used to generate and standardize the magnitude of field anomalies in the reanalysis. This is accomplished by first calculating the WMO standard 30 year (1971-2000) climatological mean ($\bar{X}$) and standard deviation (σ) from all NCEP reanalysis and derived fields for each 6-hr period within a calendar year. In addition, a 21-day centered running mean is applied using values valid only at the corresponding 6 hr period (00Z, 06Z, 12Z, and 18Z) for those 21 days. This process preserves the climatological features associated with diurnal cycles. The climatological mean is then subtracted from the observed field (X) and divided by the standard deviation to arrive at a normalized climatological anomaly (N) (Eq 1). This technique is applied for all 6-hr periods from 1948-2004 (~82,000 analyses).

$$N = \frac{X - \bar{X}}{\sigma} \qquad (1)$$

Generating and normalizing the anomalies converts a pseudo normal distribution into a normal standard distribution (Hart and Grumm, 2001). Additionally, the normalization process is beneficial for recognizing extreme events because the magnitude of the standardized anomalies are independent of season and location, both of which are important in the recognition of patterns conducive to producing significant weather events. The compiled datasets are then used in an objective statistical model to examine the significant event and the linked patterns.

ii) Cluster Analysis

Creating the fingerprints for these weather event types is an iterative process. The more event incidents added to the database, the clearer the fingerprints become. As specific events occur, they are entered into an Event-List waiting to be processed. Each significant weather event is represented by the date and time the event started, the name of the significant weather event type (i.e. snow, fire, hail, tornado, etc.), and the latitude and longitude of the event location. Initially, a massive database collection was done in order to build a historical Event-List file. The fingerprints of the anomalies are determined by processing the Event-List with a Perl script that employs a GrADS (Doty and Kinter 1992) library function. The GrADS script determines the anomalies, by variable and level over the domain. Anomalies are accumulated for each event based on the date, time, and location of the anomaly.

Based on previous research (Hart and Grumm 2001), a list of pre-determined variables and data levels are used to analyze each type of weather event. GrADS is used to locate the primary and secondary "Peaks" and "Valleys" in the gridded standard anomaly data for each variable and level. The location of a "Peak" is defined as the location where the standard anomaly values of all the gridpoints surrounding the location are less than or equal to the standard anomaly value at that point. Conversely, the location of a "Valley" is based on the same test, but for a local minimum. The secondary peaks and valleys are determined using the same logic. If no peaks or valleys are found for the variable and level at a time, or only one peak or valley is found, the values were assigned as null. The vector (distance and direction) of the peak or valley relative to the specified latitude and longitude of the observed significant weather event are then calculated and included in the relational database for that event.

Clustering algorithms are then used to discern and rank anomalies based on their magnitude and location for each significant weather event. The clusters are discerned using a method called "Strong Point Analysis" developed for this project. The method determines which grid points are likely members of a cluster. The technique takes into account both the spatial extent and the total number of occurrences in the determination of whether a point is part of a cluster. A cluster is defined as a network of strong points with neighboring weak points; that is points with weaker values surrounding it.

In order to use these clusters for forecasting, an understanding of the importance of each field to the weather event type is necessary. To accomplish this, several analysis techniques are employed. A field strength analysis is developed which assesses the overall value of a data field as a potential predictor. The overall value of a data field as a potential predictor is broken down into two parts. *Alpha* (α) is the measure of the cluster's spatial signal for the field while Phi (ψ) is the measure of the cluster's standard anomaly signal for the field.

In the proposed operational forecasting of these significant weather events, a program is implemented that acquires the top two peaks and the bottom two valleys for each field available from the ETA forecasts initialized at 0000, 0600, 1200, and 1800 UTC. These peaks and valleys are determined across the eastern two-thirds of the United States.

3. RESULTS

i. COLD EVENTS:

Table 2 lists the parameters with the strongest signals in predicting unseasonably cold events in the MARFC domain. The

Predictor	ALPHA	PHI
PWAT_Valley1	0.823633	0.598655
SpecHum_925_Valley1	0.815965	0.518204
Uwnd_850_Valley1	0.809627	0.525457
SpecHum_925_Valley2	0.782571	0.680353
PWAT_Peak1	0.776741	0.804199
Temp_700_Valley1	0.753086	0.375572
Thk_1000_850_Valley1	0.731907	0.43831
Uwnd_850_Valley1	0.725943	0.508321
PWAT_Valley2	0.725926	0.752453
Uwnd_850_Peak1	0.72381	0.710007
SpecHum_925_Peak2	0.713233	0.44445
SpecHum_925_Peak1	0.710769	0.482851
Temp_850_Valley1	0.703549	0.47969
Vwnd_850_Peak1	0.70119	0.391523
Thk_1000_500_Valley1	0.698183	0.364074
MSLP_Peak1	0.687965	0.540658
Hgt_500_Valley1	0.667454	0.367081

Table 2. Key predictors for events of record or unseasonably cold weather. Data include the predictors, the strength of the field with respect to the spatial cluster (Alpha), and the strength of the field with respect to the clusters standard anomaly (PHI).

parameters for cold events suggest that atmospheric moisture has the most utility, with a precipitable water minimum (PWAT_VALLEY1) showing the strongest signal. Directly related to PWAT, the 925 hPa specific humidity shows a similarly strong signal. There is also a strong signal in the 850 hPa U-wind values. Thermal fields, such as the 700 hPa and 850 hPa temperature negative anomalies (valleys) and 1000-850 hPa thickness minimums were also useful predictors of significant cold outbreaks.

The key values are listed in order of Field strength. With the exception of below normal U-wind values, moisture related variables appear to dominate the signal. The mean sea-level pressure peak (indicative of high pressure) made the list but was less reliable than most of the moisture and low-level thermal fields.

The implication of moisture (dry air) and the U-winds suggests that local effects, with favorable radiational cooling conditions, are

Predictor	ALPHA	PHI
Thk_1000_850_Peak1	0.809891	0.772666
SpecHum_925_Peak2	0.804613	0.735071
PWAT_Valley2	0.801655	0.51474
SpecHum_925_Valley1	0.79146	0.580606
PWAT_Peak1	0.780852	0.439101
Temp_850_Peak1	0.779502	0.798009
Uwnd_850_Peak1	0.769848	0.612459
SpecHum_925_Peak1	0.76455	0.286129
Thk_1000_500_Peak1	0.752228	0.8107
Vwnd_850_Peak1	0.751333	0.6496
Uwnd_850_Valley1	0.746448	0.604293
SpecHum_925_Valley2	0.746448	0.379143
Hgt_500_Peak1	0.724095	0.775329
Temp_700_Peak1	0.716909	0.809757
MSLP_Peak1	0.703362	0.504203

Table 3. As in Table 2 except for key predictors associated with record heat or unseasonably warm events.

the dominant pattern regimes during cold outbreaks. Along with the presence of a high pressure, it is important that a weak pressure gradient exist to allow for decoupling of the boundary layer during cold outbreaks.

ii. HEAT EVENTS

Important predictors for heat waves are shown in Table 3. Similar to cold events, moisture variables appear to dominate the list of more effective predictors. The 1000-850 hPa thickness maximum was a strong predictor as was the 850 hPa temperature maximum. Other features often associated with heat waves, such as strong surface anticyclones also showed some skill as forecast predictors of significant heat events.

Figure 2 shows the location of the maximum PWAT anomaly associated with heat waves. These data imply that during heat events, above normal PWAT tends to be observed to the north and west of the domain. The larger concentric circles show grid points where the anomaly maxima have occurred more than once.

Predictor	ALPHA	PHI
Uwnd_850_Valley1	0.806705	0.392421
PWAT_Peak1	0.802015	0.488817
SpecHum_925_Peak1	0.783837	0.508674
SpecHum_925_Valley1	0.768681	0.732768
SpecHum_925_Valley2	0.759259	0.692313
Uwnd_850_Peak1	0.755083	0.489589
PWAT_Valley1	0.751493	0.80692
MSLP_Valley1	0.74969	0.485518
Vwnd_850_Peak1	0.731244	0.554725
Vwnd_850_Valley1	0.709376	0.480579
Temp_850_Valley1	0.707483	0.366164
Uwnd_250_Valley1	0.705622	0.415792
Vwnd_250_Peak1	0.703191	0.450226
PWAT_Valley2	0.701692	0.604345
Uwnd_850_Valley2	0.700785	0.428762
SpecHum_925_Peak2	0.700034	0.419009
Uwnd_250_Peak1	0.660649	0.531551
MSLP_Peak1	0.640502	0.318435

Table 4. As in Table 2 except for key predictors associated with snow events.

iii. SNOW EVENTS

Key predictors for snow events are given in Table 4. Moisture variables continued to dominate; however the most coherent signal appeared to be the 850 hPa U-wind valley (east wind anomaly). Such a feature is common during large winter storms as easterly upslope flow enhances lifting and precipitation. Coupled with the dynamical and mechanical lifting, ample low-level moisture is widespread during significant snow events. This is indicated by PWAT anomalies with values comparable to the 850 hPa U-wind anomalies.

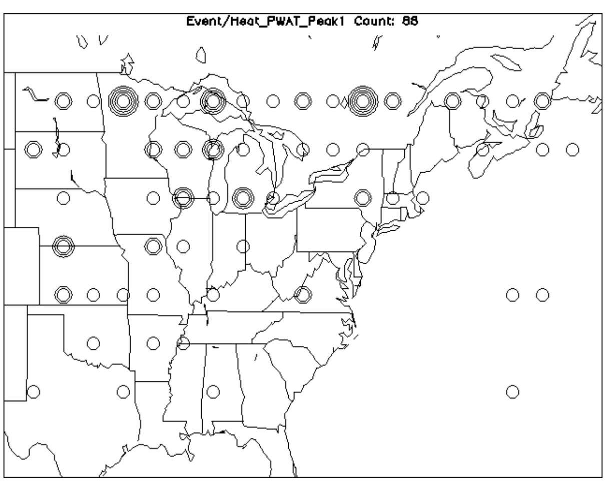

Figure 2. The location of the first maximum anomaly of precipitable water during a significant heat event. Circles show the location. Concentric circles show multiple occurrences of an anomaly at the same grid point.

The MSLP valley also provided some signal suggesting some significant snow events are associated with a strong and deep surface cyclone. The MSLP positive anomaly, an anchor in forecast lingo for the requisite cold air did not show as strong of a signal. These data suggest that not all heavy snow events require a significant surface cyclone, sometimes they occur with large high pressure systems. The presence of a signal in both 850 and 250 hPa wind anomalies shows the importance of the jet circulations in the area of interest. The fact that the first 250 hPa wind variable is a negative anomaly in the U-wind component might imply that slow moving systems produce heavy snow.

The position of the strongest negative MSLP anomalies with snow events is shown in Figure 3. These data show the tendency of negative pressure anomalies to occur south and east of the MARFC region, typically east of the Delmarva coast.

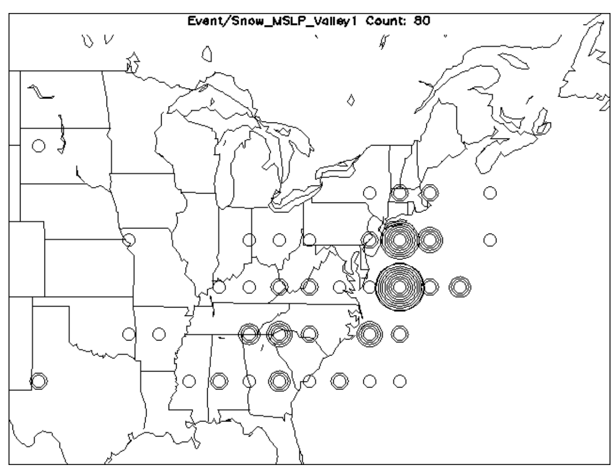

Figure 3. As in Figure 2 except for surface cyclone negative anomaly locations during heavy snow events.

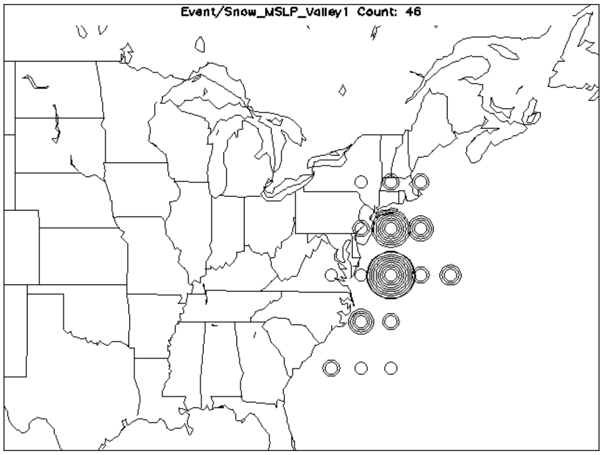

Figure 4. As in Figure 3 except clustered locations of negative pressure anomalies during heavy snow events.

iv. ICE EVENTS

The anomaly parameters for ice events are not shown. These data were very similar to those associated with snow event with moisture variables dominating as the key predictors. The only significant differences were the emergence of a strong 850 hPa V-wind positive anomaly as the third strongest predictor.

Predictor	ALPHA	PHI
Vwnd_700_Peak1	0.814746	0.588234
Vwnd_250_Peak1	0.784897	0.517484
MSLP_Valley1	0.766212	0.490729
Vwnd_850_Peak1	0.753168	0.543049
Uwnd_850_Peak1	0.750092	0.478905
Vwnd_850_Valley1	0.74076	0.46708
Uwnd_850_Valley1	0.740201	0.385135
Uwnd_250_Peak1	0.737903	0.489141
Uwnd_700_Valley1	0.724929	0.447031
Uwnd_700_Peak1	0.704768	0.635345
Vwnd_700_Valley1	0.698293	0.559381
Vwnd_250_Valley1	0.697926	0.55995
Hgt_500_Valley1	0.688378	0.402538
Uwnd_250_Valley1	0.684416	0.497832
MSLP_Peak1	0.611081	0.363993

Table 5. As in Table 2 except for key predictors associated with non-convective high wind events.

v. HIGH WIND EVENTS

Table 5 shows the key predictors associated with non-convective high wind events. High wind events appeared to be the only event where moisture variables were not significant identifiers. Not surprisingly, low-level wind anomalies were critical in forecasting these events. The MSLP valleys and peaks making the list may confirm the known connection between the isallobaric wind component and high wind events. The event data exposes the critical role of couplets of wind anomalies of opposite values which are indicative of a highly amplified pattern.

vi. TORNADOES

Table 6 exhibits the key predictors associated with tornadoes. Due to the mesoscale nature of these events, it is not surprising that the overall value of strength field is lower than with other fields. This tendency was also found for hail events (not shown). Perhaps stratifying based on weak, strong and severe tornadoes might have produced a different signal. Similar to other event types, moisture variables tended to dominate. Strong low-level winds and strong wind shear in the surface to 850 and 700 hPa layers also proved to be valuable predictors of these events. The high moisture value is likely related to the high CAPE often associated with tornadic events. The presence of mid-level low moisture values, as indicated by the 700 hPa valley in specific humidity, may suggest the need for some dry air intrusion into systems which produce tornadoes.

The MSLP valley was also a predictor. This variable was originally used due to the association of strong surface cyclones moving into the Great Lakes in association with many large tornado outbreaks. The strong signal may validate this casual observation.

4. CONCLUSIONS

A study is conducted with the goal of augmenting the use of climatic anomalies to forecast significant weather events. A database is developed to categorize and obtain information on events such as heavy snow, ice, high wind, extreme heat and cold, and tornadic events. These events are then used to produce a database of climatic anomalies from the NCEP-NCAR reanalysis data.

Preliminary results suggest that certain climatic anomalies may provide a fingerprint to help predict the occurrence of these events when the method is applied to real-time prognostic model data. With the exception of high wind events, the moisture variables appear to be critical predictors in all event types. Clearly, above normal moisture plays a role with events associated with precipitation.

Both heat and cold events are associated with anomalous moisture. During cold events below normal moisture appears to be a critical predictor. With light winds, this parameter indicates the importance of lowering the heat capacity of the air and also modulating the long-wave radiation energy budget. Several thermal variables, such 850 hPa temperatures and 1000-850 hPa thickness were also useful in predicting these events. During warm episodes, above normal moisture and above normal low-level temperatures were critical predictors.

Snow and ice events showed signals involving moisture and anomalous low-level wind fields. The anomalously low 850 hPa U-wind showed the importance of easterly flow during snow events. For ice events, the low-level anomalous southerly flow was also a critical factor.

Overall, the signal for tornadic events was not as strong. Based on the mesoscale nature of these events this is not surprising. However, moisture, wind, and shear anomalies did show promise as predictors for these types of events.

Predictor	ALPHA	PHI
PWAT_Peak1	0.673448	0.653191
SpecHum_700_Peak1	0.666451	0.698335
Vwnd_850_Peak1	0.659182	0.541871
MSLP_Valley1	0.659061	0.4257
Uwnd_850_Peak1	0.655215	0.669678
Shear_850_Peak1	0.652305	0.570545
SpecHum_700_Valley1	0.643825	0.627973
PWAT_Valley1	0.627323	0.624699
Uwnd_850_Valley1	0.616372	0.567247
Shear_700_Valley1	0.612302	0.647669
Vwnd_850_Valley1	0.599825	0.542428
Shear_700_Peak1	0.599087	0.677639
Vwnd_300_Peak1	0.596827	0.594354

Table 6. As in Table 2 except for key predictors associated with tornado events.

These data and results suggest that a method using climatic anomalies could be used to predict future events of each type. The combination of the anomaly value and its location from a numerical model could be used to assess the risk of a specified event type.

5. ACKNOWLEDGEMENTS

This work was funded in part by UCAR COMET partnership grant S03-44672.

6. REFERENCES

Doty, B. and J.L. Kinter III, 1992: The Grid Analysis and Display System (GrADS): A practical tool for earth science visualization. *Eighth International Conference on Interactive Information and Procession Systems*, Atlanta, Georgia, 5-10 January, 1992.

Knight, P. G. and Evans, M. E., 2000: Prediction of excessive rainfall in the Middle Atlantic region. *Preprint for Extreme Precipitation Conference.*

Hart, R. E., and R. H. Grumm, 2001: Using Normalized Climatological Anomalies to

Rank Synoptic-Scale Events Objectively. *Mon. Wea. Rev.,* **109**, 2426–2442.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, B. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K.C. Mo, C. Ropelewski, J. Wang, R. Jenne, D. Joseph, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc*., **77**, 437–472.

Mesinger, F., G. DiMego, E. Kalnay, P. Shafran, W. Ebisuzaki, D. Jovic, J. Woollen, K. Mitchell, E. Rogers, M. Ek, Y. Fan, R. Grumbine, W. Higgins, H. Li, Y, Lin, G. Manikin, D. Parish, and W. Shi, 2004: NCEP North American Regional Reanalysis, 15th Symp.On Global Change and Climate Variations, Seattle, WA, 11-15 Jan 2004.

Significant Event Type	Criteria	Location	Dataset Period	# Events
Snow Storms	≥ 4-6" 2-day Total	1 of 14 Cities	11/6/1953 - 2/17/2003	189
Ice Storms	≥ ¼ Ice Accretion	Counties	1/26/1961 - 4/9/2003	1400
Heat Waves	2 Consecutive Record Highs	2 of 10 Cities	1/24/1950 - 8/19/2002	245
Cold Waves	2 Consecutive Record Lows	2 of 10 Cities	1/30/1948 - 5/22/2002	222
Convective Winds	≥ 10 Reports of winds > 50kts	Counties	1950-2003	100's
Tornadoes	≥8 F0-F1 ≥6 with ≥ 2 F2	Counties	7/5/1950 – 9/24/2001	1568
Hail	10 Reports ≥ ¾ " Diameter	Counties	3/5/1955 – 3/17/1996	2078
Floods	≥ 2" in 24 Hrs	≥ 1 Basin	1955-2001	~2,800
Flash Floods	≥ 2" in 6 Hrs	≥ 1 Sub-basin	1950-2003	~500
Fire	≥ 100 Acres	≥ 1 Occurrence	1970-2001	275
Non-Convective Winds	> 2 reports of winds > 50kts	2 of 14 cities	1967-2003	~250

Table 1. Event types used in the data base, the criteria used to define an event, the method to identify the location of the event type, the period of record of each event-type data base and the number of events used for this study.