Extended Abstract
27th Conference on Interactive Information Processing Systems (IIPS)
Amer. Meteor. Soc., Seattle, WA, January 23-27, 2011

A SPATIOTEMPORAL ASSESSMENT OF TORNADO WARNINGS WITHIN STORM PREDICTION CENTER CONVECTIVE OUTLOOKS USING GEOGRAPHIC INFORMATION SYSTEMS

Darrel M. Kingfield[*,1,2,3], May Yuan[3], Harold E. Brooks[4], and Michael B. Richman[5]

[1] Cooperative Institute for Mesoscale Meteorological Studies, Norman, Oklahoma
[2] NOAA/NWS Warning Decision Training Branch, Norman, Oklahoma
[3] Geoinformatics Program, University of Oklahoma, Norman, Oklahoma
[4] NOAA National Severe Storms Laboratory, Norman, Oklahoma
[5] School of Meteorology, University of Oklahoma, Norman, Oklahoma

1. Introduction

The National Weather Service (NWS) is the only government organization tasked with the issuance of tornado forecasts and warnings nationwide. The Storm Prediction Center (SPC) is the primary agency tasked with forecasting the risk of organized severe convective storms and its associated hazards (wind, hail, tornadoes). Days in advance of a potential severe event, the forecasting process begins with environmental evaluations and numerical modeling on multiple spatial scales (Kain et al. 2003). An output of this process is the “SPC Outlook”, a publically disseminated product which provides guidance on the spatiotemporal locations of high-impact weather. The product provides both categorical risk classes and probabilistic risk percentages for tornadoes, wind, and hail (Table 1).

Outlook Probability	Tornado Risk	Wind Risk	Hail Risk
2%	SEE TEXT	Not Used	Not Used
5%	SLGT	SEE TEXT	SEE TEXT
10%	SLGT	Not Used	Not Used
15%	MDT	SLGT	SLGT
30%	HIGH	SLGT	SLGT
45%	HIGH	MDT	SLGT/MDT
60%	HIGH	MDT/HIGH	MDT

Table 1: SPC Probabilistic to Categorical Outlook Conversion (from NWS, 2010)

NWS Weather Forecast Offices (WFOs) will use this information, alongside local forecasting techniques, to prepare for the event. If warranted, the WFO will issue severe weather warnings for specific storm-based hazards. One such meteorological hazard which has its own warning category is the tornado.

Tornado warnings are one of the most focused spatiotemporal products issued by the NWS. These warnings can encompass a county or a fraction of one and last up to an hour. Each of the 122 WFOs is tasked with issuing warnings to their domain of responsibility or county warning area (CWA). A WFO can issue a tornado warning if there is “radar or satellite indication and/or reliable spotter reports of a tornado” (NWS, 2005). Since October 1st, 2007, tornado warnings, along with other NWS short-fuse warnings, have been issued for the immediate threat area. This replaces a decades-old warning system that used geopolitical boundaries as its perimeter.

Using Geographic Information Systems (GIS), spatial characteristics from these forecaster-defined warnings can be used to develop associations between warnings and other external factors, such as physical observations (e.g. storm size) or human factors (e.g. forecaster uncertainty). In this paper, tornado warning area, storm motion, and performance metrics will be compared with SPC tornado probabilities and the resulting trends will be summarized.

2. Data & Methods

A two-year data set, 12z on 1 January 2008 to 12z on 1 January 2010, was selected

* Corresponding author address: Darrel Kingfield, WDTB/CIMMS, 120 David Boren Blvd., Suite 2640M, Norman, OK, 73072 Email: Darrel.Kingfield@noaa.gov

to: (1) provide a sufficient sample of polygon-based tornado warnings, and (2) compare two years with a different frequency of tornado activities. In total, 7923 tornado warnings, 3228 tornado events, and 3645 SPC Day 1 Outlooks were issued during this period and used in the analysis.

2.1 Plotting the Dataset

The SPC website contains an archive of their Day 1 Outlook text products. These files contain a …TORNADO… probabilistic section with the vertex points for each tornado outlook class. These points were extracted from each forecast period and plotted as polygons (Figure 1). The two year dataset provided five possible hazard categories for plotting: 2%, 5%, 10%, 15% and 30%; there were no 45% or 60% probability groups.

The NWS Performance Management website maintains a detailed record of both tornado warnings and events. Each tornado warning text product contains a LAT…LON… tag with the vertex points for each forecaster defined polygon. A TIME…MOT…LOC tag in the warning contains information on the direction, motion, and "current location" of the warned feature.

In order to minimize a warning being within two or more SPC risks, a method for simplifying the polygon down to a single point was implemented. For each warning in the dataset, the vertex points from the LAT…LON… tag were plotted along with the "current location" point in a single domain (Figure 2). If the "current location" point was inside the polygon, this point would define the warning location (Figure 2a). If the "current location" point was outside the polygon, a line between this point and the polygon center was drawn; the intersection point defines the warning location (Figure 2b).

Tornado events were also plotted as point features based off the location of first reported touchdown. The event data were used to both verify tornado warnings and provide the location of missed events.

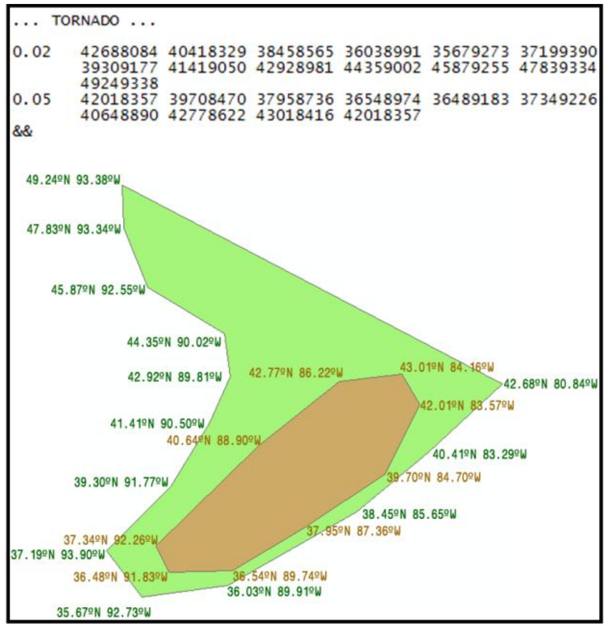

Figure 1: SPC tornado probability text to polygon conversion

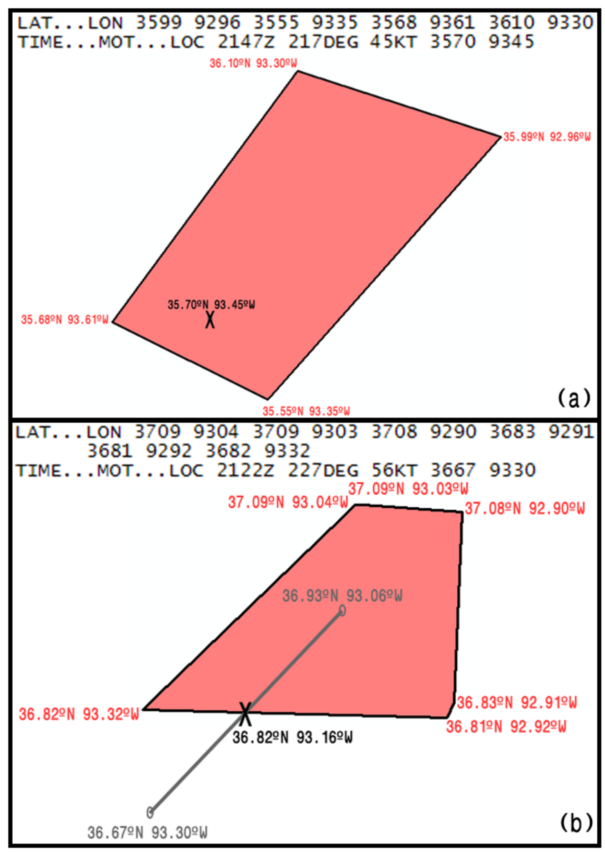

Figure 2: Classification of warning point (marked by an **X**) when **(a)** point is within warning polygon **(b)** point is outside the warning polygon

2.2 Classifying the Dataset

In order to identify what warnings were issued within the SPC risk classes, each warning point was plotted against the current SPC Outlook forecast valid at that time. We assumed the warning decision was made based off the most recently issued SPC Outlook as this would provide the most up-to-date information on the recent/forecasted state of the atmosphere. Currently, there are up to five possible Day 1 Outlooks valid at any time for a given day (Table 2).

Issuance Time (UTC)	Valid Time (UTC)	Valid Period
0600	1200 Day 1 to 1200 Day 2	24 Hours
1300	1200 Day 1 to 1200 Day 2	23 Hours
1630	1630 Day 1 to 1200 Day 2	19.5 Hours
2000	2000 Day 1 to 1200 Day 2	16 Hours
0100	0100 Day 1 to 1200 Day 2	11 Hours

Table 2: Issuance time, valid time and valid period for SPC Day 1 Outlooks (from NWS, 2010)

In addition to the five possible SPC tornado probability percentage groups within which a warning could be located, two additional classification groups were added based on the proximity of warnings that fell outside of all risks. A distance of 250 miles from the 2% tornado risk was chosen to distinguish between these two classes. In summary, a warning could fall in one of seven classes: greater than 250 miles from risk, within 250 miles from risk, 2%, 5%, 10%, 15%, or 30%.

Warning motion and area were other pieces of information paired with the warning classification to be summarized. Warning motion was grabbed from the warning text. Warning areas were calculated after plotting the tornado warning polygons with an Alber's Equal Area projection in ArcGIS.

2.3 Utilizing Performance Metrics

The NWS computes three measures of forecast skill, based off a 2x2 contingency table (Table 2), to quantify warning efficiency and performance after an event (Dept. of Commerce, 1998). These measures are: Probability of Detection (POD), False Alarm Rate (FAR), and the Critical Success Index (CSI). Detailed definitions and usage examples of these metrics can be seen in previous work (Donaldson et. al, 1975; Doswell et. al, 1990; Schaefer, 1990; Polger et. al, 1994).

		OBSERVED	
		YES	NO
FORECAST	YES	A	B
	NO	C	D

Table 3: A 2x2 contingency table for forecasted vs. observed events

The POD is the percentage of events that are forecast. This is the ratio of number of warned events (A) over the total number of observed events (A+C).

$$POD = \frac{A}{A+C} \quad (1)$$

The FAR is the percentage of forecasts that did not have an event associated with it. It is a ratio of the number of false alarms (B) over the total number of forecasts (A+B).

$$FAR = \frac{B}{A+B} \quad (2)$$

The CSI is a function of both the POD and FAR; defined as the quotient of the number of warned events (A) over the total number of false alarms (B) and events (A+C).

$$CSI = \frac{A}{A+B+C} \quad (3)$$

Counts for these metrics were obtained by plotting warning polygons and tornado event points in the same spatiotemporal domain. Overlap would count as a correct forecast (A); no overlap would result in either an incorrect forecast (B) or missed event (C). With warnings placed in different sized domains of varying tornado probability, defining what is the correct forecast of a null event (D) was left outside the scope of this project.

3. Results

3.1 *Warning Area by SPC Risk*

The data suggest that a warning issued in subsequently higher SPC tornado risks covers a larger area (Figure 3). From the dataset, the upper 20% of tornado warnings had a minimum area of: 420 square miles (sq. mi.) if it is greater than 250 miles from any risk, 500 sq. mi. if within 250 miles from any risk, 496 sq. mi. inside a 2% risk, 567 sq. mi. inside a 5% risk, 590 sq. mi. inside a 10% risk, 675 sq. mi. inside a 15% risk, and 720 sq. mi. inside a 30% risk.

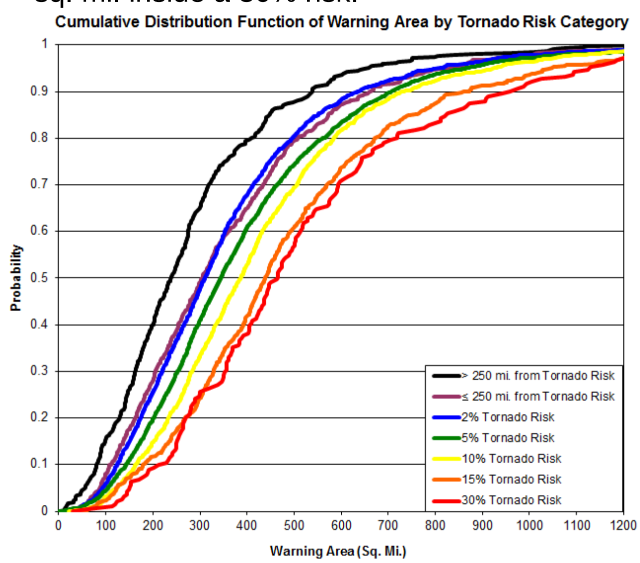

Figure 3: Cumulative Distribution Function plot of warning areas within each of the seven possible risk classification categories

To determine if the larger warning areas by group are statistically significant, statistical permutation testing was used. The null hypothesis (H_0) states that warning polygon areas were all the same between groups and the alternative hypothesis (H_1) states that warning polygon areas were significantly different between groups. Two groups were analyzed at a time with their differences in sample means compared through 100,000 iterations (Table 4). Of the 21 possible permutations, 19 yielded a p-value less than

Comparison Groups	Permutation p-value	Comparison Groups	Permutation p-value	Comparison Groups	Permutation p-value
> 250 mi. vs. ≤ 250 mi.	<0.001	≤ 250 mi. vs. 5%	0.008	2% vs. 30%	<0.001
> 250 mi. vs. 2%	<0.001	≤ 250 mi. vs. 10%	<0.001	5% vs. 10%	<0.001
> 250 mi. vs. 5%	<0.001	≤ 250 mi. vs. 15%	<0.001	5% vs. 15%	<0.001
> 250 mi. vs. 10%	<0.001	≤ 250 mi. vs. 30%	<0.001	5% vs. 30%	<0.001
> 250 mi. vs. 15%	<0.001	2% vs. 5%	<0.001	10% vs. 15%	<0.001
> 250 mi. vs. 30%	<0.001	2% vs. 10%	<0.001	10% vs. 30%	<0.001
≤ 250 mi. vs. 2%	0.833	2% vs. 15%	<0.001	15% vs. 30%	0.137

Table 4: Summary table of permutation test p-values when comparing sample means of warning area

0.01, rejecting the null hypothesis at the 99% confidence interval and showing statistical significance in warning area between groups.

There were two comparisons of note which failed at the 99% confidence interval. First, comparing warnings within the 15% tornado risk (N=740) versus the 30% tornado risk (N=192) returned a p-value of 0.137, rejecting the null hypothesis at the 85% confidence interval. This could be due to insufficient sampling as warnings within the 30% only made up 2.5% of the overall studied warnings. Second, comparing warnings less than 250 mi. from the nearest risk (N=578) versus the 2% tornado risk (N=1705) resulted in a p-value of 0.833. Further analysis of the 578 warnings within 250 mi. of the 2% risk found that 64% of them were less than 50 mi. from the 2% risk. With a majority of the warnings on the border between the two groups in this dataset, it is expected to find that the mean area between these groups would not be significantly different.

With the data suggesting that there is a statistically significant increase in warning area with increasing tornado probability, can another relationship be defined to explain this observation? One testable hypothesis with the current dataset is the role of storm motion. In practice, faster storms would require larger polygons to cover a valid time period of 30 to 60 minutes.

3.2 *Warning/Storm Motion by SPC Risk*

The data suggest that warnings issued in subsequently higher SPC tornado risks have greater storm motions associated with them (Figure 4). From the dataset, over 90% of tornado warnings issued from inside a 30% tornado risk have a storm motion greater than

30 knots (kts). Compared to the other categories, storm motions of 30 kts can be seen in: 80% of warnings in the 15% tornado risk, 65% of warnings in the 10% risk, 52% of warnings in the 5% risk, 37% of warnings in the 2% risk, 27% of warnings within 250 miles of a tornado risk, and 11% of warnings greater than 250 miles from any risk.

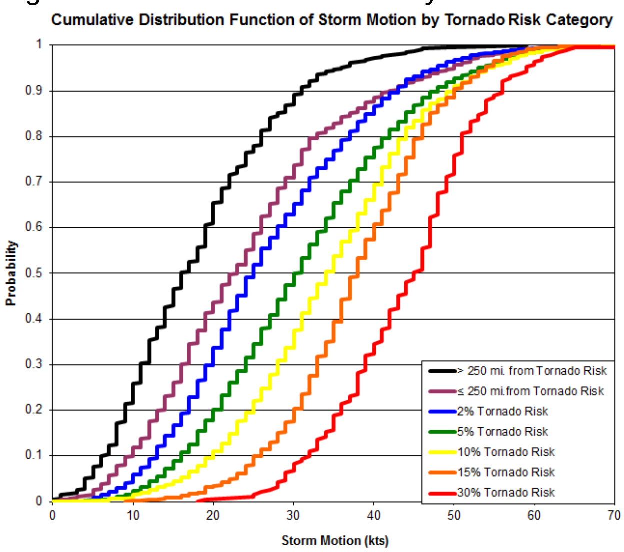

Figure 4: Cumulative Distribution Function plot of storm motion within each of the seven possible risk classification categories

Permutation testing was used to determine if differences in storm motions between groups were statistically significant. The null hypothesis (H_0) states that mean storm motions is the same between groups and the alternative hypothesis (H_1) states that mean storm motion is significantly different between categories. The sample means of storm motion were compared through 100,000 iterations for each set of 2 comparison groups (Table 5). All 21 possible permutations yielded a p-value less than 0.01, rejecting the null hypothesis at the 99% confidence interval.

With statistically significant increases in storm motion by tornado probability category, the data strongly suggests that storms in higher tornado probabilities have faster motions. Compared against the warning area information, the data infers that faster storm motions are a factor in the size (area) of warnings. Furthermore, storm motions and their interaction with the synoptic-scale environment can be a decision factor on the

Comparison Groups	Permutation p-value	Comparison Groups	Permutation p-value	Comparison Groups	Permutation p-value
> 250 mi. vs. ≤ 250 mi.	<0.001	≤ 250 mi. vs. 5%	<0.001	2% vs. 30%	<0.001
> 250 mi. vs. 2%	<0.001	≤ 250 mi. vs. 10%	<0.001	5% vs. 10%	<0.001
> 250 mi. vs. 5%	<0.001	≤ 250 mi. vs. 15%	<0.001	5% vs. 15%	<0.001
> 250 mi. vs. 10%	<0.001	≤ 250 mi. vs. 30%	<0.001	5% vs. 30%	<0.001
> 250 mi. vs. 15%	<0.001	2% vs. 5%	<0.001	10% vs. 15%	<0.001
> 250 mi. vs. 30%	<0.001	2% vs. 10%	<0.001	10% vs. 30%	<0.001
≤ 250 mi. vs. 2%	<0.001	2% vs. 15%	<0.001	15% vs. 30%	<0.001

Table 5: Summary table of permutation test p-values when comparing sample means of storm motion

placement of tornado probabilistic outlooks by the SPC.

3.3 NWS Performance Statistics by SPC Risk

Evaluating POD, FAR, and CSI for each warning classification category shows an overall improvement in skill scores as tornado probability increases (Figure 5).

Of the three performance metrics, the

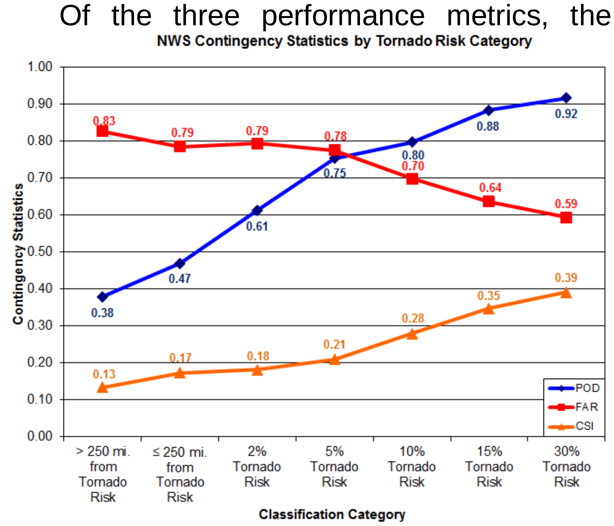

Figure 5: NWS performance metrics for warnings in each of the seven classification categories

POD score shows the greatest improvement as the tornado probability percentage increases. The POD increases from 0.378 for warnings greater than 250 mi. from any risk to 0.917 in the 30% tornado risk, an overall improvement of 0.539.

The improvement in FAR is most noticeable from 5% tornado risk onward. The FAR improves from 0.828 for warnings greater than 250 mi. from any risk to 0.594 inside the 30% tornado risk, an overall improvement of 0.234.

Similar to the trend in FAR, the CSI shows its greatest improvement from the 5% tornado

Classification Category	POD			FAR			CSI		
	2.5%	*50%*	*97.5%*	*2.5%*	*50%*	*97.5%*	*2.5%*	*50%*	*97.5%*
>250 mi. (N=418)	0.269	0.378	0.502	0.773	0.827	0.878	0.091	0.134	0.183
≤250 mi. (N=578)	0.379	0.470	0.563	0.740	0.786	0.826	0.136	0.173	0.216
2% (N=1705)	0.548	0.612	0.671	0.773	0.795	0.819	0.159	0.182	0.203
5% (N=3048)	0.704	0.753	0.799	0.758	0.776	0.793	0.191	0.209	0.227
10% (N=1242)	0.734	0.796	0.851	0.670	0.700	0.728	0.250	0.279	0.309
15% (N=740)	0.822	0.886	0.939	0.596	0.636	0.671	0.311	0.348	0.388
30% (N=192)	0.808	0.918	0.989	0.525	0.594	0.668	0.317	0.392	0.467

Table 6: Summary of tilted bootstrap at boundaries of 95% confidence internal and median for POD, FAR, and CSI scores

risk onward. The CSI improves from 0.134 for warnings greater than 250 mi. from any risk to 0.391 inside the 30% tornado risk, an overall improvement of 0.257.

While these results are positive for an improved detection of tornadoes in higher risk categories, can previously discovered factors account for uncertainty in the dataset? For example; we saw from Figure 3 that warning sizes were significantly larger in higher risk categories. Additionally, the size of the SPC forecast area decreases with each increasing risk category. Larger warnings in a smaller forecast area will ultimately yield a greater likelihood of warning verification. In order to get a range of uncertainty for these contingency statistics, a non-parametric tilted bootstrap analysis was performed on the dataset (Table 6).

When comparing the POD ranges against their neighboring classes, the data suggests that the POD at the 5% risk is statistically significantly higher than the 2% risk. Looking beyond neighbors, the 30% risk is only statistically significant when compared to all classes at the 5% risk and below.

Looking at neighboring FAR ranges shows statistical significance between the 5% and 10% groups. It should be noted that there is overlap of 0.001 between the 10% and 15% classes. This small difference could be removed if a higher number of bootstrap replicates are used.

Evaluating the CSI ranges along neighboring classes shows statistical significance between the 5%/10% and the 10%/15% class lines. This corresponds much more closely to the comparisons in the FAR scores than with the POD scores.

With a reduced frequency of issuance in addition to a smaller forecast area in these higher risk categories, there will ultimately be a reduced number of observed yes events (a+c). Ultimately, this leads to a higher amount of uncertainty and hinders statistically significant departures between neighboring risk classes in most instances.

4. Summary, Conclusions, and Implications

4.1 Summary & Conclusions

This study examined spatial characteristics of tornado warnings inside and within proximity of various tornado probabilities issued through the SPC Day 1 Convective Outlook. Two years of tornado warning products were plotted alongside tornado events and classified within seven categories. Two of these categories were proximity driven: (1) greater than 250 mi. from any risk and (2) within 250 mi. from any risk. The other five were within SPC-defined tornado probability groups: 2%, 5%, 10%, 15%, and 30%.

The data analysis suggests that warning sizes were larger in higher tornado probabilities. The mean warning size for warnings greater than 250 mi. from any risk was 279.62 sq. mi. compared to 530.80 sq. mi. for warnings inside the 30% tornado risk. One of the possible reasons for these size differences is due to storm motion. Faster storm motions would result in larger warning polygons.

By classifying storm motions based on tornado probability category, the dataset showed a positive correlation between speed and tornado probability. The differences in storm motions by category were significantly different at the 99% confidence interval through all category comparisons. Mean storm motion was 17.85 kts for warnings greater than 250 mi. from any risk and 44.18 kts for warnings inside the 30% tornado risk.

Evaluating NWS performance metrics for warnings and events in each of the seven categories showed a positive correlation in skill scores as tornado probability percentage increases. Across the seven categories there was a POD improvement of 0.539, a FAR improvement of 0.234, and a CSI improvement of 0.257. Warning size and smaller risk forecast areas play a role in the overall performance improvement, but sample deficiencies yield fewer statistically significant differences between neighboring risk categories.

Under the constraints of this study, the dataset analysis suggests that tornado warnings inside higher SPC tornado probabilities are more likely to be larger in size and have a greater chance of verifying a tornado event. In addition, faster storm motions were observed inside these higher risk categories and this could have played a role in the greater warning size.

Utilizing GIS technologies allows for complex spatial datasets to be interrogated and quantified. While SPC forecasts are directed at the probability of a specific hazard, warning characteristics (e.g. size, frequency) and their trends from these SPC forecasts could be useful for NWS partners (e.g. emergency managers, broadcasters) as a decision assistance tool. With a larger dataset (5-10 years), outputs from this study could yield meaningful trends in the form of warning analogs; providing a first guess on the frequency and coverage of warnings hours before storm development. Future work hopes to further define and evaluate the merits of this system.

4.2 Implications

Information from *Storm Data* provided the inputs for plotting the tornado events. While tornadoes are more high-profile events than other severe reports (wind, hail, etc.), there is a possibility of inaccuracy in event time and location (Witt et al., 1998; Trapp et al., 2006). The development of a "Next Generation Storm Data" program will incorporate better data assimilation techniques to increase accuracy regarding event location (MacAloney II, 2009). Benefits of this new system were seen in half of the study (post-2009). For the constraints of this study, there is a chance some of the tornado locations and times on record may be inaccurate.

5. Acknowledgements

This work was supported under the Cooperative Agreement NA17RJ1227 of the NOAA-University of Oklahoma/Cooperative Institute for Mesoscale Meteorological Studies, the NOAA/NWS Warning Decision Training Branch, and the United States Geospatial Intelligence Foundation (USGIF).

Special thanks go out to Kevin Goebbert and Chad Shafer for their comments and guidance on the performance metrics.

The views expressed in this paper are those of the authors and do no necessarily represent those of the NWS, NOAA, or CIMMS.

6. References

Brooks H. E., C. A. Doswell, and M. P. Kay, 2003: Climatological Estimates of Local Daily Tornado Probability for the United States. *Wea. Forecasting*, **18**, 626-640.

Doswell C. A., R. Davies-Jones, and D. Keller, 1990: On Summary Measures of Skill in Rare Event Forecasting Based on Contingency Tables. *Wea. Forecasting*, **5**, 576-585.

Kain J. S., M. E. Baldwin, P. R. Janish, S. J. Weiss, M. P Kay, and G. W. Carbin, 2003: Subjective Verification of Numerical Models as a Component of a Broader Interaction between Research and Operations. *Wea. Forecasting*, **18**, 847-860.

MacAloney II, B. W., 2009: The Next Generation of Storm Data Collection. *34th Annual Meeting of the National Weather Association*, Norfolk, VA.

National Weather Service, 2005: WFO Severe Weather Products Specification, *NWS Instruction 10-511*, NOAA.

National Weather Service, 2010: National Severe Weather Products Verification, *NWS Instruction 10-512*, NOAA.

Polger P. D., B. S. Goldsmith, R. C. Przywarty, and J. R. Bocchieri, 1994: National Weather Service Warning Performance Based on the WSR-88D. *Bull. Amer. Meteor. Soc.*, **75**, 203-214.

Schaefer, J. T., 1990: The Critical Success Index as an Indicator of Warning Skill. *Wea. Forecasting*, **5**, 570-575.

Trapp, R. J., D. M. Wheatley, N. T. Atkins, R. W. Przybylinksi, and R. Wolf, 2006: Buyer Beware: Some Words of Caution on the Use of Severe Wind Reports in Postevent Assessment and Research. *Wea. Forecasting*, **21**, 408-415.

U.S. Department of Commerce, Office of the Inspector General, 1998: NWS's Verification System for Severe and Hazardous Weather Forecasting Needs Modernization, *Inspection Report No. IPE-9255*.

Witt, A., M. D. Eilts, G. J. Stumpf, E. D. Mitchell, J. T. Johnson, and K. W. Thomas, 1998: Evaluating the Performance of the WSR-88D Severe Storm Detection Algorithms. *Wea. Forecasting,* **13**, 513-518.