Evaluation of NWS Watch and Warning Performance Related to Tornadic Events

Kelly Keene
National Weather Center Research Experiences for Undergraduates, and The University of Georgia, Athens, Georgia
Paul Schlatter
NOAA/NWS Warning Decision Training Branch, Norman, Oklahoma
Jack Hales
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Harold Brooks
NOAA National Severe Storms Laboratory, Norman, Oklahoma

ABSTRACT

Two organizations within the National Oceanic and Atmospheric Administration's (NOAA) National Weather Service (NWS) are responsible for disseminating critical information to their customers prior to and during severe weather. The Storm Prediction Center (SPC) focuses on conducting an overall assessment of the environmental conditions on the national scale, and issues the appropriate severe weather watches. At the local level, Weather Forecast Offices (WFO) are responsible for issuing warnings to cover imminent severe weather threats. The focus of this research is primarily on the evaluation of the NWS watch and warning performance in relation to tornadic events from 1997 to 2007. Data are obtained from the NOAA Performance Management Website and the SPC watch database. The watch and warning records are matched to each tornadic event, allowing an evaluation of tornado warning performance in relation to watch type and lead time. Statistical analysis of the data reveal interesting results on the nature of the relationship between tornado warnings and whether a watch was in effect, and if so, what type. Tornado warning performance and lead time increases with increasing F-scale intensity and also when there is a tornado watch in effect. The lowest tornado warning statistics occur when no watch is in place. Having a tornado watch in effect greatly increases the Probability of Detection (POD), while slightly decreasing the False Alarm Rate (FAR). For all tornado warnings from 1997-2007, there is an increasing of POD, while maintaining a steady FAR. Finally, the entire tornado warning data set is visualized using a new technique (Paul Roebber, 2008), and also added to a previous study (Brooks 2004).

1. Introduction

One of the primary missions of the National Weather Service (NWS) is to issue watches and warnings to ensure public safety. The Storm Prediction Center (SPC), located in Norman, Oklahoma, is responsible for issuing tornado and severe thunderstorm watches when "the risk of a hazardous weather event has increased significantly, but its occurrence, location, and/or timing is still uncertain. It is intended to provide sufficient lead time so that those who need to set their plans in motion can do so" (DOC/NOAA/NWS, 2007). An SPC issued watch typically covers an area of about 25,000 square miles (about 65,000 square kilometers) and is valid for about six hours (Corfidi, 1999). Tornado and severe thunderstorm warnings are issued by one of the 124 Weather Forecast Offices (WFOs) distributed across the country. Severe weather warnings are issued for condition(s) in which "a hazardous event is occurring, is imminent, or has a very high probability of occurring. A warning is used for conditions posing a threat to life and property" (DOC/NOAA/NWS, 2007). Typical warnings from a WFO cover an area of about 600 square miles, and are valid for 30-60 minutes. For the purpose of this study, severe thunderstorm warnings will not be evaluated.

Tornado warning performance has been the focus of numerous studies. Hales (1989) evaluated tornado warning performance for tornadoes classified as significant (an F-Scale intensity of F3, F4, or F5). His tornado warning and event data spanned the time period from 1982-1988 (before the deployment of the WSR-88D radar) and was related to the issuance of watches. He found that prior to the Weather Service Radar-88 Doppler (WSR-88D), Probability of Detection (POD) of a tornado warning inside a weather watch was two times higher than if no watch was in effect (.40 vs. .20). One of the conclusions of the study was that a tornado watch is critical "in setting the stage for timely warnings by creating a proper mind set in the field forecaster by stressing the meteorological parameters driving the tornado threat" (Hales, 1989).

A comprehensive study by Brooks (2004) analyzed all tornado warnings and events from 1986-2002. He computed POD, False Alarm Ratio (FAR), and Probability of False Detection (POFD) for all tornado events and warnings, concluding that the quality of the tornado warning system does not allow for a significant reduction in FAR without an accompanying significant reduction in POD. The study found that POD has increased significantly for tornado warnings from around 0.30 in 1986 to about 0.75 in 2002, and FAR has not significantly changed over the same time period.

Ram et al. (2007) related tornado and severe thunderstorm warning performance to watch type for a dataset from January 1 – April 19, 2006. Though working from data limited to part of one year, they found that warning performance improves with tornado watches in effect relative to no watch. They found that 75% of all tornado events occurred in either a tornado watch or a Particularly Dangerous Situation (PDS) tornado watch, and tornado warning performance is maximized for PDS tornado watches. Ram et al. (2007) attempted to determine if warning performance improved because of the watches, by surveying 54 NWS forecasters. Survey results were inconclusive, and the present study likewise does not attempt to address or quantify all potential factors that may have led to an improvement in tornado warning performance.

The focus of this project is to evaluate all tornado warnings and events from January 1, 1997 through September 30, 2007, in relation to watches issued by SPC. An assessment was performed to measure different aspects of how warning performance might be affected by the issuance of watches, such as the type of watch (tornado, severe thunderstorm, or no watch). We also evaluated how tornado warning lead times are affected by the three different watch types. In addition, we broke the event data into three different Fujita Scale intensity levels.

It is important to evaluate all potential hypotheses for this study, whether the results show a positive impact, negative impact, or no impact on tornado warning performance by the issuance of watches prior to warnings. Thus, the primary goal of this study is to reveal how WFO tornado warning performance behaves in relation to SPC watches.

Section 2 will provide information regarding data collection and the methodology used to carry out this research evaluation. Section 3 provides the detailed results, and section 4 will give a brief overview of the conclusions attained from the results.

2. Data and Methodology

Data from two sources are used for this project. Information regarding tornadic events and warning verification comes from the NOAA Performance Management Website (https://verification.nws.noaa.gov), and watch information is collected from the SPC watch database. The NOAA Performance Management Website contains tornado warning verification statistics from January 1986 through September 2007 (October 1, 2007 the NWS implemented storm-based warnings, which are verified differently than county-based warnings). This study examines tornado events and warnings during the county-based era after all WSR-88D's were deployed, a dataset from January 1, 1997 through September 30, 2007. We gathered data from every WFO, every county, and every F-Scale Intensity, giving us a collection of 15,393 tornadic events. Each tornado event is county-based, meaning it is counted as a separate event for each county through which a tornado travels. For each event information was given regarding the date and time of the event, and whether or not a tornado warning was in effect prior to the first official report of the tornado. The database also provides information regarding each tornado warning issued in which no tornado was reported, or false alarms.

Once the data are placed in a spreadsheet, the data collected from the SPC watch database are matched to each event and warning. There are three possible watch categories that are matched to each tornado warning and event: tornado, severe thunderstorm, or no watch.

Observed

Forecast		Yes	No
	Yes	A	B
	No	C	D

Figure 1: 2 x 2 Contingency Table, (Doswell, et. al., 1990)

The primary tool used for data analysis is the 2 x 2 contingency table, which has been well established in previous work (i.e Donaldson et al., 1975; Doswell, 1990; Schaefer, 1990). The 2 x 2 table compares tornado events and forecasts, as yes or no events. An example of the table is given by figure 1. A hit (A) in this table is defined as a yes/yes event, meaning a tornado was reported, and a tornado warning was in effect. A miss (C) is defined as a yes/no event, meaning a tornado was reported, but there was no tornado warning prior to the report time. A no/yes event is known as a false alarm (B), indicating that there was no reported tornado, but a tornado warning was in effect. The information for no/no events, known as a correct null (D), is not collected. The 2 x 2 table provides the information required to compute warning performance statistics POD, FAR, critical success ratio (CSI), success ratio (SR), and Bias. POD, FAR, CSI, SR, and Bias are defined as:

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

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

$$CSI = \frac{1}{\frac{1}{(1-FAR)} + \frac{1}{POD} - 1} \qquad (3)$$

$$SR = 1 - FAR \qquad (4)$$

$$Bias = \frac{A+B}{A+C} \qquad (5)$$

The resulting statistics are plotted into different line graphs for analysis.

An area that was not evaluated in this study was that of the psychological impact the issuance of a watch prior to a tornado warning may have on WFO forecasters. This concern has been previously examined, (Ram, et. al., 2007), and provided inconclusive results. Ram et al., 2007 distributed numerous surveys to NWS WFO forecasters across the country. Analysis of those surveys suggested that the affect of watches in place prior to a warning cannot be determined as having a positive or negative impact, due to various reasons, such as difference in location, confidence levels, and even personalities of the forecasters surveyed. Due to the inconclusive results and time restraints, this area was not taken into consideration for this project.

A potential problem with the data gathered for this study is the method in which the NWS uses to validate warnings. Validation of an event is obtained from contact of a witness to the event, whether the NWS contacts the witness, or the witness calls to report the event. This would mean that there could potentially be error in the number of events, causing it to be lower than the actual number, due to lack of reports from sparsely populated areas (Doswell et al., 2005).

Another area of concern is that the NOAA Performance Management Website gives double credit if more than one tornado is reported in the same county, while under the same warning. This can slightly inflate the POD. It should also be considered that other basic errors can occur in the NWS official tornado events dataset. For example, there can be inaccuracy in "the reporting or recording of time and location information, spatial and temporal variability in the efforts to collect severe-weather reports for warning verification programs, [and] changes in the nature of detailed damage surveys" (Brooks, et. al., 2003).

Another type of forecast system performance plot, the Roebber Method, (Paul Roebber, 2008) was also used to visualize forecast quality. The visualization technique mathematically and geometrically relates POD, FAR, CSI, SR, and Bias. Figure 2 gives an example of the graph in which multiple forecasts from his study were measured.

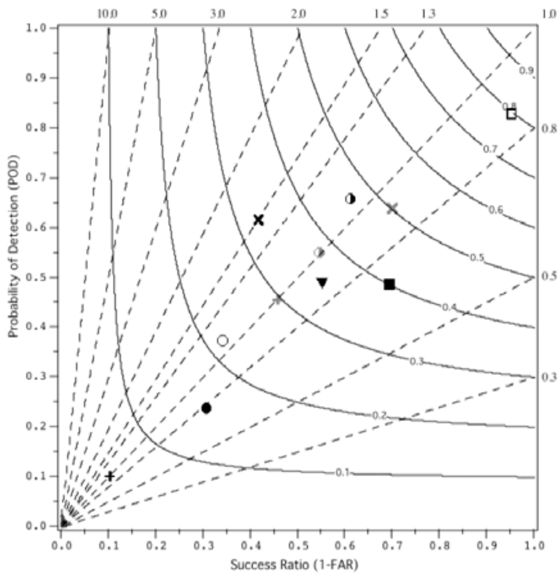

Figure 2: Roebber method to visualize forecast quality, (Roebber, 2008). Along the X-axis is the SR. The Y-axis gives POD. The solid contours represent CSI, while the dashed lines give Bias results, with the values indicated on the outward extension of the lines. In this diagram moving to the right is a reduction in FAR or moving up, which is an increase in POD. A perfect forecast would then lie in the upper right corner. This graph shows that there are typically trade-offs to improving certain measures of forecast quality. For example, the solid triangle and X in the graph both have comparable values of CSI. The X, however, has fewer misses, but more false alarms than the triangle. The open is closest to the top right corner, making it the highest quality forecast in this particular diagram.

3. Results

a. Type of Watch

Figure 3 shows the POD of tornado warnings in relation to watches over the 11 year time period. Based on figure 3, it is clear that when a tornado watch is in effect, the POD of tornado warnings is substantially higher than when a severe thunderstorm watch is in effect and even more so when no watch is in effect. The difference in the average POD of tornado warnings while under a tornado watch and POD while under a severe thunderstorm watch is .132, and there is a .327 difference when no watch is in effect. This suggests that having any watch in effect, and especially a tornado watch, prior to a tornado event, will likely lead to significant improved chances of detection by the WFO. Also indicated in figure 3 is a slight increase in POD from 1997 to 2007.

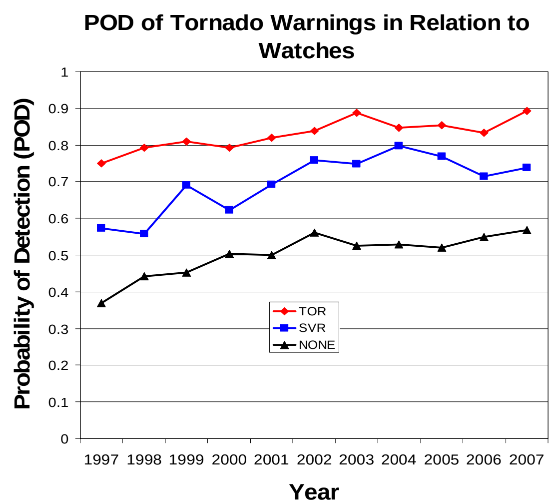

Figure 3: POD of tornado warnings in relation to watches. Note that tornado watches are indicated by lines with diamonds as points to indicate years. Severe thunderstorm watches are indicated by squares, while triangles symbolize when no watch was in effect.

Figure 4 shows the FAR of tornado warnings in relation to watches. While a higher POD is desired, the goal for FAR is a lower number. This figure shows that, with the exception of the years 2004 and 2006, when a tornado watch is in effect, the FAR slightly decreases. With the exception of the years 1997, 1998, and 2006, the FAR of tornado warnings was lower if the warning was issued during a severe thunderstorm watch. The overall difference in FAR between tornado watches and no watches is .081. The improvement in FAR is not as impressive as is the increase in POD with tornado watches. Nevertheless, it can be inferred that having any type of watch in place prior to a tornado warning will likely lead to a decrease in overall

FAR. Unlike POD, there is no visible indication of an increasing or decreasing temporal trend in overall FAR over the 11 year period of record.

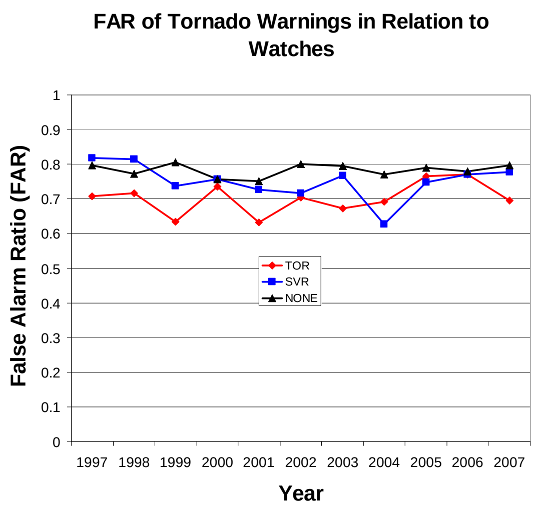

Figure 4: FAR of tornado warnings in relation to watches.

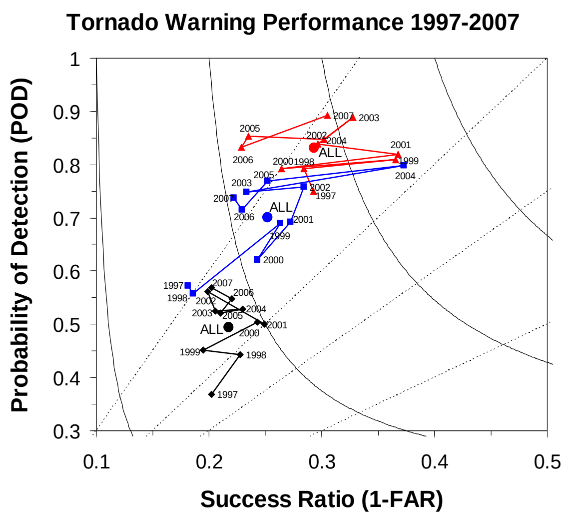

Figure 5: Roebber technique of plotting tornado warning performance from 1997-2007. Each area is shown as clusters of the 11 years, linked together respectively, and indicated by triangles (tornado watch), squares (severe thunderstorm watch), and diamonds (no watch). There is a larger point within each cluster, which is symbolic of the average performance from 1997 through 2007 for each watch type.

Figure 5 portrays the Roebber visualization technique with our dataset. The diagram has been magnified to emphasize areas of high POD and high FAR, characteristic of NWS tornado warnings. The three areas of interest include warning performance within a tornado watch, a severe thunderstorm watch, and no watch. Following the lines between the years, it is noted that over time POD tends to have an upward trend for all watch categories, with some variability from year to year, regardless of watch type. There is a substantial increase in tornado warning quality between a severe thunderstorm watch and having no watch, mainly because of an increased POD. There is more significant separation between tornado warning performance under a tornado watch compared to no watch. The primary outcome of this plot indicates that with a tornado watch in effect, the overall performance of warnings are moving toward the desired upper right-hand corner, closer toward perfection.

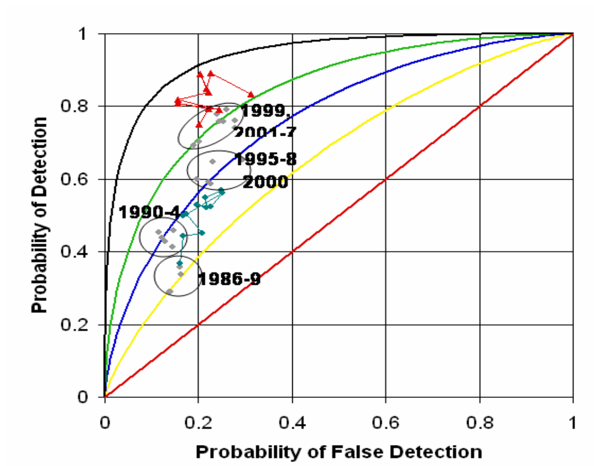

Figure 6: Adapted from Brooks (2004). The free-standing plotted points (diamond-shaped) indicate all years of tornadic event data, with or without any type of watch in place. Representative of tornadic event data within a tornado watch are the years, marked by triangles, with lines drawn between the years to indicate change from year to year. The years are given by squares, with connecting lines between for tornadic event data while no watch is in effect. The curved contour lines give a smoothed pattern of different time spans.

Figure 6 is adapted from Brooks (2004). It merges our 1997-2007 data with his, dating back to 1986. In figure 6, a perfect score would lie in the top left corner, giving a probability of false detection (POFD) of zero and a POD of one. Following Brooks (2004), POFD is computed as 0.10 of all events. The contour farthest to the left and top of the graph is a proposed goal for warning performance, to be reached by the year 2025. The next contour (to the right) signifies where the NWS is now in performance, which is essentially a 20 year difference in time. This is roughly the same as the distance between events when a tornado watch is in place and when no watch is in place. This implies that the difference between having a tornado watch in effect and having no watch in effect equals 20 years of improvement in overall tornado warning performance. The implications of this will be discussed in section 4.

b. F-Scale Intensity

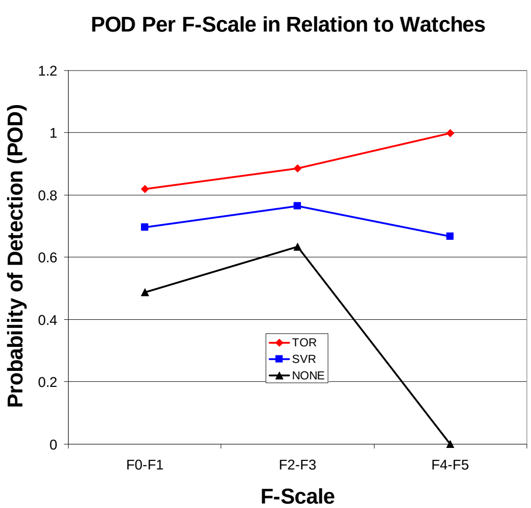

Figure 7: POD per F-Scale in relation to watches

Another goal of this study is to examine whether POD changes in relation to tornado intensity based on the type of watch in effect. Figure 7 shows an evaluation of the POD at three different F-Scale intensity levels (weak, strong, and violent), in relation to the three watch categories. Following Atlas (1976), this study defines weak tornadoes as F0-F1, strong as F2-F3, and violent as F4-F5. Note the substantial drop in POD with increasing tornado intensity when no watch is in place. The zero POD for violent tornadoes not in a watch is the result of a single event in 1998, where there was no watch and no warning prior to an F4 tornado event. Consistent with figure 3 there is an upward trend in the POD when a severe thunderstorm watch is in place, and more so when a tornado watch is in place. It is also notable that the POD of warnings actually increases for more intense tornadoes while under a tornado watch. Figure 7 also shows that the issuance of a tornado or severe thunderstorm watch prior to a warning leads to an increase of the POD, regardless of the intensity level of the tornado.

Figure 8 shows the POD of tornado warnings per F-Scale intensity by year, while in a tornado watch. Based on this figure, there is a general increasing temporal trend of POD with increasing tornado intensity. There is an exception in 2001, in which the overall POD during weak tornadic events was higher than that of strong tornadic events. It is unknown as to why this is the case. A significant result displayed in figure 8 is that for violent tornadoes under a tornado watch, there has never been a missed event from 1997-2007.

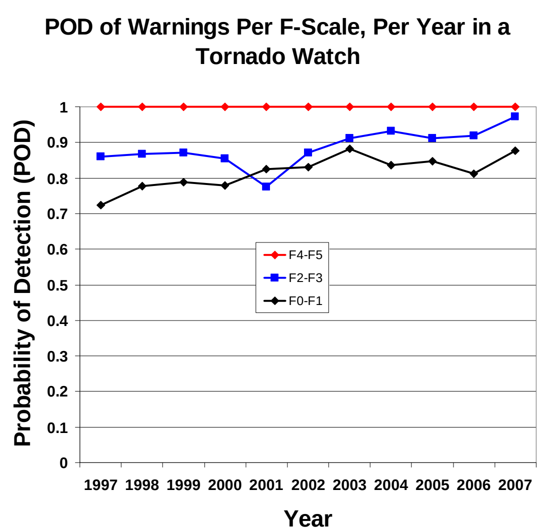

Figure 8: As in figure 3, except per F-Scale.

c. Lead Time

Another measure of forecast quality important to the general public is tornado warning lead time. This allows the public more time to take appropriate precautions once a tornado warning is issued. Figure 9 displays the average lead time of tornado warnings prior to a tornado event over the 11 year course, and broken down into the three watch-type categories. A lead time is defined as the amount of time in minutes between the issuance of a warning and the reported start time of a tornadic event. The NOAA Performance Management Website calculation of lead times includes missed events, which are given a lead time of zero.

While under no watch, there is an average lead time of 6.22 minutes for all tornado warnings from 1997-2007. The average tornado warning lead time in a severe thunderstorm watch is 10.61 minutes, while under a tornado watch it is 14.75 minutes. Thus, lead time is improved by more than eight minutes when a tornado watch is in effect compared to no watch.

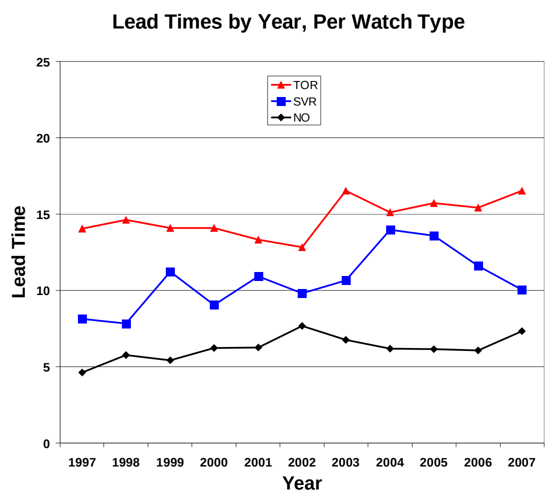

Figure 9: Lead times per year, per type of watch.

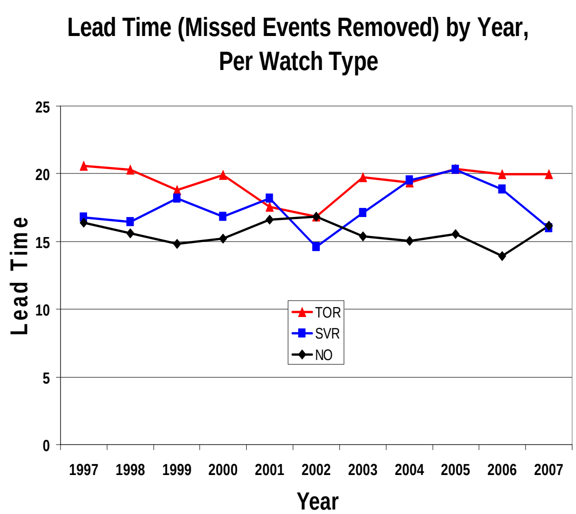

Figure 10: As in figure 9, except with missed events removed.

Figure10 is similar to figure 9, except that missed events are removed from the calculation. As mentioned above, collected data include missed tornado events which attain a zero lead time, significantly biasing the average lead times. Figure 9 is the official way NWS evaluates their tornado warnings, but it is also meaningful to examine lead times without the effects of POD. Removing that aspect in this graph now gives an average lead time of 17.80 minutes for tornado warnings while under a tornado watch, 15.17 minutes while under a severe thunderstorm watch, and 12.40 minutes when no watch had been issued. Having some sort of watch in place produces an average lead time of five to six minutes higher than with no watch. Figure 10 shows that there is no increase in average lead time from 1997-2007 when missed events are removed. Thus, comparing figure 9 to figure10, it is evident that an increase in POD from 1997-2007 is the reason the lead times increase over time in figure 10.

4. Conclusions

This demonstrated the following results:

- The average POD of tornado warnings increases by .327 when a tornado watch is in effect, as opposed to when no watch is in effect. From 1997-2007 there is also indication that the POD is steadily increasing for both watch types as compared to when no watch is in effect.

- When a tornado watch is in effect, there is a slight decrease (in the amount of .081) in FAR compared to when no watch is in effect. From 1997-2007 there is no evidence of an increase or decrease in FAR for any watch type, and for no watch.

- Lead time from tornado warning to occurrence is improved by an average of five to six minutes when a tornado watch is in effect, as opposed to no watch.

This study strongly suggests that the issuance of any SPC watch improves the overall tornado warning performance. In particular, tornado watches have been shown to improve warning performance by an amount equivalent to the overall improvements made in the last 20 years. In essence, not having a tornado watch in effect would be equivalent to removing all improvements made resulting from the science, training, and technology, such as the Doppler radar, the Warning Decision Training Branch, and the Verification of the Origin of Rotation in Tornadoes Experiment (VORTEX), for the past 20 years.

Both the improvement in warning performance and lead time prior to tornadoes (especially for stronger and violent tornadoes) during a tornado watch point to the vital role the SPC plays in the NWS during the tornado warning process.

Acknowledgements. This material is based upon work supported by the National Science Foundation under Grant No. ATM-0648566. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We greatly appreciate the help we have received from Andy Dean.

References

Atlas, D., 1976: Severe local storms. *Bull. Amer. Meteor. Soc.,* **57,** 398-435.

Brooks, H.E., Doswell, C.A. III, Kay, M.P., 2003: Climatological estimates of local daily tornado probability for the United States. *Wea. Forecasting,* **18**, 626-640.

Brooks, H.E., 2004: Tornado-warning performance in the past and future: a perspective from signal detection theory. *Bull. Amer. Meteor. Soc,* **85,** 837-843.

Corfidi, S.F., 1999. The birth and early years of the Storm Prediction Center, *Wea. Forecasting,* **14,** 507-525.

DOC/NOAA/NWS, 2007: National Severe Weather Products Specification. National Weather Service Instruction 10-512, May 2007, 65 pp.

DOC/NOAA/NWS, cited 2008: National Weather Service Verification/Storm Data. [Available online at https://verification.nws.noaa.gov/verification/severe/svrverif.html.]

DOC/NOAA/NWS, cited 2008: Storm prediction Center Product and Report Archives. [Available online at http://www.spc.noaa.gov/archive/.]

Donaldson, R.J., R.M. Dyer and M.J. Krauss, 1975: An objective evaluator of techniques for predicting severe weather events. Preprints: *9th Conf. Severe Local Storms,* Norman, Oklahoma. Amer. Meteor. Soc., 321-326.

Doswell, C.A. III, R.P. Davies-Jones, and D.L. Keller, 1990: On summary measures of skill in rare event forecasting based on contingency tables. *Wea. Forecasting,* **5,** 576-585.

Hales, J.E. Jr., 1989: The crucial role of tornado watches in the issuance of warnings for significant tornadoes. *Natl. Wea. Dig*., **15, 4**, 30-36.

Ram, J., Johns, R., Quoetone, E., Schlatter, P., 2007: National Weather Service Warning Performance Associated With Watches. [Available online at http://www.caps.ou.edu/reu/reu07/projects.html]

Roebber, P.J., 2008: Visualizing Multiple Measures of Forecast Quality. Submitted to *Wea. Forecasting.*

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