1.1 THE SKILL OF FORECASTING RELATIVELY ISOLATED SEVERE THUNDERSTORM EVENTS

Andrew R. Dean[1, 2] and Joseph T. Schaefer[2]
1 – Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, OK
2- NOAA/NWS Storm Prediction Center, Norman, OK

1. INTRODUCTION

The mission of the NOAA/Storm Prediction Center (SPC) prominently includes the forecasting of organized severe convection over the contiguous United States. "Severe" convection is officially associated with at least one of the following:

- Hail ¾-inch or greater in diameter,
- Convective wind gusts of 50 knots or greater, or wind damage commensurate with such wind speeds, or
- Tornadoes

When there appears to be an enhanced, imminent risk of severe convection, a "watch" is issued. According to National Weather Service (NWS) guidelines (National Weather Service 2005), a watch should be issued when severe convection is expected to occur over an area of at least 8,000 mi^2 and duration of at least 2 hours. The type of watch depends on the form of severe weather expected:

- a "Severe Thunderstorm Watch" (SEV) should be issued when six or more wind and hail reports are expected;
- a "Tornado Watch" (TOR) should be issued when two or more tornadoes, or at least one strong or violent (F2 or greater on the Fujita Scale) tornado, is expected.

These categories are not mutually exclusive. A TOR watch will frequently also meet the criteria of a SEV watch.

In order to evaluate forecast performance and to provide useful feedback to forecasters, watches are routinely verified after official storm reports, which are collected by NWS field offices, become available. Several important aspects of watch verification must be considered, including probability of detection (POD), false alarm rate (FAR), and the time between forecast and report (lead time). In this study, the focus will be mainly on trends in the POD.

Traditionally, watch POD is defined as "the fraction of reports that occurred in a watch." However, as the watch criteria above suggest, some reports are not expected to occur in a watch, i.e., they do not sufficiently cluster together with other reports in space and time. A relatively simple clustering technique will be used to determine which reports sufficiently cluster to meet the criteria of a watch and to compute PODs.

2. TRENDS IN SEVERE WEATHER REPORTS AND WATCHES

SPC maintains a database of severe weather reports, with tornado data dating back to 1950 and wind/hail data dating back to 1955. Figure 1 clearly shows a dramatic increase in the number of severe hail and wind reports since around 1980, with a smaller increase in the number of tornadoes over the same period. This does not necessarily imply that the actual frequency of such events has increased; advances in radar technology, more aggressive reporting policies, an increase in "storm chasing", the advent of cell phone communications, and numerous other factors have contributed to a significantly greater ability to detect severe weather when it occurs (Schaefer and Edwards 1999).

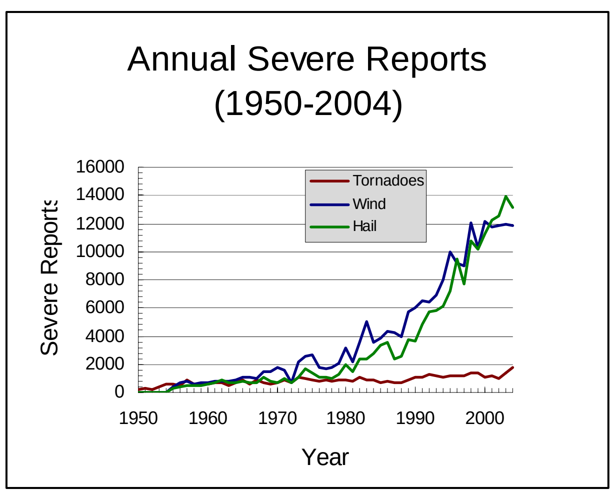

Figure 1: The number and type of severe thunderstorm reports by year (Note: 1972 wind and hail data are incomplete.)

* Corresponding author: Andrew R. Dean, CIMMS, University of Oklahoma, 1313 Halley Circle, Norman, OK 73069-8493; E-mail: andy.dean@noaa.gov

One side effect of the increased reporting efficiency of recent years is that many reports are nearly identical in location and time to other reports, the result of the same event being reported multiple times. Before 1996, wind and hail reports that occurred in the same county as another report of the same type and within 10 miles and 15 minutes of each other were treated as one event in the SPC database unless separate reports were of significant magnitude (2+ inch hail, 65+ knot wind gusts) or were associated with injuries or fatalities. This policy was changed in 1996 so that all reports collected by the NWS are now included in the database. Figure 2 shows that since 1996, there has been a drastic increase in the number of "redundant" reports, i.e., those that are within 10 miles and 15 minutes of other reports of the same type. For the remainder of this study, 10 mile/15 minute filtering is applied to all of the data so that no bias is introduced into the results from this change in policy.

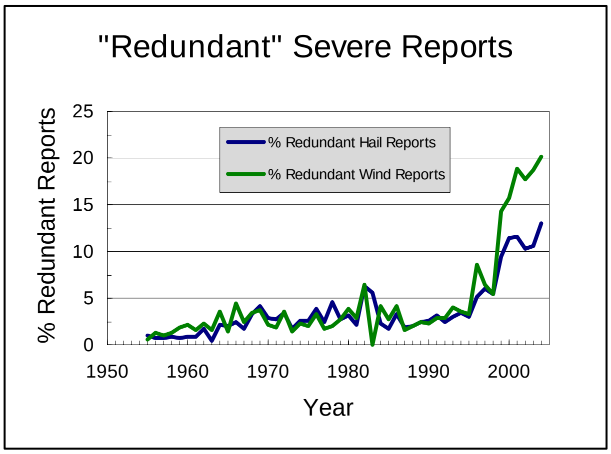

Figure 2: Percent of Severe Thunderstorms that are "redundant."

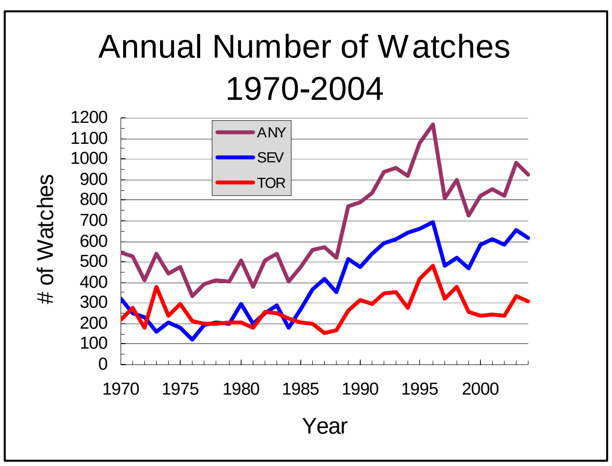

Figure 3: The annual number of watches by type from 1970–2004.

The annual total of watches issued (Fig. 3) shows a corresponding increase in the number of watches with time, though the response is slightly lagged; while the number of non-tornadic reports began to increase in the 1970s, the total number of watches begins to increase rapidly beginning in 1985. Beginning in 1985, severe thunderstorm watches become increasingly more common than tornado watches. A peak in the number of all types of watches in noted in the mid-1990s, though the total number of severe reports has continued to increase since then.

3. REPORT CLUSTERING

A simple grid-based method is used to determine if individual reports cluster together in space and time to meet the criteria of a watch. To approximate an area of 8,000 mi^2, a neighborhood of radius 50.5 miles (81.3 km) is defined around each grid point. Reports that fall within that neighborhood are sorted by their time of occurrence and those that group together in intervals of two hours or less are considered to be a cluster. This process is repeated for each grid

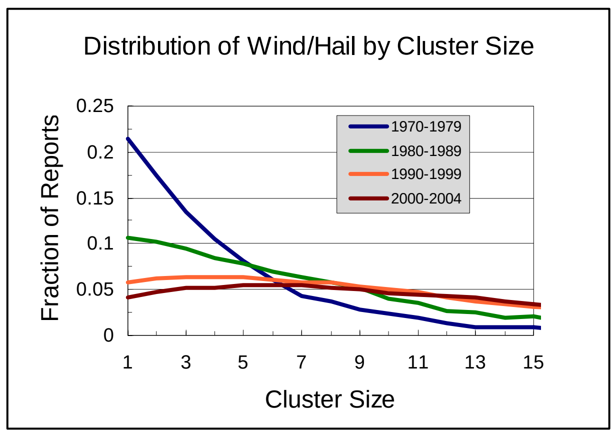

Figure 4: Distribution of non-tornadic severe weather reports according to their maximum cluster size. Clusters of combined and hail reports are used.

point. A report can fall into more than one cluster; the maximum cluster size that each report is associated with is noted and used in the results below. A 10-km grid is used in this case as a compromise between high resolution and computing efficiency.

Clustering results for each decade since 1970 are presented in Figure 4, which shows the fraction of non-tornadic (wind and hail) reports that are associated with a given maximum cluster size. A dramatic decrease with time in the fraction of isolated events (cluster size = 1) is evident since the 1970s, with 21% of non-tornadic reports being isolated during the period 1970–1979 compared with only 4% isolated reports in the 5-year period 2000–2004. A similar trend is seen in tornado reports (clustered with other tornado reports only). Figure 5 shows that 41% of tornado reports were isolated from 1970–1979, compared with 21% from 2000–2004.

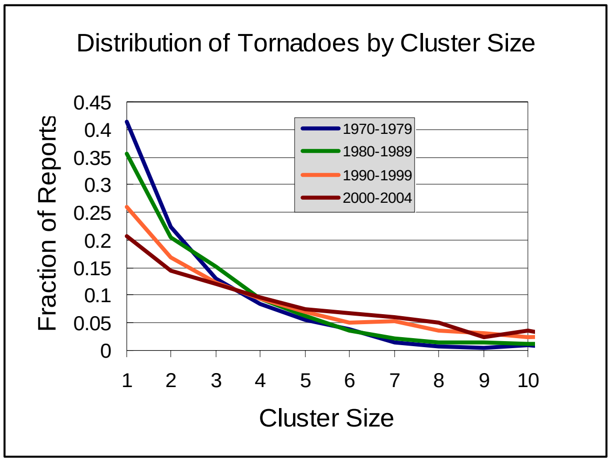

Figure 5: Distribution of tornado reports according to their maximum cluster size. Clusters of only tornado reports are used.

The cumulative distributions of reports by cluster size (Figures 6 and 7) are useful for determining the fraction of reports that cluster to meet the criteria of a watch. The percentage of non-tornadic reports that cluster together in a group of six or more (SEV criteria) has steadily increased with time from 29% over the period 1970–1979 to 75% from 2000–2004. While less marked than for wind and hail reports, there has also been a steady increase in the percentage of tornado reports that cluster together in a group of two or more (TOR criteria) from 50% from 1970–1979 to 79% from 2000–2004. Note that in the 1970s, the fraction of non-tornadic reports that clustered to meet watch criteria was much smaller than the fraction of tornado reports that clustered to meet watch criteria. However, in recent years, the fraction of non-tornadic reports that should be in a watch has become nearly equal to the fraction of tornadic reports that should be in a tornado watch.

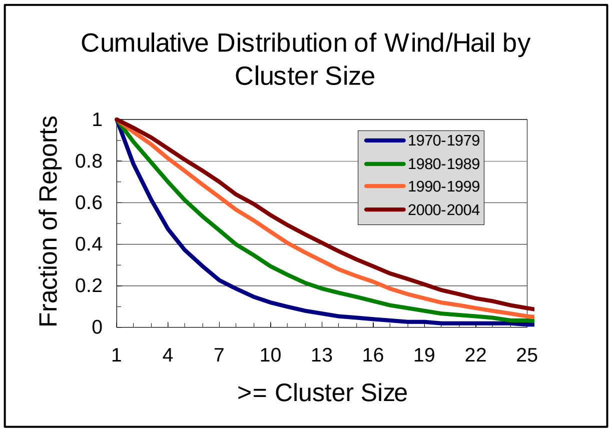

Figure 6: Cumulative distribution of non-tornadic severe reports by cluster size. The value on the y-axis corresponds to the fraction of reports that cluster at the size given on the x-axis or larger.

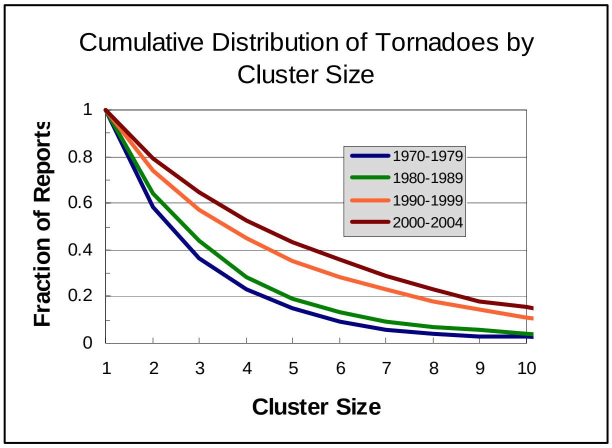

Figure 7: Cumulative distributions of tornado reports by cluster size. The value on the y-axis corresponds to the fraction of reports that cluster at the size given on the x-axis or larger.

4. WATCH VERIFICATION IN THE CONTEXT OF REPORT CLUSTERING

Annual values of the overall watch POD, the fraction of watches that verify according to watch criteria (WatchVerf), the fraction of reports that cluster to watch criteria (ClustFrac), and the POD for clustered reports (PODclust) from 1970–2004 for filtered, non-tornadic reports in any type of watch are shown in Figure 8. POD values generally increase in time until peaking in the mid-1990s, corresponding with the peak in total number of watches issued (Fig. 3). Since then, POD values have leveled off or slightly decreased. The fraction of watches that verify according to watch criteria has steadily increased throughout the period.

The same measures for tornadoes in tornado watches are shown in Figure 9. POD shows a general increase through the period, with some year-to-year variation, before leveling off since the mid-1990s. Again, there is also a general increase in the fraction of tornado watches that verify using the TOR criteria, though not as dramatic as when all watches are considered.

In some cases, local minima in the fraction of clustered tornado reports correspond with local minima in all of the verification measures, including those measures which only consider clustered reports. This is seen in both the “all severe” and the “tornado” measures, and is particularly evident in 1982, 1987, and 2000. This indicates that forecasting severe thunderstorms is more difficult in years when clustered activity is anomalously low.

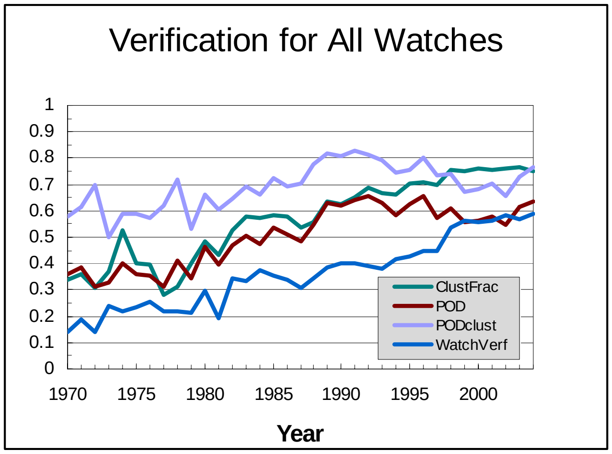

Figure 8: Verification measures for all reports in all watches. ClustFrac = fraction of reports that cluster to watch criteria, POD = overall POD, PODclust = POD of clustered reports only, WatchVerf = fraction of watches that contained a sufficient number of reports to meet watch criteria.

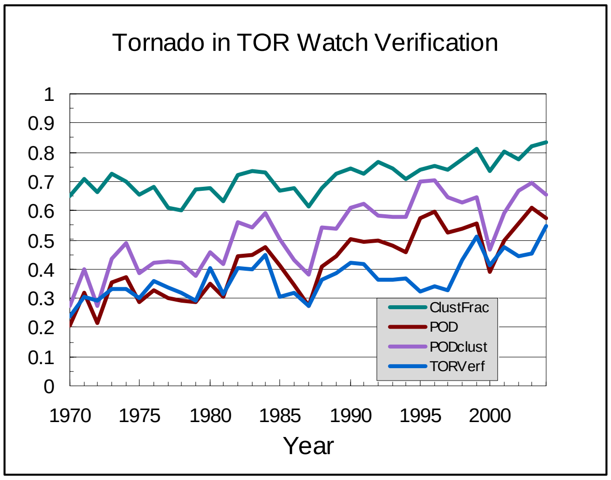

Figure 9: Verification measures for tornado reports in tornado watches. ClustFrac = fraction of tornado reports that cluster to meet TOR criteria, POD = POD of tornado reports in tornado watches, PODclust = POD of only clustered tornado reports in tornado watches, TORverf = fraction of tornado watches that met TOR criteria.

The effect of report clustering on POD can be further investigated by examining distributions of POD by cluster size, as shown in Figures 10 and 11 for non-tornadic reports and Figures 12 and 13 for tornadic reports. Results are grouped together by decade to show how the distributions have changed with time. In the case of non-tornadic reports, the POD for reports that occur in smaller cluster sizes is actually lower for the period 2000–2004 than in any of the other three decades considered (Fig. 10), even though the overall POD remains rather high for that period. This implies that relatively isolated non-tornadic events have become more difficult to forecast in recent years. This issue will be explored later.

In contrast to the non-tornadic statistics, the POD distribution for tornadoes by cluster size (Fig. 12) does not show a decline at smaller cluster sizes in recent years. The distribution for 2000–2004 is similar to the distribution for the periods 1980–1989 and 1990–1999, even though the anomalously low tornado verification statistics from 2000 are adversely affecting the 5-year average. There is no apparent increase in the difficulty of forecasting relatively isolated tornado events, in contrast to the results for non-tornadic reports, though the POD distribution suggests that weak (F0 and F1) tornado events remain difficult to forecast as a whole.

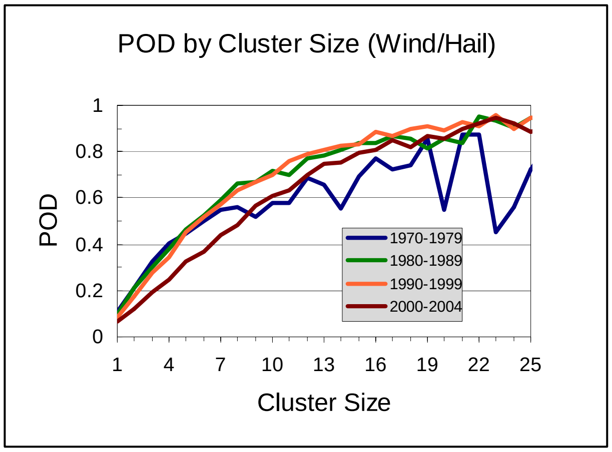

Figure 10: POD of non-tornadic reports by cluster size.

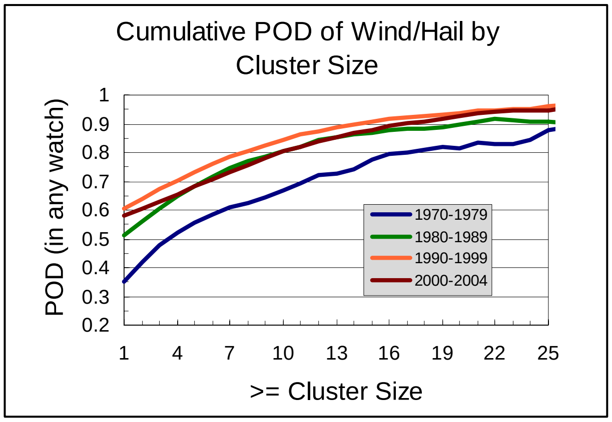

Figure 11: Cumulative POD of non-tornadic reports by cluster size.

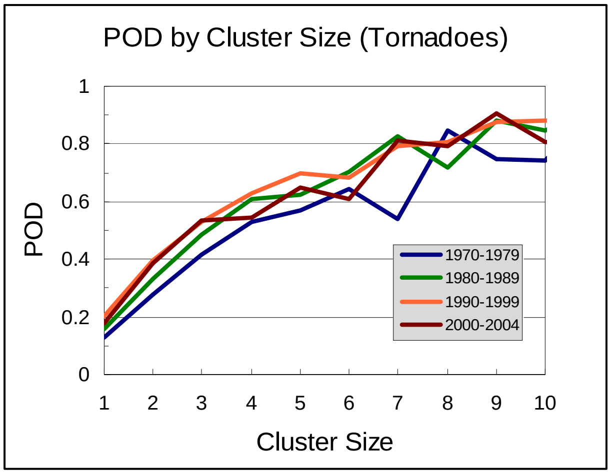

Figure 12: POD of tornadoes by cluster size, where POD is defined as the fraction of tornadoes in tornado watches.

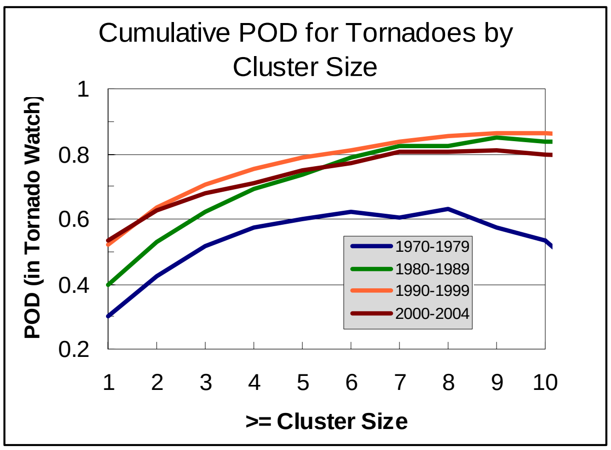

Figure 13: Cumulative POD of tornadoes by cluster size, where POD is defined as the fraction of tornadoes in tornadoes watches.

5. A FURTHER APPLICATION OF REPORT CLUSTERING FOR WATCH VERIFICATION

Considering the dramatic increase in the total number of reports and report clustering in recent years, are the original criteria for watches still appropriate? One potentially useful measure for investigating this issue can be defined as the fraction of reports that are correctly forecast to either be in a watch if they cluster sufficiently or not in a watch if they do not cluster sufficiently. In the context of a 2x2 contingency table, this measure is traditionally referred to as the *proportion correct* (Wilks 1995). By computing the proportion correct over a range of cluster size criteria, we can examine how well the watches are calibrated with the watch criteria.

The proportion correct for non-tornadic reports (Fig. 14) peaks at a cluster size of six for the three 10-year periods from 1970–1999, which corresponds with the SEV criteria. However, from 2000–2004, the proportion correct peaks at a cluster size of eight. This result is not surprising in light of the decrease in POD for non-tornadic reports in smaller clusters over the same period noted in Figure 10. Given recent trends in watch verification, raising the SEV criteria to eight reports would result in more non-tornadic reports being correctly classified as being in a watch or not, if there was no other change in the philosophy or practice of issuing watches or in the criteria for classifying a non-tornadic event as severe.

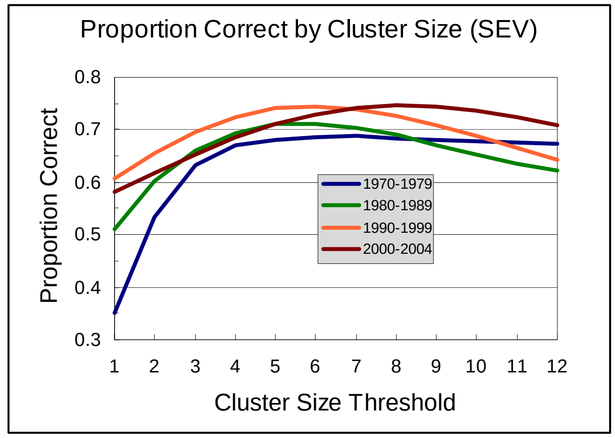

Figure 14: Proportion of correctly forecast non-tornadic reports at different cluster size thresholds.

The proportion correct for tornadoes (Fig. 15) peaks at a cluster size of four for 1970–1979 and 1980–1989, while peaking at a cluster size of three for 1990–1999 and 2000–2004. This suggests that the tornado watches have never quite reached the goal set by the TOR criteria. As above, the proportion correct could be maximized by changing the watch criteria. However, raising the criteria, particularly for tornado watches, may not be desirable, as discussed below.

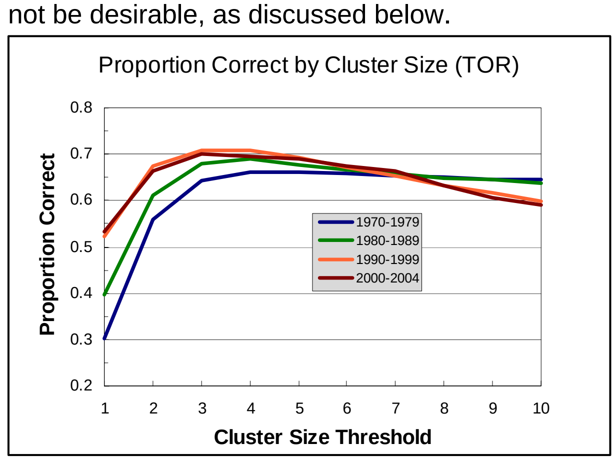

Figure 15: Proportion of correctly forecast tornado reports at different cluster size thresholds (tornado clusters only).

6. DISCUSSION

The report clustering results presented here suggest that, as a fraction of the total number of reports, relatively isolated reports are becoming increasingly rare. Around 70% of non-tornadic reports occurred in clusters of 10 or less from 1980–1989, but that number has dropped to less than 50% for the period 2000–2004. Similarly, around 70% of tornado reports occurred in tornado clusters of four or less from 1980–1989, but that number has dropped to less than 50% for 2000–2004.

While the POD for tornadoes at a given cluster size has been relatively consistent, POD for non-tornadic reports at smaller cluster sizes has decreased, suggesting that small clusters of non-tornadic reports have apparently become more difficult to forecast. A more rigorous study is necessary to fully investigate why this is the case. One possibility is that because of increased reporting, events that would not have produced enough reports to justify a watch in the 1970s and 1980s are producing more reports in recent years. As a result, less-predictable events, such as weakly-forced diurnal convection in the southeastern United States during the summer, are increasingly producing enough reports to meet watch criteria.

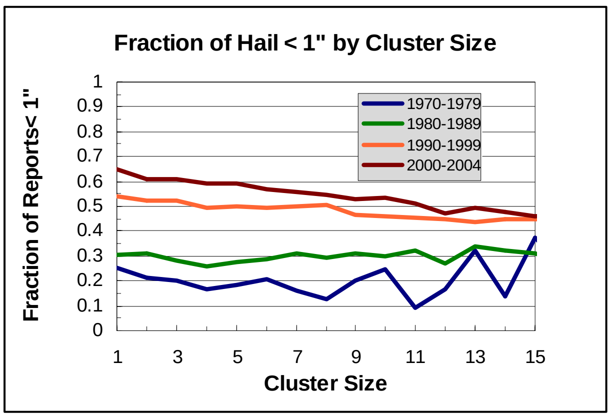

Figure 16: Fraction of hail reports less than 1-inch (2.54 cm) by cluster size.

Another notable trend is that the intensity of relatively isolated events has become weaker over the past decade and a half. Figure 16 (hail less than 1-inch in diameter) and Figure 18 (F0 and F1 tornado reports) show that hail and tornado reports of lesser magnitude comprise a greater fraction of the total number of reports both with time and with decreasing cluster size since 1990. Figure 17 shows that the fraction of wind reports of lesser magnitude (less than 55 knots) decreases with cluster size since 1990. However, a notable decrease in the fraction of wind reports less than 55 knots at all cluster sizes occurred during the period 2000–2004. This anomaly is at least partially due to a NWS policy change requiring that an estimated wind speed be given with all wind damage reports. Also notable is the decrease in recent years in the fraction of F2 or greater tornado reports at all cluster sizes. For the period 2000–2004, at least 90% of the tornado reports that cluster in groups of five or less have been F0 and F1 tornadoes

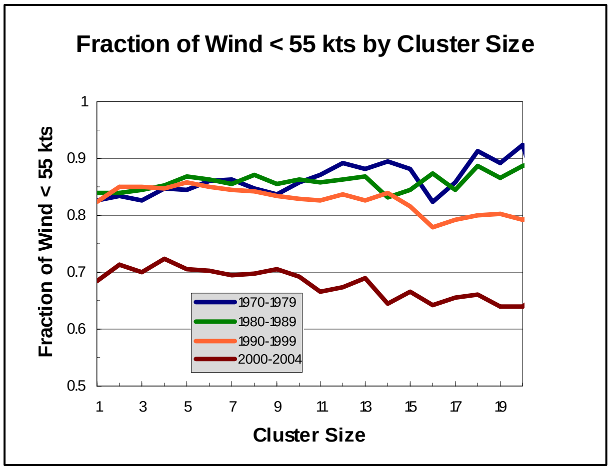

Figure 17: Fraction of severe wind reports with either no reported magnitude or a magnitude of less than 55 knots (28.3 m/s) by cluster size.

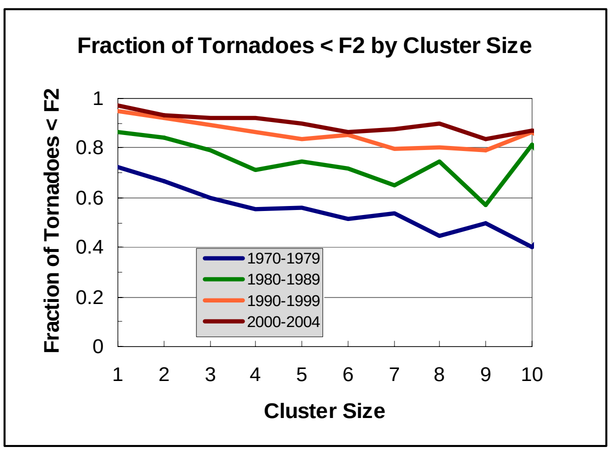

Figure 18: Fraction of tornado reports with a Fujita-scale rating of F0 or F1 by cluster size.

Given that (1) POD generally increased until the mid-1990s and has mostly held steady since then and (2) the fraction of watches that meet watch criteria has continued to increase (fewer watches have successfully covered the same number of reports), the overall skill of watches has continued to increase over time with respect to those measures. However, watch verification is a multi-faceted process and not all of the information necessary for an exhaustive analysis has been presented here. The main gap in forecast performance at present is a lack of watches to cover relatively isolated events which still cluster to meet watch criteria. It is possible to increase the POD by increasing the number of watches issued, but this would potentially cause a decrease in the fraction of watches verified. This is the forecast dilemma of trying to forecast as many events as possible without issuing too many "false alarms".

In light of the increase in report clustering with time and the results presented in section 5 above, altering the watch criteria to require a greater coverage of reports may be appropriate from a statistical perspective. Prior to any changes, further study is needed to determine whether a change in criteria is desirable from a societal perspective and whether a change in the frequency of watches will improve public response.

In the case of tornadoes, the public safety hazard combined with the mission of the Storm Prediction Center—to provide accurate forecasts of as many events as possible—make watch criteria revision inappropriate. Thus, the goal of capturing any group of two or more tornadoes in a tornado watch should not be changed, even though current science does not allow that goal to be met perfectly.

Acknowledgements: Numerous discussions with Russ Schneider and the SPC staff helped clarify the ideas presented. Dan McCarthy maintains the SPC database of severe reports that we used in this study. In addition, we would like to thank Linda Crank, an SAIC employee who works at the SPC, for her editing and formatting assistance.

Andrew R. Dean was funded under NOAA–University of Oklahoma Cooperative Agreement #NA17RJ1227, U.S. Department of Commerce. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of NOAA or the Department of Commerce.

7. REFERENCES

National Weather Service, 2005: National Weather Service Instruction 10–512. [Available online at http://www.nws.noaa.gov/directives/010-pd01005012d.pdf.]

Schaefer, J.T., and R. Edwards, 1999: The SPC tornado/severe thunderstorm database. Preprints, *11th Conf. on Applied Climatology,* Dallas, TX, Amer. Meteor. Soc., 215–220.

Wilks, D.S., 1995: Statistical Methods in the Atmospheric Sciences: An Introduction, Academic Press, 467 pp.