8A.6A ON THE PRECISION OF THREATS IN NATIONAL WEATHER SERVICE SEVERE WEATHER WARNINGS

Ray Wolf*
NOAA National Weather Service
Davenport, Iowa

Sara Christensen
Valparaiso University
Valparaiso, Indiana

1.0 Introduction

The National Weather Service (NWS) implemented storm-based warnings in the fall of 2007 to provide a greater level of precision in depicting the actual areal threat of severe weather in contrast to the former county-based method. The national verification program provides statistical measures for these "polygon" warnings such as probability of detection (POD), false alarm rate (FAR), and lead time. This form of assessment is referred to as administrative verification, since the results are used by NWS leadership to track agency performance.

Two other types of verification exist, social and scientific. Social verification assesses performance using measures based on the end users viewpoint. These measures would take into account different values and perspectives that users have in comparison to the NWS. For example, while the occurrence of only a funnel cloud during a tornado warning warrants a missed event under the NWS's administrative verification, it may yield a hit from a user's perspective, such as an emergency manager, because the *risk* of a tornado was high enough to justify taking protective measures even though a tornado did not ultimately occur.

Most NWS severe thunderstorm warnings also indicate the type and/or magnitude of the expected threat by specifying a hail size and/or wind speed. Indeed that has been the case at the NWS office in Davenport, Iowa (DVN) since at least 2005. Of course, the tornado threat is implicit via issuance of a tornado warning. However, there is no formal effort to verify the accuracy of these warning parameters in severe thunderstorms. An assessment to measure the forecast skill of these parameters would constitute scientific verification, since the ultimate goal would be to improve the forecasting of tornadoes, hail and straightline winds. Thus, the goal of this study is to quantify the current level of skill in threat forecasting to develop a baseline from which future training and research efforts can build and improvements be measured.

----

*Corresponding author address: Ray Wolf, NOAA/NWS; e-mail: ray.wolf@noaa.gov

2.0 Methods

Severe thunderstorm warnings (SVR) and tornado warnings (TOR) in the DVN county warning area (CWA) from 2005-2008 were paired with severe weather reports occurring during the valid time of the warning. Thus missed events and warnings followed by no severe weather reports of any type are not included in this study (they are already tracked by NWS administrative verification). The maximum hail size and/or maximum measured wind gust, or estimated wind gust (based on the type and degree of damage reported), occurrence of a tornado or funnel cloud, were compared to the parameter and value (SVR wind speed or hail size) stated in the initial warning issuance. Tornado is the implicit parameter when a tornado warning is issued. Verification statistics were generated from those comparisons.

Some challenging data quality issues are faced when developing a data set of severe weather reports, and these must be considered when evaluating the results. For example with tornadoes, some tornadoes may occur and not be observed, especially at night. Straightline wind damage could be misdiagnosed as tornado damage (and vice versa) if no one observed the event. And there is potential error in discriminating between the occurrence of a tornado or funnel cloud, particularly when no damage ensues and viewing may occur from a distance or be obscured by low visibility, trees, etc.

With severe thunderstorms, the issue of sampling is potentially significant, i.e., did we get the report of the maximum hail size or wind gust the storm actually produced. Did we have spotters in the correct locations and did we make an effort to get the maximum report after verifying the warning? Did the spotter correctly estimate the size of hail or speed of the wind? How accurate is the wind speed estimate based from damage? This latter issue can be mitigated somewhat by application of the EF scale, though these are still estimates.

Most of these error sources appear to have no bias, except that of maximum hail size or wind gust. In these cases, an underestimate of reality would be expected rather than overestimate.

3.0 Results

3.1 *Tornado Warnings*

Outcomes from TORs were sorted in the following order as depicted in Figure 1: no severe weather reports = true false alarm, tornado occurred, funnel cloud occurred, wind and/or hail occurred. The True False Alarm is indeed that, at least based on the lack of reports, i.e., no severe weather or funnel clouds occurred. The presumption is made that since a warning was issued, an effort to verify it was made and the lack of reports likely represents reality. Warnings falling into this category indicate a failure to recognize a *warning threshold*, and thus imply a certain approach to training. This is the same training issue faced when addressing a missed event.

In comparison, tornado warnings followed only by severe criterion winds and/or hail indicate a failure to recognize the *type of threat*, thus a different training approach is needed to show improvement in this area. For example, a forecaster could issue a TOR and observe hail from a storm occurring in an environment with a stable boundary layer where the tornado threat would be minimal and the hail threat understandable.

Statistics were tabulated for each category as indicated in the flow chart in Figure 1. Approximately half of the TORs were true false alarms (52 of 101), and 49 verified with some sort of severe weather report and/or funnel cloud. Figure 2 depicts this and shows that about half of the TORs were followed by wind or hail reports, while only one-third were followed by tornadoes and 14% with funnel clouds. Interestingly, combining the tornado and funnel cloud categories show that half of the TORs had a distinct tornado threat, while the degree of tornadic threat in the remaining wind/hail category is unclear from this data set and analysis. A comparison of wind vs. hail indicates that each occurred with nearly equal frequency (figure not shown).

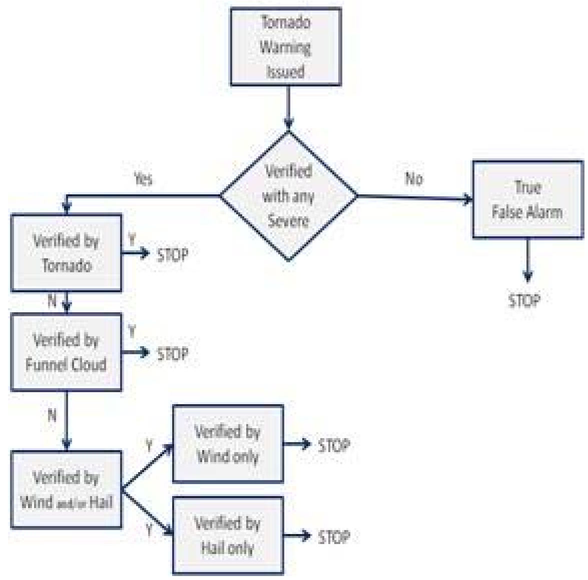

Figure 1. Flow chart depicted the decision model for sorting severe weather reports during tornado warnings.

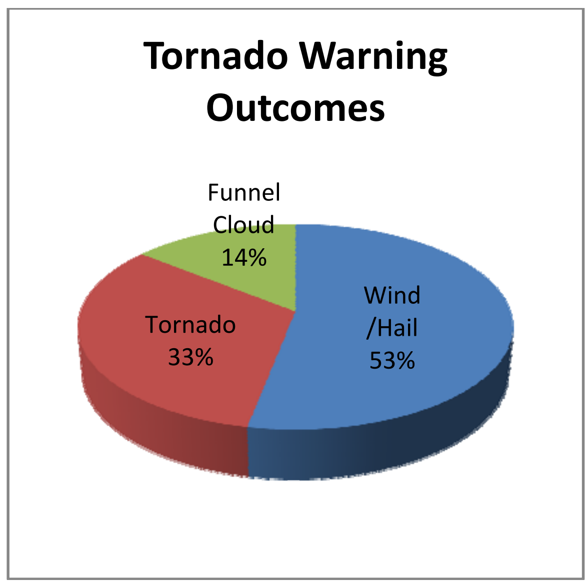

Figure 2. The percent occurrence of severe weather type during a tornado warning when severe weather is observed.

3.2 *Severe Thunderstorm Warnings*

For each SVR issued, the warnings were segregated into hail and wind forecasts then compared to the observed reports. Both the correctness of the threat type and the precision of the threat forecast were assessed.

3.2.1 Wind

There were 228 SVRs issued mentioning wind as a threat or with severe criterion wind observed. Of these, 67% (153) had severe criterion wind or wind damage observed. Tornadoes and funnel clouds each occurred in 4% (9) of the SVRs. Five percent (11) of the warnings did not mention wind, but severe criterion wind or wind damage was observed. About one-third of the SVRs mentioning wind speeds verified only by hail, thus the severe wind threat is overforecast.

A quantitative assessment of wind speed forecasts when SVRs mentioned wind and severe criterion winds were observed yielded the following: mean absolute error 9.6 mph, standard deviation 8.3 mph, bias -1.5 mph. A frequency distribution of the errors is shown in Figure 3. The practice of forecasting and commonly reporting wind speeds to the nearest 10 mph is apparent in the data set. The vast majority of the errors are 10 mph or less, though there is a cluster at 20 mph. This may reflect forecasters relying on the default 60 mph forecast too often, forecasters' conservative approach to forecasting extreme winds, or a more efficient reporting of extreme winds due to the significant damage likely to occur. While infrequent, errors of 25 mph or more suggest an opportunity to study unique cases which likely proved challenging to the warning forecasters.

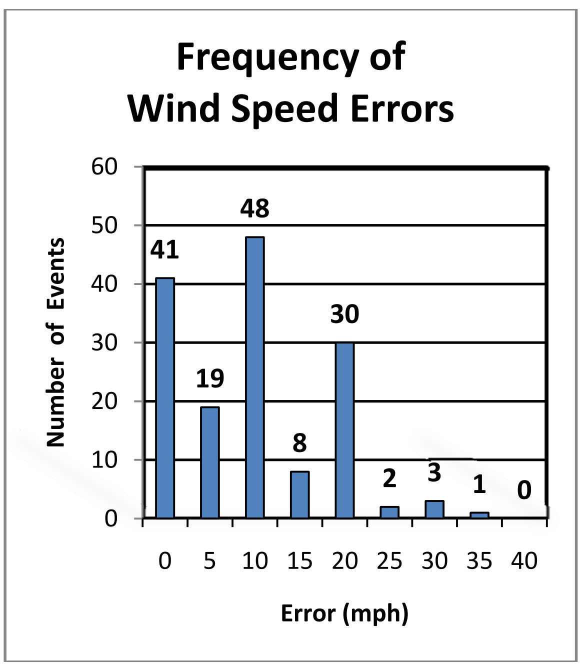

Figure 3. Frequency of wind speed errors (absolute value, mph) for SVRs when wind is forecast and observed.

3.2.2 Hail

There were 260 SVRs with hail forecast and/or observed. Seventy-one percent (184) had hail observed and forecast while only 3% (7) had hail occur and not forecast. About one-fourth of the SVRs mentioning hail verified only by wind, thus overwarning on hail is an issue. This especially seems to be true for late summer cases.

The largest proportion of hail size forecast error falls into the category of .25 inch or less (Figure 4). About 30% of the errors fall into the ranges .25-.75 inches, with many fewer errors of greater than .75 inch. This in part reflects the climatology of very large hail being rare, and when those conditions exist, they are typically readily identifiable. However, the bias of underreporting the maximum hail size which actually occurs in a severe thunderstorm could be apparent here.

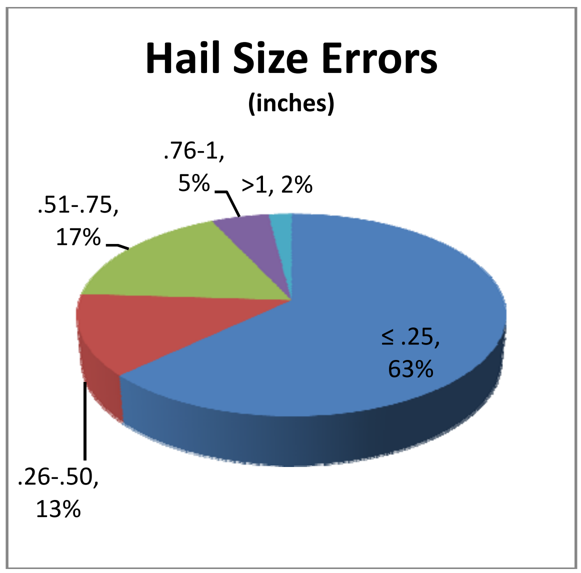

Figure 4. Frequency of hail size forecast errors by range (inches).

4.0 Conclusions

This study has served to increase understanding of the quality of severe thunderstorm and tornado warnings issued by DVN beyond the traditional measures of POD, FAR, and lead time. These results suggest continued effort is needed to understand and diagnose environmental and radar signals of tornadic and non-tornadic environments and storms. Early experience with WSR-88D super resolution data suggests an opportunity to improve tornado POD, and perhaps improve hail and wind threat identification too. Super resolution data may at least initially increase FAR until forecasters adjust to it, although that has not yet been observed since its implementation earlier this year.

Overwarning on hail and wind threats may be improved by increasing forecaster awareness to more selectively and deliberately choose the appropriate threat rather than settling for default values. The importance of accurately quantifying the wind threat in SVRs is apparent by statistics associated with severe weather fatalities in the NWS Davenport CWA since the inception of the WSR-88D in 1995. Four fatalities have occurred due to straightline winds from severe thunderstorms (all tree related, implying wind speeds around or above 75 mph), while only one fatality has occurred in a tornado (mobile home). Thus improvement in quantitative wind speed forecasts is a local area of focus.

Finally, these data provide a baseline comparison for assessing performance in light of WSR-88D upgrades. During the next two years, sufficient data should be collected to assess a potential impact of super resolution data which arrived in the middle of the 2008 severe weather season. Moreover, baseline performance will be quantified well before installation of the dual polarization upgrade. Thus we will eventually be able to assess and document the improvement in warning threat identification and quantitative assessment resulting from this significant technology upgrade.