A Comparative Analysis of Severe Weather Warnings in the Continental United States

Rachel Reeves
Jennifer Tate
Aaron Treadway

Harold Brooks
National Severe Storms Laboratory
Forecast Research and Development Division

Kim Klockow
University of Oklahoma

1. Project Summary

In the United States, the National Weather Service (NWS) is in charge of issuing warnings and advisories for the entire country. It achieves this by splitting the contiguous U.S. into 4 NWS regions and 122 different county warning areas (CWAs), each with their own weather forecast office (WFO). Warnings issued by these offices are the primary way information is communicated to the public during life-threatening severe weather events.

Each office follows certain criteria that qualify a storm as severe (50 knot winds/1 inch hail) or tornadic, but the text of the warnings, size and shape of the warnings, and when the warnings are issued on the storm are all decisions left up to the local WFO and the forecaster. These factors, along with forecast accuracy, play a role in the public's perception of the warnings issued by a WFO. By examining each of these factors, differences between offices and the 4 NWS regions will arise.

This project will serve to quantify those differences and focus on a few key questions: Do warnings issued by individual offices carry different meanings? Should the public have to worry about which CWA they live in? How do variations in warnings between offices affect forecast accuracy? Do adjacent WFOs warn severe storms in the same way? These questions are important to consider because different warning policies could possibly confuse the public that live on CWA boundaries. If one WFO warns a storm but another office does not, it could result in confusion for the general public as well as emergency managers who are in charge of disaster response.

To address these questions, data from the Iowa State University Department of Agronomy's IEM Cow (an online database that contains information regarding NWS warning size, duration, and verification information) will be collected from the past eight years, and statistical analysis on size, shape, length, lead time, and forecast accuracy will be performed. In addition, the text of some selected warnings will be analyzed to determine differences in intensity of warning wording, which will provide information regarding how the public understands and acts upon warnings.

The results of this project will have many positive effects. Forecasters will be better equipped to issue beneficial warnings when they are aware of the manner in which WFOs around them warn. Furthermore, the public will have gained valuable insight into how their local WFO warns a severe thunderstorm or tornado threat. If widespread discrepancies are found, further research may be warranted to create across-the-board warning procedures for the NWS.

2. Project Narrative

a. Introduction and Background

Several studies have been performed regarding NWS warning procedures, decision making processes, and forecast accuracy. This research plans to expand on what has been done and provide insight particularly into differences between regions and WFOs.

In Hoium et al. (1997), a study of warning procedures at the Raleigh office in North Carolina aimed to identify the key events that forecasters see before they issue a warning. These two events were an initiator and a trigger. The first (initiator) is the feature that catches a forecaster's attention and results in a forecaster taking a closer look at the storm. Examples of an initiator could be a velocity couplet beginning to form, a hook echo on radar, indications from radar that a storm is producing large hail, or strong straightline winds on velocity scans. The second (trigger) is added meteorological information that makes it clear to a forecaster that a warning needs to be issued. A majority of the time, a forecaster is waiting for another radar scan or two for the feature to become more pronounced. The article does not discuss whether office procedure, personal experience, or education play a role in the forecaster's recognition of an initiator or how long the forecaster should wait for a trigger before making a decision on the warning.

Call (2009) examines watch and warning procedures of the NWS with regard to ice storms through a survey of warning coordination meteorologists (WCMs). Many of these discoveries can be applied to severe thunderstorm and tornado warnings. Most WFOs generally follow established procedures for warning ice storms, though there is a "high level of autonomy" in text of warnings and some freedom in when to issue warnings, depending especially on region. Forecasters seem to take into account the specific hazards of their area when making decisions. Most have contact with emergency managers and the local media, which is important for warning dissemination. As Call states, "An accurate warning that is not heard or acted upon is of limited value," which is certainly the case for all hazardous weather warnings. In one 2007 winter weather event, there were three main differences in WFOs that were compared: length, intended audience, and tone. One office tended toward a longer warning, while another felt that a short warning would get the message across. One warning seemed to be directed toward emergency managers and local government officials, while one was intended for the public as a whole. This also relates to tone. The warning meant for the public was more forceful and contained “six imperative sentences and two other sentences with the subject “‘you,’” while the other warnings had less intense language. Overall, it is worth noting that WCMs take location and needs of the public into account when issuing products regarding ice storms, which should be and likely is the case for issuance of severe weather warnings as well.

Forecast skill is another important factor when looking at warnings issued by forecasters. No matter what the office policies are or how good a forecaster is at making decisions, they must exhibit some meteorological skill to be able to properly issue a warning. Roebber wrote three papers on skill in 1996. In The Contributions of Education and Experience to Forecast Skill, he examines the relationship between skill and a forecaster’s education, as well as his experiences forecasting in the past, using student and professor forecast data from a university forecast competition. Several important facts about experience are explored in the paper. The first of these is that forecasting experience is not just about being able to forecast on a daily basis, but rather the ability to focus on events that weather models poorly forecast. These events, such as extreme events like large snow storms and some severe weather events, allow the forecaster to apply his knowledge and think on his toes. In these situations a forecaster must be able to take what is being observed and let these observations be clues to the forecaster as to what is coming next. This is much like the forecaster seeing an initiator and being on the lookout for the trigger as discussed in the paper by Hoium et al. Roebber found that the experience needed to make the right decision is quickly gained in the field by implementing basic forecasting skills. As technology continues to get better and the gap in skill level closes between the weather models and human forecasters, it is important to remember that regardless of the model skill, there is always a need for a human aspect to be able to refine the forecasts made by models. In the same way humans can increase their skill by taking advantage of the weather models, radar tools, and other technological advances (Roebber and Bosart 1996).

Roebber et al. (1996) discussed another impact on forecasting skill: distance of the forecast point from the forecaster. Another national forecasting competition was analyzed to see if distance had any impact on forecast skill and the quality of forecast. It was found that local forecasters do know model biases for their immediate area so they are able to correct for these in their forecast, making them more skillful. It is generally believed that the farther away the forecast point is, the less accurate the forecast, but this article found otherwise. While there were significant differences in temperature forecast skill with distance, there was little difference in skill concerning precipitation forecasting (Roebber et al. 1996). Little variation in skill with respect to distance is logically expected for issuing warnings as well. All NWS forecasters know what makes a storm severe or tornadic, so they should be able to identify these features no matter how far away the storm is. While distance may not impact warning procedure much, the article does note that it may be important for NWS offices to keep a few forecasters at a particular office for several years so they become accustomed to the local model biases and storm trends.

Adverse weather often has economic impacts that must be considered when decision making, whether by the NWS or by local companies. In Roebber 1996a he examines several situations in which this is the case, including newspaper delivery or pouring concrete when it might rain. It provides a formula based on his Fig. 1 that takes into account net payoff of taking protection or not taking protection from a weather event, with respect to the cost and benefits of that protection. However, it qualifies the infallibility of this formula, saying that it is often useful to make “sequential decisions” instead, which would allow for reevaluation of those decisions as circumstances change. These ideas are applicable to warning issuance, as forecasters must take the societal and economic impacts their warnings generate into account as they issue them, and must continually consider changing conditions to issue new warnings or update existing warnings.

Forecaster accuracy is difficult to measure. One way to simplify forecast verification is through the use of a 2x2 contingency table that categorizes events into four categories based on whether or not the event was forecasted and whether or not the event actually happened. It is important to note that with this table, it is impossible to determine the value of D in the contingency table (Brooks 2004). This difficulty is due to the fact that one cannot accurately assess the number of correct forecasts made for events that do not exist.

2x2 Contingency Table		**Event Observed**	
		Yes	**No**
Event Forecasted	**Yes**	A	B
	No	C	D

Table adapted from Roebber 2009

From this table, five statistics can be determined that help measure the strength of a forecast: probability of detection (POD, equation 1), false alarm ratio (FAR, equation 2), bias (equation 3), critical success index (CSI, equation 4), and success ratio (SR, equation 5) (Roebber 2009). Using these statistics, one can gauge the forecast performances for different WFOs. For example, a forecast office that averaged high forecast strength would have a high POD, a low FAR, a high CSI, and a high SR. It is also possible to see how forecast strength changed over time with these statistics. A WFO that decided to start issuing more warnings would have an increase in POD, but there would also likely be an increase in FAR (Brooks 2004). Thus there is a delicate balance between trying to increase a forecast's POD while at the same time decreasing its FAR.

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

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

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

$$CSI = \frac{A}{A + B + C} \quad (4)$$

$$SR = 1 - FAR \quad (5)$$

b. Objectives

The major goal of this project is to quantify severe thunderstorm and tornado warnings in various ways to determine differences in warnings due to NWS region, WFO, and other factors. Warnings will be classified by size and shape, length, lead time, and forecast accuracy. In addition, select cases of warning text will be studied, particularly for warnings that crossed WFO boundaries. Proposed research questions include:

1. How does warning issuance differ by NWS region and WFO? This includes variations in size, length of time, and text of warnings. Extensive literature is not readily available addressing this question on such a large scale.

2. Do discrepancies in warning issuance between offices affect forecast accuracy (e.g. probability of detection, false alarm ratio, and lead times)? Statistics have been created to quantify these differences, and utilized by Roebber (2009) to analyze terminal aviation forecasts, but this work will focus on severe thunderstorm and tornado warnings issued by various WFOs.

3. How do differences in warning issuance and forecast accuracy influence risk communication and warning verbiage used by WFOs? Call (2009) examines text of warnings from three offices during a significant ice storm event, but this research will explore warning text differences in cases of severe storms.

c. Description of Project

To quantify NWS Warnings is a tremendous undertaking, but tools already exist to make the process easier. To be able to achieve the objectives set forth in the proposal, warnings from 2005-2012 will be studied across the continental United States. This encompasses the Southern, Eastern, Western, and Central regions of the NWS. Undertaking such a large-scale project is made possible by the Iowa Environmental Mesonet's Storm Based Warning Verification system called IEM COW. Through this system, online information about a warning's issuance and cancellation time, the warning's size and county coverage size, as well as whether or not the warning was verified by a Local Storm Report (LSR) is accessible. Using this tool, the data can be analyzed through Microsoft Excel and other plotting methods in a variety of ways. These include maps comparing warning size to county size, charts showing warning length trends, and graphs comparing success ratio to probability of detection. An example of this last plot can be found in Fig. 2 of the Roebber 2009 paper.

Once the warning information is presented in visual form, it will be broken down and compared by NWS region, CWA, and, in some cases, on an event scale. Noting similarities and differences in how warnings are issued allows conclusions to be made about a WFO's or NWS region's warning policies.

Another clue to a region's or WFO's warning policy is in the text of the warning. The NWS has directives that inform WFOs what products should look like (specifically Directive 10-511 for severe weather products), but there is also some leeway given for the offices to write different action statements and include other information within the warnings. Since warnings are the direct method of disseminating life-threatening information to the public, they also have the power to affect the public's perception of the NWS and of the warnings themselves. Warning text will be analyzed by looking at an NWS warning verification database. Specific cases will be examined where notable severe weather events crossed WFO boundaries to see how each office worded warnings concerning the same storms.

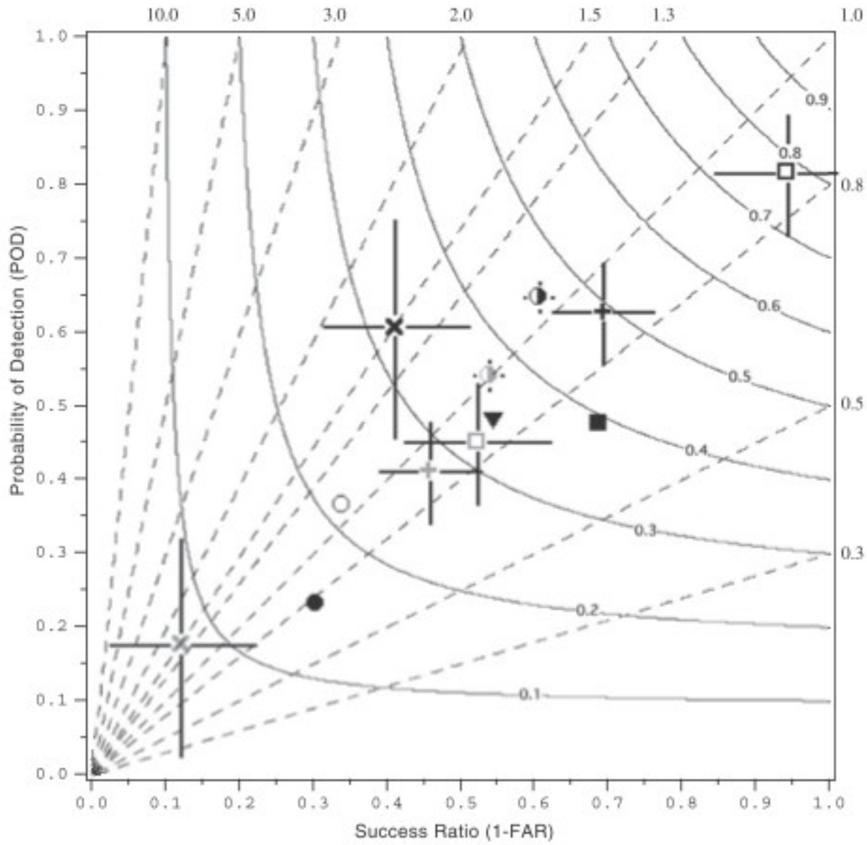

FIG. 2. Performance diagram summarizing the SR, POD, bias, and CSI. Dashed lines represent bias scores with labels on the outward extension of the line, while labeled solid contours are CSI. Sampling uncertainty is given by the crosshairs and reference "sample frequency" forecasts (where available) are in gray. Shown are 48-h forecasts of convective occurrence (bold square; Fowle and Roebber 2003), SPC forecasts of significant severe weather (filled square) and significant tornadoes (filled triangle), HPC 6–24-h forecasts of heavy precipitation (amounts in excess of 12.5 mm in 6 h) for cold (open circle) and warm (filled circle) seasons, heavy and light snow densities from a neural network [boldface multiplication and plus symbols, respectively; Roebber et al. (2003)], TAFs from NWS WFOs (bold half-filled circle), and MOS (gray half-filled circle).

Figure 2 from Roebber 2009. We will adapting the performance diagram to apply to warning statistics.

d. Broader Impact

This project's analysis of warning size, length of warnings and lead times, and statistics on forecast accuracy can help the public in understanding various warnings issued by different WFOs. It is possible that research obtained in this project will show that different WFOs have different standards for issuing warnings. It could be, for example, that some offices prefer to issue long, rectangular-shaped warning cones so that lead times would be longer, while other offices prefer to issue smaller warning cones that expire sooner so that the cone could be easily adjusted as the storm progresses. A study of survey responses from visitors to the National Weather Center in Norman, OK found that on average, the preferred tornado lead time is 34.3 minutes (Hoekstra et al 2010). Knowing this, certain offices with shorter average lead times might be inclined to adjust their warning procedure to match that of other offices whose average lead time is longer.

In addition, it is expected that adjacent WFOs with similar weather patterns have a similar number of warnings issued per year, as well as a similar POD, FAR, etc. When adjacent offices have statistics that are vastly different, there is cause for concern. Someone traveling across county lines into another WFO should expect the same (or at least relatively similar) issuance of warnings from the two forecast offices, whether they are relocating because of a job or simply going shopping.

3. Statement of Work

The initial research will be split into three categories: length of warning and lead time analysis, forecast accuracy statistics, and warning size analysis. Jennifer will work on collecting data related to warning lengths and lead times. Rachel will collect information on forecast accuracy to create statistics on strength of forecasts. Aaron will study the differences in size of warnings and compare the polygon system to the county-based system. Initial data collection will be completed by the beginning of March. After initial research is finished, each person will work on analyzing his/her data and making comparisons between different WFOs and NWS Regions. The text of warnings for specific weather events will also be studied, and that work will be split evenly between the group as well. This should be completed near the beginning of April, when all the data collected will be assimilated together and conclusions collectively drawn by the group as a whole. Finally, a report of findings and a poster presentation will be jointly written and presented by all group members.

References:

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.

Call, D. A., 2009: An assessment of national weather service warning procedures for ice storms. *Wea. Forecasting*, **24**, 104–120.

Hoekstra, S., H. Brooks, J. Brotzge, S. Erickson, K. Klockow, and R. Riley, 2010: A preliminary look at the social perspective of warn-on-forecast: preferred tornado warning lead time and the general public's perceptions of weather risks. *Wea. Climate Soc.*, **3,** 128-140.

Hoium, Debra K., A. J. Riordan, J. Monahan, K. K. Keeter, 1997: Severe thunderstorm and tornado warnings at Raleigh, North Carolina.*Bull. Amer. Meteor. Soc.*, **78**, 2559–2575.

Roebber, P. J. and L. F. Bosart, 1996: The complex relationship between forecast skill and forecast value: a real-world analysis. *Wea. Forecasting*, **11**, 544–559.

Roebber, P.J., L. F. Bosart, 1996: The contributions of education and experience to forecast skill. *Wea. Forecasting*, **11**, 21–40.

Roebber, P. J., L. F. Bosart, G. S. Forbes, 1996: does distance from the forecast site affect skill?. *Wea. Forecasting*, **11**, 582–589.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608