FORECASTER'S FORUM

Subjective Tornado Probability Forecasts in Severe Weather Watches

MICHAEL D. VESCIO* AND RICHARD L. THOMPSON

Storm Prediction Center, Norman, Oklahoma

(Manuscript received 6 April 2000, in final form 3 October 2000)

ABSTRACT

An experiment was conducted at the Storm Prediction Center (SPC) to assess the accuracy of subjective probability forecasts for tornadoes within individual convective watch areas. Probability forecasts for one or more and three or more tornadoes were produced for 166 severe weather watches during 1997 and 1998. Categorical forecasts of maximum tornado intensity, as indicated by F-scale damage ratings, were also performed. The probability and intensity forecasts were made in an operational setting prior to the issuance of each watch to simulate the decision making process that might be employed if the SPC were to begin including probabilities in their watch products. Results indicate considerable skill in forecasting tornado probabilities, though the maximum intensity forecasts were not particularly accurate. It is hypothesized that accurate tornado intensity forecasts will be difficult to achieve until storm-scale processes are more fully understood.

1. Introduction

The Storm Prediction Center (SPC) has the responsibility of issuing tornado and severe thunderstorm watches for the contiguous United States. Because the knowledge of severe convective storms and their antecedent conditions has increased dramatically in the past decade, the opportunity may now exist for SPC forecasters to provide additional and more specific information in convective watches, similar to the maximum hail size and wind gust estimates already provided in the public severe thunderstorm watch product. An experiment was conducted at the SPC in 1997 and 1998 to determine the accuracy of subjective probability and intensity forecasts for tornadoes. During the 2-yr period, probability forecasts accompanied 166 watches (108 tornado and 58 severe thunderstorm), with a total of 229 tornadoes occurring within the watches sampled here. Although several hundred watches per year are issued by the SPC, only the two authors participated in this experiment. Further, due to the labor intensive process of collecting and archiving the necessary data in real time, probability and intensity forecasts were made only when there was still sufficient time to complete our required forecast duties. Although a larger sample size would have been preferable, enough cases were collected to draw some preliminary conclusions on the potential skill in making probability and intensity forecasts for tornadoes in an operational setting. Results from this experiment are discussed herein, with a goal of incorporating probabilities into severe thunderstorm and tornado watches issued by the SPC.

2. Methodology

For each watch, probability forecasts were made for one or more tornadoes and three or more tornadoes occurring in the watch area. The median watch size in this study was 25 000 mi^2 with 70% of the watches falling within the range of 15 000–30 000 mi^2. It is recognized that the probability of tornadoes is typically not uniform throughout the watch area, and that the probability of a tornado per unit area will increase with increasing watch size given identical meteorological conditions. The intent of this study, however, was to determine the accuracy of probability forecasts *anywhere* within the watch area, with the watch size dictated by the meteorological situation. Further, an analysis of the verification of watches by size from this dataset indicates that larger watches were no more likely to contain tornadoes than their smaller counterparts. For example, 30% of the watches with an area less than 25 000 mi^2 verified with at least one tornado, while watches greater than 25 000 mi^2 had a 34% verification rate.

---

* Current affiliation: NWS Forecast Office, Fort Worth, Texas.

Corresponding author address: Michael D. Vescio, National Weather Service, Forecast Office, 3401 Northern Cross Blvd., Fort Worth, TX 76137.
E-mail: Michael.Vescio@noaa.gov

Probabilities were assigned before each watch was transmitted to simulate the procedure that might be followed if these forecasts were to be disseminated as part of the public watch product. At the inception of the experiment, it was decided that the following arbitrary probabilities would be used: 1%, 5%, 25%, 50%, 75%, and 95%. As mentioned in section 1, there were 58 severe thunderstorm watch cases in this experiment (35% of the database). These watches were included because, even in severe thunderstorm watches, the probability of a tornado is greater than zero. For example, in 1999 about 25% of the severe thunderstorm watches issued by the SPC had one or more tornadoes in them. Eleven of the 58 severe thunderstorm watches (19%) in this dataset contained at least one tornado. However, all of these tornadoes were brief and weak. The SPC has developed considerable skill at differentiating between tornadic and largely nontornadic environments during the past several years, as the scientific understanding of storm-scale processes has increased. For example, the environments conducive for supercells can be readily determined by examining such parameters as the bulk Richardson number (BRN) shear (Weisman and Klemp 1982), and measures of lifted parcel buoyancy such as convective available potential energy (CAPE). Within supercell environments, the threat for tornadoes can be assessed through examination of parameters such as storm-relative helicity (Davies Jones et al. 1990), midlevel storm-relative winds (e.g., Brooks et al. 1994; Thompson 1998), and various combinations of CAPE and low-level shear such as the energy helicity index (Hart and Korotky 1991; Rasmussen and Blanchard 1998). With this increased knowledge, some forecasters at the SPC have begun to issue severe thunderstorm watches in situations when supercells are expected, but the threat for more than brief weak tornadoes is minimal.

There has been little published research on the operational forecasting of tornado intensity. Colquhoun and Riley (1996) did develop a regression equation to forecast F scale (Fujita 1971) based on environmental buoyancy and shear from tornado proximity soundings. However, they did not sample environments when severe convection developed but tornadoes failed to occur. Although testing the regression equation operationally would be useful, the intent of this study was to *subjectively* forecast tornado intensity. Maximum tornado intensity forecasts were made for each watch using the following categories: weak (F0–F1), strong (F2–F3), and violent (F4–F5). It is recognized that there are problems when forecasting tornado intensity based on the F scale. Because the F scale is a damage scale, a tornado must cause damage before it can be rated anything other than F0. Despite the inherent difficulties in using the F scale to judge the intensity of tornadoes, tornado intensity forecasting is already being done to some extent in SPC outlooks and watches, and this experiment provided some insight into the potential skill (discussed in section 3) of making these forecasts.

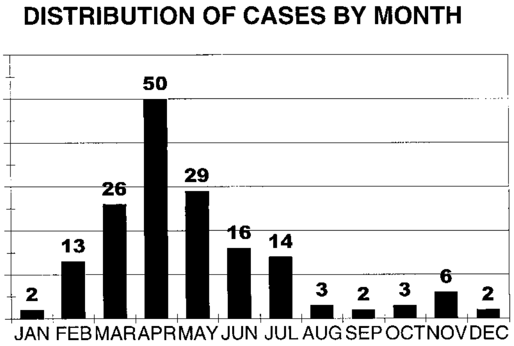

FIG. 1. Distribution of watch cases by month.

Probability and intensity forecasts were subjectively determined based on the examination of several observed and forecast parameters in the watch area, in a manner consistent with other SPC severe weather forecast products. For example, BRN shear, environmental buoyancy, convective inhibition, storm-relative helicity, and storm-relative flow were routinely examined using observed and short-term model data (e.g., Johns and Doswell 1992; Davies 1993). The presence (or absence) of mesoscale boundaries was also considered. Studies by Maddox et al. (1980) and Markowski et al. (1998), as well as the personal forecast experience of the authors, suggest that storms interacting with preexisting boundaries are more likely to produce tornadoes. Finally, the expected mode of convection affected the probability and intensity forecasts. Low probabilities and weak intensities were generally assigned in situations when squall lines were expected, whereas the highest probabilities and intensities were assigned when long-lived classic supercells were anticipated.

3. Results

Before proceeding further, it is important to note, for comparison purposes, that during the 1990s approximately 50% of all tornado watches verified with at least one confirmed tornado (compared to 49% in this study). Figure 1 displays the monthly distribution of watch cases during the 2-yr experiment. Although all months were represented in this study, most of the cases occurred between March and July with a peak of 50 watches in April. This distribution is not surprising because most severe weather (and therefore the most watches) occur in the spring and early summer. Figure 2 shows the distribution of probability forecasts for one or more and three or more tornadoes. The most common probability used for one or more tornadoes was 50% (68 cases), followed by 25% (49 cases), with all of the arbitrarily designated probabilities used in the actual forecasts. In contrast, the most frequent probability used for forecasting three or more tornadoes was 5% (75 cases), followed by 25% (55 cases), with 95% not used during the experiment. The lower probability forecasts of three

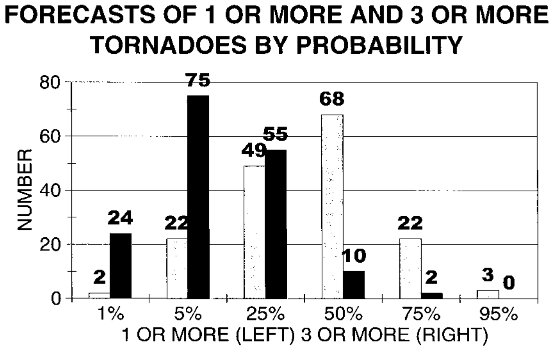

FIG. 2. Forecasts of (left) one or more tornadoes and (right) three or more tornadoes by probability.

or more tornadoes compared to one or more tornadoes reflects the natural tendency to have less confidence that multiple events will occur. When stratifying the probability forecasts by watch type, it is clear (Figs. 3 and 4) that lower probabilities were typically forecast for severe thunderstorm watches. It is also important to note that 21 tornado watches had a forecast probability of 25% or less for one or more tornadoes. At first glance, it might appear that some of these watches could have been severe thunderstorm watches. However, most of these low probability tornado watches were issued in the cool season (Oct–Mar) when environmental wind shear is very strong but instability is marginal. In these situations, the probability of tornadoes may be low, yet strong–violent tornadoes can occur.

Figure 5 is the reliability diagram (Wilks 1995) for the probability forecasts of one or more tornadoes. What is immediately obvious from the figure is that the frequency of occurrence (in percent) of one or more tornadoes increased with increasing probability forecasts, which suggests that these forecasts were reliable. For example, at least one tornado occurred 47% of the time when a 50% probability was forecast. Reliability (R) can be calculated using the following equation:

$$R = \frac{1}{N}\sum_{i=1}^{I} n_i(f_i - o_i), \qquad (1)$$

where N is the total number of forecasts, n the number of forecasts for a given probability, f the probability forecast value (in this case 0.01, 0.05, 0.25, 0.50, 0.75, and 0.95), and o the relative frequency of occurrence for a given probability forecast. Values of reliability can range from 0 (perfectly reliable) to 1. The reliability for the probability forecasts of 1 or more tornadoes was 0.004.

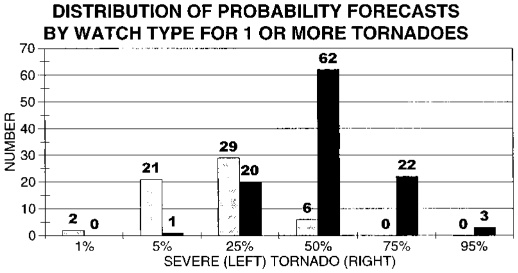

FIG. 3. Distribution of probability forecasts for one or more tornadoes by watch type: (left) severe thunderstorm watches and (right) tornado watches.

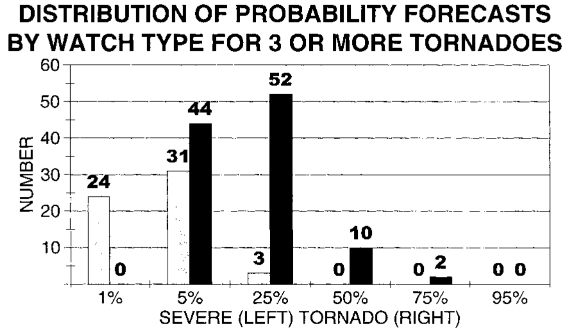

FIG. 4. Same as Fig. 3 except probability forecasts are for three or more tornadoes.

Another statistic commonly used to assess the accuracy of probability forecasts is the Brier score (Wilks 1995). The National Weather Service uses the Brier score in precipitation probability forecast verification. The Brier score is the mean squared error of probability forecasts. A perfect forecast has a Brier score of 0, whereas a Brier score of 1 indicates no forecast skill. The Brier score for the probability forecasts of one or more tornadoes was 0.19. Results were similar for the forecasts of three or more tornadoes (Fig. 6), with a reliability of 0.003 and a Brier score of 0.12.

Table 1 summarizes the intensity category forecasts and verification. Weak maximum intensity was by far

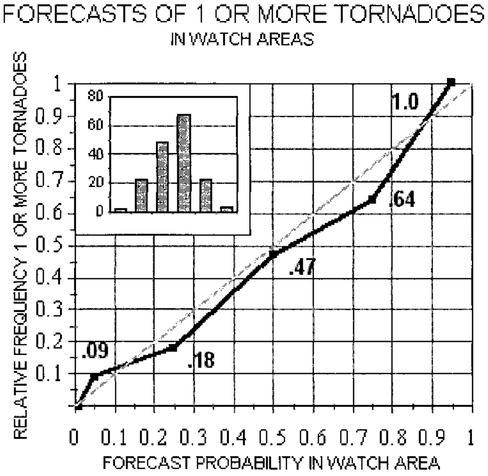

FIG. 5. Reliability diagram for the forecasts of one or more tornadoes in watch areas. The heavy line represents the actual frequency of occurrence of tornadoes for each probability forecast group. Straight line ($y = x$; in lighter shading) indicates perfect reliability. Inset bar graph displays the number of forecasts of each probability (1%, 5%, 25%, 50%, 75%, and 95% from left to right).

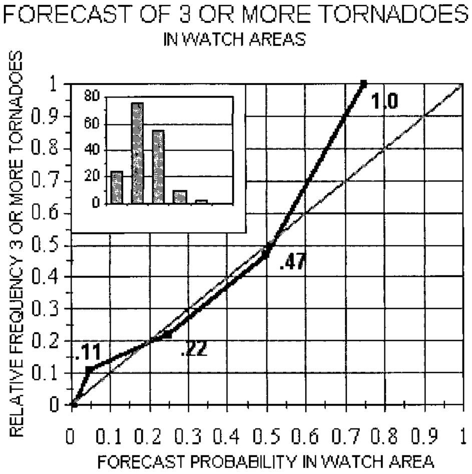

FIG. 6. Same as in Fig. 5 except for forecasts of three or more tornadoes in watch areas.

the most common forecast (114 of 166 cases, or 69%), and there were no forecasts of violent tornadoes (no violent tornadoes occurred in any of the watches used for this study). Strong tornadoes occurred in only 10 watches, weak tornadoes occurred 53 times, and no tornadoes were reported in 103 watches. These results clearly indicate a tendency to overforecast strong tornadoes. For example, only 8 of the 31 (26%) watches with tornadoes in them and a strong intensity forecast actually verified with strong tornadoes.

4. Discussion

It is clear from this experiment that tornadoes can be forecast accurately within convective watch areas. Reliability values (Brier scores) for forecasts of one or more tornadoes and three or more tornadoes were 0.004 (0.19) and 0.003 (0.12), respectively. Given such reliability, a logical next step would be for the SPC to provide probability information to the user community. In fact, the SPC plans to incorporate tornado probabilities into the public watch product after further study. In future experiments, the forecast probability values in watches may be derived in a manner similar to the SPC experimental probabilistic outlooks. Specifically, the probability values in the watch area could represent the probability of the occurrence of a tornado within 25 miles of any point in a watch (rather than the probability of a tornado *anywhere* in the watch).[1] This would, in effect, normalize watches of varying size. It would also be useful to provide probability contours within watch areas to highlight the region of greatest threat. However, most watches are issued early in the formative stage of severe convection, and at present operational models do not simulate convection reliably enough to consistently delineate areas of enhanced risk within watch areas.

TABLE 1. Verification of intensity forecasts by damage category: weak (F0–F1) and strong (F2–F3).

Forecast	Observed		
	None	Weak	Strong
Weak	82	30	2
Strong	21	23	8

Examination of Table 1 suggests a substantial tendency to overforecast maximum tornado damage intensity, especially for the strong tornado forecasts. Though the overforecasting bias would likely be reduced with a much larger sample size, tornado intensity and duration forecasts will continue to have only limited success until most tornadoes are clearly and consistently documented and tornadogenesis is more thoroughly understood.

Acknowledgments. The authors would like to thank Mike Kay and Bob Johns of the SPC, Harold Brooks from the NSSL, and the two anonymous reviewers for providing suggestions that greatly improved this manuscript.

REFERENCES

Brooks, H. A., C. A. Doswell III, and R. B. Wilhelmson, 1994: The role of midtropospheric winds in the evolution and maintenance of low-level mesocyclones. *Mon. Wea. Rev.,* **122,** 126–136.

Colquhoun, J. R., and P. A. Riley, 1996: Relationships between tornado intensity and various wind and thermodynamic variables. *Wea. Forecasting,* **11,** 360–371.

Davies, J. M., 1993: Hourly helicity, instability, and EHI in forecasting supercell tornadoes. Preprints, *17th Conf. on Severe Local Storms,* St. Louis, MO, Amer. Meteor. Soc., 107–111.

Davies-Jones, R., D. Burgess, and M. Foster, 1990: Test of helicity as a tornado forecast parameter. Preprints, *16th Conf. on Severe Local Storms,* Kananaskis Park, AB, Canada, Amer. Meteor. Soc., 588–592.

Fujita, T. T., 1971: A proposed characterization of tornadoes and hurricanes by area and intensity. SMRP Paper 91, University of Chicago, 42 pp.

Hart, J. A., and W. D. Korotky, 1991: The SHARP workstation—V1.5. A Skew-T/hodograph analysis and research program from the IBM and compatible PC. User's manual. NOAA/NWS Forecast Office, Charleston, WV, 62 pp.

Johns, R. H., and C. A. Doswell III, 1992: Severe local storms forecasting. *Wea. Forecasting,* **7,** 588–612.

Maddox, R. A., L. R. Hoxit, and C. F. Chappell, 1980: A study of tornadic thunderstorm interactions with thermal boundaries. *Mon. Wea. Rev.,* **108,** 322–336.

Markowski, P. M., E. N. Rasmussen, and J. M. Straka, 1998: The occurrence of tornadoes in supercells interacting with boundaries during VORTEX-95. *Wea. Forecasting,* **13,** 852–859.

Rasmussen, E. N., and D. O. Blanchard, 1998: A baseline climatology of sounding derived supercell and tornado forecast parameters. *Wea. Forecasting,* **13,** 1148–1164.

Thompson, R. L., 1998: Eta Model storm-relative winds associated with tornadic and nontornadic supercells. *Wea. Forecasting,* **13,** 125–137.

Weisman, M. L., and J. B. Klemp, 1982: The dependence of numerically simulated convective storms on vertical wind shear and buoyancy. *Mon. Wea. Rev.,* **110,** 504–520.

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

---

[1] This capability was not available at SPC during the time of this study.