False Alarms and Close Calls: A Conceptual Model of Warning Accuracy

LINDSEY R. BARNES* AND EVE C. GRUNTFEST

Department of Geography and Environmental Studies, University of Colorado at Colorado Springs, Colorado Springs, Colorado

MARY H. HAYDEN

Institute for the Study of Society and Environment, National Center for Atmospheric Research, Boulder, Colorado

DAVID M. SCHULTZ+

Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, and NOAA/National Severe Storms Laboratory, Norman, Oklahoma

CHARLES BENIGHT

Department of Psychology and the CU Trauma, Health, and Hazards Center, University of Colorado at Colorado Springs, Colorado Springs, Colorado

(Manuscript received 6 September 2006, in final form 20 December 2006)

ABSTRACT

The false alarm rate (FAR) measures the fraction of forecasted events that did not occur, and it remains one of the key metrics for verifying National Weather Service (NWS) weather warnings. The national FAR for tornado warnings in 2003 was 0.76, indicating that only one in four tornado warnings was verified. The NWS's goal for 2010 is to reduce this value to 0.70. Conventional wisdom is that false alarms reduce the public's willingness to respond to future events. This paper questions this conventional wisdom. In addition, this paper argues that the metrics used to evaluate false alarms do not accurately represent the numbers of actual false alarms or the forecasters' abilities because current metrics categorize events as either a hit or a miss and do not give forecasters credit for close calls. Aspects discussed in this paper include how the NWS FAR is measured, how humans respond to warnings, and what are alternative approaches to measure FAR. A conceptual model is presented as a framework for a new perspective on false alarms that includes close calls, providing a more balanced view of forecast verification.

1. Introduction

The fraction of all tornado warnings issued by the National Weather Service (NWS) across the United States in 2003 that did not verify was 0.76 (NWS 2006). In other words, given four tornado warnings, only one was associated with a reported tornado. This fraction is called the false alarm rate (FAR), where a false alarm is "an event . . . forecast to occur but did not" (Wilks 2006, p. 261). An ideal forecast would have an FAR of 0.00, but the uncertainties in forecasting technology, uncertainties in forecasting science, and uncertainties in verification likely make this an unattainable goal.

Values for FAR differ by the type of weather event, seemingly indicating the varying levels of difficulty of forecasting, or verifying, different weather phenomena. For example, the 2005 national FAR for flash floods was 0.46 (M. Mullusky 2006, personal communication), the 2004–05 (October 2004–September 2005) national FAR for winter storm warnings was 0.31, the 2004–05 (October 2004–September 2005) national FAR for high wind warnings was 0.31, the 2005 (January–December 2005) national FAR for severe thunderstorm warnings

* Current affiliation: Hazards and Vulnerability Research Institute, Department of Geography, University of South Carolina, Columbia, South Carolina.

+ Current affiliation: Division of Atmospheric Sciences, Department of Physical Sciences, University of Helsinki, and Finnish Meteorological Institute, Helsinki, Finland.

Corresponding author address: Lindsey R. Barnes, Dept. of Geography, University of South Carolina, Columbia, SC 29208.
E-mail: lindseyrbarnes@yahoo.com

DOI: 10.1175/WAF1031.1

© 2007 American Meteorological Society

was 0.48, and the 2005 (January–December 2005) national FAR for combined severe thunderstorm and tornado warnings was also 0.48 (B. MacAloney 2006, personal communication). The FAR for forecasts of tropical cyclone genesis in 2003 was 0.43 for the eastern North Pacific Ocean basin and 0.32 for the North Atlantic Ocean basin (Brown et al. 2004). In contrast, the higher FAR associated with tornado warnings indicates the difficulties forecasters face in determining whether convective storms will produce a tornado or not (e.g., Moller 2001, p. 447) and verifying whether such events actually occur (e.g., Speheger et al. 2002).

One question is whether an FAR of 0.00 is even attainable. One event that was generally considered to be forecast well was the 3 May 1999 tornado outbreak in Oklahoma (see the June 2002 issue of *Weather and Forecasting* devoted to this event; Vol. 17, No. 3). Even for this widely regarded, well-forecast outbreak, the FAR for tornado warnings was still 0.29 (Andra et al. 2002).

Repeated overwarnings, or a high FAR, are often viewed as problematic in the warning community because of anticipated complacency among the population being warned. Consequently, NWS policy actively seeks to reduce this measure. For example, by 2010, the goal of the NWS is to reduce the tornado FAR by 0.06 to 0.70 (NWS 2006). But, as demonstrated by Brooks (2004), efforts to reduce the FAR may lead to the unintended consequence of not warning for events that do occur.

This paper is intended to initiate discussion about the problems in the methods of calculating the FAR and the implications of the stated FAR values and goals. What is a reasonable goal of FAR for various weather phenomena for the NWS? What is an attainable goal for FAR? Does an FAR of 0.76 for tornado warnings imply the NWS is doing a poor job, or does it adequately represent the generally excellent job that forecasters do?

Section 2 of this paper discusses how the verification statistics like FAR are calculated by the NWS, including factors that may introduce bias into these statistics. Section 3 discusses the human factors associated with the response to warnings. Section 4 presents a conceptual model that describes warning accuracy, incorporating close calls into the verification statistics. This conceptual model suggests that a more balanced measure of the success of warnings is possible. Section 5 concludes this paper.

2. NWS FARs

The FAR is calculated using a 2 × 2 contingency table (Table 1). The FAR is defined as the ratio of false alarms, or unverified warnings, to all warnings issued: FAR = $Z/(X + Z)$, where X represents all forecasts that were verified to have occurred with the described intensity, spatial extent, and temporal extent, and Z represents all forecasts that did not occur with the forecasted intensity, spatial extent, and temporal extent (e.g., Schaefer 1990). The values in the contingency table are obtained from two separate databases maintained by each of the 122 NWS Weather Forecast Offices (WFOs); one contains warnings issued, and the other contains verification information for those warnings. Once a month, the two databases are compared, verification statistics (like FAR) are calculated, and the results are reported to NWS headquarters for computation of national averages.

TABLE 1. A 2 × 2 contingency table for forecasting of events, where X = fraction of correct forecasts of events, Y = missed forecasts of events, Z = incorrect forecasts of events that did not occur, and W = correct forecast for events that did not occur.

		Forecast	
		Yes	No
Observed event	Yes	X	Y
	No	Z	W

Verifying observations are collected from official NWS storm surveys, official NWS observations from surface observing stations, and unofficial observations from amateur radio operations, newspaper clippings, emergency managers, trained spotters, law enforcement, and the public. The resources that are allocated to verification are not as great as those that are allocated for forecasting and warning issuance, hence the reliance on volunteers and unofficial observations. Resources for verification are also a function of staffing and time. In contrast to isolated weather events, workloads increase when subsequent severe weather days occur. With attention focused on issuing timely warnings and with more warnings being issued, verifying all warnings issued under such duress may be challenging. Spotter networks for verification are sparse in many rural areas across the United States; thus, many actual events may go unreported in the verification database. The difficulty in verifying weather phenomena in rural areas may affect the verification statistics, even on a national level. Some have argued that forecasters may be reluctant to issue warnings in sparsely populated areas because the warning may not verify (e.g., Edwards 2006). Furthermore, because verification is done by the WFO rather than by an impartial entity, verification is often biased. For example, when a warning is issued, the WFO may make extra efforts to search the volunteer network to find a verifying report. In con-

trast, if no warning is issued for a potentially missed event, similar efforts to find a report that would verify a missed event are likely not as substantial. Such inconsistencies have the potential to introduce biases into the statistics (e.g., Weiss et al. 2002; Trapp et al. 2006).

Although the calculation of FAR is well posed mathematically, we argue that the way the statistic is calculated by the NWS fails to give proper credit to forecasters. Specifically, the statistics computed by the NWS are classified as either hits or misses, and the statistics fail to account for close calls. Close calls are defined as "something achieved (or escaped) by a narrow margin" (see the Web site http://dictionary.reference.com/). Each warning is associated with a specific phenomenon, specific area, and specific time. If an event occurs with slightly less intensity than predicted, just outside of the warned area, or immediately after the warning expiration time, no credit is received for these events. Such events would be considered close calls.

Consider two days with tornado warnings issued for a specific county in the United States. On the first day, a mesocyclone moves across the county producing damaging winds and large hail, but no tornado. On the second day, sunny skies prevail and no severe weather occurs. Both days would be counted as false alarms because tornadoes were not verified on either day. Clearly, the forecast on the first day is better than that on the second day, but the statistics compiled by the NWS would count these two false alarms exactly the same.

3. Perceptions of false alarms

The human response to warnings for natural hazards is affected by many factors. One factor that academics and practitioners debate is the effect of false alarms. Conventional wisdom is that overwarning reduces the public's willingness to respond to future warnings. In contrast, more recent research indicates the public may have a high tolerance for false alarms. The conventional wisdom is also known as the *cry-wolf effect*, adapted from Aesop's well-known fable. Similarly, Breznitz (1984) defines the *false alarm effect* as "the credibility loss [of a warning system] due to a false alarm." Breznitz (1984) used laboratory experiments to study physical reactions to repeated false alarms, finding that repeated false alarms reduce subjects' willingness to respond.

Evidence for the cry-wolf effect in natural hazards research, however, has not been forthcoming. Drabek (1986, 77–78) found that Breznitz's (1984) research supporting the cry-wolf effect fails to account for the effects of social context or media attention that would lend credibility to an event. In studies conducted over a 2-yr period of several earthquake "near predictions" in Los Angeles County, Turner (1983) found that a threat is more credible the more frequently it is discussed, both through media and informal discussion. Atwood and Major (1998) found evidence of both a false alarm effect and a mobilization effect after Iben Browning's 1990 unofficial false earthquake prediction for the New Madrid region of the central United States. Although 46.1% of survey respondents reported a false alarm effect, 16.7% reported a greater concern for future earthquakes. Janis (1962) found that false alarms may not reduce peoples' willingness to take protective actions in future warnings and may even create a higher level of vigilance if there is an understanding of the event and the reason for the warning. Janis (1962) emphasized that an increased tendency to follow future warnings is affected by two major factors: an increase in information and an increase in the understanding of one's vulnerability to the hazard. These results were corroborated more recently by Dow and Cutter (1998) in their examination of evacuation behaviors of residents of South Carolina who had experienced false alarm–near-miss situations for Hurricanes Bertha and Fran. Dow and Cutter (1998) determined that these experiences did not affect residents' perception of risk; residents stated that they would make few changes in future evacuation plans. Although Dow and Cutter (1998) did not find evidence of the cry-wolf effect, they highlight that the cry-wolf effect is often a widespread source of speculation and concern within the warning community.

Other studies reiterate that false alarms or close calls are not necessarily detrimental to appropriate responses. A study of a dam-failure false alarm in which 14 000 people were in the inundation zone in Ventura, California, found that, although surveyed populations may have experienced frustrations, the respondents were not negatively affected by the false alarm (Carsell 2001). Rather, the false alarm provided a learning opportunity of appropriate responses such as attaining knowledge of evacuation plans for future events.

While Breznitz (1984) examined repeat false alarms in a laboratory setting and Dow and Cutter (1998) examined the impacts of repeated near misses for hurricane evacuations, little has been done to examine how repeated actual false alarms affect warning response in the context of real events. Carsell (2001) found that an isolated false alarm was not detrimental to an appropriate response. However, instances of repeated actual false alarms are rare, and there has been little opportunity to research the impacts of multiple false alarms on warning response. If and when such events occur,

future research would be valuable to understanding the societal response to multiple false alarms, particularly if there is a threshold at which a number of actual false alarms within a specific time period affects the willingness to respond.

Studies surveying emergency management personnel have found that internal false alarms or close calls do not have negative effects; rather, they can provide learning opportunities to improve warnings, responses, and protocols, as well as to test lines of communication and new technologies (Gruntfest and Carsell 2000; Weaver et al. 2000; Rhatigan et al. 2006). Although emergency management personnel's confidence may not be lessened by false alarms, Rhatigan et al. (2006) found that some respondents perceived that the public's confidence would be lessened by a false alarm. In fact, those issuing warnings may be more reluctant to issue warnings for the fear of issuing a false alarm (e.g., Gruntfest and Carsell 2000; Weaver et al. 2000).

Although overwarning is often viewed as a problem, there may be reasons to warn a larger area than that which will be directly affected. For example, global positioning system (GPS) technology allows warnings to be delivered to cell phones at a precise location. Although a person may not be in the path of the tornado, his or her family or friends may be. For instance, children at a friend's house, a spouse on the way home from work, or an elderly relative may be in the path and be unaware. In this situation, knowing about an event that will not affect your exact location can enable one to contact family members to make sure they are safe and taking protective actions or ensure that people will not drive from a safe location into the path of the tornado. How to account for this beneficial aspect of overwarning, however, is not simple.

Is a high FAR for some types of weather phenomena more acceptable than for other types? For example, a much higher FAR may be tolerated for short-fuse weather events where the amount of time and effort expended by the public may be limited to tens of minutes for a false alarm tornado warning. In contrast, a high FAR for hurricane evacuations where the disruption and financial expense could be significant may not be viewed in the same way by the public. Clearly, this factor also needs to be considered.

Another factor to consider in the impact of false warnings on the public response is how many people actually receive all NWS warnings issued. Many challenges exist to effectively disseminating warning messages to all people at risk. Unless a person is watching television, listening to the radio, using a National Oceanic and Atmospheric Administration (NOAA) weather radio with tone alert turned on, subscribing to a notification service, or hearing a warning siren, the warning may never be received. The ability of warning messages to penetrate a person's normal activities is highly dependent on what activities people are engaged in and the time of day or night the warning is issued (Lindell and Perry 2004, p. 76). If a person is at work, shopping, or sleeping, the likelihood of not receiving a warning message is greater than if a person is watching prime-time television. Additionally, language barriers can limit the number of warnings received for non-English-speaking immigrant and minority populations (Lindell and Perry 2004, p. 193).

Despite all this concern about high FARs, the public may not sense that high FARs are a problem. The NWS customer satisfaction survey (NOAA 2006) found that, overall, people are very satisfied with NWS products and warnings and, specifically, they were very satisfied with hazardous weather information. Yet, this survey does not ask any questions about perceptions of false alarms or FAR. Although the results of this survey did show that the public is satisfied with the NWS, the report did not discuss whether the performance measures that are used by the agency as a yardstick to measure the success of warnings are the best measures. The high rating the NWS received in this survey seems to be contrary to the high FARs. Would the public still have high confidence in the NWS if they knew that 76% of tornado warnings were not verified?

4. Conceptual model

Common definitions and public understanding of false alarms do not mirror the NWS definition of false alarms. The FAR includes events that could not be verified because they occurred in sparse spotter networks, do not meet intensity thresholds, or occurred but not within the time frame or geographic locations specified in the warning.

We have developed a conceptual model that presents a broader, more general depiction of warnings for possible events, including false alarms and close calls. This conceptual model provides a new framework for viewing the accuracy of warnings, suggesting that close calls are different from false alarms and should be categorized differently (Fig. 1). This conceptual model envisions hazardous weather events on a spectrum, ranging from events that occur but are not warned for (unwarned events) on one end to events that do not occur but for which a warning is issued (false alarms) on the other. An event that occurs and for which there is a perfect warning, lies in the center of the spectrum (perfect forecast). This model is a way to map a spectrum of events similar to the four-panel contingency table in

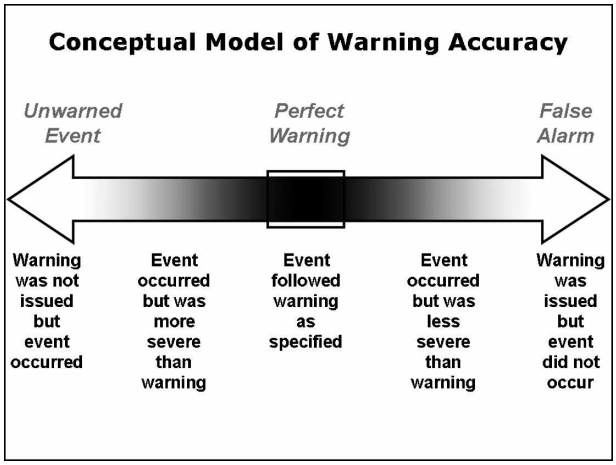

FIG. 1. Conceptual model of warning accuracy.

Table 1, where a perfect warning is represented by the panel with *X*, a false alarm is represented by the panel with *Z*, and an unwarned event is represented by the panel with *Y*. This model does not include the panel represented by *W* because unforecasted nonevents are unlikely to affect public perception.

Although there are similarities between this conceptual model in Fig. 1 and the contingency table in Table 1, there is an important difference. Rather than discrete boxes of yes or no, a spectrum is used to demonstrate the range of accuracy of the warnings. This spectrum emphasizes that most events do not neatly fit into a yes–no categorization, and can characterize events falling between an actual false alarm and a perfect warning (an overwarned event) or between an unwarned event and a perfect warning (an underwarned event).

The following examples show the elements of this conceptual model:

- *False alarm:* An example of a false alarm occurred over Labor Day weekend 1985 for the predicted track of Hurricane Elena. Nearly 1 million people were evacuated all along the coastline from Tampa, Florida, to New Orleans, Louisiana. The storm in the Gulf of Mexico was initially headed for Florida, but made a loop in the Gulf and eventually made landfall in Biloxi, Mississippi, as a category 3 hurricane. Four deaths were attributed to Elena, and more than 250 homes were destroyed with economic losses totaling $1.25 billion (Case 1986). Although not a perfect false alarm (as a storm was threatening a large stretch of coastline), examples of perfect false alarms in the meteorological literature are rare.
- *Unwarned event:* On the other side of the spectrum, an unwarned event occurred in the 28 August 1990 Plainfield, Illinois, tornado outbreak where many people received no warnings of multiple F3–F5 tornadoes. This unexpected tornado outbreak caused more than $500 million in damage, took 29 lives, and injured hundreds of people (NWS 1991; Seimon 1993).
- *Underwarned event:* An example of an event that was more severe than predicted was the April 1997 flood of the Red River of the North in Grand Forks, North Dakota, and East Grand Forks, Minnesota. The Red River flows through the broad, flat Red River basin from south to north into Canada where slower ice melting causes ice damming and flooding to the south. Winter snow and high melt rates contributed to the 1997 flood. The NWS issued a flood outlook for Grand Forks and East Grand Forks for the 1997 flood season. Two numbers were given in the flood outlook for expected flood stage: one based on a scenario of average temperature and no additional precipitation (47.5 ft); the other based on a scenario of average temperatures and additional precipitation (49 ft). Many people interpreted these numbers to be the range of possible maximum flood height or that 49 ft would be the absolute maximum flood height, rather than an outlook that gave two possible scenarios with substantial uncertainty. Because many people in these communities interepreted the flood outlook to be the *maximum* height, people openly blamed the NWS flood outlook for the devastation from the flood and viewed the outlook as an underwarned event when the actual flood topped out at 54 ft. However, Pielke (1999) determined that the blame for the $2 billion devastation should be shared. The NWS should have more clearly communicated the uncertainty in the flood outlooks, and government officials should have sought to understand the outlooks and accounted for the forecast uncertainty in their own decision making.
- *Overwarned event:* Examples of an overwarned event occurred in South Carolina for Hurricane Bertha (July 1996) and Hurricane Fran (September 1996). There were evacuation orders for South Carolina for both hurricanes, but the storms both made landfall farther north, in North Carolina. In both of these hurricane evacuations, the event was less severe than predicted in South Carolina (Dow and Cutter 1998). Total United States damage was $270 million for Bertha and $3.2 billion for Fran. Twelve lives were lost with Bertha and 26 with Fran (Pasch and Avila 1999).
- *Perfect warning:* An example of a warning in which the event was more or less as predicted was the 3 May 1999 Oklahoma tornado outbreak. This outbreak had 66 tornadoes, with 58 in the Norman, Oklahoma,

WFO warning area (Andra et al. 2002). This outbreak resulted in 45 deaths and 645 injured persons (Brown et al. 2002). The NWS issued 48 tornado warnings during that event (Andra et al. 2002). Although many accurate warnings were issued and the event was generally well forecast, this outbreak was not considered a perfect warning according to verification statistics. The FAR for this event was 0.29 (Andra et al. 2002). This example indicates the conceptual difficulty in classifying events as perfect warnings when FAR is nonzero.

Does this statistic, that 29% of all warnings for the 3 May 1999 Oklahoma tornado outbreak were false alarms, detract from the successful warnings issued? Or are such statistics only meaningful to administrators and statisticians? Were any of these unverified warnings actual false alarms? Does an FAR of 0.29 accurately reflect public perception? Perhaps the success of such events could be communicated more clearly with an improved performance measure that calculates close calls differently from false alarms. A close-call performance measure would give credit to forecasters for the many times close calls occur, rather than becoming part of the high FAR. To expect all forecasts to be either a "hit" or "miss," or simply "right" or "wrong," is not realistic; meteorological forecasts do contain uncertainty, and the science of forecasting is not perfect. Although fear of issuing a false alarm, or having a high FAR, may reduce the willingness of forecasters to issue warnings, Gruntfest and Carsell (2000) warn that a focus on reducing false alarms may have the unintended, and far more dangerous, outcome of a greater number of missed events. Events that fall on either side of a perfect warning should be viewed separately from hits or misses and should be viewed as valuable forecasts. Although these events do not occur exactly as predicted, forecasters should be given credit for alerting at-risk populations of a potential hazard.

When examining the national FAR for different weather phenomena, the FAR for tornado warnings is the highest. Yet, tornado warnings can only be issued when a tornado has been spotted or if there is a strong indication on radar; thus, there is always a threat present when a tornado warning is issued. Under a close-call warning metric, the tornado FAR would be reduced because of the strict requirements for issuing tornado warnings.

This conceptual model deals primarily with false alarms in relation to the intensity of the event. However, false alarms also occur because of spatial and temporal dimensions. For example, Hurricane Fran and Hurricane Bertha were close calls in both intensity and spatial dimensions. Because all three dimensions define the accuracy of a warning, all three dimensions should be considered in the calculation of FAR. An event that occurs within the forecasted intensity, time frame, and geographic location should be considered a perfect warning. An event that occurs with a lesser intensity, after the specified time frame, or outside the warning area should be considered an overwarned event. An event that occurs with a greater intensity, a longer time frame or covers a larger area should be considered an underwarned event. The degree of deviation from the specified intensity, time frame, and geographic location should be considered in the quantification of the warning performance measures. However, the current metrics do not account for these factors. For example, a tornado occurring 1 min after the warning expiration should be quantified differently than a tornado event occurring 3 h after the warning expiration. Similarly, a tornado event occurring 100 m outside the warning area should be quantified differently than an event occurring 100 km from the warning area. Rather than a hit or miss, we suggest that a different value be assigned based on the range in severity, time, and/or space. We suggest that a false alarm metric that accounts for all three factors (or three metrics that account for one factor each) would be a more accurate representation of forecaster abilities and the overall success of NWS warnings.

5. Conclusions

Conventional wisdom indicates that false alarms reduce confidence in future warnings and that overwarning is problematic. This perception has contributed to institutional goals of the NWS to reduce its FAR. While reducing the number of false alarms is a worthy goal, focusing on reducing this metric could lead to a greater chance of an unwarned event. We suggest two needed improvements in verification and in the FAR metric. First, the NWS metrics should be revised to represent the actual number of false alarms and the actual number of close calls according to the three warning criteria (intensity, time, and space). Second, the spotter network for event verification presents a major challenge, especially in areas where populations are sparse. Forecasters should not be hesitant to issue a warning because of fear it will be unverified. Further, FAR in rural areas should be comparable to FAR in urban areas; yet difficulties in verifying events in data-sparse regions of the country can contribute to higher FARs in rural areas. A metric that addresses the uneven distibution of spotters should be developed so as not to penalize forecasters unfairly in data-sparse regions of the country.

The conceptual model discussed in this paper demonstrates, through the use of actual events, that there is a gradual difference between a missed event, a perfect warning, and a false alarm. The conceptual model allows us to view warning accuracy through a new lens and suggests that warning verification should be quantified in a different, more complete, manner. Specifically, FAR would more closely represent the success of NWS warnings if a close call were categorized differently than a false alarm or a missed event.

This paper is not intended to be a solution to the high FAR. Rather the conceptual model presented in this paper is designed to serve as a framework to reevaluate the current methodology for determining FAR. Whether one examines the national FAR for tornado warnings (0.76) or the FAR of an event (e.g., 0.29 for the 3 May 1999 Oklahoma tornadoes), this measure does not represent the success of the warnings as perceived by the public. Rather than focusing on improving the national FAR for tornado warnings to 0.70 within the current evaluation system, perhaps this system should be examined and improved metrics should be developed to be aligned with NWS (2006) goals and to be "more useful to our customers and. . . more accurately represent NWS performance."

Consequently, the next step is to use this conceptual model to propose revised metrics that more accurately represent the success of warnings, thereby providing the NWS customers with a more useful measure of the performance of the NWS. We envision a spectrum of performance measures that ranges from very stringent (as the present system is now where the event must occur exactly in the space, time, and intensity it was predicted to be) to more lenient measures that reflect public perceptions of forecast success through overwarned events or close calls. By doing so, we feel the NWS will provide a more accurate depiction of the success of its warning program to the public.

Acknowledgments. Funding was provided by National Science Foundation Grant CMS-0301392. Research was performed under the CU Trauma Center, Departments of Geography and Psychology, at the University of Colorado at Colorado Springs. We thank Mary Mullusky, NWS National Flash Flood Program coordinator, for providing the false alarm values for flash floods. We thank Brent Macalony, meteorologist in the NWS Warning Verification Program, for providing the false alarm values for winter storm warnings, high wind warnings, severe thunderstorm warnings only, and severe thunderstorm and tornado warnings combined. We thank Susan Cutter, director of the Hazards and Vulnerability Research Institute, for her input on the conceptual model. Funding for DMS was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreement NA17RJ1227.

REFERENCES

Andra, D. L., Jr., E. M. Quoetone, and W. F. Bunting, 2002: Warning decision making: The relative roles of conceptual models, technology, strategy, and forecaster expertise on 3 May 1999. *Wea. Forecasting,* **17,** 559–566.

Atwood, L. E., and A. M. Major, 1998: Exploring the "cry wolf" hypothesis. *Int. J. Mass Emergency Disasters,* **16,** 279–302.

Breznitz, S., 1984: The false alarm effect. *Cry Wolf: The Psychology of False Alarms,* Lawrence Erlbaum, 9–16.

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.

Brown, D. P., R. Molleda, and R. J. Pasch, 2004: Verification of tropical cyclone genesis forecasts from the TCP's Tropical Analysis and Forecast Branch (TAFB). Preprints, *26th Conf. on Hurricanes and Tropical Meteorology,* Miami, FL, Amer. Meteor. Soc., CD-ROM, P1.2.

Brown, S., P. Archer, E. Kruger, and S. Mallonnee, 2002: Tornado-related deaths and injuries in Oklahoma due to the 3 May 1999 tornadoes. *Wea. Forecasting,* **17,** 343–353.

Carsell, K., 2001: Impacts of a false alarm: The January 29, 2000 Ventura, California experience. M.S. thesis. Dept. of Geography and Environmental Studies, University of Colorado at Colorado Springs, 86 pp. [Available from Dept. of Geography and Environmental Studies, University of Colorado at Colorado Springs, P.O. Box 7150, Colorado Springs, CO 80933-7150.]

Case, R. A., 1986: Atlantic hurricane season of 1985. *Mon. Wea. Rev.,* **114,** 1390–1405.

Dow, K., and S. L. Cutter, 1998: Cry wolf: Repeat response to hurricane evacuation orders. *Coastal Manage.,* **26,** 237–252.

Drabek, T. E., 1986: Warnings. *Human System Response to Disaster: An Inventory of Sociological Findings,* Springer-Verlag, 70–99 pp.

Edwards, R., cited 2006: Proposals for changes in severe local storm warnings, warning criteria and verification. [Available online at http://www.stormeyes.org/tornado/verf/index.html.]

Gruntfest, E., and K. Carsell, 2000: The warning process toward an understanding of false alarms. Report for the U.S. Bureau of Reclamation, Dept. of Geography and Environmental Studies, University of Colorado at Colorado Springs, 47 pp. [Available from Dept. of Geography and Environmental Studies, University of Colorado at Colorado Springs, P.O. Box 7150, Colorado Springs, CO 80933-7150.]

Janis, I. L., 1962: Psychological effects of warnings. *Man and Society in Disaster,* G. Baker and D. Chapman, Eds., Basic Books, 55–91.

Lindell, M. K., and R. W. Perry, 2004: *Communicating Environmental Risk in Multiethnic Communities*. Sage Publications, 262 pp.

Moller, A. M., 2001: Severe local storms forecasting. *Severe Convective Storms, Meteor. Monogr.,* No. 50, Amer. Meteor. Soc., 433–480.

NOAA, cited 2006: The National Weather Service customer satisfaction survey. [Available online at http://www.nws.noaa.gov/com/files/NWS_Public_survey050608.pdf.]

NWS, 1991: The Plainfield/Crest Hill tornado, northern Illinois, August 28, 1990. Natural Disaster Survey Rep., National Weather Service, 43 pp. plus 8 appendixes.

——, cited 2006: Working together to save lives: National Weather Service strategic plan for 2005–2010. [Available online at http://www.weather.gov/sp/NWS_strategic_plan_01-03-05.pdf.]

Pasch, R. J., and L. A. Avila, 1999: Atlantic hurricane season of 1996. *Mon. Wea. Rev.,* **127,** 581–610.

Pielke, R. A., Jr., 1999: Who decides? Forecasts and responsibilities in the 1997 Red River flood. *Appl. Behavior Sci. Rev.,* **7,** 83–101.

Rhatigan, R., L. R. Barnes, and E. Gruntfest, cited 2006: Looking back at Superstorm '93: Perspectives of emergency managers in southern Appalachia. [Available online at http://www.ccb.ucar.edu/superstorm/ss93Final.pdf.]

Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting,* **5,** 570–575.

Seimon, A., 1993: Anomalous cloud-to-ground lightning in an F5-tornado-producing supercell thunderstorm on 28 August 1990. *Bull. Amer. Meteor. Soc.,* **74,** 189–203.

Speheger, D. A., C. A. Doswell III, and G. J. Stumpf, 2002: The tornadoes of 3 May 1999: Event verification in central Oklahoma and related issues. *Wea. Forecasting,* **17,** 362–381.

Trapp, R. J., D. M. Wheatley, N. T. Atkins, R. W. Przybylinkski, and R. Wolf, 2006: Buyer beware: Some words of caution on the use of severe wind reports in postevent assessment and research. *Wea. Forecasting,* **21,** 408–415.

Turner, R. H., 1983: Waiting for disaster: Changing reactions to earthquake forecasts in southern CA. *Int. J. Mass Emergerceny Disasters,* **1,** 307–334.

Weaver, J. F., E. Gruntfest, and G. M. Levy, 2000: Two floods in Fort Collins, Colorado: Learning from a natural disaster. *Bull. Amer. Meteor. Soc.,* **81,** 2359–2366.

Weiss, S. J., J. A. Hart, and P. R. Janish, 2002: An examination of severe thunderstorm wind report climatology: 1970–1999. Preprints, *21st Conf. on Severe Local Storms,* San Antonio, TX, Amer. Meteor. Soc., 446–449.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences.* 2d ed. Academic Press, 627 pp.