P1.2

RATING TORNADO DAMAGE: AN EXERCISE IN SUBJECTIVITY

Roger Edwards
Storm Prediction Center
Norman, OK 73069

1. INTRODUCTION AND BACKGROUND

Since the advent of the Fujita tornado damage scale (Fujita 1971, hereafter, F-scale), the judgmental and subjective nature of damage assessment has been well documented, perhaps most prominently by Doswell and Burgess (1988, hereafter DB88), who strongly emphasized the nature of F-scale ratings as driven by *damage and not intensity* of tornadoes.

Tornado ratings from the 1950s into the early 1970s were based almost entirely upon remote, post-event review of media articles and photographs (Schaefer and Edwards 1998). Starting in the middle 1970s, local National Weather Service (NWS) offices have assigned F-scale ratings to most tornadoes based on a blend of first-hand surveys by their meteorologists and anecdotal evidence from news media, emergency management, storm observers, law enforcement personnel and the general public. Under this system variations in local rating practices from office to office become an intrinsic part of the modern *Storm Data* tornado database, which is used by the NWS and by most researchers studying aspects of tornado climatology. The dangers of such nonuniformity in rating procedures to scientific utility of tornado data were also well elucidated in the DB88 discussion, and remain such a concern 14 years later.

Further, the NWS tornado “climatology” has reflected more intensive reporting of weak tornadoes related to NWS warning verification practices, a trend observed through the 1990s (e.g., Weiss and Vescio 1998). The only database incorporating a relatively consistent standard of rating, owing to its nature as a single-person effort, is that of Grazulis (1993 and 1997). Still, the ratings therein are based on its author’s interpretation of others’ accounts of each event, whether in the form of news articles, *Storm Data* ratings, NWS survey reports, Fujita surveys or other secondhand sources.

These characteristics of tornado data imply a strong impact of subjective judgment on the rating process, but little has been done previously to document this effect on the data. [A concurrently published preliminary work by Guyer and Shea (2003) has utilized a different rating exercise as the one published here, with similar results.]

2. F-SCALE DAMAGE RATING EXERCISE

2.1. Background and Method

As part of a series of workshops at the NWS/Texas Tech University Severe Storms Conference (Lubbock, TX, 9 February 1998), the author conducted an exercise where individuals in the audience were asked to assign F-scale damage ratings based on photos displayed on a scene-by-scene basis, and to justify assigned ratings with brief explanations. They viewed a sequence of photographic slides representing six scenes of damage along the path of the Spencer, SD, tornado (USDOC, 1998). Each scene was at a separate site rated in the initial damage survey by the author and his field survey partners from F0 to F4. Most scenes contained still imagery from both ground and aerial surveys, as in the Fig. 1 example. The scenes are generally described as follows:

1. Unanchored wood frame house moved whole off its cinder block foundation, into a mobile home anchored by steel cables. Mobile home missing portion of roof but otherwise intact.

2. One of several adjacent unanchored and poorly anchored frame homes with slab foundations and basements, flattened with no walls standing.

3. Removed metal shed (no floor), had been anchored into bare ground by cables and bolts.

4. Six adjacent, circular, steel grain bins bolted to slab foundations, containing varying levels of grain. Nearly empty bins were mostly to completely gone; remaining bins were about full.

5 (See Fig. 1). Frame fire station completely removed from slab foundation to which it had been anchored by .25 inch (.64 cm) diameter bolts. Adjacent steel water tower was toppled amidst large debris (automobile and nearly intact wall and large roof segments from fire station). Minimal damage to trucks parked in fire station, some fire gloves on floor unmoved.

6. Two story, split-level apartment building with brick facade, top floor partially collapsed into part of lower level and partly scattered into a cyclonically curved fan northeastward.

Multiple photos were provided for each scene along with verbal descriptions, to aid the rating process. The audience members rated the damage on multiple-choice worksheets, with space in each damage scene's section for justification and comment.

2.2. Audience

This exercise was taken by people with a wide range of damage survey experience, in an audience of meteorologists, engineers, storm spotters and chasers, and lay people – each representing groups whose input has been considered in actual tornado ratings. Many were NWS forecasters with greatly varying experience in conducting damage surveys. Like NWS forecasters as a whole, few had personally seen F5 damage, so they tended not to have a mental benchmark for the full spectrum of tornado damage. [Spencer was officially rated F4.]

Included in the rating exercise was a large contingent of storm chasers and several professional wind engineers. Except for the wind engineers, only a few of the audience members had any formal training in damage analysis techniques outside of rudimentary illustrative memoranda that some of the NWS personnel had read on-the-job (e.g., Bunting and Smith 1993). No NWS personnel in attendance had formal structural or architectural engineering education, to our knowledge. One audience member was a professional engineer with a Master' s degree in meteorology and who had extensive experience in damage analysis. For a few others, this was the first time they had ever attempted to rate damage. So, as is collectively true with those who are commonly tasked to survey wind damage events, the spectrum of expertise was very broad.

Figure 1. “Scene 5" imagery from the Spencer, SD fire station, used in the Lubbock F-scale exercise. Credit Brian E. Smith (aerial, A) and Roger Edwards (ground, B-D).

	Scene 1	Scene 2	Scene 3	Scene 4	Scene 5	Scene 6
Contents	**House and mobile home**	**Frame homes**	**Anchored metal shed**	**Grain bins**	**Fire station, water tower**	**Apt. bldg.**
Sample Size	64	64	62	62	64	62
Max F Scale	3	4.5	4	4	5	4
Min F Scale	0	2	1	1	1	2
# F Scales	4	3.5	4	4	5	3
Mean F	1.44	3.07	2.35	2.16	3.03	3.04
Std. Dev. of F	.68	.70	.96	.79	.78	.57

Table 1. Simple statistical analysis summary of numerical F-scale assignments for the damage rating exercise.

2.3. Exercise Results and Responses

Table 1 summarizes some statistics based on the multiple-choice responses; and the next subsection contains sample responses from scene 5 (Fig. 1). From 62 to 64 respondents rated each scene, with statistical results shown in Table 1. The greatest range in ratings occurred with Scene 5, from F1 to F5. However, all the other scenes also displayed a large range of ratings, typically across three to four values of F-scale. The standard deviation ranged from 0.57 (scene 6) to 0.96 (scene 3), indicating that participants often had trouble assigning F-scale values in a consistent manner.

To illustrate the subjective variability in justification of ratings, an answer for each part of the spectrum in assigned damage levels for scene 5 is quoted here:

F5: “Displaced auto near water tower suggests sustained transport of very large projectile.”
F4: “Water tower most likely well anchored; however, fire truck remains. Still, possible F5.”
F3/F4: “Borderline. Would be F5 but attachment was very suspect. Fire station may have just popped loose intact, & even F4 not consistent with unmoved trucks.”
F3: “Structures not as sound as they could be (no washers etc.). Water tower probably got hit by debris, destroying supports.
F2/F3: “But car was moved? Trucks were still there, gloves still there, no washers or nuts.”
F2: “Poor anchoring. Could have been lifted off easily.”

There are some important caveats to this analysis. A few people did not rate certain scenes, or alternatively, chose to assign "F-unknown." This accounts for the small variation in sample size among scenes. Where respondents circled two adjacent ratings, uncertainty and/or the intent to assign borderline ratings were assumed. Though actual damage rating is in whole F-scale integers only, we allowed for some uncertainty by permitting entries where two adjacent ratings were circled. For statistical purposes only, such choices were assigned the half-integer level between circled ratings (e.g., 3.5 where F3 and F4 were both circled), realizing that only whole-integer ratings are assigned in actual events. Also, due to time constraints, no attempt was made to categorically sort responses by any measure of expertise or experience.

Another version of this exercise was presented at the Eumetsat Workshop on Severe Storms (Pruhonice, Czech Republic, August 2002), with similar results, from an audience of virtually no experience in surveying storm damage (Doswell 2002).

3. DISCUSSION AND RECOMMENDATIONS

The results of the rating exercise illustrate the large range of ratings assigned to damage scenes and the large subjectivity inherent in the process. As expected, there was substantial disagreement among participants on the F-scale rating of damage scenes. Although scene 5 generated the biggest range of ratings, the relatively large standard deviation and of scene 3 indicates people had the most overall difficulty rating the damaged shed. Those scenes represented "nonstandard" structures or loose anchoring for which explicit rating guidance is scant in the literature (e.g., Grazulis 1993, Bunting and Smith 1993). Scene 6 may have been the most straightforward, with a distinct pattern of damage to one common structure type (two story apartment building), and accordingly, tighter clustering of responses. Though some people rated scene 4 as F5, unlike with scene 3, several also rated it as low as F1, again a possible function of its complexity. Hence, scene 3 (non-anchored house swept away) had the highest average damage rating.

It is apparent from these results that rating damage using the F-scale is highly subjective and variable, as suggested by DB88, and is often performed by people with limited expertise. **Hence, a single, nationally deployable, interdisciplinary team of expert assessors should be developed from which individuals are selected for aerial and/or ground surveys of individual events.** This team should consist of specialists from the meteorological and engineering communities who have extensive experience in multiple tornado damage surveys. Economic constraints may prevent all tornadoes from being surveyed; however, intensive training of NWS field meteorologists in damage survey techniques and principles is proposed for events not covered by the national team. Such training may be conducted by national survey team members at a centralized location as part of a course of at least one week in duration, dedicated specifically to this topic. Course materials and exercises should cover a great variety of structural types and situations beyond the "well-built" houses upon which the F-scale is based.

Finally, because of the subjectivity and variability inherent to the rating process, each *Storm Data* tornado listing should contain a discrete categorical entry representing the *specific source* for the rating (site survey, media report, spotter or chaser account, law enforcement report, public report, etc.) in order that tornado data may be more representatively analyzed and sorted in scientific studies, hazard and risk assessments and climatological compilations.

4. ACKNOWLEDGMENTS

Thanks to Dan McCarthy for his strong encouragement in presenting this angle on the F-scale issue, to Steve Weiss for his insightful review of the article, and for their insight and ideas during the exercise and/or in subsequent discussions: Chuck Doswell, Tom Grazulis, Jared Guyer, Tim Marshall, Kishor Mehta, Long Phan, Rich Thompson and Mike Vescio. Great gratitude goes to Greg Harmon for inviting the author onto the Spencer SD damage survey and for his aid in delivering the original exercise.

5. REFERENCES

Bunting, W.F. and B.E. Smith, 1993: A Guide For Conducting Damage Surveys. NOAA Tech. Memo. NWS SR-**146**. Available from National Technical Information Service, U.S. Department of Commerce, Springfield, VA 22161.

Doswell, C.A., III, 2002: Personal communication.

Doswell, C.A., III, and D. W. Burgess, 1988: On some issues of United States tornado climatology. *Mon. Wea. Rev.*, **116**, 495-501.

Fujita, T.T., 1971: Proposed characterization of tornadoes and hurricanes by area and intensity. *Satellite and Meteorology Research Paper*, **91**, The University of Chicago, Chicago, IL 42 pp.

Grazulis, T.P., 1993: *Significant Tornadoes: 1680-1991.* Environmental Films, St. Johnsbury VT, 1326 pp.

_____, 1997: *Significant Tornadoes: Update, 1992-1995.* Environmental Films, St. Johnsbury VT, 117 pp.

Guyer, J.L. and T.J. Shea. 2003: An Assessment of the Variability in Operational Assignment of F-Scale Damage. Preprints, *Symposium on F-Scale and Severe-Weather Damage Assessment*, Long Beach CA, this volume.

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.

U.S. Department of Commerce, 1998: *Storm Data*, **40** (5). [Available from National Climatic Data Center, Asheville, NC 28801.]

Weiss, S.J., and M.D. Vescio, 1998: Severe local storm climatology: 1955-1996: Analysis of reporting trends and implications for NWS operations. Preprints, *19th Conf. On Severe Local Storms*, Minneapolis MN, Amer. Meteor. Soc., 536-539.