On the Implementation of the Enhanced Fujita Scale in the USA

Charles A. Doswell III [1,*], Harold E. Brooks [2], and Nikolai Dotzek [3,4]

[1] Cooperative Institute for Mesoscale Meteorological Studies, National Weather Center, 120 David L. Boren Blvd., Norman, OK 73072, U.S.A.

[2] NOAA/National Severe Storms Laboratory, National Weather Center, 120 David L. Boren Blvd., Norman, OK 73072, U.S.A.

[3] Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, 82234 Wessling, Germany

[4] European Severe Storms Laboratory (ESSL), Münchner Str. 20, 82234 Wessling, Germany

[*] *Corresponding Author*: Dr. Charles A. Doswell III, Cooperative Institute for Mesoscale Meteorological Studies, National Weather Center, 120 David L. Boren Blvd., Norman, OK 73072, USA. Tel: +1 (405) 325-6093, Fax: +1 (405) 325-3098, eMail: cdoswell@gcn.ou.edu, http://www.cimms.ou.edu/~doswell/

Abstract

The history of tornado intensity rating in the United States of America (USA), pioneered by T. Fujita, is reviewed, showing that non-meteorological changes in the climatology of the tornado intensity ratings are likely, raising questions about the temporal (and spatial) consistency of the ratings. Although the Fujita scale (F-scale) originally was formulated as a peak wind speed scale for tornadoes, it necessarily has been implemented using damage to estimate the wind speed. Complexities of the damage-windspeed relationship are discussed.

Recently, the Fujita scale has been replaced in the USA as the official system for rating tornado intensity by the so-called Enhanced Fujita scale (EF-scale). Several features of the new rating system are reviewed and discussed in the context of a proposed set of *desirable* features of a tornado intensity rating system.

It is concluded that adoption of the EF-scale in the USA may have been premature, especially if it is to serve as a model for how to rate tornado intensity outside of the USA. This is in large part because its degree of damage measures used for estimating wind speeds are based on USA-specific construction practices. It is also concluded that the USA's tornado intensity rating system has been compromised by secular changes in how the F-scale has been applied, most recently by the adoption of the EF-scale. Several recommendations are offered as possible ways to help develop an improved rating system that will be applicable worldwide.

Keywords: Tornado, F-scale, EF-scale, Intensity distribution

1 Introduction

The National Weather Service (NWS) of the United States of America (USA) has recently implemented the so-called Enhanced Fujita (EF) scale (e.g., Potter 2007) for damage-based rating of tornadoes. In contrast, the Fujita (F) scale (Fujita 1971; 1981), which it has replaced, was originally created as a wind speed scale. The advantage of a scale based on wind speeds is that it doesn't depend on construction practices in any particular part of the world; it is completely transferable anywhere. However, as Doswell and Burgess (1988) point out, a wind speed scale is just not useful in practice, because wind measurements from tornadoes are relatively rare. Damage continues to be the best and most useful indicator of tornado intensity on a routine basis, despite the complex relationship between damage and wind speed.

All of the tornadoes in the USA affect only a small total area annually (of order 500-1000 km^2), so that the probability of having measurements from in situ anemometers is quite small, and such sensors are destroyed in most tornadoes anyway. Historically, only a handful of anemometer measurements of tornadic winds have ever been obtained (e.g., see Figs. 75 and 77 of Fujita et al. 1970) and the *strongest* winds in a significant tornado could never be measured this way.

Remote sensing of tornado winds by using the Doppler principle is possible. An operational network of WSR-88D Doppler radars covers most of the USA, but physical limitations (e.g., beam spreading and the radar horizon) and the operating characteristics of the radars (e.g., the spatial and temporal sampling resolution) make the possibility of obtaining useful tornado wind speed measurements from them for the purpose of rating tornado intensity quite unlikely. Since the late 1980s, the technology for occasional probing of tornadoes by *mobile* Doppler radars and lidars has been developed to overcome some of the operational radar limitations (Bluestein and Unruh 1989; Wurman et al. 1997). The relationship between the velocities sensed by mobile radars (typically at or above heights of around 50-100 m) and the actual winds near the surface (i.e., where the damage occurs, at heights of 10 m or less above ground level) remains to be determined. Some recent studies (e.g., Wurman and Alexander 2005) have begun to explore this topic. Unfortunately, even if a reliable and accurate method for extrapolating mobile Doppler radar measurements downward to within 10 m can be developed, it will be some time before we have wind speed estimates from mobile Doppler radars for even a tiny fraction of the lifetimes of another tiny fraction of all tornadoes. In the USA, more than 1000 tornadoes are reported annually, but at present, only around 20 tornadoes are sampled by mobile Doppler radars every year. Therefore, the damage-wind speed relationship is going to be used for some time to come.

Herein we review some of the changes in the practice of rating tornadoes in the USA that have occurred over the years and their impact on the ratings. Some of these changes were intentional, while others were not. The implications for continued applicability of comparisons of ratings across time and space past are troubling (e.g., Brooks and Doswell 2001; Dotzek et al. 2003, 2005; Feuerstein et al. 2005).

The paper is organized as follows: Section 2 presents a history of the tornado rating system in the USA. In section 3, challenges for use of the F-scale are described and desirable criteria for any rating system are described. Section 4 provides the conclusions, along with our recommendations.

2 History

2.1 The creation and implementation of the F-scale

Professor T. Theodore Fujita developed the F-scale in the late 1960s. Prior to this, there had been no formal attempt to differentiate tornado occurrences by intensity, although it certainly was known that tornadoes are not uniformly intense. The F-scale was implemented nationally with the support of Mr. Allen D. Pearson, then head of the National Severe Storms Forecast Center in Kansas City, MO (predecessor to the current Storm Prediction Center) – part of the NWS. The F-scale became the official basis for rating tornadoes in the early 1970s.

Shortly after its official adoption by the NWS, the Nuclear Regulatory Commission sponsored an effort to develop F-scale ratings for historical tornadoes from 1950 through 1976 as part of a study to safeguard the nation's nuclear power generating stations,. This was done by paying researchers (mainly college students) to review newspaper accounts and come up with an estimate of tornado intensity for every tornado in the record. The researchers were given what materials then existed to document how to make F-scale ratings. Results of this project were summarized in a paper by Kelly et al. (1978), providing the first climatological information about tornado intensity distributions in space and time. Since F-scale ratings were to be determined thereafter for all tornado reports in the official record – *Storm Data* (available from the National Climatic Data Center) – this provided for a continuing expansion of the database supporting the climatology of tornado intensities based on their F-scale ratings.

2.2. Post-Event Surveys of Tornadoes Since 1950

Prior to the development and operational implementation of the F-scale, the responsibility for providing input for *Storm Data* had been assigned to the NWS state climatologists within each state. In the early 1970s, however, those Federal state climatologist positions were abolished, so the task of providing input to *Storm Data* became an additional duty for the staff members at the local NWS offices in whose area of responsibility tornadoes (and other severe weather) was reported. For many years thereafter, there was essentially no training program for the NWS staff on how to estimate F-scale ratings.

Fujita did occasional detailed post-event analyses for selected tornado cases from the 1950s until his retirement in 1992; he and his graduate students developed a multifaceted storm survey methodology, using both ground-based and aerial survey methods for assessing the distribution of tornado intensities along a tornado's path (e.g., Forbes and Wakimoto 1983). This effort was limited to no more than a handful of events every year, typically major outbreaks of tornadoes (and other types of storms). Fujita's team gained experience in doing such surveys, although some uncertainty about their ratings was inevitable. The National Severe Storms Laboratory (NSSL) also did occasional scientific damage surveys for events within or close to Oklahoma, as part of their tornado-related research. The NWS is not obligated officially to use the findings of surveys done by external agencies, but they certainly have used this information to produce F-scale estimates whenever such surveys have been done and the results made available. At the same time, the NWS was doing fewer of its own detailed scientific surveys of major tornado events, presumably because it was expected that Fujita's team (or someone else) would do this for them – such surveys are not free. The main concern for the increasingly infrequent formal NWS post-storm "surveys" has evolved toward assessing the quality of the service provided by the NWS during the event, rather than focusing on the scientific and/or engineering issues. Individual NWS offices are responsible for establishing the intensity rating for every tornado, whether or not an official NWS post-event service assessment is conducted.

In May of 1970, a powerful tornado struck Lubbock, Texas, passing near the campus of Texas Tech. University (TTU). Largely as a result of that devastating event, a wind engineering research program was created at TTU, with a primary emphasis on structural engineering issues. The TTU researchers began doing surveys of their own on selected nearby tornado events, mostly seeking to refine the wind speed-damage relationship and to answer questions about how to design structures to resist tornadic winds. By 1977, this program provided its first major contribution to the topic (Minor et al. 1977), with many more to follow. Eventually, the TTU wind engineers began to do surveys nationally (for a few events per year), although still with an emphasis on events within and near the state of Texas.

Following Fujita's retirement in 1992, the number of scientifically-oriented post-event surveys dropped precipitously (Speheger et al. 2002). Many important tornado events were not being given a careful review by science teams, although the TTU wind engineers and NSSL scientists continued to do occasional surveys, including the events of 3 May 1999 in Oklahoma and Kansas.

In April of 2002, a tornado that struck La Plata, Maryland was initially rated by the local NWS office team as an F5 tornado. Subsequent review suggested that this likely was an overrating of this tornado, and its official rating eventually was downgraded to F4. In response, after some deliberations, the NWS created the so-called Quick Response Team (QRT), a group of volunteers with experience at damage assessments for violent tornado cases. The establishment of the QRT was intended to provide "expert" assistance to any local NWS survey team in cases involving one or more tornadoes that *might* be rated F4 or F5. In practice, the national QRT has been called upon only rarely after its first early deployments following tornadoes in May 2003. The impacts of these changes in the application of the F-scale concept to the ratings will be detailed further in section 2.4.

2.3 Development of the EF-Scale

Roughly a decade ago, structural engineers led by the TTU group initiated a series of discussions that began with a "Fujita Scale Forum", whose participants were invited based on their established professional involvement with the tornado intensity ratings, with the goal to "enhance" the F-scale. The engineers long had felt that the lack of calibration for the F-scale's wind speed-damage relationship, notably at the high end, was associated with overestimates of the wind speeds for F3-F5 damage. The structural engineers have believed steadfastly that virtually all of the observed damage to frame homes could be accounted for by wind speeds that would at most be somewhere near the transition from F3 to F4 (i.e., about 90 m s^{-1}).

However, mobile Doppler radar-measured velocities at the high end of the F5 class (~142 m s^{-1}) have actually been observed within about 100 m of the ground on 3 May 1999 (Burgess et al. 2002). In fact, velocities approaching that high end were observed by mobile Doppler radars as far back as 1991 (Bluestein et al. 1993). Furthermore, there is theoretical evidence to support the transient occurrence of extreme wind speeds near the surface in the range of Fujita's original F5 category or perhaps even beyond – see Fiedler and Rotunno (1986), Fiedler (1998), and Lewellen and Lewellen (2007). Still, it continues to be particularly difficult to determine just what wind speeds are associated with the "high-end" *damage* produced by tornadoes. We have relatively little direct observational information about the very complex interaction between tornadic winds and the structures they damage. For reasons already discussed, we must continue to use damage in lieu of the desired wind speed measurements.

Most structures damaged by tornadoes are not *engineered* to resist high wind speeds. For such objects, it is especially challenging to assign wind speeds to the damage, as we will discuss shortly. On rare occasions, however, engineered structures are found within the tornado damage path and these can, to some extent, serve to "calibrate" the damage-wind speed relationship. If a structure designed to resist wind speeds of V fails, then the wind speeds must have exceeded V. Unfortunately, such unambiguous indicators are rare, and like all damage indicators when the degree of damage is “completely destroyed”, provide only a lower bound on the wind speeds.

A complicating factor in the use of any damage indicator is that each example of any particular indicator likely will not fail at exactly the same wind speed. Not all frame homes are identical and specific failure points are never identical, either. Further, there is some suggestion that the four-dimensional (three spatial dimensions and time) structure of the wind field in tornadoes might be quite complex, with the temporal character of the high winds an important issue. Thus, for example, after the Jarrell, Texas tornado of 27 May 1997, some engineers (e.g., Phan and Simiu 1999) disputed its F5 rating, proposing that its relatively slow movement meant that the *duration* of the tornadic wind speeds contributed significantly to the complete destruction of homes in a Jarrell subdivision. According to their analysis, much lower wind speeds than those associated with minimal F5 rating (117 m s^{-1}) could have caused *all* the observed damage. Although we can offer no evidence to dispute their findings, the wind speed necessary to produce complete destruction of a home is, again, only a lower bound to the actual wind speed. As yet, no one has conducted any experiments to determine the relationship between duration of the wind and the damage produced, especially at the upper end of the F-scale.

Eventually, the effort to modify the wind speeds associated with the Fujita scale resulted in the adoption of the EF-scale by the NWS, effective 1 February 2007 (Potter 2007). An important part of the EF scale is the notion of *damage indicators* (cf. Fujita 1992). Participants in the process of "enhancing" the F-scale were asked to propose what they considered were useful indicators of the wind speeds in tornadoes, primarily to create new indicators in addition to the "well-constructed" frame home that formed the primary indicator for the F-scale as originally adopted. The synthesis of that input was a list of 28 damage indicators to allow the members of a local NWS survey team to estimate the wind speeds associated with an observed *degree of damage* for each indicator. That is, the observed damage can fall somewhere between no damage and complete destruction of the indicator. Files containing documentation of the indicators and degrees of damage recently have been carried on a hand-held computer by local NWS survey teams, many of whom now have had some limited training in the rating task. The scientists and engineers who developed the EF-scale assigned a windspeed estimate to each degree of damage for every damage indicator. These windspeed estimates were not done entirely objectively but rather were based primarily on the opinions and experience of the participants. Of particular note is that the wind speeds associated with the high-end indicators, including "well-constructed" USA frame homes were revised substantially – downward.

Further, the minimum criteria for producing EF5 damage effectively have been increased: complete destruction of a *typical* frame home in the USA would no longer be considered adequate for an EF5 rating and perhaps not even for EF4. The home would have to be constructed to a higher standard than in the era when the F-scale was the official rating scale to qualify for an EF5 rating. This change in practice is without regard to the associated wind speed estimates assigned to the EF-scale. The change occurred despite an informal agreement among the original Forum participants that the EF-scale ratings should be identical to F-scale ratings from the past, in order to maintain historical continuity. Actually, the tendency to impose higher standards on F4+ damage began in the late 1970s, when structural engineers began to emphasize the importance of considering the structural integrity of frame homes in the path of specific tornadoes. Thus, we show next that there has been a continuing evolution in tornado intensity ratings, especially for the F4+ events, that began well before the adoption of the EF-scale.

2.4 Documentation of Rating System Evolutionary Changes

Although the overall number of reported tornadoes has increased dramatically since the early 1950s, the number of tornadoes rated F1 or greater (F1+) has been relatively constant, albeit with considerable interannual variability, since 1953 (see Fig. 1). Most of the increase in the annual tornado numbers is associated with an increase in tornadoes rated F0. Based on linear regression, a slightly downward slope (corresponding to a decrease of 1.5 reports per year) is present from 1953 to 2006, but is not statistically significant (a p-value of 0.15). In the database, there are 27 885 tornadoes rated F1+ in the period 1950-2006. In order to see any secular trends in damage reporting (cf. Brooks and Dotzek 2008), it is illustrative to consider the number of tornadoes at higher thresholds normalized with respect to 1000 F1+ tornadoes. That number of F1+ tornadoes corresponds to about 2-3 years of reports.

Early in the record, 500 or more of any run of 1000 F1+ tornadoes were rated F2+ (Fig. 1). However, since the early 1980s, that number has fallen to about 300. Although causes for this cannot be known conclusively, it is pertinent to observe that the F-scale was first implemented in real time by some NWS offices on a trial basis in 1972, and by the late

1970s it had been adopted throughout the NWS (McCarthy et al. 2006). Verbout et al. (2006) have called attention to the possibility that the retrospective ratings for tornadoes before the adoption of the F-scale produced a bias in the early record. A plausible description summarizing the behavior seen in Fig. 1 is as follows: a period of relatively consistent ratings into the early 1970s, followed by a period of inconsistent practices in the time near the adoption of the F-scale that persisted into the 1990s, followed by a decade of relatively consistent standards through the end of the 20th century. In particular, runs beginning in 1991 through 2000 were remarkably consistent, ranging from 276 to 339 F2+ tornadoes per 1000 F1+ tornadoes. Note that the run beginning at the end of 2000 includes tornadoes through early 2003. We believe it may not be coincidental that early in the 1990s, the NWS produced a formal guide for conducting damage surveys (Bunting and Smith 1993[†]) and was included as part of the Doppler radar training course that all NWS forecasters were taking at the time.

No run of 1000 F1+ tornadoes beginning after the middle of April 2002 has had more than 264 F2+ tornadoes. The lowest number to date was 201 for the period of July 2003 through June 2005. The reduction by one-third in the number of F2+ tornadoes is comparable to that seen during the period following adoption of the F-scale. It began without a comparable official change in rating practice and followed a decade of relatively consistent ratings.

The unusual nature of the ratings from 2003 to the present is illustrated dramatically when considering violent tornadoes (F4+) in the Southern Region of the NWS.[‡] We have extended the record of F4+ tornadoes in the Southern Regions back to 1904 by using the record of Grazulis (1993) for the period 1904-1949, and also have included the as-yet preliminary F-scale ratings though September of 2007 for this analysis. During the period, 440 F4+ tornadoes were reported in the Southern Region, or approximately 4.2 annually. If we consider the gaps between consecutive violent tornadoes, most of the gaps are less than one year, indicating multiple violent tornadoes in a given year. Another clustering of gap lengths is bounded on the high end by approximately one year, and a few longer gaps up to about two years in length. By far, the longest gap is the 1393-day hiatus between the F4 tornadoes on 8 May 2003 and 1 March 2007. Only nine years in the 104-year period of record 1904-2007 (inclusive) did not have at least one F4+ tornado in the NWS Southern Region, but four of them are in the recent period 2002-2006. Assuming that consecutive years are statistically independent (the calculated autocorrelation of the annual number of

[†] The Bunting and Smith text was originally written in 1990 and available to NWS forecasters but was not published as a technical memorandum until 1993. Brian Smith had been part of Fujita's graduate student survey team, participating in several surveys with Fujita before joining the NWS.

violent tornadoes is -0.07, so this is a reasonable assumption), the probability of three consecutive years without a violent tornado, based on the 1904-2007 data, is approximately 1 in 10 000. Although meteorological causes cannot be ruled out definitively, it seems likely that non-meteorological causes have to be considered likely for this low probability event, given that overall tornado numbers have not changed dramatically.

3 Challenging issues for tornado intensity rating systems

3.1. Recognized issues with the F-scale

After the introduction and adoption of the F-scale in the 1970s, some troubling aspects of the system became apparent. Perhaps the most glaring problem was that the F-scale is based on only one primary damage indicator: a "well-constructed" wood frame home, which in the USA is the typical structure in the path of a tornado when it crosses a populated area. Apart from the ambiguity of just how the term "well-constructed" is defined, the fact that many tornadoes do *not* strike populated areas raises serious challenges for estimating the intensity of such events. If a tornado fails to hit a recognized damage indicator, a rating nevertheless is required. In practice, this means that many tornadoes are given a "default' rating – often either F0 or F1, unless there is some compelling reason in the opinion of the person doing the ratings to give such an event a rating other than the default value. In the absence of any information, it seems more appropriate to have the option to assign an intensity rating of "unknown", but official NWS policy mandates that *every* tornado be assigned an F-scale rating, irrespective of what it hits.

Moreover, the existing database for tornadoes currently does not provide any way to document the *source* for the rating. Without knowing the source(s) for the information used to make the rating (which could include a diverse set of possibilities), the level of uncertainty in the rating cannot be determined. If the rating is based on a detailed ground and aerial survey by a team of scientists and engineers, the rating has a much lower uncertainty than if the rating is estimated by an untrained person interpreting local newspaper accounts well after the event.

In the few cases where an *engineered* structure is in the path, it is possible to assign a wind speed (albeit, a lower bound) to the failure of this structure and so provide objective information for assigning an F-scale rating. Also, if something extraordinary is observed during a survey – such as pavement scoured from the roads or a heavy object (e.g., a railroad car or a large farm implement) documented as having been airborne – a high rating could be

‡ The NWS Southern Region includes the states of New Mexico, Oklahoma, Texas, Arkansas, Louisiana, Tennessee, Mississippi, Alabama, Georgia, and Florida.

assigned. Unfortunately, there is as yet no consensus about how to interpret these extreme occurrences in terms of the wind speed necessary to produce them.

Most of the known challenges of applying the F-scale in actual practice are associated with the wind speeds assigned to the degree of damage. The wind speeds originally were defined by Fujita for each F-scale category without overlap. A windspeed of 157 mph[§] (70.2 m s^{-1}) is at the top of the F2 category, whereas a wind speed of 158 mph (70.6 m s^{-1}) is at the bottom of the F3 category. This gives the illusion of great precision (1 mph or roughly 0.5 m s^{-1}) in the associated wind speeds that is not justified by our knowledge of the actual wind speeds in a tornado. As already noted, any particular example of a damage indicator will not fail in exactly the same way, at exactly the same wind speed as every other example of that indicator. Flying debris impacts can change the response of a structure to a given wind speed; the orientation of the structure with respect to the wind can mean different degrees of damage; the duration of the wind, the temporal acceleration of the wind, the presence (or absence) of nearby structures, and many other factors can all influence the damage. The relationship between damage and wind speed for any particular event involves the nonlinear interaction of a complex wind field in space and time with a unique set of structures. We observe that meteorologists tend to interpret variations in the damage to variations in the wind speed, whereas structural engineers tend to interpret the same variations in damage as variations in the structural integrity of the objects in the path. In reality, it is likely that *both* are always involved to some degree, but it can be difficult to separate the contributions from wind and structural variability.

The decades-old concern of structural engineers has been to determine the wind speeds actually needed to produce a given degree of damage to a "well-constructed" frame home. It is difficult to imagine putting a whole house into a wind tunnel and doing comprehensive tests to calibrate the degree of damage as a function of wind speed, for homes incorporating a variety of construction practices. Besides, the cost of building and then destroying dozens of homes appears prohibitive. Even if it were feasible to do such a set of experiments, it is impossible to simulate in a wind tunnel the actual evolution of the wind as a tornado encounters a real home. It is likely that every particular tornado-structure interaction is different in detail from any other. Further, including the effects of flying debris, as well as rapid changes in the speed and direction of the wind, would be difficult to simulate in a wind tunnel.

That variations in structural integrity make the notion of a "well-constructed" frame home difficult to apply in practice is widely known now. When the F-scale first was adopted, this effect was not widely recognized among meteorologists. Increasing awareness of structural issues evidently has influenced the ratings over time, as noted above. When homes in the USA are actually built, there is wide variation in how well the key attachment points in the load path are secured. In places where building codes have been imposed (mainly cities), home builders sometimes depart from the codes to increase profitability – some of those code departures have been approved by local government as "variances," but many are not. Enforcement of building codes is not always effective, and much rural construction is done in the absence of any building codes.

On a survey after a tornado has struck, those doing F-scale ratings need to be aware of what to look for in terms of structural integrity, but they often have little or no experience with violent tornado events and have been given only limited training in structural issues, if any. Because NWS local office survey teams generally are meteorologists, not structural engineers, structural engineering is not typically part of their education. A formal guide for doing F-scale ratings was published by the NWS (Doswell 2003), coincidentally during the time when the implementation of the EF-scale was being considered. Fujita (1992) was aware of the problem with construction practice and developed his own proposed solution to this problem by adding a separate damage scale, the "f-scale", to the original, wind speed scale-based, F-scale (Fig. 3). In his proposed methodology, the degree of damage to a damage indicator was modified by knowledge of the structural integrity to arrive at a final rating. This proposal was never adopted officially, but it does raise some points that we discuss in the next section.

[§] As originally defined by Fujita, the F-Scale wind speed units were in miles per hour (mph).

3.2. Desirable properties of a tornado intensity rating system

There are three fundamentally important properties of tornado *intensity* rating systems, and improving the quality of any one of them can degrade the quality of the others. As a result, changes in the systems can have unintended consequences and require careful consideration of the trade-offs.

The first desirable property is that it should resolve all physically possible wind speeds and provide enough damage indicators to be broadly *applicable*, whatever the local conditions along a given tornado path (cf. Brooks, 2004). Obviously, it would be optimal to have observations of winds covering the time and space volume for every tornado but, as admitted previously, in practice we have to fall back on damage to infer wind speeds.

Secondly, it should be *accurate*, in order to provide a climatology of intensity for all reported tornadoes. Given the difficulty of estimating wind speeds from damage, this is a challenging requirement. Clearly, there can be a fundamental trade-off between applicability and accuracy – highly accurate estimates may not be possible in most cases, for lack of appropriate indicators.

The third property is *consistency*. Ideally, the same process for ratings should be used everywhere through all time, to remove secular trends in the database. Again, this may not be feasible; differences in construction between countries and even within countries can make consistent evaluation difficult, to say nothing of past inconsistencies. Further, our methods inevitably evolve as the associated science, engineering, and technology change.

The recent changes in the USA's historical rating system illustrate the trade-offs. In principle, deploying the QRT as frequently as possible should help with accuracy and consistency for the rating of violent (F4+) tornadoes. The contributions of experienced, knowledgeable experts should lead to more accurate estimates done in consistent ways for surveyed events. Unfortunately, the relatively small group of such experts, as well as the cost of doing detailed ground and aerial surveys, limits the sample of events that can be surveyed to violent tornadoes – it would be impractical to use them for every tornado. Implicit in using the QRT is the unproven assumption that each of the experts would rate the same events equally. It is likely that the local survey teams, generally characterized by relatively little experience at the task, produce larger variability in how events are rated than the QRT members. These hypotheses have never been tested, however.

Again, in principle, the EF-scale should improve the accuracy and breadth of applicability in the USA. With more damage indicators, it becomes more likely that something will be damaged that can be compared to a database of expert judgment. Assuming that the expert judgments are accurate (which has not been tested), then that accuracy should be reflected in the ratings. To some extent, one major strength of the F-scale was its simplicity in having only one primary damage indicator, and it remains to be proven that the relative complexity of the EF-scale rating system is really an improvement over the simpler f- or F-scale systems.

Adoption of the EF-scale also raises disconcerting issues about consistency. Only if NWS offices use the portable database appropriately is it likely that the ratings will be done in similar ways around the USA, assuming that adequate training is provided. However, the apparent reluctance within the NWS to utilize the QRT procedure for possibly violent tornados has contributed to their climatologically implausible near-extinction in the recent record. It can be argued that without a period of overlapping use between the F-scale and the EF-scale, it is impossible to know whether the final ratings have changed because of the new guidance. However, we have shown that rating practices started to change well *before* the official adoption of the EF-scale in early 2007, after a period of consistent ratings in the decade of the 1990s. Hence, it is doubtful that a period of dual operation would have been worthwhile. The temporal consistency of the USA's tornado record evidently has been compromised in several different ways, with the adoption of the EF-scale being another example of evolving practice that likely can be attributed to the poorly understood relationship between wind and damage.

3.3. Implications for tornado ratings outside the USA

The present design of the EF-scale has also serious implications for intensity ratings of tornadoes and other small-scale damaging wind events worldwide. The original development of the F-scale in the 1960s and 1970s, accompanied by the proposal of the conceptually similar T-scale (Meaden, 1976)[**], occurred during a "dark age" of tornado research in Europe, in strong contrast to a very active period of such research between about 1850 and 1950 (cf. Wegener 1917; Dotzek et al. 2008a). In this "dark age" period, recording and rating of tornadoes was not consistently done on a routine basis in many European countries, as well as most nations around the world outside of North America. Initiatives to advance and update climatology and tornado hazard assessments mainly relied on the voluntary efforts of individual scientists and thus were not sustainable: data were gathered only for particular studies and not continued thereafter, or data collection ended with the retirement of the dedicated person.

[**] The T-scale is essentially the same as the F-scale but has twice as many categories, which implies greater precision. It has not been shown that this implied precision increase can be justified.

A gradual improvement began in the early 1980s. At this time, perhaps encouraged by the formal overview publication by Fujita (1981), the F-scale gained acceptance outside the USA. Authors like Fuchs (1981) had already proposed a tornado rating system with steps of intensity comparable to the F-scale classes F1–F3, but soon the F-scale became the most widely applied intensity scale. The data used by Dotzek et al. (2003, 2005) and Feuerstein et al. (2005) illustrate the F-scale's worldwide application. However, in contrast to the development in the USA, tornado ratings in Europe never have been tied to one particular damage indicator like the "well-constructed frame house"; rather, they have been based on *all* the available damage information for each case, including damage to vegetation (cf. Wegener 1917). It is significant to note that in cases with neither damage nor windspeed information, consequently no intensity rating had been assigned to the event.

To provide the link between the velocity intervals of the F-scale to the locally observed damage, regional descriptions of typical damage were created in Europe, relying on the fact that building construction standards were more homogeneous and generally higher than in the central part of the USA. Dotzek et al. (2000) had set up such a damage description valid for central Europe, in cooperation with Munich Re Group. Its basic treatment of vegetation damage was later augmented by Hubrig (2004) and applied by Svabik and Holzer (2005) in their analysis of tornadoes in Austria. The resulting description has been made available online (www.tordach.org/pdf/FT_scales.pdf , with an English version augmented with exemplifying damage photographs to appear under www.essl.org/research/scales/). The experience with having only one definition of the wind speed intervals and then adding regionally valid damage descriptions has been seen as beneficial, helping to ensure that international tornado ratings refer to a uniform wind speed range and thereby remain climatologically consistent and comparable.

Over the last ten years, awareness of tornadoes and other severe thunderstorm phenomena has increased significantly, leading to increasing reports of tornado occurrence in Europe (Dotzek 2003). We can expect several hundred tornadoes over land in Europe each year, and the recently established European Severe Weather Database (ESWD, www.essl.org/ESWD/, see Dotzek et al. 2008b) confirms these numbers. Presently, four European national meteorological and hydrological services (NMHS) are collaborating with the ESWD, but its main strength is to allow for public severe weather reports as well. This strongly enhances the data density, especially in regions where the operational observing networks are coarse or increasingly reliant on automatic stations. There are no default intensity ratings in the ESWD for tornadoes with no or insufficient damage information, and the source of information forming the basis of any intensity rating is part of the metadata accompanying the report. Furthermore, if additional evidence becomes available for a particular severe weather case later on, its ESWD record and potentially also its intensity rating, can be revised in the quality-control procedure.

Dotzek et al. (2008b) have compared the intensity distribution of all rated tornadoes in Europe to those from the USA in the time period 1920 to 1999. The two distributions are very similar, except for a greater underreporting of weak tornadoes (F0 on the F-scale) that persists in Europe. The similarity of the distributions is reassuring and gives us confidence that worldwide homogeneity of tornado ratings is possible, so long as there is an agreed-upon worldwide wind speed scale with regionally-adapted degree of damage descriptions tied to those wind speeds.

By switching to the EF-scale with its revised wind speed estimates in the USA, the consistency of ratings in Europe and worldwide is at stake. The F-scale has only recently become an international standard, and many European nations still lack tornado records based on F-scale of sufficient length to assess if introduction of a modified EF-scale – specifically, adapted to local European construction practices – could bring any improvement. Some persons doing the initial ratings have only limited experience and training. Yet, even though the European Severe Weather Database will continue to also depend on volunteer reports from the public, there is an increasing involvement of NMHS employees and ESSL staff in the provision and quality-control of the ratings. Nevertheless, no European counterpart to the QRT exists to date.

It is logically possible (but as yet unproven) that adoption of the EF-scale has produced more accurate estimates of winds that cause damage in the USA. As noted, the EF-scale is more complicated to apply and is directly applicable only to USA construction practices. The effort to produce its decision matrix was considerable and it is not yet clear that its benefits justify carrying out a similar effort in Europe to modify the EF-scale to incorporate sufficient local knowledge of construction practices under the upcoming EU building code. So, it is likely that for practical reasons, use of the F-scale in Europe will have to continue, at least for some time.

Dotzek (2008) recently proposed the Energy- or "E-scale" as a wind speed scale that can be calibrated and is coupled to physical quantities X like mass flux or momentum density ($M = \rho\, v$, where ρ is air density and v is wind speed), kinetic energy density ($E = \rho\, v^2/2$) or the kinetic energy flux density ($P = \rho\, v^3/2$). In short, a nonlinear scaling in these quantities

$$X_* (\mathrm{X} - \mathrm{X}_0)^n = a_x\, v^n \qquad (1)$$

results in a universal windspeed-scale relation which is always linear in v:

$$v(X) = v_* \ (X - X_0) \quad , \quad \text{with} \quad v_* = [a_x^{-1} X_*]^{1/n} \quad . \tag{2}$$

In Eq. (1), the scaling quantity X_*, the prefactor a_x and the exponent n depend on either of the physical observables M, E, P. The scaling velocity v_* in Eq. (2) is determined by the choice of the critical values M_*, E_*, or P_*, allowing for calibration of the scale. The well-known Mach scale is a special case of the E-scale.

In the initial proposal by Dotzek (2008), the scaling velocity was chosen to facilitate conversion of existing worldwide F-scale data to their E-scale intervals. Results suggested that mainly the high-F4 and F5 tornadoes would have to be re-rated[††] and that for tornadoes stronger than F3, the new thresholds lie at lower wind speeds than proposed in the original F-scale. Thus, some of the requirements set up in developing the EF-scale are fulfilled by the E-scale. Further, the E-scale wind speeds are applicable worldwide, they are open to calibration, and they avoid the subjectivity of "expert elicitations" as done for the EF-scale.

For these reasons, coupling the E-scale concept to detailed regional damage descriptions as done with the EF-scale for the USA may provide a way to overcome many of the F-scale shortcomings without endangering the international consistency and the physical basis for tornado ratings worldwide. With a suitable dialogue between atmospheric scientists and wind engineers, this should be a manageable task.

4 Conclusions and recommendations

4.1 Conclusions

This paper has reviewed and identified the shortcomings of the original F-scale, despite its greatest strength: simplicity. The shortcomings of the EF-scale have also been identified, as well as its major strength: provision of a larger set of damage indicators. Although North America has the highest tornado occurrence rate worldwide, and the USA continues to run the most advanced programmes in tornado research and forecasting, it is evident that the methods used for rating tornado intensity in the USA have been changing ever since the F-scale was adopted. Replacement of the F-scale by the EF-scale is only the latest episode in the story of that evolution. We have shown evidence for major secular trends in the data that are unlikely to originate in real climatological changes. Therefore, we conclude that the USA tornado intensity ratings have been compromised. We have shown this began prior to the adoption of

[††] If the necessary metadata were available for the US record, see Sec. 3.1 and Dotzek (2008).

the EF-scale. It is likely that formal implementation of the EF-scale was premature, given the continuing research efforts in relating wind measurements to observed damage levels.

Further, the EF-scale is openly associated with USA-specific construction practices. This raises more concerns about its adoption. Although the most desirable tornado intensity scale would be tied either directly (as was the original F-scale) or indirectly (as with the proposed E-scale) to wind speeds, it is apparent that this continues to be impractical for doing tornado intensity ratings. Before the adoption of the Richter scale by seismologists around the world, which measures the magnitude of earthquakes by the energy released, it was preceded by a subjective, damage-based intensity scale. The development and adoption of the Richter scale was a great advance for seismology and we believe that ultimately some objective measure of tornado wind speeds would be of similar value to tornado science. Nevertheless, barring some unforeseen breakthrough in technology, a damage-based scale remains the only practical alternative.

4.2 Recommendations

We have argued it would be highly desirable to find a procedure for tornado ratings open to detailed, *regional* damage indicators and degree of damage descriptions and which relies on a wind speed range categorization that encompasses the full range of wind speeds physically possible in tornadoes. This procedure needs to have the flexibility to be recalibrated with new findings from either wind engineering or mobile Doppler radar data, for instance.

It likely would be beneficial to establish formal international communication channels to discuss rating issues. In the USA, there is an online forum for experts and NWS personnel, although it is not evident that it is being used to its full extent. In Europe, similar fora exist, mainly tied to the developing Skywarn network, but not yet fully established within the European NMHSs. Although to obtain high-resolution wind speed measurements for tornadoes anywhere in the world will remain impractical, we maintain that an accurate wind speed-damage relationship as part of the tornado intensity rating scale should be continued. The debate over that relationship will go on, but it seems likely that the existing EF-scale's high-end wind speeds have been revised too far downward from the F-scales's original values for what is physically possible in tornadoes. Adoption of the EF-scale appears to pre-empt continuing debate on the topic, which we don't believe is a correct perception. The official recognition of the EF-scale by the NWS does *not* signify that any formal process exists within the NWS for making changes to the EF-scale, if needed. In fact, it is unclear just how such changes could be implemented.

In this situation, the new E-scale concept (section 3.3) can help the scientists and engineers to come to valid conclusions what a universal windspeed relation could be. Therefore, we recommend a continued discussion between atmospheric sciences and wind engineering in order to develop a synthesis of a (calibrated) E-scale and regionally adapted damage indicator / degree of damage decision matrices.

We further recommend that if large changes are being considered in rating practice outside the USA, a parallel period of rating with both systems should be used to gauge the effects of the changes. There should be considerable dialog between those who will be making tornado intensity ratings abroad and those with experience who are doing so in the USA. Although countries outside the USA can and should develop their own methods, being aware of the experiences from the USA seems valuable. We also urge the use of "unknown" or "unrated" as a damage category for those cases in which insufficient evidence exists to assign a rating with any confidence. We also recommend that some formal process for continuing revision of the EF-scale needs to be established.

Finally, we believe that any database for documenting tornado occurrences should include the capability for providing extensive metadata information about the sources used in the documentation – as prescribed, for instance, in the ESWD data format (www.essl.org/reports/tec/ESSL-tech-rep-2006-01.pdf). If it is accepted that *any* rating, including those based on direct wind measurements, inevitably have some degree of uncertainty, then source information is critical in estimating that uncertainty. This applies not only to the tornado intensity rating – it also applies to all the other documentation (e.g., path width, path length, etc.).

Acknowledgments

ND's contribution to this paper was partly funded by the German Ministry for Education and Research (BMBF) under contract 01LS05125 in the project RegioExAKT (Regional Risk of Convective Extreme Weather Events: User-oriented Concepts for Climatic Trend Assessment and Adaptation, www regioexakt.de) within the research programme *klimazwei*. Our reviewers offered helpful comments, which we appreciate.

References

Bluestein, H. B., and W. P. Unruh, 1989: Observations of the wind field in tornadoes, funnel clouds, and wall clouds with a portable Doppler radar. *Bull. Amer. Meteor. Soc.*, **70**, 1514-1525.

Bluestein, H. B., J. G. Ladue, H. Stein, D. Speheger, and W. F. Unruh, 1993: Doppler radar wind spectra of supercell tornadoes. *Mon. Wea. Rev.*, **121**, 2200-2222.

Brooks, H. E., 2004: On the relationship of tornado path length and width to intensity. *Wea. Forecasting*, **19**, 310-319.

Brooks, H. E., and C. A. Doswell III, 2001: Some aspects of the international climatology of tornadoes by damage classification. *Atmos. Res.*, **56**, 191-202.

Brooks, H. E., and N. Dotzek, 2008: The spatial distribution of severe convective storms and an analysis of their secular changes. In: *Climate Extremes and Society*. H. F. Diaz and R. Murnane, Eds., Cambridge University Press, Cambridge, 340 pp, 35-53.

Bunting, W. F., and B. E. Smith, 1993: *A Guide for Conducting Damage Surveys*. NOAA Tech. Memo. NWS-SR-146, 44 pp.

Burgess, D. W., M. A. Magsig, J. Wurman, D. C. Dowell, and Y. Richardson, 2002: Radar observations of the 3 May 1999 Oklahoma City tornado. *Wea. Forecasting*, **17**, 456-471.

Doswell, C.A. III, 2003: *A Guide to F-Scale Damage Assessment*. U.S. Dept. of Commerce, NOAA/NWS, 94 pp. [Available at http://www.flame.org/~cdoswell/publications/F-scale_Assess_Guide.pdf]

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

Dotzek, N., 2003: An updated estimate of tornado occurrence in Europe. *Atmos. Res.*, **67-68**, 153-161.

Dotzek, N., 2008: Derivation of physically motivated wind speed scales. *Atmos. Res,* this volume.

Dotzek, N., G. Berz, E. Rauch, and R. E. Peterson, 2000: Die Bedeutung von Johannes P. Letzmanns "Richtlinien zur Erforschung von Tromben, Tornados, Wasserhosen und Kleintromben" für die heutige Tornadoforschung. (The relevance of Johannes P. Letzmann's "Guidelines for research on tornadoes, waterspouts, and whirlwinds" for contemporary tornado research). *Meteorol. Z.*, **9**, 165-174. [In German, available under „Publications" at http://www.essl.org]

Dotzek, N., J. Grieser, and H. E. Brooks, 2003: Statistical modeling of tornado intensity distributions. *Atmos. Res.*, **67-68**, 163-187.

Dotzek, N., M. V. Kurgansky, J. Grieser, B. Feuerstein, and P. Névir, 2005: Observational evidence for exponential tornado intensity distributions over specific kinetic energy. *Geophys. Res. Lett.*, **32**, L24813, doi:10.1029/2005GL024583.

Dotzek, N., R. E. Peterson, B. Feuerstein, and M. Hubrig, 2008a: Comments on "A simple model for simulating tornado damage in forests". *J. Appl. Meteor. Climatol.*, **47**, 726-731.

Dotzek, N., P. Groenemeijer, and B. Feuerstein, 2008b: Overview of ESSL's severe convective storms research using the European Severe Weather Database ESWD. *Atmos. Res*., this volume.

Feuerstein, B., N. Dotzek, and J. Grieser, 2005: Assessing a tornado climatology from global tornado intensity distributions. *J. Climate*, **18**, 585-596.

Fiedler, B. H., 1998: Wind-speed limits in numerically simulated tornadoes with suction vortices. *Quart. J. Roy. Meteor. Soc.*, **124**, 2377-2392.

Fiedler, B. H., and R. Rotunno, 1986: A theory for the maximum windspeeds in tornado-like vortices. *J. Atmos. Sci*., **43**(21), 2328-2340.

Forbes, G. S., and R. M. Wakimoto, 1983: A concentrated outbreak of tornadoes, downbursts and microbursts, and implications regarding vortex classification. *Mon. Wea. Rev.*, **111**, 220-235.

Fuchs, D., 1981: Gefährdung des Tiefflugs durch Tornados (Tornado hazard to low-flying aircraft). *Promet*, **81**(4), 8–10. [In German]

Fujita, T. T., 1971: Proposed characterization of tornadoes and hurricanes by area and intensity. SMRP research paper no. 91, University of Chicago, 42 pp.

Fujita, T. T., 1981: Tornadoes and downbursts in the context of generalized planetary scales. *J. Atmos. Sci.*, **38**, 1511-1534.

Fujita, T. T., 1992: *Mystery of Severe Storms*. Chicago University Press, Chicago. 298 pp.

Fujita, T. T., D. L. Bradbury, and C. F. van Thullenar, 1970: Palm Sunday tornadoes of April 11, 1965. *Mon. Wea. Rev.*, **98**, 29-69.

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

Hubrig, M., 2004: Analyse von Tornado- und Downburst-Windschäden an Bäumen (Analysis of tornado and downburst wind damage to trees). *Forst und Holz*, **59**, 78-84. [In German]

Kelly, D. L., J. T. Schaefer, R. P. McNulty, C. A. Doswell III, and R. F. Abbey, Jr., 1978: An augmented tornado climatology. *Mon. Wea. Rev.*, **106**, 1172-1183.

Lewellen, D. C., and W. S. Lewellen, 2007: Near-surface intensification of tornado vortices. *J. Atmos. Sci.*, **64**, 2176-2194.

McCarthy, D. W., J. T. Schaefer, and R. Edwards, 2006: What are we doing with (or to) the F-Scale? *Preprints, 23rd Conf. on Severe Local Storms*, St. Louis, MO, Amer. Meteor. Soc. [Available at http://www.spc.noaa.gov/publications/mccarthy/slsc23.pdf].

Meaden, G. T., 1976: Tornadoes in Britain: Their intensities and distribution in space and time. *J. Meteor.*, **1**, 242-251.

Minor, J. E., J. R. McDonald, and K. C. Mehta, 1977: The tornado: An engineering-oriented perspective. NOAA Tech. Memo ERL NSSL-82, NTIS Accession No. PB 281860/AS, 196 pp.

Phan, L. T., and E. Simiu, 2002: The Fujita scale: A reassessment from a structural engineering perspective. *Proceedings, U.S./Japan Natural Resources Development Program*, Joint meeting 31st Tech. Memo. of PWRI 3653, Tsukuba, Japan, 469-474. [Available at http://www.fire.nist.gov/bfrlpubs/build99/art111.html]

Potter, S., 2007: Fine-tuning Fujita. *Weatherwise*, **60**(2), 64-71.

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.

Svabik, O., and A. M. Holzer, 2005: Kleinräumige, konvektiv verursachte Stürme und Wirbelstürme (Tornados) in Österreich (Small-scale, convectively induced windstorms and tornadoes in Austria). *Österr. Beitr. Meteorol. Geophys.*, **36**, ZAMG Publ. Nr. 417, Vienna, 91 pp. [In German]

Verbout, S. M., H. E. Brooks, L. M. Leslie, and D. M. Schultz, 2006: Evolution of the U.S. tornado database: 1954-2003. *Wea. Forecasting*, **21**, 86-93.

Wegener, A., 1917: *Wind- und Wasserhosen in Europa (Tornadoes and Waterspouts in Europe)*. Verlag Friedrich Vieweg und Sohn, Braunschweig, 301 pp. [In German, available under "References" at http://www.essl.org].

Wurman, J., J. Straka, E. Rasmussen, M. Randall, and A. Zahrai, 1997: Design and deployment of a portable, pencil-beam, pulsed, 3-cm Doppler Radar. *J. Atmos. Oceanic Technol.*, **14**, 1502-1512.

Wurman, J., and C. R. Alexander, 2005: The 30 May 1998 Spencer, South Dakota, storm. Part II: Comparison of observed damage and radar-derived winds in the tornadoes. *Mon. Wea. Rev.*, **133**, 97-119.

Figure captions

Fig. 1: Annual counts of tornadoes rated F1 or greater (F1+) in the USA from 1950-2006 (solid circles) and the number of tornados rated F2 or greater (F2+) for consecutive runs of 1000 F1+ tornadoes (line). The F2+ count is for the period beginning with the date on the horizontal axis, continuing until 1000 F1+ tornadoes are reported.

Fig. 2: The number of days until the next violent tornado (F4 or F5) occurs in the Southern Region of the NWS from 1904-2007. The value of the points is the number of days from the date of one violent tornado to the next violent tornado – for example, the maximum value shown (1393 days) is plotted at the date of 8 May 2003 and represents the gap between that date and 1 March 2007.

Fig. 3: Fujita's *f*-scale matrix from 1992 (units adapted).

Figures

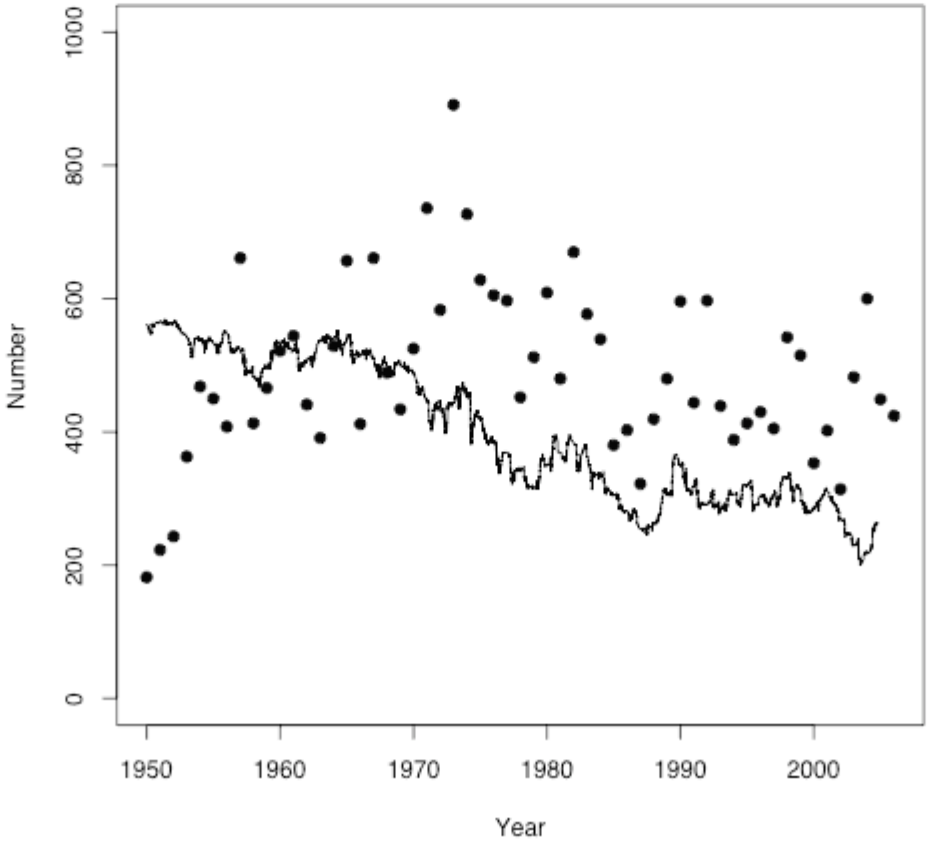

Fig. 1: Annual counts of tornadoes rated F1 or greater (F1+) in the USA from 1950-2006 (solid circles) and the number of tornados rated F2 or greater (F2+) for consecutive runs of 1000 F1+ tornadoes (line). The F2+ count is for the period beginning with the date on the horizontal axis, continuing until 1000 F1+ tornadoes are reported.

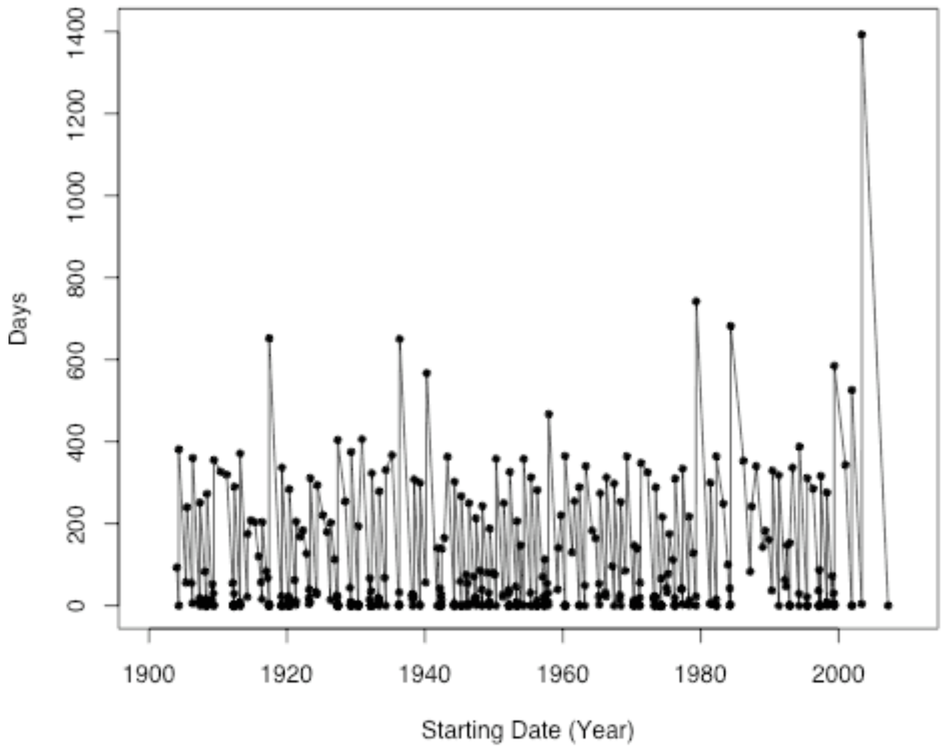

Fig. 2: The number of days until the next violent tornado (F4 or F5) occurs in the Southern Region of the NWS from 1904-2007. The value of the points is the number of days from the date of one violent tornado to the next violent tornado – for example, the maximum value shown (1393 days) is plotted at the date of 8 May 2003 and represents the gap between that date and 1 March 2007.

Damage: f scale		Little Damage	Minor Damage	Roof Gone	Walls Collapse	Blown Down	Blown Away
		f0	f1	f2	f3	f4	f5
Windspeed: F scale		F0	F1	F2	F3	F4	F5
		18 m/s – 33	33 – 50	50 – 70	70 – 93	93 – 117	117 – 143
		64 km/h – 118	118 – 181	181 – 254	254 – 333	333 – 420	420 – 513
To convert f scale into F scale, add the appropriate number							
Weak Outbuilding	−3	f3	f4	f5	f5	f5	f5
Strong Outbuilding	−2	f2	f3	f4	f5	f5	f5
Weak Framehouse	−1	f1	f2	f3	f4	f5	f5
Strong Framehouse	0	F0	F1	F2	F3	F4	F5
Brick Structure	1	–	f0	f1	f2	f3	f4
Concrete Building	2	–	–	f0	f1	f2	f3

Fig. 3: Fujita's *f*-scale matrix from 1992 (units adapted).