Paper 9.4, $2^{nd}$ *Conf. Environ. Appl.*, Long Beach, CA (8-12 January 2000), Amer. Meteor. Soc., 212-219.

# CLIMATOLOGICAL RISK OF STRONG AND VIOLENT TORNADOES IN THE UNITED STATES

Peggy R. Concannon
Northern Illinois University
DeKalb, Illinois

Harold E. Brooks and Charles A. Doswell III
NOAA/ERL/National Severe Storms Laboratory
Norman, Oklahoma

## 1. INTRODUCTION

Knowledge of tornado climatology is valuable to a variety of groups, especially weather forecasters, emergency management officials, and the public. It is crucial for such individuals to understand the threat posed by tornadoes in the United States, particularly the threat of strong and violent tornadoes. In this study, the term strong refers to those tornadoes producing F2 or F3 damage while violent refers to those producing F4 or greater damage. Significant refers to any F2 or greater tornado. Even though only about 10% of tornadoes are significant, these tornadoes are responsible for the majority of deaths caused by tornadoes in the country, with violent tornadoes claiming 67% of the total. Furthermore, with the aftermath of such events, the US suffers millions of dollars in damage costs—an important consideration for the insurance industry. Due to this destructive potential toward life and property, we chose to consider these tornadoes only in our study, using data from 1921-1995. Furthermore, the significant tornado dataset is likely to be more reliable than that for any tornado since they are more likely to be observed. Typically, they are larger and thus more visible, with longer path lengths, and they cause the most damage. We attempted to estimate the daily climatological probability of an F2 or more damaging tornado occurring near any location in the US.

Our results have implications for planning for natural disasters as well as setting a baseline for answering the question of detecting the effects of possible climate change on severe weather events.

## 2. DATASET

The dataset we have used is from Grazulis (1993, hereafter G93), looking at the years of 1921-1995, containing over 10,000 tornadoes listed by date, damage classification, and location of touchdown by county. G93 attempted to produce a homogeneous dataset. Although any decisions made in judging the damage classification of any tornado are necessarily arbitrary, particularly for historic events, we believe it is advanatageous to use this dataset since, we believe, those decisions should be more internally self-consistent. A comparison of the official National Weather Service dataset and G93 will be made a later date. There are differences in the two as noted by G93, but the question of differences in important statistical properties is still open.

It is important to note the limitations of the entire raw tornado report dataset. Intrinsic problems exist in the Fujita rating scale itself. This scale is damage-based; analysis of wind speed, which categorizes each tornado intensity rating, is possible only if a tornado causes damage. Additional problems related to a system based on damage are the quality of the building construction and materials used. Readers are referred to G93 for supplemental discussion of the F-scale problems.

However, the damage scale is obviously not the only source of error in tornado data. Numerous dilemmas exist which have contributed to fallacious and erroneous reports. Among these are recent increased detail in damage surveys, misinterpreted observations, population biases, and influences of public awareness. Doswell and Burgess (1988) discuss these issues in further detail.

By choosing a dataset of only significant tornadoes, we have encountered both limitations and benefits. The problems specific to the dataset of significant and violent tornadoes is that the number of strong tornadoes only comprises about 30% of all tornadoes, while the number of violent tornadoes only make up about 2% of the total. This fact demonstrates how small the sample size is for these more damaging tornadoes. Thus, with such small sample sizes, one must take caution when interpreting the results. On the other hand, the dataset for significant and violent tornadoes may have more reliability over that for the weak tornadoes. The fact that more people are likely to observe an intense tornado due to its larger size and longer path length reduces the number of "missing" significant or violent tornadoes within the dataset. According to G93, the majority of these intense tornadoes were either on the list of known tornadoes or could be located.

## 3. METHODOLOGY

Due to such limitations in the raw dataset, we have taken a conservative approach to the data analysis. Rather than dealing with issues of path length or width, we have started by considering only the date of tornado touchdown and its approximate location because we believe these are the most reliable and temporally consistent aspects of the reports. This allows us to find the probability of touchdown rather than probability of occurrence. Although this leads to limitations in the results, looking at these two pieces of data should provide a good estimate of what we are considering. Future work can include additional complexity.

In order to place the locations of tornado touchdowns onto a grid, we converted the locations listed by counties to that of the latitude and longitude of the centroid of the county and mapped the data onto a Lambert conic conformal projection with a nominal grid spacing of 80 km (true at 30° and 60° N). Our basic unit of data is whether a tornado touched down or not on any particular day in the dataset. If more than one tornado occurred in a particular grid box, only the most damaging is mapped. We emphasized "tornado days" rather than the actual number of tornadoes since, although changes in the reporting of severe weather has had a large impact on the raw number of tornadoes, the more conservative variable, tornado days, has been more consistent through time. The use of tornado days, so that the value at any grid point on any day is either 0 or 1 allows for a simple interpretation of the final products as either the mean value on any particular day or the probability of the event on that day. The product is a 62 x 49 grid covering the US and southern Canada. The area of each grid box is roughly equivalent to the area of a circle 25 statute miles in radius.

The approach we have taken effectively assumes that there is some unknown, underlying statistical distribution of tornadoes. We have attempted to recover a distribution that is consistent with that underlying statistical distribution by creating smoothed fields in space and time. We have used Gaussian smoothers in space and time (Silverman 1986) that are relatively large. By doing so, we cannot capture small-scale variability. However, we can see relatively large-scale features that are reasonably reliable. We were interested only in these strong signals.

To perform the smoothing, we first calculated the mean number of tornado-days that occurred at each grid point for each day of the year for certain time periods. We smoothed in the time dimension to find the mean value on any day of the year, assuming the data were periodic. We used a Gaussian smoother which assumes that data from one particular day of the year will provide information about the probability of tornado touchdown on days close to that particular day. The standard deviation of the Gaussian in the temporal smoothing parameter was 15 days to provide a slowly varying cycle.

After we smoothed in time, we smoothed in space in both north-south and east-west directions to determine the probability of a tornado touchdown being reported in the grid box at any location on our grid. The standard deviation of the Gaussian in the spatial smoother was 120 km. Again, these strong smoothers mean that we can only see strong signals. The technique can be applied to any period of time from one year up to the whole 75 years, providing smoothed estimates of the underlying probability of tornadoes on any day of any year for any location. For more details, see Brooks (1999).

## 4. RESULTS

Perhaps the most basic and important quantities that can be derived from the data are the total threat of tornado touchdown, which, for our definition, is described by the mean number of days per year with at least one tornado at each grid point. An L-shaped maximum in the number of days per century with a significant tornado (F2 or greater) occurring stretches from southeastern Mississippi to southern Oklahoma, then north-northeastward to western Iowa (Fig. 1). Throughout this area, there are more than 25 significant tornado days per century, with a peak value of near 40 days, just southeast of Oklahoma City. The probabilities fall off rapidly to the west and north of the "L" and more slowly in the northeastern section. It is possible that poor reporting associated with low population in the region through the Texas and Nebraska Panhandles and the Dakotas may limit the western extent. The threat east of the Appalachians is very low.

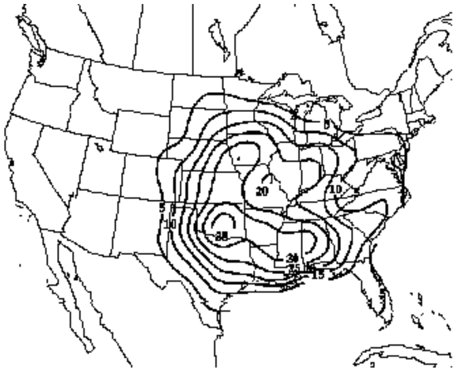

Fig. 1: Mean number of days per century with at least one F2 or greater tornado touching down in grid box. Based on data from 1921-1995. Contour interval 5 days, with lowest contour = 5.

If we limit our attention to violent (F4 or greater) tornadoes, the gross features are the same (Fig. 2). Values are roughly 10% of those for signficant tornadoes, but the location of the maximum reported threat stays the same in south-central Oklahoma. In general, the "L" is still present, although the base is less prominent. The peak of the threat is slightly over 50 days per millenium or, equivalently, the maximum climatological probability of a violent tornado within approximately 25 miles of any point during the course of a year in the US is about 5%.

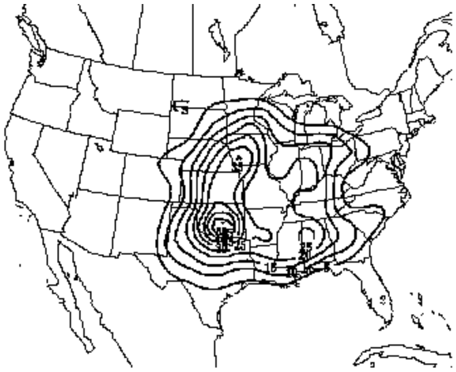

Fig. 2: Mean number of days per millenium with at least one F4 or greater tornado touching down in grid box. Based on data from 1921-1995. Contour interval 5 days, with lowest contour = 5.

Separation of the data into shorter sub-periods of record is revealing. Looking at separate 15-year periods shows that the gross overall pattern of the distribution of significant tornadoes is similar, although details are different (Fig. 3). In particular, the earliest 15 years of the record (1921-1935) are very similar to the last 15 years (1981-1995). It is especially noteworthy that this resistant variable (significant tornado days) shows no long-term continuous increase, in contrast to less resistant variables, such as the raw number of tornadoes (see Schaefer and Brooks, 2000). In fact, there are slightly more significant tornado days in the early record (Fig. 3a) in southern Kansas and Oklahoma than in the last par t of the record (Fig. 3c). The biggest difference in the record is in the 1951-1965 period (Fig. 3b), when the number of significant tornado days was much higher than in any of the other periods, with approximately 20% more significant tornado days occurring then.

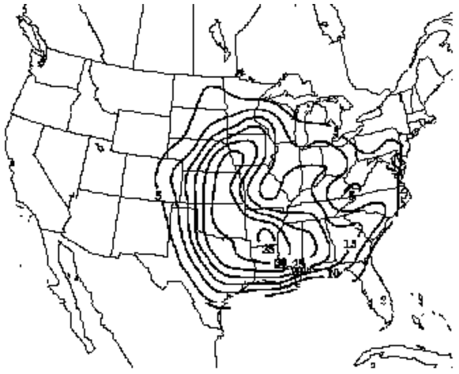

Fig. 3a: Same as Fig. 1 except for period 1921-1935.

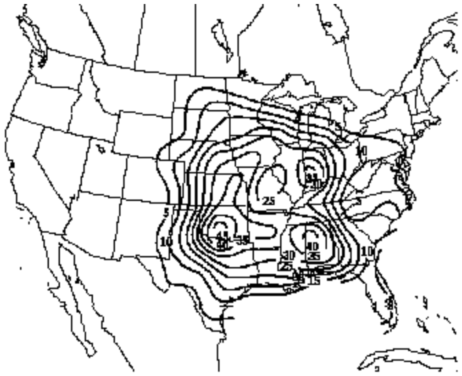

Fig. 3b: Same as Fig. 1 except for period 1951-1965

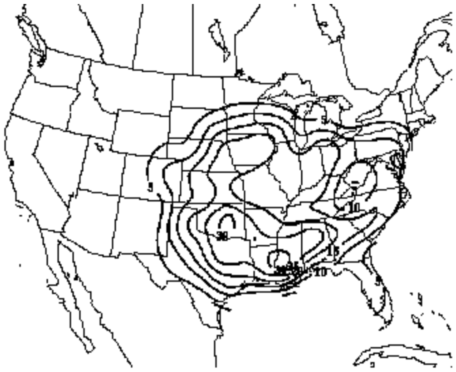

Fig. 3c: Same as Fig. 1 except for period 1981-1995.

Looking at the annual cycle of tornado threat at any individual location is also insightful. The relative consistency of the spring maximum in Oklahoma is evident, while the low-grade fall threat is very incon-sistent there (Fig. 4a). Based on the 15-year periods, four of the five periods see the maxima in the annual cycle occur within two weeks of 1 May. Maximum probabilities on any particular day for any period are less than 0.1%. The 3 May 1999 tornado occurred very near the peak (1 May) in the annual cycle based off the full record. Thus, in some sense, it was the most probable violent tornado that could occur.

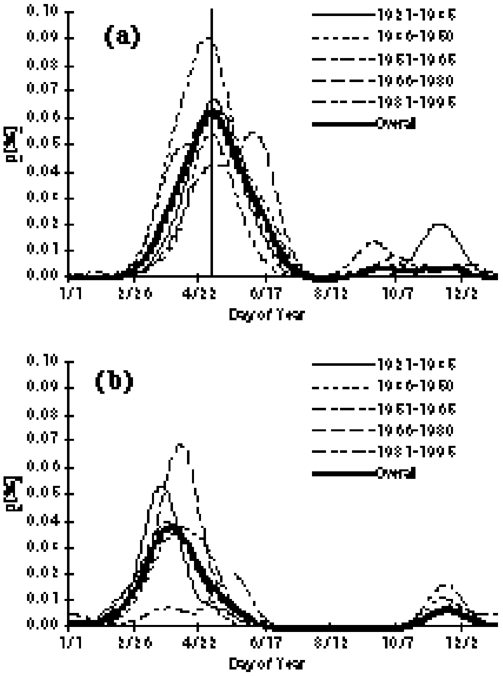

Fig. 4: Annual cycle of probability in percent of violent tornado touchdown based off of 5 different 15-year periods, as well as all 75 years of record (Overall). a) Near Oklahoma City with vertical line is at 3 May. b) In northwestern Alabama near local maximum in Fig. 2.

In contrast to the Oklahoma point, a point from northwestern Alabama shows a less consistent spring maximum and a more consistent fall maximum centered in mid-to-late November (Fig. 4b). At the time of the spring maximum (late March-early April) in fact, one of the five sub-periods (1981-1995) shows a relative minimum, so that the estimates of when the peak threat is varies from the middle of March until late May.

In an effort to assess objectively the variability of the timing of the season, we have calculated 'climatological' annual cycles for every grid point for every year in the record for both significant and violent tornadoes. We then calculate the day of the peak for each of the 75 years. In some locations, such as near Dallas, Texas, the dates of the peak threat of a significant tornado occurring are tightly clustered from year to year (Fig. 5a), so that almost 80% of the years have peaks occurring during a 50 day period from late March until late May (Julian dates 87-137). In contrast, eastern Mississippi shows much less clustering with only 40% of the years having a peak within a 50 day period from late February until late April (Julian dates 57-107). A map of the percentage of years that have maxima within any 50-day window at each location shows a region with little variability in the date of the peak extending from the northern half of Texas up through Nebraska and western Iowa (Fig. 6). At least 60% of years in this region have their peak probability of significant tornado occurring in a 50-day window. East of this area, the clustering of the maxima falls fall to less than half of the years. This makes it hard to define when the tornado season is, in some sense.

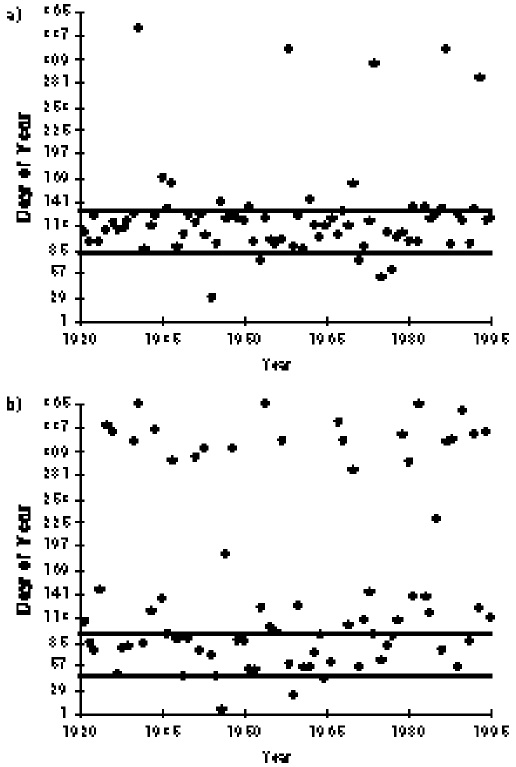

Fig. 5: Day of year of maximum in significant tornado probability for each year at single point. a) Point near Dallas, Texas illustrating strong clustering of maxima. b) Point in eastern Mississippi illustrating weak clustering. Horizontal lines enclose 50-day period with most years with maximum day in period.

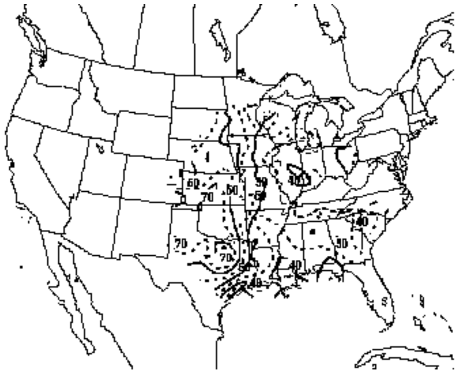

Fig. 6: Percentage of years with date of maximum threat of F2 or greater tornado falling in any 50-day long window. Contour interval is 5% with lowest contour value 30% and highest value 75%. Every other contour is dashed. Only locations with 10 days per century are contoured.

The timing of the peak threat is obviously a feature of interest. Keeping in mind that, as a result of the interannual variability, estimates of when the peak is may not be robust in the eastern US, we can make an estimate of the timing by considering the timing of the peak in the mean annual cycle (Fig. 7). In the southeast US, the peak occurs in March and April. As the year progresses, the maximum moves outward from Alabama. It slowly progress north from Texas to Canada from April to July. The seasonal transition is sharper in the northeast, going from April to July across Pennsylvania. It is important to note that the number of tornado days in that part of the country is much smaller than to the west and south (see Fig. 1), so that the estimates of the timing may not be reliable and, even if they are, may not be particularly meaningful. In contrast to the climatology of all tornadoes (Brooks 1999) that shows a maximum in the seasonal cycle in November in the southern half of Mississippi, no location has a peak in its mean cycle in the fall of the year.

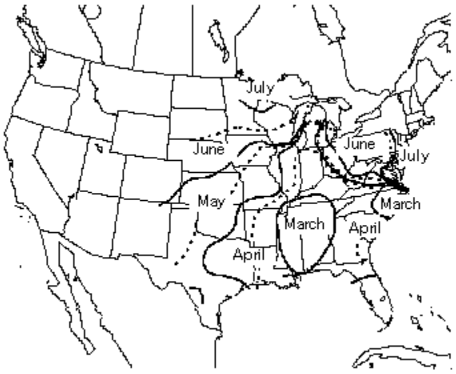

Fig. 7: Progession of maximum threat of significant tornado during year. Date based on mean of 75-year sample. Solid lines indicate first day of month, dashed lines 16th day of Month.

## 5. DISCUSSION

This work is part of a larger effort to estimate and describe the threat of a variety of weather hazards in the US and, eventually, to make estimates of those threats around the world. We would like to emphasize several points that are particularly important.

First, any efforts to make climatological estimates require careful consideration of the nature of the problem at hand, and of available datasets. One of the most challenging aspects of dealling with convective storm reports is gathering a sufficiently large sample to have confidence in the meaning of the results, but to avoid the problems of large changes in the reporting database over time. Clearly, using the overall reports of severe weather back until 1950 will cause significant difficulties unless the temporal changes in the reports are accounted for (Schaefer and Brooks, 2000). Blindly analyzing data without regard for the ways in which the data were collected can lead to serious problems. In this case, we have attempted to use as homogeneous of a dataset as possible, but using G93. It is not a perfect representation of what occurred, but we believe it to be as consistent of a dataset as there is. We have enhanced its consistency even further by considering our 'event' to be a tornado day.

As far as results of the study are concerned, the primary area of the US in which significant tornadoes occur most often is in a L-shaped region from Iowa to Oklahoma to Mississippi, with the highest threat in Oklahoma. It is important to remember that this is based solely upon the reports in G93. It is possible that low population densities and the accompanying small number of structures (Rasmussen and Crosbie 1996), particularly west of 100° W longitude, may lead to an underreporting of events. This problem is likely to be most severe for the violent tornadoes, since the sample size is much smaller for them. With these caveats in mind, we believe the overall general pattern is reasonable.

The movement of the peak in tornado threat during the year is consistent with changes in the annual cycle of meteorological variables. As moisture from the Gulf of Mexico is advected northward and westward over the Great Plains during the spring and early summer, the timing of the maximum in the annual cycle moves with it. The threat at southern locations weakens in summer as the jet stream retreats northward.

We can find no evidence for a long-term increase or decrease in the threat from significant tornadoes. The evidence of variability between different subperiods in the record indicates that changes in the frequency of significant tornado days on the order of 25% have occurred in this century. Detecting any changes related to climate change may be very difficult, given the apparently high natural variability.

The maximum frequency for having strong tornadoes come close to any location, in our case approximately 25 miles, is roughly once every two to three years. It is, obviously much less outside of the peak region and, if a smaller area of concern is defined, the return time between events is even longer. If we limit our concern to violent tornadoes, such as the 3 May 1999 Oklahoma City tornado, they occur near a location on one day once every 20 years or longer. As a result, people experience these events very rarely and it is a difficult challenge to keep people prepared for these events. Education in the schools appears to be an excellent way to increase knowledge of how to respond to tornadoes and it may have played a major role in the fact that there were no fatalities between the ages of 4 and 24 in the Oklahoma City tornado.

Insurance companies could increase the tornado-resistant properties of buildings by giving advantageous rates to structures that are built to be survivable. While individual residents may not be in a particular house long enough to experience a tornado in their vicinity, it is much more likely that, over the longer lifetime of the house, a violent tornado will occur relatively close to the house. Tying the preferred rates to the property and not to the individual who makes the improvements would increase the use of such structures. Similarly, development and enforcement of building codes designed to enhance tornado survival could be useful in high-risk regions.

The conjunction of high frequency of strong and violent tornadoes and the relative consistency of the season from year to year from north Texas up into western Iowa is a natural, objective way to define "Tornado Alley". The concept of Tornado Alley may be very important for the emergency management community. It is relatively easy to keep awareness up in a region where events happen frequently and where the threat is confined to a relatively short period of time. In addition, it is typically easier to recruit volunteer storm spotters in such an area and to maintain their enthusiasm. For instance, public awareness was extremely high in the 3 May 1999 Oklahoma City tornado. Despite damaging almost 8000 structures, fewer than 40 direct fatalities occurred. In contrast, heightening awareness in an area where tornadoes rarely occur or occur over a broader season of the year is much more difficult.

We believe that the problem of public awareness in regions where the climatological threat of a tornado on any particular day is low is one reason for many of the high death toll events over the last 20 years. During that period of time, only two of the 22 tornadoes in the US that have caused at least 8 fatalities (representing the highest 10% of death tolls) in what we have defined objectively as "Tornado Alley" with this dataset (26 April 1991-Andover, Kansas, and 3 May 1999-Oklahoma City). One of those had its fatalities in a trailer park and the other was the (inflation-adjusted) biggest property damage tornado in US history.

Almost one-fourth of all significant tornadoes occur in this objective Tornado Alley, but only 9% of the major killers have. Thus, the vast majority of high fatality tornadoes in recent years have occurred in areas where tornadoes are an especially rare event on any given day. Eliminating those events will be extremely difficult, given the challenge of getting people to respond when their basic state of awareness is very low. As the recent Salt Lake City tornado reminded people, tornadoes can occur almost anywhere in the US and there is no reason to believe that if the atmosphere is capable of producing a tornado somewhere, it is also capable of producing a strong tornado. Rare events do occur. Preparing the public to be ready for them is a difficult task, but recognizing the nature of the threat has long-term potential rewards for emergency management and response and insurance interests.

6. ACKNOWLEDGMENTS The lead author participated in the Research Experiences for Undergraduates program at the Oklahoma Weather Center in the summer of 1999. We thank Tom Grazulis of the Tornado Project for providing us with the data and Mike Kay of the Storm Prediction Center for his assistance in gridding the data.

## 7. REFERENCES

Brooks, H. E., 1999: Severe thunderstorm climatological probabilities. URL: http://www.nssl.noaa.gov/hazard/

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

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

Rasmussen, E. N., and C. Crosbie, 1996: Tornado damage assessment in VORTEX-95. Preprints, *18th Conf. Severe Local Storms*, San Francisco, CA, Amer. Meteor. Soc., 153-157.

Schaefer, J. T., and H. E. Brooks, 2000: Convective storms and the impact. Preprints, *2nd Conf. Environ. Appl.*, Amer. Meteor. Soc., Long Beach, CA, this volume.

Silverman, B. W., 1986: *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, 175 pp.

Damages from violent tornadoes seem to be increasing, similar to the trend for other natural hazardsâ€”in part due to changing population, demographics, and more weather-sensitive infrastructureâ€”and some analysts indicate that losses of $1 billion or more from single tornado events are becoming more frequent.Â  Severe thunderstorms and tornadoes affect communities across the United States every year, causing fatalities, destroying property and crops, and disrupting businesses.Â  This report focuses on the risk from severe thunderstorms and tornadoes to the... â€œClimatological risk of strong and violent tornadoes in the United Statesâ€. Preprints, 2nd Conf. On Environmental Applications , Amer. Meteor. Soc., Long Beach, California. American Meteorological Society, 212-219. Doswell, C.A. III, and D.W. Burgess, 1988. â€œOn some issues of the United States tornado climatologyâ€.Â  â€œA HAZARD MODEL FOR TORNADO OCCURRENCE IN THE UNITED STATES 2005â€ Internet publication. Schaefer, J.T., D.L. Kelly, R.F. Abbey, 1986. â€œA minimum assumption tornado-hazard probability modelâ€. Journal of Climate and Applied Meteorology, 25, 1934-1945. Wilks, D.S., 1995: â€œStatistical Methods in the Atmospheric Sciencesâ€. Academic Press, Inc., 467 pp. SEE ALSO Climatological Risk of Strong and Violent Tornadoes in the United States (Paper 9.4, Second Conference on Environmental Applications)] * [http://www.nssl.noaa.gov/hazard/ Severe Thunderstorm and Tornado Climatology (NSSL)] * [http://www.tornadoproject.com/fscale/fscale.htm#top Fujita Scale]. Canadian sources.Â  Tornadoes of 2008 â€” Tracks of all United States tornadoes during 2008 Timespan January to December 2008 Maximum rated tornado â€¦ Wikipedia. Tornadoes of 2007 â€” Infobox Tornado year (EF scale) name = Tornadoes of 2007 caption = Tracks of all U.S. tornadoes in 2007. timespan = January to December 2007 max EF = EF5 max location = Greensburg, KansasElie, Manitoba (using Fujita scale) max date = May 4June 22 â€¦