Widen, H. M., J. B. Elsner, C. Amrine, R. B. Cruz, E. Fraza, L. Michaels, L. Migliorelli, B. Mulholland, M. Patterson, S. Strazzo, and G. Xing, 2013: Adjusted tornado probabilities. *Electronic J. Severe Storms Meteor.*, **8** (7), 1–12.

Electronic Journal of
SEVERE STORMS METEOROLOGY

Adjusted Tornado Probabilities

HOLLY M. WIDEN, JAMES B. ELSNER, CAMERON AMRINE, RIZALINO B. CRUZ, ERIK FRAZA, LAURA MICHAELS, LOURY MIGLIORELLI, BRENDAN MULHOLLAND, MICHAEL PATTERSON, SARAH STRAZZO, AND GUANG XING
Florida State University, Tallahassee, Florida

(Submitted 10 June 2013; in final form 4 December 2013)

ABSTRACT

Tornado occurrence rates computed from the available reports are biased low relative to the unknown true rates. To correct for this low bias, the authors demonstrate a method to estimate the annual probability of being struck by a tornado that uses the average report density estimated as a function of distance from nearest city/town center. The method is demonstrated on Kansas and then applied to 15 other tornado-prone states from Nebraska to Tennessee. States are ranked according to their adjusted tornado rate and comparisons are made with raw rates published elsewhere. The adjusted rates, expressed as return periods, are <1250 y for four states, including Alabama, Mississippi, Arkansas, and Oklahoma. The expected annual number of people exposed to tornadoes is highest for Illinois followed by Alabama and Indiana. For the four states with the highest tornado rates, exposure increases since 1980 are largest for Oklahoma (24%) and Alabama (23%).

1. Introduction

Reliable tornado hazard assessment is an important application of the tornado database. A tornado report reaches the database only if a manual observation of damage is made and verified. The precision on the genesis location is specified to two decimal places (latitude and longitude) until 2009 and to four decimal places afterwards. Technological changes, greater awareness of tornadoes, as well as more spotters and chasers have improved the probability that a tornado will be reported (Doswell et al. 1999; Verbout et al. 2006). Therefore, the number of reports in the historical database is a lower bound on the true number of tornadoes. In fact, the difference between the observed and true number of tornadoes is shrinking (Elsner et al. 2013). Since the tornado dataset is imprecise and inhomogeneous, hazard assessments need to account for these changes. If the raw numbers are used directly, the estimated risk of encountering a tornado will be too low.

The purpose of this paper is to illustrate a method for tornado hazard assessment that attempts to improve estimates of tornado risk. In short, the improvement is made by using a statistical model for report density as a function of distance from nearest city or town center (Elsner et al. 2013). The methodology produces a bias-corrected annual probability (rate) of being struck by a tornado. We demonstrate the procedure using tornado reports from Kansas first, and then apply the methodology to 15 other states. The rate estimates are made at the state level and comparisons are made with statewide raw rates published elsewhere. Since two tornadoes of different intensities traversing the same area can produce different damage paths, we repeat the analysis using strong and violent tornadoes.

In section 2, we provide a brief discussion of the tornado database. In section 3, we examine

Corresponding author address: Holly M. Widen, Florida State University, Tallahassee, 113 Collegiate Loop, Tallahassee, FL 32306, E-mail: hmr12b@my.fsu.edu

the path statistics of Kansas tornadoes as an illustration. In section 4, we describe the methodology used to estimate the risk of encountering a tornado. In section 5, we repeat the analysis for 15 additional states and rank them according to risk and exposure, based on statewide population. In section 6, we summarize the study and the conclusions and provide caveats to help improve the estimates. All the code used to generate the results of this paper (figures and tables) is available in PDF and CSV format.

2. Tornado data

NOAA's Storm Prediction Center (SPC) maintains the most reliable dataset of all reported tornadoes in the United States from 1 January 1950 to the present. The database originally was organized by the SPC (then known as the National Severe Storms Forecast Center) from newspaper accounts of all tornado reports (Kelly et al. 1978; Schaefer and Edwards 1999). Earlier records exist, but there has not been a consistent effort to investigate, document, or maintain a database of these earlier occurrences (Galway 1977; Grazulis 1993).

Data are compiled on each tornado's path length and width, formation and dissipation locations (latitude and longitude), intensity on the (enhanced) Fujita scale (EF scale), and other characteristics. The EF scale is a subjective rating system which assigns a category of intensity according to the amount, type, and appearance of tornado damage. Originally, the damage scale was related physically to the tornado wind speed (Fujita 1981). Currently, wind speed is phenomenologically related to the observed damage (Feuerstein et al. 2005). For instance, EF1 damage corresponds to wind speeds in the range of 38–49 m s^{-1} (peak 3-s) and EF5 damage corresponds to wind speeds between 89–105 m s^{-1} (derived EF scale). The EF scale replaced the F scale in February 2007 with slightly different and more specific criteria for assessment (Potter 2007; Edwards et al. 2013). The F scale and the EF scale are considered equivalent for climatological applications such as this one.

We download the dataset from http://www.spc.noaa.gov/gis/svrgis/ as a Geographic Information System (GIS) shapefile. Tornado path length is converted from miles to meters and tornado path width is converted from yards to meters. Tornado paths represent the full path and not state segments as used in Simmons and Sutter (2011). A description of the data attributes is available from http://www.spc.noaa.gov/wcm/SPC_severe_database_description.pdf. The shapefile is read by the software R (R Core Team 2013) as a spatial points data frame using the readOGR() function from the **rgdal** package (Bivand et al. 2013) for the R computing environment. The data have a Lambert conformal conic (LCC) projection with parallels of 33° and 45° N and a center longitude of 96° W. The projection uses the GRS80 ellipsoid.

3. Kansas tornadoes: 1950–2011

A boundary file for the state of Kansas is obtained from the U.S. Census Bureau. The boundary is projected using the LCC of the tornado database. Kansas reports are extracted from the database by including only tornadogenesis points contained within the boundary.

Here we illustrate the method using statewide tornado data, but the procedure can be applied to a dataset covering a smaller area (e.g., county warning area). Indeed our methodology may be more generally applicable to smaller areas where the assumption of spatial homogeneity is more tenable, although there is a tradeoff because of decreasing sample size.

a. Path length and width

The above extraction method results in 3713 Kansas tornado reports over the period 1950–2011. Table 1 shows the median path length (m) and width (m) as well as the number of tornadoes distributed by EF-scale category. A total of 266 reports have an unknown EF-scale category from earlier in the study period.

On average, the damage rating is higher for longer and wider tornadoes. The path width variable changed from an estimate of the average path width to the maximum path width in 1994. However, according to Brooks (2004), this change does not appear to significantly influence the overall statistics of path width in the database. Paths within the state from tornadoes that begin outside of the state are not included.

Table 1: Kansas tornado report statistics by EF scale category. The lengths and widths are the median values and the quartile values are subset below.

	All Counts (1950–2011)	Path Length (m) (.25 & .75 Quartiles)	Path Length (m) (.25 & .75 Quartiles)
EF0	1969	805 (161, 1609)	27 (16, 50)
EF1	862	3219 (483, 9656)	46 (9, 91)
EF2	419	4828 (483, 16093)	91 (9, 197)
EF3	159	16093 (8047, 35808)	274 (72, 640)
EF4	35	37015 (18017, 49093)	457 (229, 805)
EF5	6	60197 (38145, 76444)	1006 (503, 1207)
Unknown	266	161 (161, 161)	9 (9, 9)

However, this discrepancy is likely more or less compensated by including the complete path length of tornadoes that move out of the state.

The median path length for Kansas tornadoes is 853 m and the median path width is 46 m. On average, stronger tornadoes travel longer distances. Thus, the subset of strong tornadoes (EF2 or higher) has a median path length of 9576 m and a median width of 91 m. There are 41 violent tornadoes (EF4 and EF5) (1.1% of the total number of reports) in Kansas over the 62-y period. The median path length and width of these tornadoes is 38 624 m and 457 m, respectively. The correlation between path length and width is .343 (.314, .371) (95% confidence interval, CI) for the set of all tornado reports and .246 (.171, .319) (95% CI) for the set of strong tornadoes. For the set of violent tornadoes, the correlation is statistically insignificant at .209 (–.105, .486) (95% CI). The rank correlation between path area and EF scale is .462 for the set of all tornado reports, .452 for the set of strong tornadoes, and .408 for the set of violent tornadoes.

b. Spatial distributions

The spatial distribution of tornado reports across Kansas is shown by EF-scale thresholds in Fig. 1, where the points indicate genesis locations. The annual statewide density for all tornadoes is 2.84 reports per 10^4 km^2. For the set of strong tornadoes, the density is .47 reports per 10^4 km^2.

With the exception of the violent tornadoes (EF4 and EF5), there is no obvious spatial trend in the report densities. In other words, the spatial report density is approximately the same regardless of location. However, violent tornadoes appear to be relatively more probable over the eastern half of the state. The apparent absence of violent tornadoes in the western half could be due partly to a lack of damage indicators in areas without structures. On the other hand, the elevation is higher in western Kansas. This limits the amount of low-level moisture resulting in higher cloud bases on average. Development of near-ground rotation is inhibited by relatively colder downdrafts (through greater evaporation), thereby limiting convergence and upward accelerations (Markowski and Richardson 2013) needed for violent tornadoes. Despite the spatial uniformity in tornado reports in the cardinal directions, an overlay of city centers (Fig. 2) shows that reports appear more numerous in the vicinity of cities (Elsner et al. 2013). There are 871 Kansas cities in the U.S. cities database obtained from http://www.nws.noaa.gov/geodata/catalog/national/data/ci08au12.zip.

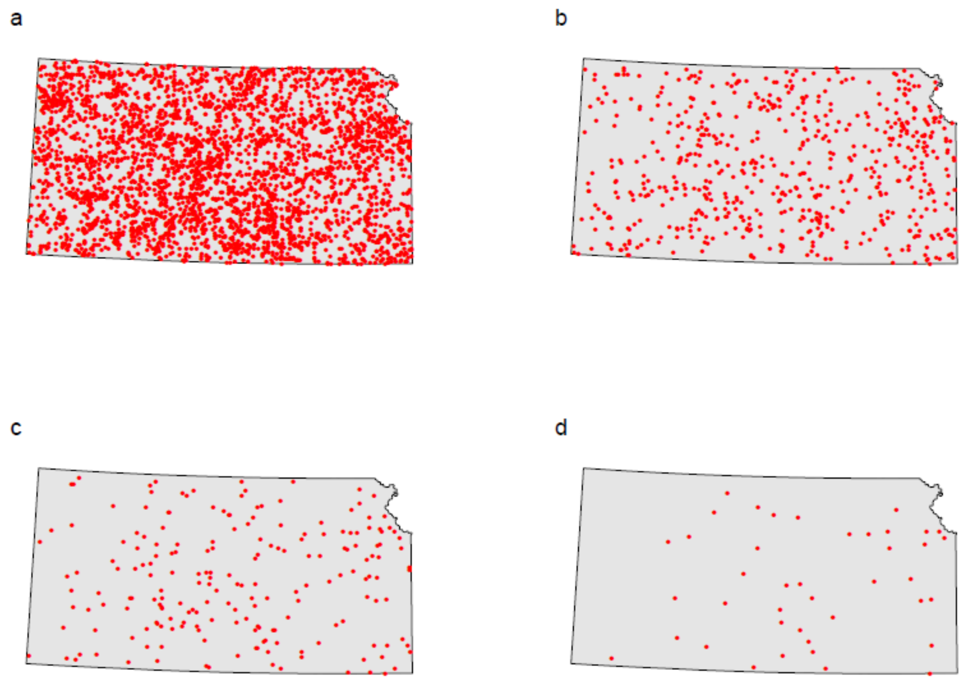

Figure 1: Kansas tornado reports over the period 1950–2011. Red points are the tornadogenesis locations for: a) all (includes those without an assigned category), b) EF2 and higher, c) EF3 and higher, and d) EF4 and EF5.

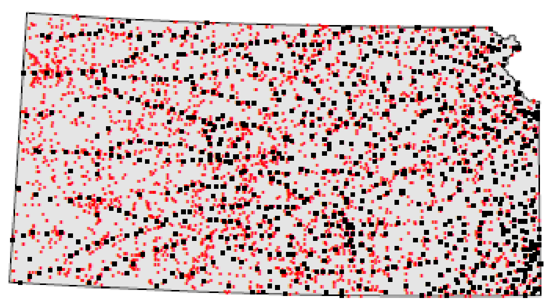

Figure 2: Kansas tornado reports from 1950–2011 (red) and city centers (black).

c. Distance from nearest city center

A statistical description of tornado report clusters near cities and towns is obtained by estimating the spatial report density as a smoothed function of distance from nearest city center. First, a 128 × 128 grid containing distances from the nearest city center is computed. Distances range from .04–33.7 km with a median of 7 km. Fifty percent of all grids are between 4.6–10.3 km from the nearest city.

Second, let $Z(u)$ be the distances from nearest city on grid u, then the estimated tornado report density in the grid is given by

$$\hat{\lambda}(u) = \hat{\rho}(Z(u)) \qquad (1)$$

where $\hat{\rho}(Z)$ is estimated using a kernel smoothing. This technique is implemented by applying the probability integral transform to the distance-from-nearest-city value (yielding values in the range 0–1), then applying edge-corrected density estimation on the interval [0, 1], and back-transforming (Baddeley and Turner 2006). The probability integral transform uses the empirical cumulative distribution function for the covariate Z $[P(Z(u) \leq z)]$ for a random selection of pixels). We set the bandwidth to be .25 standard deviations of the kernel, which is chosen through trial and error to obtain a smooth, monotonic relationship.

Figure 3 shows the annual report density in reports per 10^4 km^2 as a function of distance (km) from nearest city center. For all tornadoes (Fig. 3a), the smoothed density peaks at 4 reports at zero distance but drops to <2 reports at distances >15 km. Thus, although the statewide average density is 2.8 reports per 10^4 km^2, there are significantly more reports near population centers. More details on this procedure, including how the function is changing with time, are available in Elsner et al. (2013).

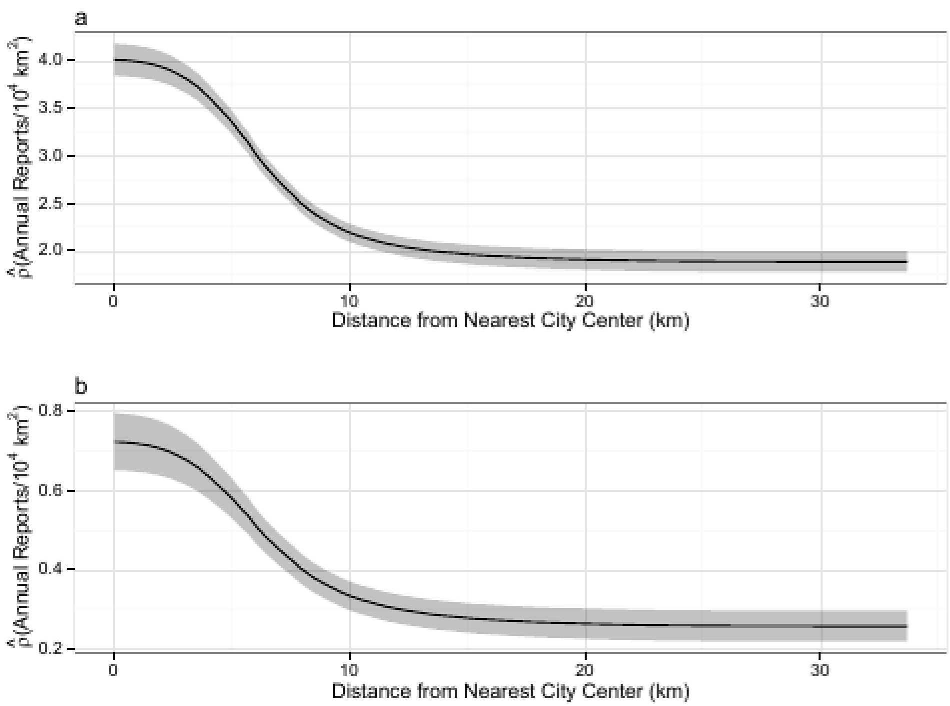

Figure 3: Tornado report density as a function of distance from nearest city center for: a) all tornadoes and b) strong and violent tornadoes (EF2 or higher). The gray band is the 95% Confidence Interval on the density estimate.

4. Risk of a tornado encounter

With the assumption of a uniform statewide tornado distribution, the probability of encountering some part of a tornado is obtained by adding the damage area (path length times path width) of each report and then dividing by the total area of the state (Thom 1963). The probability is expressed per annum as a result of dividing by the number of years in the database. Since the number of reports is a lower bound on the actual number of tornadoes, we multiply the total damage area by $\hat{\alpha}$, which is the average report density at the city center multiplied by the area of the state divided by the observed number of tornado reports.

Let P be the annual probability of a tornado striking any point in Kansas given by

$$P = \frac{A_c}{AY} \qquad (2)$$

where A_c is the corrected total area of tornado damage, A is the area of Kansas, and Y is the number of years in the database (62 y). The corrected total damage area is given by

$$A_c = \hat{\alpha} A_r \qquad (3)$$

where A_r is the total tornado area given by $\sum_{i=1}^{n} l_i w_i$ in which n is the number of tornado reports and l_i, w_i are the path length and width of tornado i, respectively. The coefficient $\hat{\alpha}$ is given as the ratio of $\hat{\rho}$ at distance zero to $\hat{\rho}$ at maximum distance. A value of $\hat{\alpha} = 1$ indicates no undercount.

The method results in a statewide hazard probability of .0661 (.0633, .0689)% (95% CI) per year. The estimate is the annual probability of getting hit by a tornado at any location in Kansas. The confidence interval is based on the standard error on $\hat{\rho}$. A direct comparison can be made by considering the raw annual probability given in Simmons and Sutter (2011, their Table 2.7). They estimate an annual probability for Kansas of .0329 based on the tornado database over the years 1950–2009, which is low by a factor of 2 relative to our adjusted estimate.

The return period is expressed as the inverse of the annual probability. Our method estimates a return period for Kansas of 1512 y. Thus, any location in the state can expect to be hit by a tornado once every 1512 y, on average.

5. Other states

The method can be applied to any tornado area. However, the assumption of a homogeneous spatial distribution of reports is untenable for states with pronounced variations in tornado-occurrence density such as Texas, where tornadoes are much more likely in the north than in the south. In contrast, Kansas tornado frequency can be described to a first order by a single-rate parameter. The rate of tornadoes does not vary significantly by compass direction; although on average, there are more tornado reports near towns and cities.

Table 2: Tornado report count per EF category per state. States are ranked by total number of reports.

State	Total	EF0	EF1	EF2	EF3	EF4	EF5	Unknown
Kansas	3713	1969	862	419	156	35	6	266
Oklahoma	3406	1372	1067	655	184	48	7	73
Nebraska	2583	1255	728	281	76	23	1	219
Iowa	2244	928	673	414	100	39	6	84
Illinois	2152	977	659	353	97	23	2	41
Missouri	1991	741	787	325	92	33	1	12
Mississippi	1861	524	736	385	129	24	4	59
Alabama	1811	576	678	385	123	35	7	7
Louisiana	1728	542	795	293	87	9	1	1
Arkansas	1626	451	602	389	142	27	0	15
Minnesota	1590	819	501	189	50	19	2	10
Indiana	1277	393	466	269	82	25	2	40
Wisconsin	1247	404	483	255	47	17	3	38
Tennessee	1049	434	225	82	25	1	0	282
Ohio	936	293	390	181	41	14	3	14
Kentucky	785	321	173	64	16	1	0	210

Here we consider 15 additional states where the assumption of homogeneity is also a good first-order approximation and repeat the analyses on the tornado reports in the database from these states. Table 2 shows the number of tornado reports by EF-scale category for each state.

Kansas stands out for the most tornado reports during this period with 3713 (previously mentioned), closely followed by Oklahoma with 3406 reports. Oklahoma has the highest number of strong tornado reports with 894 and Kansas is second with 616. Oklahoma also experiences the most violent tornadoes—55 during this study period. Iowa, Alabama, and Kansas trail closely behind with 45, 42 and 41 violent tornado reports, respectively.

The adjusted annual probabilities of a tornado strike for each state are shown in Table 3 and ordered by decreasing rate. Alabama leads the list with an annual probability of being struck at .098 (.092, .104)% (95% CI) followed by Mississippi with an annual probability of .097 (.091, .103)% (95% CI) and by Arkansas with an annual probability of .093 (.087, .099)% (95% CI). The return periods for the top four states are <1250 y, including 1020 y for Alabama, 1031 y for Mississippi, 1075 y for Arkansas, and 1235 y for Oklahoma. Kentucky has the lowest annual probability at .017 (.015, .019)% (95% CI) (return period of 5882 y) of the 16 states considered in this analysis.

Table 3 also lists the annual probability of a strong or violent tornado strike. Alabama also leads this ranking with an annual probability of .087 (.077, .096)% (95% CI) and Mississippi is second with an annual probability of .080 (.071, .089)% (95% CI). These are overestimates of the chance of EF-scale damage at any location because the EF-scale rating given to a tornado represents the worst damage somewhere along the path. The adjusted probabilities are considerably higher than the raw probabilities listed in Simmons and Sutter (2011) that are shown here.

The return period for a tornado strike at any location in Alabama using the raw probability is 3817 y or about 3.7 times longer. The adjusted probabilities are also correlated with the normalized statewide killer events listed in Ashley (2007). The resulting correlation is .739 (.385, .904) (95% CI) indicating a significantly strong positive relationship.

Table 3: Annual probability (Ann. Pr.) of a tornado strike. SS11 refers to Simmons and Sutter (2011).

State	All Ann. Pr. (%) (CI)	Strong/Violent Ann. Pr. (%) (CI)	All Ann. Pr. (%) (SS11)	Strong/Violent Ann. Pr. (%) (SS11)
Alabama	0.098 (0.092, 0.104)	0.087 (0.077, 0.096)	0.0262	0.0227
Mississippi	0.097 (0.091, 0.103)	0.080 (0.071, 0.089)	0.0415	0.0355
Arkansas	0.093 (0.087, 0.099)	0.080 (0.072, 0.089)	0.0410	0.0360
Oklahoma	0.081 (0.078, 0.084)	0.072 (0.066, 0.077)	0.0387	0.0326
Kansas	0.066 (0.063, 0.069)	0.059 (0.053, 0.064)	0.0329	0.0276
Indiana	0.065 (0.060, 0.069)	0.060 (0.052, 0.067)	0.0299	0.0272
Iowa	0.058 (0.055, 0.062)	0.052 (0.046, 0.057)	0.0326	0.0290
Louisiana	0.055 (0.052, 0.058)	0.042 (0.037, 0.047)	0.0180	0.0132
Tennessee	0.049 (0.046, 0.053)	0.038 (0.032, 0.043)	0.0236	0.0190
Nebraska	0.047 (0.045, 0.050)	0.042 (0.037, 0.048)	0.0232	0.0202
Illinois	0.041 (0.038, 0.043)	0.032 (0.029, 0.036)	0.0239	0.0188
Missouri	0.037 (0.035, 0.040)	0.029 (0.025, 0.032)	0.0176	0.0145
Wisconsin	0.035 (0.032, 0.037)	0.030 (0.026, 0.034)	0.0235	0.0200
Ohio	0.025 (0.023, 0.027)	0.020 (0.017, 0.023)	0.0161	0.0131
Minnesota	0.023 (0.021, 0.024)	0.019 (0.016, 0.022)	0.0129	0.0101
Kentucky	0.017 (0.015, 0.019)	0.013 (0.010, 0.015)	0.0120	0.0095

6. Exposure

Finally, we multiply the statewide tornado rate by the population to obtain an estimate of the expected number of persons exposed to tornadoes each year. This number has little intrinsic value as it assumes the population is uniformly distributed across the state. Moreover, the expectation is not useful for a highly skewed distribution. Nevertheless, it provides a useful metric of *relative* exposure that allows for a comparison between states.

The populations of each state from 1980 and 2010 are obtained from the U.S. Census Bureau, to calculate exposure and assess changes over time. Figure 4 is a slopegraph (Tufte 1983) displaying the states in order of tornado exposure, expressed as the number of people per year, in 1980 (left) and 2010 (right). Between the two columns is a sloped line demonstrating how their exposure has changed due to the fluctuation in population over the 20-y period.

Three separate groupings of exposure are apparent. The group demonstrating the highest exposure consists of Illinois, Alabama, and Indiana. Illinois leads the list with about 5207 people exposed annually by 2010. Alabama is second with 4662 people and Indiana is third with 4196 people. Kentucky, Nebraska, and Minnesota form the group exhibiting the lowest exposure with 736, 867 and 1206 people exposed annually, respectively. However, care must be exercised in interpreting the exposure values. For instance, Cook County, IL, which includes the city of Chicago, contains a large portion of the state's population. As such, the statewide annual exposure is overestimated for much of the state and underestimated in Cook County.

The percent change in exposure from 1980–2010 is calculated and can be observed in the sloped lines in Fig. 4. Each state demonstrates a positive change in tornado exposure. Tennessee exhibits the highest increase since 1980 (38%), followed by Minnesota and Arkansas (30% and 28%, respectively). Of the top 5 tornado-exposed states in 2010, Oklahoma and Alabama show the largest increase from 1980 (24% and 23%, respectively).

7. Summary and conclusions

Tornado hazard assessment is hampered by incomplete records and reporting practices that have improved with time. Estimates of tornado occurrence rates computed from the database of available reports will be biased low, relative to the true rate. Here we demonstrate a method to estimate the annual probability of getting hit by a tornado that uses the average tornado report density as a function of distance from nearest city or town, to correct statewide tornado probabilities.

The probability of encountering a tornado is obtained by adding an estimate of damage area (path length × path width) of each report, then dividing by the total area of the state. The total damage area is corrected by the ratio of the report density at the city center to the report density at maximum distance from the city, to account for underreporting in areas away from cities.

Results show Alabama with the highest annual probability of experiencing a tornado, as well as of experiencing a strong and violent tornado. Alabama is followed closely by Mississippi, Arkansas and Oklahoma, placing the highest rates in the south-central and southeastern parts of the country. Simmons and Sutter (2011) have four out of these five states at the top of their list based on raw probabilities, but in different order. Alabama, ranked 1st here, is ranked 7th by Simmons and Sutter (2011). More importantly, our corrected rate of experiencing a tornado is considerably higher than the raw rate. The top four states all have tornado return periods <1250 y. This information might be important for building-code requirements.

We multiplied the corrected rate by the state population to estimate the expected number of people exposed to a tornado strike per year. Although this statistic assumes a uniform population distribution, it is useful for comparing exposures between states. Illinois is the most exposed state by a considerable amount, although this is due to the large population of Chicago. Alabama and Oklahoma, two of the states with high annual probabilities, are among the top five exposed states. Indiana is the second most exposed state while Ohio and Tennessee alternate for the fifth making the Midwest one of the most exposed regions in the U. S. State populations from 1980 are used to make comparisons of exposure over time (1980–2010). Every state displays an increase over the 20-y period. The highest increases in exposure are in Tennessee, Minnesota, and Arkansas.

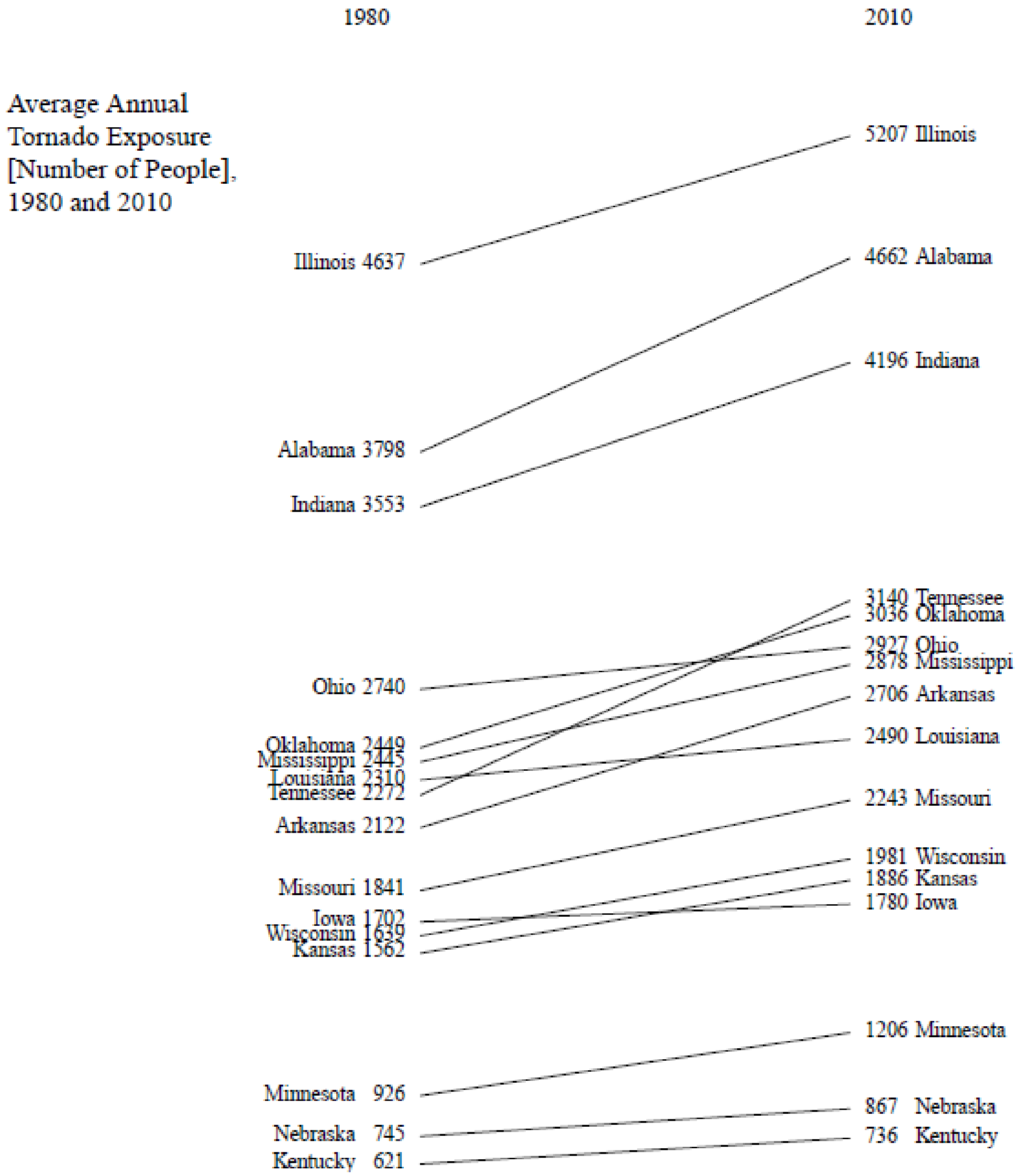

Figure 4: Slopegraph of statewide tornado exposure between 1980–2010. Exposure is the statewide population multiplied by the statewide adjusted tornado rate.

The study can be improved with a better estimate of path area than assuming a rectangle from path length and width. The study also can be improved by accounting for the variation in tornado strength within the path. This is especially relevant for rates of strong and violent tornadoes. Finally, we note that state tornado segments could be used instead of the complete tornado path. Although that would make the state rates more accurate, the assumption we used here can be applied to any region.

ACKNOWLEDGMENTS

We thank Thomas H. Jagger for reviewing an earlier draft of the manuscript. We thank the R project team, especially Roger Bivand, for the sp package. The slopegraph was created with R code obtained from Nathan Yau.

REFERENCES

Ashley, W. S., 2007: Spatial and temporal analysis of tornado fatalities in the United States: 1880–2005. *Wea. Forecasting*, **22**, 1214–1228.

Baddeley, A. and R. Turner, 2006: Modelling spatial point patterns in R. *Lecture Notes in Statistics*, Vol. 185, Springer-Verlag, 306 pp.

Bivand R., T. Keitt, and B. Rowlingson, cited 2013: Rgdal: Bindings for the Geospatial Data Abstraction Library. [Available online at http://CRAN.R-project.org/package=rgdal.]

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

Doswell, C. A., A. R. Moller, and H. E. Brooks, 1999: Storm spotting and public awareness since the first tornado forecasts of 1948. *Wea. Forecasting*, **14**, 544–557.

Edwards, R., J. G. LaDue, J. T. Ferree, K. Scharfenberg, C. Maier, and W. L. Coulbourne, 2013: Tornado intensity estimation: Past, present, and future. *Bull. Amer. Meteor. Soc.*, **94**, 641–653.

Elsner, J. B., L. E. Michaels, K. N. Scheitlin, and I. J. Elsner, 2013: The decreasing population bias in tornado reports across the central Plains. *Wea. Climate Soc.*, **5**, 221–232.

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

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

Galway, J. G., 1977: Some climatological aspects of tornado outbreaks. *Mon. Wea. Rev.*, **105**, 477–484.

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

Kelly, D., J. Schaefer, R. McNulty, C. Doswell III, and R. Abbey Jr., 1978: An augmented tornado climatology. *Mon. Wea. Rev.*, **106**, 1172–1183.

Markowski, P., and Y. Richardson, 2013: How to make a tornado. *Weatherwise*, **66** (4), 12–19.

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

R Core Team, cited 2013: R: A language and environment for statistical computing. R version 3.0.1 “Good Sport”. R Foundation for Statistical Computing. [Available online at http://www.R-project.org.]

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.

Simmons, K. and D. Sutter, 2011: *Economic and Societal Impacts of Tornadoes.* Amer. Meteor. Soc., 282 pp.

Thom, H., 1963: Tornado probabilities. *Mon. Wea. Rev.*, **91**, 730–736.

Tufte, E. R., 1983: *The Visual Display of Quantitative Information*. Vol. 914, Graphics Press, 197 pp.

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.

REVIEWER COMMENTS

[Authors' responses in *blue italics.*]

REVIEWER A (Harold E. Brooks):

Initial Review:

Recommendation: Accept with minor revisions.

General comments: The methodology and conclusions seem reasonable. I have only minor comments to offer. *[Editor's note: some reviewer comments appeared crucial enough to be included as substantive.]*

In addition to the technological changes, awareness, etc., the biggest factor in changes in tornado numbers is the increased emphasis on verification of warnings. Changes in verification and assessment practices mean that the reported intensity, length, and width need to be used with caution. Width is particularly problematic. The authors should note that the reported value changes from mean width to maximum width during the period they use the data. The fact that the date of change is not obvious when a time series of width data is created makes understanding the width data more difficult.

Yes. We now note that the width variable changed from an estimate of the average path width to the maximum path width in 1994. According to Brooks (2004) this change does not appear to significantly influence the overall statistics of path width in database.

[Re:] assumption of uniform population. I'd like to see a little more discussion on this, particularly given the prominence of Illinois in the table. Given that a large fraction of the population of the state lives in a corner (Cook County has ~40% of the population) and there reasonably may be an expectation of gradients of tornado occurrence (Cook County had ~1% of the total Illinois county reports of tornadoes in the SPC database), it's quite possible that the Illinois exposure is vastly overestimated. I'd just like to see more caveats.

Yes. This is a good point. We now state that care must be exercised in interpreting the exposure values. For instance, Cook County, Illinois, which includes the city of Chicago, contains a large portion of the state's population. As such the statewide annual exposure is overestimated for much of the state and underestimated in Cook County.

Note: Over the period 1950–2011, Cook County, IL had 45 tornado reports (2% of all Illinois tornado reports over this period). This amounts to 0.018 per km^2 per 62 ys. The median value over all 102 counties in the state is 0.012 per km^2 per 62 years with an interquartile range of .007 per km^2 per 62 y.

[Minor comments omitted…]

REVIEWER B (Patrick T. Marsh):

Initial Review:

Recommendation: Accept with minor revisions.

General comments: Generally speaking, I'm extremely pleased with the quality of this paper. It is straightforward, concise, and free of obvious deficiencies. My comments are of the minor variety: a handful of questions and/or clarifications. On the whole, most of my comments could be classified as "nitpicky". I would also like to commend the authors for publishing the code used for their analyses. As a huge proponent of reproducible research, I cannot state how refreshing this is.

Thank you. We agree that it is important for scientific research to be reproducible.

Substantive comments: At the end of paragraph 1, the authors give some statistics from the Joplin tornado. In particular, they state that the path length was 10 kilometers and the fatality count was 162. Although the authors give a citation for this information, the official SPC tornado database (which the authors use for their analyses) gives a path length of 21.62 miles (34.79 kilometers) and a fatality count of 158. I would suggest that the authors use these values as they are what are in the historical record.

Fixed in the revision. Thank you.

At the start of paragraph 2, the authors state, "Reliable tornado hazard assessment is thus an important application of the tornado database." Although I believe I understand what the authors are intending, however, how "reliable" for hazard assessment is a dataset that only records the starting and ending points of tornadoes, and does so only to two decimal places (yield a precision of only 1.1 km). Additionally, I fail to see how what the authors have stated up to this point illustrate "thus"ly that this is an application of the tornado database.

Yes, thank you. The revised manuscript now reads: "The precision on the touchdown location is specified to two decimal places (latitude and longitude) until 2009 and to four decimal places afterwards. Although the tornado dataset is imprecise, the need for a reliable assessment remains."

In the very next sentence, the authors use the phrase "manual observation" as a criterion for inclusion into the official tornado database. I interpret this statement to mean that a tornado must be seen, which is not a requirement.

Yes, thanks. We changed the phrase to "manual observation of damage".

I am curious as to why the authors chose to use the state of Kansas to illustrate their technique. Is there a technical reason or was it merely because, "you have to start somewhere"?

More or less the latter. However, Kansas is historically known for tornado activity and thus seemed like a good place to start.

In the last paragraph of section 3.2, the authors state "However, violent tornadoes appear to be relatively more probable over the eastern half of the state." Would the authors care to speculate as to why this is? My guess is that this has to do with population density, as the number of damage indicators for (E)F5 tornadoes (generally) requires the tornado to traverse an area with substantial population, rather than anything meteorological.

We agree but also think meteorology could play a role. We added [clarifying text] to the manuscript on this discussion. *[Large block quote omitted...]*

At the end of page 4, the authors give their bandwidth as 0.25 standard deviations. Would the authors care to give a reason as to why this value was chosen?

Yes, thanks. Our reasoning is made clearer in the revision which now states: "We set the bandwidth to be .25 standard deviations of the kernel which was chosen through trial and error in order to obtain a smooth, monotonic relationship."

Could the authors provide a more detailed figure caption for figure 3? The first time this figure is mentioned in the text, the authors only mention the figure in general, not figure 3a. As such when I looked at the figure I was confused as to what was different between the two figures. I eventually saw the answer on the next page, but still feel that this could be better conveyed in the figure caption.

Yes, thank you. We added detail to the caption of Figure 3 and mentioned Figure 3a specifically in the associated text in the revision.

[Minor comments omitted...]