The Decreasing Population Bias in Tornado Reports across the Central Plains

JAMES B. ELSNER AND LAURA E. MICHAELS
The Florida State University, Tallahassee, Florida

KELSEY N. SCHEITLIN
Longwood University, Farmville, Virginia

IAN J. ELSNER
University of Florida, Gainesville, Florida

(Manuscript received 22 August 2012, in final form 19 February 2013)

ABSTRACT

Tornado–hazard assessment is hampered by a population bias in the available data. Here, the authors demonstrate a way to statistically quantify this bias using the ratio of city to country report densities. The expected report densities come from a model of the number of reports as a function of distance from the nearest city center. On average since 1950, reports near cities with populations of at least 1000 in a 5.5° latitude × 5.5° longitude region centered on Russell, Kansas, exceed those in the country by 70% [54%, 84%; 95% confidence interval (CI)]. The model is applied to 10-yr moving windows to show that the percentage is decreasing with time. Over the most recent period (2002–11), the tornado report density in the city is slightly fewer than 3 reports (100 km$^2$)$^{-1}$ (100 yr)$^{-1}$, and this value is statistically indistinguishable from the report density in the country. On average, the population bias is less pronounced for Fujita (F) scale F0 tornadoes, but the bias disappears more quickly over time for the F1 and stronger tornadoes. The authors show evidence that this decline could be related in part to an increase in the number of storm chasers. The population-bias model can enhance the usefulness of the Storm Prediction Center’s tornado database and help create more meaningful spatial climatologies.

1. Introduction

An important application of tornado climatology is tornado–hazard assessment. Reliable assessment requires records over several decades or more. But a tornado record depends on an observer making a report and on an official report being entered into the database. Over the long run, this introduces a bias whereby reports outside of towns and cities tend to be less numerous. An additional complication arises from the fact that public awareness of tornadoes was low before 1990 (Doswell et al. 1999). Tornado–hazard assessments are hampered by this changing population bias in the available data.

The population bias is well known (Snider 1977; Doswell et al. 1999). Methods for identifying, quantifying, and modeling it have been proposed (King 1997; Ray et al. 2003; Anderson et al. 2007). These studies use climatologically uniform regions and estimate tornado frequency in a subset of the region likely to have the most accurate reports. The reliable frequency is then compared with the frequency elsewhere. Sometimes the frequency elsewhere is adjusted using a statistical or empirical model. Tornado reports are typically aggregated into areal units, and some models include a term to account for spatial autocorrelation. All previous studies assume the bias is constant over time.

Here we demonstrate a statistical model for the population bias that does not use spatial aggregations. The method allows us to quantify the changing nature of the bias and to speculate on the cause. The study is different from previous research on this topic in two important ways. First, we analyze the touchdown locations as point data rather than as areal aggregations (counties, grids, or circular domains). Second, we quantify the change in the population bias over time. The research is the first of its
kind to study the changing nature of the population bias in the occurrence of tornado reports.

The paper is outlined as follows: In section 2 we present the tornado data used and our partition of it, focusing on the months of May and June over the central plains. In section 3 we model an inverse relationship between distance from the nearest city and tornado report density. In section 4 we show the population bias decreases with time as the number of tornado reports in country areas has risen substantially, so that the spatial density of reports over the period 2002–11 is statistically indistinguishable from the density of reports in the vicinity of cities. In section 5 we argue that an increase in the number of storm chasers could be, at least in part, related to the decrease in the population bias. In section 6 we provide a summary and some concluding remarks.

2. Tornado data

The Storm Prediction Center (SPC) maintains a dataset of all reported tornadoes in the United States from 1 January 1950 to the present. Earlier records exist, but there has not been a consistent effort to investigate, document, or maintain a record of these earlier occurrences (Galway 1977). The SPC dataset is the most reliable archive available for tornado studies. We download the dataset from the SPC’s website (http://www.spc.noaa.gov/gis/svrgis/). At the time of download, the number of tornado reports was 56,221.

a. Study domain

For this study we consider tornado reports within a region centered on Russell, Kansas, during the months of May and June. The region’s domain is selected by first geocoding the city name and then downloading a map from Google Maps with a zoom of 7 (0 is for the entire world and 21 is for individual buildings) centered on the city.

The map domain is bounded by 36.10° and 41.57°N latitudes and 102.37° and 95.34°W longitudes, and it encompasses the central plains from northern Texas to central Nebraska. This is the region and time of year with the most tornadoes (see Fig. 1), and there is no large spatial gradient in occurrence. It also corresponds to when and where storm chasers roam looking for tornadoes. If there is a population bias, then this is the region and time of year where we would most likely detect it.

Figure 2 shows the touchdown locations of the 3879 tornadoes used in this study. Tornadoes occur throughout the region, and there appear to be areas where reports are concentrated. For instance, there is a group of touchdowns near Great Bend, Kansas, and Grand Island, Nebraska. On the other hand, fewer tornadoes have been reported in southeastern Nebraska.

Increasing public awareness of tornadoes has led to an overall increase in the number of reported tornadoes. Figure 3 plots the reports by Fujita (F) scale and year. The F scale is a rating system that categorizes tornadoes according to the amount, type, and appearance of tornado damage. F4 and F5 tornadoes are too rare to show meaningfully as an annual time series. It appears that many of the smaller tornadoes were not reported in earlier years, creating a noteworthy upward trend over time. The upward trend calls into question the homogeneity of the dataset. Moreover, it stands to reason that the trend is likely most pronounced in remote areas. The question that motivates the present work is, are we now catching them all? It should be noted that we make no attempt here to examine issues associated with quantifying tornado intensity through damage estimates.

b. Touchdowns as a spatial point pattern

We consider the touchdown locations as events in a point pattern dataset (Diggle 1983) and use functions from the spatstat package of R (Baddeley and Turner 2005). We obtain the tornado reports from the SPC as shapefiles. The points are on a Lambert conformal conic (LCC) planar projection with secant parallels of 33° and 45°N and a meridian of 96°W longitude and with a North American Datum of 1983 (NAD83).
There are a total of 3879 touchdown reports during May and June over an area of 379,759.4 km², resulting in an occurrence rate of 1.02 reports (100 km²)⁻¹ over the 62-yr period from 1950 to 2011. The area covers the region from central Nebraska in the north to central Oklahoma in the south to extreme eastern Colorado in the west to eastern Kansas in the east.

3. Are rural tornado reports less numerous?

The occurrence rate of reports computed above represents the mean spatial density over the entire study area, but the apparent clustering of touchdowns in Fig. 2 suggests reports are more numerous near cities. We examine this hypothesis using a statistical model. We begin with an overlay of cities on a map of the spatial-varying tornado density estimates.

a. Report density and city centers

Local tornado densities are estimated using a Gaussian smoother with edge correction. The local rates are computed on a 128 × 128 grid of pixels by a convolution of the isotropic Gaussian kernel with point masses at each of the tornado report locations.

We obtain city locations across the United States from online (at http://www.nws.noaa.gov/geodata/catalog/national/html/cities.htm). We subset these for our study area, and for populations exceeding 1000 residents based on the 1990 census. We find 313 cities and towns with populations ranging from 1001 in Arapohoe, Nebraska, to 367,302 in Tulsa, Oklahoma. We project the locations using the same LCC projection as the tornado reports. Figure 4 shows the city locations on top of the tornado densities.

A town or city does not guarantee a concentration of tornado reports, but in regions without a city or town, there are no significant report concentrations. This fact (if a tornado occurs, then its detection is more likely near a city, but a city does not make it more likely that a tornado will occur) limits the utility of correlation statistics for studying the population bias.
b. Model of touchdown density as a function of distance from city center

To quantify the relationship between tornado report density and distance from city, we use a statistical model. We first create a map consisting of distances from the nearest city for points on a $128 \times 128$ grid of pixels. The distance map is shown in Fig. 5. Most pixels are less than 30 km from the nearest city with a median distance of 17 km. Fifty percent of all pixels are between 11 and 24 km from the nearest city.

Next, we model the mean report density as a smoothly varying function of the distance from the nearest city center. Let $Z(u)$ be our distances on grid $u$ and then the model is given by

$$\lambda(u) = \rho[Z(u)],$$

where $\rho(z)$ is estimated using a kernel smoothing, implemented by applying the probability integral transform to the distance-from-nearest-city value, yielding values in the range of 0 to 1, then applying edge-corrected density estimation on the interval $[0, 1]$, and back-transforming (Baddeley and Turner 2005). 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 0.25 standard deviations of the kernel to ensure a smooth relationship. This approach smoothes the relationship between reports and distance from the nearest city rather than smoothing the reports like in Brooks et al. (2003) using space–time grids and in Dixon et al. (2011) using kernels. Our approach has the advantage of working directly with report locations and providing a level of uncertainty on the model parameters. A larger bandwidth increases the smoothing.

Results from the model are shown in Fig. 6, where $\rho$ is plotted as a function of distance from the nearest city. This is a statistical model indicating the average report density as a function of distance from the nearest city. Report density peaks above 1.35 reports ($100 \text{ km}^2$)$^{-1}$ at zero distances with a slow and then rapid decrease at greater distances. Report densities saturate at 0.8 reports ($100 \text{ km}^2$)$^{-1}$ at the farthest distances. The disparity between tornado report density at zero distance and the report density at maximum distance provides a quantitative description of the population bias in the tornado database.

For comparison we randomly relocate the city centers across the study area and rerun the model. Results are shown in Fig. 7. As expected, unlike the model using distance from the nearest city center, there is no significant variation in report density as a function of distance from the nearest random location. The report density is nearly constant at the study area rate of 1.02 reports ($100 \text{ km}^2$)$^{-1}$ over the 62-yr period.
4. Changing population bias

Figure 6 is based on all years in the dataset. By taking the ratio of the report density at zero distance (city) to the report density at maximum distance (country) we can say that, on average, over the 62 yr from 1950 to 2011 reports near cities and towns exceed those in the country by 70% with a 95% confidence interval on this percentage of 54%–87%.

a. By decade

Here, we examine the changing nature of this percentage. Figure 8 shows results from the model run on 42 consecutive 10-yr moving intervals beginning with the period 1961–70. The 95% confidence band is shown in each panel. Variations from one decade to the next are slight, but overall there is a tendency over time for the report density in the country to match the report density in the city (the line is becoming more horizontal). There are clear differences in the curve over time, with the most recent years showing a significantly flatter curve compared with the early decades. Indeed, during the first decade of the twenty-first century, there is no significant difference between the city and country report densities, as can be seen by imagining a horizontal line that stays completely within the confidence band.

Frequently, the reporting bias in the tornado record is removed by detrending the time series. Figure 9 shows time series of the report density in the city and the country using the same 10-yr moving intervals. Again, the city report density is taken at zero distance and the country report density is taken at the farthest distance.

FIG. 4. Tornado report density and city centers for cities with a 1990 census population exceeding 1000 in a 5.5° latitude × 5.5° longitude region centered on Russell, KS. Tornado report densities are computed on a 128 × 128 grid of pixels.

FIG. 5. Distance map showing the distances to the nearest city center. Distances are computed on a 128 × 128 grid of pixels. Region encompasses a 5.5° latitude × 5.5° longitude area centered on Russell, KS.

FIG. 6. Tornado report density as a function of distance from the nearest city center. Model is for tornado reports in a 5.5° latitude × 5.5° longitude region centered on Russell, KS.
There are increasing trends for city and country report densities. The trend in the city densities levels off starting in the mid-1990s, while the trend in the country densities continues upward. In fact, the country density statistically matches the city density in the most recent decade (2002–11). Differences between city and country tornado reporting trends should be taken into account when attempting to remove reporting biases.

b. City-versus-country report density

The model allows us to define the ratio of report density in the city to that in the country as the population bias. Specifically, we take the report density estimate at zero distance and divide it by the density estimate at the farthest distances. This ratio of city to country report density varies from one decade to the next, but it should show a decreasing trend.

Figure 10 shows this ratio as a percentage by which tornado reports near cities and towns exceed that in the countryside with uncertainty lines on the estimates. The decreasing trend in population bias is substantial. During the early decades, the percentage by which tornado reports in cities exceeded those in remote areas is about 150%. By the mid-1990s, it had leveled out to about 2 to 1 in favor of city reports; however, during the twenty-first century the decline resumed, so that during the most recent decades the percentage is not significantly above zero.

We divide the tornado reports into F0, F1, and F2 and higher categories and repeat the analysis on the three subsets of reports. Results using all years in the database are summarized in Table 1. The uncertainty on the percentage is quite large, especially for the stronger tornadoes, where the number of reports is lower. Interestingly, the weak tornadoes show a smaller population bias. This finding is consistent with Anderson et al. (2007), who show that F0 and F1 tornado reports in Oklahoma vary less with population density than do the F2–F5 reports. They suggest that this might be a consequence of some strong tornadoes being misclassified as weak ones, especially early in the record, but we find a relatively smaller population bias for the weaker tornadoes, even in the most recent decade of reports. Another interpretation of this finding is that some F0 tornadoes go unreported even in the vicinity of towns, whereas the probability of nondetection of a strong tornado in a population center is relatively lower. If this is the case, then the percentage difference between the city and country densities will be smaller for F0 tornadoes.

Figures 11–13 show the percentages by decade for the three subsets of tornado reports. It is clear that the population bias is decreasing for all three categories (F0, F1, and F2+) and that during the most recent decades, the difference between the report densities near and far from cities is about zero given the uncertainty on the model estimates. The population bias decreases more quickly for the stronger tornadoes with little, if any, bias since the early 1990s for the F2 and stronger tornadoes.

c. Sensitivity to population threshold

Here, we examine the sensitivity of the results to the population threshold. Table 2 shows the results from the model using data over all years and all F-scale categories and for three different population thresholds, including cities with at least 500, 1000, and 2000 people. The population bias is highest with the lowest population threshold. This makes sense, since with a lower threshold there are more cities and the relationship captures the bias locally. Higher report densities in the vicinity of the smaller towns that are not included result in a less pronounced population bias.

City population values are available from the 1990 and 2000 decennial censuses. Comparing the list of cities with populations exceeding 1000, we find that 7 of the 313 cities that met the 1000-person threshold in 1990 did not in 2000. However, an additional nine towns surpassed the threshold for a net pickup of two towns. We rerun the model on the 315 cities and towns using the 2000 population values and find no statistically significant difference in the population bias.

5. More storm chasers

Here we speculate on one factor that might be related to the twenty-first century decline in the tornado report population bias. In particular, we focus on the increasing
surveillance of tornadoes by storm spotters and chasers. Storm chasing has been around for 50 yr or more, but public awareness of chasing increased dramatically at the time of the theatrical release of *Twister* in May of 1996 (Robertson 1999). The Discovery Channel’s reality series *Storm Chasers* beginning in 2007 has further tracked the popularity of storm chasing. Once a scientific endeavor, storm chasing has become a tourist business in the plains and an option for risk recreation (Cantillon and Bristow 2001). Could an increasing number of chasers roaming Tornado Alley each year help to explain, at least in part, the decrease in the population bias of tornado reports?

While there are no hard data on the number of storm chasers, it is possible to use proxy data to track the surge in popularity. Here, we obtain data from online (http://www.news.google.com) the number of Google News pages mentioning storm chasers. We use the advanced

---

**Fig. 8.** Tornado report density as a function of distance from the nearest city in 10-yr moving intervals. Model is fit to the tornado reports in a 5.5° latitude × 5.5° longitude region centered on Russell, KS.
search options and under the This Exact Phrase option type the search term “storm chaser” for specified dates. The same news story covered by multiple sources is counted only once. Although public awareness and interest in storm chasing might not be a good proxy for the actual number of chasers, it is possible that there is some correlation. Anecdotally, increasing public attention to the hazard caused by ‘‘chaser convergence’’ suggests an increasing number of chasers.

Figure 14 shows the city and country density of reports in 10-yr intervals as a function of a 10-yr moving average of the number of news pages mentioning storm chasers. The relationship is excellent for the country densities but poor for city densities. We repeat the analysis using the number of newspaper articles that mention the phrase ‘‘storm chaser’’ or ‘‘tornado chaser’’ on NewspaperArchive.com and find nearly identical results.

In addition to a likely increase in storm chase activity, today’s chasers might be more successful in navigating storms and observing and documenting tornadoes than those in the past. Improvements in technology, including mobile Internet and GPS, might make chasers more successful at reporting a tornado. The possible increase in tornado observations by storm chasers and spotters in the remote countryside might be contributing to the decline in population bias of tornado reports.

The hypothesis predicts that in regions without storm chasers, a decline in the population bias would be minimal. We repeat the analysis using tornado touchdown points and city locations across similar sized areas (5.5° latitude/longitude) centered on Terre Haute, Indiana (Midwest), and Armory, Mississippi (South), and find no significant decline in the population bias since 1990, consistent with our prediction. Although we have no data on the changing numbers of storm chasers in these regions, we would speculate that there has not been a marked increase. This finding supports our hypothesis and argues for more research on this topic.

6. Summary and concluding remarks

Population bias, defined here as the ratio between the in-town (city) and out-of-town (country) report densities, hampers efforts to construct meaningful climatologies or to provide accurate hazard assessments from tornado databases. Past efforts have focused on analyzing this bias by spatially aggregating tornado reports and population. Here, we present a fundamentally different approach grounded in the theory of spatial statistics. The method uses a statistical model for report density as a function of distance from the nearest city center. The relationship is examined in aggregate and over 10-yr moving intervals using a 5.5° latitude × 5.5° longitude region centered on Russell, Kansas.

<table>
<thead>
<tr>
<th>F scale</th>
<th>Percentage (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0</td>
<td>60.3 (40.8, 82.9)</td>
</tr>
<tr>
<td>F1</td>
<td>84.1 (52.1, 123.8)</td>
</tr>
<tr>
<td>F2+</td>
<td>85.2 (45.1, 138.0)</td>
</tr>
</tbody>
</table>
Our principal findings include:

- Historically, the number of reported tornadoes across the premiere storm chase region of the central plains is lowest in the countryside.
- The number of tornado reports in the countryside has increased dramatically since the 1970s, but especially since 1996.
- During the period from 2002 to 2011, tornado report density in the countryside is statistically indistinguishable from the report density in the cities, at slightly less than 3 reports $(100 \text{ km}^2)^{-1} (100 \text{ yr})^{-1}$.
- On average this population bias is less pronounced for F0 tornadoes, but the bias disappears more quickly over time for the F1 and stronger tornadoes.

We speculate that an increase in the number of storm chasers roaming the central plains during tornado season might help explain the increase in tornado reports. Advances in reporting technology might multiply this
The hypothesis is supported by the fact that we do not find a decline in the population bias over similar sized areas in the South or Midwest, where storm chasing is less prevalent.

Our statistical method of report density as a function of distance from the nearest city could help researchers develop more realistic spatial tornado climatologies. The methodology can also be applied to test hypotheses concerning the influence of other covariates, such as the distance from the nearest radar site and natural features such as lakes and hills. It can also be used to estimate site-specific population adjustments to the historical database for things such as nuclear power plants and wind farms, thereby increasing the usefulness of the SPC tornado database. As a next step, we suggest the method be applied over this same region during months outside the peak season and elsewhere where tornado risk is relatively uniform. All the code used in this study is available online (at http://rpubs.com/jelsner/1223).

The study can be improved by considering the influence tornado outbreaks have on the results. Further,
the study could be expanded to examine the influence of city size on the results. In this regard it would be interesting to try two covariates to examine the population bias in tornado reports near a small town (e.g., fewer than 5000 residents) relative to the population bias in a large city. New work will focus on using the spatial distribution of nonviolent tornadoes to predict the distribution of the rare violent tornadoes after accounting for the population bias.

Finally, our results demonstrate that there are valid things that can be said about tornadoes (or other extreme events) even if the sample of reports is biased. We have shown here that the sample of tornado reports is biased in a systematic way that we understand and can be corrected. In contradistinction, problems occur when we know our sample is biased, but we fail to

<table>
<thead>
<tr>
<th>Population threshold</th>
<th>Number of cities</th>
<th>Percentage (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
account for it or fail to account for it in a proper quantitative way.

Acknowledgments. The Department of Geography at The Florida State University provided some financial support for this research.

REFERENCES