A geographical information system (GIS) perspective is taken to examine conceptual and methodological complications present in tornado density and probability mapping. Tornado density is defined as the inverse-distance-weighted count of tornado touchdown points or tornado-affected cells within a neighborhood area. The paper first adds a few geographic elements into the tornado definition and then characterizes tornado density as a density field in GIS that depends on predefined, modifiable areas to exist. Tornado density is therefore conceptually distinguished from both individual tornadoes and tornado probability. Three factors are identified to be vital in tornado density mapping: the neighborhood size, the distance decay function, and the choice of tornado properties. Correspondingly, 12 neighborhood sizes ranging from 20 to 360 km are tested, four distance decay functions are compared, and two tornado properties—tornado touchdown locations and pathlengths—are separately incorporated in mapping. GIS interpretations, clarifications, and demonstrations are provided for these factors to reach a thorough understanding of how the factors function and affect the resultant tornado density maps. Historical tornado data of the eastern half of the United States from 1973 to 2013 are used in these demonstrations. Uncertainty and propagation analyses are recommended for future tornado density and probability mapping, and a Monte Carlo simulation using tornado pathlength data is conducted as an example of uncertainty modeling. In all, tornado density mapping is diagnosed as a largely subjective activity, and the mapper needs to make multiple choices according to the mapping purpose, scale, and the involved tornado record data.
tornado touchdown points or tornado-affected cells (when pathlength data are used). Though not addressed directly in this paper, we believe the density values derived from the past records of tornadoes have implications for future tornado risk or probability.

We examine how GIS factors function in the process of tornado density mapping, in hopes that better density maps can be produced in the future by wisely manipulating these factors. The next section presents GIS definitions of tornado and tornado density as the conceptual support for tornado density mapping. The third section demonstrates a few GIS complications and their effects on tornado density mapping. Conclusions and suggestions are given in the last section.

2. Tornado and tornado density viewed in GIS

a. A GIS definition of tornado and its representation

Tornado density mapping is a GIS activity even though it may be conducted by a climatologist, as geographic space plays a vital role in these efforts. As a result, this mapping activity needs conceptual guidance regarding the questions “what is a tornado?” (see American Meteorological Society 2015) and “what is a tornado in GIS?” To accomplish this, we add the following elements in the tornado definition:

1) a tornado is a dynamic event in space–time (Yuan 1998) with a rather discrete geographic boundary and life span;
2) a tornado is truly three-dimensional in geographic space in that it covers a land surface area with horizontal shape and size and it has a complete vertical dimension; and
3) a tornado has its own geometric (e.g., pathlength, width, area, and direction), thematic (e.g., intensity, wind speed), and temporal (e.g., occurring time and speed/pattern of traveling) properties (Schaefer et al. 1986) that can be tied both to the geographic space–time (e.g., wind speed at a certain space–time point during the tornado) and to the tornado itself (e.g., intensity of a tornado).

The above definition allows the GIS representation of tornadoes to be evaluated more completely. Specifically, when individual tornadoes are conceptualized as discrete objects in geographic space [see Bian (2007) and Goodchild et al. (2007) for a definition of “objects” in GIS], they are usually mapped as various vector features: touchdown points, track polylines, or damage area polygons. Their third or vertical dimension is usually ignored, at least for the purpose of tornado density mapping. The temporal dimension is either ignored (such as in this study) or greatly aggregated (e.g., Concannon et al. 2000). Occasionally, an individual tornado (or tornado area) may also be conceptualized as a continuous field (Goodchild et al. 2007) and mapped using raster [e.g., see Fig. 4 of Standohar-Alfano and van de Lindt (2014)] or isoline (e.g., Yuan et al. 2002) data models. Tornado properties can be linked to any of the above spatial representations to serve various tornado mapping purposes.

According to the above tornado definition, all tornado density maps are simplifications of the real-world “tornado” phenomena. Users’ choices often vary (and vary in quality as well) with regard to how to simplify tornadoes into tornado density maps. However, it is difficult to assess which of the above-listed tornado simplifications or representations is absolutely “better,” because the assessment has to be linked to the actual mapping situation and the application purpose. For instance, to evaluate tornado hazards of “places” across the United States, it would be useful to map tornadoes to (named) counties and states (Boruff et al. 2003), but to identify the tornado alley, it would be better to map tornado density continuously using (small enough) grid cells or isolines (e.g., Dixon et al. 2011).

b. A GIS interpretation of tornado density mapping

When examined in GIS, differences between the two concepts of “tornado” and “tornado density” are essential. Tornado density is a typical density field. To calculate and report a density value at a location, we must first specify an area (e.g., a county, a circle, or a rectangular cell) surrounding this location. The adjustment of the area in shape and size would result in a different density reading for the same location. Therefore, all tornado density values are relative (to the choice of the area) instead of absolute. In contrast, the existence of tornadoes and most other weather phenomena is fully independent of human definitions. A value on a density field usually describes a two-dimensional geographic area rather than an individual point (Goodchild et al. 2007), even though the value may be assigned to a point (e.g., central point) within the area. In this sense, tornado density is also different from many climatic variables that are ordinary fields in GIS, such as mean air temperature. Ordinary fields are strictly properties of individual points in geographic space and have nothing to do with the surrounding areas of these points (even with the existence of spatial autocorrelation). Tornado probability should be viewed as an ordinary field rather than a density field as well, since it is a property, unknown as it is, of individual locations (i.e., points), even though tornado probability of points often needs to be estimated from area-based tornado density.
Meanwhile, tornado density—once defined—is expected to vary across geographic space in a more or less continuous manner. Sudden beginnings and endings of tornadoes are common, but sudden changes of tornado density (and tornado probability) in geographic space are hardly natural, at least in vast, relatively low relief areas such as the central United States. This observation may explain why some researchers would smooth the crisp tornado density field they created (e.g., Ray et al. 2003; Ashley 2007). Nevertheless, how exactly tornado density should vary (continuously) across geographic space—such as its rate of change and its geographic scale (wavelength; see Dixon et al. 2014) of variability—is again relative and is controlled by human choices. From this perspective, tornado density is again different from many other climatic variables, including tornado probability.

More specifically, a tornado density value depends on at least four geographic factors:

1) the shape of the neighborhood area to which this value refers,
2) the size of the neighborhood area,
3) the relative importance (i.e., distinguishable weights or contributions) of individual tornadoes within the neighborhood area in the computation of the density value (e.g., because of their variable relationships with the value point), and
4) the thematic and geometric properties of individual tornadoes considered for the derivation of the density value.

The first two factors correspond neatly to the well-known modifiable areal unit problem (MAUP)—both scale and zoning problems—in GIS (Openshaw 1984; Jelinski and Wu 1996), just in a special computational environment (e.g., in a raster neighborhood computation rather than vector environment) and application background (i.e., a GIS problem hidden in climatology). These two factors determine that the tornado density computation is relative and modifiable because of the modifiable shape and size of the neighborhood area. Neighborhoods of three kinds are common in tornado density mapping, necessarily resulting in different outputs: circular areas (e.g., Dixon et al. 2011; Coleman and Dixon 2014), rectangular (or similar) grids (e.g., Kelly et al. 1978; Brooks et al. 2003; Ashley 2007), and irregular polygons such as counties or states (e.g., Paul 2001; Boruff et al. 2003; Broyles and Crosbie 2004). The neighborhood size adopted in previous works varies greatly as well, typically from 40 km (e.g., Coleman and Dixon 2014) to 200 km or larger (e.g., Kelly et al. 1978; Ray et al. 2003; Dixon and Mercer 2012; Smith et al. 2012).

An additional complication in tornado density literature is that neighborhoods (both the shape and size) are sometimes defined in separate stages. For instance, Ashley (2007) counted tornado variables into cells of 60 km in size and then applied a $3 \times 3$ (i.e., a $180 \text{ km} \times 180 \text{ km}$ rectangle window) low-pass filter to generate smoother surfaces. Coleman and Dixon (2014) first accumulated tornado pathlengths into 1-km$^2$ cells using a 250-km circular window and a kernel function and then aggregated the results into a 5-km grid. “Neighborhood” in these cases is a compound term combining the effects of area (e.g., window, kernel, and grid) choices in both of the stages. When a regular neighborhood is used, it would become incomplete and irregular in both shape and size along the edge of the study area (e.g., along the coastline), causing so-called edge effects (Yamada and Rogerson 2003) that need to be corrected (see section 3a).

The third and fourth factors involve how tornado density is reported. Ideally, the numbers should be read to spatially continuous points or sufficiently small raster cells (e.g., many times smaller than the neighborhood area) to better represent the spatial continuity of tornado density. The choice of small cells also implies that few cells contain multiple tornadoes so that most tornadoes can be differentiated in the density computation. In return, the density difference from a cell to its neighboring cells is primarily because they are associated with different neighborhood areas, but the significant overlapping of these areas would result in a density pattern that is more or less continuous in geographic space. When the geographic distribution of tornadoes is positively autocorrelated (meaning tornadoes tend to have a clustered distribution, as is the case in the United States), this spatial autocorrelation will strengthen the spatial continuity of the resultant density pattern.

For a cell to carry the density value of its surrounding area, though, tornadoes closer to this cell should be considered more important and be given a heavier weight simply because they are more closely related to the density reading at this cell, and vice versa. This weight adjustment is equivalent to distance decaying or inverse-distance weighting in GIS (Franke 1982; Silverman 1986; Mitas and Mitasova 1999) and has been implemented in multiple previous works (e.g., Brooks et al. 2003; Dixon et al. 2011), although the used terms vary. Distance decaying further enhances the geographic continuity of density transition between neighboring cells, as the weights of the same tornadoes would only vary slightly when they are assigned for density values of neighboring cells.

When distance decaying needs to be implemented, the distinction between a circular neighborhood area and a
rectangular or irregular area for density computation is obvious. In a circular neighborhood case, the distance decaying rate and distance limit (of the neighborhood area) would be uniform in all directions, or isotropic. There is consequently the least amount of artifacts or need of justification for the use of circular neighborhoods. When rectangular or irregular shapes are used, however, artifacts related to the zoning problem of MAUP (Jelinski and Wu 1996) will be more severe. For instance, there will be cardinal versus diagonal differences in weight reduction rates within a rectangular neighborhood that are hardly justifiable.

Some researchers counted tornadoes into large non-overlapping grids (Kelly et al. 1978; Ashley 2007) or polygons (Paul 2001; Boruff et al. 2003; Widen et al. 2015), with the benefit of easier and more meaningful (e.g., tornadoes in a county or in a rectangular area) interpretation of resultant patterns, as well as easy integration with socioeconomic or demographic data. However, it necessarily causes loss of geographic details in tornado density variability, because there are too few neighborhoods defined and each tornado can only be counted into one of these neighborhoods with one of two weights: 0 or 1. Sudden transitions of tornado density will be produced across the boundaries of these geographic units, but no changes will be mapped within. A subsequent spatial interpolation or smoothing is sometimes applied to these crisp density readings to generate a more continuous density variability (e.g., Ray et al. 2003; Ashley 2007), but this secondary step is less objective and less respectful of the original data in comparison with a continuous density field that is directly derived from data.

3. GIS-related methodological complications and interpretations

We used the Storm Prediction Center (SPC) tornado database (1973–2013) of the eastern half of the United States and experimented on a few key methodological issues in tornado density mapping. We only included tornadoes at or above an enhanced Fujita (EF) scale of 2 (EF2; 6455 in all; see Fig. 1) because of a higher level of data stability and consistency of these more significant tornadoes (e.g., Coleman and Dixon 2014).

a. Neighborhood size choice, cell size choice, and edge effects

Dixon et al. [2011; also see Dixon and Mercer (2012)] indicate that the choice of the neighborhood size is more important than that of distance decay functions. Indeed, neighborhood size is essential because it reflects the scale problem of MAUP (Openshaw 1984; Jelinski and Wu 1996) in tornado density mapping and controls the
scale (i.e., level of detail) of the resultant density pattern [e.g., see Figs. 5–7 of Dixon et al. (2014)]. However, it is still a subjective choice and is modifiable. When a precise choice of the neighborhood size is difficult to justify, an exploratory attitude and an examination of multiple sizes may become necessary. For example, Dixon et al. (2014) used global Moran’s I as an objective reference to identify the “ideal” radii as the distances at which the strongest spatial autocorrelation of tornado pathlengths was observed, since these distances correspond to the maximum spatial stability of tornado distributions.

While “one size fits all” does not exist (Dixon et al. 2014), as we believe as well, it is important to understand what exact differences exist between various sizes (also what the tornado density map will be used for) before judging which choice is an appropriate one. The range of the tested neighborhood sizes needs to be specified. Dixon et al. (2014) identified a scale range of 55–180 km that can serve as a useful reference. Putting the methodological and data quality factors aside, we believe the minimum tested neighborhood size should be chosen in consideration of the distance between neighboring tornadoes, so that the density readings can be supported by more than one tornado record. On the other hand, the maximum neighborhood size should be much smaller than the study area so as not to oversmooth the geographic variability of the density surface. We calculated the distance between each pair of nearest touchdowns in the study area to guide the choice of the neighborhood size, obtaining an average distance of 11.70 km and a maximum of 127.43 km in the eastern half of the United States with a high standard deviation (9.06 km, see Table 1).

The above results were influenced by adjacency to large water bodies and to mountainous areas where there are few tornado reports (see, e.g., Elsner et al. 2013), as well as by the inclusion of tornado-rare states. To remove the influence, we did the same distance computation in a more tornado-frequent region of the United States and included only a selection of contiguous, inland, and largely low relief states—Arkansas (AR), Iowa (IA), Kansas (KS), Missouri (MO), Nebraska (NE), and Oklahoma (OK)—involving 1896 tornado records. A moderate yet consistent distance reduction was observed (Table 1). A distance of 20 km, which corresponds to the 88.5th percentile of tornado–tornado distances (i.e., distance from each touchdown to its nearest neighbor) in the six selected states (86.2% in the entire study area) but is much smaller than the maximum, was chosen to be the minimum neighborhood size (i.e., radius) tested. With this radius, most neighborhood areas contain more than one tornado to meaningfully report tornado density variability at a local scale.

To understand how tornado density patterns vary with the change of neighborhood size, we calculated tornado density using touchdown point data across a range of 12 neighborhood sizes from 20 to 360 km: 20, 40, 60, 80, 100, 120, 150, 180, 210, 240, 300, and 360 km. The size 360 km was chosen in consideration of the size of the study area (see Fig. 1). We used a linear inverse-distance-weighting function [see Eq. (4) in section 3b] to count the weighted number of tornadoes within the moving neighborhood area of the tested radii.

Cell size is another important parameter, which determines how well the role of each tornado in tornado density mapping can be distinguished from its nearest neighbor(s). We used 1-km (the same for future maps unless otherwise specified) cells. Cells of this size are sufficiently small to distinguish almost all (≥97%) neighboring tornadoes for tornado density computations. Edge effects artificially reduce tornado density readings (e.g., weighted counts in neighborhood areas) along the boundary of the study area (e.g., in Florida). We removed these effects with a scale-specific edge effect remover (for this and all future density computations) in three steps:

1) Convert the entire study area (only) into a raster layer of 1-km cells and count the number of cells within an r-radius circular area of each cell, where r is the chosen neighborhood size.

2) Calculate the percentage of each circular area (r) covered by the study area (the count obtained in step 1 divided by the number of cells in a full circle), which would be less than 100% along the edges, hence being able to serve as an edge effect remover at r.

3) Correct the original density readings by dividing them with the percentages obtained above.

Figure 2 shows the outputs from the experiments on neighborhood sizes. The legend in Fig. 2 only displays the relative difference (e.g., high vs low) in density

### Table 1. Summary statistics of distances from each tornado (touchdown point) to its nearest neighbor, for all EF2–EF5 tornadoes in the study area and for those in selected states (AR, IA, KS, MO, NE, and OK).

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std dev</th>
<th>Min</th>
<th>First quartile</th>
<th>Third quartile</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>All tornadoes</td>
<td>11.70</td>
<td>9.06</td>
<td>0</td>
<td>5.60</td>
<td>15.24</td>
<td>127.43</td>
</tr>
<tr>
<td>Tornadoes of six states</td>
<td>10.78</td>
<td>7.39</td>
<td>0</td>
<td>5.49</td>
<td>14.32</td>
<td>52.98</td>
</tr>
</tbody>
</table>
FIG. 2. Tornado density maps at 12 different scales (neighborhood sizes; km) using the linear distance decay function and tornado touchdown points: (a) 20, (b) 40, (c) 60, (d) 80, (e) 100, (f) 120, (g) 150, (h) 180, (i) 210, (j) 240, (k) 300, and (l) 360.
values on each of the maps, so as to facilitate a visual comparison of spatial patterns between the maps. The actual density values on the density maps have very different magnitudes when the neighborhood size varies. The adjustment of the neighborhood size primarily resulted in dramatically variable wavelengths (i.e., amount of details) of the mapped tornado density patterns, especially when neighborhood sizes are smaller than 210 km. These patterns are all interpretable, just at different scales (hence fitting for different applications). This is to say, the choice of the neighborhood size for tornado density mapping can (and needs to) be based on the targeted level of details or wavelength in the tornado density pattern, which is necessarily related to the mapping purpose. In this paper, the above statement is made without considering data quality (e.g., the stability and temporal span covered by the records) issues.

It should be noted that coarsening the neighborhood size did not shift the locations of tornado centers. Instead, it smoothed out many local tornado centers to reveal only those regional- to continental-scale tornado centers. With the use of the linear distance decay function, five prominent tornado centers are stably observed across most of the tested neighborhood sizes: central Oklahoma, northeastern Texas, north-central Arkansas, southwestern Mississippi, and northern Alabama.

b. The role of distance decay functions

While the incorporation of distance decaying is common in literature, the choice of distance decay functions, as well as the very role of these functions, is not well agreed on (e.g., Dixon and Mercer 2012; Marsh and Brooks 2012). If the direct mapping purpose is to depict tornado density patterns instead of estimating the underlying probability surface, we believe the choice of distance decay functions is largely a subjective one and that attention should be paid primarily to their specialty and differences [e.g., see Silverman (1986), sections 3.3 and 3.4]. Our discussion here, then, is not to evaluate which is a “better” function, but to interpret and evaluate differences between a few functions. Some of the evaluated functions were modified in this paper from their original form for comparison and interpretation purposes, so that all of the evaluated functions would assign a weight from 0 (or nearly 0) to 1 within a circular neighborhood area of radius \( r \) (200 km in Fig. 3 for the demonstration). Tornadoes near the neighborhood edge would receive a weight close to 0, whereas those at the center would receive a weight of 1. The four distance decay functions are as follows:

1) The adjusted Epanechnikov quadratic probability density function (Silverman 1986; implemented in

\[
W(x) = 1 - \left( \frac{d_x}{r} \right)^2, \quad (1)
\]

where \( W(x) \) is the weight of tornado \( x \), \( d_x \) is the distance from \( x \) to the circular neighborhood center, and \( r \) is the radius (size) of the neighborhood (\( d_x = r \)), controlling the scale of the resultant density pattern.

2) A modified Gaussian kernel function (with the normalization term being removed for comparison purposes), similar to the one used by Brooks et al. (2003), is

\[
W(x) = e^{-\left(1/2\right)d_x^2/\sigma^2}, \quad (2)
\]

where \( \sigma \) is a smoothing parameter whose role is equivalent to \( r \). The variable \( \sigma \) is 60 km in Fig. 3 for this comparison, so that the modified Gaussian function can generate a similar range of weight reduction (i.e., from 1 to close to 0) within a similar distance range (200 km in Fig. 3) to be comparable with the other functions.

3) A quasi-IDW (see Franke 1982) function, where the \( d_x \) factor has been removed from the denominator [see Eq. (2) of Franke (1982)] for a weakened distance decaying, is expressed as

FIG. 3. Four distance decay functions: 1) adjusted quadratic probability, 2) modified Gaussian, 3) quasi IDW, and 4) linear.
4) The simplest case, a linear inverse-distance-weighted function, is expressed as

\[ W(x) = \frac{r - d}{r}. \]  

(4)

The linear function generates a constant rate of weight reduction within the neighborhood and can be used as a reference for the evaluation of other functions. In comparison, the weights assigned by the quadratic function are heavily biased toward tornadoes closer to the neighborhood center for a long distance (i.e., up to 100 km, or \( r/2 \), where the weight reduction rate reached the rate of the linear function). In other words, tornadoes within \( r/2 \) receive biased high weights with relatively slow weight reduction, which allows high weights of these tornadoes to be combined to identify large tornado clusters, or large zones of high tornado density. Small but separate tornado clusters (i.e., density peaks) within \( r/2 \) may be combined while their individual prominence is reduced (e.g., see Fig. 6a, described in greater detail below). The weight reduces sharply beyond \( r/2 \), especially near the edge of the neighborhood (Fig. 3). Change of the neighborhood size (Fig. 4) would result in a new set of high-weight tornadoes that are located within a new \( r/2 \) distance, allowing this function to effectively detect the wavelengths of tornado density variability with changing neighborhood size or kernel radius (e.g., see Dixon and Mercer 2012; Dixon et al. 2014).

The smoothing parameter \( \sigma \) in the modified Gaussian function plays a similar role to \( r \) (e.g., in the quadratic function) but takes a much smaller value (60 km, or 30% of \( r \), in Fig. 3) than \( r \) when the corresponding functions spread weights within a similar distance range (see Fig. 3). Additionally, the Gaussian function assigns high weights only to tornadoes immediately next to the neighborhood center (<30 km in Fig. 3, roughly \( \sigma/2 \) or 15% of \( r \), while the rate of weight reduction reaches the rate of the linear function at \( r = 19 \) km). A sharp decline of weights follows (30–110 km). Across the second half of the radius (especially \( >120 \) km, or \( 2\sigma \)) the assigned weights stay very low. The Gaussian function therefore has a different “scale” implication compared to other functions: its use would result in strong bias or preference toward tornadoes immediately next to the neighborhood center (roughly \( \sigma/2 \) only), and tornadoes beyond \( 2\sigma \) are given very low weights in density computations. As a result, use of the Gaussian function would effectively identify small but densely packed tornado clusters such as the ones in eastern North Dakota and southeastern Nebraska (see Fig. 6b, described in greater detail below). In comparison to the quadratic function, the choice of the Gaussian function has an effect that is similar—but not equivalent—to reducing the neighborhood size since heavy weights are assigned to 15% of \( r \) rather than \( r/2 \), when the two functions distribute weights within a neighborhood of similar sizes.
Change of $\sigma$ (equivalent to change of neighborhood size $r$) in the Gaussian function primarily affects the width of the first weight reduction stage—the high weight and short range ($<\sigma/2$) stage (Fig. 5)—making this function an effective detector of scale-specific tornado clusters (spikes). Since the Gaussian function can assign some (very low) weights to much more remote stations than the other functions (which is another key difference between the Gaussian function and the quadratic function), a secondary effect of adjusting $\sigma$ is that the range of weight reduction (i.e., $\to 0$) can be greatly affected. Figure 5 displays the modified Gaussian function curves with five $\sigma$ values: 17, 32, 45, 55, and 71 km.

In comparison with the modified Gaussian function, the quasi-IDW function produces a more stable weight decline throughout. Relative to the linear function, however, its weight reduction is sharper in the first $r/2$ (the weight reduction rate lowered to the rate of the linear function at 100 km, or $r/2$, in Fig. 3), but much slower at a low level near the neighborhood edge (>140 km). In comparison with the other functions, the quasi-IDW function assigns the lowest (relative) weights to tornadoes near the neighborhood center, making it ineffective in detecting either prominent regional high-density zones or local density spikes (see Fig. 6c).

Figure 6 shows 3D tornado density maps produced using the four distance decay functions after the weights were spread (i.e., reduced to 0, or close to 0) within a neighborhood of 60 km (roughly equivalent to $\sigma$ of 17 km). This neighborhood size was chosen to demonstrate how the four functions may behave differently at a regional wavelength (i.e., within individual states of the United States), as well as to identify possible regional tornado concentrations or clusters.

It should be noted that the sum of weights assigned by each distance decay function (e.g., the area covered by each curve in Fig. 3) is unique, resulting in different magnitudes of tornado density values because of the adjustment of the distance functions (Fig. 6). On average, the quadratic function consistently reports higher density values than the other functions. However, the primary difference between distance decay functions is in how to spread weights among tornadoes within the neighborhood area, so as to distinguish the relative importance of each tornado in comparison with others. The magnitude difference of the assigned weights (hence density values) should therefore not affect the comparison and evaluation results of the functions.

c. Choice of tornado properties and uncertainty in tornado density mapping

So far, tornado density maps in this paper are all based on $\geq$EF2 touchdown points. These maps simply report inverse-distance-weighted “counts” of tornadoes. The diversity of tornado events suggests that tornado-to-tornado differences (i.e., geometric, thematic, and temporal) should be incorporated as far as possible for more informative density mapping. In fact, the produced
density maps often need to carry direct “real world” meanings to be more useful, which requires the incorporation and manipulation of more tornado properties in mapping. The side effect, however, is that stronger uncertainty in data (Verbout et al. 2006) and more complications in analyses will propagate through the analytical procedure (Heuvelink 1998) to make the output more uncertain. This uncertainty issue makes Concannon et al. (2000) and Brooks et al. (2003) advocate the use of simple data (viz., tornado occurrences and touchdown locations) in tornado density mapping because these data are comparatively reliable and consistent.

Existing efforts in literature have been abundant in reporting tornado risk or probability (e.g., Brooks et al. 2003; Doswell et al. 2005; Coleman and Dixon 2014), frequency (e.g., Kelly et al. 1978), hazards (e.g., Boruff et al. 2003), fatalities (e.g., Ashley 2007), tornado clusters (e.g., Broyles and Crosbie 2004; Dixon et al. 2011), and so on. Some of these efforts involve more complex analyses and/or more data than simply counting touchdowns, hence implying possible presence of more uncertainty sources. However, the outputs of existing works, including maps presented in this paper, are typically reported in an “as is” manner. Since “ground truth” data are unavailable [see Tyrrell (2003) for a rare exception] to support the evaluation of the mapping outputs, what is needed is an uncertainty and propagation analysis for tornado density (or probability) mapping [see, e.g., Meyer et al. (2002) for a similar effort]. We thereby suggest a backward step be taken: while the ultimate goal may still be to map tornado probability, efforts should first be made to define the uncertainty level of the involved density maps.

As an experiment, we conducted a Monte Carlo analysis after incorporating tornado pathlengths and their errors into density mapping. This analysis starts with an effort to map tornado density according to the pathlengths of involved tornadoes. In the United States, SPC data report tornado paths as straight lines connecting start and end points, regardless of the actual shape of tornado tracks. We used the linear distance decay function and created a tornado density map by incorporating tornado path data of SPC (Fig. 7), following two steps. First, we converted each tornado path (line) into one or multiple square-shaped “tornado” cells of 291 m$^2$ in size, with the tornado pathlength determining the number of cells the tornado covers. The measurement 291 m is the mean width (with a standard deviation of 383 m) of EF2 tornadoes that have the width information available, allowing the converted tornado cells to represent land areas that were affected by tornadoes. Second, we counted weighted tornado cells by use of a circular neighborhood window of 60 km in radius. The resultant density values are necessarily affected by the propagation of the length and shape error of the SPC path data. Figure 8 is a scatterplot of 146 randomly sampled density values based on touchdown point data (at a spatial resolution of 1 km) and pathlength data (at a spatial resolution of 291 m), both using the linear distance decay function and a 60-km circular neighborhood window. The two sets of values were taken from 146 randomly sampled points of the study area, with a second-order polynomial trend line fitted.

In comparison with the density map produced using tornado touchdown data (e.g., see Fig. 2c), the choice of pathlength data resulted in a different geographic
pattern (Fig. 7), exemplifying the strong impact of tornado property choices (also see Fig. 8). For instance, Mississippi, Alabama, and Arkansas are shown to be states containing the densest tornado path cells. The tornado density is reduced in Oklahoma and dramatically reduced in northeastern Texas, indicating the abundance of frequent but short-path tornadoes in both of the states. States such as Iowa, Nebraska, Indiana, and Missouri have similar density on the two tornado density maps.

Next, we introduced a random error term into the SPC path data to repeat the above density calculation, as required by the Monte Carlo procedure. In the case study, this random error term was introduced as a quantitative measure of the “length” error in tornado path data (but not the track “location” or “shape” error), and it is based on an examination of the differences between straight-line paths and the curved, “real” tornado tracks. The random error term was created and incorporated in five steps:

1) Digitize 100 tracks of historical tornadoes using NOAA satellite tornado damage images, into two sets of lines: real, often curved tornado polylines and straight lines (e.g., see Fig. 9).

2) Calculate the lengths of these two sets of lines, as well as the real/straight error ratio of the two sets of lines (expectably $\approx 1$) and then obtain summary statistics of the error ratio for the 100 sample tornadoes (Table 2).

3) For all tornadoes included in our study, create a random error term that has the same mean and standard deviation as the error ratio defined in step 2 (boldface numbers in Table 2) and then assign the random error term to all the tornadoes as a new attribute, which also becomes the attribute of the converted tornado cells.

4) Recalculate the tornado density at the selected scale and cell size (e.g., the same as those used for Fig. 7) by accumulating the random error term of the tornado cells in the same weighed manner, which will result in a new tornado density map.

5) Repeat steps 3 and 4 for 100 times to complete the 100-repetition Monte Carlo analysis (Heuvelink 1998).

From a new set of 100 tornado density maps, an average density map was obtained (Fig. 10a). Compared to Fig. 7, this map incorporated sample statistics (Table 2) of the tornado pathlength error when the actual all-population error was unavailable. It is also a more stable solution for tornado density mapping because of numerous repetitions. This density map is accompanied

---

**Table 2. Differences between digitized real tornado path and straight-line lengths of 100 tornadoes.** See the text for an explanation of the boldface numbers.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Straight (m)</td>
<td>25.818</td>
<td>31.667</td>
<td>554</td>
<td>194.888</td>
</tr>
<tr>
<td>Digitized real (curved; m)</td>
<td>26.516</td>
<td>32.257</td>
<td>608</td>
<td>199.232</td>
</tr>
<tr>
<td>Curved/straight ratio</td>
<td><strong>1.032</strong></td>
<td><strong>0.049</strong></td>
<td>1</td>
<td>1.321</td>
</tr>
</tbody>
</table>
by an uncertainty map (standard deviation of the 100 repetitions for each cell, Fig. 10b) that reports the variability (or stability) of each tornado density reading on Fig. 10a. We also defined a relative error of tornado density as the relative difference between the Monte Carlo mean density and the SPC (pathlength) density, calculated as the percentage of their difference (the former minus the latter) in the density value derived from SPC paths. The corresponding error map (Fig. 10c) reports the magnitude of error in each density reading that was caused by the error of tornado pathlengths. Although there are numerous minor errors in reported tornado pathlengths (Table 2), they resulted in relatively small errors in tornado density maps (Fig. 10c). We believe efforts like this are important in future tornado density mapping, especially when more tornado properties need to be incorporated. It should be noted that the above Monte Carlo analysis is based on the sample error of tornado pathlengths. It ignored the effects of other error sources and error-related factors, such as the possibility that long- and short-path tornadoes have different error patterns.

4. Concluding remarks

From a GIS perspective, we defined tornadoes as multidimensional real-world events within a framework of geographic space, time, and theme. Tornado density, in contrast, was recognized as a density field that depends on various human choices to exist, involving the neighborhood window, the method used to compute the density, the considered tornado properties, and the output resolution. Tornado density values can therefore only be relative. The identification of the above factors provided a conceptual foundation for the later interpretation of the tornado mapping process, to allow us to clarify the involved complications.

Three vital factors were identified to cause complications in tornado density mapping: the neighborhood size, the choice of the distance decay function, and the involved tornado properties. We tested 12 neighborhood sizes ranging from 20 to 360 km. Once a neighborhood size is chosen, the choice resulted in a density pattern that can be interpreted at the corresponding scale and may fit for applications at a particular scale. This choice should therefore be made in consideration of the application purpose.

The tested distance decay functions have “scale” implications. They also have their own specialty strengths for tornado density mapping. The adjusted Epanechnikov quadratic function can be used to delineate large zones of high tornado density at various scales and to detect the variable wavelengths of tornado density variability with
the adjustment of neighborhood sizes. The modified Gaussian function can identify prominent tornado density “spikes,” or small areas of high tornado density. The quasi-IDW function cannot effectively identify tornado clusters at any distance range or scale.

Changing the choice of tornado properties in tornado density mapping implies the change of the resultant tornado density maps in both geographic patterns and in meanings of the density readings. Northeastern Texas, for instance, was identified to be a high-density center on the touchdown-based density maps, but had a dramatically reduced importance when the tornado pathlength was incorporated in density mapping. In addition, changing from the touchdown choice to the pathlength choice resulted in adjusted meanings of tornado maps from counts to coverage areas.

The adjustment of the considered tornado properties is also directly related to the level of uncertainty in the resultant density maps. Uncertainty and propagation analyses were therefore prescribed for future tornado density mapping. An experiment based on tornado pathlength data was conducted to demonstrate how a simple Monte Carlo analysis with manageable efforts can characterize uncertainty in resultant tornado density maps. A few GIS technical issues, such as the choice of the cell size and the removal of edge effects, as well as the corresponding GIS methods, were also discussed and demonstrated in the paper.

REFERENCES


Broyles, J. C., and K. C. Crosbie, 2004: Evidence of smaller tornado clusters at any distance range or scale.


