Reply to “Comments on ‘Tornado Risk Analysis: Is Dixie Alley an Extension of Tornado Alley?’”

—P. Grady Dixon
Department of Geosciences, Mississippi State University, Mississippi State, Mississippi

—Andrew E. Mercer
Department of Geosciences, Mississippi State University, Mississippi State, Mississippi

In their comment, Marsh and Brooks (2011, hereafter MB11) have provided a clear explanation of kernel density estimation (KDE). The visual comparisons of multiple kernel types (MB11) help to reduce confusion by readers while aiding with repeatability for future studies. However, as explained in our original paper (Dixon et al. 2011, hereafter DMCA11), the kernel function (i.e., the shape of the probability density function) is much less important than the bandwidth (i.e., the kernel radius). This claim is supported by several previously published works on geospatial analyses (Silverman 1986; Schabenberger and Gotway 2005; de Smith et al. 2007; Xie and Yan 2008; Poulos 2010). A kernel function and bandwidth should be chosen with the data and the scientific question in mind. Therefore, as MB11 argue, it does not make sense to strictly compare magnitudes that were calculated using different kernels. This does not mean that the choice of kernel function or bandwidth can be expected to completely change the character of the original data. In this paper, we show that the differences in results reported by DMCA11 and Brooks et al. (2003, hereafter BDK03) are due largely to the use of tornado path data and point data, respectively.

Using tornado data from the same period (1980–99) as BDK03, we show that almost identical results are achieved with different kernel functions so long as the kernel bandwidths are similar (Fig. 1). MB11 explain that BDK03 employ an effective kernel radius of approximately 308 km (293 km is reported as the radius of the 95th percentile), so we expanded our effective kernel radius to 310 km (a 155-km Epanechnikov kernel and an additional uniform smoothing radius of 155 km). Our original effective kernel radius is 80.50 km (a 40.25-km Epanechnikov kernel and an additional uniform smoothing radius of 40.25 km). The spatial patterns, especially maxima and minima, remain mostly unchanged with variations in kernel function (Figs. 1a,b), but there are strong differences when comparing results based only on tornado points or tornado paths (Figs. 1b,d; Figs. 2a,c; Figs. 2b,d). Correlations of grid values for differing methods yield an $r^2$ of 0.63 for path densities based on different kernel bandwidths (Figs. 1c,d) and 0.49 when comparing point density to path density (Figs. 1b,d) with the same kernel bandwidth. The exact magnitude at many locations may vary with different kernel functions, which is why it is usually unwise to compare results based on different kernels, but the spatial patterns should be less affected.

The original maps presented here make use of all tornado events, whereas BDK03 use “tornado days” as counted within a fixed grid. DMCA11 made use of tornado days by selecting the longest path within 40.25 km (25 miles) of every event. The reason we did not employ tornado days for this reply is that we could not justify any particular method to select representative events when using only tornado points, and our starting grid is fine enough that tornado days for each grid cell would not effectively reduce the population bias that is known to affect tornado records.

To ensure that the differences in results displayed by the two papers in question are not due simply to differences in study periods, we applied our kernel function to point and path data from the same period (1950–2007) as DMCA11 (Fig. 2). Although magnitudes may increase with an increased bandwidth, the general spatial patterns are different only when comparing point density and path density. There are some noticeable differences between the study periods, but the point data still result in maxima restricted mostly to the southern Great Plains and northeastern Colorado, whereas the path data yield maxima across the central Great Plains and the Deep South.

We are not arguing that any particular method (points or paths, Gaussian or Epanechnikov, or large or small bandwidth) is superior to the others. Regardless of the methods, the original data should be recognizable in the patterns and conclusions and the results should be reproducible. DMCA11
Fig. 1. A comparison of (a) tornado frequency maps created by Brooks et al. (2003) and (b)–(d) those created using the kernel function of DMCAII. All maps are based on 1980–99 tornado reports. Differences in the maps are attributed to the original data being (a),(b) tornado starting points or (c),(d) tornado path data and to kernel bandwidths of (a),(b),(d) ~310 km versus (c) 80 km. Contouring style and colors were chosen to be consistent with (a) Brooks et al. (2003).

Fig. 2. A comparison of tornado frequency maps based on data for the period 1950–2007. Differences in the maps are attributed to the original data being (a),(b) tornado starting points or (c),(d) tornado path data and to kernel bandwidths of (b),(d) 310 km versus (a),(c) 80 km. Contouring style and colors were chosen to be consistent with (c) DMCAII.
made use of path data because their goal was to assess tornado risk to humans and property. We see much value in the methods of BDK03, especially to atmospheric science researchers, because predicting tornado occurrence requires knowledge about where they form most frequently. The larger bandwidth of BDK03 is helpful because it produces a product that is less affected by individual events and is relatively easy to interpret. The small bandwidth of DMCA11 is based on a threat radius that is commonly communicated by the Storm Prediction Center (“within 25 miles of a point”), and it allows local areas to be highlighted (due to relative maxima or minima) for future study.

REFERENCES


