METHODS TO RANK AND CLASSIFY SEVERE WEATHER OUTBREAKS

Chad M. Shafer and Charles A. Doswell III
University of Oklahoma/Cooperative Institute of Mesoscale Meteorological Studies, Norman, OK

1. INTRODUCTION

In recent years, a project has been undertaken to investigate the degree to which synoptic-scale processes determine the occurrence or absence of tornado outbreaks (see Mercer et al. 2009; Shafer et al. 2009; Shafer et al. 2010a,b). These studies required a means of identifying prototypical tornado outbreaks for evaluation. Doswell et al. (2006 – hereafter D06) developed a technique to rank tornado outbreaks and primarily nontornadic outbreaks (separately) using characteristics of the severe reports from the 1970-2003 period. The highest-ranked cases of each type were considered to be prototypical outbreaks of each type, and agreed with subjective notions regarding the meteorological and societal significance of such events during this period. Mercer et al. (2009) and Shafer et al. (2009) found that tornado outbreaks and primarily nontornadic outbreaks could be discriminated reasonably well using synoptic-scale data as initial conditions in mesoscale model simulations up to three days in advance of the events. Shafer et al. (2010a) cautioned that outbreak discrimination was affected by the time of year in which the events occurred, but these studies nevertheless have shown that synoptic-scale processes play a role in the type of outbreak that occurs.

Of course, most outbreaks of severe convection are not easily classified as major tornado or primarily nontornadic outbreaks. Therefore, we now must determine if synoptic-scale processes can be used to predict the relative severity of convective events. To conduct such studies, identification of prototypical major severe weather outbreaks (mostly but not only major tornado outbreaks) is necessary. This study seeks to rank severe weather outbreaks of any type based on their relative severity, using a method similar to that of D06.

Corresponding author address:
Chad M. Shafer
University of South Alabama
Earth Sciences Department, LSCB 136
Mobile, AL 36688
cmshafer@usouthal.edu

2. CONTROVERSIES

As alluded to above, severe weather outbreaks are ranked based on the characteristics of the severe reports. Several studies have recently shown the numerous nonmeteorological artifacts present in the dataset, from population biases (King 1997) to secular trends in tornado (Brooks et al. 2003; Verbout et al. 2006) and nontornadic (Doswell et al. 2005) reports. Many of these artifacts cannot be accounted for (Shafer and Doswell 2010 – hereafter SD10; Shafer et al. 2010b); however, secular trends in the data (e.g., Fig. 1) can be removed *to some degree*. D06 used a logarithmic temporal detrending technique (illustrated in Fig. 1). Any remaining artifacts in the data will affect the results to some extent, but the uncertainties associated with these artifacts never will be known completely.

Additionally, most severe weather report variables, as archived in the Storm Prediction Center severe weather database (Schaefer and Edwards 1999), indicate societal impacts rather than meteorological variables. For example, the F-scale ratings of tornadoes are direct indications of damage, and the number of fatalities and killer tornadoes clearly are based on societal impacts. Though there clearly is some association between meteorological significance and societal impacts (e.g., Brooks 2004), the correlations between the two certainly are not perfect.

Finally, selection of variables to rank severe weather outbreaks based on their relative severity is not straightforward. Many of the variables are correlated (see SD10). No completely explanatory or exhaustive list of variables has been proposed before to assess the severity of convective outbreaks. What is clear is that using an archived single report variable is inadequate in evaluating an outbreak's severity (see D06 and SD10). Therefore, a multivariate ranking index is preferred, particularly given a single variable's

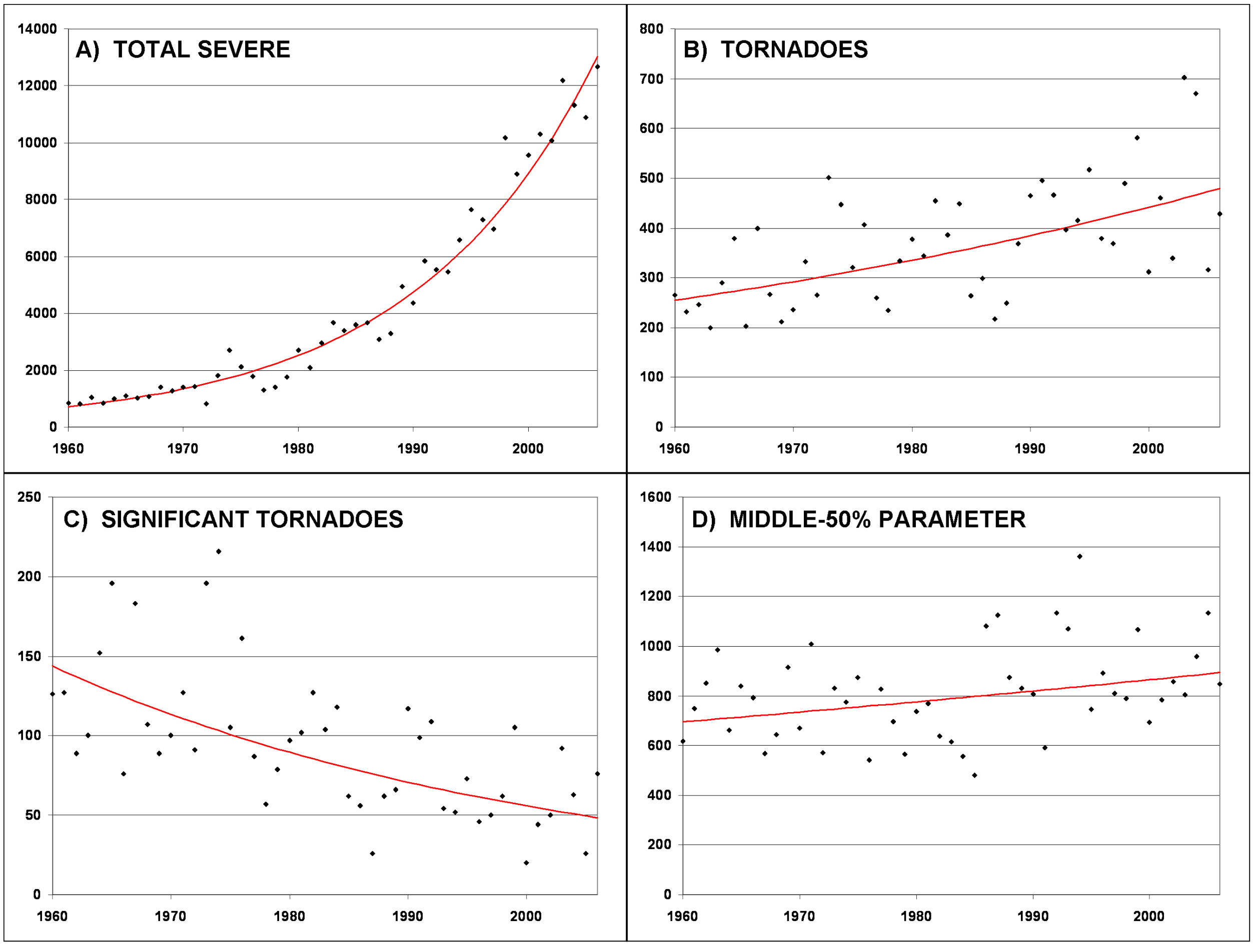

Figure 1: The total number of a) severe reports, b) tornado reports, c) significant tornado reports, and d) the middle-50% parameter, for the top 30 days of total severe reports in a given year. The regression lines fit the logarithm of the reports. The middle-50% parameter is described in Section 3 of the text.

susceptibility to secular trends, even after detrending, as shown by D06.

In order to rank severe weather outbreaks based on their relative severity, it should be shown that (1) the results are relatively robust to modifications of the perceived importance of the severe weather report variables, (2) the rankings agree with subjective notions of the events, (3) the method to rank severe weather outbreaks is repeatable, and (4) the rankings are impervious to secular trends in the dataset. The following sections demonstrate that these four criteria are met sufficiently for our needs.

3. DATA AND METHODS

From the SPC severe weather database, a total of 14 variables were selected or derived to rank the severe weather outbreaks (Table 1). Most of these variables are self-explanatory, and all are described in D06. Many of the variables required detrending (refer to Fig. 1 for examples). The procedure to detrend the variables is discussed in D06 and SD10. Note that multiple methods to detrend the variables could be implemented; the logarithmic detrending conducted herein is simply one such method and was based on the exponential temporal growth in many of the report variables.

Each case is ranked based on a linear-weighted sum of the variables included for analysis. The value of each variable for each case is computed (which is the actual value for that event, or the detrended value if the variable exhibits a substantial secular trend) and then normalized to have a mean of zero and unit standard deviation, which is necessary because of the wide-ranging magnitudes of the variables.

Table 1: Variables used for ranking the severe weather outbreaks from 1960-2006. The parameters are described in Doswell et al. (2006). The asterisk (*) indicates a detrended variable.

Variable	*Abbreviations*
Total number of severe reports*	ALL
Total number of tornadoes*	TORN
Total number of hail reports*	HAIL
Total number of wind reports*	WIND
Total number of significant hail reports*	SIGH
Total number of significant wind reports*	SIGW
Total number of significant tornadoes*	SIGT
Total number of violent tornadoes*	VIOT
Number of long-track tornadoes	LTT
Number of killer tornadoes	KT
Destruction Potential Index	DPI
Total path length	TPL
Fatalities	FTL
The middle-50% parameter*	M50

All variables were standardized as follows:

$$\tilde{x}_i^{(j)} = \frac{x_i^{(j)} - \bar{x}_i}{s_i} \quad (1).$$

In Eqn. (1), *i* represents a particular member of the *n* variables used in the ranking index, and *j* is one of the *m* cases considered for ranking. The mean is symbolized by $\bar{x}$, and the standard deviation is represented as *s*, with their well-known formulas:

$$\bar{x}_i = \frac{1}{m}\sum_{j=1}^{m} x_i^{(j)} \quad (2),$$

$$s_i = \frac{1}{m-1}\sqrt{\sum_{j=1}^{m}[x_i^{(j)} - \mu_i]^2} \quad (3).$$

The standardized variable $\tilde{x}_i^{(j)}$ is then given a weight w_i, and the final score of the index is given by:

$$I^{(j)} = \frac{\sum_{i=1}^{n} w_i \tilde{x}_i^{(j)}}{\sum_{i=1}^{n} w_i} \quad (4).$$

Thus, the score of the index is the sum of the products of the weights and standardized values divided by the sum of the weights. In this manner, it is the *relative* weights of the variables that are pertinent. Variables were weighted with values ranging from 0 to 10, as all of the parameters were associated positively with the significance of severe weather events. The only exception to this is the so-called middle-50% parameter, which is discussed below. This method also allows for easy modification of the weights, to determine the sensitivity of the rankings to various perceptions of the relative importance of the variables.

Initially, the top 30 days each year from 1960-2006 were considered for ranking, based on the total number of severe reports in a given 24-h period (1200 UTC on the nominal date to 1159 UTC the following day). Thus, a total of 1410 cases were ranked, based on the characteristics of the severe reports. This initial criterion required the inclusion of a variable to account for geographic scatter to the reports. D06 introduced the middle-50% parameter as a means of identifying cases with substantial geographic scatter to the reports or multiple clusters of observed severe weather. As such cases do not agree with subjective notions of a single severe weather outbreak, the middle-50% parameter was designed to penalize cases with reports spanning a large portion of the conterminous United States. The middle-50% parameter is the product of the interquartile ranges of the latitudes and longitudes of the reports for a given event (see SD10; their Fig. 3). The larger the product (latitude-by-longitude box), the more likely the case exhibits substantial geographic scatter or multiple clusters of reports. As such, the middle-50% parameter was given a negative weight for proper utilization.

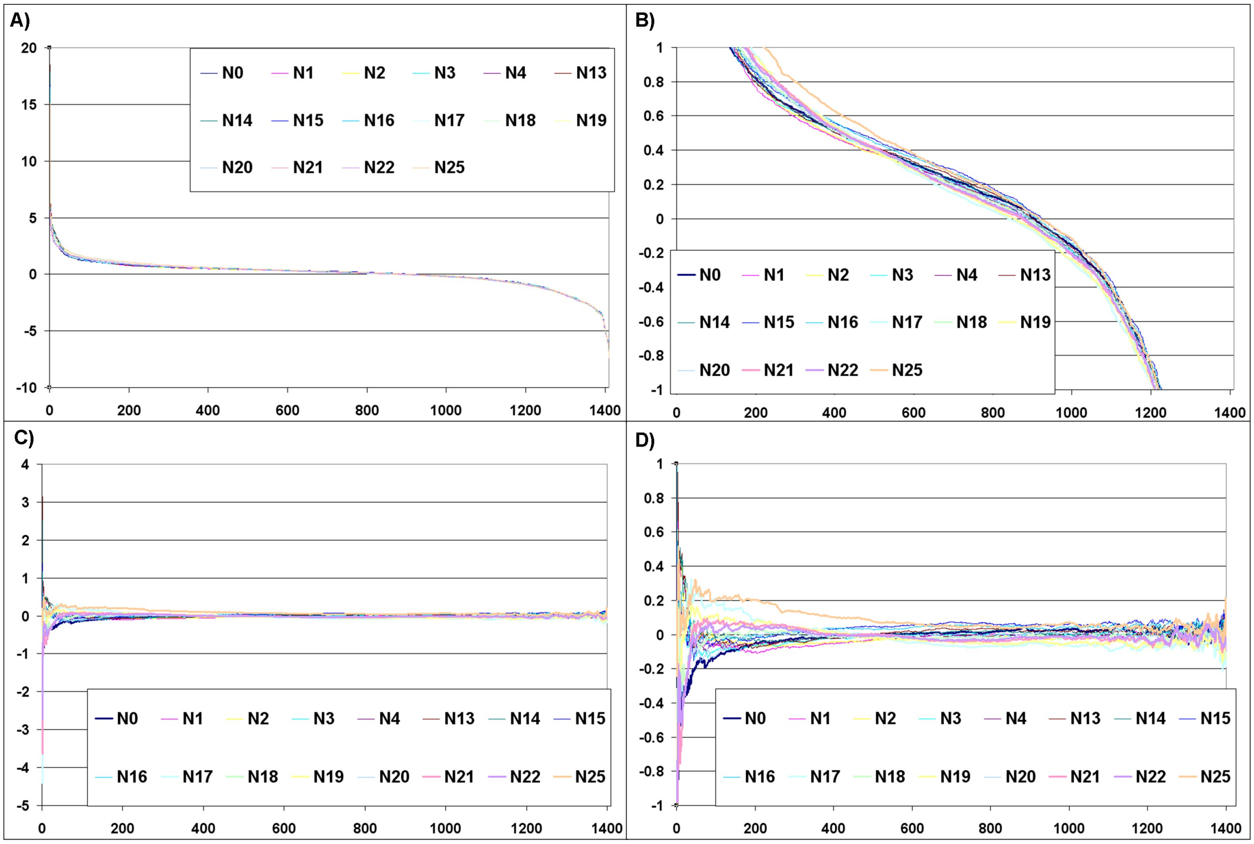

Figure 2: Plots of: a) scores for each outbreak day in order from highest ranking (1) to lowest ranking (1410) for 26 indices, as described in SD10; b) zoom-in on the index scores in the range -1 to 1 — note the different scale on the ordinate; c) deviations for each index from the mean score for each of the rankings, d) zoom-in on the deviations ranging from -1 to 1.

4. INITIAL RESULTS

The outbreak ranking index scores (Eqn. 4) of the 1410 cases, as a function of the outbreak rank (1 is the highest rank; 1410 is the lowest rank) follow remarkably similar curves from highest-to-lowest ranking, despite modifications to the severe weather report variable weights (Fig. 2; see SD10 for details on the 26 indices tested). Most of the variability with the scores occurs with the highest-ranked cases, suggesting the less significant cases exhibit very few differences case to case. This is discussed at length in SD10. As a result, the rankings for the highest-ranked cases feature relatively little variability from index to index, whereas the less significant events display large variability.

In general, three sections of the characteristic curves are observed. The highest-ranked cases are found to have a steep negative slope to the scores (as a function of rank), and are generally the top 100-200 cases of the dataset. The middle ~1000 cases exhibit a gradual to nearly negligible negative slope, and the final ~200 cases return to the steep negative slope. The top-ranked cases are nearly always major tornado outbreaks (with a very small number of high-impact derechos) and are referred to as *major outbreaks* hereafter, and the lowest-ranked cases are those with substantial geographic scatter or multiple clusters of reports. The middle set of cases are referred to hereafter as *intermediate outbreaks*, and are generally those cases with relatively few and/or low-impact tornadoes. Additionally, the method allows for rankings of events in which secular trends are not overly apparent; see SD10 for more details.

A drawback of this initial work is that the lowest-ranked cases either would not be classified as severe weather outbreaks (for those cases with substantial dispersion of the severe weather reports) or could possibly be classified as multiple outbreak events (for those cases with geographically distinct clusters of severe reports). Although the middle-50% parameter is effective in the identification of these events, subsequent discrimination studies (e.g., Shafer et al. 2010b) have struggled to account for these cases in a way that agrees with subjective perceptions of these events. Therefore, a modification to the technique is proposed herein.

5. KERNEL DENSITY ESTIMATION

Rather than using separate 24-h periods as a means of identifying potential outbreak cases for consideration, we propose to use the severe reports themselves as a means of identifying convective outbreaks. We propose using 24-h periods as a means of separating the severe weather reports, but for each period, the density of those reports and their tendencies to cluster are analyzed for event consideration. Kernel density estimation (KDE; Bowman and Azzalini 1997) is used for this purpose.

KDE is simply a means of estimating the probability density function using a variable of interest. Specifically, a one-dimensional KDE can be represented as the following:

$$f(x)=\frac{1}{n}\sum_{i=1}^{n}K_h\left(x-x_i\right) \qquad (5),$$

where n is the number of severe reports on a given day, K_h is a kernel function, and h is a tunable smoothing parameter (bandwidth). Typically, the kernel function implemented is Gaussian (e.g., Brooks et al. 1998), and that is the case for this study:

$$K_h\left(x-x_i\right)=\frac{1}{h\sqrt{2\pi}}\exp\left[\frac{-(x-x_i)^2}{2h^2}\right] \qquad (6).$$

It can be shown that for multivariate KDE, Eqn. (5) can be represented as:

$$f(x)=\frac{1}{n}\sum_{i=1}^{n}\left[\prod_{k=1}^{d}K_{h_k}\left(x^{(k)}-x_i^{(k)}\right)\right] \qquad (7),$$

where d is the number of dimensions; for this study, d = 2. The bandwidth (which can be different for each dimension, but was not in this study) and the threshold value of the approximated probability density function (PDF) can be used to determine the reports associated with a particular geographic cluster. Because d = 2, Eqn. (6) is modified for two dimensions by taking the square of itself, such that the quantity ($x - x_i$) becomes a two-dimensional distance D_i. The end result is:

$$f(x)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{2\pi h^2}\exp\left[\frac{-D_i^2}{2h^2}\right]. \qquad (8).$$

The observed reports for a given day either were associated with a grid point for various map projections using objective analysis techniques (as in Brooks et al. 1998), or were computed as distances from all of the grid points for a particular map projection directly. Thus, the distance quantity in Eqns. (6)-(8) either was defined in terms of grid point separation (i.e., the *grid point method*) or in terms of actual distance (i.e., the *distance method*). It was found, as expected, differences among the techniques and map projections were minor, so long as the bandwidth and PDF thresholds were modified accordingly.

The KDE method is illustrated using reports observed on 29 April 1991 (Fig. 3 – reports not shown). To identify clusters of reports associated with a particular severe weather outbreak, the bandwidth (h) and approximate PDF threshold [$f(x)$] must be tuned to appropriate values. The objective is to find values of bandwidth and PDF threshold such that the contours of PDF threshold are similar in smoothness to those of SPC convective outlooks (similar to the objectives of Brooks et al. 1998) and the threshold magnitude of the PDF threshold encompasses as many of the reports in proximity to the cluster of interest as possible. Smoothness of the PDF contours is determined by the bandwidth, with higher values associated with increased smoothing (Fig. 3). Contours that are too precise (noisy; Fig. 3c) are too dissimilar to SPC outlooks and are indicative of high certainty of the severe reports, which ignores many of the flaws of the severe reports database (Section 2). Conversely, oversmoothed contours tend to adjoin two distinct clusters of reports that are associated with distinct synoptic-scale systems (Fig. 3d). Therefore, bandwidths found in Figs. 3a,b are preferred. As desired, the method is capable of identifying multiple events in the same 24-h period.

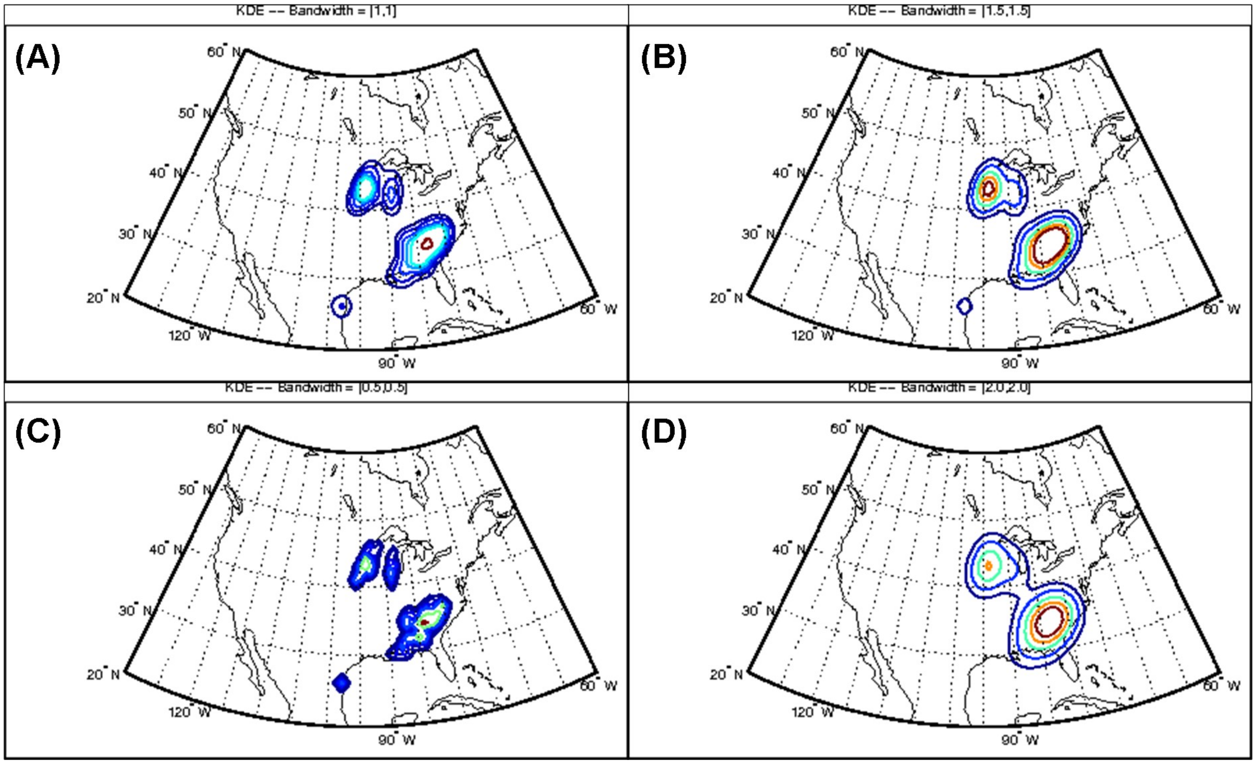

Figure 3: Two-dimensional kernel density estimations of the probability density functions for severe reports from 1200 UTC 29 April 1991 to 1159 UTC 30 April 1991, using bandwidths of (a) 1, (b) 1.5, (c), 0.5, and (d) 2 for the latitudinal and longitudinal directions. These plots use severe reports converted to a latitude-longitude map projection with 1° grid spacing. The outermost contour is 0.001, and the contour within the outermost contour (second contour) is 0.005 for each plot.

After analysis of hundreds of cases, bandwidths of grid spacing ~1° (~100-150 km) were selected. This agrees with previous work (Brooks et al. 1998; 2003) using KDE for severe weather observations and forecasts. Moreover, using either the grid point method or the distance method (cf. Figs. 4a,c) or various map projections with modified grid spacing (cf. Figs. 4a,b) did not alter these results substantially. Although this analysis clearly is subjective in nature, the selection of the same bandwidths and PDF thresholds for a particular map projection and grid spacing will always produce the same result, confirming the repeatability of the method. Moreover, slight modifications to the selection of bandwidth and PDF thresholds do not alter the subsequent rankings of cases (not shown).

The area within the PDF threshold selected is determined to be the region associated with a severe weather outbreak of interest. That is, any report falling on or within the PDF threshold selected is associated with that particular severe weather outbreak. After analyzing each day from 1 January 1960 to 31 December 2008, two additional criteria were introduced to eliminate cases with undesirably few reports or excessive dispersion of the reports. If the number of reports or the density of the reports did not exceed the annual detrended mean value of either parameter (for a given cluster) in the relevant year, the case was excluded from consideration. The end result was 6072 cases (using the 1° latitude-by-longitude map projection and the grid point method) to be ranked for the 49-year period. Not only did this technique automatically exclude cases with substantial geographic scatter, but it also increased the number of events with clustered severe weather reports for ranking outbreaks. Given the observed sample size limitations in previous outbreak studies (see Shafer et al. 2010a,b), this outcome was particularly encouraging.

The subsequent ranking of these 6072 events was conducted in the same way as described in Sections 3 and 4, and the scores as a function of ranking index (Fig. 5) were similar

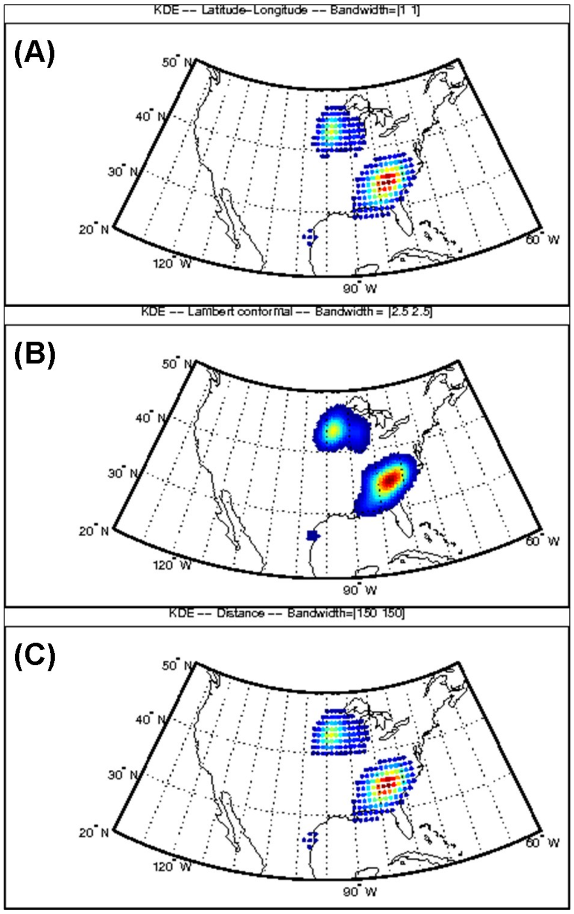

Figure 4: Scatter plots showing the grid points that exceed a specified probability density function threshold using kernel density estimation of the severe reports from 1200 UTC 29 April 1991 to 1159 UTC 30 April 1991. In (a), the grid point method (see relevant text) is used, with a latitude-longitude projection with 1º grid spacing, a bandwidth of 1, and a PDF threshold of 0.001. In (b), the grid point method is used, with a Lambert conformal projection, 54-km grid spacing, a bandwidth of 2.5, and a PDF threshold of 0.0003. In (c), the distance method (see relevant text) is used, with a latitude-longitude projection, 1º grid spacing, a bandwidth of 150 km, and a PDF threshold of 0.00014.

to the initial ranking (cf. Fig. 2). The major differences are the large sample size and the elimination of the steep negative slope for the lowest-ranked cases, attributable to the method's effective elimination of cases with substantial geographic scatter. In general, the rankings of the events were relatively consistent after varying the weights of the severe weather report variables for the highest-ranked cases, and variable for the intermediate events, as with the previous method (see also SD10). Additionally, subjective analysis of the rankings for select cases (Fig. 6) suggests the rankings are consistent with subjective perceptions (i.e., the highest-ranked cases are mostly major tornado outbreaks, with the number, areal coverage, and significance of those reports diminishing as the rankings of the cases decrease. Additional analysis of the KDE method is discussed in Shafer and Doswell (2011).

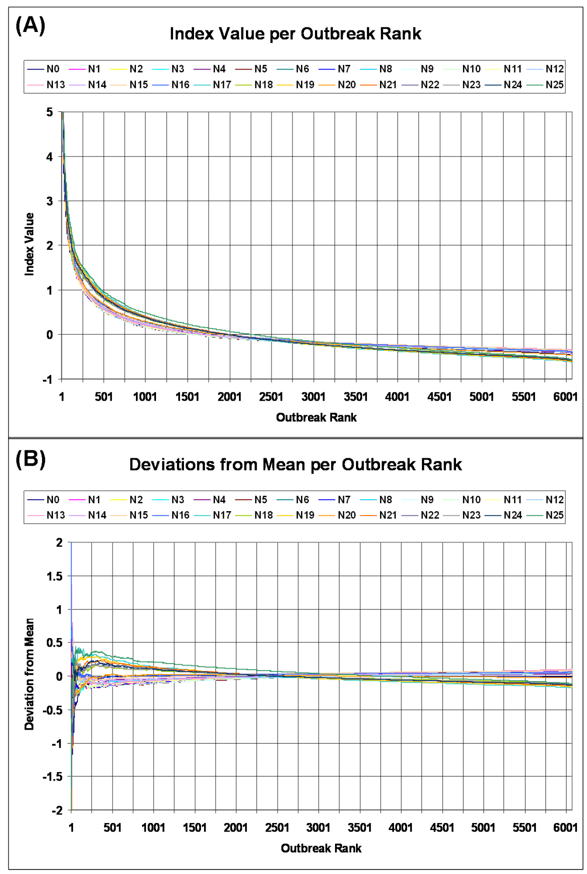

Figure 5: (a) The index scores (*y*-axis) and rankings (*x*-axis) of each of the 6072 cases for each of the indices in the study (labeled). (b) The deviations (*y*-axis) of each of the indices (labeled) from the mean score of all the indices for a particular rank (*x*-axis).

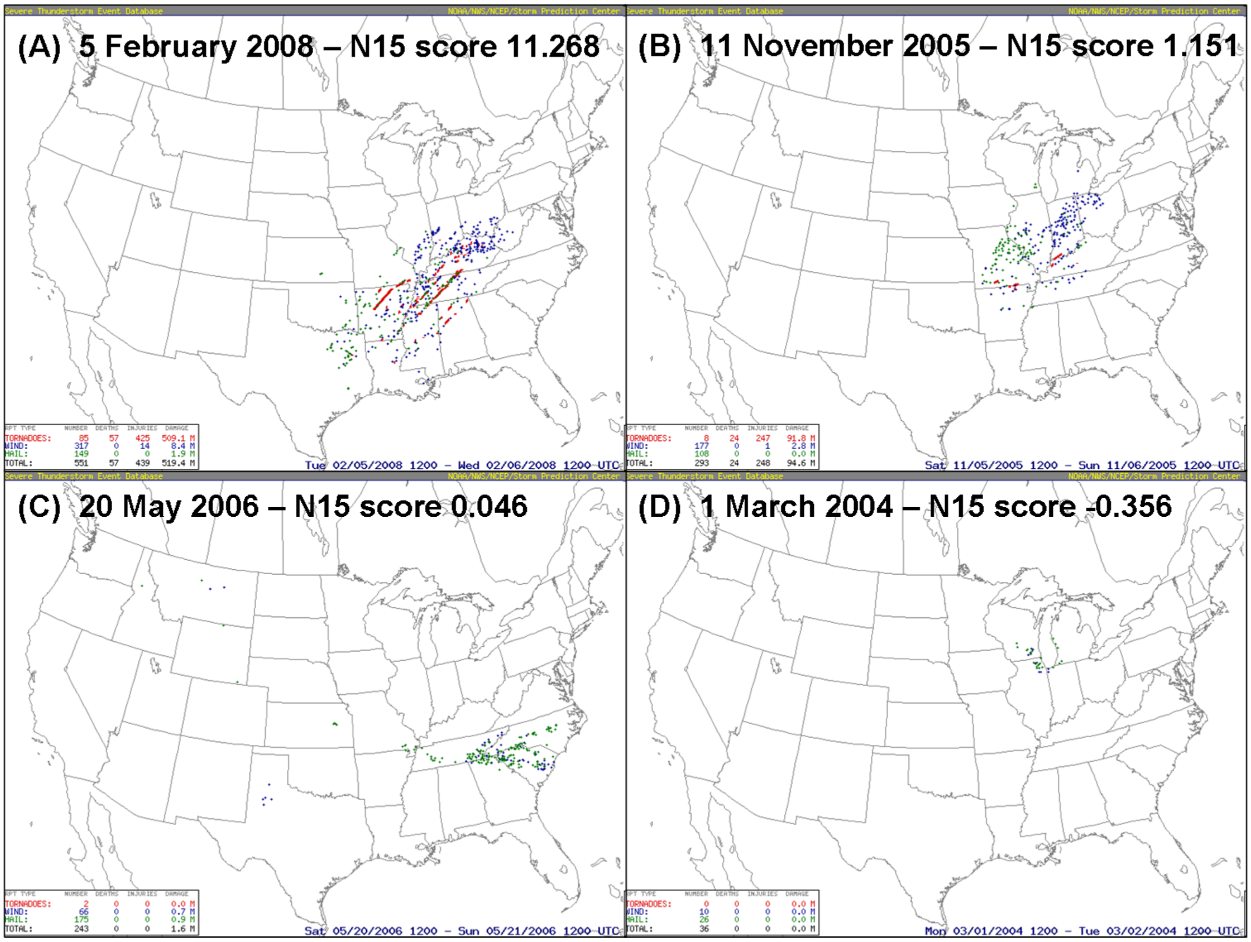

Figure 6: Severe reports on from 1200 UTC on (a) 5 February 2008, (b) 11 November 2005, (c) 20 May 2006, and (d) 1 March 2004 to 1200 UTC the following day, with severe wind gusts or wind damage in blue, severe hail in green, and tornadoes in red. The outbreak ranking index score for the N15 index (one of the 26 evaluated in this study) are indicated. Higher scores indicate higher rankings for the cases.

6. OUTBREAK CLASSIFICATION

SD10 presented a technique to classify severe weather outbreaks based on the characteristics of the severe reports. They used cluster analysis (see, e.g., Gong and Richman 1995) on a four-dimensional decomposition of the ranking indices. The four-dimensional decomposition included separate tornado, hail, wind, and geographic (miscellaneous) components. After it was determined that k-means and Ward's hierarchical clustering were the preferred CA techniques (not shown), case classification was conducted for 2-15 user-specified classes. For the results of the initial work (i.e., Section 4), see SD10. For the results of the 6072 cases obtained using the KDE method, it was found that a small number of even-integer classifications was preferred (i.e., 2, 4, or 6 – see Fig. 6). In general, two-category classifications separated events into major and intermediate severe weather outbreaks (Figs. 7a,b), whereas four-category classifications separated events into major tornado, hail-dominant, wind-dominant, and mixed-mode outbreaks (Fig. 7c). The six-category classification separated hail-dominant and wind-dominant events further into major and minor events. Such classifications agree with subjective analysis of these cases (e.g., the numbered cases in Figs. 7a, as positioned in Figs. 7c,d). This work is also described in Shafer and Doswell (2011).

7. MULTI-DAY OUTBREAKS

The remaining limitation of the present work is the occurrence of multi-day convective

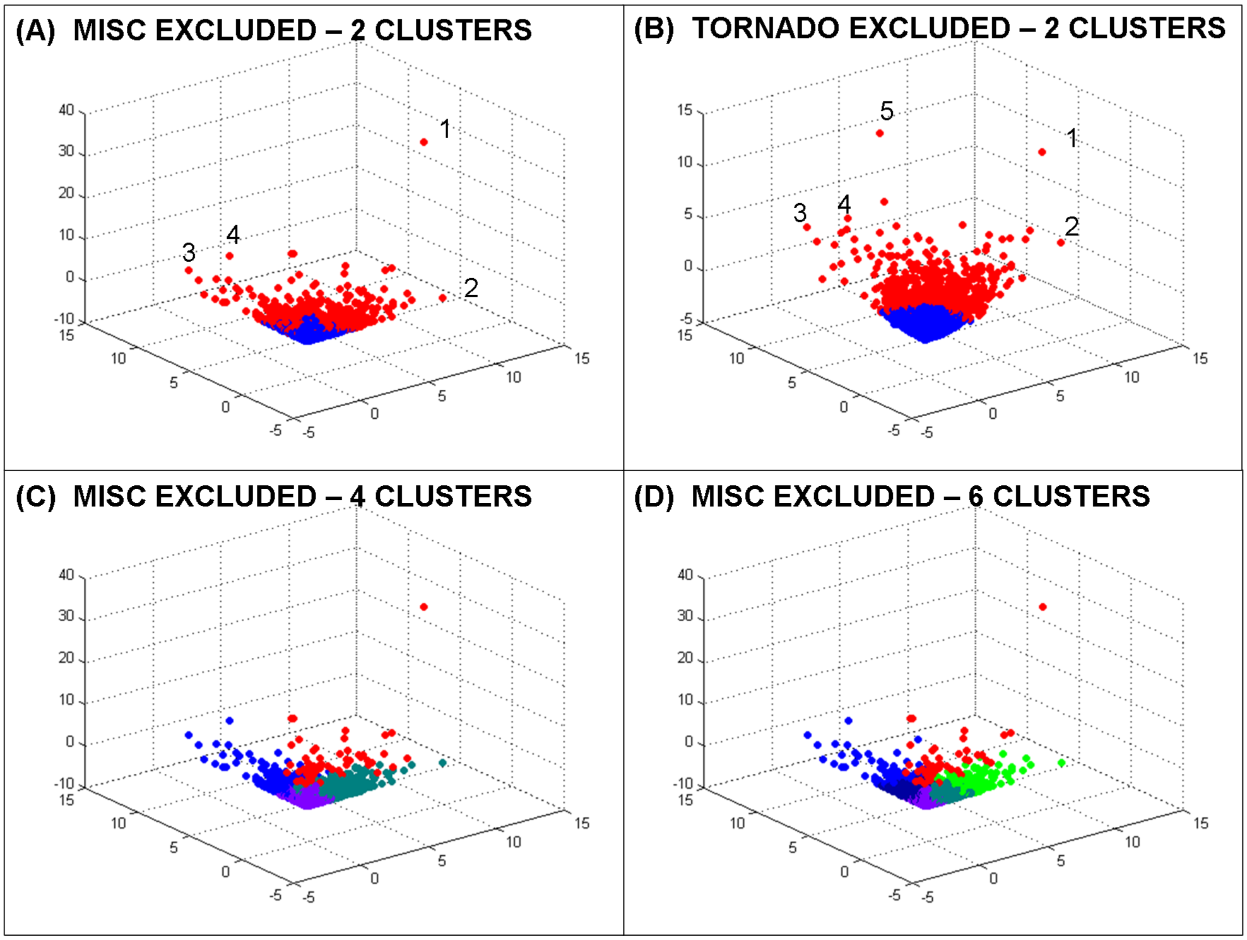

Figure 7: Clusters obtained using the four-dimensional decomposition of the N3 index and k-means cluster analysis. Clusters identified by color, and excluded components of the analysis are labeled. Cases identified include (1) 3 April 1974, (2) 21 April 1996, (3) 1 July 1994, (4) 30 May 1998, and (5) 20 June 1974. In (c) and (d), hail-dominant events are shown in shades of green, wind-dominant events are shown in shades of blue, major tornado outbreaks are shown in red, and mixed-mode events are shown in purple.

outbreaks. Some cases exist in which a severe weather event begins in one 24-h period and continues into (and occasionally well after) the next 24-h period. Other cases feature distinct events occurring within a 24-h period in the same region. Finally, some cases transpire over multiple 24-h periods and become separate events or begin as separate events and become a coherent region of severe weather, or both. Therefore, the next step in the outbreak ranking procedure is to include a time dimension in the KDE analysis to prevent these cases from being considered in a way that does not describe their evolution accurately.

A relatively crude method of accounting for multi-day outbreaks is to look at severe weather reports in small (i.e., < 24 h) intersecting time periods (Fig. 8). Recently, we have used this approach for 6-h time periods in which the final three hours of the first time period are the first three hours of the next time period. A KDE region, as outlined in Section 5, is identified for each 6-h period. If the KDE region in a specific time period intersects a KDE region in the subsequent time period, the events are considered as a single outbreak. This process continues until a KDE region in the latter time period does not intersect a KDE region in the previous time period.

The smallest time period this represents is 9 hours (two 6-h periods with the middle 3 hours considered in both). Therefore, the smallest time period resolved by the method is 3 hours. The number of events for each 3-h

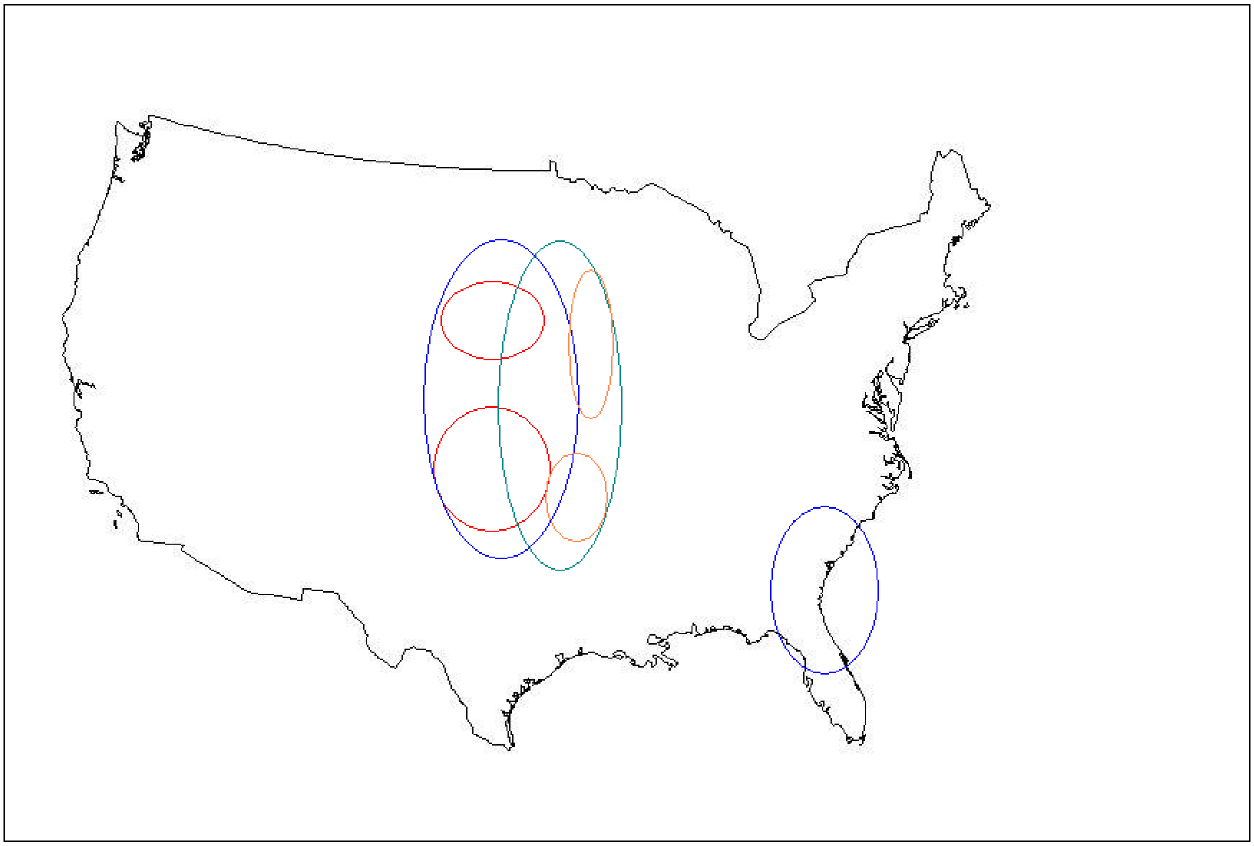

Figure 8: Illustration of the time-increment KDE method. Each ellipse represents a hypothetical region identified by the KDE method associated with a cluster of severe reports, where the color represents a specific 6-h period. For the KDE regions in the central plains, the red region is the first 6-h period, the blue region is the next 6-h period (which includes the last three hours of the red time period), the green region is the next 6-h period (including the last three hours of the blue time period), and the orange region is the next 6-h period (including the last three hours of the green period). The reports within each of the colored regions during each of those 6-h periods would be included as a single event. The blue region in the Southeast (i.e., the same time period as the blue region in the central plains) is associated with a distinct cluster of reports and is considered as a separate event. Note that the blue region in the Southeast has the exact same region in an adjacent 6-h time period, since both time periods would contain the same reports.

period up to the maximum length of an observed multi-day outbreak (Fig. 9) shows that a vast majority (>78%) have outbreak lengths of a day or less and 90.2% have durations ≤ 36 h. Of the outbreak events longer than 36 hours, two distinct types of events are observed. The first is a progressive severe weather outbreak, in which reports traverse a substantial portion of the conterminous US associated with a single synoptic-scale system. The second features a nearly stationary longwave trough, in which numerous shortwave troughs move over a similar region on multiple days. Future studies that examine the multi-day outbreaks identified using this study or similar work will require an objective means of distinguishing these events, as the phenomena associated with each type are quite different. Preliminary rankings of the 5148 events identified as severe weather outbreaks using the time-increment KDE method indicate similar results to previous work, with the multi-day events more appropriately assessed.

8. CONCLUSIONS

Despite the shortcomings of the severe weather database and the difficulties associated with defining and identifying severe weather outbreaks, the methods presented in this study were capable of producing reasonably

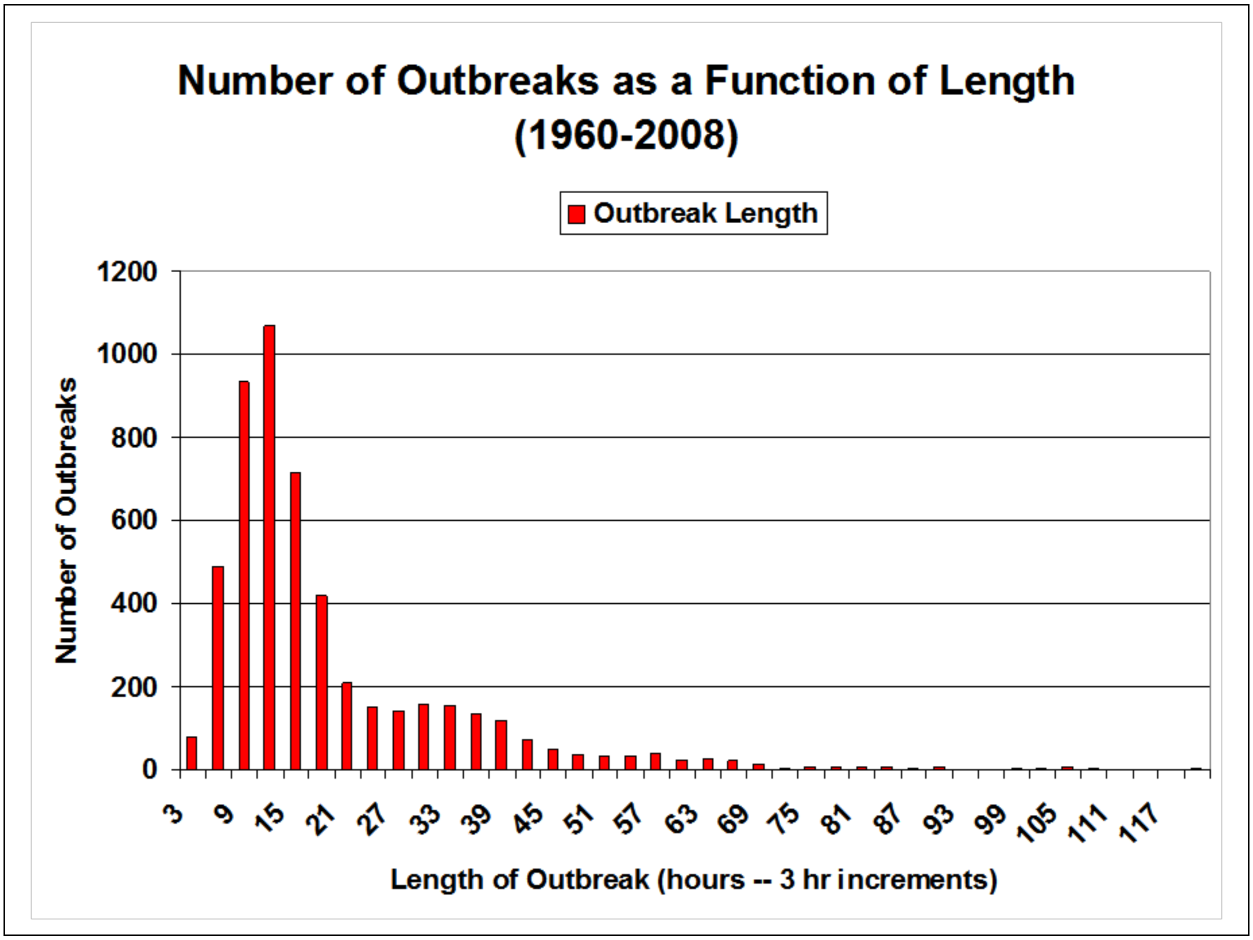

Figure 9: The number of severe weather outbreaks as a function of the length of time associated with the relevant cluster of reports, using the time-increment KDE method.

consistent rankings of major severe weather outbreaks that agreed with subjective perceptions of past events. Furthermore, the methods were repeatable and resistant to secular trends in the data, as desired. Moreover, the highest-ranked cases are ranked consistently, and are mostly major tornado outbreaks as desired. Thus, the primary objective of the study, which is the identification of prototypical major severe weather outbreaks, has been met using a modified version of the linear-weighted, multivariate ranking index first introduced in D06. Although we do not claim that our method is optimal, we suggest that optimality is not necessarily a requirement, given the shortcomings of the severe weather database, as described in Section 1.

Now that a list of major severe weather outbreaks has been identified, subsequent studies can use this study, or similar techniques, to conduct studies of outbreak discrimination. Furthermore, the methods introduced here easily can be amended for other meteorological phenomena (such as flash floods, hurricanes, and winter storms). For more details on the proposed methods and results briefly described in this paper, the reader is referred to SD10 and Shafer and Doswell (2011).

Acknowledgments

The authors wish to acknowledge John Hart and Jared Guyer, who developed SeverePlot V3.0 (used for plots of the severe reports in Fig. 6). We also express appreciation to Harold Brooks, Valliappa Lakshmanan, Kim Elmore, and Roger Edwards for reviewing our work in SD10 and Shafer and Doswell (2011). John Hart was instrumental in developing the middle-50% parameter. This work is supported by NSF Grant AGS-0831359.

References

Bowman, A. W., and A. Azzalini, 1997: *Applied Smoothing Techniques for Data Analysis: The Kernel Approach Using S-Plus Illustrations*. Oxford University Press, 208 pp.

Brooks, H. E., M. Kay, and J. A. Hart, 1998: Objective limits on forecasting skill for rare events. *Preprints*, Nineteenth Conf. on Severe Local Storms, Minneapolis, MN, Amer. Meteor. Soc., 552–555.

——, C. A. Doswell III, and M. P. Kay, 2003: Climatological estimates of local daily tornado probability for the United States. *Wea. Forecasting*, **18**, 626–640.

Doswell, C. A. III, R. Edwards, R. L. Thompson, J. A. Hart, and K. C. Crosbie, 2006: A simple and flexible method for ranking severe weather events. *Wea. Forecasting*, **21**, 939–951.

——, H. E. Brooks, and M. P. Kay, 2005: Climatological estimates of daily local nontornadic severe thunderstorm probability for the United States. *Wea. Forecasting*, **20**, 577–595.

——, R. Edwards, R. L. Thompson, J. A. Hart, and K. C. Crosbie, 2006: A simple and flexible method for ranking severe weather events. *Wea. Forecasting*, **21**, 939–951.

Gong, X., and M. B. Richman, 1995: On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. *J. Climate*, **8**, 897–931.

King, P. S. W., 1997: On the absence of population bias in the tornado climatology of southwestern Ontario. *Wea. Forecasting*, **12**, 939–946.

Mercer, A. E., C. M. Shafer, C. A. Doswell III, L. M. Leslie, and M. B. Richman, 2009: Objective classification of tornadic and non-tornadic outbreaks. *Mon. Wea. Rev.*, **137**, 4355–4368.

Schaefer, J. T., and R. Edwards, 1999: The SPC tornado/severe thunderstorm database. Preprints, *11th Conf. on Applied Climatology*, Dallas, TX, Amer. Meteor. Soc., 603–606.

Shafer, C. M., and C. A. Doswell III, 2010: A multivariate index for ranking and classifying severe weather outbreaks. *Electronic J. Severe Storms Meteor.*, **5** (1), 1–39.

——, and ——, 2011: Using kernel density estimation to identify, rank, and classify severe weather outbreaks. *Electronic J. Severe Storms Meteor.*, **accepted**.

——, A. E. Mercer, C. A. Doswell III, M. B. Richman, and L. M. Leslie, 2009: Evaluation of WRF forecasts of tornadic and nontornadic outbreaks when initialized with synoptic-scale input. *Mon. Wea. Rev.*, **137**, 1250–1271.

——, ——, L. M. Leslie, M. B. Richman, and C. A. Doswell III, 2010a: Evaluation of WRF model simulations of tornadic and nontornadic outbreaks occurring in the spring and fall. *Mon. Wea. Rev.*, **138**, 4098–4119.

——, C. A. Doswell III, M. B. Richman, and L. M. Leslie, 2010b: On the use of areal coverage of parameters favorable for severe weather to discriminate major outbreaks. *Electronic J. Severe Storms Meteor.*, **5** (7), 1–43.

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.