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

Electronic Journal of SEVERE STORMS METEOROLOGY

A Multivariate Index for Ranking and Classifying Severe Weather Outbreaks

CHAD M. SHAFER
Cooperative Institute for Mesoscale Meteorological Studies/School of Meteorology, University of Oklahoma, Norman, Oklahoma

CHARLES A. DOSWELL III
Cooperative Institute for Mesoscale Meteorological Studies, Norman, Oklahoma

(Submitted 26 August 2009; in final form 03 March 2010)

ABSTRACT

In previous work, severe weather outbreaks have been classified either as major tornado outbreaks or as primarily nontornadic outbreaks, but the large majority of such events are of a mixed character. This study proposes a reproducible method for ranking *all types of severe weather outbreaks* from the period 1960-2006. Numerous nonmeteorological artifacts exist in the severe weather reports archived during this period, and many of the variables used to formulate the multivariate indices had to be detrended to reduce the effect of secular trends. The resulting outbreak rankings indicate that the methodology presented herein is able to distinguish the most significant severe weather outbreaks from intermediate outbreak days and days with a large amount of geographic scatter in the severe reports. The rankings of the most severe outbreaks and those outbreak days with a large degree of spatial scatter exhibit only limited variability when the selection of parameters and their weights are modified, but a relatively high degree of volatility is noted with the intermediate cases. This result suggests there is relatively little difference in the severity of these intermediate events. However, the particular *modes* of severe weather in these events can be quite different. A *k*-means cluster analysis of the outbreak days, using a four-dimensional representation of the multivariate indices, indicates that outbreak days can be separated into five groups: major tornado, wind-dominated, hail-dominated, multi-modal, and days with considerable spatial scatter of the severe reports.

1. Introduction

Severe weather forecasters face numerous challenges when a severe weather outbreak is likely to occur. These challenges include determining the primary mode of severe convection, predicting the primary type of severe weather that will occur, and gauging the meteorological and societal significance of the event. Although some outbreaks of severe weather can be categorized dichotomously as major tornado outbreaks or primarily nontornadic outbreaks (Doswell et al. 2006; hereafter, D06; Shafer et al. 2009), most outbreaks consist of severe weather reports of a mixed nature. That is, many severe weather outbreaks feature a moderate number of tornadoes and a large number of large hail and/or severe wind reports. D06 developed a method for ranking major tornado outbreaks and primarily nontornadic outbreaks separately, using two different multivariate indices (the O index for tornado outbreaks; the S index for primarily nontornadic outbreaks). The focus of this study is to rank *all* types of severe weather outbreaks, including those of mixed character, using scores calculated from a multivariate index. If a meteorological model is capable of predicting the scores of future outbreak days accurately and skillfully, the characteristics of the potential outbreak can be predicted based on similarly scored past outbreaks.

Corresponding author address: Chad Shafer, School of Meteorology, University of Oklahoma
120 David L. Boren Blvd., Suite 5900
Norman, OK 73072-7307
E-mail: cmshafer@ou.edu

a. Controversies surrounding ranking outbreaks

1) Meteorological versus societal significance

Ranking outbreaks of severe weather can be controversial for a number of reasons. For example, using variables that measure societal impact of a severe weather outbreak can be a measure of the significance of an event. However, there is not a one-to-one correlation between the *meteorological* significance and the societal impact of severe weather outbreaks — societal impacts often are maximized when meteorological events strike areas of relatively high population density. Many tornado outbreaks occur over the sparsely populated western plains of the United States, resulting in relatively little societal impact of the event regardless of the number and intensity of tornadoes reported during the event. When this happens, there is a tendency for the number and intensity of the tornadoes to be underestimated (Paruk and Blackwell 1994; King 1997). It is also possible for a meteorologically minor severe weather outbreak to occur over a densely populated region (e.g., the East Coast of the United States), resulting in a large societal impact for the event. Similar biases are observed with nontornadic reports (e.g., Billet et al. 1997; Weiss et al. 2002; Doswell et al. 2005). D06 explained that this is a result of the nature of severe convection, with spatial and temporal scales that are relatively small. Severe convection is relatively infrequent and difficult to forecast, rendering the prediction of societal impacts of these events even more challenging than the weather itself. However, it is clear that there is some association between an outbreak's meteorological significance and its societal impact. For example, Brooks (2004) documented a correlation between a tornado's intensity and its path area. A large path area means a relatively high probability that a tornado will cause damage and injury.

2) Nonmeteorological artifacts

A second source of potential controversy in ranking severe weather outbreaks is the plethora of nonmeteorological artifacts found in severe storm databases. The number of annual tornadoes observed over the past four decades has increased substantially (Speheger et al. 2002; Brooks et al. 2003; D06; Verbout et al. 2006), with the total number of reports nearly doubling from the 1960-2003 period (Fig. 1, D06). The number of nontornadic severe weather reports has an even sharper increase (Doswell et al. 2005; D06), with the number of severe reports increasing by an order of magnitude from 1960-2003 (Fig. 4, D06). However, several tornado variables actually have declined over the period, including the number of significant (≥F2) tornadoes, the number of violent (≥F4) tornadoes, the number of long-track tornadoes, and the number of fatalities (e.g., Verbout et al. 2004; D06). The information archived for these events also has changed over time, with increased emphasis on tornado damage intensity (e.g., Fujita 1981) and wind speeds (e.g., Abbey and Fujita 1975; 1979), and with frequent policy changes incorporated into both tornado and nontornadic severe weather reports (see Brooks et al. 2003 and Doswell et al. 2005 for more information). Additionally, the lack of consistency in the recording of these events has been well-documented (e.g., Doswell and Burgess 1988; Weiss et al. 2002; Doswell et al. 2005); and changes in population density undoubtedly have affected the trends in the severe reports (e.g., King 1997). Any scheme developed in the ranking of these outbreaks must take these factors into account, especially the secular trends in the data (D06).

3) Parameter selection

A third source of controversy in ranking severe weather outbreaks involves the selection of parameters to use in developing a ranking scheme. Although the benefits of using a multivariate index are clear (cf. Tables 4 and 5; D06), a mutually exclusive and completely exhaustive list of parameters to develop such an index does not exist. Many of the variables selected to develop the O index and S index in D06 very likely are correlated. For example, the destruction potential index (DPI; Thompson and Vescio 1998) and tornado path length undoubtedly are related. The DPI is the sum of the individual tornado damage areas multiplied by the Fujita-scale rating (plus one, to account for values of F=0) for each tornado during the valid time of an event. The tornado damage area is the path length multiplied by the path width of the tornado. As D06 explained, the nonlinear nature of the DPI suggests the variable can contain some independent information about the outbreak, however. Because of the limited number of observed parameters in archived severe weather reports and the obvious associations among many of the parameters, inclusion of correlated parameters in a ranking scheme is unavoidable. The resulting rankings reported by D06 suggested that including correlated parameters in the development of the indices did not impact the rankings negatively, in a subjective sense.

b. Goals of the study

Whatever issues might be of concern in developing an outbreak ranking scheme, if it can be shown that a candidate scheme is relatively robust to parameter selection and the choice of weights for those parameters, then those concerns may be of only minor importance (Section 2). In this study, several different combinations of parameters and weights for these parameters have been tested to show that the scheme developed is adequately resistant to these subjective changes (Section 3). Furthermore, the final rankings should agree reasonably well with rankings developed subjectively by meteorologists familiar with the events. Subjectivity is an inevitable consequence of a ranking scheme, but it also provides a necessary and reasonable check on the utility of such a scheme. Although it is tempting to seek to determine a "best" ranking scheme, the controversies surrounding the ranking schemes and the limitations of the data analyzed (Section 1a) suggest that this is not possible. Further, the notion of what scheme is best depends on the purpose for ranking the events.

It is not our objective to *define* severe weather outbreaks. The goal of this study is to create a systematic, repeatable methodology to rank the most important severe weather outbreaks to meet the needs of a project related to this topic (e.g., Shafer et al. 2009). Exclusion of days from our study's time period should not be interpreted as a day in which an event meeting *someone's* definition of an outbreak did not occur. The definition of "outbreak" has been a continuing source of contention and debate (e.g., Pautz 1969; Blechman 1975; Galway 1977; Grazulis 1993; Verbout et al. 2006), but we see no need to attempt a formal definition. As with the ranking scheme developed in D06, those wishing to define outbreaks differently should be able to develop a ranking scheme to suit their particular needs.

2. Data and methodology

The data and methodology utilized for the ranking scheme proposed herein were nearly identical to D06. Because the methods incorporated by D06 are scientifically reproducible and produced subjectively favorable results for their work, we saw no compelling reason to alter the techniques they employed substantially. Only the major details will be discussed briefly below; the reader is referred to D06 for more specific details on the methods chosen and the reasons for these choices. The Storm Prediction Center severe storm database is described in detail by Schaefer and Edwards (1999). The data document the type of severe weather reported (tornado, large hail, severe wind), the intensity or, for tornadoes, damage rating, the location, the time of occurrence, and various societal impacts (including number of injuries, number of fatalities, damage costs for events occurring during or after 1998, etc.) for each report. An outbreak day is defined as 1200 UTC on the date indicated to 1159 UTC on the following date. For example, the 3 May 1999 severe weather outbreak is valid from 1200 UTC 3 May 1999 to 1159 UTC 4 May 1999. Each outbreak day was considered independently; consecutive outbreak days were not considered together.

The variables used in the present study (Table 1) are a combination of the parameters used to develop the O index and S index in D06. No additional variables outside of those used in D06 were considered but, as will be illustrated in this study, not all variables listed in Table 1 were used in each multivariate index evaluated. The selection of these specific variables was essentially arbitrary, though the variables are clearly associated with the meteorological significance and societal impacts of severe weather. The selected variables cannot be shown to be optimal, but as D06 stated, demonstration of optimality is not necessarily a requirement.

The development of each index in this study follows the same procedure as that in D06. Specifically, a set of n variables has a value for each of the m cases to be ranked by the index. Each of the variables has its own mean and standard deviation, so all of the variables are standardized by subtracting the relevant mean from each of the m values and dividing the difference by the standard deviation (Eqns. 1-3, D06). The index for the i^{th} case is the sum of the weighted parameters divided by the sum of the weights of the parameters (Eq. 4, D06). The weights of the parameters are determined subjectively. The 26 candidate indices use different combinations of weights to investigate the robustness of the ranking scheme. Weights ranged from 0 to 10. As a consequence of the methodology, whereby the sum of the weighted variable is divided by the sum of the weights, the *relative* weights of the parameters affect the results (D06). More information on the candidate indices, their corresponding weights, and the reasons behind the selections is discussed in Section 3a.

Table 1: Variables used for ranking the severe weather outbreaks from 1960-2006. The parameters are described in Doswell et al. (2006) and in Sections 2 and 3 herein. The asterisk (*) indicates a detrended variable.

Variable	*Abbreviations used in subsequent tables*
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

The top 30 days from each year, *according to the total number of reports of all types of severe weather*, were selected from the period 1960 to 2006 (inclusive) for the ranking scheme developed in this study[1]. Such a large period of record allows a considerable number of outbreak days to be analyzed (1410 = 47 years times 30 days per year) and increases the number of relatively rare events, such as major tornado outbreaks. Detrending several of the variables in Table 1 was necessary for the 47-year period of record, owing to nonmeteorological artifacts in the data (discussed in Section 1), including a substantial increase in the total number of reports for each type of severe weather (see Fig. 1 for examples). Detrending was conducted in exactly the same manner as that of D06, as follows. The sum of the reports for the particular variable to be detrended was computed for the 30 days selected for the relevant year. These sums were calculated for each of the 47 years, and a linear regression is fit to the logarithm of the 30-day sums. This technique was chosen for simplicity, but the trends observed with many of the variables appeared exponential, justifying the use of this method. The result is a nonlinear fit when plotted on a linear scale. Examples of the trend line for several variables are illustrated in Fig. 1. The daily average value was computed by dividing the annual value on the regression line by 30. The detrended variable was the reported value for a specific case during that year divided by the daily average value for that year. Detrended variables then were normalized using the same procedure as that described above.

Approximately 4% of the cases analyzed in this study featured a large number of severe reports scattered throughout the country and evidently were not associated with a single synoptic-scale system (e.g., Fig. 2a). On approximately 6% of case days, two or more spatially separate clusters of events occurred (e.g., Fig. 2b). Other cases (approximately 3%) exhibited a relatively small number of events closely tied to a synoptic-scale system, with numerous others scattered throughout much of the rest of the United States (e.g., Fig. 2c). Collectively, these cases accounted for slightly fewer than 200 of the 1410 cases analyzed (see Sections 3 and 4).

To reduce the rankings for cases such as those observed in Fig. 2, as subjective perceptions of severe weather outbreaks include days featuring a large number of reports over a geographically small area, the method adopted in

[1] The period of record for this study (1960-2006) differs from D06 for both tornado outbreaks (1970-2003) and primarily nontornadic outbreaks (1980-2003).

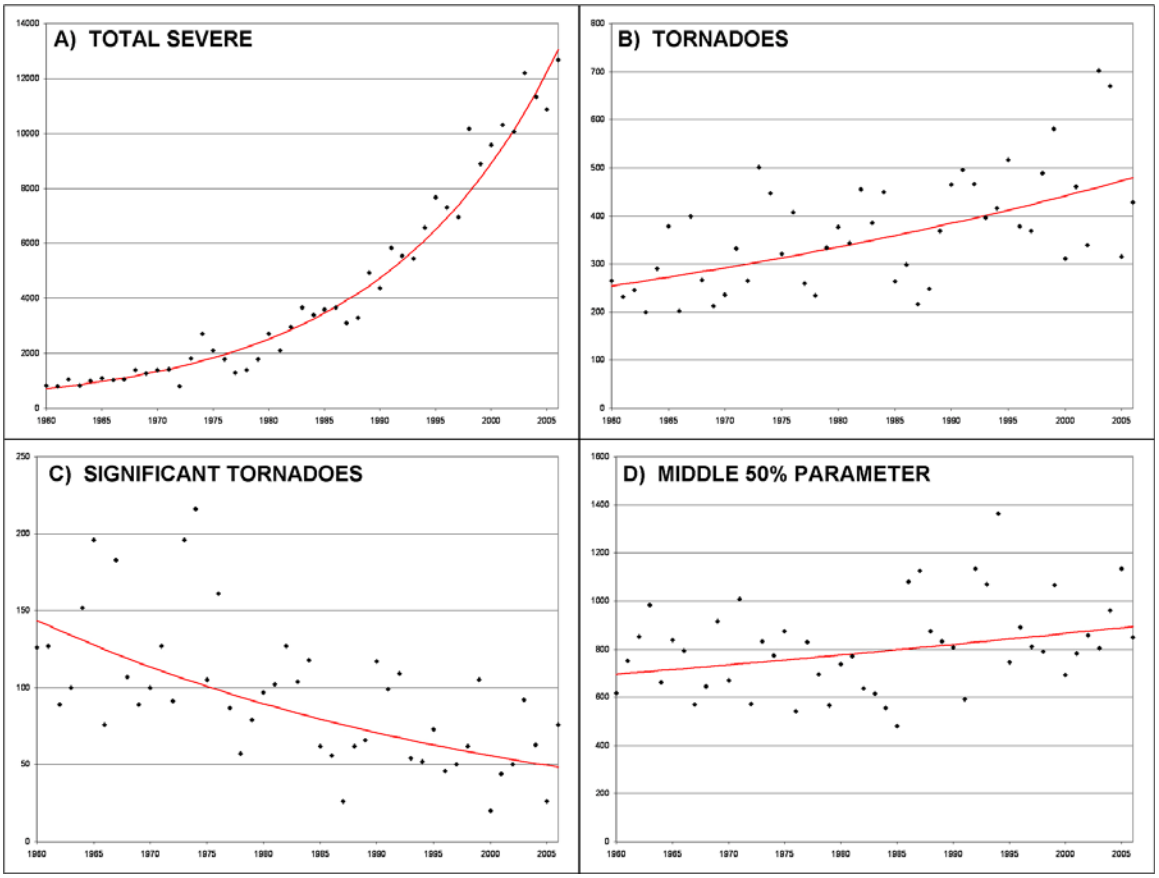


Figure 1: The total number of a) severe reports, b) tornado reports, c) significant tornado reports, and d) the middle-50% parameter, for the top 30 days of total severe reports in a given year. The regression lines fit the logarithm of the reports. *Click image to enlarge*.

D06 was used in this study as well. That is, the 25th and 75th percentiles of the latitudes of the reports are determined and the difference between the two latitude values is calculated. The same procedure is conducted for the longitudes of the reports. The two differences obtained are multiplied together, resulting in a latitude-longitude box (Fig. 3), and this value is used as the "middle-50% parameter". This variable required detrending, due to the large increase in total severe reports over the last five decades (Doswell et al. 2005; D06). Computation of this detrended parameter is done the same way as other detrended variables.

After an index is calculated using the other variables, the detrended middle-50% parameter is subtracted from the original index computed for the outbreak day. Outbreaks that are geographically clustered would have negative middle-50% parameter standardized scores (an area smaller than the mean for that year) and, as a result, would increase the value of the original index computed for that day. On the other hand, cases with a large amount of scatter in the locations of the severe reports would result in a positive middle-50% parameter score and would decrease the original index value. This method proved effective for most of the cases featuring scatter in the reports for a particular day; however, cases exhibiting two highly clustered regions of severe reports (e.g., Fig 2b) may be more appropriately represented as two separate potential severe weather outbreaks. More sophisticated techniques could be employed to account for this tendency (D06, p. 948), but were not developed in this study since cases of this sort are relatively infrequent (see above).

The selection of the parameters for the multivariate indices, the weights of the parameters, the detrending technique, and the development of a linear-weighted ranking scheme were all essentially arbitrary. We are not

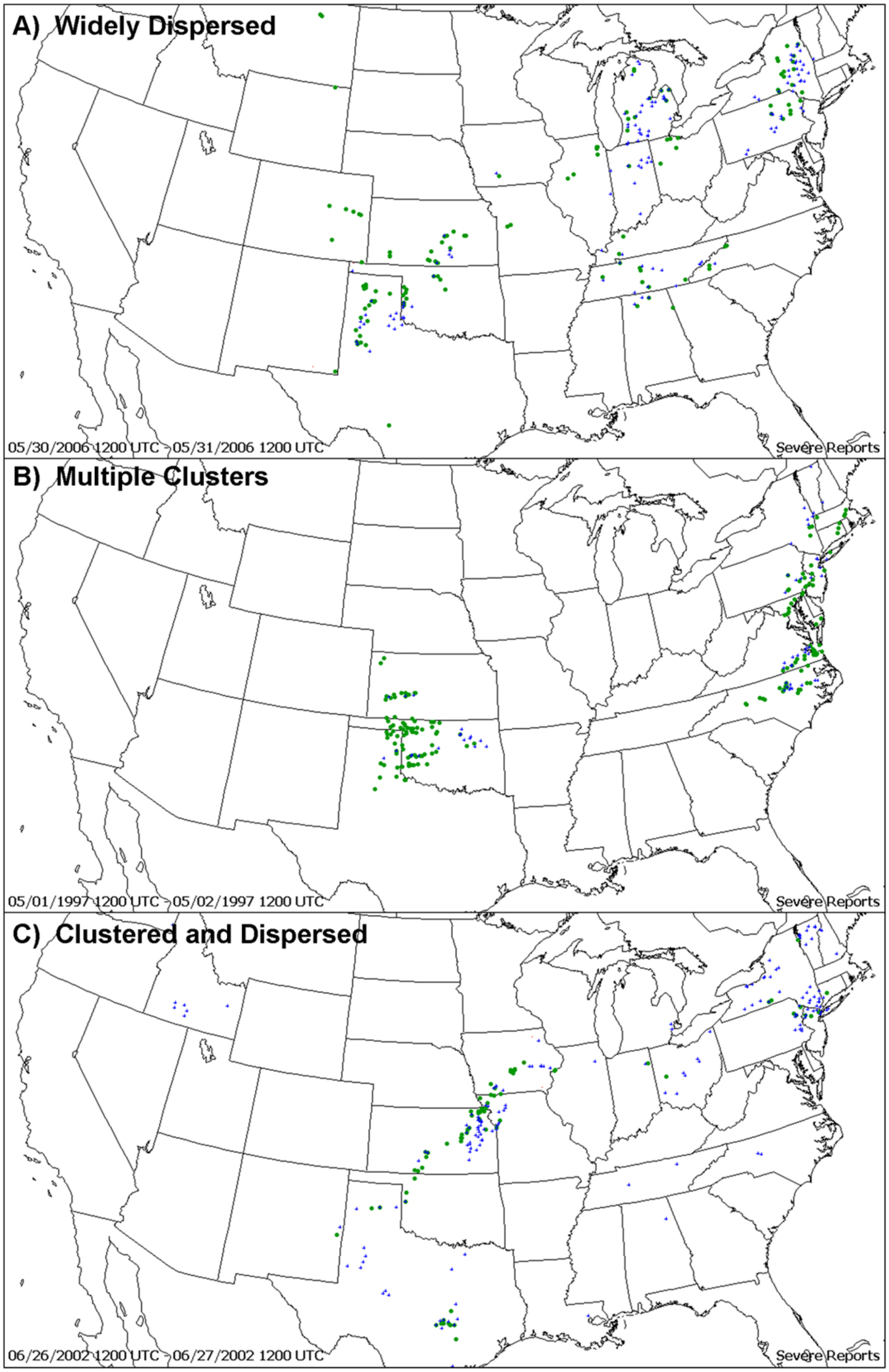
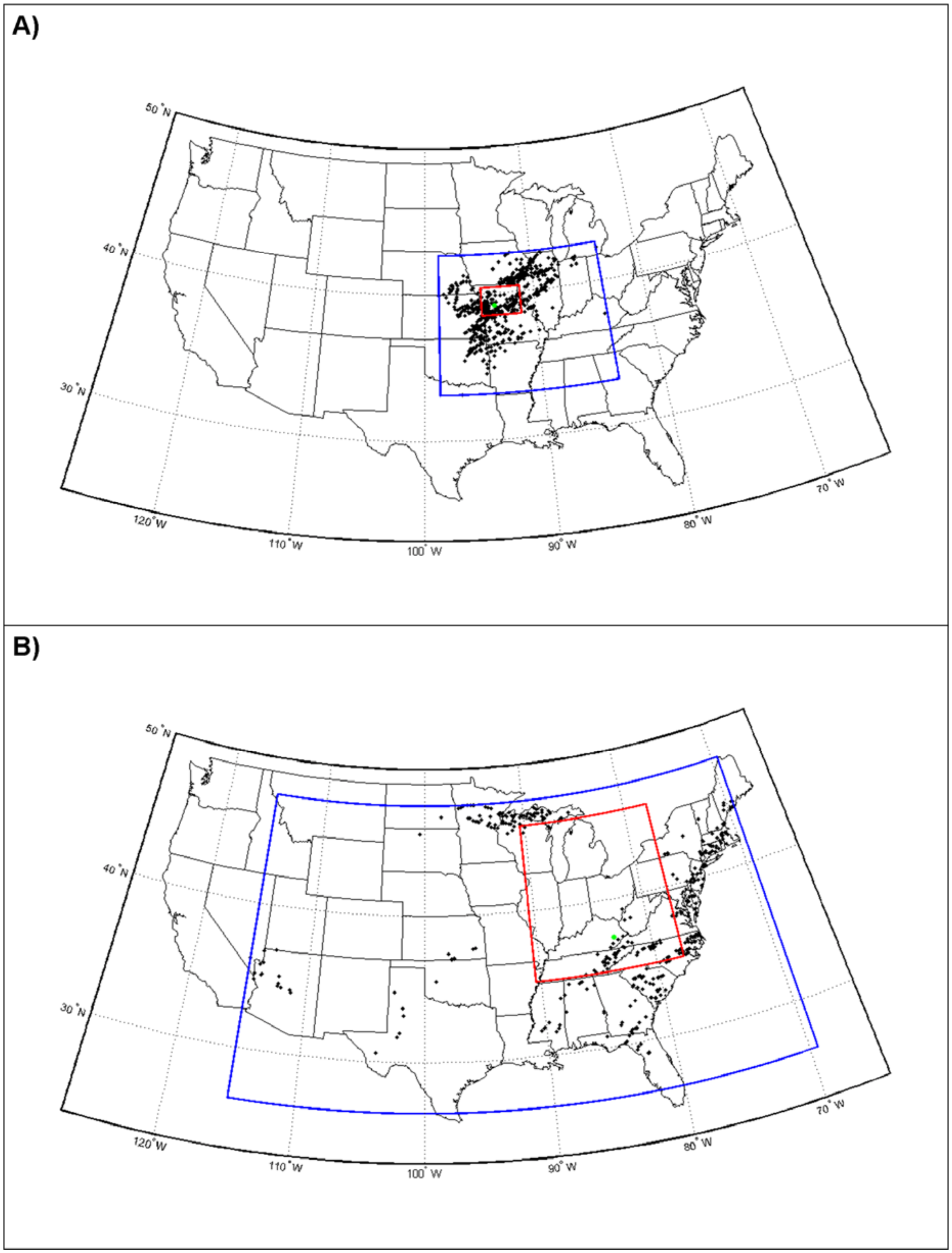


Figure 2: All severe reports for: a) 30 May 2006, b) 1 May 1997, and c) 26 June 2002. Green dots indicate large hail reports, blue plus signs indicate severe wind reports, and red dots (lines) indicate tornado reports (tracks). *Click image to enlarge.*



<u>Figure 3</u>: Examples of the middle-50% parameter for: a) 12 March 2006 and b) 28 July 2006. The blue box indicates the maximum and minimum latitudes and longitudes of severe weather reports. The red box indicates the 25th and 75th percentiles of the reports' latitudes and longitudes. The green dot indicates the median latitude and median longitude of severe reports. *Click image to enlarge*.

claiming that this procedure is optimal. However, the success of D06 in developing indices to identify the most significant tornado and primarily nontornadic outbreaks provides support for their implementation in this study, as the technique is reproducible and in accordance with subjective preconceptions.

3. Severe weather outbreak day rankings

a. Analyzing the robustness of the ranking schemes

Various combinations of weights for the variables in Table 1 were developed to compute scores for each outbreak day using 26 ranking indices (Fig. 4). Parameters thought to be of greater meteorological significance (e.g., DPI; path length; significant tornadoes; significant hail; significant severe winds; etc.) commonly were given high weights. The severe hail and wind report variables always were given relatively low weights. The indices are labeled N0-N25 and feature the following characteristics:

- **N0**: All variables were given equal weights (control).
- **N1-N2**: All variables in Table 1 were included, with relatively higher weights given to the tornado parameters. Subtle differences were incorporated in the weights of specific tornado parameters.
- **N3-N4**: N3 (N4) featured the same weights as N1 (N2), except higher for the significant hail and wind reports.
- **N5-N8**: Same weights as N1-N4, except the hail variables were removed.
- **N9-N12**: Same weights as N1-N4, except the wind variables were removed.
- **N13-N16**: N13 was the maximum score of N1, N5, and N9, so as to counteract negative biases for outbreaks in which one type of nontornadic report was dominant (e.g., Fig. 5). N14 was the maximum of N2, N6, and N10, and so on.
- **N17-N19**: All but two tornado variables were removed. The highest weight was given to significant tornadoes (N17), DPI (N18), and total path length (N19). The other tornado variable retained was the number of tornadoes, which was given a smaller weight.
- **N20**: Similar to N3 and N4, with subtle differences in weights for the tornado parameters.
- **N21-N22**: Similar to N18, with differences in the weights for total numbers of all reports, tornado reports, and significant nontornadic reports. DPI was emphasized because of its particular capability of ranking outbreaks similarly to indices including all tornado variables (see below).
- **N23-N25**: Same as N22, except the hail (N23) or wind (N24) parameters are removed. N25 is the maximum score of N22, N23, and N24.

The computation of such a large number of ranking indices allows for investigation into the volatility of rankings of the outbreaks when altering weights of specific parameters, with emphasis on significant hail and wind reports (e.g., comparing N1 and N3), use of a small versus a large number of tornado variables (e.g., comparing N1-N4 versus N17-N19), and possible biases for cases with a dominant storm type versus those with no clear dominant storm type (e.g., comparing N1, N5, N9, and N13).

Only a subset of the indices is included in the following discussion, figures, and tables. The reasons for this reduction include:

1. Indices N5-N8 (N9-N13) featured positively biased outbreak days dominated by wind (hail) reports and negatively biased outbreak days dominated by hail (wind) reports. This result is by design to create the "maximum" indices N13-N16. Indices N23 and N24 are also excluded for this reason.
2. Analysis of a subset of the indices sufficiently indicates the basic behavior of the multivariate indices, including the volatility of the rankings, the variability of the scores, and the sensitivities of the scores to altering weights of specific parameters. Redundancies of these characteristics prompted the exclusion of N20 (similar to N1-N4) and N21 (similar to N22).
3. Readability of figures and tables becomes increasingly cumbersome with a larger number of indices included.
4. Inclusion of N0 (all variables equally weighted – see Fig. 4) allows for comparison to a "control" index.

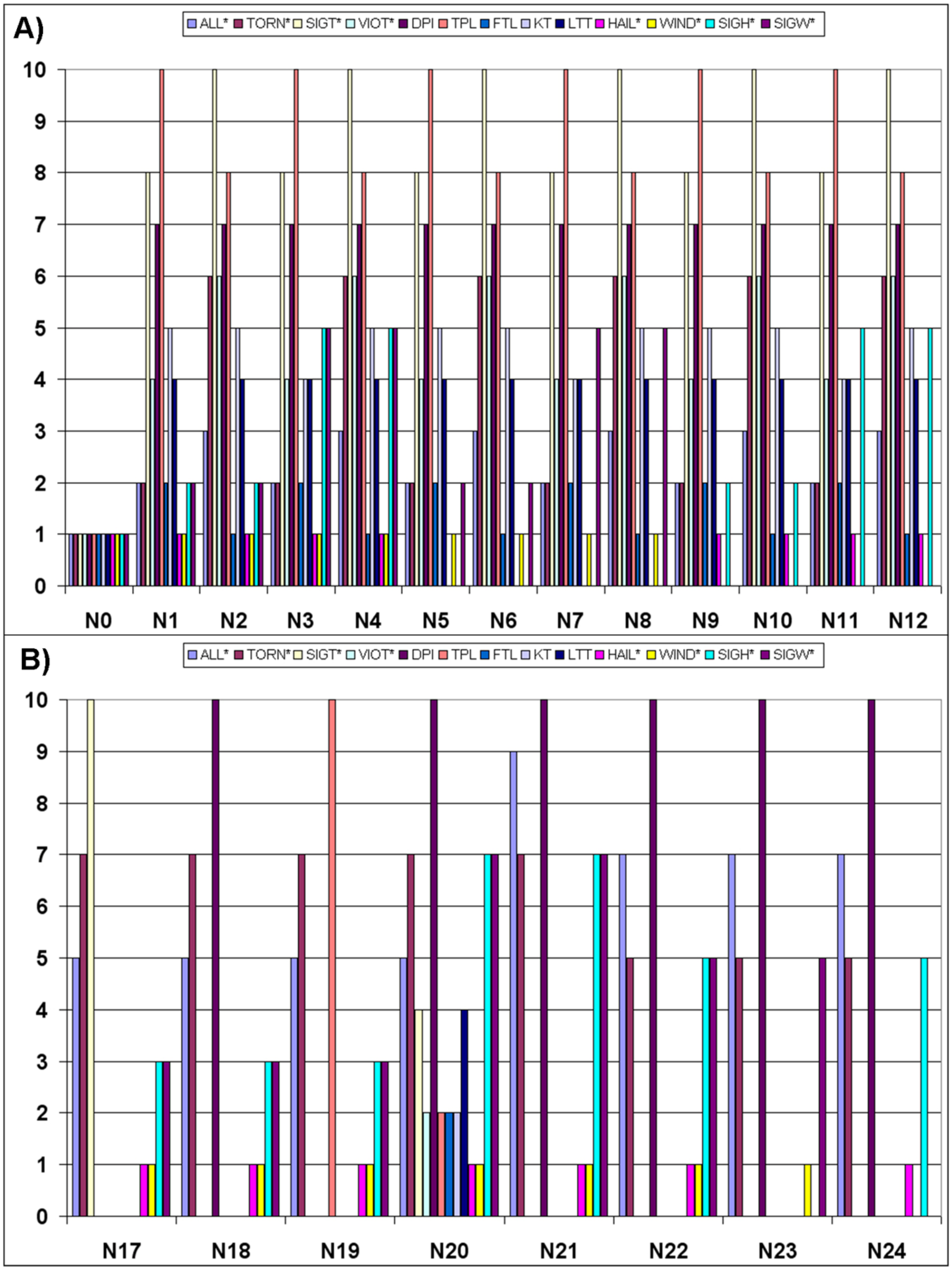


Figure 4: Weights for the variables listed in Table 1 for each index as described in the text. The "maximum" indices are not included; as these are determined by the maximum scores of three indices (see text for more information). An asterisk (*) indicates a detrended variable.

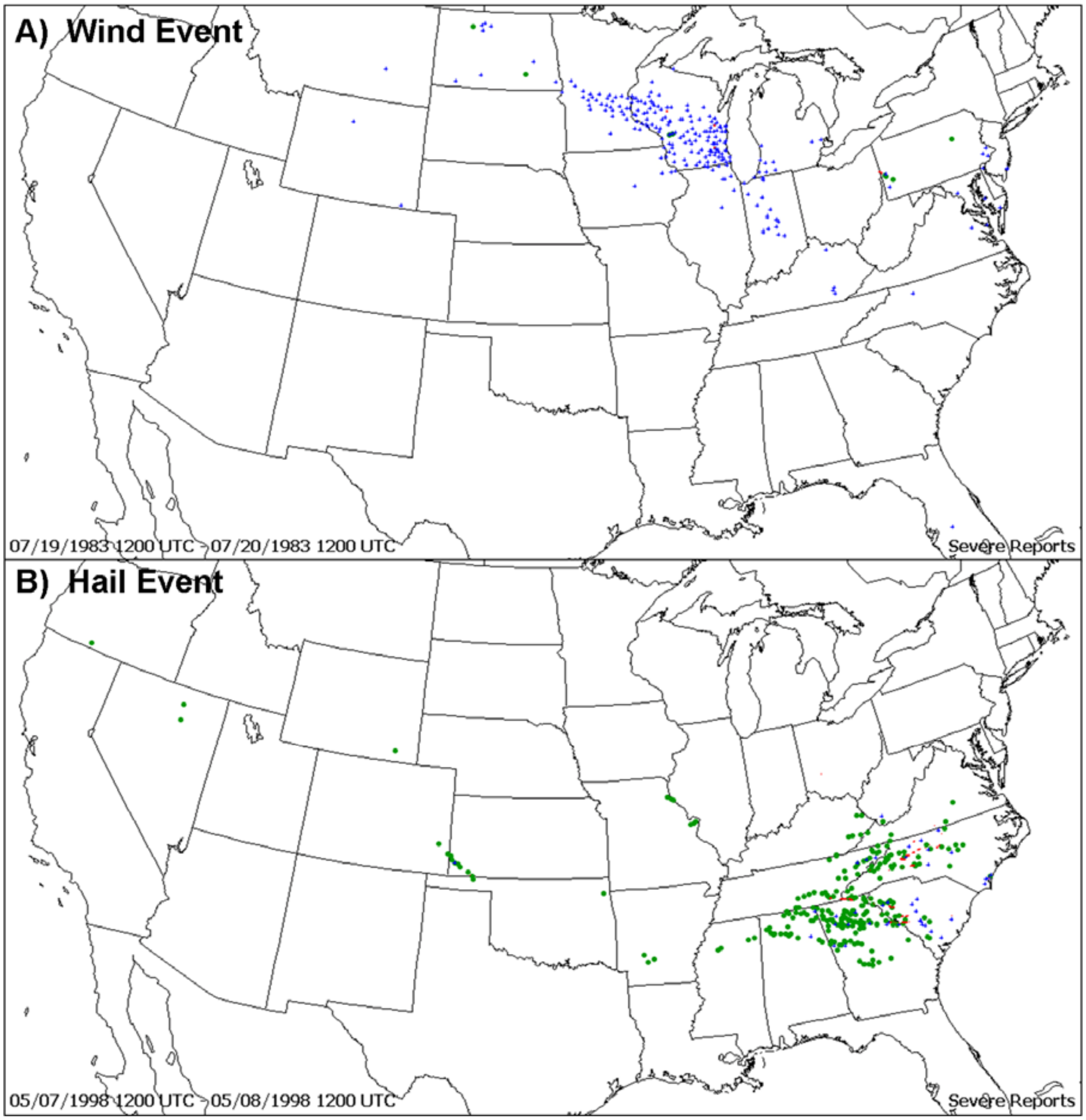


Figure 5: Example of cases that feature: a) primarily severe wind reports (19 July 1983) and b) primarily severe hail reports (7 May 1998). Plotting conventions as in Fig. 2. *Click image to enlarge.*

Investigating the resistance of the rankings for each combination of weighted parameters involved two basic procedures. The first considered the basic "structure" of the scores for each of the days, with the hope of observing that the scores for the different indices followed similar patterns from highest to lowest ranking. The second procedure involved analysis of particular portions of the rankings among the various indices to determine the volatility of the rankings for particular cases.

The scores for each case *for each of the indices in Fig. 4 and the "maximum indices" described above* clearly follow the same basic pattern (Fig. 6). The first 100-200 cases generally feature a sharp decline from very high values (scores of 10 to 20) to scores of about 1, after which the slope of the curve decreases markedly. The middle 1000 cases feature scores between 1 and –1, with a relatively small slope for each of the indices. The magnitudes of the slopes increase substantially again for the final

200 cases. Most of the variability regarding the magnitudes of the scores is found in the highest ranked cases; however, as is shown below, the cases included in this steeply-sloped portion of the curve are quite similar no matter which index is used. The specific ordering changes with different weighting schemes, but the first 100-200 cases remain highly ranked by *all* the weighting choices. This is also true for the other steeply-sloped portion of the curves (i.e., for the low-ranked cases). In between, where the range of variation of the index between cases is relatively small, the rankings generally are volatile, being strongly dependent on the choice of weights. However, this segment of the population also tends to remain within this zone of the distribution. Intermediate cases do not get "promoted" to the highly-ranked section of the curve, nor are they "demoted" to the low-ranked section by different choices for the parameter weights. Outbreaks within each of the three sections have distinctive characteristics (sections 3b and 4), which could have important implications for forecasting future severe weather outbreaks (section 5).

Computation of the sorted sums of the i^{th} value of 14 random normal distributions (the number of variables used to develop the indices) of 1410 values (the number of cases considered) results in a characteristic curve (not shown) similar in appearance to the sorted scores for each outbreak day (Fig. 6). The characteristic curves of the sorted scores from the multivariate indices feature a smaller slope in the middle and steeper slopes at the extremes (especially the high-end severe weather days). These differences suggest that correlations among the variables considered are influencing the results.

A correlation matrix of the *standardized variables* used in the development of the multivariate indices (Table 2) indicates considerable positive correlations among most of the parameters, particularly the tornado variables. Specifically, correlations among the number of fatalities, the number of killer tornadoes, the number of F4 or greater tornadoes, total path length, and DPI are all greater than 0.6. Correlations are slightly lower between the number of tornadoes and other tornado variables, suggesting the inclusion of one tornado variable in addition to the total number of tornadoes (e.g., indices N17-N19, N21-N25) is valuable. Correlations between the total number of wind (hail) reports and the total number of significant wind (hail) reports are substantially higher than between wind and hail reports. In general, there is little correlation between tornado and nontornadic variables and little correlation between the middle 50% parameter and other variables.

The above results may be indicating that the relatively steep slope of the high-end portion of the multivariate index scores is a result of the inclusion of several highly correlated tornado variables. Interestingly, indices N17-N19 and N21-N25 appear to be displaced upward near the transition between major and intermediate outbreaks (Fig. 6). These were the indices in which a number of the tornado variables were removed (Fig. 4). The relatively increased slope at the low end of the intermediate outbreaks and the more gradual increasing slope of the outbreak days with large scatter may be indicative of the middle 50% parameter having little correlation with other variables.

b. Characteristics of the three groups of severe weather outbreaks

Rankings of the top 25 outbreak days (Table 3) and the bottom 25 outbreak days (Table 4) for some of the indices listed in Table 1 indicate modest variability of case rankings for the top 25 days and very little variability for the bottom 25 days. No matter which index is analyzed, including those not shown in Table 3, the 3 April 1974 tornado outbreak (Fig. 7a; see also Fujita 1974) always is positioned highest. Similarly, the 6 June 1981 outbreak day (Fig. 7b) is positioned lowest for all of the indices.

Indices N0 and N13-N16 strongly favor scoring tornado outbreaks highest (cf. Table 5, D06). Many of these outbreaks are well-known for their large societal impacts, including the 11 April 1965 (Fujita et al. 1970), 31 May 1985 (Ferguson et al. 1987), 26 April 1991 (Johns and Hart 1993), 3 May 1999 (Thompson and Edwards 2000), and 4 May 2003 (Lindsey and Bunkers 2005) tornado outbreaks. The positioning of these outbreaks at the top of the ranking scheme is desirable and expected, since many of the parameters used in the calculation of these indices involve the number, intensity, and societal impacts of tornadoes.

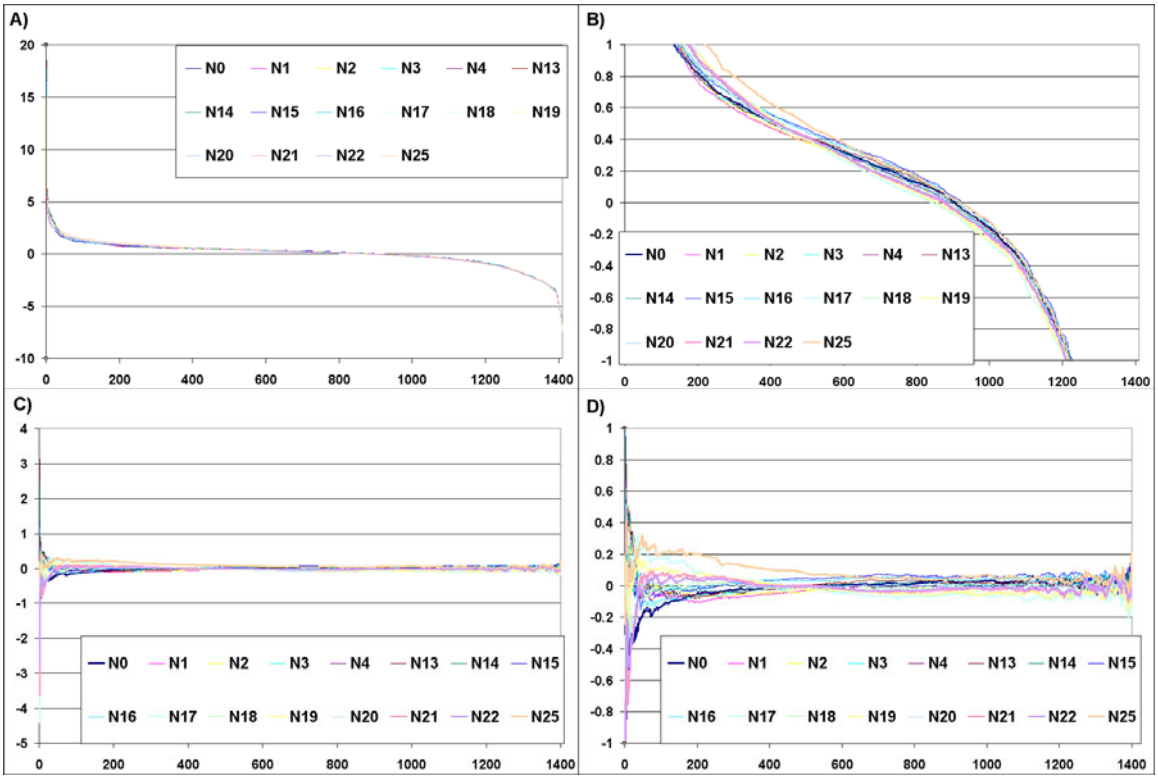


Figure 6: Plots of: a) scores for each outbreak day in order from highest ranking (1) to lowest ranking (1410) for each of the indices listed in Table 1; b) zoom-in on the index scores in the range -1 to 1 — note the different scale on the ordinate; c) deviations for each index from the mean score for each of the rankings, d) zoom-in on the deviations ranging from -1 to 1. Indices N0 (control), N17-N19, N21-N22, and N25 (indices with two tornado variables) are emphasized in (b)-(d) for clarity. *Click image to enlarge.*

However, high-end severe weather outbreaks with relatively few tornadoes predominantly are excluded from the highest rankings for these indices. To counteract this tendency, several of the tornado parameters were removed from indices N17-N19 and N21-N25; however, the remaining tornado parameters were given relatively high weights compared to the nontornadic parameters. Interestingly, many cases listed in the top 25 for these indices are similar to those of "tornado-weighted" indices, indicating the overall robustness of the ranking scheme developed, even with removal of many of the parameters. However, a few cases characterized by a very large number of total severe reports and significant severe weather reports appear in the top 25 lists for these indices. Some of these cases include 14 April 1974 (Fig. 8a), 20 June 1974 (Fig. 8b), and 30 May 1998 (Fig. 8c). Other days that are ranked higher (but not necessarily in the top 25) because of the exclusion of numerous tornado parameters include 19 July 1983, 1 July 1994, 21 April 1996 (Fig. 8d), and 15 May 1998. Many of these events feature a very large number of significant wind (e.g., 20 June 1974) or significant hail (e.g., 21 April 1996) reports. Comparison of indices N17-N19, N22, and N25 indicates that retaining DPI results in the inclusion of tornado outbreaks in the top 25 that more closely resemble the rankings created by maintaining a larger number of tornado parameters.

The strong resistance to the change in rankings for the bottom 25 cases (Table 4) can be attributed to the effectiveness of the middle-50% parameter. The middle-50% parameter's standardized scores can be quite large if the tabulated latitude-by-longitude area is very large. The appearance of the cases in the bottom 25 lists for each of the indices suggests mostly widely scattered severe reports (e.g., Figs. 9a,b), or multiple clusters of severe reports (e.g., Figs. 9c,d). Most of these cases occured in the summer months.

Table 2: Correlation matrix of the 14 standardized variables used in the calculation of the indices. Abbreviations as in Table 1.

	FTL	**KT**	**LTT**	**DPI**	**TPL**	**M50**	**ALL**
FTL	1	0.8936	0.58	0.7825	0.6659	-0.052	0.3418
KT	0.8936	1	0.6185	0.8151	0.7806	-0.0819	0.3975
LTT	0.58	0.6185	1	0.6845	0.7446	-0.0785	0.3041
DPI	0.7825	0.8151	0.6845	1	0.7932	-0.0904	0.3162
TPL	0.6659	0.7806	0.7446	0.7932	1	-0.1285	0.4072
M50	-0.052	-0.0819	-0.0785	-0.0904	-0.1285	1	-0.0597
ALL	0.3418	0.3975	0.3041	0.3162	0.4072	-0.0597	1
TORN	0.4498	0.5718	0.4679	0.5581	0.762	-0.1356	0.5136
WIND	0.1228	0.142	0.0728	0.0702	0.0925	0.02	0.7376
HAIL	0.1695	0.2075	0.1689	0.1779	0.2222	-0.0674	0.5676
SIGW	0.0271	0.0201	0.0336	-0.0012	0.0475	-0.0187	0.3642
SIGH	0.1669	0.2026	0.1683	0.2085	0.2415	-0.0858	0.3895
F2	0.5096	0.6747	0.5109	0.6634	0.8524	-0.1503	0.3969
F4	0.8128	0.8392	0.5863	0.8434	0.7476	-0.0795	0.3209

	TORN	**WIND**	**HAIL**	**SIGW**	**SIGH**	**F2**	**F4**
FTL	0.4498	0.1228	0.1695	0.0271	0.1669	0.5096	0.8128
KT	0.5718	0.142	0.2075	0.0201	0.2026	0.6747	0.8392
LTT	0.4679	0.0728	0.1689	0.0336	0.1683	0.5109	0.5863
DPI	0.5581	0.0702	0.1779	-0.0012	0.2085	0.6634	0.8434
TPL	0.762	0.0925	0.2222	0.0475	0.2415	0.8524	0.7476
M50	-0.1356	0.02	-0.0674	-0.0187	-0.0858	-0.1503	-0.0795
ALL	0.5136	0.7376	0.5676	0.3642	0.3895	0.3969	0.3209
TORN	1	0.1446	0.2519	0.0977	0.2823	0.8155	0.5746
WIND	0.1446	1	-0.0338	0.4197	-0.0728	0.1024	0.0764
HAIL	0.2519	-0.0338	1	0.0527	0.7171	0.2047	0.1799
SIGW	0.0977	0.4197	0.0527	1	0.089	0.027	0.0185
SIGH	0.2823	-0.0728	0.7171	0.089	1	0.2245	0.2035
F2	0.8155	0.1024	0.2047	0.027	0.2245	1	0.6851
F4	0.5746	0.0764	0.1799	0.0185	0.2035	0.6851	1

The most variability with respect to rankings of case days occurred for the middle 1000 cases. The index values for these cases generally produced only small differences in the scores for different rankings. For example, for index N25, the score for position 400 is 0.626, for position 500 is 0.494, and for position 600 is 0.378. These differences are quite small compared to the differences in the top 200 and bottom 200 cases. For example, the score for position 10 is 3.559, for position 30 is 2.525, and for position 50 is 1.994. This suggests the following alternative explanations (not necessarily exclusive):

1. The ranking scheme developed in this study is not able to rank robustly the "intermediate outbreak" days compared to the most severe outbreaks and those days with widely scattered or multiple geographic clusters of reports (i.e., days that could be classified as "non-outbreak" days despite a relatively high number of reports).
2. The rarity of high-end severe weather outbreaks compared to the relatively frequent occurrence of intermediate severe weather outbreaks makes ranking the intermediate cases more difficult, given the aforementioned limitations of archived storm reports (Section 1).
3. Forecasting the relative severity of intermediate severe weather outbreaks may not be as feasible as forecasting the nature of the severe reports for these outbreaks (Section 4).

Table 3**:** List of rankings and scores for the top 25 outbreaks for some indices identified in Fig. 4 and the text. The dates 3 April 1974, 4 May 2003, 21 January 1999, and 2 April 2006 are in boldface to illustrate variability in the rankings.

	N0		**N13**		**N14**		**N15**		**N16**
3 Apr 1974	**15.17**	**3 Apr 1974**	**18.47**	**3 Apr 1974**	**17.84**	**3 Apr 1974**	**16.71**	**3 Apr 1974**	**16.50**
11 Apr 1965	7.17	11 Apr 1965	8.15	11 Apr 1965	7.54	11 Apr 1965	7.32	11 Apr 1965	6.93
31 May 1985	4.15	31 May 1985	5.50	31 May 1985	5.23	31 May 1985	5.04	31 May 1985	4.86
5 May 1960	3.64	12 Mar 2006	5.26	12 Mar 2006	4.91	12 Mar 2006	4.97	12 Mar 2006	4.65
12 Mar 2006	3.57	**4 May 2003**	**4.36**	**4 May 2003**	**4.52**	26 Apr 1991	4.09	**4 May 2003**	**4.24**
26 Apr 1991	3.43	13 Mar 1990	4.32	26 Apr 1991	4.39	**4 May 2003**	**4.07**	26 Apr 1991	4.19
4 May 2003	**3.40**	26 Apr 1991	4.27	13 Mar 1990	4.30	13 Mar 1990	4.00	13 Mar 1990	3.97
28 Mar 1984	3.22	28 Mar 1984	4.26	28 Mar 1984	4.17	5 May 1960	3.83	28 Mar 1984	3.83
10 Nov 2002	3.15	10 Nov 2002	4.09	2 Jun 1990	4.12	28 Mar 1984	3.83	5 May 1960	3.77
15 May 1968	3.15	5 May 1960	3.99	10 Nov 2002	4.08	10 Nov 2002	3.70	10 Nov 2002	3.77
20 Jun 1974	2.94	27 Mar 1994	3.85	5 May 1960	3.89	27 Mar 1994	3.58	2 Jun 1990	3.72
13 Mar 1990	2.91	2 Jun 1990	3.83	3 May 1999	3.86	3 May 1999	3.46	3 May 1999	3.62
3 May 1999	2.88	3 May 1999	3.71	22 Nov 1992	3.67	2 Jun 1990	3.44	**21 Jan 1999**	**3.46**
5 May 1964	2.88	22 Nov 1992	3.71	**21 Jan 1999**	**3.66**	22 Nov 1992	3.32	27 Mar 1994	3.36
2 Apr 2006	**2.71**	**21 Jan 1999**	**3.51**	27 Mar 1994	3.58	**21 Jan 1999**	**3.31**	22 Nov 1992	3.32
2 Apr 1982	2.71	27 May 1973	3.31	2 Apr 1982	3.34	5 May 1964	3.16	5 May 1964	3.06
21 Jan 1999	**2.56**	17 Apr 1970	3.31	27 May 1973	3.18	17 Apr 1970	3.10	2 Apr 1982	2.99
2 Jun 1990	2.54	2 Apr 1982	3.18	7 Jun 1984	3.16	**2 Apr 2006**	**3.01**	**2 Apr 2006**	**2.96**
27 Mar 1994	2.54	7 Jun 1984	3.17	8 Apr 1999	3.10	27 Mar 1973	2.98	16 Jun 1992	2.94
17 Apr 1970	2.49	5 May 1964	3.15	16 Jun 1992	3.09	7 Jun 1984	2.90	7 Jun 1984	2.90
23 Apr 1968	2.42	**2 Apr 2006**	**3.08**	5 May 1964	3.09	8 Apr 1999	2.80	27 May 1973	2.88
21 Apr 1967	2.36	8 Apr 1999	3.07	**2 Apr 2006**	**3.02**	2 Apr 1982	2.77	8 Apr 1999	2.83
20 Mar 1976	2.34	15 May 1968	2.87	17 Apr 1970	2.98	15 May 1968	2.65	17 Apr 1970	2.82
18 May 1995	2.26	23 Nov 2004	2.75	18 May 1995	2.93	16 Jun 1992	2.63	18 May 1995	2.82
22 Nov 1992	2.24	16 Jun 1992	2.72	15 May 1968	2.88	18 May 1995	2.57	15 May 1968	2.71

Table 3**:** *(Continued)*

	N17		N18		N19		N22		N25
3 Apr 1974	**10.91**	**3 Apr 1974**	**13.53**	**3 Apr 1974**	**12.98**	**3 Apr 1974**	**12.58**	**3 Apr 1974**	**15.08**
16 Jun 1992	4.45	11 Apr 1965	5.88	12 Mar 2006	5.50	11 Apr 1965	5.48	11 Apr 1965	6.54
12 Mar 2006	4.37	31 May 1985	4.70	11 Apr 1965	4.99	31 May 1985	4.28	31 May 1985	5.23
4 May 2003	**4.32**	27 Mar 1994	4.33	5 May 1960	3.79	20 Jun 1974	3.98	27 Mar 1994	4.81
21 Jan 1999	**4.17**	26 Apr 1991	3.82	**4 May 2003**	**3.77**	27 Mar 1994	3.93	20 Jun 1974	4.62
26 Apr 1991	3.98	28 Mar 1984	3.81	**21 Jan 1999**	**3.68**	5 May 1960	3.74	26 Apr 1991	4.36
10 Nov 2002	3.97	5 May 1960	3.74	**2 Apr 2006**	**3.67**	26 Apr 1991	3.68	28 Mar 1984	4.15
19 Apr 1996	3.91	20 Jun 1974	3.51	13 Mar 1990	3.52	5 May 1964	3.45	5 May 1960	4.01
20 Jun 1974	3.59	13 Mar 1990	3.44	20 Jun 1974	3.50	28 Mar 1984	3.42	5 May 1964	3.83
18 May 1995	3.57	5 May 1964	3.41	30 May 2004	3.40	13 Mar 1990	3.06	13 Mar 1990	3.56
2 Apr 2006	**3.53**	3 May 1999	3.38	10 Nov 2002	3.29	3 May 1999	2.96	3 May 1999	3.54
11 Apr 1965	3.48	17 Apr 1970	3.12	5 May 1964	3.23	17 Apr 1970	2.95	20 Mar 1976	3.42
5 May 1960	3.33	20 Mar 1976	3.06	26 Apr 1991	3.22	20 Mar 1976	2.85	17 Apr 1970	3.40
13 Mar 1990	3.28	18 May 1995	2.91	23 Nov 2004	3.20	15 May 1968	2.70	15 May 1968	3.35
2 Jun 1990	3.20	**21 Jan 1999**	**2.81**	8 May 1988	3.10	24 May 1962	2.70	18 May 1995	3.21
2 Apr 1982	3.13	7 May 1993	2.76	20 Mar 1976	3.02	18 May 1995	2.64	24 May 1962	3.18
3 May 1999	3.09	8 Jun 1974	2.72	18 May 1995	2.97	**2 Apr 2006**	**2.54**	8 Jun 1974	3.12
20 Mar 1976	2.96	15 May 1968	2.72	6 May 2003	2.84	8 Jun 1974	2.52	**21 Jan 1999**	**2.99**
7 Jun 1984	2.92	**4 May 2003**	**2.62**	19 Apr 1996	2.81	**21 Jan 1999**	**2.47**	7 May 1993	2.87
24 Jun 2003	2.85	27 May 1973	2.54	29 May 2004	2.79	19 Apr 1996	2.47	19 Apr 1996	2.86
29 May 2004	2.84	19 Apr 1996	2.52	15 May 1968	2.78	7 May 1993	2.44	23 Apr 1968	2.84
23 Nov 2004	2.81	**2 Apr 2006**	**2.50**	7 Jun 1984	2.74	**4 May 2003**	**2.38**	26 Apr 1984	2.76
10 May 2003	2.77	21 Apr 1976	2.42	7 Apr 2006	2.69	23 Apr 1968	2.34	**4 May 2003**	**2.71**
15 May 1968	2.76	2 Apr 1982	2.36	2 Jun 1990	2.63	30 May 1998	2.29	12 Mar 2006	2.69
7 Apr 1980	2.75	10 Nov 2002	2.33	2 Apr 1982	2.60	14 Apr 1974	2.29	**2 Apr 2006**	**2.67**

Table 4: List of rankings and scores for the bottom 25 cases for some indices identified in Fig. 4 and the text. The cases 3 July 1960, 5 June 1973, 2 August 1993, and 6 June 1981 are in boldface to illustrate the variability in the rankings.

	N17		**N18**		**N19**		**N22**		**N25**
9 Jul 1975	-3.55	9 Jul 1975	-3.44	20 Jun 1969	-3.51	20 Jun 1969	-3.44	23 Jun 1974	-3.40
20 Jun 1969	-3.55	20 Jun 1969	-3.45	9 Jul 1975	-3.51	9 Jul 1975	-3.48	9 Jul 1975	-3.43
14 Jun 1963	-3.61	6 Jun 1963	-3.51	28 Jul 2006	-3.57	28 Jul 2006	-3.49	28 Jul 2006	-3.43
28 Jul 2006	-3.62	28 Jul 2006	-3.52	14 Jun 1963	-3.65	6 Jun 1963	-3.56	6 Jun 1963	-3.48
1 Jul 2005	-3.77	14 Jun 1963	-3.58	6 Jun 1963	-3.68	14 Jun 1963	-3.62	14 Jun 1963	-3.56
6 Jun 1963	-3.78	1 Jul 2005	-3.67	1 Jul 2005	-3.74	27 May 1965	-3.66	1 Jul 2005	-3.65
27 May 1965	-3.80	27 May 1965	-3.69	27 May 1965	-3.76	1 Jul 2005	-3.67	27 May 1965	-3.66
13 May 1980	-3.93	13 May 1980	-3.86	13 May 1980	-3.90	8 Jun 1961	-3.83	13 May 1980	-3.77
8 Jun 1961	-3.94	8 Jun 1961	-3.90	8 Jun 1961	-3.97	13 May 1980	-3.88	8 Jun 1961	-3.83
17 May 1986	-4.23	17 May 1986	-4.12	17 May 1986	-4.17	17 May 1986	-4.15	17 May 1986	-4.05
16 Jun 1971	-4.35	16 Jun 1971	-4.24	16 Jun 1971	-4.24	16 Jun 1971	-4.17	16 Jun 1971	-4.17
3 Jul 1960	**-4.69**	**3 Jul 1960**	**-4.58**	**3 Jul 1960**	**-4.65**	**3 Jul 1960**	**-4.49**	**3 Jul 1960**	**-4.35**
14 Jul 1969	-4.79	14 Jul 1969	-4.73	14 Jul 1969	-4.75	14 Jul 1969	-4.70	14 Jul 1969	-4.50
5 Jun 1973	**-4.87**	**5 Jun 1973**	**-4.76**	**5 Jun 1973**	**-4.82**	**5 Jun 1973**	**-4.77**	**5 Jun 1973**	**-4.69**
16 May 1965	-5.00	16 May 1965	-4.97	16 May 1965	-5.03	16 May 1965	-5.03	16 May 1965	-4.95
15 Jun 1987	-5.42	15 Jun 1987	-5.31	15 Jun 1987	-5.39	6 Aug 1989	-5.30	6 Aug 1989	-5.21
6 Aug 1989	-5.44	6 Aug 1989	-5.33	2 Aug 1986	-5.39	15 Jun 1987	-5.32	15 Jun 1987	-5.24
2 Aug 1986	-5.44	2 Aug 1986	-5.33	6 Aug 1989	-5.39	2 Aug 1986	-5.33	2 Aug 1986	-5.24
2 Aug 1993	**-5.88**	**2 Aug 1993**	**-5.78**	**2 Aug 1993**	**-5.83**	**2 Aug 1993**	**-5.77**	**2 Aug 1993**	**-5.69**
8 Jun 1971	-5.90	8 Jun 1971	-5.88	8 Jun 1971	-5.90	8 Jun 1971	-5.85	8 Jun 1971	-5.83
20 Jun 1977	-5.98	20 Jun 1977	-5.92	20 Jun 1977	-5.95	20 Jun 1977	-6.05	20 Jun 1977	-5.99
25 Jun 1988	-6.13	25 Jun 1988	-6.09	25 Jun 1988	-6.15	25 Jun 1988	-6.13	25 Jun 1988	-6.03
14 Jun 1966	-6.38	14 Jun 1966	-6.36	14 Jun 1966	-6.43	14 Jun 1966	-6.32	14 Jun 1966	-6.25
18 Jun 1972	-6.71	18 Jun 1972	-6.89	18 Jun 1972	-6.92	18 Jun 1972	-7.02	18 Jun 1972	-6.93
6 Jun 1981	**-7.55**	**6 Jun 1981**	**-7.50**	**6 Jun 1981**	**-7.50**	**6 Jun 1981**	**-7.57**	**6 Jun 1981**	**-7.52**

Table 4: *(Continued)*

	N17		**N18**		**N19**		**N22**		**N25**
9 Jul 1975	-3.55	9 Jul 1975	-3.44	20 Jun 1969	-3.51	20 Jun 1969	-3.44	23 Jun 1974	-3.40
20 Jun 1969	-3.55	20 Jun 1969	-3.45	9 Jul 1975	-3.51	9 Jul 1975	-3.48	9 Jul 1975	-3.43
14 Jun 1963	-3.61	6 Jun 1963	-3.51	28 Jul 2006	-3.57	28 Jul 2006	-3.49	28 Jul 2006	-3.43
28 Jul 2006	-3.62	28 Jul 2006	-3.52	14 Jun 1963	-3.65	6 Jun 1963	-3.56	6 Jun 1963	-3.48
1 Jul 2005	-3.77	14 Jun 1963	-3.58	6 Jun 1963	-3.68	14 Jun 1963	-3.62	14 Jun 1963	-3.56
6 Jun 1963	-3.78	1 Jul 2005	-3.67	1 Jul 2005	-3.74	27 May 1965	-3.66	1 Jul 2005	-3.65
27 May 1965	-3.80	27 May 1965	-3.69	27 May 1965	-3.76	1 Jul 2005	-3.67	27 May 1965	-3.66
13 May 1980	-3.93	13 May 1980	-3.86	13 May 1980	-3.90	8 Jun 1961	-3.83	13 May 1980	-3.77
8 Jun 1961	-3.94	8 Jun 1961	-3.90	8 Jun 1961	-3.97	13 May 1980	-3.88	8 Jun 1961	-3.83
17 May 1986	-4.23	17 May 1986	-4.12	17 May 1986	-4.17	17 May 1986	-4.15	17 May 1986	-4.05
16 Jun 1971	-4.35	16 Jun 1971	-4.24	16 Jun 1971	-4.24	16 Jun 1971	-4.17	16 Jun 1971	-4.17
3 Jul 1960	**-4.69**	**3 Jul 1960**	**-4.58**	**3 Jul 1960**	**-4.65**	**3 Jul 1960**	**-4.49**	**3 Jul 1960**	**-4.35**
14 Jul 1969	-4.79	14 Jul 1969	-4.73	14 Jul 1969	-4.75	14 Jul 1969	-4.70	14 Jul 1969	-4.50
5 Jun 1973	**-4.87**	**5 Jun 1973**	**-4.76**	**5 Jun 1973**	**-4.82**	**5 Jun 1973**	**-4.77**	**5 Jun 1973**	**-4.69**
16 May 1965	-5.00	16 May 1965	-4.97	16 May 1965	-5.03	16 May 1965	-5.03	16 May 1965	-4.95
15 Jun 1987	-5.42	15 Jun 1987	-5.31	15 Jun 1987	-5.39	6 Aug 1989	-5.30	6 Aug 1989	-5.21
6 Aug 1989	-5.44	6 Aug 1989	-5.33	2 Aug 1986	-5.39	15 Jun 1987	-5.32	15 Jun 1987	-5.24
2 Aug 1986	-5.44	2 Aug 1986	-5.33	6 Aug 1989	-5.39	2 Aug 1986	-5.33	2 Aug 1986	-5.24
2 Aug 1993	**-5.88**	**2 Aug 1993**	**-5.78**	**2 Aug 1993**	**-5.83**	**2 Aug 1993**	**-5.77**	**2 Aug 1993**	**-5.69**
8 Jun 1971	-5.90	8 Jun 1971	-5.88	8 Jun 1971	-5.90	8 Jun 1971	-5.85	8 Jun 1971	-5.83
20 Jun 1977	-5.98	20 Jun 1977	-5.92	20 Jun 1977	-5.95	20 Jun 1977	-6.05	20 Jun 1977	-5.99
25 Jun 1988	-6.13	25 Jun 1988	-6.09	25 Jun 1988	-6.15	25 Jun 1988	-6.13	25 Jun 1988	-6.03
14 Jun 1966	-6.38	14 Jun 1966	-6.36	14 Jun 1966	-6.43	14 Jun 1966	-6.32	14 Jun 1966	-6.25
18 Jun 1972	-6.71	18 Jun 1972	-6.89	18 Jun 1972	-6.92	18 Jun 1972	-7.02	18 Jun 1972	-6.93
6 Jun 1981	**-7.55**	**6 Jun 1981**	**-7.50**	**6 Jun 1981**	**-7.50**	**6 Jun 1981**	**-7.57**	**6 Jun 1981**	**-7.52**

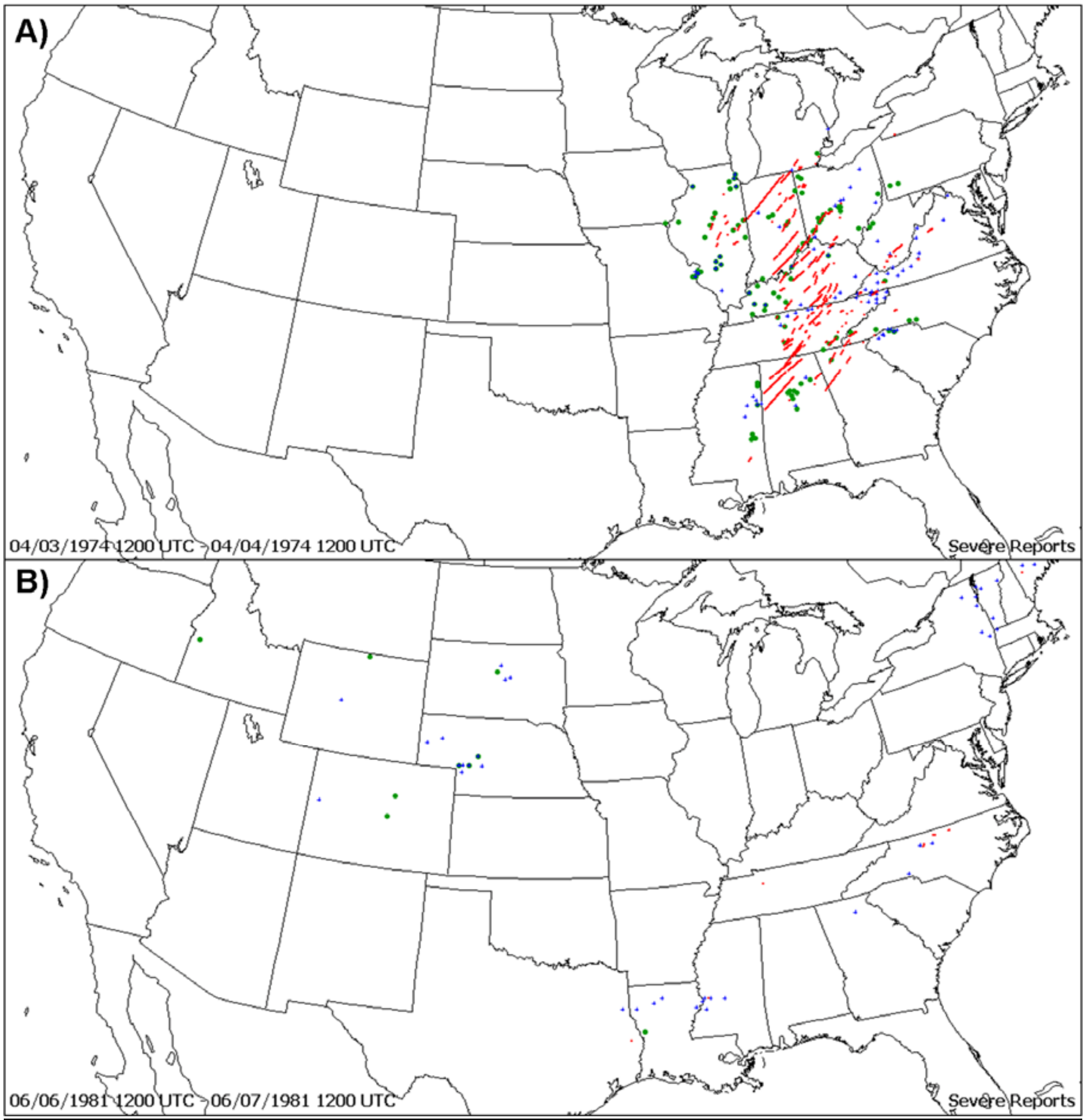


Figure 7: As in Fig. 2, for a) 3 April 1974 and b) 6 June 1981. *Click image to enlarge.*

A consideration of the case days ranked 399 (20 April 1995; Fig. 10a), 499 (17 April 1995; Fig. 10b), and 599 (2 May 1997; Fig. 10c) for index N25 indicates only subtle differences in the character of the outbreaks (including total number of reports, total number of significant events, and DPI; not shown). Table 5 illustrates the volatility of the rankings for these three cases with several of the indices presented. The range of scores for a particular case rarely exceeds 0.25, but the rankings can differ by as much as 100 to 200 positions. Furthermore, the robustness of the ranking schemes in terms of distinguishing the high-end events (the left portions of the curves in Fig. 6) from the intermediate events (the middle portions of the curves in Fig. 6) and from the potentially "non-outbreak days" (the right portions of the curves in Fig. 6) suggests that discriminating outbreaks in terms of these three classes may be useful, rather than using the value of the index to attempt to distinguish one case from another.

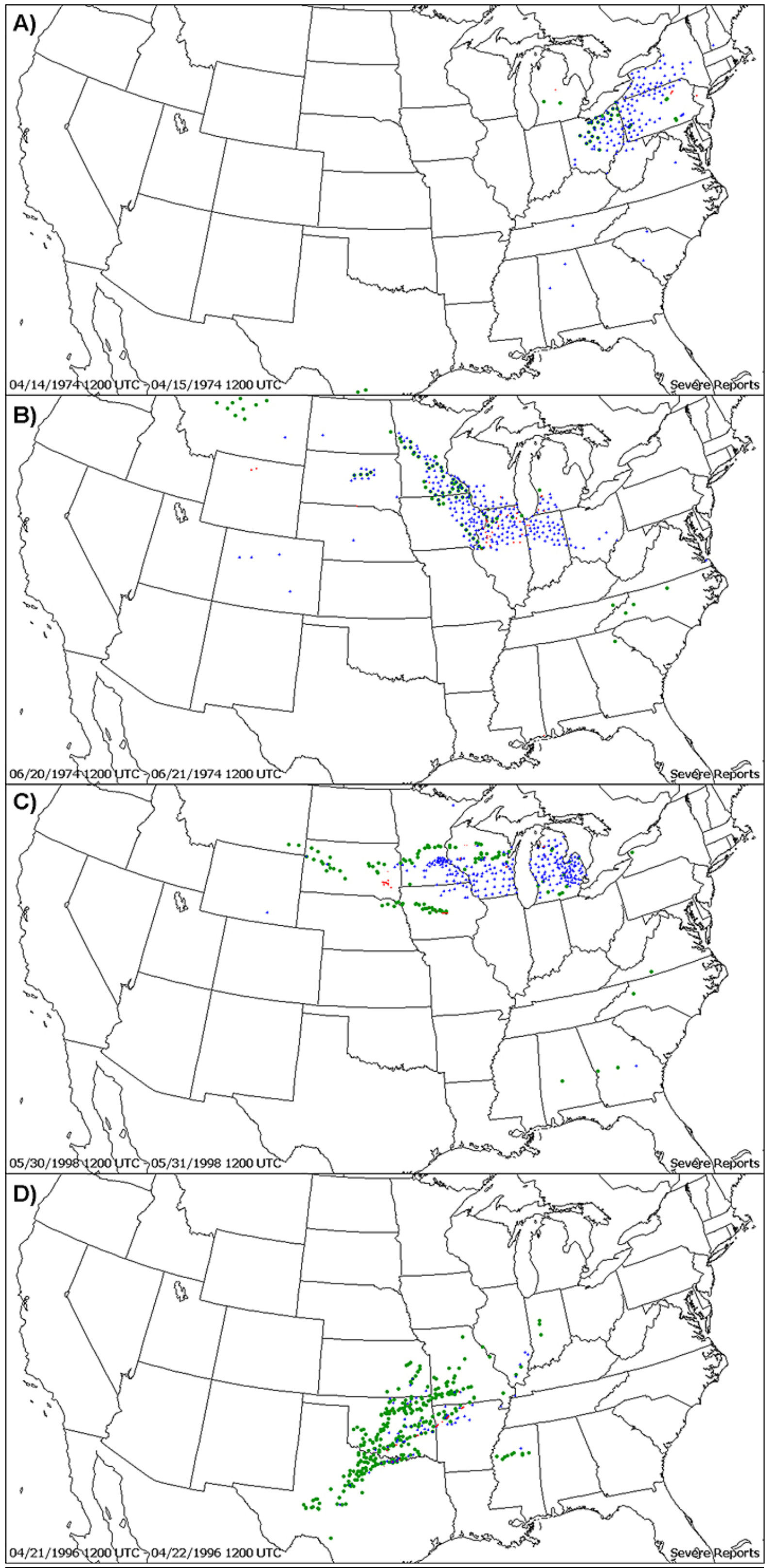


Figure 8: As in Fig. 2, for a) 14 April 1974, b) 20 June 1974, c) 30 May 1998, and d) 21 April 1996. *Click image to enlarge.*

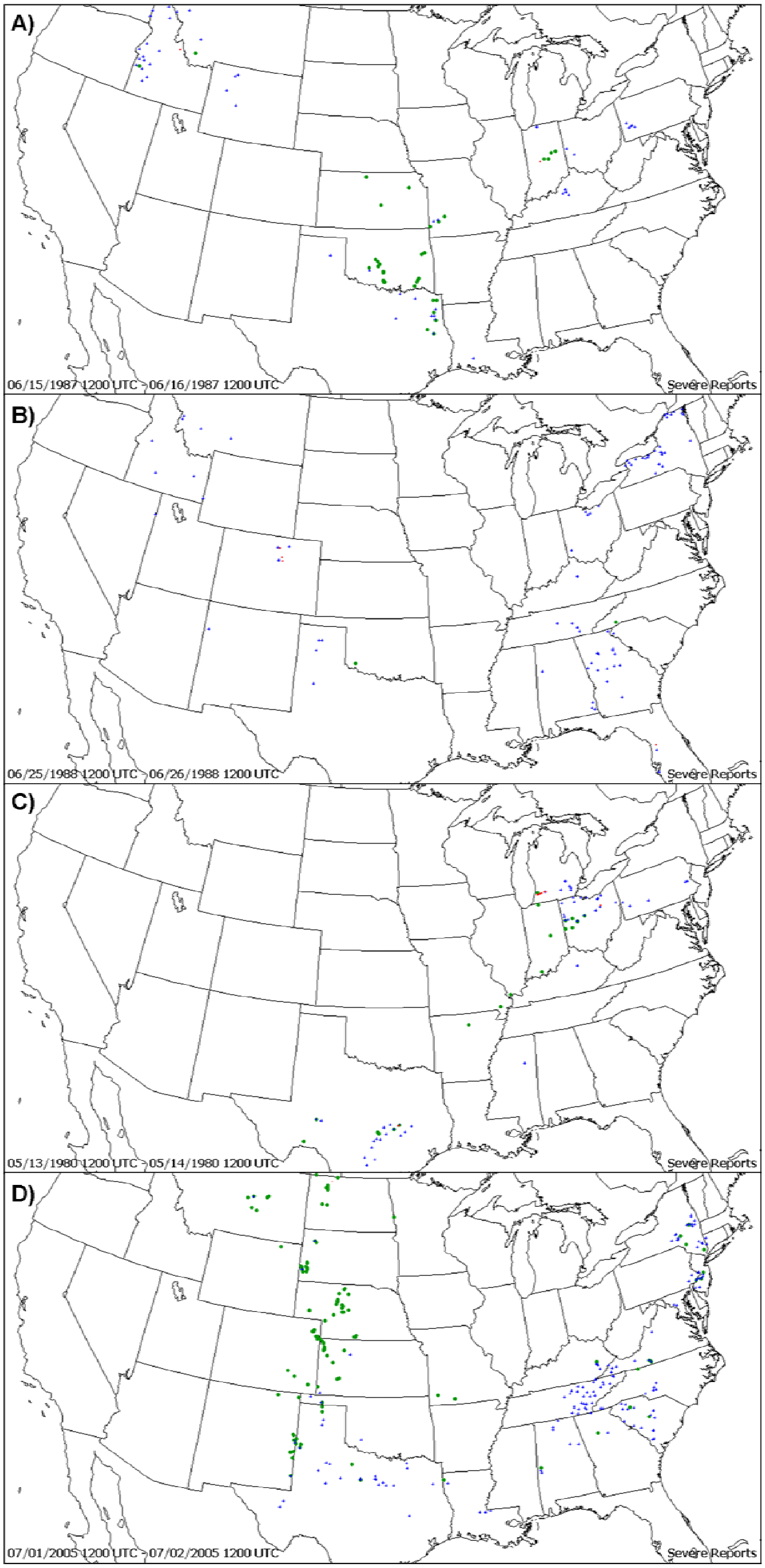


Figure 9: As in Fig. 2, for a) 15 June 1987, b) 25 June 1988, c) 13 May 1980, and d) 1 July 2005.
Click image to enlarge.

Table 5: Rankings and scores for the cases shown in Fig. 10 for some of the indices described in Fig. 4 and the text.

	N0	N13	N14	N15	N16	N17	N18	N19	N22	N25
20 Apr 1995	377; 0.54	281; 0.66	259; 0.70	312; 0.66	293; 0.69	259; 0.79	353; 0.59	318; 0.68	416; 0.50	399; 0.63
17 Apr 1995	513; 0.40	478; 0.44	424; 0.48	495; 0.47	447; 0.50	397; 0.55	404; 0.51	436; 0.49	462; 0.45	499; 0.49
2 May 1997	455; 0.46	355; 0.56	389; 0.52	422; 0.53	443; 0.51	534; 0.36	562; 0.34	522; 0.37	580; 0.32	599; 0.38

4. Cluster analysis of the outbreak days

Because the multivariate indices show relatively small variation in the scores but considerable volatility in the rankings for the intermediate severe weather outbreak days, the utility of the ranking scheme as an analog for forecasters in future predictions of severe weather outbreaks could be interpreted to be quite limited. However, the individual cases exhibit several distinct types of outbreak days. For example, there are some outbreak days that feature primarily wind reports, whereas other days are dominated by hail reports (cf. Fig. 5). To determine if these characteristic outbreak days can be distinguished from each other using the multivariate indices developed in this study, a *k*-means cluster analysis[2] (e.g., Gong and Richman 1995) was employed on the outbreak days analyzed in this study.

The methodology for the cluster analysis is relatively simple. The multivariate index for each of the cases was decomposed into four parts: a hail component, a wind component, a tornado component, and an "others" component. Each component (except for the "others" component) comprises the sum of the standardized scores of each variable used for the relevant type of severe report (e.g., for hail, the severe hail and significant hail reports were used), weighted as indicated in Fig. 4, divided by the sum of the weights. The "others" component contains the total number of reports and the middle-50% parameter. This component is determined by subtracting the standardized middle-50% parameter value from the standardized score of the total number of reports for each outbreak day. This procedure essentially expands the multivariate index into a four-dimensional vector; each outbreak case is represented as a point in that four-dimensional space.

A *k*-means cluster analysis (squared Euclidean distance, random sample seeding) was performed on the four-dimensional vectors for each of the outbreak days for each of the indices analyzed in this study. After analyzing silhouette plots (Kaufman and Rousseeuw 1990) for a range of designated clusters (2 through 15; not shown), a 5-class categorization was chosen as representing the data best for each of the indices. Reviewing the cluster analyses of the four-dimensional vectors necessitated producing four three-dimensional scatter plots, with the four-dimensional data projected onto this three-dimensional "cross section".

The scatter plots of the *k*-means cluster analysis for two of the four-dimensional vectors (N3 and N22, which were selected to compare an index with all tornado variables with an index with a small number of tornado variables) indicate that the clusters illustrate the distinctive characteristics of the severe reports for the outbreak days quite well (Figs. 11 and 12). These clusters are described easily by eliminating one of the components of the vector in the three-dimensional scatter plots. For example, in the plots that remove the "others" component of the vector (top left plots in Figs. 11 and 12), the red points are elevated (*z*-component) above the rest of the clusters. The green points are strongly associated with the *x*-axis, and the blue dots are strongly associated with the *y*-axis. The first component (*x*-component) of the three-dimensional vector represented in the scatter plot is the "hail"

[2] Most hierarchical clustering techniques were found to cluster single outliers (e.g., 3 April 1974) rather than characteristic outbreak types. Because the ultimate goal is to classify outbreak days by characteristic types of events, hierarchical techniques were not employed. However, Ward's method (Ward 1963) produced clusters similar to those developed by the *k*-means technique (not shown) and could be used as an alternative.

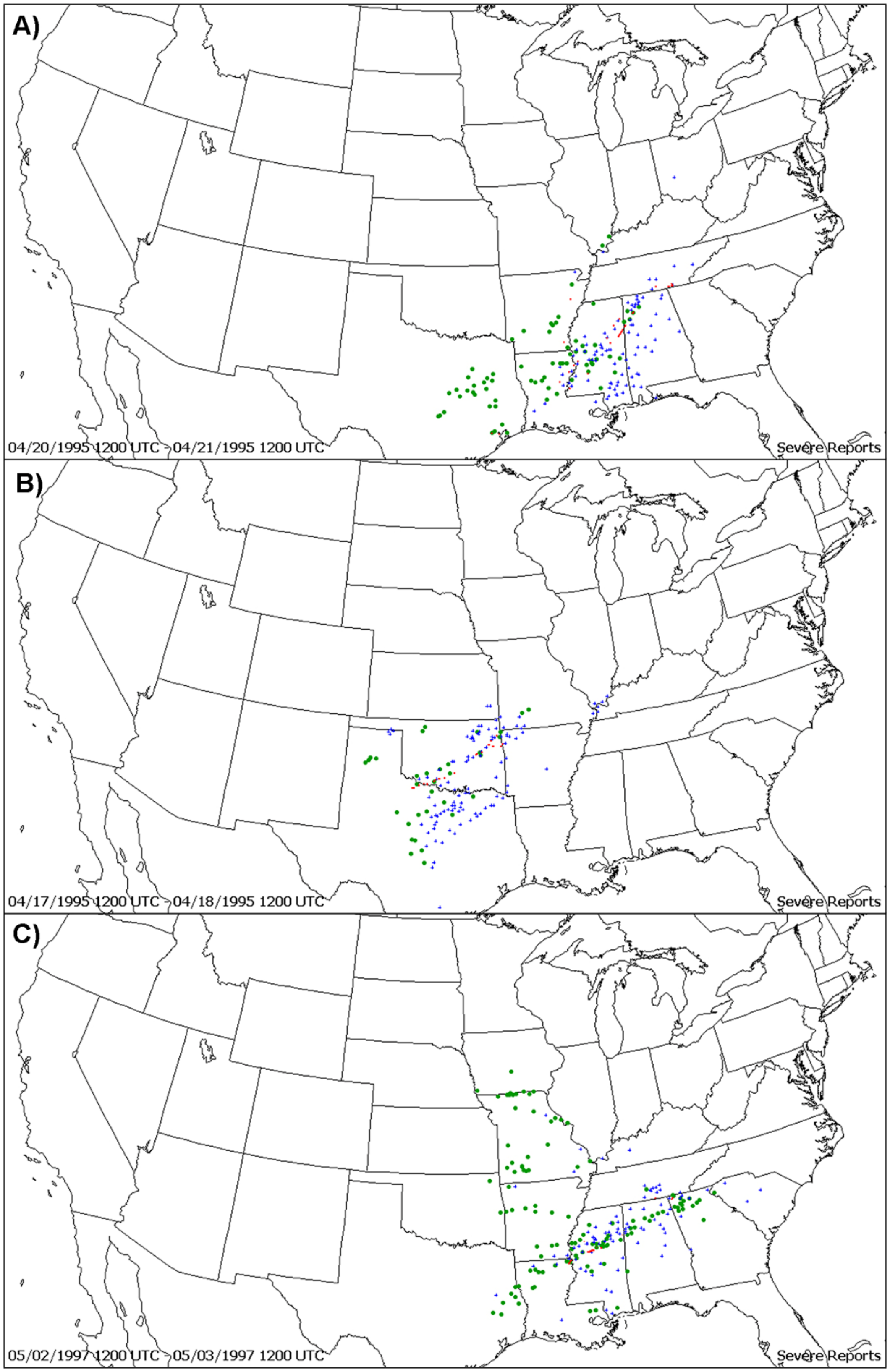


Figure 10: As in Fig. 2, for a) 20 April 1995, b) 17 April 1995, and c) 2 May 1997. *Click image to enlarge.*

component, the second component (*y*-component) is the "wind" component, and the third component (*z*-component) is the "tornado" component. In the plots that remove the "tornado" component of the four-dimensional vector (top right plots in Figs. 11 and 12), the orange dots are clearly separated from the purple dots and contain negative *z*-component values. The "others" component is the *z*-value of the three dimensional vector created for display. The green dots (hail component) are aligned with the *x*-axis, and the blue dots (wind component) are aligned with the *y*-axis, as before. The purple dots are clustered near the origin for each component that is excluded in the three-dimensional plots. Thus, the red dots are associated with major tornado outbreaks, the green dots are associated with outbreak days with hail as the dominant storm report, the blue dots are associated with outbreak days with wind as the dominant storm report, and the purple dots are associated with outbreak days featuring no dominant storm type, relatively little geographical scatter, and an intermediate number of total storm reports.

These descriptions of the clusters can be illustrated further by analyzing the cases represented by each of the dots. The classifications for each of the cases in Tables 3 and 4 for N3 and N22 illustrate the strong tendency for the top-ranked and bottom-ranked outbreaks to be in the same category (Table 6). The top-ranked cases are primarily tornado outbreaks, and the bottom-ranked cases are primarily cases with a large degree of scatter in the severe reports. This is desirable based on the relatively high weights on tornado variables and the middle-50% parameter, and illustrates the utility of indices with a small number of tornado variables (specifically DPI and number of tornadoes) to reproduce similar rankings to those indices that include all of the tornado variables.

Analysis of the 19 July 1983 and 7 May 1998 outbreak days (Fig. 5) also illustrate the utility in the four-dimensional representation developed to elucidate a multivariate index. The 19 July 1983 severe outbreak is classified "blue", or primarily a wind event, in both the N3 and N22 four-dimensional representations; and the 7 May 1998 outbreak is classified "green", or primarily a hail event, in both representations. The 17 April 1995, 20 April 1995, and 2 May 1997 outbreak days (Fig. 10) are all classified as "purple", or outbreak days featuring severe reports with little preference for wind or hail. These results are consistent with the plots of the severe reports of these cases.

Some tendencies in the clustering are not desirable. Many outbreaks of severe weather occurring early in the 1960-2006 period consist of a relatively large number of tornadoes with very few hail or wind reports (e.g., 26 May 1973; Fig. 13a). This is likely a nonmeteorological artifact in the reporting of these outbreaks, as a result of the tendency during these older events to report the tornadoes and not hail and/or wind. Commonly, these events are classified by the cluster analysis as "purple", or the mixed-mode severe report classification, if the tornado and "others" components of the four-dimensional index are not excessively large (i.e., values below 2) and the values for wind and hail are negative. Other examples include 24 January 1967, 3 April 1968, and 18 April 1969.

Outbreak days with a moderate to large number of tornadoes, but an unusually large number of wind or hail reports (according to the standardized scores, as described in Section 2), commonly are classified as either "wind" or "hail" cases. A "wind" example is the 30 May 2004 outbreak day (Fig. 13b), and a "hail" example is the 7 Apr 2006 outbreak day (Fig. 13c). If all three types of severe weather reports are unusually large in number, the event commonly is classified as a "tornado" case (e.g., 3 April 1974; Fig. 7a). Finally, some outbreak days with a small to moderate number of tornadoes but a large number of total reports over a geographically small area are sometimes classified in the "tornado" cluster (e.g., 20 June 1974 for the N22 index; see Fig. 8b), whereas this tendency is diminished for indices with a larger number of tornado parameters. The cluster analysis of the N3 four-dimensional vector identifies the 20 June 1974 outbreak day as a wind event.

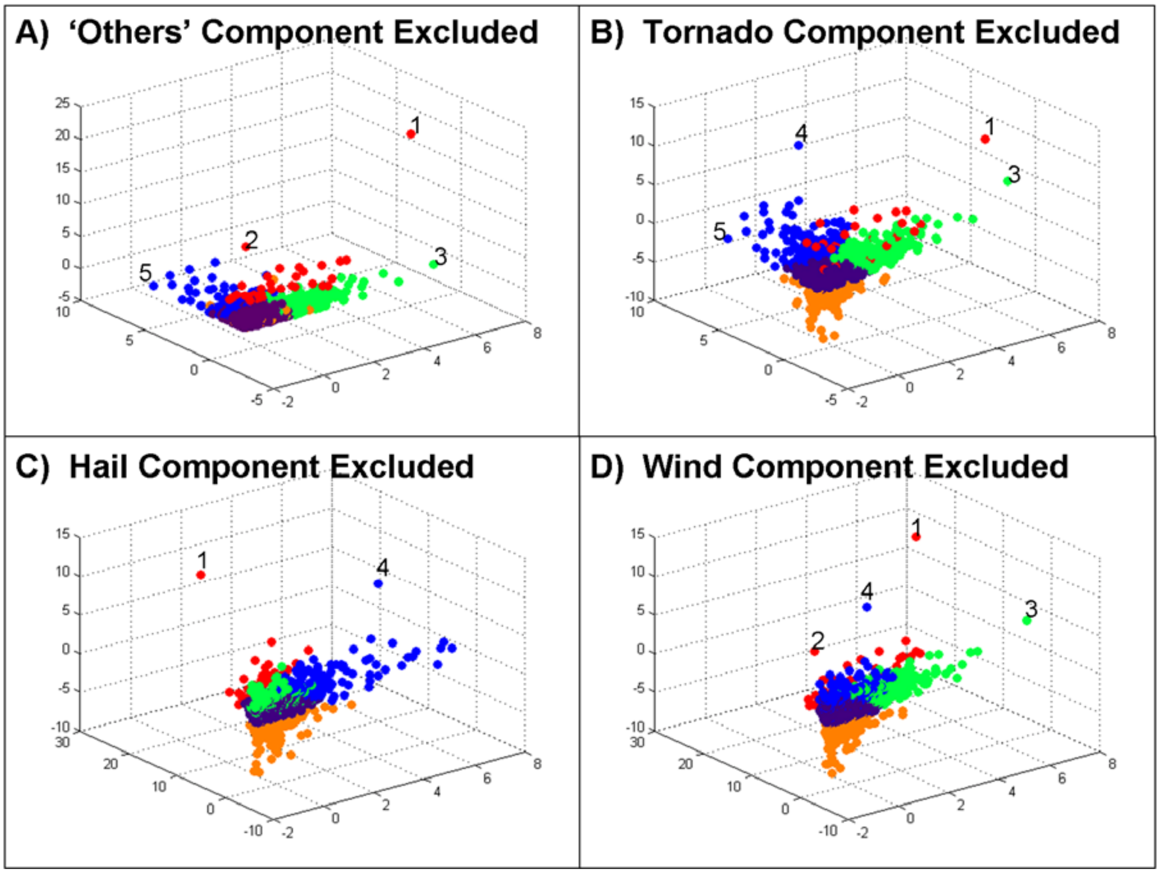


Figure 11: Scatter plots of the *k*-means cluster analysis for the four-dimensional representation of index N3 for each of the outbreak days in the study. The four-dimensional vector representation is described in the text. Excluded components are labeled. Points are associated with clusters that represent outbreak days as follows: red for major tornado outbreaks or days with an excessive number of total severe reports, green for days with an excessive number of hail reports, blue for days with an excessive number of wind reports, purple for days with mixed modes of severe weather, and orange for days with a relatively small number of total severe reports, and/or with a large degree of scatter or lack of geographical clustering of the severe reports. Specific cases are identified by numbers: 1) 3 April 1974, 2) 11 April 1965, 3) 24 May 1962, 4) 20 June 1974, and 5) 1 July 1994.

5. Discussion

Ranking severe weather outbreaks is controversial (Section 1), and the challenge is exacerbated by the numerous nonmeteorological artifacts contained within the archive of storm reports (Sections 1 and 2). Despite these difficulties, the multivariate indices developed in this study, which rank the top 30 days of each year from the period 1960-2006 according to the total number of severe reports on a given day, seem to be capable of ordering these events in a way that is relatively consistent with subjective notions regarding the severity of the outbreaks. The ranking schemes are robust in determining the most severe outbreaks and days with a lack of geographical clustering in the reports, whereas the ordering of the intermediate severe weather outbreaks is relatively volatile. However, a four-dimensional representation of the multivariate indices, based on components describing parameters for tornado, hail, wind, and the total number of reports and the geographical spacing of those reports, illustrate that the events could be clustered into "types". A *k*-means cluster analysis grouped the outbreak days into five categories: major tornado, hail-dominant, wind-dominant, "mixed-mode" (days with no dominant storm type), and "marginal" (days with a relatively small number of reports and/or a large degree of scatter).

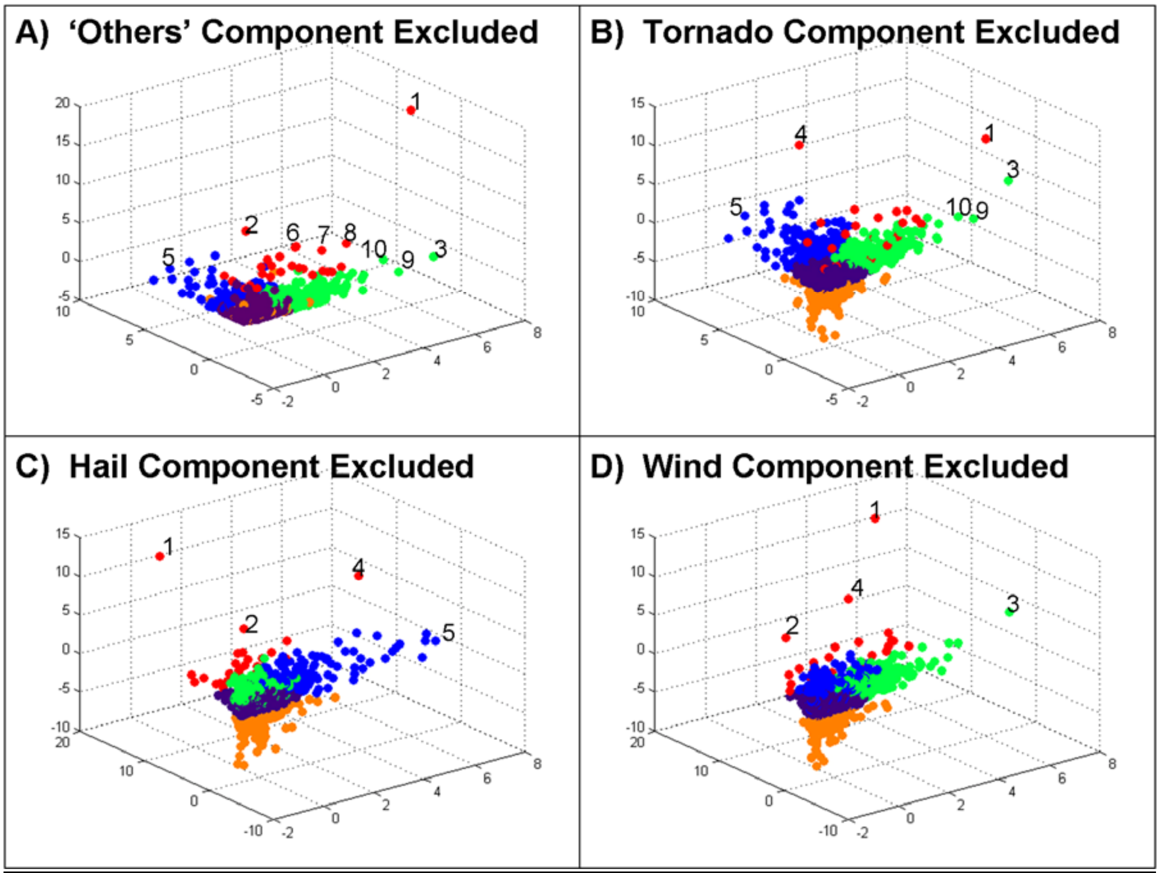


Figure 12: As in Fig. 11, for the N22 four-dimensional representation. The colors of the clusters correspond to those in Fig. 11. Specific cases are identified by numbers: 1-5) as in Fig. 11, 6) 31 May 1985, 7) 5 May 1964, 8) 26 April 1991, 9) 21 April 1996, and 10) 26 April 1994.

Subjective knowledge of the major tornado outbreaks proved to be useful in selecting parameters to incorporate into the indices and choosing the weights given to each of the parameters. However, subjective awareness of the intermediate events, including the primarily nontornadic outbreaks discussed in D06 and Shafer et al. (2009), is limited. Because of this limited knowledge and the existence of nonmeteorological artifacts in the storm data (e.g., Doswell and Burgess 1988; Brooks et al. 2003; Doswell et al. 2005), it is unlikely that the detrending techniques developed in D06 and used in this study have eradicated these artifacts completely. For example, many of the bottom 25 cases (Table 4) for each of the multivariate indices occur before 1990.[3] Doswell et al. (2005) show that nontornadic severe reports rapidly increased during the 1980s and continued after the implementation of the Weather Surveillance Radars-1988 Doppler (WSR-88Ds). It is likely that less significant, primarily nontornadic events are represented relatively poorly prior to this time.

Unfortunately, other inconsistencies exist in the data after 1990, which also may be influencing the representation of outbreaks during this period (Weiss et al. 2002; Doswell et al. 2005). It is not at all clear that reducing the period of analysis in this study would result in an improved set of rankings. The differences in the rankings by the indices that included more tornado parameters, versus those with a small number of tornado parameters, are apparent (Tables 3-5). However, there is a clear

[3] Although not shown, the cases ranked 26th to 50th from the bottom, with many of the multivariate indices presented in this study, consist of a larger number of cases after 1990, which suggests that no such artifact may exist in the resulting rankings.

Table 6: Categorizations of the top 25 and bottom 25 outbreaks for indices N3 and N22 using a *k*-means cluster analysis. Color designations are described in the text.

Top 25	N3	Top 25	N22	Bottom 25	N3	Bottom 25	N22
3 Apr 1974	Red	3 Apr 1974	Red	29 Jun 1994	Orange	20 Jun 1969	Orange
11 Apr 1965	Red	11 Apr 1965	Red	20 Jun 1969	Orange	9 Jul 1975	Orange
31 May 1985	Red	31 May 1985	Red	14 Jun 1963	Orange	28 Jul 2006	Orange
12 Mar 2006	Red	20 Jun 1974	Red	28 Jul 2006	Orange	6 Jun 1963	Orange
26 Apr 1991	Red	27 Mar 1994	Red	1 Jul 2005	Orange	14 Jun 1963	Orange
4 May 2003	Red	5 May 1960	Red	6 Jun 1963	Orange	27 May 1965	Orange
13 Mar 1990	Red	26 Apr 1991	Red	27 May 1965	Orange	1 Jul 2005	Orange
5 May 1960	Red	5 May 1964	Red	13 May 1980	Orange	8 Jun 1961	Orange
28 Mar 1984	Red	28 Mar 1984	Red	8 Jun 1961	Orange	13 May 1980	Orange
10 Nov 2002	Red	13 Mar 1990	Red	17 May 1986	Orange	17 May 1986	Orange
27 Mar 1994	Red	3 May 1999	Red	13 Jun 1971	Orange	16 Jun 1971	Orange
3 May 1999	Red	17 Apr 1970	Red	3 Jul 1960	Orange	3 Jul 1960	Orange
2 Jun 1990	Red	20 Mar 1976	Red	5 Jun 1973	Orange	14 Jul 1969	Orange
21 Jan 1999	Red	15 May 1968	Red	14 Jul 1969	Orange	5 Jun 1973	Orange
22 Nov 1992	Red	24 May 1962	Green	16 May 1965	Orange	16 May 1965	Orange
5 May 1964	Red	18 May 1995	Red	15 Jun 1987	Orange	6 Aug 1989	Orange
17 Apr 1970	Red	2 Apr 2006	Red	6 Aug 1989	Orange	15 Jun 1987	Orange
2 Apr 2006	Red	8 Jun 1974	Red	2 Aug 1986	Orange	2 Aug 1986	Orange
27 May 1973	Red	21 Jan 1999	Red	2 Aug 1993	Orange	2 Aug 1993	Orange
7 Jun 1984	Red	19 Apr 1996	Red	20 Jun 1977	Orange	8 Jun 1971	Orange
2 Apr 1982	Red	7 May 1993	Red	25 Jun 1988	Orange	20 Jun 1977	Orange
8 Apr 1999	Red	4 May 2003	Red	8 Jun 1971	Orange	25 Jun 1988	Orange
16 Jun 1992	Red	23 Apr 1968	Green	14 Jun 1966	Orange	14 Jun 1966	Orange
15 May 1968	Red	30 May 1998	Blue	18 Jun 1972	Orange	18 Jun 1972	Orange
18 May 1995	Red	14 Apr 1974	Blue	6 Jun 1981	Orange	6 Jun 1981	Orange

separation of the more significant outbreaks from the less significant outbreaks no matter which index is used, and the cases included in these distinct categories are quite consistent among the various indices presented. Depending on the research topic being investigated, a multivariate index including more or fewer tornado parameters may be desirable. Preference of one multivariate index to others is not obvious and was not expected, given the preexisting controversies and subjective interpretations discussed in Section 1.

As shown in Fig. 6, the grouping of severe weather outbreaks into three categories (major, intermediate, or marginal) is a very robust outcome. It likely would be of some benefit to severe storm forecasters to have some objective guidance about whether a given day would fall into this three-category classification based on the ranking index. For example, our three-category outbreak day rankings in this study could be interpreted as having a relationship to the SPC's categorical outlooks (see Ostby 1992; a description of these categorical outlooks effective as of 14 February 2006 can be found online:
http://www.spc.noaa.gov/misc/SPC_Prob_Conv_Otlk_Change_20060214.html).
Events falling in our major outbreak class all seem to correspond to what would be considered "high risk" days. Those days with intermediate outbreak rankings seem to correspond to "moderate risk" days. Those days with marginal outbreak rankings might be either "slight risk" or just below that category. It is noteworthy that the average annual number of days ranked as major outbreaks is similar to the average annual number of "high risk" outlooks. Further, the average annual number of intermediate outbreak-ranked days corresponds roughly to the average

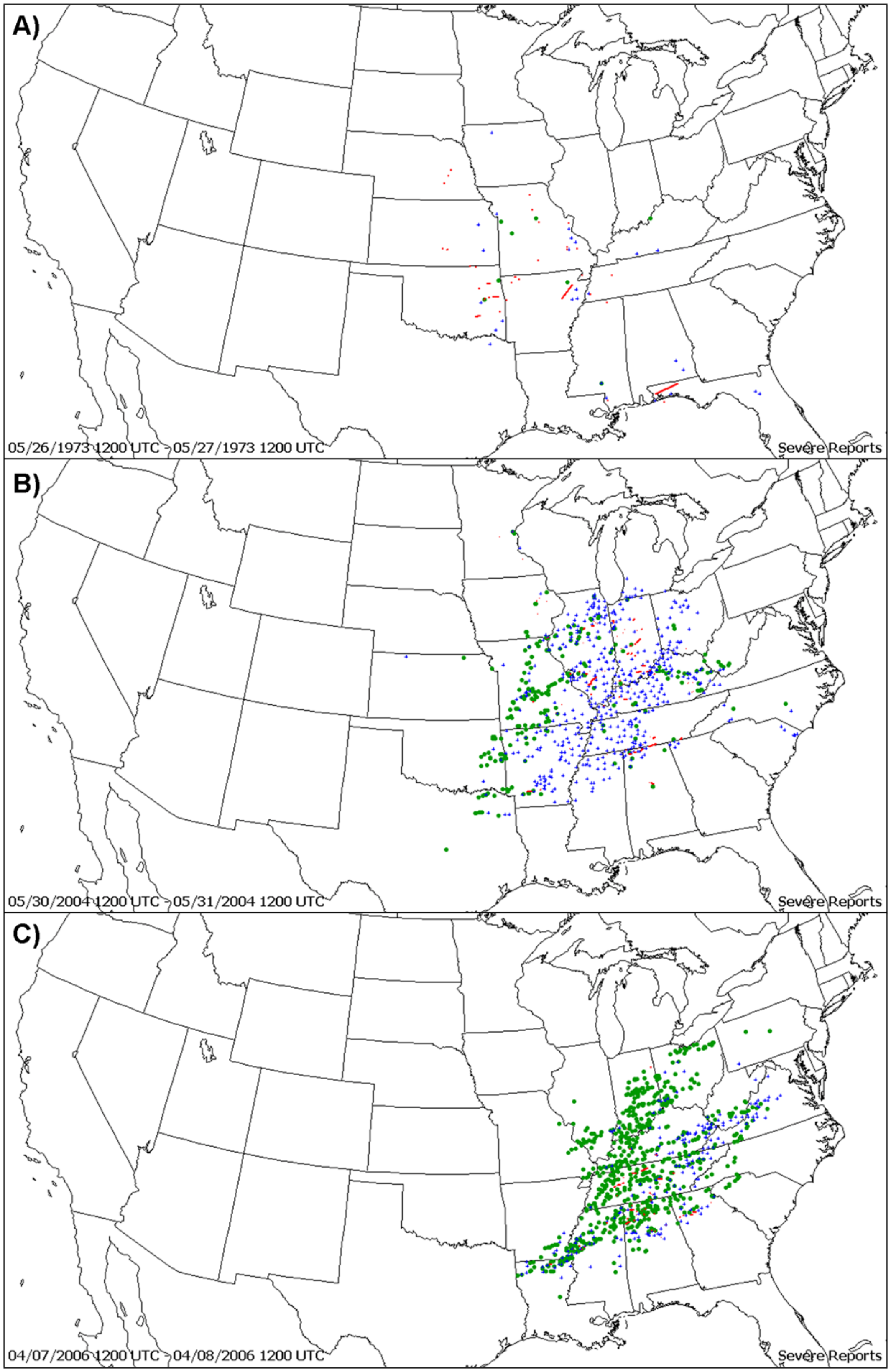


Figure 13: As in Fig. 2, for the 26 May 1973 (a), 30 May 2004 (b), and 7 April 2006 (c) outbreak days. *Click image to enlarge.*

annual number of "moderate risk" days. It would be useful to evaluate the utility of the proposed ranking system as guidance for categorizing the threat level, at least for high- and moderate-risk situations.[4]

Moreover, it also seems likely that objective guidance about the "mode" of the outbreak, in terms of the four-dimensional classification scheme for the outbreak ranking index, would benefit severe storm forecasters. The issue of whether or not the ranking index and its location in the four-dimensional outbreak mode space can be forecast is outside the scope of this study, however.

We reiterate that this study is not attempting to *define* what constitutes a "severe weather outbreak". The threshold used to develop the list of cases analyzed in this study (the top 30 days in a year, according to the total number of severe reports collected) is arbitrary. Interpretation of a day excluded from this study during the 1960-2006 period as a "non-outbreak" is not appropriate, as it is certainly possible for more than 30 severe weather outbreaks to occur in a given year. Similarly, inclusion of an "outbreak day" in this study does not necessarily imply that a severe weather outbreak occurred on that day. Again, it is possible that fewer than 30 outbreaks of severe weather (according to some definition of an outbreak) will occur in a given year.

We also are not claiming that the methodology developed and the various indices presented herein represent the optimum way to rank and classify severe weather outbreaks. The choice of parameters, the choice of weights for these parameters, the choice of a method for combining the variables included in the indices, and the choice of detrending methods developed are all based on subjective preconceptions about the characteristics of the relevant problem. However, the results of the ranking scheme are reproducible and do not appear to contradict subjective perceptions regarding the meteorological significance of the outbreaks included in the study. Moreover, the schemes are robust enough that similar results can be expected from most reasonable choices for a scheme to rank severe weather outbreaks.

[4] The major (intermediate) outbreak days classified in our study were not always forecast with high-risk (moderate-risk) outlooks.

The methods used to rank and classify severe weather outbreaks in this study can be employed elsewhere with multiple research topics. Examples include: using a similar technique to rank severe weather outbreaks in other portions of the world, determining the effects of changing the Fujita scale on more recent outbreak years, and investigating other types of meteorological hazards (e.g., hurricanes, flash floods, and winter storms). Additionally, studies investigating sensitivities of ranking schemes to various data archives, correlations among the archived parameters, different detrending techniques, and various methods of quantifying geographical scatter should be considered.

ACKNOWLEDGMENTS

This research was supported by NSF Grant ATM-080567. Plots of the severe reports in this manuscript were created using the SVRPLOT software described in Hart (1993). We thank Harold Brooks, Roger Edwards, and Valliappa Lakshmanan for their insightful reviews. We also appreciate the suggestion by John Hart to use the "middle-50%" variable, which is a simple but elegant way to evaluate the spatial coherence of severe weather reports.

REFERENCES

Abbey, R. F. Jr., and T. T. Fujita, 1975: Use of tornado path lengths and gradations of damage to assess tornado intensity probabilities. Preprints, *Ninth Conf. on Severe Local Storms,* Norman, OK, Amer. Meteor. Soc., 286–293.

——, and ——, 1979: The DAPPLE method for computing tornado hazard probabilities: Refinements and theoretical considerations. Preprints, *11th Conf. on Severe Local Storms,* Kansas City, MO, Amer. Meteor. Soc., 241–248.

Billet J., M. DeLisi, and B. G. Smith, 1997: Use of regression techniques to predict hail size and the probability of large hail. *Wea. Forecasting*, **12**, 154–164.

Blechman, J. B., 1975: The Wisconsin tornado event of April 21, 1974: Observations and theory of secondary vortices. Preprints, *Ninth Conf. on Severe Local Storms*, Norman, OK, Amer. Meteor. Soc., 344–349.

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

——, 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, and D. W. Burgess, 1988: On some issues of United States tornado climatology. *Mon. Wea. Rev.*, **116**, 495–501.

——, 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.

Ferguson, E. W., F. P. Ostby, and P. W. Leftwich Jr., 1987: The tornado season of 1985. *Mon. Wea. Rev.*, **115**, 1437–1445.

Fujita, T. T., 1974: Jumbo outbreak of 3 April 1974. *Weatherwise*, **27** (3), 116–126.

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

——, D. L. Bradbury, and C. F. van Thullenar, 1970: Palm Sunday tornadoes of April 11, 1965. *Mon. Wea. Rev.*, **98**, 29–69.

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

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.

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

Hart, J. A., 1993: SVRPLOT: A new method of accessing and manipulating the NSSFC Severe Weather Database. Preprints, *17th Conf. on Severe Local Storms,* St. Louis, MO, Amer. Meteor. Soc., 40–41.

Johns, R. H., and J. A. Hart, 1993: Differentiating between types of severe thunderstorm outbreaks: A preliminary investigation. Preprints, *17th Conf. on Severe Local Storms,* St. Louis, MO, Amer. Meteor. Soc., 46–50.

Kaufman, L., and P. Rousseeuw, 1990: *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley and Sons, 342 pp.

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

Lindsey, D. T., and M. J. Bunkers, 2005: Observations of a severe, left-moving supercell on 4 May 2003. *Wea. Forecasting*, **20**, 15–22.

Ostby, F. P., 1992: Operations of the National Severe Storms Forecast Center. *Wea. Forecasting*, **7**, 546–563.

Paruk, B. J., and S. R. Blackwell, 1994: A severe thunderstorm climatology for Alberta. *Natl. Wea. Dig.* **19,** 27–33.

Pautz, M. E., 1969: Severe local storm occurrences, 1955–1967. ESSA Tech. Memo. WBTM FCST12, Washington, DC, 77 pp. [Available from National Technical Information Service, 5285 Port Royal Rd., Springfield, VA 22161.]

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., 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.

Speheger, D. A., C. A. Doswell III, and G. J. Stumpf, 2002: The tornadoes of 3 May 1999: Event verification in central Oklahoma and related issues. *Wea. Forecasting*, **17**, 362–381.

Thompson, R. L., and M. D. Vescio, 1998: The destruction potential index—A method for comparing tornado days. Preprints, *19th Conf. on Severe Local Storms*, Minneapolis, MN, Amer. Meteor. Soc., 280–282.

——, and R. Edwards, 2000: An overview of environmental conditions and forecast implications of the 3 May 1999 tornado outbreak. *Wea. Forecasting*, **15**, 682–699.

Verbout, S. M., L. M. Leslie, H. E. Brooks, and S. L. Bruening, 2004: Leveling the field for tornado reports through time: Inflation-adjustment of annual tornado reports and objective identification of extreme tornado reports. Preprints, *22d Conf. on Severe Local Storms*, Hyannis, MA, Amer. Meteor. Soc., CD-ROM, 7B.3.

——, 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.

Ward, J. H., 1963: Hierarchical grouping to optimize an objective function. *J. Amer. Stat. Assoc.*, **58**, 236–244.

Weiss, S. J., J. A. Hart, and P. R. Janish, 2002: An examination of severe thunderstorm wind report climatology: 1970–1999. Preprints, *21st Conf. on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., 446–449.

REVIEWER COMMENTS

[Authors' responses in *blue italics.*]

REVIEWER A (Harold E. Brooks):

Initial Review:

Recommendation: Accept with minor revisions

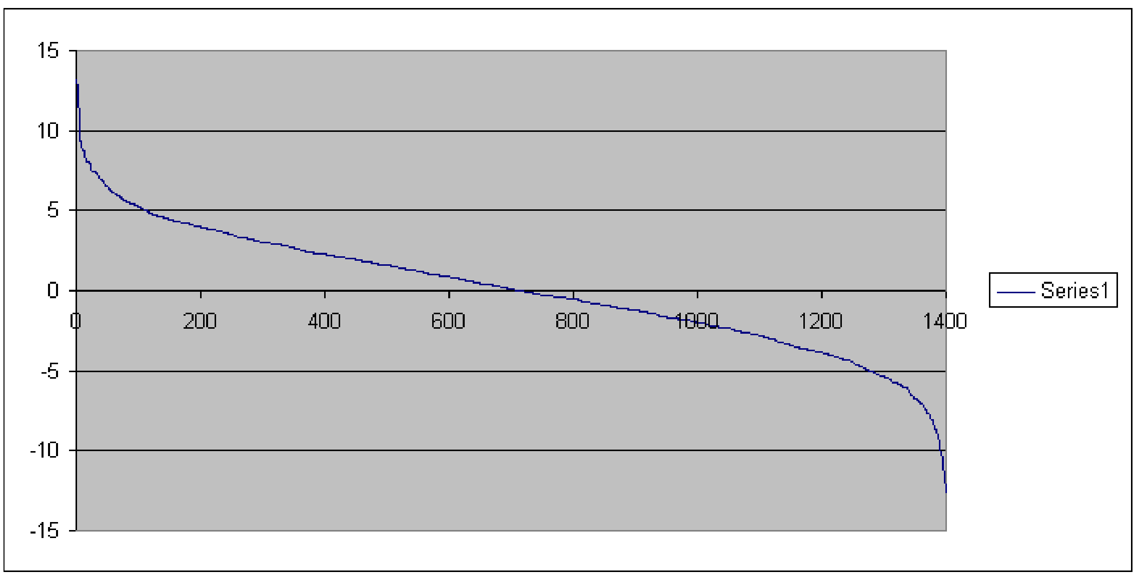
General Comments: In general, I like this paper quite a lot. The scatterplots in Figs. 9 and 10 are nicely illustrative of the results. My comments are minor. The longest comment is what I think is a possible explanation of one of the questions that arises in the work.

Substantive Comments:

1. p. 9: Since 3 April 1974 and 6 June 1981 are the two extremes, I'd really like to see the maps from those two days. Although 3 April 1974 is familiar, 6 June 1981 certainly isn't and the contrast would be interesting.

We have included a figure that provides the severe reports for each of these outbreak days (Fig. 7). However, we did not feel much discussion of the contrasts in these particular outbreak days was necessary or desirable, because of the separation in years for these two outbreaks (which would contribute to differences in the collection of reports) and because we go into detail about the characteristics of high-end severe weather outbreaks versus outbreak days with large scatter in a more general sense.

2. Figure 5 and discussion on p. 14: I don't think the shape is particularly surprising. Below is the result of a simple experiment. I created 14 series of 1400 random numbers from a normal distribution with mean 0 and standard deviation 1. I then summed the i^{th} elements from each series. Ordering those sums gives the following distribution:



This is the result of using uncorrelated series. Adding in some correlation to the series tends to flatten the middle or increase the slopes at the ends, depending upon how it's imposed. Since I imagine that the distributions that the authors have used have large departures from normal and that the correlation structure between the different variables has interesting features, it's quite likely that much of the shape is simply the result of adding correlated random series. I'm not sure what the impact of the normalization process is on this, as well. It's possible that it's not so much that the ranking scheme is not robust, but that the correlation between some of the variables associated with the most severe outbreaks (is) relatively well-correlated.

As a result, I think a lot of what you're seeing in this figure is a result of the nature of the statistical properties of the variables chosen. That doesn't invalidate any of the results or the speculation of the importance for forecasting/diagnosing outbreaks, but it might change the search for physical significance.

It might be interesting to take the 14 series and look at the statistical relationships between the variables. That's beyond the scope of what the authors are trying to do here, but it could be explored in the future.

We have added several paragraphs to section 3a detailing the experiment you describe above, the correlations among the parameters selected for the indices we developed, and possible implications of these correlations. We thank you for the enlightening comments!

[Minor comments omitted...]

Second review:

Recommendation: Accept

REVIEWER B (Roger Edwards):

Initial Review:

Reviewer recommendation: Accept with minor revisions

The "Rasmussen table" below summarizes my evaluation of this study. General and specific comments follow the table.

Criterion	**Satisfied**	**Deficient, but can be remedied**	**Deficient; *cannot* be remedied by modifying the paper**	**Deficient, *not known* if it can be remedied by modifying the paper**
1. Does the paper fit within the stated scope of the journal?	X			
2. Does the paper 1) identify a gap in scientific knowledge that requires further examination; 2) repeat another study to verify its findings; or 3) add new knowledge to the overall body of scientific understanding?		X		
3 Is the paper free of errors in logic?	X			
4. Do the conclusions follow from the evidence?	X			
5. Are alternative explanations explored as appropriate?				X
6. Is uncertainty quantified?	X			
7. Is previous work and current understanding represented correctly?	X			
8. Is information conveyed clearly enough to be understood by the typical reader?		X		

Overview: The manuscript represents a follow-up of sorts to the Doswell et al. (2006) study on ranking of U.S. severe local storm outbreaks within 24-hour daily temporal bounds bracketed by 12 UTC. The authors similarly detrended the data on the basis of yearly averages of top-30 events, created an array of 26 severe indices N0-N25 (each with at least slightly different weights of input variables), then mapped those indices of outbreak days into 4-dimensional vector space using a cluster analysis method. From this, they could determine and classify "types" of outbreaks by dominance of: hail (green), wind (blue) and tornado (red) modes, mixed-mode without obvious dominance by wind or hail (purple), or nebulous severe days with great amounts of scatter (orange). Outbreak days then could be ranked across years of occurrence and within classifications (duly acknowledging some limitations of detrending imposed by secular artifacts in reporting practices).

My assessment of the level of revision required is "minor" in a relative sense, given that:

1) There aren't any major "show-stoppers" here that either would entail intensive reanalysis or impugn the integrity of the results, and
2) The number of overall suggestions for improvement, while certainly nontrivial, isn't anywhere nearly as massive as the bulk of papers I've reviewed during the past several years. As such, this submission already is off to a fine start!

Overall, with mostly minor exceptions, the authors appeared to have approached this study in a very logical, statistically reproducible and defensible way, while potentially setting a positive precedent for a standardized approach to sorting and ranking severe weather outbreak days across years. Such methods potentially can be applied to sorting and ordering severe storm days in other geographic domains (e.g., portions of Canada or Europe) where data can be accumulated for a multi-year period over the same areas. I reckon there are unforeseen benefits to this approach also, perhaps in other subdisciplines in earth sciences that involve ranking or ordering of frequent-event datasets that change markedly over time because of secular artifacts. May main concerns (below) are organizational and presentational, not so much methodological, and should be addressable in relatively straightforward ways, therefore the "minor" revision tag. I also have embedded numerous minor comments concerning figures, clarity, formatting, organization, and otherwise, in the Word file.

I am quite confident that this paper will take its rightful place in the formal literature following a moderately intensive degree of minor revision, and that the concerns below can be ameliorated satisfactorily for prompt publication. Still, I wish to peruse both the revised version of this document and the authors' point-by-point responses to this review.

General Comments (significant concerns):

<u>1. Hypothetical:</u> While methodologically solid, the flow of the description of methods suffers via somewhat disjointed, missing or misplaced explanations and justifications. Frame the problem(s), say why you're trying to solve them, and with what hypothesized result(s). Otherwise, it gives the appearance of a statistical fishing expedition. For example, why was k-means clustering employed, especially in lieu of another clustering technique? What motivated the use of a 26-variable array in the first place – a burst of divine light, random stabbing in the dark, or legitimate hypotheses that such methods would sort the events in an effective and useful manner (with, presumably, some idea as to how, and with what result)? I strongly suspect the latter in each case, so…say so and present the driving hypotheses behind these analyses. Fortunately the deficiency is remedied easily in a way that adheres closer to the ideal of the scientific method: State your various major methodological hypotheses up front before diving into the analyses (i.e., actually provide each hypothesis before testing it).

The paper is clearly divided into two basic parts: ranking and classifying (more in response to comment 3 below). The original task, and the driving question of this study, is if a multivariate index like those in Doswell et al. (2006; hereafter, D06) could be developed that consistently ranks any type of outbreak in a way that agrees with subjective notions. Because the results of this original goal suggested that the ranking of such outbreaks presented limitations because of the volatility within the intermediate outbreaks, the next task was to determine if the indices developed could be used to classify the outbreaks in terms of severe report characteristics. In a sense, we have created a paper that is somewhat chronological rather than the standard "introduction-methodology-results-conclusions" format of most journal articles in meteorology. As such, some of the methodology is located in section 4 (specifically, the clustering techniques). Framing the manuscript in this matter seems more reasonable, since the objectives with each task were quite different.

Similarly, there are several examples in the original manuscript of "framing the problem", including the two topics specifically mentioned above:

> Because the multivariate indices show relatively small variation in the scores but considerable volatility in the rankings for the intermediate severe weather outbreak days, the utility of the ranking scheme as an analog for forecasters in future predictions of severe weather outbreaks is limited. However, the individual cases exhibit several distinct types of outbreak days. For example, there are some outbreak days that feature primarily wind reports, whereas other days are dominated by hail reports (cf. Fig. 5). To determine if these characteristic outbreak days can be distinguished from each

other using the multivariate indices developed in this study, a k-means cluster analysis (e.g., Gong and Richman 1995) was employed on the outbreak days analyzed in this study. (Section 4 – modified for figure number).

Regarding the candidate indices, we state that tests of several variations of the weights of the parameters were analyzed to investigate the robustness of the ranking scheme. In these examples, we frame the problem explicitly.

Hypothesizing what we expect(ed) to find with the results can surely be reported in more detail in the manuscript, and we have made a considerable effort to do so on many of the suggestions provided with the substantive and technical comments. However, in some circumstances, and with some scientific questions, specific mentioning of hypotheses seems obvious and/or unnecessary. For example, a binary scientific problem has two possible results, and if the problem is framed clearly, these results should be obvious. It's entirely possible for a scientist to be completely unsure of the answer beforehand, thereby having no idea what the end result will be. One specific example mentioned in the technical comments involved the variable DPI and its relatively effective performance in ranking the most severe outbreaks similarly to indices which maintained all of the tornado variables analyzed in this study. The result easily could have been total path length, number of significant tornadoes, etc., and a priori scientific/educated guesses regarding which one would be best are, in the end, simply guesses. Though we have made a substantial effort to mention many of our hypotheses explicitly, there is a point when such detail becomes tedious and unnecessary.

The main problem in our original manuscript appears to be with the middle step: explanations of choice of methods (see more with comment 2 below), and we have altered the manuscript substantially in this respect.

To clarify these choices, we have modified the manuscript for the two specific examples provided as well as many examples provided in the technical comments. Regarding the choice of 26 candidate indices, verbiage is added in sections 2 and 3a (see comment 4 below) regarding the specifics of the 26 candidates, the reasons for weight selections, and the reasons for using 26 at all. This choice is certainly somewhat subjective (as there is no means of determining the "best" number of indices to analyze), but it allows for investigation of multiple changes to the weights. For example, what happens if some of the variables are removed (such as the tornado parameters – N17-N19; N21-N25), if some cases are subject to lower rankings because only one type of severe weather is prevalent, when significant wind/hail reports are emphasized, etc?

For the k-means cluster analysis, we have included a footnote detailing why this particular method was chosen. Essentially, it was because of its simplicity in computing, its relative resistance to cluster individual outliers, and its subjectively meaningful results. Hierarchical techniques were strongly susceptible to clustering outliers (e.g., using "average" and "single" linkage techniques). The Ward's technique did compare well to the k-means cluster analysis and could just as readily be used, and we have mentioned this as well.

<u>2. Mysteries of reasoning:</u> This is somewhat related to #1, and is a function of lack of clear communication instead of analytic flaws. There are several places through the manuscript – the "Data and methodology" section in particular – where the reason for an analytically important choice is left unstated, and as such, a minor mystery to the reader. These include, but aren't limited to, the reasons for

- selection of data thresholds in time (starting at 1960) or
- scaling technique (logarithmic scaling of reports for the yearly top-30 regression graphs), or
- selection of particular indices for case-specific examination (e.g., Tables 2-4) or deviation analysis (Fig. 5) from the broader set, or
- the assumption that synoptic systems are "distinct" (what does that mean, anyway?) for disparately clustered report areas.

The authors safely can assume that some readers will guess the answers to such little mysteries right away. I've had some experience with these as a co-author on an earlier paper of this ilk, and as such, can make confident inferences to understand what's going on, and for what reasons. But why make the readership at large guess or infer at all, when it's no big deal just to be more up-front about the methods and their rationale? If such choices are arbitrary, say so; otherwise, concisely explain in the text. It's about the ideal of scientific reproducibility, which applies not only to the actual statistical techniques but to the reasoning behind them. I've also pointed out some examples where this sort of explanation already is done, to the credit of the authors.

Considerable effort was made to clarify reasons for the methodology in the revised manuscript. For example, the selection of the time period 1960-2006 was simply to increase the sample size of the cases analyzed. An increase in the time periods from those used in D06 allowed for a greater number of relatively rare types of outbreaks within the 1410 cases analyzed (such as major tornado outbreaks, outbreaks dominated by hail or wind reports, etc.). This will likely be beneficial for future work investigating a mesoscale model's capability of predicting the type of outbreak, as determined by the clustering analysis in section 4.

However, some specific examples you mentioned were explained in D06. For example, the scaling technique used for detrending was arbitrary, as stated in the conclusions section of D06 and in the conclusions section of this paper. (One look at Fig. 1, however, may provide insight on why a logarithmic scaling was performed: many of the functions appear exponential. We have mentioned this in the revised text.) D06 is referenced, with additional wording in the revised manuscript, to explain some of these somewhat arbitrary choices. Other examples include the choice of weights for the parameters, the use of a linear-weighted multivariate index as opposed to other methods, etc.

Section 3a was substantially altered to discuss the indices more clearly (see response to comment 4) and to explain why some of the indices were not selected for analysis.

We have removed wording regarding the "distinct" synoptic-scale systems, because the major point of this portion of the methodology section was to emphasize that some outbreaks feature substantial geographical scatter.

<u>3. Description of purpose:</u> The processing of these events through the development of the 26 variables and clustering in 4-D space seems to be a robust method for sorting and describing these events in the most objective way possible (given the inherent subjectivity of the choice of weights and of numerous individual severe reports). The manuscript's consistent description of the purpose for doing so (ranking), however, appears incomplete based on what actually is done (ranking and binning). I suggest, therefore, that more balance is given in how the goal is described, not just it the text but in the title as well, i.e., "A Multivariate Index for Ranking and Classifying Severe Weather Outbreaks."

Referring back to our response to comment 1, the paper is presented in two parts: ranking and classifying. The classification aspect came about as the analysis of the ranking schemes indicated that there was substantial volatility of the rankings with the intermediate cases. Though the paper is presented in a somewhat unusual format, the intentions of the classification aspect are clearly illustrated in sections 3b and 4. (See, e.g., the text from the manuscript we included in our response to comment 1.) However, our original intent was to rank the outbreaks.

Nevertheless, we have altered the manuscript title as suggested and have modified the manuscript in various places to broaden the proposed intent to "ranking" and "classifying".

<u>4. Indices explained poorly and/or not shown:</u> Unless shown (as partly by Fig. 3) or clearly explained somewhere, the remaining "not shown" indices are black boxes, left to the reader to deduce. Given their core importance to the research, and in the interest of reproducibility, all indices need to be shown. This is an ideal use for an Appendix.

We agree that the descriptions of the indices, the various differences among them, and the reasons for using a subset in our figures need to be clarified. We have substantially modified section 3a in an effort to eliminate these deficiencies. However, an appendix is not appropriate, since the results are all based on the indices we developed – too fundamental to the results to place outside of the main text. It is also not correct to assume that all of the indices are equally valid, but this assumption may have been based on our lack of clarity in describing these indices initially. Some of the indices have biases by design, as these indices are used for the calculations of the "maximum" indices: N13-N16 and N25. For example, indices N5-N8 are positively biased toward "wind" events and negatively biased toward "hail" events since the hail variables were given a weight of zero. Hopefully, section 3a is much more thorough and comprehensible.

Furthermore, the verbiage describing interrelationships of the various indices Nx early in Section 3a is "N"-credibly convoluted. The text description is literally accurate, but functionally useless. Read it aloud into a sound recorder and play it back, and you'll see what I mean. It sounds like "N"-comprehensible gibberish. I had to get out a pencil and paper to perform a very slow, careful, diagrammatic shell game with N-this and N-that to keep up with what was what, and where, and still succeeded only after some backtracking and repetition. In case the problem was my own obtuseness, I then read the text aloud to several fellow meteorologists, including one of the other co-authors on D06. None fully could keep track of what on Earth was going on. Solution: Treat a picture as better than a hundred words. Direct the reader to a set of clear Appendix illustrations akin to Fig. 3, but covering all the indices. Group the indices logically by their weighting pattern in some way, e.g., all the zero wind-weight indices are grouped together, then all the zero hail-weight indices, etc., climaxing with all the max-score indices. Use of a flow chart containing all the indices, akin to diagrams of electrical circuitry (also placed in the proposed Appendix), may further aid the reader in associating N-this to N-that correctly. Then considerably shorten and rephrase Section 3a accordingly.

Refer to our response above. Figure 4 (old Fig. 3) now includes all of the indices except the "maximum" indices, for reasons explained in the text.

<u>5. Utility for forecasters and for severe storm climatology:</u> In one sense, these somewhat related matters are more nebulously addressed in the manuscript than they should be; and where they are addressed, it is with some apparent misunderstanding about the translation of this work into SPC outlook categories. The connection between the result bins of this research and SPC categorical outlooks is unclear, aside from the seemingly coincidental rough matching of their frequency of occurrence per year, and is incompletely elucidated. Please refer to specific comments embedded in the document for more discussion. Also, since SPC categorical outlooks ostensibly are driven by probabilistic threat assessments, there may be an even greater benefit to assigning probabilistic ranges for each event type (tornado, wind, hail).

We are unsure how the "rough matching" of the three types of outbreaks (based on relative severity) can be described as "seemingly coincidental" to the frequency of the categorical outlooks each year. The cases in our ranking scheme that fall on the high-end portion of the outbreaks are all examples of the most severe outbreaks, which one would hope deserve the "high risk" categorical outlook. The selection of 3 April 1974, 11 April 1965, and 31 May 1985 at the top of nearly every index we developed is certainly a desired, if not expected, outcome of our work. Surely such days are deserving of a "high risk" outlook, are they not? A cursory look at recent cases in our top 200 for every index that also were forecast as high risk days include but are not limited to 28 March 1984, 26 April 1991, 27 March 1994, 19 April 1996, 31 May 1998, 21 January 1999, 8 April 1999, 3 May 1999, 4 May 2003, 10 May 2003, 22 May 2004, 29-30 May 2004, 12 March 2006, and 7 April 2006. We believe the fairly high number of high risk days (which verified as major severe weather outbreaks) that end up in our top 200 outbreaks hardly seems coincidental but is instead quite intentional, which at least provides motivation for investigating how our three-category designations of outbreak severity correspond to definitions of categorical outlooks. That is, the highly ranked days should correspond roughly to "high risk" days, the intermediate ranked days should correspond roughly to "moderate risk" days, and the low ranked days might be considered either as "slight risk" or "see text" days, depending on the degree of scatter on a particular day.

Despite our objections/concerns to the above comment, our inclination was to remove much of the wording regarding SPC categorical outlooks within the text. However, we maintain that this idea is not based merely on coincidence and so deserves some brief mention in the conclusions.

There's fertile ground here for the authors to propose potential utilities of their methods (or adaptations thereof) in analysis and normalization of hazardous weather report climatology – SLS reports before 1960 or in the future, in other parts of the world, with other kinds of Storm Data weather records (flash floods, non-thunderstorm wind damage, fatal lightning, etc.) – that deserves at least superficial attention in the final Discussion section as areas for further work.

We have added a paragraph in the conclusions to address this.

<u>6. Interpretation of results.</u> Borderline "major-minor" comment, but it is potentially important: A little more conceptual interpretation is needed for the results of the cluster analyses. For example, what

significance, if any, is there to the tight neighboring of wind-dominant days with hail vector component excluded, and vice versa (per the colored plots in Fig. 9)?

We have added some wording in this section, expanding on the descriptions of the five clusters. Be careful reading these graphs; the scales change with some of the panels.

The tight clustering is a result of the "wind" and "hail" clusters featuring a predominance of these events without a particularly large number of other types of reports and may be emphasized by a scale change on the graph. Also note the z-axis modifications from panel (a) to panels (b)-(d). Panels (b)-(d) have a z-component associated with the "others" component. The "hail" and "wind" clusters remain tightly packed on the axes because of the lack of tornado reports (y-axis) and lack of geographical scatter (z-axis) (panels (c) and (d)).

[Minor comments omitted...]

Second review:

Reviewer recommendation: Accept with minor revisions. No further formal review by me is needed. The "Rasmussen table" below summarizes my evaluation of this study. General and specific comments follow the table.

Criterion	**Satisfied**	**Deficient, but can be remedied**	**Deficient; *cannot* be remedied by modifying the paper**	**Deficient, *not known* if it can be remedied by modifying the paper**
1. Does the paper fit within the stated scope of the journal?	X			
2. Does the paper 1) identify a gap in scientific knowledge that requires further examination; 2) repeat another study to verify its findings; or 3) add new knowledge to the overall body of scientific understanding?	X			
3 Is the paper free of errors in logic?	X			
4. Do the conclusions follow from the evidence?	X			
5. Are alternative explanations explored as appropriate?	N/A			
6. Is uncertainty quantified?	X			
7. Is previous work and current understanding represented correctly?	X			
8. Is information conveyed clearly enough to be understood by the typical reader?	X			

Overview: The revision of this article is greatly improved from the original in the areas in which it was most deficient – namely, clarity and flow of explanations, "framing the problem" (and exemplifying it in several instances), describing choices of methods, and including testable hypotheses. Admittedly, the act of comprehensibly yet concisely describing such a big set of variables with similar-sounding names N1, N2, N3, and so on, with their numerous flavors of the same general theme, is challenging undertaking. The authors seem to have succeeded on the second try. They scrapped the original pile of dense and confusing verbiage and instead produced a much clearer and better-organized approach for doing so in the main text body (Section 3a). While I still think a somewhat more detailed, flow-chart style appendix would help, the latest approach is thoughtful, satisfactory and acceptable. I thank the authors for their careful attention to such improvements and the nontrivial amount of work needed to undertake them in a relatively short amount of time.

Only a very few minor and technical concerns and corrections remain; and it appears that those (below) can be addressed easily by the authors without need for any further perusal on my part in the review stage.

[Minor comments omitted...]

REVIEWER C (Valliappa Lakshmanan):

Initial Review:

Recommendation: Accept with minor revisions

Major Comments:

(1) The authors develop an index for ranking all types of severe weather outbreaks and discover that the rankings are relatively impervious to the choices and weights of parameters. What they mean by this is that the outbreaks can be broken into three tiers based on the rankings -- the rankings are volatile within the tiers, but rarely do outbreaks move from one tier to another. To me, this indicates that their "ranking index" is really just a classification index, i.e., it can be used as a screen to identify significant outbreak days but that the rankings themselves have no meaning.

This interpretation is not completely correct. The robust result of the ranking schemes developed creating three tiers of outbreaks is absolutely true, but we believe there is reason not to dismiss the rankings of each multivariate scheme developed. The ordering of the high-end and low-end outbreak days is relatively consistent (particularly the low-end days, in which the middle 50% parameter appears to be particularly effective); Tables 3 and 4 illustrate this nicely. The rankings of the outbreaks are particularly volatile with the intermediate outbreak days, as a result of the scores for approximately 1000 of the 1410 case days falling in the -1 to 1 score range, depending on the parameters selected and their weights for a specific index (e.g., Table 5).

Furthermore, the ordering of the high-end outbreak days agrees with subjective preconceptions [e.g., the 3 April 1974 tornado outbreak is generally considered to be the worst outbreak in recent history and is listed first in every index developed; the highest-ranked outbreaks are clearly more significant (in terms of meteorological significance and societal impacts) than the lowest-ranked outbreaks within the high-end outbreak days, etc.]. It's unreasonable to expect negligible volatility in the rankings, based on the controversies expressed in section 1, and the lack of "absolute truth" regarding such rankings. However, the trend from rank 1 to rank "much greater than 1" should make intuitive sense. The removal of several tornado parameters for indices N17-N19 and N21-N25 resulted in a shift of several outbreak days with few tornadoes but very large numbers of hail and wind reports toward the high-end events, which agreed with our preconceptions. It is possible (though outside the scope of the current work) that a forecast model may be more capable of predicting the scores of an outbreak day using a certain index. Such a result would be beneficial for operational forecasting, archiving of severe storm reports, etc.

Finally, as section 3 discusses, the rankings of the intermediate outbreak days are volatile (e.g., Table 5), but the scores using the various multivariate indices feature fairly limited variability. This suggests that there indeed may not be much difference among these outbreaks (in terms of overall severity), which is mentioned in the manuscript.

A rank has no meaning/purpose unless it can be used for comparison -- i.e., an event with rank of 10 is more significant than an event with a rank of 15. However, the authors find that the ranks are tiered and it is these tiers that are impervious to the choices and weights of parameters. This would indicate that the resulting index should NOT be used for ranking outbreaks, but only to screen/filter significant outbreaks. I would suggest that the authors make this clear.

This was already suggested in the manuscript multiple times. For example, see the last sentence of section 3b, the first sentence of section 4, and several of the paragraphs in section 5. See above for further comments on interpretation of the rankings.

(2) The authors should expand on their motivation for *ranking* severe weather outbreaks. They start out with the claim that such a ranking would be useful to severe weather forecasters, but do not explain how a ranking would help. I'm the furthest thing from a severe weather forecaster, of course, so I could be missing something obvious. But it seems to me that an index that is based on severe weather *reports* would not have much prognostic value. The second motivation that they provide is that this would help studies such as the choice of cases to evaluate WRF runs. But such a study would not require a ranking -- just a screen to select significant cases.

Determining the relative severity of past outbreaks based on a ranking scheme would be greatly beneficial to forecasters if it can be shown that a meteorological model is capable of predicting the score of the multivariate indices. In a sense, the predicted score could then be used as an analog for past events. Past outbreaks with scores near the value predicted for a forecast outbreak could be analyzed to determine the possible characteristics of the outbreaks in terms of severe reports and potential societal impacts. We have added some wording in the introduction to emphasize.

(3) A cursory examination of the storm reports database would be enough to convince someone that most outbreak days are dominated by one or the other type of severe weather report. The k-means clustering only formalizes this observation in the aggregate.

From what we have observed with the 1410 cases we have analyzed, the contention that most outbreak days are dominated by one type of severe weather report is not accurate. The largest cluster in our analysis is the "mixed-mode" cluster. The point of this cluster analysis was to determine if the multivariate index we have developed can provide information in terms of the type of outbreak in addition to its relative severity.

However, it is equally clear from an examination of the storm reports that an outbreak day often consists of a high number of hail reports *and* a relatively large fraction of wind reports. It would have been more interesting if the authors had modeled each outbreak day as a mixture of distributions (of the 5 categories that resulted from their cluster analysis) and reported on the fractions involved in the mixtures. Is this different on significant vs. non-significant days? In other words, mixture modeling would permit a more fine-grained, statistical look at the data.

We agree that this work would be interesting to do, but this is well beyond the original scope of the study. We think this is more appropriate for a follow-up investigation.

[Minor comments omitted...]

Second review:

Recommendation: Accept

The authors have addressed my concerns but I'm not fully satisfied with this response:

ME: A rank has no meaning/purpose unless it can be used for comparison -- i.e., an event with rank of 10 is more significant than an event with a rank of 15. However, the authors find that the ranks are tiered and it is these tiers that are impervious to the choices and weights of parameters. This would indicate that the resulting index should NOT be used for ranking outbreaks, but only to screen/filter significant outbreaks. I would suggest that the authors make this clear.

AUTHORS: This was already suggested in the manuscript multiple times. For example, see the last sentence of section 3b, the first sentence of section 4, and several of the paragraphs in section 5. See above for further comments on interpretation of the rankings.

Note that the TITLE of the manuscript is "A multivariate index for RANKING and classifying severe weather outbreaks". No amount of clarification can counteract what the title seems to promise. If the editor is satisfied with this title, then I have no further comments on this manuscript. I recommend that it be accepted for publication.