Gensini, V. A., and W. S. Ashley, 2011: Climatology of potentially severe convective environments from North American regional reanalysis. *Electronic J. Severe Storms Meteor* **6** (8), 1–40.

Electronic Journal of
SEVERE STORMS METEOROLOGY

Climatology of Potentially Severe Convective Environments from the North American Regional Reanalysis

VICTOR A. GENSINI
The University of Georgia, Athens, Georgia

WALKER S. ASHLEY
Northern Illinois University, DeKalb, Illinois

(Submitted 1 March 2011; in final form 22 December 2011)

ABSTRACT

This study establishes a U. S. climatology of potentially severe convective environments for the 30-y period 1980–2009 from the North American Regional Reanalysis. Variability of environments supporting significant severe weather is examined for four active severe-weather regions in the U. S. Regional comparisons illustrate potentially significant-severe environments varied greatly both spatially and temporally over the 30-y period of record. The spatial and temporal distributions of significant severe-weather environments and reports are subjectively examined for comparison purposes. While one has to be cautious when linking environments and reports, average calculated significant severe-weather environments show similarities to the annual cycle of significant severe-weather reports. Additionally, mean center analysis indicates that there is no significant shift in the average position of these environments during the period of record.

1. Introduction

The 32-y period 1980–2011 exhibited 110 weather-related disasters causing damages totaling over one billion dollars (NCDC 2011). Of those, about 30% (32 disasters) were the result of severe thunderstorms. While events of this magnitude only occur on average once per year, the impacts are realized at regional scales. For instance, recent severe thunderstorms in April 2011 spawned tornadoes responsible for $17.3 billion in damages and over 350 fatalities across 20 states. The increasing trend of losses from severe thunderstorms (Changnon 2001) and tornadoes (Brooks and Doswell 2001; Changnon 2009) can be attributed to societal and economic changes rather than an increase in event frequency (Bouwer 2011). However, recent research has indicated that the potential for severe thunderstorm environments may increase under future anthropogenic emissions scenarios (Trapp et al. 2007; Van Klooster and Roebber 2009). The combination of increasing societal vulnerability (Cutter et al. 2003) and severe thunderstorm environment frequency may lead to greater severe thunderstorm hazard impacts in the future.

Forecasting severe-weather events involves techniques such as pattern recognition, climatology, and parameter evaluation (Johns and Doswell 1992). However, since pattern recognition and climatology are based on forecaster experience and climatological averages, they can fail in atypical circumstances. Routinely, events occur that lie outside of classic spatial severe-weather patterns, peak temporal severe-weather climatology, and rule-of-thumb forecasting values. Therefore, arguably the most effective forecasting technique is parameter evaluation using an ingredients-based approach advocated by Doswell et al. (1996). Initially introduced to aid forecasters in predicting flash-

Corresponding author address: Victor A. Gensini, Department of Geography, The University of Georgia, 210 Field Street, Athens, GA 30602. E-mail: vgensini@uga.edu

flood events, this method assesses the ingredients that are necessary for deep, moist convection (DMC) and can be used to develop a set of variables that are necessary for DMC, based upon physical principles.

Brooks et al. (2003b; hereafter B03) sought to understand the global distribution of severe DMC environments using an ingredients-based approach coupled with course-resolution global reanalysis data; however, no study has employed newer, high-resolution reanalysis datasets such as the North American Regional Reanalysis (NARR; Mesinger et al. 2006) to examine convective environments. NARR permits researchers to examine historical DMC environments in more detail than ever before. NARR data are used in this research to examine the variability of potentially significant[1] severe DMC environments across four DMC-active regions in the U. S. In turn, this allows forecasters to understand the spatial and temporal aspects of DMC environments in their respective region. A climatology of significant severe-weather ingredients for each of the four regions will promote discussion of inter/intraregional variability, as well as inter/intrannual variability in single domains. Comparisons of interregional variability allow us to see if trends are consistent across multiple domains. Understanding this variability is vital for hypotheses about future organized DMC environments in various climate change scenarios. In fact, the report of the 2002 IPCC Workshop on Changes in Extreme Weather and Climate Events (Brazdil et al. 2002) states that reanalysis techniques will be vital in determining how convective parameters vary and how they will affect future distributions of hazardous convective weather.

2. Background

a. DMC environments

Diagnostic parameters such as CAPE, convective inhibition (CIN), storm-relative helicity (SRH), 0–6-km bulk wind difference (BWD), and lifting condensation level (LCL) are all useful in determining the potential for DMC (Rasmussen and Blanchard 1998; Rasmussen 2003; Craven and Brooks 2004). Essentially, the ingredients-based forecasting methodology (Doswell et al. 1996) uses parameters analyzed by forecasters in a prognostic sense through extrapolation and numerical weather prediction. Forecasters historically placed thresholds of these parameters on a composite chart (Miller 1972; Crisp 1979) to represent the greatest threat for DMC. Combinations of different ingredients found on composite charts and sounding presentations have been used to develop composite indices. Composite indices often are employed in discriminating between atmospheric environments favorable for certain types of severe-weather events.

Doswell and Schultz (2006) emphasize that forecasters must exercise caution when employing indices and parameters. They argue that these indices seek to simplify the nonlinear atmosphere and should not be treated as a simple solution for where DMC will occur. Furthermore, it is vital to understand exactly which variables enter into the calculation of composite indices, and precisely how they are combined, in order to understand their strengths and weaknesses.

Previous research has used various DMC ingredients to discriminate between significant-severe and severe environments (B03), supercell and non-supercell environments (Thompson et al. 2003, 2007), and tornado and significant tornado environments (Thompson et al. 2003, 2007). This study uses results from B03 by using the product of CAPE and 0–6-km BWD to determine a potentially significant severe-weather environment. In particular, B03 show that when the product of 100-hPa mixed-layer CAPE and 0–6-km BWD is >20 000, the environment favors significant severe-weather events. This composite index was chosen over other candidates because of its simple, yet effective, calculation of a potentially significant severe-weather environment for climatological purposes. This study also incorporates CIN into the Craven and Brooks (2004) Significant Severe, or C composite index, to identify areas most favorable for significantly severe DMC.

b. Geographic variation of DMC

Lee (2002) showed that global reanalysis data provides a good approximation of severe-storm parameters when compared to collocated observed soundings. It is probable, but not

[1] Significant severe weather is defined as hail at least 5 cm (~2 in.) in diameter, convective wind gusts ≥120 km h^{-1} (65 kt), or a tornado of at least EF2/F2 damage (Hales 1988).

guaranteed, that other reanalysis datasets behave similarly. If the NARR fails to reproduce convective variables accurately, then it is possible that plots presented herein could be inaccurate. NAR has some documented biases, mainly corresponding to small temperature biases (Kennedy et al. 2011) and precipitation fields (Ruane 2010a; Ruane 2010b).

B03 were first to use reanalysis data to approximate convective environments using an ingredients-based method. Specifically, B03 developed spatial distributions of global severe thunderstorm and tornado environments for the period 1997–1999 using global reanalysis proximity soundings. Since B03 considered only 3 y, temporal aspects of convective environments were addressed in a later study by Brooks et al. (2007; hereafter B07). B07 used 7 y of global reanalysis data to construct annual cycles of four convectively important variables. These annual cycles provide insight into the mean convective season that a particular location may experience (e.g., high probabilities of severe storms during a focused part of the year, or lower probabilities of severe storms throughout the entire year).

Both B03 and B07 suggest the need for more research using reanalysis. For example, B03 propose that it may be possible for reanalysis to address issues of spatiotemporal changes in the distribution of environments favorable for severe thunderstorms, and to provide a framework for investigating possible effects of climate change scenarios on severe thunderstorm distribution and frequency. B07 began investigating this topic by addressing the current state of convective annual cycles, but additional research is needed to draw meaningful conclusions about future global convective regimes. As illustrated in B03 and B07, the central U. S. is home to the one of the world's highest probabilities of significant-severe convective weather. Therefore, detecting changes or trends in the spatiotemporal distribution of convective environments in this region could have significant implications for such probabilities.

3. Methods

a. Data

Stability and vertical wind-shear variables from the NARR were examined for 1980–2009 at 0000 UTC, to develop climatologies of environments favorable for significant severe-weather events. The NARR offers a consistent climate data suite for North America (Mesinger et al. 2006), and is preferred to global reanalysis data for this study, owing to its relatively high spatial resolution. Native NARR gridded binary (GRIB) data has a horizontal grid spacing of 32 km, and 45 vertical σ layers (Black 1994). The NARR uses the 2003 operational Eta model as part of the assimilation cycle (G. Manikin 2010, personal communication). In comparison, B03 and B07 employed the U. S. National Center for Environmental Prediction (NCEP) and NCAR global reanalysis (Kalnay et al. 1996).

The NCEP/NCAR global reanalysis used in B03 has 210-km horizontal grid spacing and 28 vertical σ layers. Although NARR resolution is superior, the domain only encompasses North America and thus does not allow for global environment examination. Furthermore, the NARR begin in 1979, 30 y later than the NCEP/NCAR global reanalysis. For this study, the resolution benefits of the NARR outweighed the length of temporal record. This study focuses on the most active organized DMC region in the world, justifying the highest spatial resolution data available to examine environments in greater detail.

Calculation of a proximity C composite index in this study uses 0–6-km BWD (derived by vertically interpolating winds at constant pressure levels to AGL height coordinates) and MUCAPE (directly available from the NARR dataset). The threshold for a significant-severe environment was computed as follows:

$$\text{0–6-km BWD} \times \text{MUCAPE} = 20\,000 \qquad (1)$$

Note that this formula is different than that used in B03:

$$2.86\log(\text{0–6-km BWD}) + 1.79\log(\text{MLCAPE}) = 8.36 \qquad (2)$$

While these are qualitatively similar, the B03 discriminator emphasizes shear more. In addition, the B03 line is generally less than Eq. (1) and is more conservative in the depiction of significant-severe environments. Depending upon the vertical profile, this difference could be important, particularly in low-CAPE situations. Thus, we refer to our index as a proximity C composite index in this manuscript.

NARR calculations of MUCAPE and MUCIN entail dividing the lowest 180-hPa of the atmosphere into six 30-hPa deep layers. Average physical properties then are computed for each 30-hPa layer. Using properties of the 30-hPa layer with the largest θ_e, MUCAPE and MUCIN are calculated (G. Manikin 2010, personal communication). Therefore, some similarities exist between this definition of MUCAPE and MUCIN compared to conventional CAPE calculations for mixed-layer parcels. Calculating CAPE using the virtual temperature correction is most appropriate (Doswell and Rasmussen 1994); however, theoretical parcel calculations in the NARR do not use it (G. Manikin 2010, personal communication). From a climatological perspective, the calculations of CAPE and CIN from the NARR should not be a major source of error in the proximity C composite index calculation.

The NCAR command language (NCAR 2011) was used to post-process NARR files into netCDF format for ease of examination in a geographic information system (GIS). Raster netCDF images were grouped by year and summed to create frequency climatologies for environments with MUCAPE ≥2000 J kg^{-1}, MUCAPE × 0–6-km BWD ≥20 000 (proximity C composite index), and MUCAPE × 0–6-km BWD ≥20 000 in the presence of MUCIN ≥–75 J kg^{-1}. The MUCIN ≥–75 J kg^{-1} threshold was chosen based on results from Bunkers et al. (2010), who found that most significant-severe reports occurred in environments with MUCIN ≥–75 J kg^{-1}. Next, raster files were organized by month to analyze the annual cycle. Gridded significant-severe environments underwent two passes of a Gaussian (3 × 3) low-pass filter to help with revealing spatial patterns. Readers should use caution when evaluating filtered data because spatial smoothing tends to mask fine-scale details, while broad patterns are retained (Brooks et al. 2003a; Doswell et al. 2005; Ashley 2007).

b. Parameter variability

Four spatial regions (Fig. 1) were identified to examine the regional variability of CAPE and the proximity C composite index. The North American Albers Equal Area Conic projection was chosen to examine derived fields in order to compare regions of equal area. Each region is roughly 1.6 × 10^6 km^2 and contains 1550 reanalysis grid points. Regional means were calculated for each year, smoothed with a 5-y running mean. The domains were chosen arbitrarily to analyze if there were any latitudinal (e.g., region 3-to-1) or longitudinal (e.g., region 1-to-2) shifts in regional trends of proximity significant severe-weather environments.

Determining if there has been a shift in significant severe-weather environments over time would benefit assessment of climatological risk. To investigate this, regional averages of the frequency of proximity composite C index values ≥20 000 were used. Annual averages are then compared to the 1980–2009 average to create departures from the 30-y mean. This offers the opportunity to identify areas that may have been anomalously active (or inactive) in terms of severe weather in a given year.

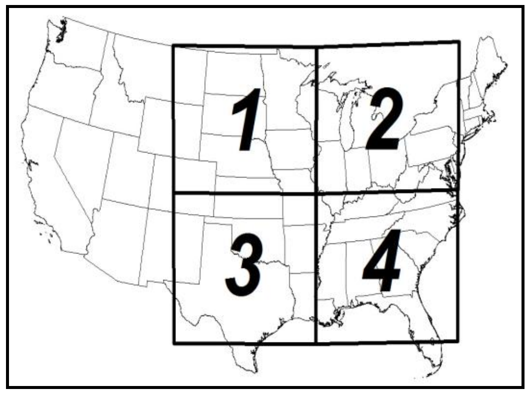


Figure 1: Four regions analyzed in this study: 1) Northern Plains; 2) Great Lakes; 3) Southern Plains; 4) Southeast.

c. Comparison to reports

In order to assess the validity of these environment climatologies, it is desirable to verify that they are capturing significant severe-weather events. First, significant severe-weather reports in a 6-h window from 2100–0300 UTC for the period 1980–2009 were queried from the SPC's SVRGIS database (Smith 2006). Although there are many referenced issues with the severe-weather report database (e.g., Doswell and Burgess 1988; Grazulis 1993; Brooks and Doswell 2001; Brooks and Doswell 2002; Verbout et al. 2006; Doswell 2007), they are the only ground-truth report data available. To gain a clearer picture of report frequency, smoothing in the form of a kernel density function was used at a 32-km grid spacing (same as the NARR), and a 250-km search radius. A kernel density function calculated the density of point features around each output raster cell. The density at each cell then was calculated by adding the values of all the kernel surfaces where they overlay the raster circle center. The kernel function used in this analysis is based on the quadratic kernel function described in Silverman (1986, p. 76, Eq. 4.5). By comparison, a Gaussian kernel with approximately half that value for its σ would have similar structure. Therefore, this is close to a 125-km Gaussian kernel, analogous to what was used in Brooks et al. (2003a).

Maps then were compared subjectively to identify discrepancies or similarities between the frequencies of significant severe-weather reports and modeled significant severe-weather environments. Because significant severe weather occurs in a variety of conditions and is dependent on more than just the variables examined in this study, it would be unwise to expect that the modeled environments herein capture all events. Instead, of greater importance are the patterns and trends that may (or may not) be identified throughout the temporal period. Although outside the purpose of this study, a potential source for future research would be to examine other verification metrics (e.g., false-alarm environments and reports that occur outside of significant-severe environments).

4. Results

a. CAPE

Similar to results shown in B03 (cf. their Fig. 6), most areas east of the Rocky Mountains experience five or more days per year with CAPE values ≥2000 J kg^{-1} (Fig. 2). The frequency of days with CAPE values ≥2000 J kg^{-1} is maximized near the Gulf Coast, where the proximity to surface moisture plays a dominant role in creating large CAPE.

Of more significance is the annual cycle of these large-CAPE environments (Fig. 2 animation). During the winter months, large-CAPE environments are most frequent near the Gulf of Mexico as moisture plays a larger role than typically weak mid-tropospheric lapse rates. As higher surface θ_e values begin to spread poleward in the spring, the juxtaposition of steep mid-tropospheric lapse rates from the elevated terrain to the west (i.e., Rocky Mountains, Mexican Plateau) and low-level moisture advected poleward from the Gulf of Mexico frequently creates large-CAPE environments during the spring and early summer across the Great Plains. Such lapse rates characterize a feature known as the elevated mixed layer (Lanicci and Warner 1991). July contains the onset of the southwest U. S. monsoon season (Douglas et al. 1993; Stensrud et al. 1995; Adams and Comrie 1997). The increase in convection associated with the monsoon assists in mixing the troposphere and inhibits the formation of steep lapse rates over the elevated terrain. By July and August, the highest frequencies of large-CAPE environments have shifted off the High Plains and into the eastern Great Plains and Mississippi Valley. The northward shift of large-CAPE environments can be attributed primarily to the poleward progression of the polar jet stream. It is also possible that transpiration from corn and soybean fields during the peak of the growing season (July–August) in the Midwestern Corn Belt acts to enhance near-surface moisture (Mahmood et al. 2008). Although most of the Midwest averages one day in September with CAPE values ≥2000 J kg^{-1}, frequencies quickly shift equatorward during October.

b. Deep-layer shear

McNulty (1978) emphasized the importance of mid- and upper-tropospheric wind maxima to DMC forecasting. Although these local maxima, known as jet streaks, can be used to identify areas favored for DMC development (McNulty 1978; Maddox and Doswell 1982; Uccellini and Johnson 1979; Clark et al. 2009), they also assist in providing deep-layer wind shear necessary for the organization and sustenance of DMC (Doswell 2001). Although many different layers over which wind shear is measured can be used as diagnostic parameters to discriminate supercells from nonsupercells (Ramsay and Doswell 2005; Houston et al. 2008), we use an 18 m s^{-1} (~35 kt) threshold for the 0–6-km BWD based on results from Rasmussen and Blanchard (1998) and Rasmussen (2003). The annual cycle of 0–6-km BWD ≥18 m s^{-1} mimics that of the Northern Hemisphere polar jet stream (Fig. 3 animation).

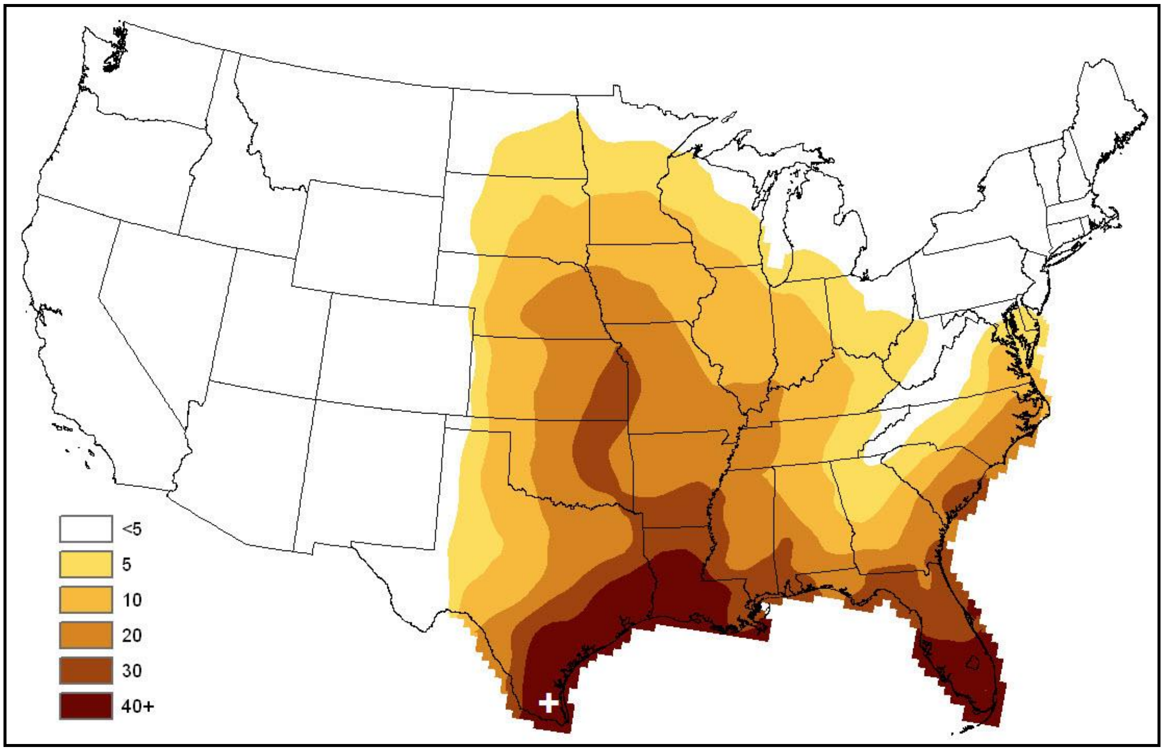


Figure 2: The average (1980–2009) number of 0000 UTC soundings per year with MUCAPE values ≥2000 J kg^{-1}. White "+" indicates maximum grid-cell value. *Click image to enlarge. Click here for an animation of the annual cycle of large CAPE environments.*

c. C composite index

The combination of CAPE and deep-layer wind shear, as assessed via the C proximity composite index, discriminates well between severe and significant severe-weather environments (B03). Similar, large-CAPE environments (Fig. 2) show some important differences from significant severe-weather environments (Fig. 4). Many locations along the Gulf Coast that exhibited high frequencies of large-CAPE environments now illustrate reduced frequencies by including shear in the presence of CAPE. From a large-scale perspective, adequate deep-layer shear environments tend to be most frequent in the eastern and northern U. S., whereas large-CAPE environments are most frequent in the south-central U. S. As a result, the area most favored for significant severe weather occurs in the eastern Great Plains where these two ingredients frequently overlap.

d. Regional variability

When analyzing the 5-y significant severe-weather environment running means, similar regional trends are found (Fig. 5). First, all but Region 2 show a decreasing trend in the number of potentially significant severe-weather environments since the late 1990s. Meanwhile, Region 2 has remained mostly unchanged. Although the trend for most regions is decreasing, it is not out of the range of earlier frequency values experienced in the early 1980s. Individual analysis of each variable and its contribution to the proximity C composite parameter indicates that CAPE is the main governing variable determining the frequency of significant-severe days. With only a 30-y period of record, it is difficult to assess the significance of these trends. A similar preliminary study (Gensini and Brooks 2008) using global reanalysis and a 49-y record showed comparable trends during the overlapping periods of record. It appears that the significant severe-weather environments Gensini and Brooks (2008) analyzed starting in the early 1970s peaked in the late 1990s, and has been declining since. Still, it is unlikely that the current length of observed reanalysis can capture the natural variability in such environments. For example, the small decline (from 15 to 10 environments y^{-1} since 1999) is only over a 10-y period and likely not capturing true natural variability.

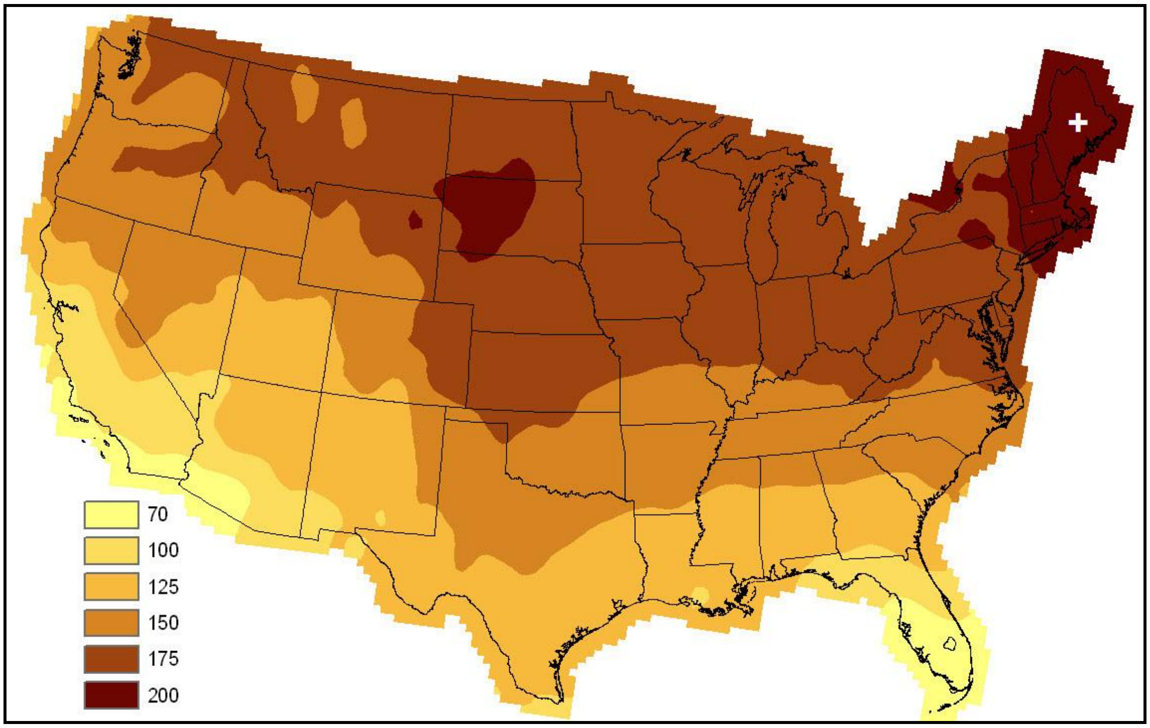


Figure 3: 1980–2009 average annual number of days with 0–6-km BWD $\geq$18 m s^{-1}. White "+" indicates maximum grid-cell value. *Click image to enlarge. Click here for an animation of the annual cycle.*

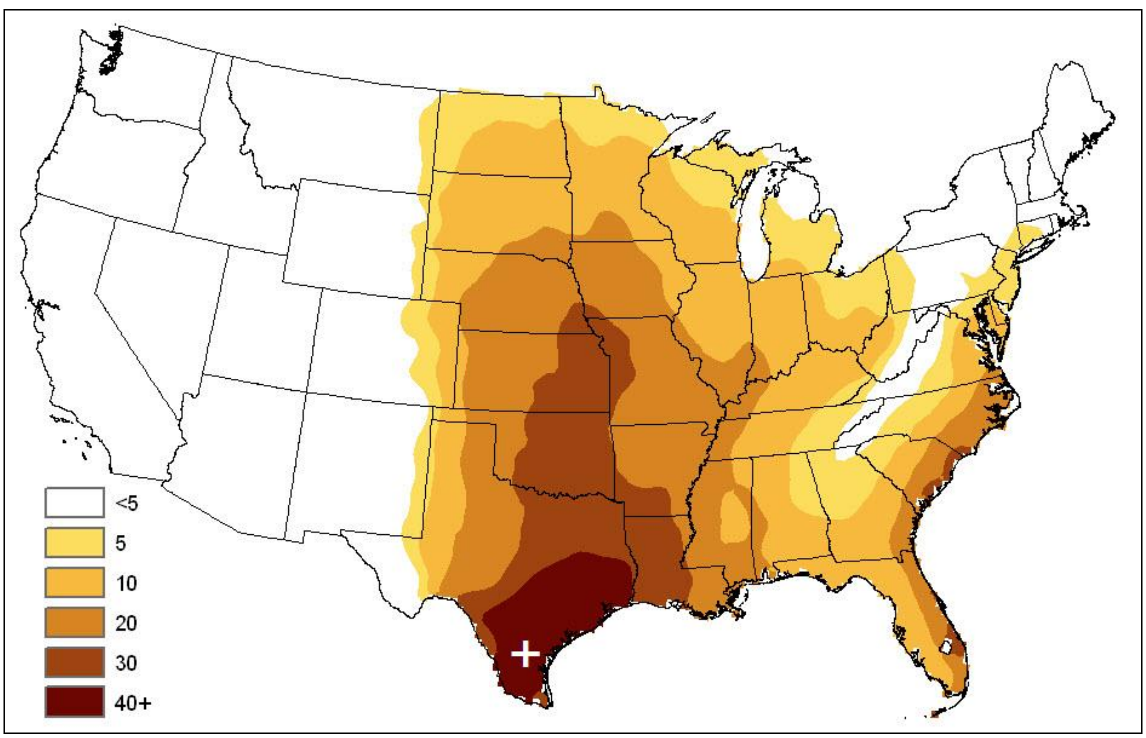


Figure 4: 1980–2009 annual average number of 0000 UTC reanalysis grid points with proximity C composite index values $\geq$20 000 in the presence of CIN values $\geq$–75 J kg^{-1}. White "+" symbol indicates maximum grid cell value. *Click image to enlarge. Click here for an animation of the annual cycle of significant- severe environments in the presence of minimal CIN*.

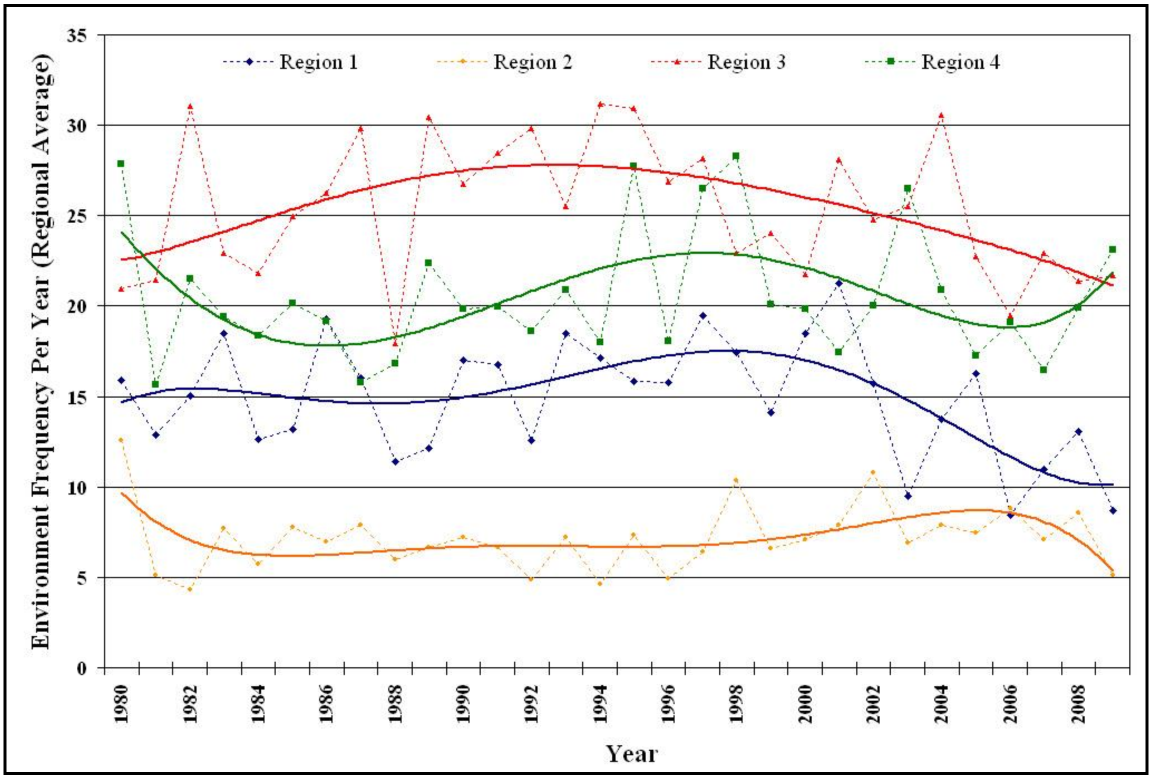


Figure 5: 1980–2009 regional average comparisons of significant severe-weather environment frequencies (dashed). Fourth degree best-fit polynomials are plotted as solid lines. Units are environments y^{-1}. *Click image to enlarge*.

Although regional mean analysis is useful to examine trends over time and allows one to compute an average with which to compare, departures and trends are occurring across numerous spatial and temporal scales. In addition, the atmosphere is not restricted to a user-defined domain. To illustrate this point, departure maps were created for a below- and above-average significant severe-weather environment year (Fig. 6). In 1988 (Fig. 6a), a central U. S. drought leading to below-average surface moisture conditions, coupled with deep- layer wind shear displaced north into Canada (Trenberth et al. 1988; Trenberth and Guillemot 1996) resulted in a large area experiencing below-average significant severe-weather environment frequency. In contrast, 2001 (Fig. 6b) was extremely active in terms of significant severe-weather environments, especially across the central and northern Plains. There were 1323 significant severe-weather reports in 2001, but only 447 in 1988. Some of this difference could be due to changes in reporting efficiency (Fig. 7). It is not the purpose of this study to examine environment anomaly correlation to any other convective variable(s).

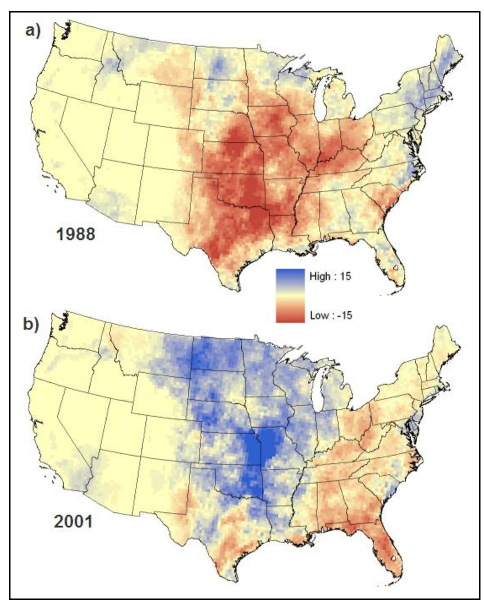


Figure 6: a) 1988 and b) 2001 significant severe-weather environment departure from the 1980-2009 annual average. Positive (blue)/negative (red) values correspond to above/below average departures respectively. *Click image for animation*.

Rather, this result demonstrates that even during spatially large below-average environment years, some locations still experience above-average frequencies and vice versa, illustrating the importance of understanding the difference in scale when examining severe convective environments.

e. Comparison to reports

Although the frequency of U. S. averaged potentially significant severe-weather environments decreased in the last 10 y of the data, there was a substantial increase in number of significant-severe reports and nearly a constant trend in significant-severe days (a day on which a significant-severe report occurred; Fig. 7). When comparing the annual cycle of significant-severe environments (Fig. 4) to observed significant-severe reports (Fig. 8), two primary results were found: 1) the main areal axis coverage of significant severe-weather events is positioned in the same area as the main axis of potentially significant severe-weather environments; and 2) southern Plains significant-severe environments are overestimated in a stepwise fashion roughly beginning south of Interstate 40 (blue dash denotes approximate location on Fig. 8) and again south of Interstate 20 (red dash on Fig. 8). Although it may seem trivial to explain this discrepancy as a result of comparison between different dataset types, there are a few other important points to mention. First, as discussed in B03, environments presented herein are essentially times when the atmosphere is favorable for organized DMC, not implying that it necessarily will occur. This index should not be used to forecast significant severe-weather occurrence; rather, it is beneficial in discriminating

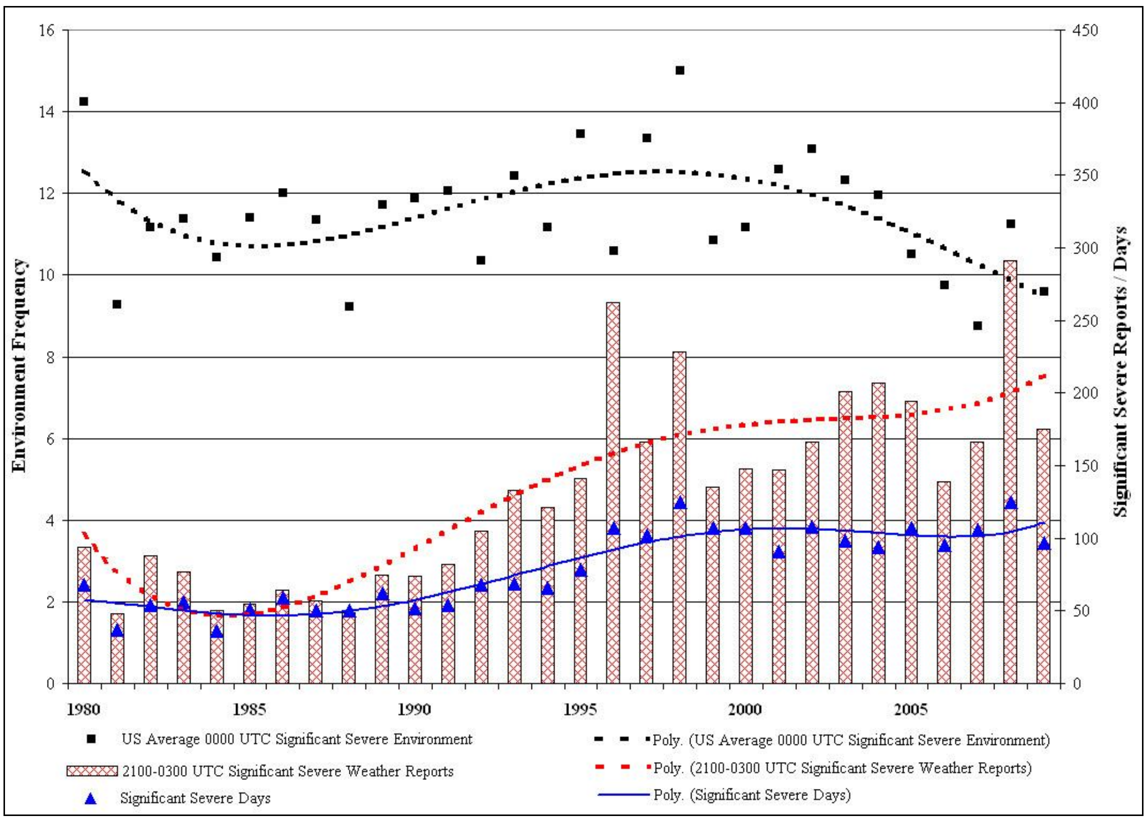


Figure 7: Comparison of 2100–0300 UTC significant severe-weather reports (bar; red), significant-severe days (blue; triangle) and U. S. averaged significant severe-weather environments (box; black). "Best-fit" lines are fourth degree polynomial functions. *Click image to enlarge*.

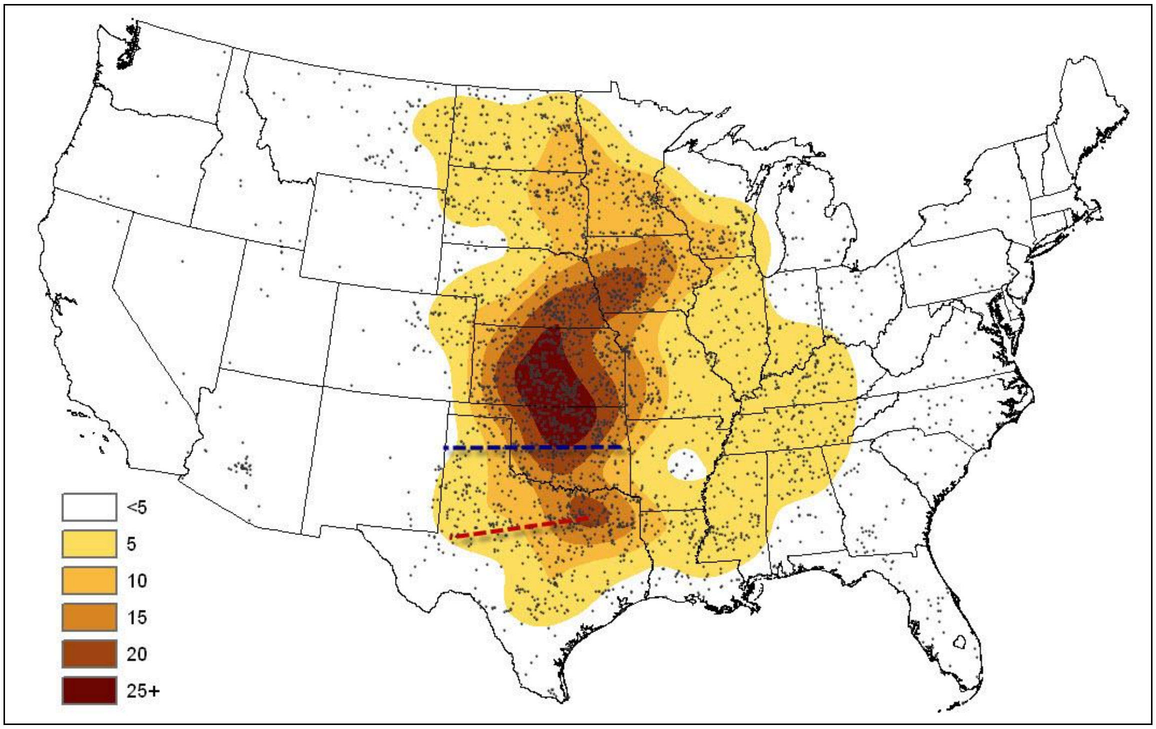


Figure 8: 2100–0300 UTC significant-severe reports (dots) and associated kernel density estimation (fill) for the period 1980–2009. Blue and red dashed lines approximate the locations of Interstates 40 and 20 respectively. Scale refers to reports per $km^2 \times 10^{-4}$. *Click image to enlarge. Click here for an animation of the annual cycle of significant-severe reports*.

between potentially severe and significant-severe environments, as shown in B03. Mesoscale factors such as convective initiation are obviously important, but are not examined in this study owing to scale and variable issues with the dataset employed. Therefore, environments portrayed in this study do not produce severe reports equally. For example, a large outbreak of significant severe weather on a given day may contribute greatly to the climatology of reports, but still would count as only one potentially significant severe-weather environment.

One possible reason that significant-severe environments do not represent reports well in southern Texas could be the threshold used for CIN in the presence of proximity C composite index values ≥20 000. Results from this study indicate that some of the largest decreases of potentially significant severe-weather environments, after incorporating CIN, were located in this region (not shown). To address this issue further, 0000 UTC MUCIN from reanalysis and collocated observed soundings were compared during May for the period 2002–2010 for the upper-air site at Brownsville, TX.

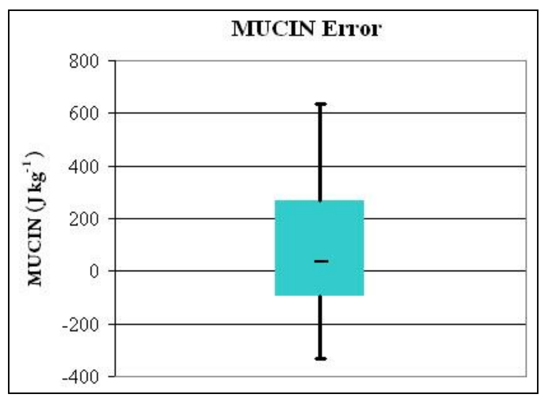


Figure 9: Box and whiskers plot of MUCIN error (J kg^{-1}) between observed and NARR 0000 UTC reanalysis soundings during May for the period 2002–2010 at Brownsville, TX. The shaded box encloses the 25th–75th error percentiles (interquartile range). Whiskers extend to the minimum and maximum values. Median value is denoted by black hash mark. Positive error values indicate an overestimation of MUCIN by the NARR. *Click image to enlarge.*

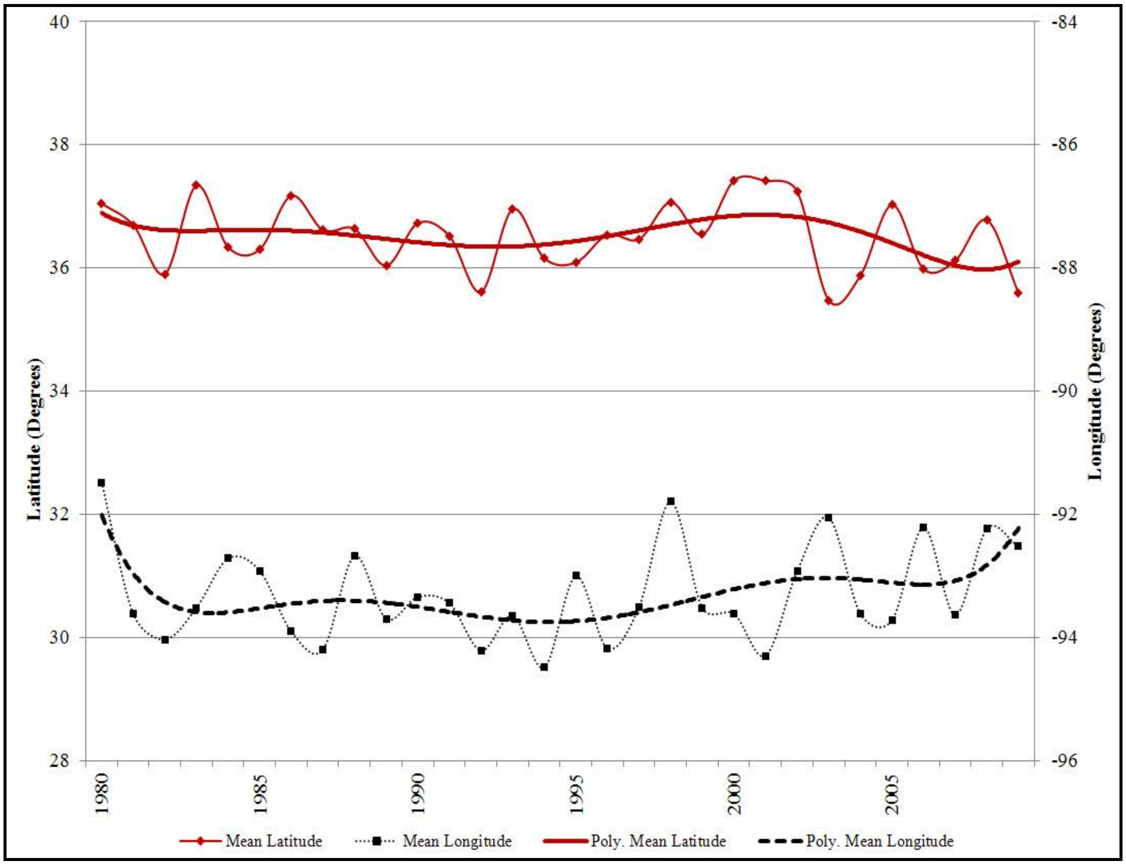


Figure 10: Mean center analysis of potentially significant severe-weather environments for the period 1980–2009. "Best-fit" lines are fourth degree polynomial functions. *Click image to enlarge*.

During this period, NARR soundings overestimated CIN by a mean value of 41 J kg $^{-1}$ (Fig. 9). While a robust, large-scale examination was not conducted, the results at least indicate that misrepresentation of CIN by the NARR could contribute to the discrepancy between proximity significant-severe environments and significant-severe reports. A subjective look at various soundings from this analysis indicates that major errors in MUCIN can occur with sharp temperature changes between the vertical pressure levels resolved in the NARR. A lack of event reporting also may contribute to this discrepancy, but the authors have not examined such an explanation due to the lack of a competing ground-truth report-assessment tool.

Monthly significant-severe reports also were analyzed to assess the performance of potentially significant-severe environments throughout the annual convective cycle (cf. Figs. 4 and 8). During the cool-season months, the proximity C composite parameter ≥20 000 in the presence of minimal CIN has difficulty capturing events. First, there are few reports during the cool season, and these events usually involve strong synoptic-scale forcing (Galway and Pearson 1981). Therefore cool-season reports likely would show a weaker diurnal signature and be less likely to be represented by the 0000 UTC reanalysis environments. These reports also may be grouped with synoptic-scale wind events, thereby not being individually classified in reporting databases (van den Broeke et al. 2005). In addition, cool-season events are climatologically characterized by a low-CAPE and high-shear environment (B07) that may not exceed the proximity C composite index threshold used. Thus, this parameter should not necessarily be used to *forecast* a significant-severe event, but rather, to discriminate between a potentially severe and significant-severe environment.

During the warm season, the potentially significant severe-weather environment climatology closely mimics the migration of significant severe-weather reports. In fact, a majority of significant severe-weather reports do coincide with the maximum frequency of significant severe-weather environments during April–September (Figs. 4 and 8). Presumably, this is because diurnally driven severe convection is captured adequately by the 0000 UTC reanalysis used. The largest difference between observed reports and potential significant-severe environments during the annual cycle again appears in southern Texas. This may be a factor of lack of lifting mechanisms (e.g., frontal passages) in this area, as well as reasons previously discussed. Subjective analysis of a few specific reports that fell outside of the significant-severe environment threshold on given days were characterized by either very large CAPE values with little deep-layer shear or vice versa.

f. Mean center

To assess whether potentially significant severe-weather environments have shifted over the 30-y study period, annual mean-center analysis was conducted. The potentially significant severe-weather mean center is the average *x* and *y* coordinate of all the annual environments for the conterminous U. S., weighted by the frequency of occurrence. This is especially useful for tracking changes in a spatial distribution over time, or for comparing the distributions of different types of spatial datasets. Mean-center analysis of potentially significant severe-weather environments suggests that there has been little to no change in the average position of the distribution (Fig. 10). Similar results were found (not shown) when the mean-center analysis was restricted to the active severe convective months of April, May and June.

5. Conclusions

This study establishes a U. S. climatology of significant severe-weather environments for the period 1980–2009 using reanalysis data. This has been accomplished on a global scale for a shorter period, but no prior study has employed new, high-resolution reanalysis datasets, such as the NARR, to analyze convective environments. Although the potential exists for an increase in severe convective environments under future climate scenarios (Trapp et al. 2007; Van Klooster et al. 2009), results indicate that significant severe-weather environments have been quite variable over the 30-y period, leading to no significant trend. To instill confidence in the climatologies presented herein, favorable significant-severe environments generated from reanalysis were compared to significant severe-weather reports. Unsurprisingly, potentially severe convective environments do not match reports. Overestimation tends to occur, as this climatology contains environments favorable for significant-severe weather, not necessarily environments that actually will produce such a report. Although difficult from a climatological perspective, future work should incorporate convective-scale lift into these types of climatologies, to better assess where severe convection may be favored.

For most locations, significant-severe environments from reanalysis show a strong annual cycle similar to that of observed reports and thus serve as proxy of locations that would favor significant-severe weather during a given time of the year. However, overestimations of significant-severe weather are most notable in southern Texas.

Because of the inherent problems with the storm-reporting process and resultant database (e.g., Doswell and Burgess 1988; Grazulis 1993; Brooks and Doswell 2001; Brooks and Doswell 2002; Doswell 2007), a better way to examine historical convective trends is by the use of environments that are known to favor severe convection via objective reanalysis techniques. Using methods similar to those herein, downscaling techniques similar to those used in Trapp et al. (2010), or other objective climatologies using remote sensing, are likely to aid in a better depiction of the climatology of significant-severe weather.

ACKNOWLEDGMENTS

The authors would like to thank Harold Brooks, Mace Bentley, and David Changnon for their comments and suggestions during early stages of this research. Reviewers Chuck Doswell, Harold Brooks, Jeff Trapp, and David Schultz all provided input that greatly enhanced the quality of this manuscript. Additionally, we appreciate editorial efforts by Roger Edwards during the submission process.

REFERENCES

Adams, D. K. and A. C. Comrie, 1997: The North American monsoon. *Bull. Amer. Meteor. Soc.*, **78**, 2197–2213.

Ashley, W. S., 2007: Spatial and temporal analysis of tornado fatalities in the United States: 1880–2005. *Wea. Forecasting*, **22**, 1214–1228.

Black, T. L., 1994: The new NMC mesoscale Eta model: Description and forecast examples. *Wea. Forecasting*, **9,** 265–278.

Brazdil, R. and Coauthors, 2002: IPCC workshop on changes in extreme weather and climate events. IPCC, Beijing, China, 107 pp. [Available online at http://www.ipcc.ch/pdf/supporting-material/ipcc-workshop-2002-06.pdf.]

Brooks, H. E., and C. A. Doswell III, 2001: Normalized damage from major tornadoes in the United States: 1890–1999. *Wea. Forecasting*, **16**, 168–176.

——, and C. A. Doswell III, 2002: Deaths in the 3 May 1999 Oklahoma City tornado from a historical perspective. *Wea. Forecasting*, **17**, 354–361.

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

——, J. W. Lee, and J. P. Craven, 2003b: The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data. *Atmos. Res.*, **67–68**, 73–94.

——, A. R. Anderson, K. Riemann, I. Ebbers, and H. Flachs, 2007: Climatological aspects of convective parameters from the NCAR/NCEP reanalysis. *Atmos. Res.*, **83**, 294–305.

Bouwer, L. M., 2011: Have disaster losses increased due to anthropogenic climate change? *Bull. Amer. Meteor. Soc.*, **92**, 39–46.

Bunkers, M. J., J. R. Wetenkamp Jr., J. J. Schild, and A. Fischer, 2010: Observations of the relationship between 700-mb temperatures and severe weather reports across the contiguous United States. *Wea. Forecasting*, **25**, 799–814.

Changnon, S. A., 2001: Damaging thunderstorm activity in the United States. *Bull. Amer. Meteor. Soc.*, **82,** 597–608.

——, 2009: Tornado losses in the U. S. *Nat. Haz. Rev.*, **10**, 145–150.

Clark, A., C. Schaffer, W. Gallus Jr., and K. Johnson-O'Mara, 2009: Climatology of storm reports relative to upper-level jet streaks. *Wea. Forecasting*, **24**, 1032–1051.

Craven, J. P. and H. E. Brooks, 2004: Baseline climatology of sounding derived parameters associated with deep moist convection. *Natl. Wea. Digest*, **28**, 13–24.

Crisp, C. A., 1979: Training guide for severe weather forecasters. Air Weather Service Tech. Note 79/002. Air Force Global Weather Central, Offutt Air Force Base, NE, 73 pp. [Available online at http://chubasco.niu.edu/projects/miller/.]

Cutter, S., B. Boruff and W. L. Shirley, 2003: Social vulnerability to environmental hazards. *Soc. Sci. Quarterly*, **84**, 242–261.

Doswell, C. A. III, 2001: Severe convective storms—An overview. *Severe Convective Storms*, *Meteor. Monogr.*, No. 50, Amer. Meteor. Soc., 1–26.

——, 2007: Small sample size and data quality issues illustrated using tornado occurrence data. *Electronic J. Severe Storms Meteor.*, **2** (5), 1–16.

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

——, and E. N. Rasmussen, 1994: The effect of neglecting the virtual temperature correction on CAPE calculations. *Wea. Forecasting*, **9**, 625–629.

——, H. E. Brooks, and R. A. Maddox, 1996: Flash flood forecasting: An ingredients-based methodology. *Wea. Forecasting*, **11**, 560–581.

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

Douglas, M. W., R. A. Maddox, K. W. Howard, and S. Reyes, 1993: The Mexican monsoon. *J. Climate*, **6**, 1665–1677.

Galway, J. G., and A. Pearson, 1981. Winter tornado outbreaks. *Mon. Wea. Rev.*, **109**, 1072–1080.

Gensini, V. A., and H. E. Brooks, 2008: Regional variability of CAPE and deep shear from reanalysis. Preprints, *24th Conf. on Severe Local Storms*, Savannah, GA, American Meteorological Society, P12.2.

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

Hales, J. E., 1988: Improving the watch/warning system through use of significant event data. Preprints, *15th Conf. on Severe Local Storms*, Baltimore, MD, American Meteorological Society, 165–168.

Houston, A. L., R. L. Thompson, and, R. Edwards, 2008: The optimal bulk wind differential depth and the utility of the upper-tropospheric storm-relative flow for forecasting supercells. *Wea. Forecasting*, **23**, 825–837.

Johns, R. H., and C. A. Doswell III, 1992: Severe local storms forecasting. *Wea. Forecasting*, **7**, 588–612.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77,** 437–471.

Kennedy, A. D., X. Dong, B. Xi, S. Xie, Y. Zhang, J. Chen, 2011: A comparison of MERRA and NARR reanalyses with the DOE ARM SGP data. *J. Climate*, **24**, 4541–4557.

Lanicci, J. M. and T. T. Warner, 1991: A synoptic climatology of the elevated mixed-layer inversion over the southern Great Plains in spring. Part I: Structure, dynamics, and seasonal evolution. *Wea. Forecasting*, **6**, 198–213.

Lee, J. W., 2002. Tornado proximity soundings from the NCEP/NCAR reanalysis data. MS Thesis, University of Oklahoma. 61 pp.

Maddox, R. A., and C. A. Doswell III, 1982: An examination of the jet stream configurations, 500 mb vorticity advection and low-level thermal advection patterns during extended periods of intense convection. *Mon. Wea. Rev.*, **110**, 184–197.

Mahmood, R., K. G. Hubbard, R. D. Leeper, and S. A. Foster, 2008: Increase in near-surface atmospheric moisture content due to land use changes: Evidence from the observed dewpoint temperature data. *Mon. Wea. Rev.*, **136**, 1554–1561.

McNulty, R. P., 1978: On upper tropospheric kinematics and severe weather occurrence. *Mon. Wea. Rev.*, **106**, 662–672.

Mesinger, F. G.,, and Coauthors, 2006: North American regional reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360.

Miller, R. C., 1972: Notes on the analysis and severe-storm forecasting procedures of the Air Force Global Weather Central. Air Weather Service Tech. Rep. 200 (Rev.), 190 pp. [Available online at http://chubasco.niu.edu/projects/miller/.]

NCAR, cited 2011: NCAR Command Language. [Available online at http://www.ncl.ucar.edu.]

NCDC, 2011: Billion dollar U. S. weather disasters, 1980– November 2011, 4 pp. [Available online at http://www.ncdc.noaa.gov/img/reports/billion/billionz-2011.pdf.]

Ramsay, H., and C. A. Doswell III, 2005: A sensitivity study of hodograph-based methods for estimating supercell motion. *Wea. Forecasting*, **20**, 954–970.

Rasmussen, E. N., 2003: Refined supercell and tornado forecast parameters. *Wea. Forecasting*, **18**, 530–535.

——, and D. O. Blanchard, 1998: A baseline climatology of sounding-derived supercell and tornado forecast parameters. *Wea. Forecasting*, **13**, 1148–1164.

Ruane, A. C., 2010a: NARR's atmospheric water cycle components. Part I: 20-year mean and annual interactions. *J. Hydrometeor.*, **11**, 1205–1219.

——, 2010b: NARR's atmospheric water cycle components. Part II: Summertime mean and diurnal interactions. *J. Hydrometeor.*, **11**, 1220–1233.

Silverman, B. W., 1986. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, 175 pp.

Smith, B. T., 2006: SVRGIS: Geographic Information System (GIS) graphical database of tornado, large hail, and damaging wind reports in the United States (1950–2005). Preprints, *23rd Conf. on Severe Local Storms*, St. Louis, MO, American Meteor. Soc., P2.6.

Stensrud, D. J., R. L. Gall, S. L. Mullen, and K. W. Howard, 1995: Model climatology of the Mexican monsoon. *J. Climate*, **8**, 1775–1794.

Thompson, R. L., R. Edwards, J. A. Hart, K. L. Elmore, and P. Markowski, 2003: Close proximity soundings within supercell environments obtained from the Rapid Update Cycle. *Wea. Forecasting*, **18**, 1243–1261.

——, C. M. Mead, and R. Edwards, 2007: Effective storm-relative helicity and bulk shear in supercell thunderstorm environments. *Wea. Forecasting*, **22**, 102–115.

Trapp, R. J., N. S. Diffenbaugh, H. E. Brooks, M. E. Baldwin, E. D. Robinson, and J. S. Pal, 2007: Changes in severe thunderstorm environment frequency during the 21st century caused by anthropogenically enhanced global radiative forcing. *Proc. Natl. Acad. Sci. USA*, **104**, 19 719–19 723.

——, E. D. Robinson, M. E. Baldwin, N. S. Diffenbaugh, B. R. J Schwedler, 2010. Regional climate of hazardous convective weather through high-resolution dynamical downscaling. *Clim. Dyn.*, doi: 10.1007/s00382-010-0826-y.

Trenberth, K. E. and C. J. Guillemot, 1996: Physical processes involved in the 1988 drought and 1993 floods in North America. *J. Climate*, **9,** 1288–1298.

——, G. W. Branstator, and P. A. Arkin, 1988: Origins of the 1988 North American drought. *Science*, **242**, 1640–1645.

Uccellini, L. W., and D. R. Johnson, 1979: The coupling of upper and lower tropospheric jet streaks and implications for the development of severe convective storms. *Mon. Wea. Rev.*, **107**, 682–703.

van den Broeke, M. S., D. M. Schultz, R. H. Johns, J. S. Evans, and J. E. Hales, 2005: Cloud-to-ground lightning production in strongly forced, low-instability convective lines associated with damaging wind. *Wea. Forecasting*, **20,** 517–530.

Van Klooster, S., and P. Roebber, 2009: Surface-based convective potential in the contiguous United States in a business-as-usual future climate. *J. Climate*, **22**, 3317–3330.

Verbout, S. M., H. E. Brooks, L. M. Leslie, and D. M. Schultz, 2006: Evolution of the U. S. tornado database: 1954–2003. *Wea. Forecasting*, **21**, 86–93.

REVIEWER COMMENTS

[Authors' responses in *blue italics.*]

We would like to thank all of the reviewers for their dedication to making this a better manuscript. Additionally, Roger Edwards deserves a big thank you for his e-mail responses and behind the scenes effort during the review process.

REVIEWER A (Charles A. Doswell III):

Initial Review:

Recommendation: Accept with major revision.

Overview: This manuscript is certainly based on worthwhile work that ultimately deserves publication. Unfortunately, the presentation has a fairly large number of issues that need to be considered and I've done a lot of wordsmithing suggestions, as well. See the attached document.

I would like to see the revised manuscript again. I'm leaning heavily toward acceptance, but I'd like to review the response by the authors.

The authors wish to thank you for your constructive and careful review that has contributed to a much-improved product. We have made substantial changes, both in organization and wordsmithing, to the manuscript. Only minor comments that were not accepted in the annotated version are elaborated on below.

[Replies to minor comments omitted...]

Substantive comments:

Although my general reaction to this work is to be supportive, I find the presentation to be rather poorly executed. There's a host of distracting minor annoyances that by themselves are not major, but their number is a major problem. This paper has the makings of an excellent contribution, but it definitely needs more work, both in wordsmithing and organizationally. There are several conclusions that appear before the evidence has even been presented! There are some generalizations that I don't believe are justified. And so on. I hope the authors will consider my suggestion for heavier smoothing of their spatial-field figures.

It sounds as if this [C composite index] is yet another attempt at a "magic" parameter, again without any evident physical basis (as noted by Doswell and Schultz, 2007).

The issues surrounding the use of a "magic" parameter from a forecasting perspective are discussed in section 2, including a reference to the mentioned article. We argue that this is not a forecasting approach per say, rather, we are developing climatology of the environments shown in Brooks et al. (2003b) to statistically favor significant severe weather. The further incorporation of CIN in our calculation was necessary to represent environments most favorable for such events [Brooks et al. (2003b) was unable to use CIN because it was not readily available in the NCEP/NCAR global reanalysis.]

Why would the region of highest frequency be most sensitive to climate change?

Is it possible you have misinterpreted our statement? We are not stating that U.S. significant severe weather environments are most sensitive to climate change. Rather, detecting statistically significant change in environment frequency would be easiest in a location that already observes a high frequency of said environments. Based on similar concerns from Reviewer D, we have opted to rework this sentence.

Just how big a step is the difference between a 3-year and a 7-year study?

The temporal difference was not the main addition to Brooks et al (2007) from Brooks et al. (2003b). The emphasis in Brooks et al. (2007) was addressing the interannual variability of DMC ingredients for specific locations.

What does "conservative" [smoothing] mean? A 3 × 3 Gaussian filter is a pretty light smoother! *My* notion of "conservative" smoothing would call for an even heavier filter and I believe the figures bear out my expectation of relatively "noisy" spatial contours! Attempting to identify "fine scale detail" in these fields is not justifiable, so I'd recommend a heavier smoothing than the one used here.

Yes, the term conservative was used rather loosely here. We have re-filtered all of our data using a two-pass scheme of the 3 × 3 Gaussian filter to produce less "noisy" spatial contours for all environments examined. Care was taken to preserve data integrity while producing the smoothest fields possible. Some of the shear fields may still seem a little noisy, but one has to remember that these fields were produced at a spatial resolution of 32 km. Therefore, some of the "noise" may rather be signal.

Do you really think you could detect, with accuracy and reliability, a climate change over a 30-y period of record?

Of course not. Nor is that the purpose of that statement. The idea here is that one may/may not be able to see a shift in severe weather environments over the 30-y period sampled. The reality is that the mean climatology of these environments is changing every day. If there was a change, it could prove beneficial for interested parties, especially those who may use these results in conjunction with other reanalysis datasets that have a longer temporal record.

Why [quadratic]? There are many other choices, including the Gaussian.

In this case, the choice was limited to quadratic as it is currently the only KDE available in our current version of ESRI's ArcGIS 9.3.1. However, our 250-km search radius used with a quadratic function is analogous to the 125-km Gaussian kernel used in Brooks et al.(2003a).

I don't understand the intent of the disclaimer here [treating every forecast scenario separately]. To what extent does BWD alone serve as a proxy for environments capable of producing significant severe weather? This sweeping generalization with regard to BWD seems unwarranted to me, and issuing a disclaimer doesn't justify the apparent claim. There might be ways to describe the importance of BWD on its own, but *this* narrative fails to be even marginally convincing.

After reexamining this section [4b. Deep-layer shear], we understand your concern. We have modified this section to omit the last portion of the paragraph. However, we would like to point out that the particular statement in question does not claim that BWD alone serves as a proxy for environments capable of producing significant severe weather. Our statement is making the point that organized DMC will not occur in certain BWD regimes.

[Overestimation of South Texas severe environments] is a big caveat that has not received herein the attention it deserves. Although the NARR has more vertical levels than the NCAR reanalysis, it still is not likely to result in an accurate representation of CIN.

We agree that it may not be the most accurate representation of CIN. However, as you are likely well aware, the NARR (and other types of reanalysis datasets) provide the best guess of the atmospheric state at a historical time by assimilating many different sources into the calculation of such derived variables.

[Minor comments omitted...]

Second and third review (green, combined due to dependent replies):

Recommendation: Accept with major revisions.

[Minor comments omitted…]

REVIEWER A (Charles A. Doswell III)

General Comments: This paper continues to frustrate me, because I believe the authors have done worthwhile work, but the resulting manuscript has a number of important issues that need to be resolved. I really want to recommend publication, but at this point, it's simply not possible for me to do so.

The authors would like to thank you for your careful review of the manuscript. We were a little surprised by your responses during this round of revisions, as it was certainly not clear to us during the first round that you had so many "major" issues with the paper. In fact, you mentioned during the first round of reviews that you did not think any of your recommendations were necessarily major. Nevertheless, we have tried to address all of your comments and incorporate them into the revised document. Your review was certainly thorough and no doubt contributed to the quality of the manuscript.

The "major" annotation was not intended to imply that the comment necessarily was a major problem with the paper, but rather to distinguish it from comments that wouldn't belong in the exchanges that will be published at the end of the final paper in EJSSM. It was for the editors. My apologies for this ambiguity.

The intended meaning of "to develop a theoretical model" is ambiguous. If by this the intent is a conceptual model of a favorable environment for severe weather, then this is precisely what ingredients-based forecasting is intended to *avoid*! To do this creation of a "theoretical model" is simply another pattern added to the collection of patterns used in pattern recognition. Ingredients-based forecasting uses ingredients that are *necessary* according to physical principles, not some "theoretical" model.

This section has been modified to remove "theoretical."

The difficulty with this response is that it misses the point. Ingredients-based forecasting is not associated with some sort of "model" (whatever its character might be: theoretical, conceptual, statistical, etc.). Rather, it's based on what are known to be necessary for the event in question to occur. Imposing a model of any sort on the process is contrary to the intent of ingredients-based forecasting.

CAPE is a nonlinear combination of the two proper ingredients: moisture and conditionally unstable lapse rates. Thus, it's *not* a proper ingredient! There's nothing wrong with using it in the way Harold et al. have done, but if this distinction isn't made clearly, then people will continue to misunderstand what "ingredients-based forecasting" means.

Good point. We agree that it is important to note (especially in this case) that CAPE is not an ingredient per se. Rather, it is potluck of other necessary thermodynamic characteristics that enhance the potential for theoretical updraft velocity.

See my first comment.

Unfortunately, I'm going to have to be convinced that the relatively coarse vertical resolution of the NARR data permits a *meaningful, reasonably accurate* estimate of CIN. I don't think having 45 instead of 28 layers represents an important difference. I know Harold felt that CIN couldn't be estimated effectively with the global reanalysis and I don't think it can be done using the NARR, either.

We disagree with this. The differences between the vertical and horizontal resolutions of the NCEP/NCAR Global Reanalysis and NCEP's NARR are quite different. It is worth noting that the additional levels are maximized in the lowest levels of the model sigma coordinates. Additionally, the RMSE for many of the variables (see http://www.emc.ncep.noaa.gov/mmb/rreanl/narr.ppt) is smaller. However, this indeed is *much of the reason why we only chose to look at 0000 UTC; when upper air observations were available to supplement the initialization of the ETA model.*

I don't see how anything in this response can be used to provide convincing evidence of a substantial basis for the authors' disagreement. By this I mean direct validation that CIN as determined directly from observed soundings is reasonably accurately reproduced by the pseudo-soundings based on the NARR data. Absent such a demonstration, the authors are, of course, free to disagree, but they have failed to convince me.

It's disturbing that this important piece of background information [global reanalysis data as a good approximation of severe-storm parameters compared to collocated observed soundings] is documented herein using only substandard publication (Lee 2002)!

Perhaps the reviewer or interested readers would be interested in examining convective variable relationships between collocated observed and reanalysis soundings with the authors? We have only taken the first steps in order to look into some of these values, but certainly not in enough detail for publication purposes.

This continues to be a weak point in this presentation.

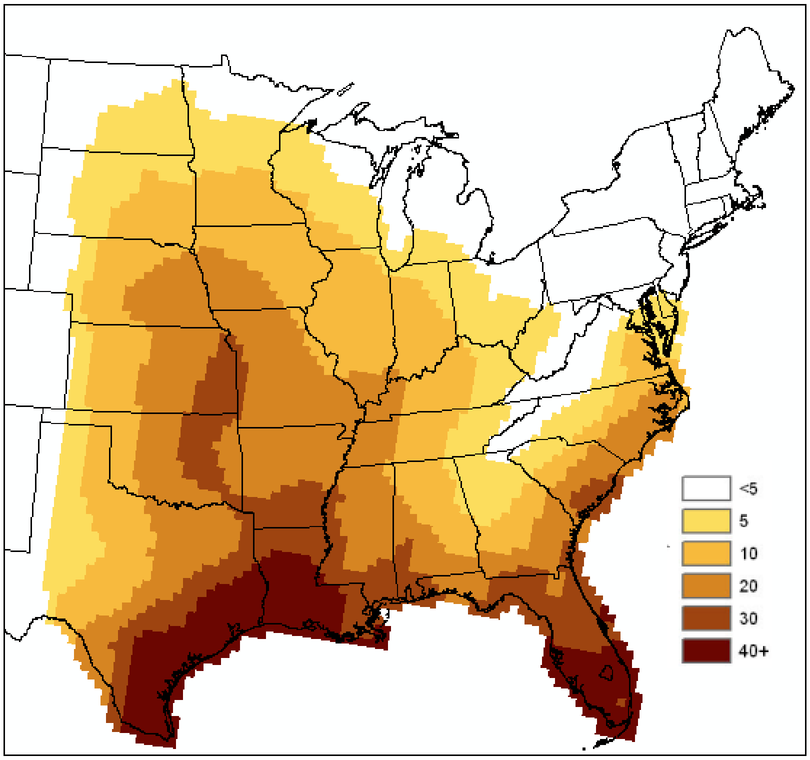
If you believe this [NARR CAPE and CIN should not be a major source of error in the C composite index] to be the case, why even bother mentioning the virtual correction?

While it may not be a large source of error from a climatological perspective, we felt it was worth mentioning as a general caveat to this CAPE climatology (at least for the reader's sake).

I still don't see the point.

A 3x3 Gaussian filter is a pretty light filter and that conclusion is evident in your figures, showing a *lot* of high-wavenumber noise still retained in the fields. Your cautionary statement to your readers is a perfect example of a strong bias I feel is common in meteorology—many of us wish to see either the raw data (full of all sorts of noise), or data only very lightly smoothed so we can see the "detail". A significant amount of the "detail" in meteorological data is simple noise (either sampling error or instrument error, or both). In the case of model fields, there are other sources of error. Meteorologists often want to retain as much detail as possible, even though the results contain a lot of physically meaningless variability (noise!). I assert that your fields need more smoothing, not less!

This has become a struggle for us, as one reviewer has suggested that all of the fields stay in the native NARR resolution as to take full advantage of the high-resolution motivation we discussed in the Introduction. We assert that our fields have already undergone substantial smoothing from their native resolution, and that this comes down to personal preference and "visual display" of the data. This is shown by a "raw" version of our Fig. 2 below. We have tried to keep both reviewers happy by coming to a happy medium that has left the departure of significant severe environments at the native resolution of the NARR, and have left our CAPE/Shear fields smoothed. While this is not likely to please everyone, it appears to be a solid compromise.



The fact that this is a struggle for the authors is not of concern to me. Regardless of the assertions of the authors, one need merely to look at the fields presented to see that they contain spatial features that are unsupportable by the data. The "raw" figure certainly is pixelated, but the "smoothed" version validates my concerns for the "details" of the fields. I understand that to some extent, this is a matter of opinion, and the authors are entitled to theirs. The authors seem to have taken an adversarial position—my intention with this criticism is to help the authors produce a better paper. If they insist on publishing figures with analyses that show unsupportable "detail", then they're free to do so. If they were to show the "raw" fields, I'd actually be happier than I am with this feeble attempt at smoothing.

[A 5-y running mean] is a pretty primitive temporal filter for a time series. It's often used because it's simple (I've done so!), but it doesn't usually give aesthetically pleasing results. A wider Gaussian filter provides a much smoother result and if you extract any small linear trend, you can repeat the data series beyond both the endpoints to obtain a smoothed version all the way to the endpoints. I can provide details, if you wish.

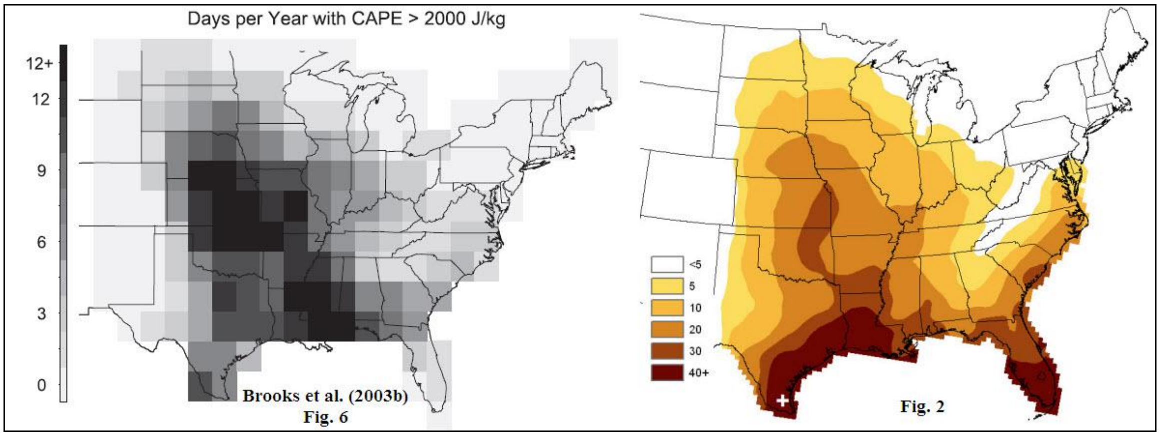
In an effort to produce such results, we have fit all of our graphs with fourth degree polynomials. The text has been modified accordingly.

As I noted earlier, an assessment of [a shift in significant severe weather environments] over a 30-yr period is not likely to offer much insight into the impact of changing climate on the severe weather environment.

We agree with this statement. However, at the least, we have been able to create a climate normal for the past 30-years in which comparisons can be made for future climate simulations.

When I look at, say, Fig. 2, it definitely appears to me that the contours are not as smooth as those presented in Brooks et al. (2003a). Hence, I'm inclined to dispute [that the figure is analogous to that of Brooks] and would like to see a comparison of the response functions for the filter used here and that used in Brooks et al. (2003a).

In fairness, I'm sure you have examined Brooks et al. (2003b). This work is much more comparable to the work presented in this manuscript (see figure below). There are other spatial filtering techniques (similar to ones used in our Fig. 8) that can be used on scattered vector based spatial data. Since we used gridded raster products (produced in netCDF format from the NARR), we were limited in the spatial filtering approaches we could use. In short, we think the filtering presented represents a reasonable representation of the "truth."



[Fig. 3] shows even more noise than Fig. 2, likely because vertical shear is a noisier variable (involving a vertical *derivative*) than CAPE (involving a vertical *integral*).

Exactly. Not much we can say other than some of the noise may be signal.

That statement is, taken at face value, an oxymoron. When the analysis permits unresolved features to be retained, then the fact that some of that might be "signal" is irrelevant. As noted above, I'd prefer "raw" data to this inadequately smoothed version. The comparison with Fig. 6 from Brooks et al. (2003) is entirely inappropriate because this particular figure shows *unsmoothed* results from a coarse grid. I was referring to the contoured figures in that paper, of course.

[Fig. 4] illustrates the usual noise problems, but most significantly, it also shows very clearly the inadequacy of the CIN estimates in the C-index to control the CAPE contribution. If you compare Figs. 2 and 3, it's quite evident that the best combination of CAPE and shear will exclude the maximum in CAPE in south TX and can be found in the central Great Plains. Your C-index fails to demonstrate properly what your data show. Compare this to Fig. 8, for instance.

We disagree. In our opinion, this is a misinterpretation of Fig. 4. If you look at the annual cycle, you will see that the reason southern TX shows a peak is due to favorable CAPE/CIN combinations in the cool season. Furthermore, examining variables separate from one another (as you mentioned by comparing Fig. 2 and 3) can be misleading. Remember that it does not really matter if shear is present but CAPE is not, or vise-versa. Perhaps the number of times CAPE and shear is juxtaposed is quite different!

I maintain that it's likely that the CIN variable is not being represented adequately by the NARR gridded fields. If you believe the problem is that I've not seen the variables (CAPE, shear, and CIN) together, then perhaps it would be best to show them together. Until I see some evidence to the contrary, I'll continue to suspect that the prevalence of apparently favorable environments far into south TX is likely the result of that problem. Since the observed severe weather doesn't peak in far south TX, it seems quite unlikely that a variable purporting to define severe weather-favorable environments that does peak in that area is working as it should. There must be some explanation for its failure in this regard—if it's not the vertical resolution affecting the CIN estimates, there has to be answer. If the author's can't provide one, then that needs to be made clear in the text. Unless the authors are willing to provide appropriate caveats to accompany this figure, I can't recommend publication.

Even a 49-y record is pretty limited in what it can show about climate change. 30-y averages are the “industry standard” for what is “normal” and so if you wanted to say something about how “normal” is changing, you’d need a 60-y record as a bare minimum, and it wouldn’t be very enlightening. Hence, the [trends here comparable to those in Gensini and Brooks (2008)] sentence probably needs to be more cautiously stated than it is now.

This is a fair statement. We have softened up the wording.

Re- "Even during spatially large below-average environment years, there are still locations that experience above-average frequencies and vice versa, illustrating the importance of understanding the difference in scale when examining severe convective environments." I understand what is being said here, and agree with it fully, but it seems to me that this begs further explanation. It might be better to omit this entirely, rather than to mention it but not give it adequate treatment.

Since we have added all of the departure years to the animation in Fig. 6, we have left this statement as-is. However, we are willing to delete it if you strongly feel it should not be included.

I think the discrepancy [severe environments highly overestimated south of Interstate 20 in Texas] is a consequence of your use of the C-index! You seem to be indulging in a kind of hand-waving, even to the point of suggesting it’s “trivial” to even suggest comparing the two maps. I don’t think this discrepancy is trivial at all – it begs a resolution.

We are confused by this comment. We also argue that this is not “trivial.” Perhaps you misunderstood the context of the sentence. For example, we state that:

> *“It is stressed that this particular index should not be used to forecast significant severe weather occurrence; rather, this index is beneficial in discriminating between potentially severe and significant severe environments as shown in B03. Mesoscale factors such as convective initiation are obviously important, but are not examined in this study owing to scale and variable issues with the dataset employed. Therefore, environments portrayed in this study do not produce severe reports equally. For example, a large outbreak of significant severe weather on a given day may contribute greatly to the climatology of reports, but would still only count as one potentially significant severe weather environment.”*

This is a very different argument than that which concerns me. Using this index to forecast would have the usual challenges associated with any such index, about which I’ve written. Rather, the arguments that concern me are about the discrepancies between the climatology of the observations versus the climatology of the C-index. This discrepancy is glaring and, while I have offered what I believe might be an explanation, either some convincing explanation should be presented or the deficiency of the C-index needs to be acknowledged.

If [a more representative climatology could be made with CIN thresholds of –50 to –25 J kg^{-1}], why was this not done?

It was not done because results from Bunkers et al. suggested that –75 J kg^{-1} would be the best choice for a threshold. We have removed this wording from the manuscript.

It’s beginning to look rather bleak for your C-index, in my opinion. A lot of this discussion seems determined to save the appearances and rationalize its inadequacies, rather than acknowledging that the wrong parameter may have been used.

That’s quite a claim. Based on results by Brooks et al. 2003b, this parameter certainly shows skill of discriminating between severe and significant severe environments. Climatologically speaking, we disagree with your notion that the C-index is not useful for delineating regions that favor significant severe *weather. Future research can certainly examine any parameter of choice, but all indices and parameters will still have issues! As you know, if we had a magic parameter, we would not be having this discussion.*

I've neither denied that it can be accurate (skill is a relative quantity), nor that it can be useful. But your results reveal what is obvious to me as a glaring discrepancy, I've already described in detail. What I find bothersome in this paper is the attempt to rationalize this discrepancy without offering a convincing explanation for its origin. If there is no convincing explanation available, then I'm seeking a reasonable assessment of the index without seeming to hand-wave the discrepancy away.

By what process has the smooth line been created? It's obviously not a 5-y moving average. Details of this "mean-center analysis" should be provided.

In an effort to produce such results, we have fit all of our graphs with fourth degree polynomials. The text has been modified accordingly.

[Minor comments omitted...]

REVIEWER B (Harold E. Brooks):

Initial Review:

Recommendation: Accept with minor revisions.

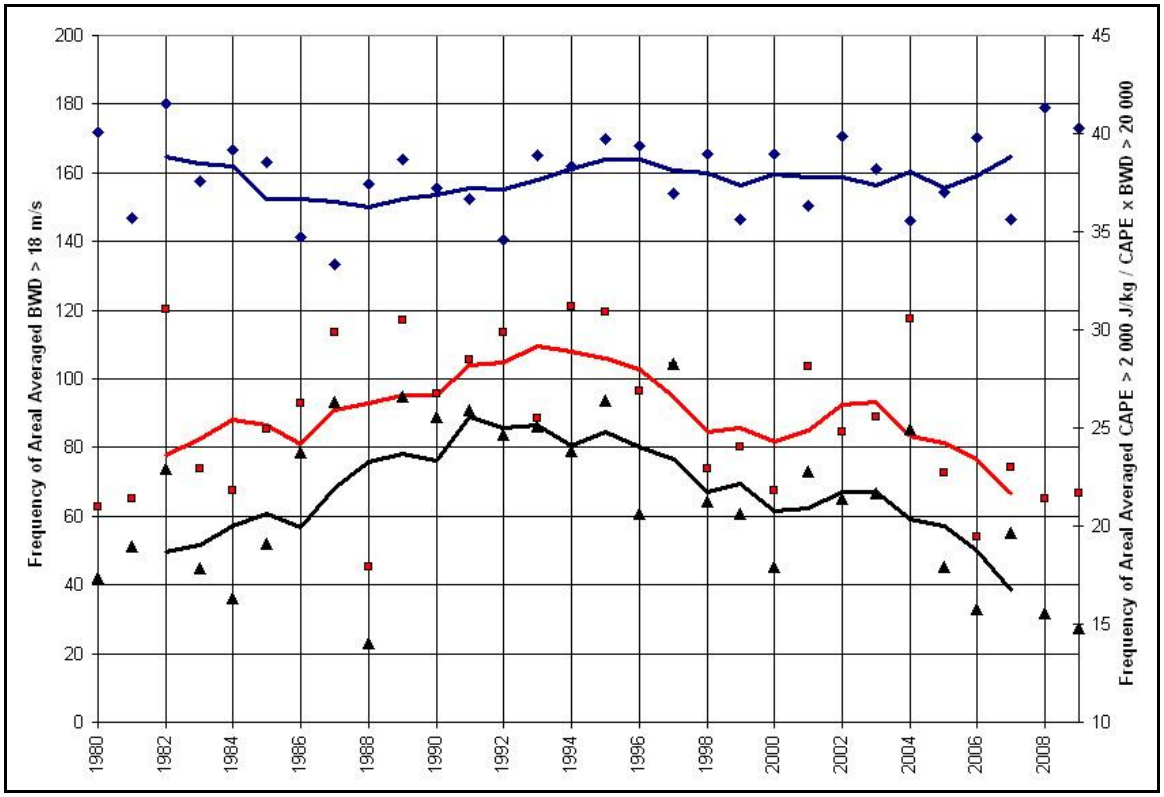
All minor comments and suggestions you provided were worked into the manuscript. The authors would like to thank you for your feedback, which undoubtedly enhanced the initial version of this manuscript. Specific major comments are elaborated on below.

The discriminator used in B03 is 2.86log(S6) + 1.79log(CAPE) = 8.36, not S6*CAPE >20 000. Although these are qualitatively similar, the B03 discriminator is sloped so that it emphasizes shear more. In addition, the B03 line is generally lower than S6*CAPE=20 000. Until CAPE >~4800, the B03 line is below 20 000. Depending upon the nature of the environments, this difference could be important, particularly in low-CAPE situations. For CAPE of 1000, the B03 line is at S6=11.1, instead of 20.

*While we were aware of the correct formula for discriminator, for ease of calculation we defaulted to the use of S6*CAPE >20 000 (i.e., we processed approximately 3.5×10^{7} grid points for each year of the NARR dataset and would have had to add four extra steps to our model that would have resulted in additional computing resources. While it is certainly possible to reproduce this exact discriminating threshold, this study fails to do so. This may explain why some of the low CAPE / high shear events subjectively examined (discussed in section 4e) were not captured. After careful consideration, we feel it would be better to present our approach as a proximity C-composite index. Wording throughout the manuscript has been adapted to reflect this change. In addition, we have included a small discussion concerning the difference of the calculations at the 2nd paragraph of section 3.*

What's the basic cause for the decline in the latter part of the record in most regions? Is it a CAPE effect, a shear effect, or the way the two come together?

Great question! The answer (as you probably figured) is a CAPE effect. To illustrate this, we have included a graph of our region 3 (Southern Plains).



This graph represents interannual variability of BWD *>18 m s^{-1} (blue), significant severe environment (red), and CAPE > 2000 J kg $^{-1}$ (black). Symbols correspond to actual values, while solid lines represent 5-y running means.*

Similar trends are found in the other three regions. 0–6-km BWD has remained relatively constant over the period of record. Thus, from a climatological perspective, CAPE is the driving factor behind the "trend" in significant severe weather environments during the period of record. We are willing to incorporate this into the manuscript if the editor deems appropriate.

In addition to the overestimate of reports, based on environmental estimates, in south Texas, it looks like the reports in the central Plains are displaced west of the environmental estimate. Any thoughts as to why?

One factor is likely the lack of accounting for a lifting mechanism (e.g., dryline advancement westward off of the high plains). The other is easterly upslope flow traversing the High Plains to the lee of the Rocky Mountains can accommodate organized DMC, especially under less than ideal conditions (Doswell 1980; Weaver and Doesken 1991).

*Doswell, C.A. III, 1980: Synoptic-scale environments associated with High Plains severe thunderstorms. Bull. Amer. Meteor. Soc., **61**, 1388–1400.*

*Weaver, J.F., and N.J. Doesken, 1991: High Plains severe weather- Ten years after. Wea. Forecasting, **6**, 411–414.*

[Minor comments omitted...]

REVIEWER C (R. J. Trapp):

Initial and second review (green, combined due to dependent replies):

Reviewer recommendation: Accept with minor revisions.

Summary: In a logical extension of the type of research initiated by B03, the authors of the current paper use NARR data, over a 30-y period, to examine a few parameters that characterize environments of severe convective storms. Perhaps the most significant finding is that regionally averaged parameters (like the C-index) exhibit little to no trend over the analysis period 1979–2009.

General comments: The research presented in this manuscript is sufficiently novel to merit publication. Although I have a large number of specific comments on the motivation, execution, and interpretation of the work, I don't view any of them as particularly major. One comment that arises in several of the points below regards a co-analysis of some of the variables that contribute to key parameters such as CAPE. For example, how does the trend of surface specific humidity, or of 700–500 hPa temperature lapse rate, compare to the trend of CAPE? In analyses of climate model data [see, e.g., Trapp et al. (2007, PNAS; 2009, GRL)], long-term positive trends and changes in CAPE correlated well with surface specific humidity, and thus helped to form conclusions about the possible effects of anthropogenic climate change. Regarding the quality of the presentation, I found the Introduction section to be somewhat lacking, especially in terms of motivating the study. For example, the third paragraph of this section provides a nice roadmap of the paper, but does not establish the true novelty of the work. Otherwise, the presentation is generally clear, and free of major grammatical errors.

We would like to thank you for your careful inspection and beneficial comments. We have incorporated all of your suggestions into the current version of the manuscript.

Are the monetary losses owing to these events adjusted for inflation? If not, the authors might want to re-consider how to discuss the frequency of such events over the 30-y period.

Yes, the losses are normalized to 2007 dollars using a GNP inflation index.

Can the authors be certain that no changes in tornado/severe thunderstorm event frequency have occurred over this period?

Given that events can only be verified (at the current time) if there is someone to witness the event occur, and then relay that report to the interested party, we cannot (nor can anyone else we argue) be certain that there have been any changes in tornado/severe thunderstorm event frequency unattributable to an increase in reporting.

My point here is that the authors are claiming that the trend is *only* from reporting, and do not allow for the possibility that an actual increase in frequency could be occurring. In other words, it is possible that the "increasing trend from losses" could indeed be due to an increase in event frequency, but we do not know how to separate out physical from nonphysical contributions to the trend. We have a paper in press in Climate Dynamics (online version is available) that suggests that in some regions, the trend may in fact be physical.

Although B03 considered environments over the globe, they didn't ignore the U.S., and indeed, the U.S. data were critical in the development of their methodology. The fact that "no study has focused solely on...environments across the U.S." is relatively weak motivation for the current work. What's much more relevant is that B03 used coarse-resolution global reanalysis: Finer-scale regional reanalysis are now available over a sufficiently long time to merit this type of analysis.

We agree with this statement. We have made substantial changes in this section to emphasize the higher spatial resolution available in the NARR as compared to the NCEP/NCAR global reanalysis.

Despite the discussion in this paragraph, the authors don't completely follow an ingredients-based approach. As Doswell et al. (1996) discuss, the basic ingredients for deep moist convection are ambient moisture, instability, and a "lifting" mechanism. Whereas CAPE is useful to assess parcel buoyancy, it does not give specific information about moisture, and it is not a strict measure of static instability. Parameters such as CAPE and CIN also do not yield information about the existence of a lifting mechanism, although the apparent need for one can be deduced from CIN. The lack of such information has been a main criticism of the environmental proxy approach (e.g., see Trapp et al. 2009, *GRL*) who used the existence of parameterized convective precipitation as evidence of sufficient lift.

We also agree with this assessment. While CAPE does not give specific information about absolute moisture content of a volume of air, when CAPE values exceed 2000 J kg^{-1} (our Fig. 2), there is likely sufficient moisture present to support DMC. From a climatological perspective, it is much harder to use an absolute ingredients based approach with thresholds for each absolute measure, as you would be sure to miss many environments.

Our experience with incorporating lifting mechanisms into this climatology (whether vertical velocity or convective precipitation is examined) encountered problems with the large degree of spatial variability that was present. This is likely realistic in an organized DMC scenario; however, using thresholds of these values was inappropriate, as it has never been examined in a discriminate sense in a study like B03.

Along these lines, are there other potential biases in the NARR data that we as readers should know? I seem to recall a number of papers published over the past 5-years or so addressing this very issue.

Can you be more specific or provide examples? We are only aware of the general caveats of reanalysis (e.g., while reanalysis serves as a "best guess" of the atmospheric state, it is only as good as the data being ingested).

The authors should, in the least, search abstracts with keywords "NARR" and "bias" in the AMS journal search engine. There are some biases related to the NARR's representation of the water cycle.

Some of this information should be used to help motivate the current study (see #3).

Wording in the last paragraph of the introduction has been modified to emphasize the benefits of using NARR data v. global reanalysis.

What is meant by a 0–6 km BWD "derived from the entire sounding" ?

This was a miscommunication on our part. Since 0–6 km BWD is not a variable in the NARR, we vertically interpolated the constant pressure levels to AGL height coordinates. We reworked this sentence to clarify.

I would assume that the authors used respective thresholds on CAPE and 0–6 [-km] BWD in the calculation of this C index? i.e., for CAPE >100 J kg^{-1} perhaps?

No restriction on CAPE or BWD was utilized in this study. While low CAPE environments are included, they are not likely to satisfy the CAPE × 0–6-km BWD threshold used in this study. However, they may indeed meet "significant severe" criteria based on the original discriminate equation by B03. Please see comments in reply to Reviewer B for more information.

In a related question, were the occurrences of 0–6-km BWD >18 m/s from days with nonzero CAPE? If not, this distribution may be reflective of non-convective environments.

This is a nice example of the annual cycle of CAPE, but the various attributions (to surface equivalent potential temperature, for example) would be more convincing with a corresponding analysis of surface equivalent potential temperature, or surface temperature and specific humidity. Also, why does transpiration contribute to late-summer CAPE ~2000 maxima in MO–KS–NE but not in IL–IN? Again, some additional evidence would bolster this interpretation. Finally, this argument seems to almost entirely ignore the seasonal progression of synoptic-scale storm tracks and mean jet-stream locations.

We believe that attributing changes in CAPE to contributing variables (i.e., lapse rates, surface θ_e, specific humidity, etc.) is a novel question worth further examination. Unfortunately, this aspect is simply outside the realm of this particular work. We have added a sentence to future work discussion calling attention to this topic.

The transpiration argument was mostly speculation, although the spatial resolution of NARR would likely be able to resolve such features. We have adjusted wording to reflect more emphasis on the mean jet-stream location.

I think it's unfortunate that the authors aren't willing to dig a little deeper. The data are readily available. These attributions would really strengthen the results.

I would encourage the authors to rethink this line of reasoning. As I recall, this C-index relates specific environments to significant severe weather occurrences but not necessarily to storm type. Why must C >20,000 necessarily indicate a supercell environment? Large C could arise from small shear but large CAPE, which would be more supportive of a bow echo that produces a significant number of severe reports along its track.

*This is a fair argument. We were basing our interpretation solely on discussions in Doswell (2001) and Doswell et al. (1993). These papers discuss that a **majority** of significant severe weather is associated with supercells. Since there is no study to our knowledge that addresses this issue exact issue, we have left this statement as-is. Other suggestions from you, other reviewers, or the editor are welcome in addressing this particular statement.*

Neither of these papers by Doswell involves a comprehensive examination of radar data which would be necessary to make such a conclusion. There are now papers by B. Gallus and collaborators (WAF) and R. Thompson and collaborators (SLS preprints, etc.) that more relevant. But why is such a statement necessary? You're relating a convective environment to severe thunderstorm occurrence.

Along the lines of what I mentioned [above], individual trends of CAPE and BWD would help explain these trends in C. This approach was taken by Trapp et al. (2007, 2009), and revealed high correlations between (climate-model) trends in surface specific humidity and CAPE.

See response to 2nd comment by Reviewer B.

Why didn't the authors show an annual cycle of CIN, and regional time series of CIN?

Since CIN is not independent of CAPE, we felt it was unnecessary to show such plots. Also, it would likely be argued (as with CAPE) that trends in favorable CIN are futile to help indentify organized DMC environments unless they are in the presence of favorable BWD regimes.

[Minor comments omitted...]

Third Review:

Reviewer recommendation: Accept.

Overview: I've read through the responses and revised manuscript, and am satisfied with how the authors addressed the comments from my previous two reviews. I think it's publishable in its current form.

REVIEWER D (David M. Schultz):

Initial Review:

Reviewer recommendation: Accept with major revisions.

General Comments: My comments on this manuscript appear as tracked changes and comments in the Word document uploaded to the EJSSM site. My concerns generally fall into the following categories [bulleted list below].

We thank you for your critical review that has greatly improved the initial version of the manuscript. We have made major changes to the preciseness of our wording and have made justifications for the methods employed clearer.

We were frustrated by your inability to see value added from this high-resolution reanalysis. Not only are the fields better resolved, but this study also examined the temporal component of the spatial products. Below, we have elaborated on the value added by this study and your other specific concerns within the categories you provided.

Substantive Comments:

* Lack of justification for the methods employed.

I don't understand the rationale for this statement [NARR resolution benefits outweighed the length of temporal record]. Why is having the highest resolution desirable?

Previous work has only examined potentially severe convective environments using relatively course global reanalysis data. Higher spatial resolution data from the NARR allows for the representation of such environments in more detail.

What is the justification for these four regions? These seem awful large and inhomogeneous in terms of severe weather.

These four regions were chosen so that trends in large regional areas could be examined. If the size of the region becomes too small, spatial variability may lead to a large coefficient of variability in the results. Regions were purposely made inhomogeneous to provide equal area coverage for all regions, and account for interregional spatial variability. Since we only examined four regions, we cannot make assumptions about sub regional scale trends (e.g., it may be possible that portions of a region are trending upward, while other areas in the same region are declining. This would lead to stable net average. Sub regional scale issues are mentioned in this work (section 4 d.), but would have to be addressed through future work by some other spatial index approach.

I don't really see the justification for these particular two years [1988 and 2001 in the environment-departures illustrations]. Why these two years? Why not other years? What is the point of showing these plots? That the environments vary from year to year? So what?

From our perspective, this seems like an overly critical observation. The purpose of Fig. 6 was to show that significant severe environments are much more local in scale than those represented by the regional average approach. It also provides a pathway for future work to examine a spatial index of areal coverage for these environments. As for the years chosen, the better question is "Why not?" These two years mark highly anomalous severe weather environment frequencies that readers can easily relate.

* Conflation of environments and reports.

Changes have been made throughout the manuscript to the wording as to not confuse the reader with terms such as "overestimation" when discussing environments and reports.

* Inappropriate speculation or uncited claims.

Please provide a reference for [“Parameter evaluation is the most popular and effective technique used in DMC forecasting.”]. Doswell and Schultz (2006), in fact, say the opposite and advocate ingredients-based approaches, not parameter/index evaluation.

Forecasters evaluate parameters that strive to describe the “ingredients” for DMC formation. We see nothing wrong with this statement.

* Lack of clarity in communication and precision in wording.

We have made major efforts to employ all of your suggestions into the manuscript, unless they conflicted with thoughts from other reviewers.

θ_e is a function of temperature and moisture. Therefore, it is not fair to single out just the solar heating component. Reword to be more precise.

Correct. However, the theoretical ability of a volume of air to contain water vapor is dependent upon its temperature. We have left the statement in question as-is.

The environments are not overestimated. You set the criteria and calculated the frequencies of the environments. Whether severe weather occurs during those environments is an entirely different question. Please reword to be more precise.

Done.

* Misinterpreting figures.

I don’t agree with [area most favored for significant severe weather in the eastern Great Plains where two ingredients overlap]. I think Fig. 4 looks very much like Fig. 2. I don’t see that Fig. 3 influences this field much at all.

Fig. 2 and Fig. 4 are different, both in magnitudes and in locations of environments. For example, after removing the CAPE >2000 J kg $^{-1}$ threshold, and incorporating 0–6 km BWD, the high Plains now see a dramatic increase in environments. Similar results are found in the northern Plains and Great Lakes regions. Moreover, the opposite is true in southern Florida. Although the annual averages may not seem that dissimilar, their differences are indeed important (perhaps more important are the differences found when examining the ***annual cycles*** *of Fig. 2 and Fig. 4.).*

I don’t understand [comparing the annual cycle of significant severe environments to observed significant severe reports]. Are the authors comparing Figs. 4 and 8? If so, I think their comparison is superficial.

We are not trying to verify that significant severe environments lead to significant severe reports. Rather, we are subjectively comparing the spatial distributions of two admittedly different datasets. However, we argue that a comparison of environments and reports is a fair question to ask, though we do state the caveats with such comparisons.

I am not sure that I saw that the high resolution analysis added anything to previously published results.

Would you rather see this research conducted with 2° resolution global reanalysis data? For example, compare our Fig. 2 to Fig. 6 in Brooks et al. (2003b). Not only are features better resolved, but this study examined thirty years of data compared to three years in Brooks et al. 2003b. Furthermore, Brooks et al. (2007) only analyzed the annual cycle of environments for two cross sections centered through Oklahoma City, OK. Based on initial replies from the other three reviewers, we see this critique as unjustified and unconstructive.

* Unnecessary sections of the manuscript (e.g., most of the second half of section 4).

I think the manuscript is strongest when it is limited to talking about environments. The manuscript is weakened by trying to link environments to the actual reports.

We simply disagree. We argue the manuscript would be weaker with no discussion concerning a comparison to observed reports. Please note that we are not trying to "link" environments to reports as stated a number of times in the manuscript; rather, we are simply providing a subjective comparison. Since none of the other reviewers had this argument, we have opted to leave this section as-is in the current version of the manuscript.

[Minor comments omitted...]

Second Review:

Reviewer recommendation: Accept with major revisions.

General Comments:

The manuscript has improved somewhat over the previous version, yet I felt that many of my concerns were not taken seriously. In addition, I still have concerns with the underlying motivation, the scientific method that the authors used, specific interpretations of their results, as well as numerous editorial comments to improve the clarity and precision of the text. Failure to adequately address these concerns will result in me recommending rejection.

ISSUES REMAINING FROM THE FIRST ROUND OF REVIEWS

I felt the authors did not adequately address some of my concerns in their responses. I list them in [these] groups.

1. RESOLUTION

In my first review of this manuscript, I raised the issue of why having the highest resolution was desirable for this study. I questioned whether the analysis provided by the authors added anything to previously published results. In response to my concern about the value of the NARR data for this study, the authors said, "We were frustrated by your inability to see value added from this high-resolution reanalysis. Not only are the fields better resolved, but this study also examined the temporal component of the spatial products. Below, we have elaborated on the value added by this study and your other specific concerns within the categories you provided." They also said my comments were "unjustified and unconstructive."

If the authors are frustrated by my inability to see the value in what they've done, perhaps I need to make my point more clear. I apologize for not being more clear previously.

My general concern has three specific components, which I address sequentially.

A. Have the authors motivated the reason for the use of this higher-resolution reanalysis?

We understand your concern regarding the use of temporal resolution in the current version of the manuscript. We have removed all temporal resolution wording since we chose to analyze only 0000 UTC environments.

On p. 2, the authors say, "no study has employed new high resolution [sic.] reanalysis datasets such as the [NARR] to examine convective environments. This particular reanalysis dataset permits researchers to examine historical DMC environments in more detail than ever before." On p. 3, the authors reiterate this point: "NARR...is preferred to global reanalysis data for this study due to its high spatial and temporal resolution." On p. 4, the authors say, "it was desirable to use the highest spatial resolution data available to examine environments in greater detail."

With these three statements, the authors motivate the reasons for using the NARR. The first sentence describes the "more detail than ever before" (but does not specify what kind of detail: spatial or temporal?), the second sentence mentions the "high spatial and temporal resolution", whereas the third sentence only mentions the high spatial resolution for again "greater detail."

Thus, I assert that the authors have not presented a consistent statement for whether they desire high spatial or temporal resolution; sometimes it is spatial, other times it is both. The authors need to have a clear and consistent statement as to the reason for why higher resolution is needed and whether higher spatial resolution, higher temporal resolution, or both are required.

B. Are the higher-resolution features that the NARR produces reliable?

It currently is the most accurate and reliable comprehensive climate dataset for research purposes (http://www.emc.ncep.noaa.gov/mmb/rreanl/narr.ppt).

NARR, like the NCEP/NCAR reanalysis, is a valuable tool for providing a consistent long-term dataset for climatological studies. Yet, nowhere in the manuscript do the authors address the strengths and weaknesses with using the NARR.

For example, problems with the NARR precipitation exist (West et al. 2007). Have the authors ensured that they are using the corrected version?

West, G. L., W. J. Steenburgh, and W. Y. Y. Cheng, 2007: Spurious grid-scale precipitation in the North American regional reanalysis. *Mon. Wea. Rev.*, **135**, 2168–2184.

In addition, on the NARR FAQ page <http://www.emc.ncep.noaa.gov/mmb/rreanl/>, precipitation discontinuities exist along the coastlines. Other studies on NARR precipitation include Becker et al. (2009, J. Clim.) and Bukovsky and Karoly (2007, J. Hydrometeor.). The authors do not discuss how these aspects affect their results. A presentation on the NARR site (Carrera et al. 2005) claims that the NARR has a positive bias of 1-3 deg C in 2-m temperature during the warm season in the central United States. This would change the values of CAPE presented by the authors, yet the authors do not discuss this known bias of the NARR. This study also reports that the Great Plains LLJ is well depicted, but perhaps too intense, in the NARR. As such, the authors ought to be citing this work and pointing out the benefits, but also potential places for exhibiting caution, when using NARR products.

C. Are the authors making full use of this high-resolution reanalysis?

I argue that the authors have taken a potentially valuable dataset and not used it to its fullest extent.

We are assuming this is referring to the high spatial and temporal component we stressed in our motivation. As suggested, we have removed all wording of the choice of using NARR data for its enhanced temporal resolution.

1. Indeed, the authors wish to examine the high temporal resolution data from the NARR. Yet, they only use 00 UTC analyses, not the 3-h data, which is superior to the 6-h data from the NCEP/NCAR global reanalysis. So, the authors do not use the full temporal potential of the NARR. Thus, to be precise, the authors should eliminate all claims of wanting to use the NARR because of the high *temporal* resolution.

2. In their response to me, the authors claim that they "also examined the temporal component of the spatial products." As far as I can see, this meant the annual cycle by month (Figs. 2, 3, 4, and 8) or the yearly values from 1980 to 2009 (Figs. 5, 7, and 9). The NCEP/NCAR Reanalysis could have been used for this purpose. The NARR was not needed to produce these types of plots. Thus, saying that "high temporal resolution" was why the NARR was desired for this study should be removed.

3. In their manuscript and in their response to me, the authors state that they wish to use the high spatial resolution data from the NARR. Yet, they appear to be of two minds in discussing why the NARR data is needed for their study. Consequently, they are not being consistent in their written presentation.

The authors' efforts to examine the high-resolution spatial reanalysis from the NARR are undermined by the following three smoothing operations that they perform on the data.

(a) Two passes of a "Gaussian (3 × 3) low-pass filter to help reveal spatial patterns". The authors admit that this smoothing masks fine scale details. If the authors were so intent on using the high spatial resolution of the NARR, why is smoothing applied? Are the "spatial patterns" not "revealed" at the full NARR resolution? In fact, Fig. 3 of the first draft of this manuscript shows a lot of detail in the wind difference field that is removed in the present version of this figure. To fully address my concern, the authors must address:
(i) why the full-resolution NARR data was not appropriate,
(ii) why they believe that the data in the original Fig. 3 needed to be smoothed,
(iii) why a 3 × 3 Gaussian filter was selected,
(iv) what the response function of the filter is so that the readers know how the data is being filtered (i.e., what scales are being lost from the full-resolution NARR data).

Reviewer A commented that our fields appeared too noisy and suggested further spatial filtering. We would have rather left all data at the initial native resolution of the NARR in some cases.

Please see our comment and reply from reviewer A.
[Omitted pasted text from reviewer A, Round 1 above, regarding the 3× 3 Gaussian filter]

(b) In Fig. 5, the authors apply a 5-y running mean to the yearly time series of severe-weather environments. (It is possible that the same filter is applied in Fig. 7, but that is not described in the caption.)
(i) Why five years, and not three or ten years?

Five appeared to be a good choice given the 30-y length of the dataset. These lines are simply for the reader to easily see the "smoothed" average of environments over time. We have changed this "best fit" line to a fourth degree polynomial function in response to a suggestion by Reviewer A.

(ii) Are you only interested in patterns with time-scales greater than five years? If so, then why is the NARR needed?

The interannual variability is intriguing, but not something this study chooses to examine. We are creating a frequency climatology of favorable environments for significant severe weather and examining them over the temporal record of the dataset.

(iii) What is the benefit of this filtered time series that the raw data doesn't show the reader?

The best fit polynomial function is intended to help the reader easily interpret the charts.

(c) In Fig. 1, the authors split the country into four regions. Time series for these four regions are subsequently analyzed in Fig. 5. In my first comments, I asked the authors "What is the justification for these four regions? These seem awful large and inhomogeneous in terms of severe weather." In their response, the authors said that "these four regions were chosen so that trends in large regional areas could be examined. If the size of the region becomes too small, spatial variability may lead to a large coefficient of variability in the results." The problems with their response are detailed below.

We can agree to disagree on this issue. Do you have a suggestion for new regional domains? We argue that PCA/Cluster analysis is not needed here when looking at the spatial distribution of significant severe environments. An earlier stage of this work examined COV to determine which areas were consistently the most favorable for such environments. The main issue here is that these locales were highly variable (not *surprising) from year to year! In keeping a constant (relatively large) domain size, we can capture larger scale characteristics of these parameters. In fact, we would not be surprised that at if some small scale, there is a statistically significant trend in some of these fields due to the lack of a data set longer than 30 y. This is purely speculation of course. Averaging over larger domains helps to remove some of this small-scale variability. While you may argue that this is not using the spatial resolution of the NARR to its full potential, we see it as the best way to portray trends in convective environments to readers.*

You are correct that portions of the domains fall outside of the U.S., and indeed, Storm Data was not available for such locales. Given that the regions remained constant throughout the study, this should not be a major issue. Also, as you suggested, we should be careful not to "link" environments and reports.

(i) Their response goes against their stated desire to have high spatial resolution. If the authors were ultimately going to break the central and eastern U.S. into four regions to examine trends in severe weather environments, then the global reanalyses would have been sufficient. Their response supports my contention that the authors are not using the NARR data to its fullest potential in this manuscript. Why must the authors combine data in this manner? Why not present the data at its full native resolution?

(ii) The four regions are arbitrarily chosen. They do not represent spatially coherent regions of severe weather occurrence, motivated perhaps by regions of similar severe weather occurrence in Fig. 4 (e.g., southern Texas, central Plains, east of Appalachians, New England). Why are the four regions arbitrarily chosen? More meaningful regions could be chosen that have the same area (if it were deemed important).

(iii) The authors have not motivated their choice of the four regions in the manuscript other than "to examine the regional variability of CAPE and the proximity C composite index" (except that isn't what is plotted in Fig. 5). So, this sentence needs rewording to be precise, and the four regions need to be justified (or eliminated from the manuscript).

(iv) A better approach that uses the full spatial resolution of the NARR data would be to present a spatial measure of variability. That way, the regions would be determined through an automated method (e.g., clustering), and the trends in the environments could be determined quantitatively. Books on spatial analysis techniques (e.g., clustering, PCA, EOF) may be of use here.

If these three smoothing operations (Gaussian filter, five-year running mean, four regions) remain in the manuscript, then the authors need to offer a new motivation for why the NARR data is needed. What is the point of having this high-resolution reanalysis and then smoothing, filtering, and lumping it together, reducing its effective resolution? Alternatively, and to the benefit of this manuscript, the authors should present the data at its native resolution and analyze it accordingly, with no smoothing, filtering, or lumping.

Fig. 6 shows the departures from the average significant severe-weather environment. These two figure panels are one way to show the spatial variation of the environments at the native resolution of the NARR. Why not show an animation over 1980–2009 of this field? In addition, you could normalize these departures by the mean at each point, so that regions of relatively infrequent severe weather could be contoured, as well. A third modification would be to overlay the observed severe weather plots or the kernel density plots (assuming that they have some way to address my point 3 below). Plots such as these would be several ways that this manuscript could be enhanced, without apologizing for smoothing, filtering, or lumping data. Doing so would make this manuscript much stronger.

We actually intentionally left all of the years out of this manuscript. There is enough data and analysis there to come up with a new study using a spatial index and actually coming up with some ranking index based upon such parameters [Similar to what has been done in Shafer et al. (2010) and Shafer and Doswell (2011).] Since this idea is not fully developed, we are not going to include it in this manuscript.

To summarize my point, I recognize the potential that the NARR provides for climatologies of severe-weather environments. I am disappointed that the authors do not pursue a systematic interrogation of the climatology of severe weather environments at the native resolution of the NARR. Instead, the authors have not motivated the reason for their use of the NARR, they have not demonstrated that the NARR produces reliable measures of the severe-weather environments, and they have not used this NARR data to its full potential. For this reason, I cannot recommend publication of this manuscript until these issues are addressed satisfactorily.

Please compare our Fig. 2 to Fig. 6 in B03 (included in response to Reviewer A). Perhaps it will be clearer just how much higher spatial resolution the NARR is compared to the NCEP/NCAR global reanalysis (~ 6.5 times greater).

We agree with reviewer A that some smoothing is indeed necessary (even if it just a simple bilinear interpolation) to help with the grid cell "noise." Otherwise, we are stuck between two polar opposite suggestions.

2. YEARS 1988 AND 2001 IN FIG. 6

In my initial review, I asked why the two years 1988 and 2001 were chosen for Fig. 6. I asked for clarification of why these plots were necessary. Instead the authors responded by saying that this was: "From our perspective, this seems like an overly critical observation." They also responded: "These two years mark highly anomalous severe weather environment frequencies that readers can easily relate [sic.]."

I am disappointed that the authors didn't take my questions seriously and address the questions I asked. It is the authors' job to describe their data, methods, and approaches for the reader. If the authors have not explained their method adequately, the readers may ask questions. Indeed, this is what happened in this case.

It's not that we did not take these suggestions seriously. There is indeed some miscommunication here at the fault of our description in the text. There were plenty of years that we could have chosen to compare and contrast, 1988 and 2001 just happened to be those years. Again, if necessary we can remove this figure. The point here was to illustrate how much severe environments can vary spatially across the U.S. from year-to-year.

(a) I asked why 1988 and 2001 were chosen for display in Fig. 6. The authors could have said in their manuscript that 1988 was chosen because it had the lowest value of environmental frequency in Region 2 over the entire time series and that 2001 was chosen because it was the highest frequency in Region 1 over the entire time series (Fig. 5). Consequently, as an illustration that our method demonstrates the regional variability of environments between different years, we are presenting 1988 and 2001 as illustrative examples. As a reader, I don't know if what I've written above is true, but it is one speculation as to what the authors were thinking in choosing to plot 1988 and 2001.

(b) As to what the authors mean when they say that "readers can easily relate", to what are they referring? I am a reader, and I don't understand what I am supposed to relate these two years to. Were these bad or good "storm-chasing" years? Am I supposed to relate them to the frequency of severe weather reports in Fig. 7? If so, 2001 is actually below average relative to other surrounding years. So, why do the authors claim that 2001 is "extremely active"?

You do not have to relate these environments to anything! You chose to do so in the initial review of this manuscript. However, I can contend that many people interested in severe weather climatology often talk about 1988 and its anomalously low severe weather occurrence, and thus are quite familiar. 2001 was arbitrarily chosen as it registered with anomalously large spatial areas with above average significant severe weather environments.

To summarize my point, the authors have shown two years, but they have not defended why these two years must be shown or what is significant about these two years. Adding additional text stating the authors' rationale will improve the manuscript. (In addition, I would highly recommend developing these types of plots for more years to show the variation of these fields at the native spatial resolution of the NARR data.)

We have modified the text a bit to help to try to make clear why we have chosen these years. Additionally, (as you suggested) we have added an animation with similar plots for all of the years examined to help make the manuscript stronger.

3. LINKING ENVIRONMENTS TO REPORTS

I made the argument that the manuscript is weakened by linking environments to actual reports. The authors provide reasons in their manuscript for why overlaying reports on top of calculated environments may not produce an exact correspondence, a point with which I agree. My point is that the authors have made it so clear that an exact correspondence cannot be met that what is the point of even overlaying the data in the first place? The authors argue that "we are simply providing a subjective comparison." But, they go much further with such statements as "the climatologies presented herein are indeed representative of observations" and "severe weather environments closely mimic the annual cycle of significant severe weather reports" (both from the abstract). These statements are inconsistent with denying linkages between environments and reports!

Significant changes have been made to the manuscript to make sure not to link environments and reports. The reviewer makes a fair argument and we agree that it should be made clearer that one should not try to link environments and reports.

I would argue that readers have no way of knowing why there is a relationship between reports and environmental frequencies or why there is NO relationship between reports and environmental frequencies. If that is the case, then what is the point of the overlay? To address these issues within the manuscript, the authors would need to discuss within each figure panel why there is a relationship or no relationship between the number of reports and environmental frequencies. To not do so would be to omit a crucial part of the science to this study.

I wish to make a final point about the authors' response. They say, "[Because] none of the other reviewers had this argument, we have opted to leave this section as-is in the current version of the manuscript." The authors misunderstand how the peer-review process works. Peer review is not a democracy where some comments are addressed and other minority comments are neglected. All comments must be addressed, regardless of whether the other reviewers raise the same issues or not. If changes are not made, the authors must argue their case to the satisfaction of the editor handling the manuscript, not the other reviewers.

In this case, I think I have raised sufficient concerns that the authors are not presenting a consistent message throughout the manuscript. In some places in the manuscript, the authors argue for a relationship between the reports and the environments. In other places, the authors argue for why such a relationship should not exist. They can't have it both ways.

4. θ_e

The authors make the claim that "higher surface θ_e values begin to spread northward with increasing solar declination." I argued that because θ_e was a function of temperature and moisture, it is not fair to single out just the solar heating component. The authors responded with "the theoretical ability of a volume of air to contain water vapor is dependent upon its temperature," and left the statement unchanged. The statement remains incorrect and is incorrect on many levels.

We have modified this statement. The reality is that CAPE can be thought of as a function of surface θ_e and mid-tropospheric lapse rates (our original statement).

a. θ_e is not shown in their figure, CAPE is. So, whatever the authors are inferring about surface theta-e is not shown by their manuscript.

b. As far as I can tell, "solar declination" is just a fancy synonym for latitude, so just say "latitude," if that is what you mean. However, CAPE is not a strong function of latitude at all. Clearly, CAPE is largest near the Gulf of Mexico and southern North Atlantic Ocean, disrupted by the Appalachians, as seen in the authors' Fig. 2.

c. θ_e is a function of the current temperature and mixing ratio, not the current temperature and the saturation mixing ratio (which is a function of temperature).

The statement as written needs to be deleted. Discuss the figure, not some unscientific caricature of reality.

[Minor comments omitted...]

Third Review:

Reviewer recommendation: Decline.

Final Review and Statement: I have read the revised version of this manuscript. I don't know where this peer-review process went off track, but it did. I think the authors have done some potentially interesting research. I think they are at the start of providing more insights into how severe-storm environments have changed over time. I would have liked to have seen this paper published in EJSSM pending the revisions that I have recommended.

In the previous rounds of the peer-review process, I have tried to give genuine suggestions to improve the paper. Many suggestions have been accepted by the authors and have improved the paper. Many others, I am sorry to say, haven't. I feel like I have worked in good faith to help make the paper publishable by helping the authors (a) to better describe what they did and (b) to get them to explain their results. For whatever reasons, the authors and I have not agreed about the best ways to do this.

I have found them argumentative and obstinate. They have not addressed my concerns. The result is that I don't believe the authors have brought this paper to a publishable state. My reasons are the following.

A. The manuscript as it stands in this third revision is not substantial enough to warrant publication. Some results are new and interesting, but others are mostly argued with speculation. Their hypotheses are not tested using data. If we boil down what the new and interesting results are in this manuscript, what are they?

1) Figs. 2, 3, and 4 are new and interesting. With further development and support from additional calculations in the text, this content could be publishable.

2) Fig. 5 isn't interesting because of the arbitrary way the regions were defined and because the authors did nothing in the text to discuss the reasons why storm environment frequency [what are the units?] went up and down in the various years and in the various regions. Figures like this without an explanation of what they mean are not satisfying to readers.

3) Fig. 6 would be a potentially interesting result, if the results were placed into some context (same issue as before). Other than knowing that the distributions of storm environments differ in an above-average storm year and a below-average storm year, what is the point of this figure? The results in this figure are not related to the synoptic context or climatology of those years. For example, if the authors had provided maps of the 500-mb fields during these two years, that would be one step that would help the reader better understand why these years were different. I thank the authors for including the animation of all the years.

4) Fig. 7 supports the increase in number of storm reports over time. This figure (or a citation to this result) is needed, but this figure is not a new result.

5) Fig. 8 is interesting, but is also not a new result. Furthermore, there are many reasons why this figure doesn't look like Fig. 4. The authors offer some arguments for this, but they don't provide any evidence. I believe that simple explanations about frontal passages not reaching southern Texas or threshold values of

C being too high are not sufficient to explain all the reasons why an environment can be favorable for severe storms but not produce them.

6) Fig. 9 is not a very interesting result. I don't see what this proves.

Other concerns:

6) Where is the evidence to support these arguments: "Events are usually strongly synoptically forced," "would likely show," "may by [sic] lumped in", "typically characterized...that may not exceed"? It's all too much speculation.

7) I am not convinced by the arguments offered in the manuscript as to why Fig. 4 does not look like Fig. 8. It's not south of the Dallas to Midland line that is really where the discrepancy starts. It's really in central Oklahoma and southward. In Fig. 8, south of central Oklahoma is roughly where the filled field starts to decrease. In contrast, in Fig. 4, the value of C appears to increase monotonically to the south until the white +. The southeast U.S. Atlantic coast is also highlighted in light brown in Fig. 4, but is not shaded at all in Fig. 8. In Fig. 4, there is a gradient in C from Louisiana to Florida. I don't see the same magnitude in gradient in Fig. 8 (although the small size makes it difficult to be sure). As I've argued above, the explanations offered in the text are not convincing and remain untested.

Thus, to summarize, only four figures are new results and worthy of publication. Although we are presented with interesting results in the manuscript, the explanations offered by the authors are mostly speculation rather than argued with evidence and supporting calculations. The issue is that credible scientific explanations for these fields have not been presented. Some explanations offered by the authors are plausible, others are not. Some of their explanations are not even consistent with the figures. Some are merely speculation. Some speculation is allowable when all avenues have been pursued with the data. Unfortunately, the authors have the data, but have not tested their hypotheses using the data. They have just presented figures.

B. Although I have argued against some of the interpretations in this paper, they have not been adequately revised.

1) "Regional comparisons illustrate potentially significant severe environments have changed little over the 30 y." In fact, Fig. 5 shows quite a bit of variability. Region 2 experiences a range of a factor of two variability from year to year (what they refer to as "mostly unchanged". Region 3 ranges from values near 20 to over 30, peaking in the mid 1990s. These curves, especially for regions 1, 3, and 4, have not "changed little."

2) I have argued previously that there is nothing special about these four arbitrary regions. The authors have not justified why these four regions. I don't agree that averaging incredibly rich structure from the high-resolution NARR output (as is seen in Fig. 6) into four arbitrary regions is the best way to analyze this data. The authors do not agree. But, they don't defend why they chose these arbitrary regions. Why not four other regions. Why not eight regions? I just don't get it.

C. Methods and limitations are not fully appreciated by the authors.

I agree with the authors that the NARR is a fine product for this job. I never questioned why someone would create a climatology using the NARR. That is not the argument that I am trying to have. Instead, my concern was on the validation of the model. The authors' single sentence response "most accurate and reliable...dataset" isn't approaching this issue of storm environments with the degree of skepticism that is warranted. Indeed, another reviewer also picked up on the potential problems that the NARR might have. Just because there are problems and limitations does not mean that the analysis should not be performed. Instead, the authors should be fully cognizant of its strengths and limitations. The authors have failed to provide substantive rebuttal to my arguments.

I asked whether they were using the corrected version because of precipitation errors. They did not respond. I asked the authors to comment on precipitation discontinuities and warm biases along coastlines. They did not respond. Other comments ("5. Other comments that were unaddressed..." and "Specific Comments") were not responded to.

D. I accept that the authors had to balance two different reviewer concerns: one reviewer's desire for heavy smoothing with my desire for light smoothing. I still argue, "What's the point of using a high-resolution dataset if it is going to be smoothed?" Nevertheless, I drop my objection to their smoothing approach. The other reviewer is the expert on filtering; I defer to him.

E. Other issues should be fixed.

[Minor comments omitted...]

[Editor's note: Given 1) the otherwise publishable aspects of the paper and two favorable reviews (B and C), but also, 2) several valid and crucial remaining concerns of Reviewers A and D, the editor performed a post Round 3 "editorial review", ultimate acceptance being conditional on addressing the editor's concerns satisfactorily. As EJSSM is an open-review journal, the substantive part of that review/reply follows. The paper was accepted following satisfactory completion of this process.]

Editorial Review (Roger Edwards):

MAJOR:

Throughout (SPECULATION): Every speculative statement throughout must be either 1) eliminated, 2) specifically characterized as speculation, or 3) converted into actual analytic results that are presented in support of the statement. As I often say: "Specify, not speculate."

The two biggest offenders (this is not an all-inclusive list):

- Section 4e: "…cool-season events are typically characterized by a low-CAPE and high-shear environment" based on what citation or analytic results? Please specify, not speculate. [At least you cited actual references for the other speculative-looking statements in that paragraph to which Schultz objected; as such, I can let those stand.]

We would disagree that this is speculation. [Tangential comment omitted.] Perhaps this was just more of an unsupported claim rather than speculation. Either way, we have added a citation to Brooks et al. (2007) where it is stated that, "As the spring progresses, the environments change from being high-shear, low-CAPE to being high-CAPE, low-shear" and have replaced the word "typically" with "climatologically."

- Section 4e: ". One *possible* reason significant severe environments do not represent reports well in southern Texas *could be*.." [Italics are mine for emphasis on weak, wishy-washy language.] Is it or isn't it? Provide a definitive answer derived from explicitly mentioned examination of data that can support or refute such a statement. This disconnect between South Texas severe environments and (lack of) reports is very noticeable, and important to the reader's understanding of the validity of your analytic methods. As such, it screams for more solidly justified explanation than has been provided so far. Try showing CAPE and CIN together, as Doswell suggested, to illustrate specifically whether that is a contributing factor. State clearly whether the vertical resolution affects CIN estimates, and briefly but specifically explain how you came to that conclusion.

Agreed that this is wishy-washy language. Simply, we are not sure why this distribution shows such a discrepancy. However, we have attempted to answer analytically this speculation with the addition of a figure (new Fig 9) [and referring text].

Section 3a: As Reviewer A (Doswell) requested, document, using a literature citation and/or analytic data and results, that "CIN as determined directly from observed soundings is reasonably accurately reproduced by the pseudo-soundings based on the NARR data analytic data and results." Otherwise, that is a contention with little or no basis.

There is no response to this statement other than to say that there is no peer-reviewed literature to our knowledge that addresses this specific issue. As mentioned in our manuscript, Lee (2002) addressed such concerns in a M.S. thesis, albeit non-published:

> *"Lee (2002) showed that global reanalysis data provides a good approximation of parameters pertinent to severe storms when compared to collocated observed soundings. It is strongly likely, but not guaranteed, that other reanalysis datasets will behave similarly."*

We have added an additional sentence following this statement as a general caveat: "If the NARR fails to reproduce convective variables accurately, then it is possible that plots presented herein could be at least partially incorrect." It is worth noting that the citation to Lee (2002) was also used in B03 for the very same purpose. Finally, the attempt at an analytic description of the discrepancies found is discussed in the prior review response.

Section 3b (re: Fig. 5): Specifically explain *why* you chose these regional domains and why you chose this number of them. If arbitrary, say so. I suspect there's a hypothesis behind your decision to choose the domains…was there? If so, say so: what was the hypothesis, how was it tested, and how do the results fit or counteract that testable hypothesis?

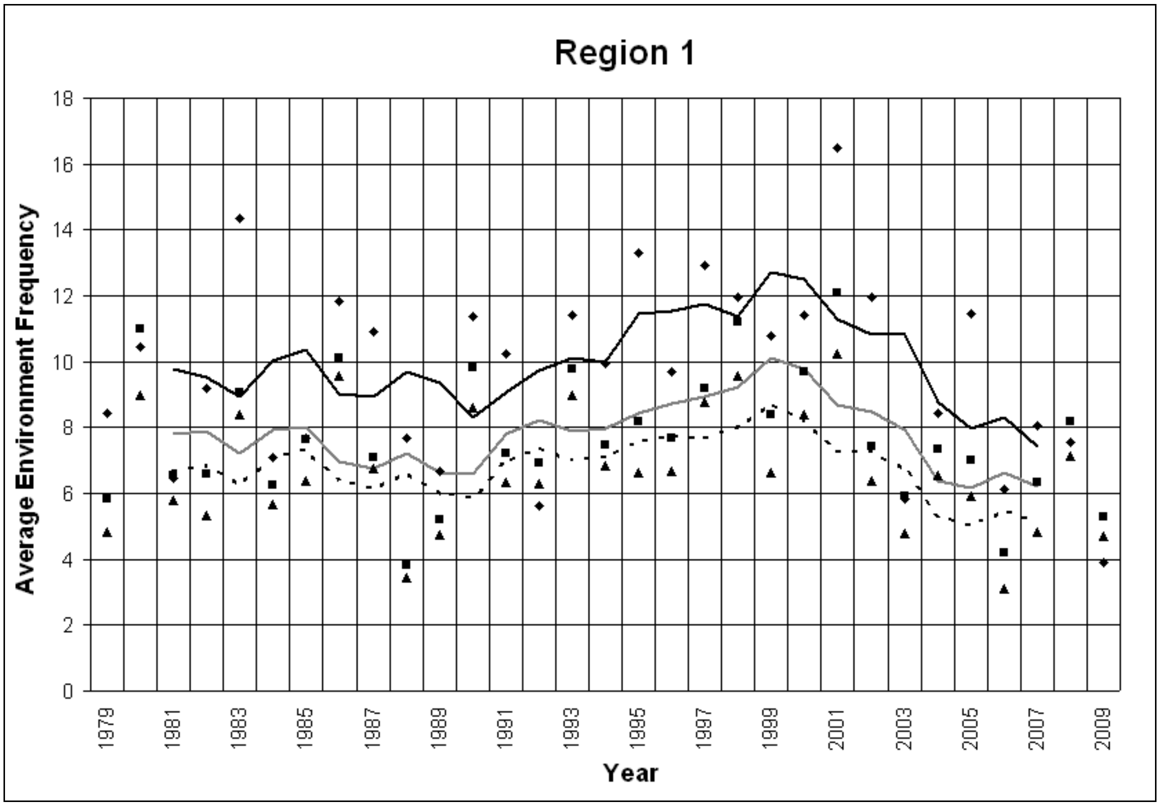
We have added a sentence regarding the choice of domain locations. These domains were actually chosen during a very early stage of the research in which Harold actually suggested four domains rather than six to get more of a "regional average" depiction of trends in environments.

> *"The domains were arbitrarily chosen to analyze if there any latitudinal (e.g., region 3-to-1) or longitudinal (e.g., region 1-to-2) shifts in regional trends of proximity significant severe weather environments."*

[Although initially termed "minor", the following comment and reply includes a potentially important and clarifying point with graphic illustration.]

Section 4d (re: Fig. 5): Specifically state why storm-environment frequency oscillated as it did, in the various years and in the various regions. Those are the trends in your data, in that figure…so: explain them. If you don't know why, say so. I suspect you can provide a justifiable reason for those trends without speculating; but if you do speculate, please do specifically note that's the nature of your statement.

Specific analysis for each year shows that it is CAPE that is predominantly the factor that governs the trend in environments on any given year. See [Supplemental Fig.]:



Supplemental Figure: Interannual variability of average large CAPE "♦", significant severe "■", and significant severe with minimal CIN environments "▲" for Region 1. Lines indicate 5-y running means of average large CAPE (black), significant severe (grey), and significant severe with minimal CIN (black dashed) respectively. Units are environments y^{-1}.

We have added a sentence in section 4.d. indicating that CAPE is the main governing variable in the trends of the proximity C-composite. We left this figure out of the final version of the manuscript and only included the above paragraph (we do not think that this figure is vital to the paper).

[Remaining minor comments omitted...]