Climatology and Analysis of High-Impact, Low Predictive Skill Severe Weather Events in the Northeast United States

MATTHEW T. VAUGHAN, BRIAN H. TANG, AND LANCE F. BOSART

Department of Atmospheric and Environmental Sciences, University at Albany, State University of New York, Albany, New York

(Manuscript received 28 March 2017, in final form 28 August 2017)

ABSTRACT

This study identifies high-impact severe weather events with poor predictive skill over the northeast United States using Storm Prediction Center (SPC) convective outlooks. The objectives are to build a climatology of high-impact, low predictive skill events between 1980 and 2013 and investigate the differences in the synoptic-scale environment and severe weather parameters between severe weather events with low predictive skill and high predictive skill. Event-centered composite analyses, performed using the National Centers for Environmental Prediction Climate Forecast System Reanalysis and the North American Regional Reanalysis, suggest low predictive skill events occur significantly more often in low-shear environments. Additionally, a plurality of low probability of detection (POD), high-impact events occurred in low-shear, high-CAPE environments. Statistical analysis of low-shear, high-CAPE environments suggests high downdraft CAPE (DCAPE) and relatively dry lower levels of the atmosphere are associated with widespread severe weather events. DCAPE and dry boundary layer air may contribute to severe wind gusts through strong negative buoyancy and enhanced evaporative cooling of descending saturated parcels.

1. Introduction

The northeast United States (hereafter the Northeast) provides a unique and challenging environment for forecasting severe weather. The densely populated urban areas of the Northeast magnify the hazards of severe convection. The Northeast harbors a varied landscape of mountains, valleys, lakes, and rivers while bordering the Atlantic Ocean and the easternmost Great Lakes (Fig. 1). Interactions between boundary layer flow and terrain create complex convective environments that can influence storm development and severity (e.g., Riley and Bosart 1987; Wasula et al. 2002; LaPenta et al. 2005; Bosart et al. 2006; Lericos et al. 2007; Tang et al. 2016). Lake and ocean boundaries, along with the aforementioned terrain variations, create additional sources of vertical motion through surface convergence and upslope flow, and may modify low-level winds and enhance low-level vertical wind shear (Bosart et al. 2006; Tang et al. 2016). Because of the combination of a challenging forecasting environment and the high population density, severe weather events can occur in synoptic-scale environments considered marginal for severe convection, leading to underpredictions of the severity and/or coverage of severe weather. We qualitatively describe such underpredicted events as having low predictive skill and seek to identify low predictive skill severe weather events that impact a large area of the Northeast in order to better understand the environments that lead to such events.

The predictive skill of severe weather events needs to be objectively quantified, and one way to quantify skill is to evaluate the performance of SPC convective outlooks. Hitchens and Brooks (2012, 2014) evaluated the skill of SPC convective outlooks on a national scale. They found increasing forecast performance for 0600, 1300, 1630, and 2000 UTC day-one convective outlooks over the last few decades. There was often little difference in skill between consecutive forecasts. They also noted an increase in severe reports over the 1973–2010 period, which is important to account for when assessing skill and the impact of events. The increase in severe reports is a part of a larger challenge when handling trends in reporting practices and frequency (Doswell et al. 2005).

We will use a technique similar to that of Hitchens and Brooks (2012) to quantify the predictive skill of severe weather events, concentrating on the Northeast.

Corresponding author: Matthew T. Vaughan, mvaughan@albany.edu

DOI: 10.1175/WAF-D-17-0044.1

© 2017 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

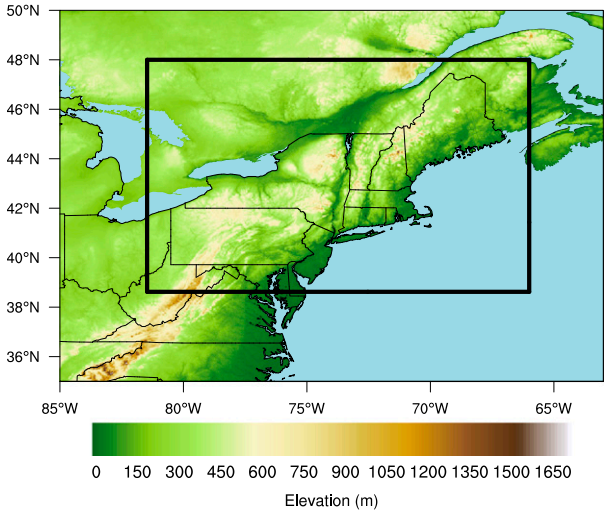


FIG. 1. Relief map (m) of the northeastern United States. The black box outlines the Northeast region used in the study.

However, our ultimate goal is not a verification of SPC convective outlooks, but to identify specific events that had low predictive skill (i.e., where a forecaster did not predict the severity or coverage of an event well). Additionally, given the inhomogeneous reporting practices with time, we need to account for the increase in severe reports in order to identify which events were also high (low) impact, that is, those that have a relatively large (small) number of reports above (below) a time-evolving threshold. Identification of low predictive skill, high-impact events will allow us to study the frequency and environmental conditions of these severe weather events and compare them to the climatology of Northeast severe weather events.

Previous studies have investigated the diurnal and seasonal variability of convection, including severe convection, across the Northeast (Lombardo and Colle 2010, 2011; Hurlbut and Cohen 2014). Hurlbut and Cohen (2014) examined severe convective environments using observed soundings over the Northeast during an 11-yr period and found that 498 of 742 (67%) severe weather days occurred during June–August (JJA). A severe weather day is any 24-h period beginning at 1200 UTC with at least one severe weather report. The peak time for the first severe report of a severe event was 1900 UTC and only 4% of first reports occurred between 0600 and 1200 UTC. Additionally, Hurlbut and Cohen (2014) documented 10 161 severe wind reports, 1732 severe hail reports (4413 reports of hail greater than 19.1 mm), and 190 tornado reports over the study period, indicating that severe wind is the primary severe weather threat across the Northeast. However, storm reports may be reported from wind damage alone, which can allow damaging wind gusts below the severe wind threshold of 25.7 $m\,s^{-1}$ to be recorded as severe wind reports in the database.

Reanalysis studies suggest the Northeast convective environment produces comparatively fewer days with favorable significant severe weather parameters than regions farther south and west (Brooks et al. 2003; Gensini and Ashley 2011). Gensini and Ashley (2011) found the Northeast had fewer than five 0000 UTC soundings with most-unstable convective available potential energy (MUCAPE) values above 2000 $J\,kg^{-1}$ each year while 0000 UTC MUCAPE values above 2000 $J\,kg^{-1}$ are more common in the Great Plains and Midwest. Lombardo and Colle (2011) investigated severe events, consisting of any convection that produced at least one severe report, in the coastal Northeast. Their results showed that the average MUCAPE was 1200 $J\,kg^{-1}$ for cellular and linear events, which accounted for more than 70% of severe events in their study period. MUCAPE provides the largest estimate of CAPE, implying the mixed-layer convective available potential energy (MLCAPE) prior to coastal Northeast severe linear/cellular events was similar to or below MLCAPE observed during nonsevere mesoscale convective systems over the central Great Plains (Cohen et al. 2007, their Fig. 8a). Additionally, Hurlbut and Cohen (2014) found most severe event days had less than 1000 $J\,kg^{-1}$ of MLCAPE present in observed Northeast proximity soundings.

Bulk shear magnitudes (i.e., magnitude of the wind difference over a layer) between severe events over the Northeast and the central United States are relatively consistent, but large variability exists in bulk shear magnitudes for individual events over the Northeast (Lombardo and Colle 2011; Hurlbut and Cohen 2014). Hurlbut and Cohen (2014) found the interquartile range (IQR) of 0–6-km bulk wind shear for severe events with more than 101 severe reports per event spans 8.5–18.5 $m\,s^{-1}$. Lombardo and Colle (2011) found similar interquartile ranges for 0–6-km bulk wind shear for Northeast coastal linear severe events. These ranges indicate a sizable portion of Northeast severe events occurred under weak bulk shear conditions less favorable for storm organization and longevity (Thompson et al. 2003; Weisman and Rotunno 2004).

Many Northeast severe events occur in more marginal severe weather environments characterized by low CAPE and/or weak vertical wind shear, warranting further region-specific study similar to the work of Hurlbut and Cohen (2014) and Lombardo and Colle (2011). To extend the work of these two studies, we seek to isolate severe weather events that have both a low predictive skill and are high impact, instead of solely studying events by report number (Hurlbut and Cohen 2014) or convective storm structure

(Lombardo and Colle 2011). These low predictive skill, high-impact events are important to understand, since these are the events forecasters struggle to predict. In particular, we seek to better understand the synoptic-scale conditions leading up to low predictive skill events. Such knowledge may serve to increase situational awareness that could allow forecasters to identify when low predictive skill events are more likely occur, thus, improving forecast performance. Therefore, we use SPC convective outlooks to identify low and high predictive skill severe weather events and compare the local and synoptic environmental conditions of each event class.

An a priori assumption is that low predictive skill severe weather events occur in low-CAPE, high-shear environments and high-CAPE, low-shear environments, when it is not clear if there are sufficient combinations of both ingredients for severe convection. Sherburn and Parker (2014) investigated severe weather events in low-CAPE, high-shear environments in the Southeast. They found that indices that used lapse rates and bulk shears in multiple layers had more skill than conventional severe weather indices. As a result, forecasting severe weather in these environments, and other low predictive skill environments, could benefit from casting a wider net to identify more useful parameters. Sherburn and Parker (2014) motivated an assessment of predictive skill based on low- and high-CAPE/shear environments in the Northeast and an evaluation of severe weather parameters that could be more useful than conventional severe weather indices in low predictive skill events.

The paper is organized as follows. A technique to identify high-impact, low-impact, and low predictive skill severe weather events, based on prior SPC convective outlook performance, and methods of environmental analysis used herein are presented in section 2. The results of this study, including a climatology, midlevel flow composites, statistical analysis of severe parameters, and a case study of a low predictive skill event, are detailed in section 3. Discussion and conclusions follow in section 4.

2. Data and methods

a. Evaluation method

Motivated by the methods employed by Hitchens and Brooks (2012), we evaluated SPC convective outlooks using severe storm reports from 1980 to 2013. This study focused on the early morning convective outlooks issued at 0600 UTC, which are valid for the 24-h period starting at 1200 UTC and ending at 1200 UTC the next day. The 0600 UTC convective outlooks were chosen over outlooks issued later in the day because the 0600 UTC outlook has been routinely issued since 1973 and allows for a larger sample size of events. While outlooks issued later in the day typically improve the skill of previous outlooks (Hitchens and Brooks 2014), we chose to investigate the "initial" predictive skill of an event before the start of the day: one verification period. Additionally, the analysis and conclusions that follow would not change appreciably using the 1300 UTC outlooks in lieu of the 0600 UTC outlooks; that is, a high-impact, low predictive skill event identified using the 0600 UTC outlook persists for the majority of the time in the 1300 UTC outlook (not shown). Only slight risk outlook areas were considered for verification due to their higher rate of occurrence (i.e., larger sample size) relative to higher-tier risk categories, especially in the Northeast, and the higher probability of severe, impactful weather communicated by the slight risk category relative to the lower-tier see text or marginal categories.[1]

TABLE 1. Contingency table for forecasts and observations.

	Observed yes	Observed no
Forecast yes	*a*	*b*
Forecast no	*c*	*d*

To evaluate the slight risk outlooks, a verification scheme similar to that of Hitchens and Brooks (2012, 2014) was developed. The 0600 UTC day-one slight risk areas over the Northeast (Fig. 1, black box) were overlaid on a 40-km grid. Individual storm reports were used to verify the slight risk outlooks (NCDC 2014). Severe weather is defined as a wind gust equal to or greater than 25.7 m s^{-1}, hail equal to or greater than 19.1 mm, or any tornado. As of 5 January 2010, the severe hail criterion changed to 25.4 mm and from that date on, the present study uses the 25.4-mm threshold. Reports were matched to the 24-h convective outlook period and gridded using a 40-km radius of influence centered on each storm report to match the intended scale of the SPC convective outlook products (i.e., "within 25 mi [~40 km] of a point"). Similar to Hitchens and Brooks (2012), grid points within 40 km of multiple storm reports do not have more influence than grid points with a single storm report within 40 km.

The verification scheme allowed spatial evaluation using a 2 × 2 contingency table (Table 1). Verification metrics calculated from this table include probability of detection (POD) and threat score (TS), defined as

$$\mathrm{POD} = \frac{a}{a+c} \quad \text{and} \tag{1}$$

$$\mathrm{TS} = \frac{a}{a+b+c}, \tag{2}$$

[1] Effective 22 October 2014, the "see text" risk category was discontinued and replaced with the marginal risk category.

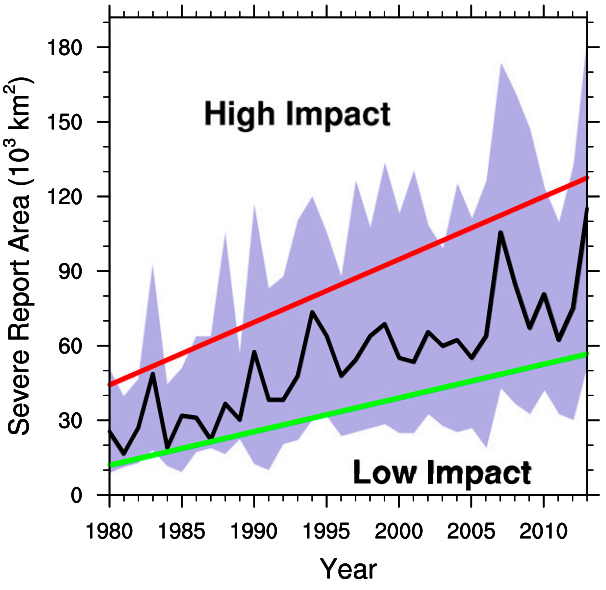


FIG. 2. The median (black) and IQR (shaded) of the severe report area on days with a slight risk in the Northeast. The red (green) line denotes linear regression on the 66th (33rd) percentile of the severe report area. Events with a larger (smaller) severe report area than the red (green) line in a given year are classified as high (low) impact.

where a is the number of grid points with hits, b is the number of grid points with false alarms, and c is the number of grid points with misses.

b. *Event classification*

The verification metrics [Eqs. (1) and (2)] can sometimes give unrepresentative results. For example, if there is only one severe report on a given day that falls 40 km outside a small slight risk area, this day would be assigned a POD and TS of zero. However, the forecast of this "low impact" event might not necessarily be poor from a subjective standpoint compared to a "high impact" event with many missed reports over a large area. Thus, we desire to subset our sample to only include high-impact events. To define a high-impact event, the increase in severe weather reports with time and report clustering around populated areas must be accounted for (Doswell et al. 2005; Hitchens and Brooks 2012; Allen et al. 2015). Increasing population, variations in population density, and the advent of digital communication are likely factors causing the increase in severe weather reports (Allen et al. 2015; Tippett et al. 2015). This increase, if not accounted for, would likely bias the sample of high-impact events toward recent years and those events that may have been confined to urban areas, if the events were chosen based on the number of severe weather reports. In lieu of the number of severe weather reports, we used an areal measure of severe weather by calculating the number of grid points within 40 km of a severe weather report. Grid points are dichotomous such that those grid points assigned multiple reports do not have more influence than grid points assigned a single report to partially mitigate the effect of report clustering in urban areas.

TABLE 2. The 25th, 50th, and 75th percentile POD and TS values for all high-impact events.

Percentile	25th	50th	75th
POD	0.14	0.63	0.88
TS	0.06	0.20	0.32

Figure 2 shows that there is an upward trend in severe report area with time, suggesting that a temporally static areal threshold for delineating high-impact events may still bias the sample of high-impact events toward recent years. Therefore, we performed a linear regression about the top tercile of the distribution of the severe report area to obtain a temporally varying lower areal threshold for classifying high-impact Northeast severe weather events (Fig. 2, red line). Thus, any Northeast severe event with more severe report coverage than the top tercile linear regression line in Fig. 2 was considered a high-impact event representative of a regional severe weather outbreak over the Northeast. Conversely, any Northeast severe event with less severe report coverage than the bottom tercile linear regression line (green line in Fig. 2) was considered a low-impact event.

Events that are poorly predicted can be grouped into two main classes: low-POD events and false alarm events. False alarm events, colloquially known as "forecast busts," have diminished through the 1980–2013 period (not shown); therefore, we restricted our examination of poorly predicted events to the low-POD class. Events with a low POD score, commonly called missed events, were defined as any event satisfying the high-impact criteria stated above and having a POD in the lowest quartile of all high-impact events (POD $\leq$ 0.14; Table 2). Missed events, as defined here, composed all of the low predictive skill events in this study. A representative example of a missed event is shown in Fig. 3.

A class of well-predicted events was defined to facilitate comparisons between high predictive skill events and low predictive skill events. Well-predicted events, hereafter called good events, were defined as any event satisfying the high-impact criteria and having a TS in the highest quartile of all high-impact events (TS $\geq$ 0.32; Table 2). A representative example of a good event is shown in Fig. 4.

c. *Environmental analysis*

The 0.5° Climate Forecast System Reanalysis (CFSR; Saha et al. 2010) was used to evaluate the local and synoptic environments associated with high-impact severe

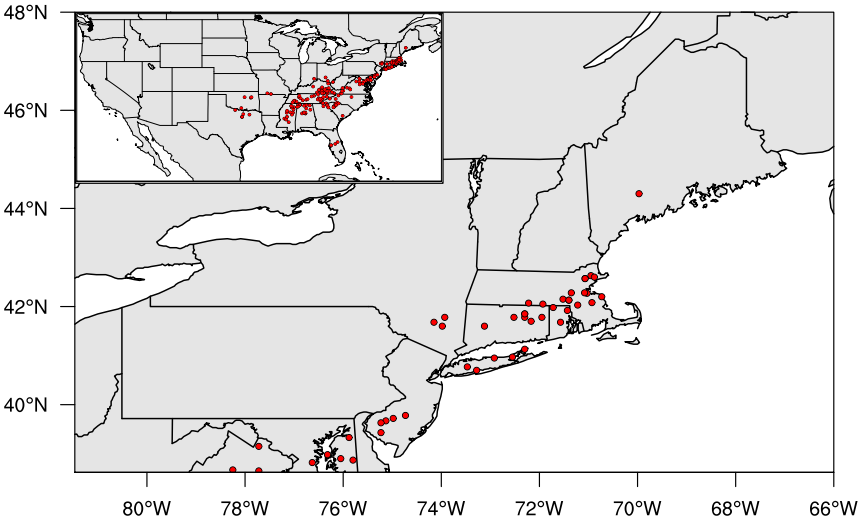


FIG. 3. Low POD or "missed" event example case from 5 Aug 2005. No 0600 UTC slight risk area was present in the Northeast or the rest of the continental United States. The red dots are missed storm reports outside a slight risk area.

weather over the Northeast. Atmospheric variables for high-impact events were taken from the average value across a 1° × 1° box centered on the point of maximum report density. We attempted to capture the preconvective environment within the CFSR for a majority of the cases and therefore used the 1800 UTC analysis for each event day as a representative time before the climatological diurnal peak in severe weather (Hurlbut and Cohen 2014).

The data-gathering process was repeated using the North American Regional Reanalysis (NARR) at 32-km grid spacing (Mesinger et al. 2006). Following Lombardo and Colle (2011), we chose the closest 3-h NARR analysis prior to the first severe report of the event day and used the average value of the atmospheric variables across a 0.6° × 0.6° box centered at the first severe report. We used two different reanalysis datasets and data selection methods in order to test the robustness of the results in the next section.

Event-centered composite maps from the CFSR were constructed to compare the spatial patterns of missed and good event classes at the synoptic scale. The composites were centered about the point of maximum report density and events were binned by 500-hPa flow direction above the same point. Event-centered composite maps

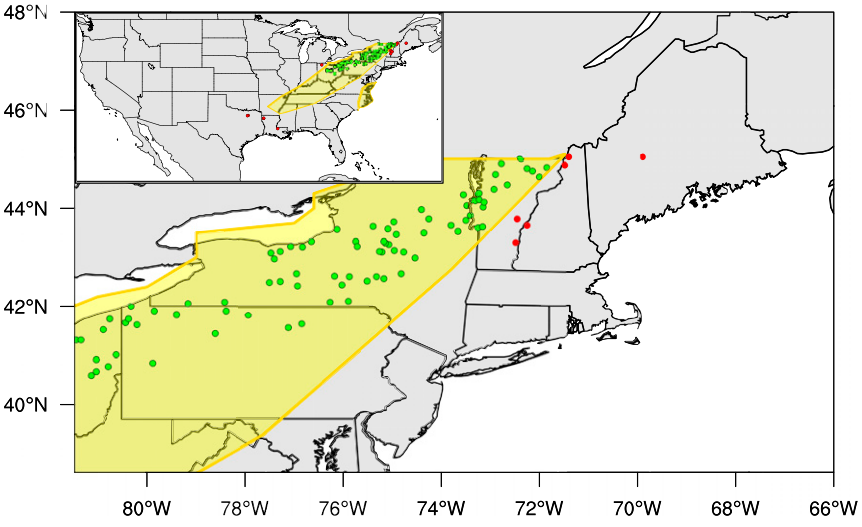


FIG. 4. As in Fig. 3, but for a "good" event on 31 Aug 1993. The yellow polygon is the 0600 UTC slight risk area valid from 1200 UTC 31 Aug to 1200 UTC 1 Sep 1993. Green dots are valid storm reports inside the slight risk area.

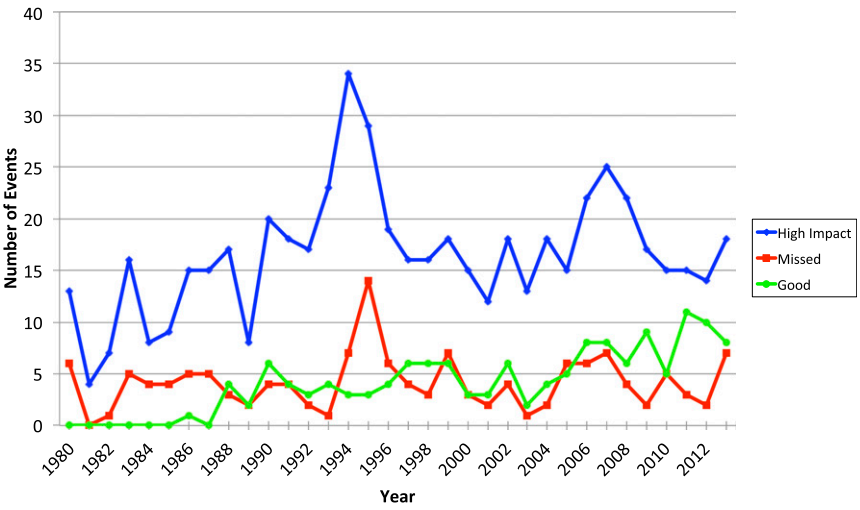


FIG. 5. Annual frequencies of high-impact (blue), missed (red), and good (green) events from 1980 to 2013 in the Northeast.

from the NARR were very similar and are not shown. Warm-season (April–September) events were chosen to focus results on environments that produce the majority of the high-impact events. Severe weather parameters, calculated from the CFSR and the NARR, were examined to help discriminate between the two event classes. In addition, high- and low-impact events under low-shear, high-CAPE conditions were compared to discriminate event intensity within the region of the CAPE–shear phase space that characterized the most missed events. Finally, a case study is shown to study the evolution of a typical missed event, using the NARR to better resolve subsynoptic-scale features at 3-h time resolution.

3. Results

a. *Annual and monthly frequency*

We first give an overview of the climatology of high-impact, missed, and good events in the Northeast. High-impact events composed 561 (21%) of the 2635 Northeast severe weather events that occurred over the study period with a maximum of 34 high-impact events in 1994 and a minimum of 4 high-impact events in 1981 (Fig. 5). Missed events composed 142 (25%) of all high-impact events with a maximum of 14 events in 1995 and a minimum of zero events in 1981. Missed events show no discernible trend through the study period. Good events composed 140 (25%) of all high-impact events and exhibited an

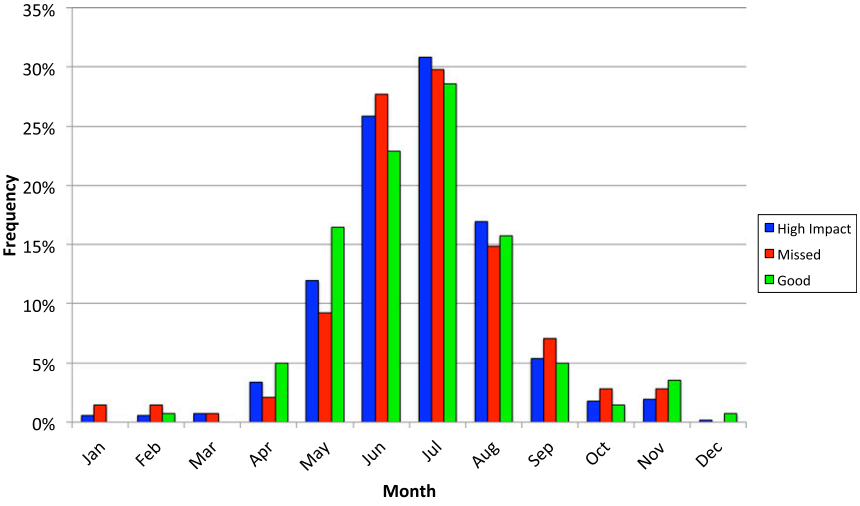


FIG. 6. Monthly frequencies of high-impact (blue), missed (red), and good (green) events from 1980 to 2013.

TABLE 3. Total severe, wind, and hail reports for missed, good, and all high-impact events. Percentages of wind, hail, and tornado reports out of the total severe report count for each event type are shown in parentheses.

Reports	Total severe	Wind	Hail	Tornado
High impact	39 094	29 129 (74.5%)	9939 (25.4%)	26 (0.0665%)
Missed	7906	5924 (74.9%)	1978 (25.0%)	4 (0.0506%)
Good	17 544	12 893 (73.5%)	4629 (26.4%)	22 (0.125%)

increasing trend with time, as 34 good events occurred before 1996 and 106 events occurred after 1996. The upward trend in the number of good events is consistent with the increase in TS after 1995 found by Hitchens and Brooks (2012) and further suggests forecasts of severe weather have improved over time. Figure 6 shows that missed and good events generally follow the monthly frequency of all high-impact events to within a few percentage points, suggesting missed and good events do not preferentially occur at different times of the year.

b. *Severe hazard type*

The distributions of severe hazard types within the high-impact, missed, and good event classes are compared in Table 3. Wind events compose approximately 75% of all three event classes while hail events compose the remaining 25%, suggesting little difference in the severe hazard type between the two event classes. However, Hurlbut and Cohen (2014) found 69% (30%) of total severe reports were wind (hail) reports and 1% were tornado reports for all severe event days. This suggests that high-impact events likely consist of a higher (lower) percentage of wind (hail) events relative to all severe events in the Northeast.

c. *Midlevel flow*

We created subsets of high-impact, missed, and good events by 500-hPa flow direction in order to determine if there was a dependence on flow direction. Most (84%) high-impact events occurred under westerly and southwesterly 500-hPa flow directions, while 11% occurred under northwesterly flow (Fig. 7). Since the missed and good event categories are mutually exclusive subsets derived using quartiles of the high-impact dataset, we can directly compare each category's frequency. The frequency of missed events was lower than the frequency of all high-impact events in westerly and southwesterly flow, whereas the frequency of missed events was higher than the frequency of all high-impact events in northerly, northwesterly, and other flow directions. These latter flow regimes composed fewer high-impact events. Using NARR data instead, missed events occurred at slightly higher (lower) rates in southwesterly (westerly) flow (not shown) but the results were otherwise similar to Fig. 7. The frequency of good events was higher than the frequency of all high-impact events in westerly flow, but good events occurred at equal or slightly lower rates than high-impact events under all other 500-hPa flow directions.

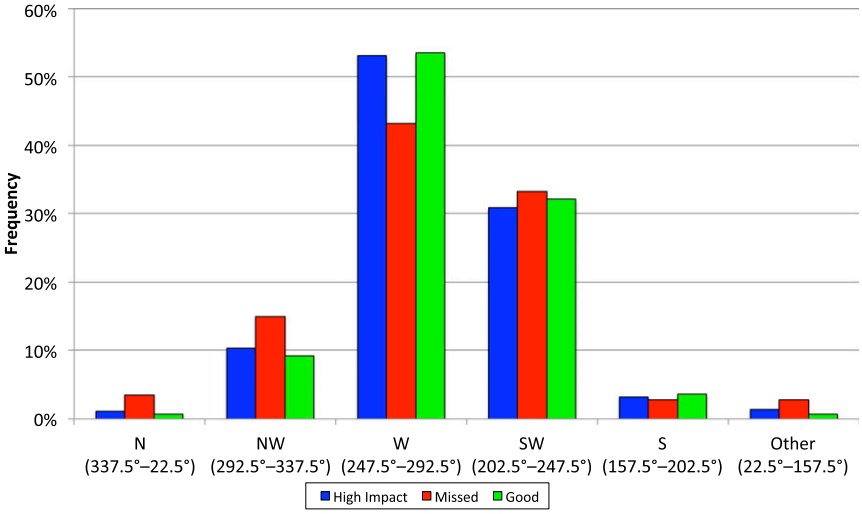


FIG. 7. Percent occurrence of high-impact (blue), missed (red), and good (green) events binned by 500-hPa flow direction from 1980 to 2013.

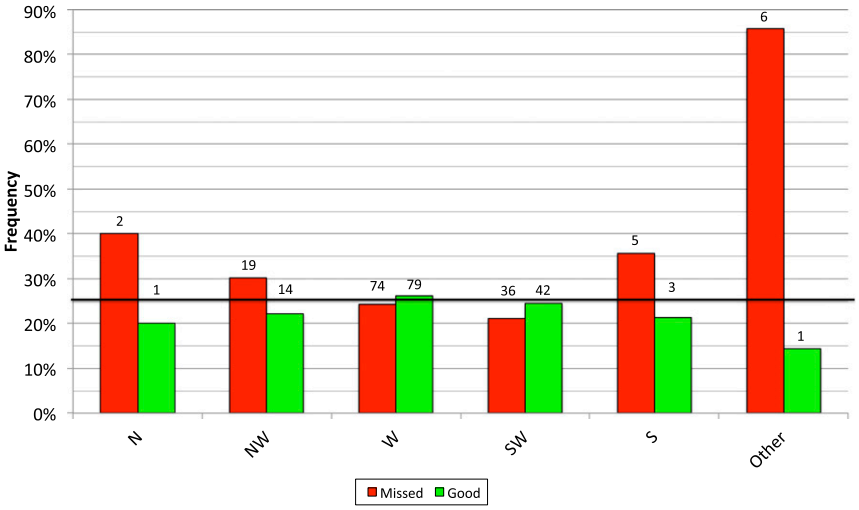


FIG. 8. Percentage of warm-season high-impact events classified as missed (red) and good (green) events binned by 500-hPa flow direction from 1980 to 2013. Sample sizes of missed and good events in each 500-hPa flow direction category are given above their respective bars. The black line indicates the expected climatological value (25%).

Figure 7 suggests that missed and good events do not occur at the same relative frequency as high-impact events for a given 500-hPa flow direction. To normalize the missed and good events with respect to the high-impact event climatology, the numbers of missed events and good events under a given flow direction were divided by the number of high-impact events under the same flow direction (Fig. 8). The quartile-based definitions of missed and good events, described in section 2, imply the 25% line in Fig. 8 represents the expected climatological percent occurrence if missed and good events do not differ from climatology. The normalized frequency of good events was approximately 25% in westerly and southwesterly flow, but occurred less often under all other flow directions. On the other hand, the normalized frequency of missed events was greater than 25% under all flow directions other than westerly and southwesterly flow. However, the small sample size of high-impact events associated with nonwesterly and nonsouthwesterly 500-hPa flow directions precludes any further conclusions relating the flow direction and predictive skill.

d. Composites

The normalized occurrence rate for missed events was higher for rarer flow regimes; despite northwesterly flow events composing 13% of all missed events, 30% of all northwesterly flow high-impact events were classified as missed events. Additionally, significant tornado events in the Northeast have occurred in northwesterly flow, often accompanied by steep midlevel lapse rates (Johns and Dorr 1996). However, most high-impact, missed, and good events occurred under westerly 500-hPa flow conditions. Therefore, event-centered composites of northwesterly and westerly missed and good events were created using 1800 UTC CFSR data to analyze the differences in the large-scale environment between the missed and good event classes.

Missed ($N = 19$) and good ($N = 14$) northwesterly flow events were characterized by a 500-hPa ridge to the southwest of the event centers and by CAPE in excess of $1000\,\mathrm{J\,kg^{-1}}$ around and upstream of the event centers (Figs. 9a–d). The compositing technique smooths shortwave troughs due to variability in trough location and intensity among events and should be considered when evaluating the 500-hPa flow field. A plume of enhanced 700–500-hPa lapse rates was advected around the periphery of a midlevel ridge and toward the event centers in both composites with $>6.5\,\mathrm{K\,km^{-1}}$ lapse rates directly below faster 500-hPa flow around $20–25\,\mathrm{m\,s^{-1}}$ in the good composite (Figs. 9b,d,f). Steeper lapse rates are indicative of a possible elevated mixed layer, which is associated with severe weather (e.g., Banacos and Ekster 2010; Johns and Dorr 1996). Deep-layer shear was significantly stronger in the good composite (using Welsh's t test[2]), especially in areas to the north and west of the event center (Figs. 9a,c,e).

The westerly missed event ($N = 70$) composite depicted CAPE between 1000 and $1250\,\mathrm{J\,kg^{-1}}$ ahead of a diffuse trough embedded in about $15\,\mathrm{m\,s^{-1}}$ of westerly flow to the northwest of the event center (Figs. 10a,b).

[2] Welsh's t test is a Student's t test that assumes two samples have different population variances.

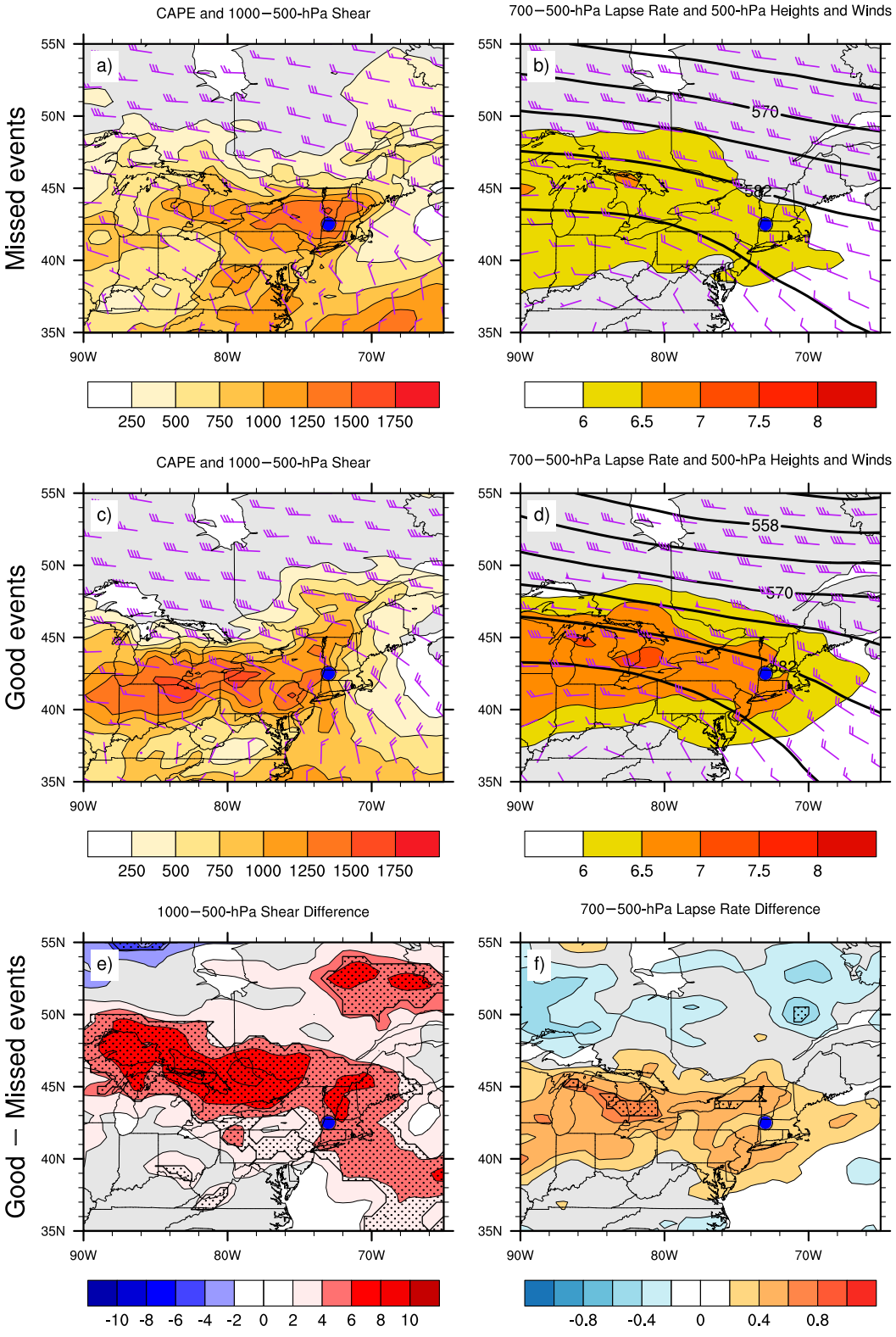


FIG. 9. Composite maps for warm-season missed and good events under northwesterly 500-hPa flow centered on the point of maximum report density (blue dot). CAPE (filled; J kg^{-1}) and 1000–500-hPa wind shear (barbs; m s^{-1}) for (a) missed events ($N = 19$) and (c) good events ($N = 14$). The 500-hPa geopotential height (black contours; dam), 500-hPa wind (m s^{-1}), and 700–500-hPa lapse rate (filled; K km^{-1}) for (b) missed and (d) good events. Difference plots of good event composite mean values minus missed composite mean values of (e) 1000–500-hPa shear (filled; m s^{-1}) and (f) 700–500-hPa lapse rate (filled; K km^{-1}). Statistically significant areas are stippled at the 95% confidence level. Background map is for scale only.

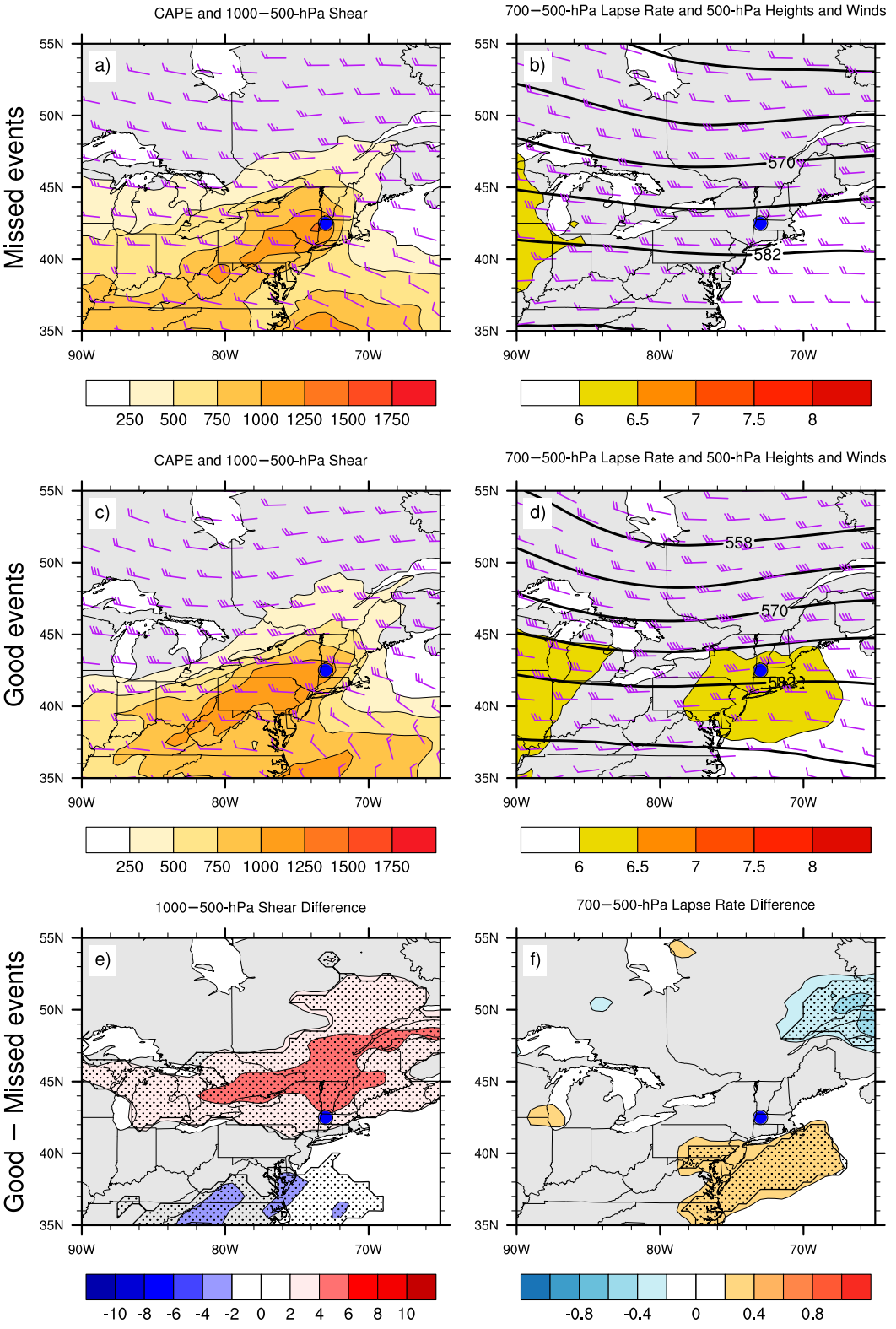


FIG. 10. As in Fig. 9, but for missed (N = 70) and good (N = 79) events under westerly 500-hPa flow.

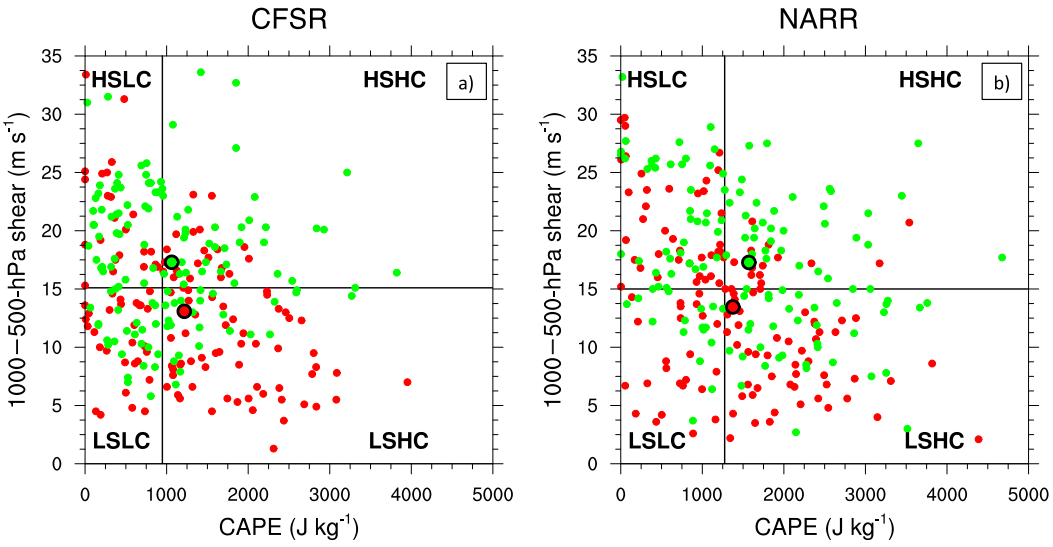


FIG. 11. Phase-space diagrams of CAPE ($J\,kg^{-1}$) and 1000–500-hPa shear magnitude ($m\,s^{-1}$) for warm-season missed (red) and good (green) events from 1980 to 2013 using the (a) CFSR and (b) NARR data. High-impact event dataset medians of CAPE and shear are given as black lines. From the top left moving clockwise, quadrants are labeled as follows: high shear, low CAPE (HSLC); high shear, high CAPE (HSHC); low shear, high CAPE (LSHC); and low shear, low CAPE (LSLC). The boldface circles denote the median CAPE and shear values for missed (red) and good (green) events, respectively.

The good composite ($N = 79$) suggests a similar trough embedded in faster, 20–25 $m\,s^{-1}$ flow to the northwest of the event center (Fig. 10d). CAPE values around the event center were similar between the missed and good composites (Figs. 10a,c). The faster midlevel flow enhanced the deep-layer shear around and to the north of the good event composite center, and steeper lapse rates existed to the south (Figs. 10e,f). Cases under southwesterly flow exhibited similar characteristics (not shown). The remaining flow directions have too few cases to perform meaningful composites.

e. *Severe parameters*

We now examine the relationship between CAPE and 1000–500-hPa shear and the frequencies of missed and good events across all flow directions. The CAPE and shear were calculated using the CFSR-based methodology from section 2c for each event. The median CAPE (948 $J\,kg^{-1}$) and shear (15.1 $m\,s^{-1}$) of all warm-season, high-impact events were used to separate the CAPE–shear phase space into four quadrants (Fig. 11a). Good events were most common under high-shear, low-CAPE conditions (43 out of 131), whereas missed events were most common under low-shear, high-CAPE conditions (54 out of 128; Table 4). The average threat scores in Table 4, and the median CAPE and shear values for missed and good events in Fig. 11, indicate that shear appears to be a better discriminator between missed and good events than CAPE.

We repeated the analysis using the NARR-based methodology from section 2c and found similar results. Table 4 shows a plurality of good events occurred under high-shear, low-CAPE conditions (40 out of 131) and a plurality of missed events occurred under low-shear, high-CAPE (51 out of 128). High-shear, low-CAPE environments have been associated with the low predictive skill of significant severe weather in the southeast United States (Sherburn and Parker 2014). While the majority of missed events in the Northeast occurred in low-shear environments, a sizable portion of missed events occurred in high-shear, low-CAPE environments, especially within the NARR (Table 4). The NARR indicated a higher median CAPE value (1274 $J\,kg^{-1}$) than the CFSR, possibly due to the higher resolution of the NARR, but the 1000–500-hPa median shear value (15.0 $m\,s^{-1}$) was similar to that of the CFSR.

TABLE 4. The number of missed events, the number of good events, and the TSs for all high-impact events in each CAPE–shear phase-space quadrant shown in Fig. 11 for the CFSR data. Values in parentheses are from the NARR data.

Event classification	Missed	Good	TS
High shear, high CAPE	22 (13)	38 (36)	0.254 (0.266)
High shear, low CAPE	23 (40)	43 (40)	0.243 (0.208)
Low shear, high CAPE	54 (51)	25 (38)	0.169 (0.199)
Low shear, low CAPE	29 (24)	25 (17)	0.190 (0.194)

TABLE 5. Warm-season, high-impact means of various parameters from the CFSR data. Values in parentheses are from the NARR data. Boldface means indicate a statistically significant difference between the good and missed categories at the 95% confidence level.

Parameter	High-impact mean	Missed mean	Good mean
1000–500-hPa shear ($m\,s^{-1}$)	15.4 (15.4)	**13.1** (**13.5**)	**17.3** (**17.3**)
1000–850-hPa shear ($m\,s^{-1}$)	6.9 (12.9)	**5.6** (**10.9**)	**7.7** (**15.5**)
700–500-hPa lapse rate ($°C\,km^{-1}$)	5.9 (5.9)	5.9 (**5.9**)	6.1 (**6.0**)

It is important to emphasize that our criteria for identifying low- and high-CAPE/shear events are determined by the median of the distributions of warm-season, high-impact events and not all severe weather events in the Northeast. One could set fixed thresholds [e.g., a low-CAPE threshold at $500\,J\,kg^{-1}$, as in Sherburn and Parker (2014)] or use different percentiles instead of the median. The median is chosen here to represent the statistics of the sample of events herein, to reflect differences in reanalyses, and to retain all missed and good events in the analysis.

Statistical significance testing, using Welsh's *t* test, was performed on an array of severe weather parameters for warm-season high-impact events to determine which, if any, parameters were significant discriminators of missed and good events in the Northeast. Very few parameters showed significance, likely as a result of the phenomenological similarity of all high-impact events; therefore, only select parameters are presented. There was a notable signal for stronger shear to be present during good events (Figs. 9e, 10e, and 11) compared with missed events. Missed events were associated with significantly lower mean shear values than good events for the 1000–500- and 1000–850-hPa layers from both CFSR and NARR (Table 5). Additionally, 700–500-hPa lapse rates were slightly lower for missed events compared with good events, but the difference is only significant in the NARR (Table 5). There is a need to better understand how these missed events occur in lower-shear conditions.

f. Low shear, high CAPE

Low-shear, high-CAPE environments were associated with a plurality of missed events (Fig. 11 and Table 4). Threat scores were relatively low under low-shear, high-CAPE conditions, especially with the CFSR (Table 4). Therefore, it is worthwhile to investigate events occurring under low-shear, high-CAPE conditions to determine whether some additional environmental parameters might be useful in discriminating the severity of events within that quadrant of the CAPE–shear phase space. To this end, we create a category of low-impact severe events, as discussed previously in section 2b (Fig. 2).

Low-impact events occurring in low-shear, high-CAPE environments are compared with high-impact events in the low-shear, high-CAPE quadrants of Figs. 11a and 11b. High-impact events under low-shear, high-CAPE conditions had significantly higher downdraft CAPE (DCAPE) and higher shear relative to low-impact events (Figs. 12a,c,d). Additionally, the average relative humidity in the lowest 150-hPa layer was significantly lower for high-impact events under low-shear, high-CAPE conditions (Fig. 12b). High DCAPE, a measure of potential energy for descending saturated parcels, indicates the potential for strong downdrafts and possible severe wind gusts at the surface. Meanwhile, low 150-hPa layer RH indicates the potential for enhanced evaporative cooling and generation of negative buoyancy, thus enhancing the risk of microbursts and wind damage. These findings corroborate the higher percentage of wind reports found in high-impact events, relative to all severe events, shown in section 3b. Differences were found to be significant (at the 99% level) in both the CFSR and NARR. However, the NARR had smaller DCAPE and RH differences between high-impact and low-impact events. It is difficult to assess which reanalysis is more "correct," but nonetheless, these results suggest that DCAPE, low-level RH, and shear magnitude may help discriminate between high-impact and low-impact events under low-shear, high-CAPE conditions.

g. 18 August 2009 case study

A brief case study is presented of the 18–19 August 2009 severe weather event to illustrate a missed event occurring under low-shear, high-CAPE conditions. CAPE values in excess of $1000\,J\,kg^{-1}$ overspread much of the Pennsylvania and the mid-Atlantic coastal region, while 1000–500-hPa shear was no higher than $10\,m\,s^{-1}$ across the same region at 1500 UTC 18 August (Fig. 13a). Despite the low shear values, severe wind reports occurred in New York, and a large cluster of severe reports, including five severe hail reports and one significant $39.6\,m\,s^{-1}$ wind report, occurred in portions of Pennsylvania, Maryland, Delaware, New Jersey, and the New York City area (Fig. 13b). Storm initiation occurred around 1500–1600 UTC along the southeast coast of Lake Erie and over the higher terrain of western Pennsylvania and West Virginia. After convective initiation, storms consolidated into small, broken-line segments with some bowing elements and progressed toward the eastern seaboard, producing severe winds

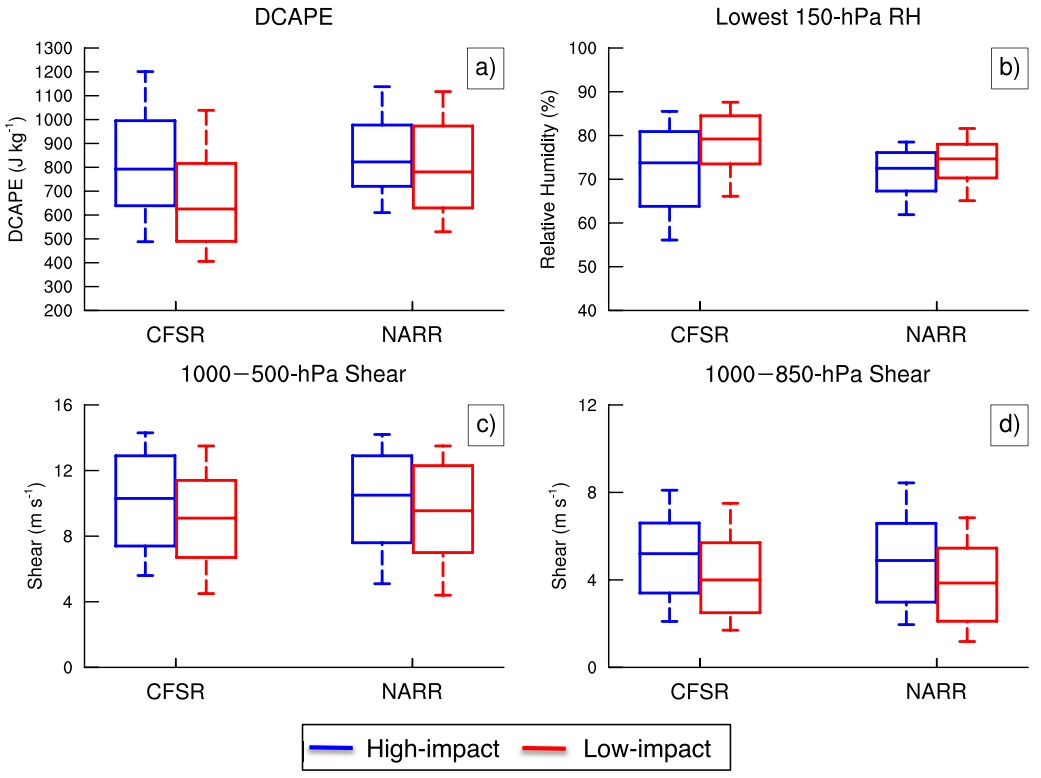


FIG. 12. Box-and-whisker plots for high-impact (blue) and low-impact (red) low-shear, high-CAPE events. The ends of the boxes represent the 25th and 75th percentiles of the distribution, the ends of the whiskers represent the 10th and 90th percentiles of the distribution, and the horizontal line within the boxes represents the median of the distribution. Distributions from the CFSR and NARR datasets are shown for (a) DCAPE (J kg^{-1}), (b) average RH within the lowest 150-hPa layer (%), and (c) 1000–500-hPa and (d) 1000–850-hPa shear (m s^{-1}).

(Figs. 14c,d and 15c,d). The event caused an estimated $1.3 million in damage (NCEI 2017).

Throughout the event, 1000–500-hPa shear remained around or below 10 m s^{-1} across the region of the severe weather reports (Figs. 14a,b and 15a,b). The 1200 UTC Pittsburgh, Pennsylvania (KPIT), sounding sampled the western edge of the stronger shear west of this region, observing 15 m s^{-1} winds within the lower troposphere

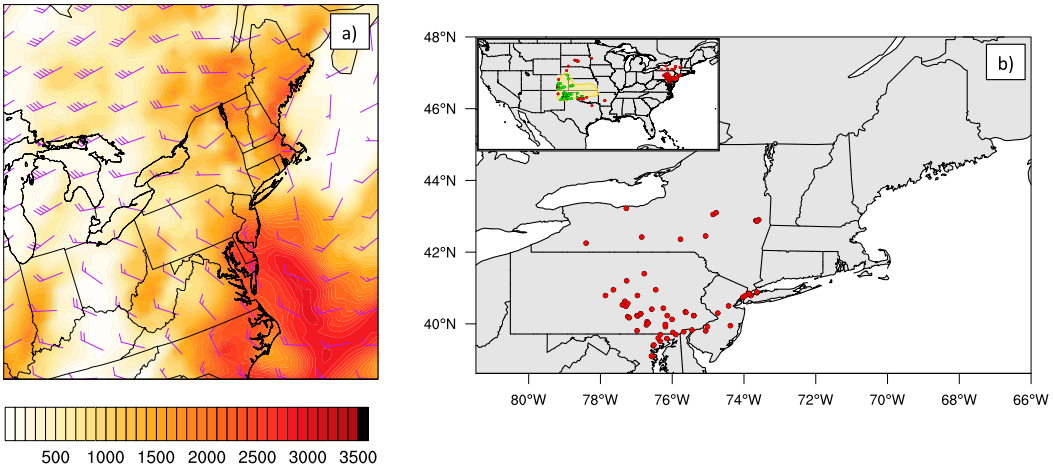


FIG. 13. (a) NARR surface-based CAPE (filled; J kg^{-1}) and 1000–500-hPa shear (barbs; m s^{-1}) valid 1500 UTC 18 Aug 2009. (b) As in Fig. 4, but for 1200 UTC 18 Aug 2009.

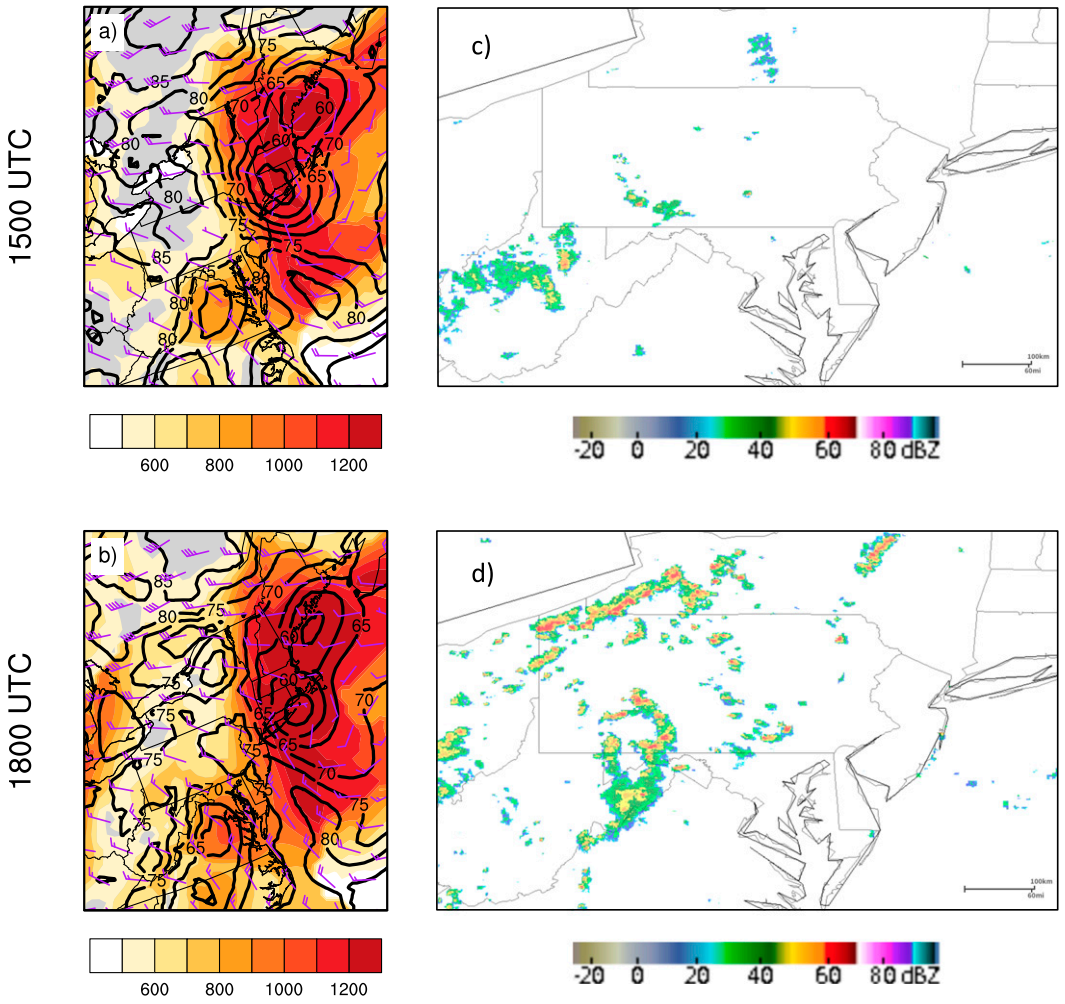


FIG. 14. NARR DCAPE (filled; J kg^{-1}), lowest 150-hPa mean RH (black contours; every 5%), 1000–500-hPa shear (barbs; m s^{-1}), and mosaic base reflectivity at (a),(c) 1500 UTC and (b),(d) 1800 UTC 18 Aug 2009.

(not shown), whereas sounding sites to the east exhibited weaker winds throughout the lower troposphere. The lack of high winds in the troposphere likely limited the downward flux of horizontal momentum by convection. Therefore, strong downdraft velocities generated in situ within the convective elements were necessary to produce damaging winds at the surface. Large DCAPE values between 600 and 1300 J kg^{-1} existed over much of the Northeast at 1500 UTC with a relative minimum in DCAPE in northwest Pennsylvania and within the higher terrain of West Virginia near the locations of convective initiation (Figs. 14a,c). Additionally, lowest 150-hPa layer mean RH values of 80%–85% indicate the low-level environments downstream of Lake Erie and within the higher terrain of West Virginia were closer to saturation compared to locations farther east. The storms progressed eastward toward higher DCAPE and lower low-level RH (Figs. 14b and 15d). The 150-hPa layer mean RH decreased over eastern Pennsylvania, likely as a result of heating of the boundary layer. Storm reports occurred as convection moved into areas with larger CAPE, DCAPE, and reduced low-level RH.

Additionally, the 0000 UTC Upton, New York (KOKX), sounding (Fig. 16) reported a DCAPE of 1199 J kg^{-1}, well above the 90th percentile for 0000 UTC 19 August (Rogers et al. 2015), prior to a 39.6 m s^{-1} wind report occurring at 0233 UTC 19 August nearby at Glen Cove, New York. This anomalously high DCAPE likely contributed to enhanced downdraft speeds associated with the severe wind reports in the region. Additional case studies of events in low-shear, high-CAPE conditions, such as the 4 July 2012 and the 24 June 2013 events, indicate similar themes of convection moving into areas of anomalously high CAPE, high DCAPE, and lower low-level RH (not shown). While this may not be the only scenario that can produce severe weather, particularly

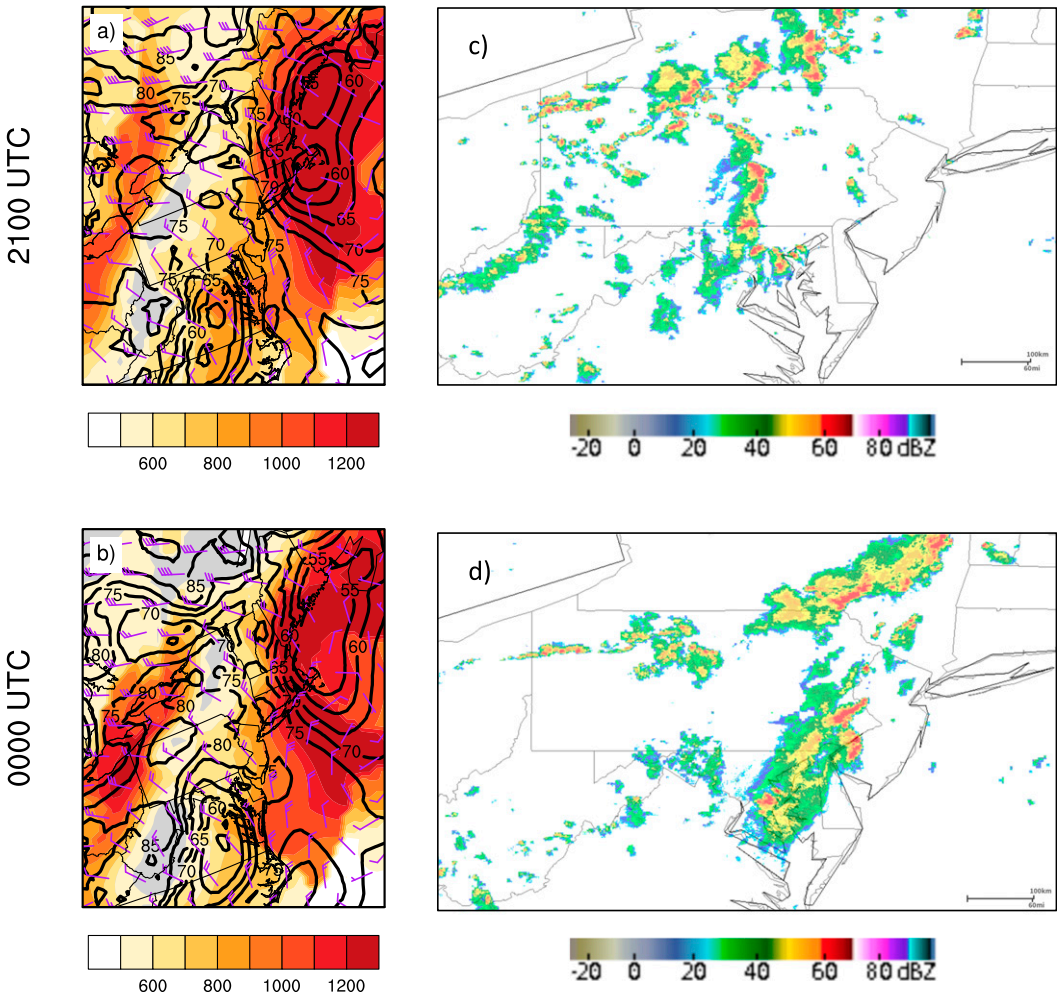


FIG. 15. As in Fig. 14, but at (a),(c) 2100 UTC 18 Aug and (b),(d) 0000 UTC 19 Aug 2009.

severe winds, it is a scenario that forecasters should be cognizant of in order to reduce missed events under low-shear, high-CAPE conditions.

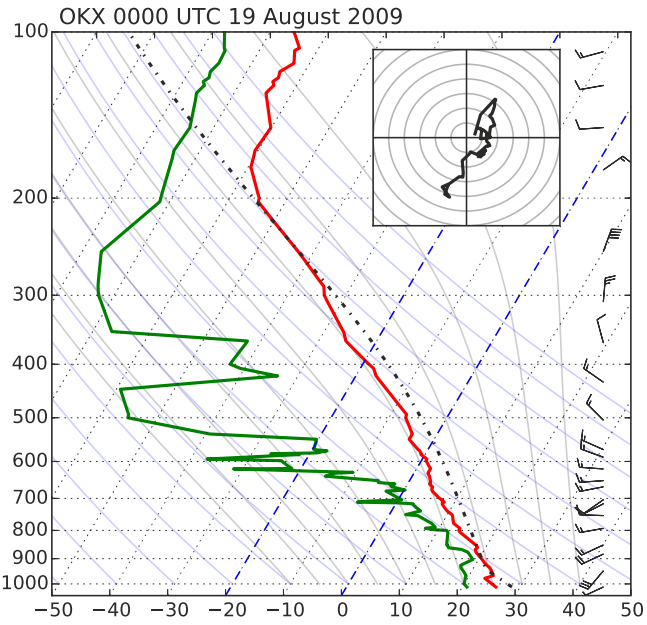
4. Conclusions

This study identified Northeast high-impact severe weather events with high spatial coverage, while controlling for the increase in severe weather reports with time. The technique was applied from 1980 to 2013 and identified 561 high-impact events representing the top 21% of severe weather events in the Northeast. Additionally, 142 high-impact events were categorized as missed events. The number of good events has increased over the period, consistent with the increasing threat scores of slight risk outlooks found by Hitchens and Brooks (2012).

The majority of missed and good events occurred during June–August and resembled the climatology of severe events in the Northeast (Hurlbut and Cohen 2014). However, there was no particular month where missed or good events occurred substantially more often than the high-impact seasonal climatology, suggesting that predictability has little or no seasonal dependence.

The frequency of missed events, compared with all high-impact events, is slightly lower in westerly and southwesterly 500-hPa flow, and higher for all other flow directions. The varying forecast skill of these events may be due to a lack of experience in recognizing the potential for severe weather events under relatively uncommon flow directions (differing from westerly and southwesterly) and the associated lack of situational awareness under flow regimes outside of the bulk of the climatological envelope (Tang et al. 2016).

Event-centered composites constructed using CFSR data allowed for a comparison of warm-season missed and good events under northwesterly and westerly 500-hPa flow. Good events had higher shear and steeper midlevel lapse rates, suggesting missed events occurred



SBCAPE (J kg^{-1})	816
MLCAPE (J kg^{-1})	1945
DCAPE (J kg^{-1})	1199
0–6-km Shear (m s^{-1})	8.23

FIG. 16. The 0000 UTC 19 Aug 2009 skew T–log$p$ sounding from KOKX. Hodograph rings are shown every 5 m s^{-1}.

under more marginal synoptic conditions conducive to severe weather. Improving situational awareness for severe weather when the shear is suboptimal for storm organization (Thompson et al. 2003; Weisman and Rotunno 2004) may lead to a reduction in the number of missed events.

Missed events occurred most frequently under low-shear, high-CAPE conditions. However, many missed events also occurred under high-shear, low-CAPE conditions, suggesting high-shear, low-CAPE environments, which present severe weather forecast challenges in the southeast United States (Sherburn and Parker 2014), also present severe weather forecast challenges in the Northeast. Rapid destabilization, occurring over periods $\leq$ 3 h, has been shown to occur prior to high-shear, low-CAPE events (King et al. 2017) and may affect predictability. However, such rapid environmental changes are not adequately resolved by the reanalyses used herein, thus assessing the influence of temporal environmental variability on forecast performance is beyond the scope of this study. Examining individual events under low-shear, high-CAPE conditions revealed high-impact events have significantly higher DCAPE, vertical wind shear, and significantly lower low-level RH compared with low-impact events under low-shear, high-CAPE conditions. The potential for enhanced evaporative cooling, indicated by the higher DCAPE and lower low-level RH, may lead to stronger downdrafts, increasing the probability of damaging winds at the surface. Additionally, the higher shear magnitudes (8–13 m s^{-1} across the 1000–500-hPa layer) for high-impact events occurring within the low-shear, high-CAPE quadrant of the phase space might suggest the convective environment may be more sensitive to localized enhancements of shear, such as channeled flow, that can contribute to storm organization and severity in local mesoscale pockets (Tang et al. 2016).

The 18 August 2009 case study presented herein depicts a missed event that occurred under low-shear, high-CAPE conditions. The storms developed in a more moist low-level environment characterized by high boundary layer mean RH around or above 75% and within 1000–500-hPa shear around or less than 10 m s^{-1}. The storms produced a cluster of severe reports upon propagating eastward into areas of higher CAPE, higher DCAPE, and lower boundary layer RH—an environment conducive to the production of microbursts, even in the absence of strong flow aloft.

The significance of lower-tropospheric shear magnitude in discriminating between missed and good events suggests future work should investigate the physical processes regarding the production of severe weather under low-shear conditions. It is possible inhomogeneities in vertical wind shear magnitude due to terrain modifications of mesoscale flows, which are not resolved by the reanalysis data used in this study, are important to locally increasing the risk of severe weather. Additionally, an evaluation of convection-allowing models and ensemble forecasts of severe weather events under low-shear conditions would be useful for assessing the predictability of these events and biases in the models that may not allow the model to predict these events with fidelity. Increased knowledge of a range of

scenarios under low-shear conditions that are capable of yielding high-impact events is necessary to improve forecast skill of severe weather events in the Northeast.

Acknowledgments. The authors thank the Storm Prediction Center's Andy Dean for providing convective outlook data and Israel Jirak and Steven Weiss for helpful comments in the early stages of this study. Additionally, the authors thank three anonymous reviewers for their constructive comments. This work was derived from a portion of the first author's M.S. thesis at the University at Albany, SUNY. The work research was funded by NOAA Collaborative Science, Technology, and Applied Research Program Grants NA13NWS4680004 and NA16NWS4680005 and by the Department of Defense (DoD) through the National Defense Science and Engineering Graduate Fellowship (NDSEG) Program.

REFERENCES

Allen, J. T., M. K. Tippett, and A. H. Sobel, 2015: An empirical model relating U.S. monthly hail occurrence to large-scale meteorological environment. *J. Adv. Model. Earth Syst.*, **7**, 226–243, doi:10.1002/2014MS000397.

Banacos, P. C., and M. L. Ekster, 2010: The association of the elevated mixed layer with significant severe weather events in the northeastern United States. *Wea. Forecasting*, **25**, 1082–1102, doi:10.1175/2010WAF2222363.1.

Bosart, L. F., A. Seimon, K. D. LaPenta, and M. J. Dickinson, 2006: Supercell tornadogenesis over complex terrain: The great Barrington, Massachusetts, tornado on 29 May 1995. *Wea. Forecasting*, **21**, 897–922, doi:10.1175/WAF957.1.

Brooks, H. E., J. W. Lee, and J. P. Craven, 2003: The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data. *Atmos. Res.*, **67–68**, 73–94, doi:10.1016/S0169-8095(03)00045-0.

Cohen, A. E., M. C. Coniglio, S. F. Corfidi, and S. J. Corfidi, 2007: Discrimination of mesoscale convective system environments using sounding observations. *Wea. Forecasting*, **22**, 1045–1062, doi:10.1175/WAF1040.1.

Doswell, C. A., III, H. E. Brooks, and M. P. Kay, 2005: Climatological estimates of daily local nontornadic severe thunderstorm probability for the United States. *Wea. Forecasting*, **20**, 577–595, doi:10.1175/WAF866.1.

Gensini, V., and W. Ashley, 2011: Climatology of potentially severe convective environments from the North American Regional Reanalysis. *Electron. J. Severe Storms Meteor.*, **6** (8), http://www.ejssm.org/ojs/index.php/ejssm/article/view/85/68.

Hitchens, N. M., and H. E. Brooks, 2012: Evaluation of the Storm Prediction Center's day 1 convective outlooks. *Wea. Forecasting*, **27**, 1580–1585, doi:10.1175/WAF-D-12-00061.1.

——, and ——, 2014: Evaluation of the Storm Prediction Center's convective outlooks from day 3 through day 1. *Wea. Forecasting*, **29**, 1134–1142, doi:10.1175/WAF-D-13-00132.1.

Hurlbut, M. M., and A. E. Cohen, 2014: Environments of northeast U.S. severe thunderstorm events from 1999 to 2009. *Wea. Forecasting*, **29**, 3–22, doi:10.1175/WAF-D-12-00042.1.

Johns, R. H., and R. A. Dorr Jr., 1996: Some meteorological aspects of strong and violent tornado episodes in New England and eastern New York. *Natl. Wea. Dig.*, **20** (4), 2–12, http://nwafiles.nwas.org/digest/papers/1996/Vol20No4/Pg2-Johns.pdf.

King, J. R., M. D. Parker, K. D. Sherburn, and G. M. Lackmann, 2017: Rapid evolution of cool season, low-CAPE severe thunderstorm environments. *Wea. Forecasting*, **32**, 763–779, doi:10.1175/WAF-D-16-0141.1.

LaPenta, K. D., L. F. Bosart, T. G. Jr, and M. J. Dickinson, 2005: A multiscale examination of the 31 May 1998 Mechanicville, New York, tornado. *Wea. Forecasting*, **20**, 494–516, doi:10.1175/WAF875.1.

Lericos, T., H. E. Fuelberg, M. L. Weisman, and A. I. Watson, 2007: Numerical simulations of the effects of the coastlines on the evolution of strong, long-lived squall lines. *Mon. Wea. Rev.*, **135**, 1710–1731, doi:10.1175/MWR3381.1.

Lombardo, K. A., and B. A. Colle, 2010: The spatial and temporal distribution of organized convective structures over the northeast and their ambient conditions. *Mon. Wea. Rev.*, **138**, 4456–4474, doi:10.1175/2010MWR3463.1.

——, and ——, 2011: Convective storm structures and ambient conditions associated with severe weather over the northeast United States. *Wea. Forecasting*, **26**, 940–956, doi:10.1175/WAF-D-11-00002.1.

Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360, doi:10.1175/BAMS-87-3-343.

NCDC, 2014: *Storm Events* database. NOAA/National Climatic Data Center, http://www.spc.noaa.gov/wcm/data.

NCEI, 2017: Storm Events database. NOAA/National Centers for Environmental Information, https://www.ncdc.noaa.gov/stormevents/.

Riley, G. T., and L. F. Bosart, 1987: The Windsor Locks, Connecticut tornado of 3 October 1979: An analysis of an intermittent severe weather event. *Mon. Wea. Rev.*, **115**, 1655–1677, doi:10.1175/1520-0493(1987)115<1655:TWLCTO>2.0.CO;2.

Rogers, J., P. Marsh, and R. Thompson, 2015: Sounding climatology page. NOAA/Storm Prediction Center, http://www.spc.noaa.gov/exper/soundingclimo/.

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, doi:10.1175/2010BAMS3001.1.

Sherburn, K. D., and M. D. Parker, 2014: Climatology and ingredients of significant severe convection in high-shear, low-cape environments. *Wea. Forecasting*, **29**, 854–877, doi:10.1175/WAF-D-13-00041.1.

Tang, B., M. Vaughan, R. Lazear, K. Corbosiero, L. Bosart, T. Wasula, I. Lee, and K. Lipton, 2016: Topographic and boundary influences on the 22 May 2014 Duanesburg, New York, tornadic supercell. *Wea. Forecasting*, **31**, 107–127, doi:10.1175/WAF-D-15-0101.1.

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, doi:10.1175/1520-0434(2003)018<1243:CPSWSE>2.0.CO;2.

Tippett, M. K., J. T. Allen, V. A. Gensini, and H. E. Brooks, 2015: Climate and hazardous convective weather. *Current Climate Change Rep.*, **1** (2), 60–73, doi:10.1007/s40641-015-0006-6.

Wasula, A. C., L. F. Bosart, and K. D. LaPenta, 2002: The influence of terrain on the severe weather distribution across interior eastern New York and western New England. *Wea. Forecasting*, **17**, 1277–1289, doi:10.1175/1520-0434(2002)017<1277:TIOTOT>2.0.CO;2.

Weisman, M., and R. Rotunno, 2004: A theory for strong long-lived squall lines revisited. *J. Atmos. Sci.*, **61**, 361–382, doi:10.1175/1520-0469(2004)061<0361:ATFSLS>2.0.CO;2.