

Evaluation of WRF Model Simulations of Tornadoic and Nontornadoic Outbreaks Occurring in the Spring and Fall

CHAD M. SHAFER, ANDREW E. MERCER,* LANCE M. LESLIE, AND MICHAEL B. RICHMAN

Cooperative Institute for Mesoscale Meteorological Studies, and School of Meteorology, University of Oklahoma, Norman, Oklahoma

CHARLES A. DOSWELL III

Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma

(Manuscript received 2 November 2009, in final form 7 May 2010)

ABSTRACT

Recent studies, investigating the ability to use the Weather Research and Forecasting (WRF) model to distinguish tornado outbreaks from primarily nontornadoic outbreaks when initialized with synoptic-scale data, have suggested that accurate discrimination of outbreak type is possible up to three days in advance of the outbreaks. However, these studies have focused on the most meteorologically significant events without regard to the season in which the outbreaks occurred. Because tornado outbreaks usually occur during the spring and fall seasons, whereas the primarily nontornadoic outbreaks develop predominantly during the summer, the results of these studies may have been influenced by climatological conditions (e.g., reduced shear, in the mean, in the summer months), in addition to synoptic-scale processes.

This study focuses on the impacts of choosing outbreaks of severe weather during the same time of year. Specifically, primarily nontornadoic outbreaks that occurred during the summer have been replaced with outbreaks that do not occur in the summer. Subjective and objective analyses of the outbreak simulations indicate that the WRF's capability of distinguishing outbreak type correctly is reduced when the seasonal constraints are included. However, accuracy scores exceeding 0.7 and skill scores exceeding 0.5 using 1-day simulation fields of individual meteorological parameters, show that precursor synoptic-scale processes play an important role in the occurrence or absence of tornadoes in severe weather outbreaks. Low-level storm-relative helicity parameters and synoptic parameters, such as geopotential heights and mean sea level pressure, appear to be most helpful in distinguishing outbreak type, whereas thermodynamic instability parameters are noticeably both less accurate and less skillful.

1. Introduction

In recent studies by Shafer et al. (2009, hereafter S09) and Mercer et al. (2009, hereafter M09), subjective and objective analysis of 50 tornado outbreaks (TOs) and 49 primarily nontornadoic outbreaks (PNOs) indicated that the Weather Research and Forecasting (WRF) model can be used to distinguish between the two types of outbreaks up to three days in advance using only

synoptic-scale data as input. These studies demonstrate that synoptic-scale processes play a substantial role in the occurrence or absence of tornado outbreaks, confirming the conclusions of, for example, Johns and Doswell (1992, hereafter JD92), Doswell et al. (1993), Stensrud et al. (1997, hereafter SCB97), and Doswell and Bosart (2001). Furthermore, some severe weather parameters associated with mesocyclones and tornadoes, and referred to as meteorological covariates (Brown and Murphy 1996; Brooks et al. 2003b), were found to be particularly effective in the discrimination of these outbreaks. These parameters include: storm-relative environmental helicity (SREH; Davies-Jones et al. 1990), low-level and deep-layer bulk shear (Bunkers 2002), the energy helicity index (EHI; Hart and Korotky 1991; Davies 1993), low-level storm-relative flow (SRFL; Kerr and Darkow 1996), and the lifting condensation level

* Current affiliation: Northern Gulf Institute, Mississippi State, Mississippi.

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

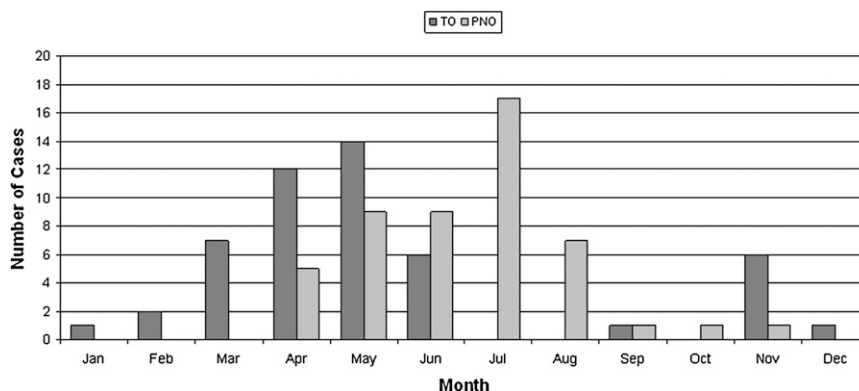


FIG. 1. Number of tornadic and primarily nontornadic outbreaks by month for the original set of cases.

(LCL; Rasmussen and Blanchard 1998). Many of these parameters have been found to be effective discriminating variables of particular storm types or observed severe weather in recent studies (e.g., JD92; Davies and Johns 1993; Johns et al. 1993; Brooks et al. 1994a,b; SCB97; Rasmussen and Blanchard 1998; Craven et al. 2002a,b; Brooks et al. 2003b; Doswell and Evans 2003; Markowski et al. 2003; Thompson et al. 2003).

The outbreaks selected in S09 and M09 (Fig. 1) were the highest ranking TOs from the period 1970–2003 and PNOs from the period 1980–2003, according to a study by Doswell et al. (2006).¹ As expected, high-ranked TOs are most common from March through June, whereas the highest-ranked PNOs were most common from June through August. This seasonal bias has been observed by JD92, Brooks et al. (2003a), Doswell et al. (2005), and others. The samples of TOs and PNOs selected in S09 and M09, and the meteorological covariates used for discrimination, may not have accounted for all the variability in the observed fields, in part because of the seasonal differences in the distributions of TOs and PNOs and their resultant influences on the simulated fields.

The results of S09 and M09 indicate that the most significant TOs and PNOs, based on the two multivariate indices developed for each type of outbreak by Doswell et al. (2006), could be distinguished easily. Convective available potential energy (CAPE) is generally higher during the warm season and lower during the cool season,

whereas wind fields supportive of supercell development and tornadoes are more frequent during the transition seasons of spring and fall than during the summer (JD92). It is not surprising that the highest-ranked TOs in S09 and M09 occurred during the spring, when the collocation of favorable instability and wind shear is more common. During the summer, when wind profiles are generally less conducive for midlevel mesocyclone development, PNO occurrence is frequent relative to TO occurrence, perhaps owing in part to commonly unfavorably high LCLs for widespread tornado development (Brooks et al. 2003b). Because of these seasonal tendencies, discrimination of TOs occurring primarily in the spring and fall from PNOs occurring primarily during the summer should be relatively simple.

Subjective analysis by S09 suggested that PNOs occurring outside of the summer were indeed more difficult to distinguish from TOs. Like the TOs, these PNO cases commonly featured strong shortwave troughs and associated wind fields, suggesting the need for a more thorough investigation of the seasonal impacts on the WRF's ability to distinguish outbreak type.² Furthermore, an investigation of the susceptibility of particular meteorological covariates to seasonal influences is required. This study is a follow-up to S09 and M09, and investigates the extent to which seasonal influences affect the statistics describing the ability of the WRF to discriminate outbreak type.

¹ These cases were chosen because they represent the most important, prototypical cases of each type of outbreak. If little or no ability to discriminate between outbreak types was found with these cases, then there would be little need to investigate more marginal cases of each type, or outbreaks of severe weather not easily classified as either type.

² The WRF's ability to distinguish outbreak type, using synoptic-scale initial conditions, is being evaluated in this study. Though the focus of this study is the determination of the degree to which synoptic-scale processes play a role in discriminating TOs from PNOs, we are not suggesting that the results will be identical using a different mesoscale model. Thus, our results should be interpreted as the WRF's ability to discriminate outbreak type, using synoptic-scale initial conditions.

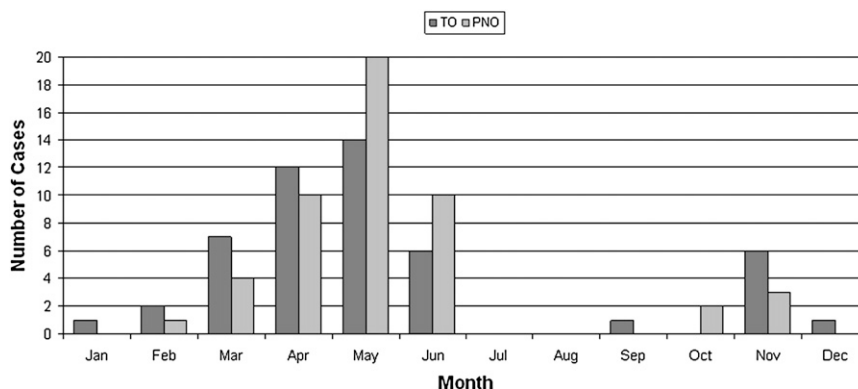


FIG. 2. As in Fig. 1, but for the revised set of cases.

Such modification of the PNO sample makes sense from a forecasting standpoint.³ Because TOs in the summer are relatively rare events, distinguishing between TOs and PNOs would not be a common forecasting concern. Outside of the summer season, both TOs and PNOs are observed, making distinction of cases during this time of year more difficult and more important. A sample of PNOs outside of summer may not be representative of the entire population of PNOs, as those occurring during the summer season would be excluded; however, the discrimination of PNOs outside of the summer months is more important operationally and may provide more complete insight into the utility of various severe weather parameters in distinguishing TOs and PNOs in synoptic-scale environments that are comparable.

2. Data and methodology

a. The cases

To determine the influence of the time of year on the ability to distinguish outbreak type, two sets of cases have been developed. The first is identical to that used in S09 and M09 (Fig. 1; S09, their Tables 1 and 2) and consists of 50 outbreaks of each type. As explained in S09, this number was selected to ensure stability and meaningfulness of conclusions, as discussed in Doswell (2007). Hereafter, this first set of cases will be referred to as the “original” set. *Because no TO in the original set of cases occurred from the period 17 June to 24 September, the second set of cases excludes PNOs that occurred during this period.* Any PNO that occurred outside this period in the original set of cases was included in the second set of cases. The TOs for each set of cases are the

same. Hereafter, the second set of cases will be referred to as the “revised” set (Fig. 2; Table 1) and includes only high-ranked PNOs that happened outside of the exclusion period (effectively, the summer season). The selection of cases is based on the ranking algorithms developed by Doswell et al. (2006) for TOs and PNOs.

An outbreak day is defined to be from the period 1200 UTC on the indicated day to 1200 UTC the following day. For example, the 3 May 1999 TO is valid from 1200 UTC 3 May 1999 to 1200 UTC 4 May 1999. Any outbreak observed to begin outside a 6-h window from 0000 UTC during an outbreak day is excluded from the study (e.g., the 5 July 1980 PNO).

b. The initial dataset

A key aspect of this study is to determine the degree to which synoptic-scale processes occurring before the outbreak influence the magnitude of the tornado component in severe convective storm outbreaks. Therefore, the data used to initialize the WRF simulations of

TABLE 1. The top 50 PNOs in the revised set of cases.

10 Apr 1981	25 Nov 1988	22 May 1999
28 Apr 1981	21 May 1989	19 Apr 2000
8 May 1981	20 Nov 1989	10 May 2000
8 Jun 1982	16 Apr 1990	13 May 2000
18 Mar 1984	19 Mar 1992	2 Jun 2000
14 Apr 1984	10 Jun 1992	6 Apr 2001
20 Oct 1984	31 Mar 1993	9 Apr 2001
12 May 1985	17 May 1993	14 Apr 2001
4 Jun 1985	12 Oct 1993	17 May 2001
5 Jun 1985	10 Apr 1994	14 Jun 2001
7 Jun 1985	4 Nov 1994	2 May 2002
6 May 1986	10 Apr 1995	15 Jun 2002
24 May 1986	24 Mar 1996	2 May 2003
14 Feb 1987	5 May 1996	13 May 2003
21 May 1987	18 May 1996	
17 May 1988	29 May 1998	
24 May 1988	12 Jun 1998	
2 Jun 1988	13 May 1999	

³ As section 2 will discuss, our study is not a forecasting study, however. This project investigates past cases, and much more needs to be done before implementation of our results in an operational setting (section 5).

TABLE 2. Model physical schemes used for this study.

Model physics	References
WRF Single Moment 6-class (WSM6) microphysics	Lin et al. (1983); Dudhia (1989); Hong et al. (1998); Skamarock et al. (2005; 2008)
Grell–Devenyi convective scheme; not used for domains 4 and 5	Grell and Devenyi (2002)
Yonsei University planetary boundary layer scheme	Hong and Pan (1996)
MM5-derived surface layer scheme	Skamarock et al. (2005, 2008)
Five-layer thermal diffusion land surface model	Skamarock et al. (2005, 2008)
Rapid radiative transfer model for longwave radiation	Mlawer et al. (1997)
Dudhia shortwave radiation scheme	Dudhia (1989)

the outbreaks should be limited to synoptic-scale features in some manner. Furthermore, TOs and PNOs are relatively rare meteorological phenomena, requiring a relatively long dataset to be available. A dataset that meets these needs is the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis dataset (Kalnay et al. 1996). A reanalysis is available at 0000, 0600, 1200, and 1800 UTC daily from 1948 onward. With a horizontal grid spacing of 2.5° latitude by longitude (approximately 200 km), 17 vertical levels, and optimal interpolation-based quality control of the data (Woollen 1991; Woollen et al. 1994) to remove erroneous data or measurements that do not represent the spatial and temporal scales intended for the analysis, the NCEP–NCAR reanalysis dataset represents “synoptic scale” data as discussed by Orlanski (1975) and Holton (1992), among others. Additional details on the development of this dataset are provided by Kalnay et al. (1996) and S09.

c. WRF model physics and domain setup

To perform the outbreak simulations, versions 2.1.2, 2.2, and 3.0.1.1 of the WRF (Skamarock et al. 2005, 2008) were used for this study. A description of the WRF and the physical setup is provided in S09 and is identical with this study. A brief summary of the most important aspects of the simulations is provided below.

The parameterization schemes selected for this study (Table 2) are those of S09. Lateral boundary conditions for each simulation were implemented using the relevant NCEP–NCAR reanalysis datasets, and the model top was at 50 hPa. Every simulation featured five domains (Fig. 3), with domain 1 fixed. Domains 2 and 5 were fixed with respect to their parent domain, whereas domains 3 and 4 were positioned to be centered on where each outbreak was observed. This positioning was performed subjectively, based on the reports as plotted by Severe Plot (SVRPLLOT; Hart 1993). Domain 1 had a horizontal grid spacing of 162 km, and the grid spacing was reduced by a factor of 3 for each child domain. There were 31 vertical levels for each domain, stretched with height to resolve the boundary layer

more adequately. Two-way nesting was incorporated. Each outbreak case was simulated for 1-, 2-, and 3-day forecasts. Because the valid time of the outbreak for each case was assumed to be 0000 UTC, a 1-day simulation was initialized at 0000 UTC the previous day. For example, for the 3 May 1999 TO, the 1-day simulation was initialized at 0000 UTC 3 May 1999 and was considered valid at 0000 UTC 4 May 1999. S09 found little decline in the ability of the WRF to distinguish outbreak type 3 days in advance of the outbreak compared to 1 day in advance using the original set of cases. A key question to be analyzed in this study is whether greater degradation in time is noted with the WRF’s ability to discriminate outbreak type using the revised set of cases, which is likely to have implications on predictability of outbreak type in operational forecast settings.

d. Permutation testing

After the simulations were performed, both subjective and objective analyses were conducted. We sought to determine if synoptic-scale processes occurring before the outbreak influenced the mesoscale environments of the two types of outbreaks and can be distinguished accordingly. With this objective in mind, the objective analysis begins by using a portion (21×21 grid points) of domain 3 (18-km grid spacing; Fig. 3) centered on the observed outbreak. The center of the outbreak was chosen subjectively, by analyzing the severe reports of the SVRPLLOT program.⁴ Objective determination of the outbreak center was not carried out, but may be considered in future studies.

The choice of domain was shown by M09 to be a practical means of objectively classifying outbreak type. Phase and timing errors in model simulations possibly could result in the severe environment not being represented in the domain at the valid time of the outbreak, and this is

⁴ The center of the outbreak was not chosen as the approximate mean or median location of all of the severe reports on a given outbreak day. Rather, the center was chosen to be the approximate mean or median location of the tornado reports for TOs and the approximate mean or median location of severe reports in a 6-h period around the outbreak valid time for PNOs.

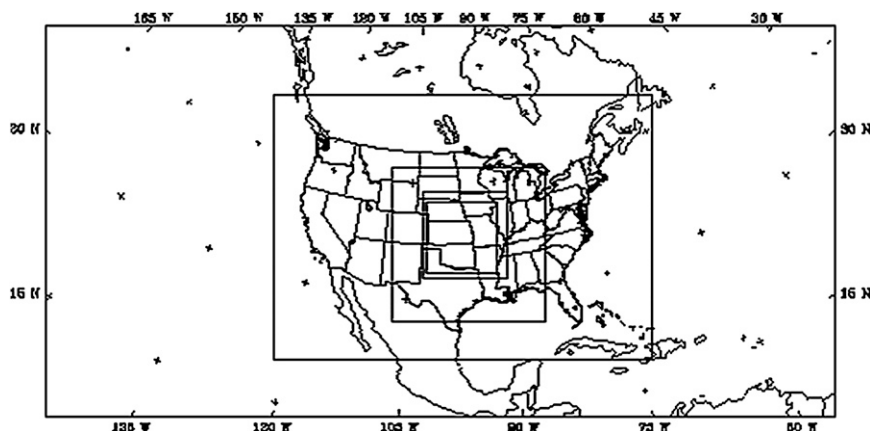


FIG. 3. Sample domain setup for WRF simulations (from S09).

not accounted for in the current study. However, S09 showed that these errors would affect the classifications for a small number ($<10\%$) of cases, suggesting that the domain size is acceptable for our purposes. An investigation into increasing the domain size to 41×41 grid points suggested that changes in the contingency statistics were not statistically significant at the 0.05 level (not shown). Section 5 elaborates upon the difficulties faced when choosing the *model's* outbreak center.

Permutation testing (Efron and Tibshirani 1993) of 1-day WRF simulated fields at each of the 441 grid points for several meteorological covariates associated with severe weather and tornadoes (as in M09) was performed, for a subset of 25 PNOs occurring only during the summer season (defined herein as 17 June–24 September; see section 2a) and for a subset of 25 PNOs occurring outside of the summer season. Permutation testing determines whether the means of two distributions are statistically distinct, with no assumptions regarding the distributions [see Mercer and Richman (2007) and M09 for more discussion]. As indicated in Table 3, the percentage of grid points in which the p value was less than 0.1 was greater than 50% for most severe weather parameters (indicating the means of the two distributions were significantly different to 90% confidence at those particular grid points). For parameters such as surface-based CAPE (SBCAPE), LCL, and deep-layer shear, a majority of grid points featured p values <0.05 , confirming the seasonal dependence of these variables as suggested by S09. The results of these permutation tests provide support for investigating the differences in discrimination capability for PNOs occurring specifically during the times of year in which TOs are relatively common.

e. Objective classification

The 441 grid points for each meteorological field were combined for each case in the original and revised lists to

create a $(441 \times n) \times N$ data matrix for a principal component analysis (PCA; Richman 1986), where n is the number of meteorological variables included in the analysis and N is the number of cases in each list (99 for the original cases; 100 for the revised cases). PCA allows for the reduction of the data matrices (i.e., the gridded fields from the “subdomain”) while maintaining information on the spatial structure of the meteorological fields. This PCA methodology is an objective means of accounting for the spatial patterns in the variables. The PC score matrix is of dimensions $N \times r$, where r is the number of PCs retained (e.g., by scree test that plots the eigenvalue magnitude against eigenvalue number;

TABLE 3. Results of permutation tests of the 1-day WRF simulations of 25 PNOs occurring from the period 17 Jun–24 Sep and 25 PNOs occurring outside this period. Values indicate the percentage of the 441 grid points analyzed in which the p values are less than that indicated. Refer to Table 4 for parameter abbreviations.

Covariate	$p < 0.1$	$p < 0.05$	$p < 0.01$
0–1-km bulk shear	56.24	43.54	13.61
0–3-km bulk shear	58.73	52.61	40.59
0–6-km bulk shear	97.73	77.55	51.70
Bulk Richardson number shear	46.03	40.36	11.79
Surface-based CAPE	100.00	99.09	76.42
Surface-based CIN	23.36	14.74	7.94
Product of CAPE and 0–1-km shear	87.53	64.40	15.87
Product of CAPE and 0–3-km shear	92.06	75.96	19.05
Product of CAPE and 0–6-km shear	89.57	56.46	14.29
0–1-km EHI	15.87	4.99	0.00
0–3-km EHI	20.41	6.12	0.00
LCL	78.00	58.96	15.87
LFC	0.23	0.00	0.00
Mean sea level pressure	63.27	8.16	0.00
Storm-relative flow	65.76	55.56	22.68
0–1-km SREH	54.42	45.35	21.54
0–3-km SREH	55.78	50.34	29.02
Vorticity generation potential (VGP)	64.63	48.07	16.33

Wilks 1995). Because there were two sets of cases to analyze, two PCAs were conducted for each meteorological variable or combination of variables analyzed.

The PC score matrices from each PCA were used as input for a statistical method known as support vector machines (SVMs; Haykin 1999; Cristianini and Shawe-Taylor 2000). An SVM is a method of machine learning that determines a hyperplane that classifies the data optimally. An overview of binary SVM classification is provided in appendix A of M09. The utility of SVMs in previous studies (e.g., Lee et al. 2004; Richman et al. 2005; Mercer et al. 2008; M09), as well as with classifying the original set of outbreak cases in M09, provides support for their use in this study.

A particularly attractive characteristic of SVMs is their ability to model nonlinear input. Kernel functions (KFs) are used to map input data that are not linearly separable into a higher-dimensional dataset (Cristianini and Shawe-Taylor 2000). Three KFs were tested in this study:

- Polynomial KF

$$k(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T \mathbf{y} + 1)^p. \quad (1)$$

- Radial basis KF

$$k(\mathbf{x}, \mathbf{y}) = e^{-1/2\sigma^2\|\mathbf{x}-\mathbf{y}\|^2}. \quad (2)$$

- Tangent hyperbolic KF

$$k(\mathbf{x}, \mathbf{y}) = \tanh(\beta_o \mathbf{x}^T \mathbf{y} + \beta_1). \quad (3)$$

The radial basis KF typically outperformed the other kernel function techniques when classifying data in this study, using the same cross-validation technique as applied in M09, and was used in the subsequent analysis described below.

The statistical model developed by the SVMs must be generalized to classify cases not supplied in a training set. A technique derived from the so-called jackknife (Efron and Tibshirani 1993) is conducted on the datasets to carry out this criterion. For the original (revised) set of cases, the 99 (100) cases were placed in random order.⁵ The first 85% of the cases were used as a training set, and the remaining 15% were used to test the

statistical model.⁶ This process was repeated, in which the first training case was placed in the testing set and the first test case was placed in the training set: 99 (100) times for the original (revised) set of cases to ensure that each case was trained and tested. The SVMs feature tuning parameters, including a cost coefficient (see Cristianini and Shawe-Taylor 2000; Mercer et al. 2008; M09) and parameters in the kernel function, such as σ in (2), that can be modified to assess the generalization of the statistical model developed. [Note that p in (1) and β_o and β_1 in (3) are also tuning parameters.] Unfortunately, determination of optimal values for these parameters is computationally demanding and commonly infeasible to calculate. Though a range of values for each of the tuning parameters was assessed for each PC score matrix analyzed, there is no guarantee that the tuning parameters selected were the “optimal solution” to the classification problem. This is a relatively minor concern, however, because of the nature of the optimality problem. In three-dimensional space (the two tuning parameters, using the radial basis kernel function, and the statistic being calculated), the optimal maximum (minimum) either has a steep or small ascent (descent). If the slope is steep, determination of values close to the optimal value is relatively easy. If the slope is small, determination of the optimal values provides little additional gain.

A large number of meteorological covariates could be analyzed when assessing the WRF’s ability to distinguish outbreak type. For each variable individually and in combination with other variables, the above technique must be performed. A preemptive reduction to a few physically meaningful variables (e.g., using permutation testing; see M09) may offer limited benefit. For example, if eight meteorological covariates were found to be physically meaningful for discriminating outbreak types, there would be $2^8 - 1$, or 255, possible combinations of variables. Clearly, this becomes a time-consuming task, as 99 statistical models have been evaluated for each of a potentially very large number of variable combinations (for the original set of cases), only after determining the best tunable parameters for the SVM. Combined with the same procedure necessary for the revised set of cases, an exhaustive analysis of WRF simulations of meteorological covariates becomes impractical.

⁵ Note that the statistical models were developed separately for the original set of cases and for the revised set of cases. That is, a subset of the original (revised) dataset was used to develop a statistical model to be tested on the remaining cases of the original (revised) dataset. The statistical models developed using the original (revised) training data were not tested on cases in the revised (original) dataset.

⁶ Various numbers of training set cases were tested with some of the meteorological parameters (specifically, 5%, 10%, . . . , 95% of the total number of cases). The results indicated that the contingency statistics improved with increased training set size, but the uncertainty in the results increased (not shown). The choice of a training set size of 85% was observed to be a good compromise for the two trends.

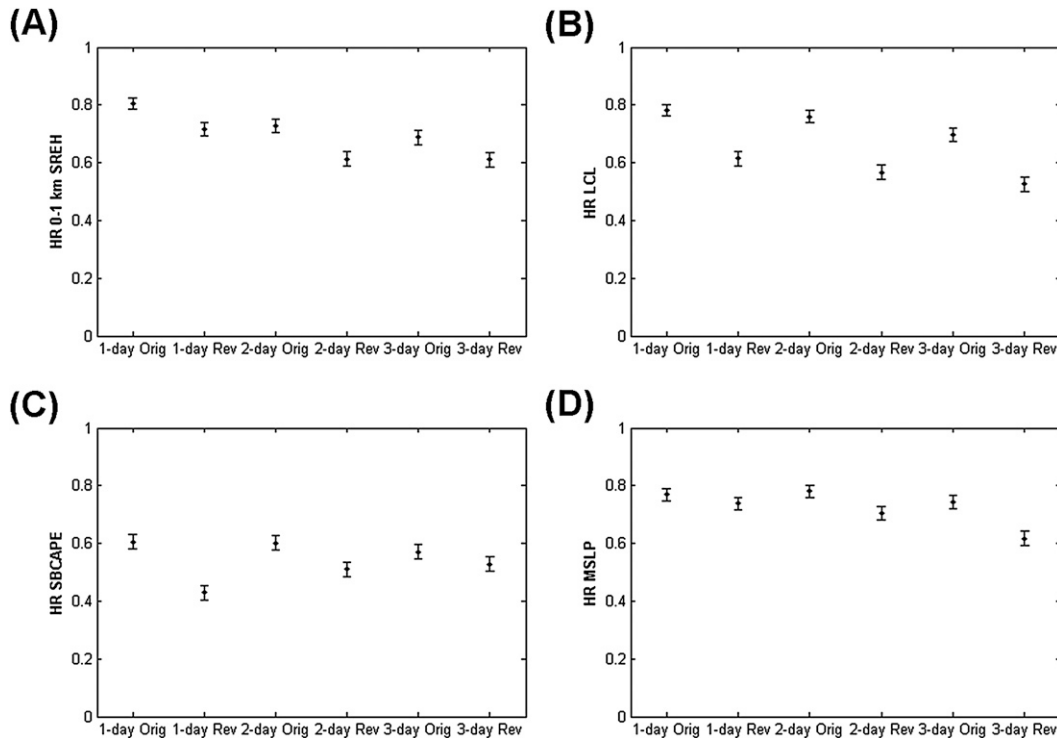


FIG. 4. Tilted bootstrap confidence intervals of the HR for 1-, 2-, and 3-day simulations for (a) 0–1-km SREH, (b) LCL, (c) SBCAPE, and (d) MSLP, for the original (orig) and revised (rev) set of cases.

This study is concerned primarily with individual meteorological fields, in contrast with M09, because of the limitations described above.⁷ This emphasis on single variables permits investigation into the relative utility of each meteorological covariate for each set of cases (original and revised) while still allowing a comparison between the WRF’s ability to distinguish outbreaks for the original set of cases with that for the revised set of cases. However, some combinations of meteorological covariates are considered in section 4 to evaluate the utility in looking at multiple variables to diagnose outbreak type.

A limitation of the discriminating algorithm used in this study is that the SVM produces 99 (100) statistical models for distinguishing outbreak type for the original (revised) set of cases. Determination of the “best” statistical model for forecast applications requires a completely independent set of cases to evaluate each statistical

model’s utility. However, TOs and PNOs are rare events, and accumulation of a large number of cases would require many more years of outbreaks to observe and analyze (Doswell 2007).

Because this study is a binary classification problem, contingency statistics (e.g., Wilks 1995) can be computed for the performance of the statistical models developed by the SVM on the test data. Numerous statistics were analyzed in this study, including hit rate (HR), probability of detection (POD), false alarm ratio (FAR), probability of false detection (POFD), the Heidke skill score (HSS; Doolittle 1888), and the Peirce skill score (PSS; Peirce 1884). Finally, 95% confidence limits based on bootstrap tilting (Efron and Tibshirani 1993) are computed for each of the contingency statistics⁸ to compare the capability of the WRF, using synoptic-scale input, at distinguishing outbreak type for the original and revised set of cases, to compare various meteorological covariates to each other for the same set of cases

⁷ We are by no means suggesting that the optimal set of meteorological parameters in distinguishing outbreak type is a single variable. The limitations of using a single variable to distinguish outbreak type are addressed in S09 and M09. For example, M09 noted that there is more variability in the bootstrap samples of the contingency statistics of the statistical model predictions for single variable fields versus fields with a large number of parameters.

⁸ The bootstrap samples are composed of matrices of $15 \times N$ rows and 4 columns, where $15 \times N$ is the total number of tests conducted for the N statistical models developed and each of the columns represents a possible forecast/observed pair in the contingency table. Contingency statistics are calculated for each bootstrap sample.

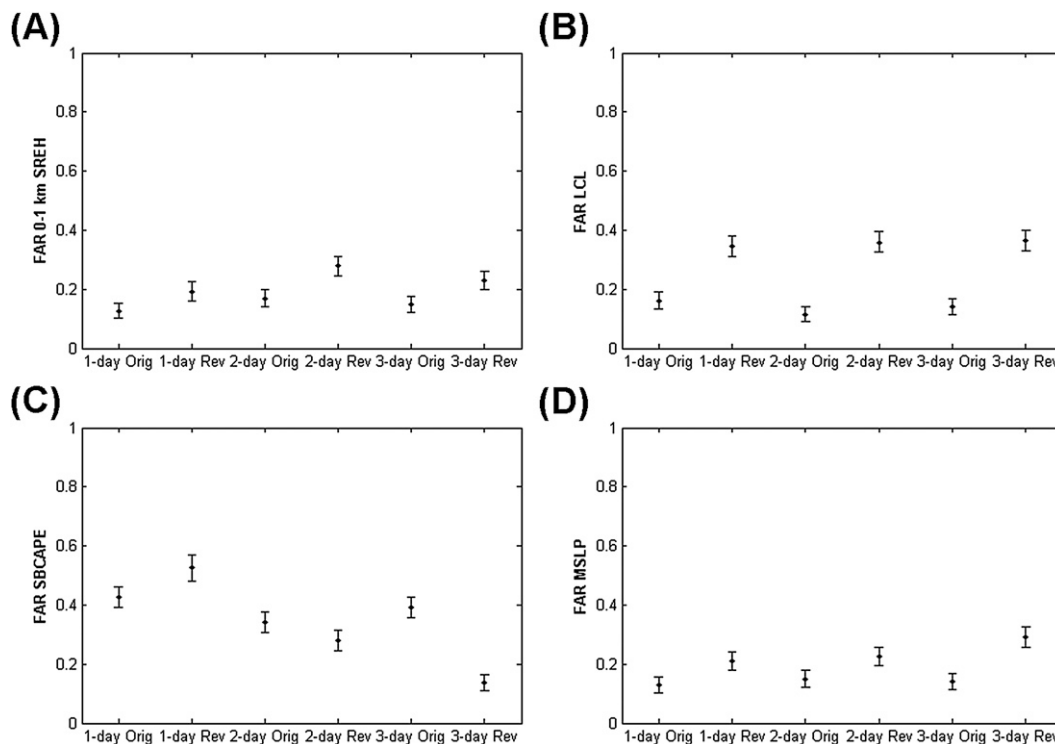


FIG. 5. As in Fig. 4, but for the FAR.

for the same forecast period, and to investigate forecast degradation or improvement as forecast period increases.

The statistical models developed by the SVM used in this study often are difficult to interpret physically (owing to nonlinearity). Subjective analysis of the individual simulations can be conducted to identify any systematic differences of TO and PNO cases that the statistical model has used to discriminate the outbreak types. Similarities between subjective and objective analyses have been observed in previous studies (cf. S09 and M09), but some examples of complementary subjective analysis will be provided in section 4 for completeness.

As noted in S09, the results of this study are conditioned on the occurrence of severe weather outbreaks. *This is not a forecasting study. Determining whether an outbreak of some type will occur, given a WRF (or any model) simulation, is beyond the scope of this work.* Furthermore, the model fields are analyzed where and when the actual outbreak occurred. Any spatial and temporal errors in the model simulations are not accounted for and very likely affect the results discussed in the following sections. However, these errors are observed in a relatively small number of cases (S09, see their sections 4e,f and their Tables 7 and 8).

3. Results

a. Comparison of discrimination capability for original and revised cases

An analysis of the tilted bootstrap 95% confidence intervals (CIs) for HR (Fig. 4), POD (not shown), FAR (Fig. 5), POFD (not shown), HSS (not shown), and PSS (Fig. 6) for four meteorological covariates indicates that the WRF, using synoptic-scale initial data, is significantly less capable of distinguishing outbreak type correctly when the list of cases is constrained to exclude events occurring in the summer months. For example, the HR CIs for 0–1 km SREH (SREH1) for 1-, 2-, and 3-day simulations do not overlap, suggesting that the WRF's ability to distinguish outbreak type using SREH1 is significantly degraded when using the revised set of cases. Differences in the median of the bootstrap CIs for the original and revised sets of cases generally are within the range of 5%–10% for each of the forecast periods. The observed HR and the medians of the HR CIs suggest that there is approximately 82%–83% accuracy in determining outbreak type for 1-day simulations of the original set of cases, with a reduction to 73%–75% for 1-day simulations of the revised set of cases. Although this degradation is nontrivial, these scores indicate that discrimination of outbreak type using SREH1 is still relatively good when

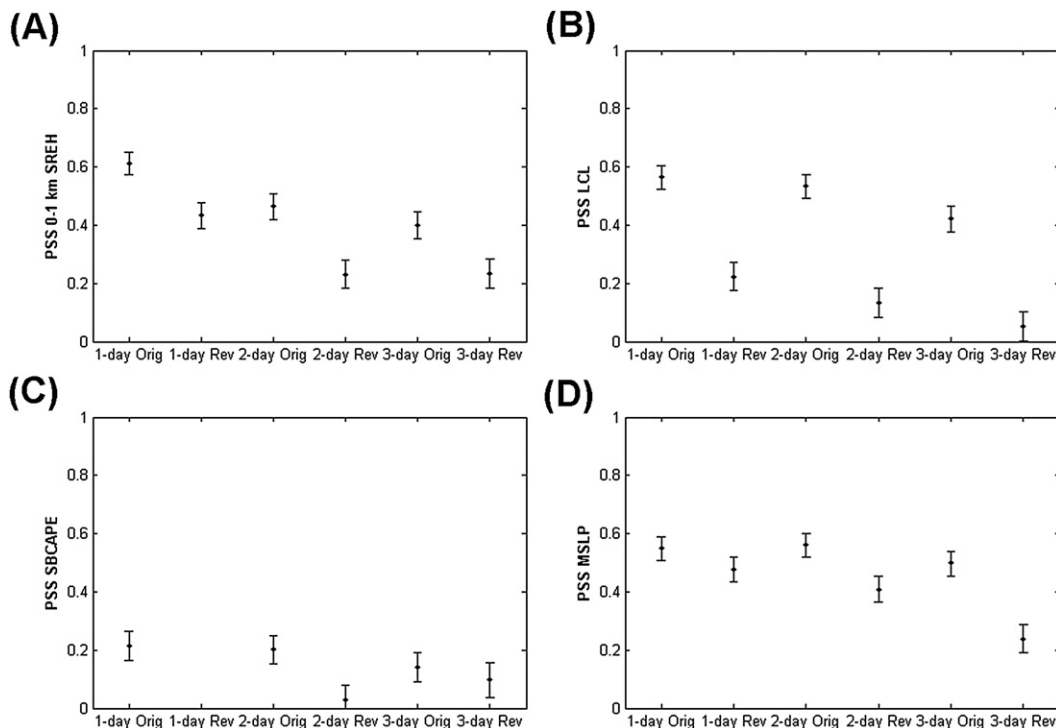


FIG. 6. As in Fig. 4, but for the PSS.

using only synoptic-scale data as initial input into the model. Even 3-day simulations show some accuracy in the revised set of cases. The median of the HR CIs for the original set of cases is above 70%, decreasing to slightly above 60% for the revised set of cases. No CIs overlap HRs of 0.5 (equivalent to flipping a coin).

However, degradation is more pronounced for the LCL, with differences in the observed HR and medians of the HR CIs ranging from 15% to 20%. This result suggests that the LCL is especially susceptible to seasonal influences. This was posited by S09, who suggested that higher LCLs were common with many PNOs during the summer months, owing to a prevalence of relatively hot, dry, well-mixed boundary layers.

The trends of the HR CIs for SBCAPE and other thermodynamic variables are much different than wind and synoptic parameters. In general, there is a gradual decline in HR with time for the original set of cases (though overlap in the CIs is commonly observed), whereas there is slow *improvement* with increased forecast period for the revised set of cases. This slow improvement is possible because simulated convection often occurs with 1-day simulations, which acts to remove instability. This tendency can be interpreted to mean that using CAPE, CIN, and other thermodynamic parameters in the short range are unhelpful if simulated convection has already developed in the simulation. However, 2- and 3-day simulations contain

more instances in which no simulated convection develops (S09, see their section 4f), making distinction of outbreaks using SBCAPE more feasible.

This trend was not observed with the original set of cases, possibly because many of the PNOs occur during the summer months. These events typically involve subtle synoptic-scale shortwave troughs and limited large-scale upward vertical motion. Simulated convection is less likely to develop in these circumstances (e.g., Stensrud and Fritsch 1994; Moller 2001), allowing the thermodynamic instability parameters to remain relatively undisturbed. Subtle degradation is observed with these cases, as more TO simulations feature poorly simulated convection with increasing forecast lead time.

As indicated by the HR, FAR, and PSS CIs, there is little accuracy or skill in using SBCAPE as a meteorological covariate in any of the forecast periods for either set of cases. This result, and those for SBCIN and other instability parameters (not shown), suggests that thermodynamic instability parameters are generally unhelpful in distinguishing outbreak type. This agrees with several previous studies (JD92; Johns et al. 1993; Monteverdi et al. 2003; S09; M09), many of which indicate that tornadoes can occur within a wide range of CAPE values.

It was also posited by S09 that the strength and spatial structure of the shortwave troughs and surface cyclones

TABLE 4. Meteorological parameters and associated abbreviations as indicated in the text and figures. References are also included where appropriate.

Severe parameter	References
Lifting condensation level (LCL)	Rasmussen and Blanchard (1998)
0–1-, 0–3-, and 0–6-km bulk shear (BULK1, BULK3, and BULK6, respectively)	Bunkers (2002)
0–1- and 0–3-km storm relative environmental helicity (SREH1 and SREH3, respectively)	Davies-Jones et al. (1990); Davies and Johns (1993); storm motions based on Maddox (1976)
0–1- and 0–3-km energy helicity index (EHI1 and EHI3, respectively)	Hart and Korotky (1991); Davies (1993)
Bulk Richardson number shear (BRNSHR)	Droegemeier et al. (1993)
Vorticity generation potential (VGP)	Rasmussen and Blanchard (1998)
700- and 850-hPa geopotential height (H700 and H850, respectively)	
Storm-relative flow ~2 km AGL (SRFL)	Kerr and Darkow (1996)
Supercell composite parameter (SCP)	Thompson et al. (2003)
Significant tornado parameter (STP)	Thompson et al. (2003)
Mean sea level pressure (MSLP)	
Surface-based convective available potential energy (SBCAPE)	SCB97
Surface-based convective inhibition (SBCIN)	Colby (1984)
Level of free convection (LFC)	Blanchard (1998)

could be used as a reliable means of distinguishing outbreak type. In this study, several “synoptic parameters,” including geopotential height, wind speed, temperature, relative humidity, and mean sea level pressure (MSLP) also were analyzed to determine if these parameters could be useful in outbreak discrimination. Indeed, results appear to indicate the utility of certain parameters, including low-level geopotential heights, low-level wind speeds, and MSLP. For example, the HR CIs for MSLP appear to have ranges of values similar to or higher than those for SREH1, indicating that using either parameter could provide similar accuracy in discriminating outbreak type.

This result should not be taken to suggest that MSLP or other synoptic parameters can predict outbreaks of severe weather as well or better than severe weather parameters. Rather, the similar results between the statistics of the severe weather parameters and synoptic parameters indicate that, if an outbreak of severe weather will occur or has occurred, the use of synoptic parameters may be just as beneficial as using severe weather parameters in discriminating outbreak type. The utility of synoptic parameters in diagnosing outbreak type should not be very surprising. Severe weather parameters generally are derived from, or are combinations of, synoptic parameters (see Doswell and Schultz 2006).

The HR CIs of MSLP for the original set of cases contain considerable overlap for the three forecast periods. This is likely because of the large number of summer PNO cases in this list of events. Many of the summer PNOs involve relatively weak synoptic-scale systems (i.e., surface cyclones) compared to the TOs. Although forecast

error increases as forecast period increases, the MSLP fields for the summer PNOs are weaker (magnitudes and gradients), in general.

This consistent accuracy does not appear in the statistics for the revised cases. However, MSLP remains a very useful meteorological parameter to discriminate outbreaks, especially in 1- and 2-day simulations. Therefore, the strength and spatial structure of synoptic-scale systems (viz., for MSLP, the surface cyclone, and surrounding pressure field) provides useful information regarding the discrimination of outbreak type. This finding can be interpreted to support the use of pattern recognition, at least for highly ranked outbreaks, a common method for operational forecasters in predicting the potential for severe weather outbreaks (Miller 1972; JD92).

In general, the tilted bootstrap POD (not shown), FAR (Fig. 5), POFD (not shown), HSS (not shown), and PSS CIs (Fig. 6) are similar in trend to the HR CIs. Many of the same conclusions regarding the sensitivities of these parameters to season, simulation errors, and simulated convection using HR CIs can be observed with the POD, FAR, and PSS CIs, as well. Particularly noteworthy results for these three statistics include 1) the improvement of FAR for SBCAPE for 2- and 3-day simulations of the revised set of cases, complementing the improved HR and PSS with increased forecast period; 2) the especially significant degradation in scores for LCL when using the revised case list, confirming this parameter’s relatively high sensitivity to seasonal influences; 3) the relatively higher skill scores of 0–1-km SREH and MSLP compared to LCL and SBCAPE,

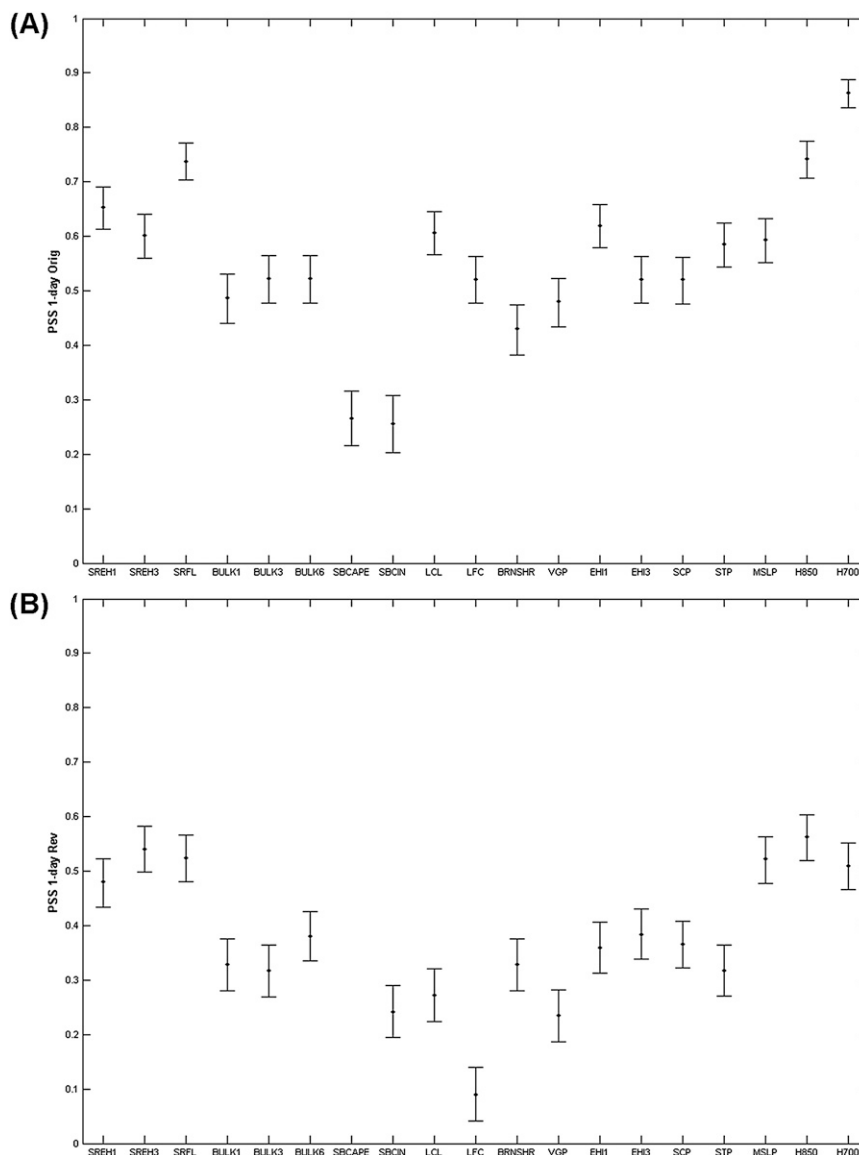


FIG. 7. Tilted bootstrap CIs for the PSS of 1-day simulations for the (a) original and (b) revised set of cases for various meteorological covariates. Refer to Table 4 for parameter abbreviations.

providing evidence that wind shear and synoptic parameters are more skillful when discriminating outbreak types; and 4) the lack of overall trends in most of the FAR CIs, which combined with the lack of trends noted in the POFD CIs (not shown), indicate most degradation in accuracy and skill scores is through a reduction in the probability to detect tornado outbreaks.

b. Comparing meteorological covariates

The accuracy and skill with which simulated fields of numerous meteorological parameters can be used to distinguish outbreak type vary with the type of parameter,

with the list of cases being analyzed, and with forecast period. The tilted bootstrap PSS CIs for a number of these parameters (see Table 4 for abbreviations) for 1-day simulations of the original set of cases (Fig. 7a) indicate that storm-relative parameters (e.g., 0–1 km SREH; 0–3 km SREH; SRFL approximately 2 km above ground level) and low-level synoptic parameters (such as 850- and 700-hPa geopotential heights and MSLP) appear to be the most useful parameters in distinguishing outbreak type. In particular, the CIs for 700-hPa geopotential heights exceed 80% and do not overlap any other parameter analyzed for this study. SRFL and 850-hPa geopotential

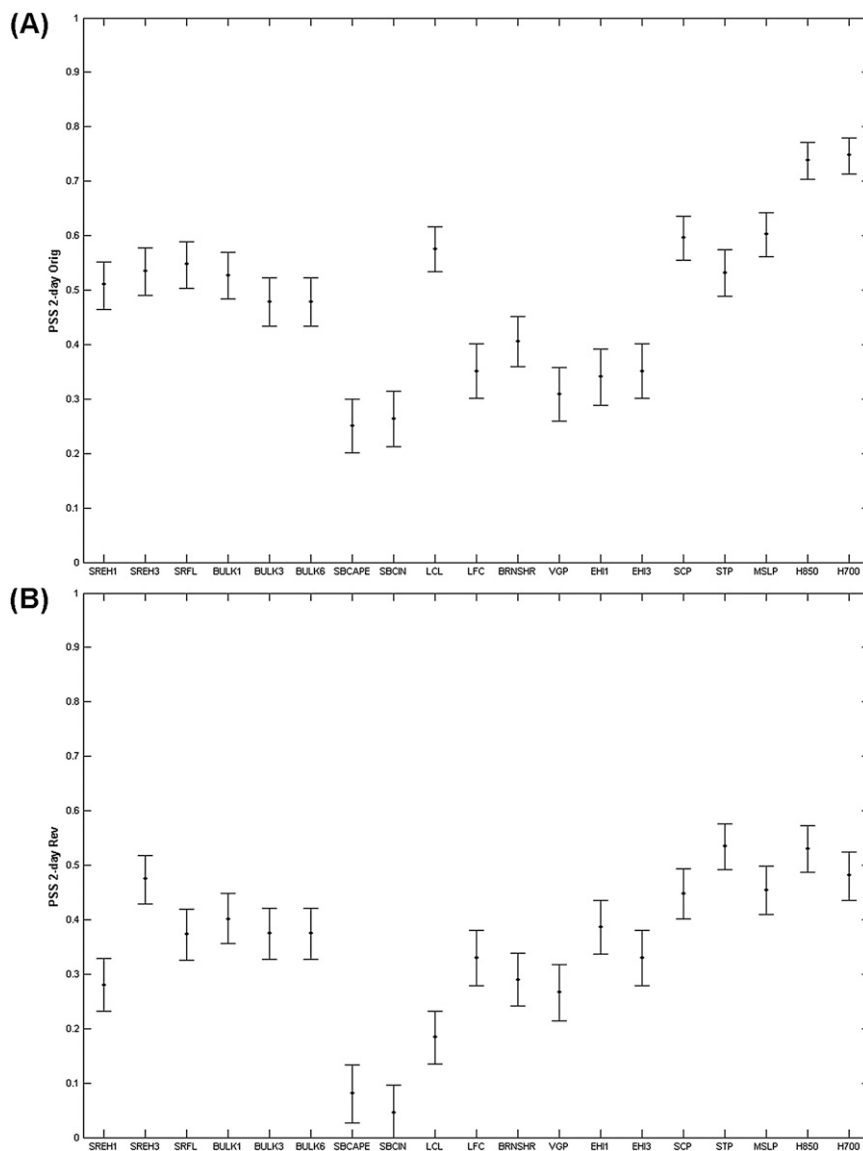


FIG. 8. As in Fig. 7, but for 2-day simulations.

heights both exceed 70% and also do not overlap the rest of the parameters analyzed in the study. The CIs for 0–1- and 0–3-km SREH do overlap other parameters, including LCL, 0–1-km EHI, MSLP, and significant tornado parameter (STP), but clearly are superior to other severe weather parameters analyzed [including Bulk Richardson number shear (BRNSHR); low-level and deep-layer bulk shear; vorticity generation potential, or VGP; SBCAPE; surface-based convective inhibition, or SBCIN; etc.]. The favorable performance of storm-relative helicity parameters agrees with many previous studies (e.g., Davies-Jones et al. 1990; Davies and Johns 1993; Drogemeier et al. 1993; SCB97; Rasmussen and Blanchard 1998; Thompson et al. 2003; S09).

Some differences are noted, however, for the PSS CIs of the 1-day simulations for the revised list of cases (Fig. 7b). For example, the 700-hPa geopotential heights no longer have a distinct advantage over any other parameter analyzed. The CIs for 0–1-km SREH, 0–3-km SREH, SRFL, MSLP, 700-hPa geopotential height, and 850-hPa geopotential height are relatively similar with considerable overlap. These 6 parameters are generally outside the 95% CIs for the remaining parameters, indicating that the observed advantage of these parameters with the original set of cases remains when modifying the case list. However, the degradation in the WRF's ability to distinguish outbreak type is evident, with the best PSS CIs only near or slightly above 0.5.

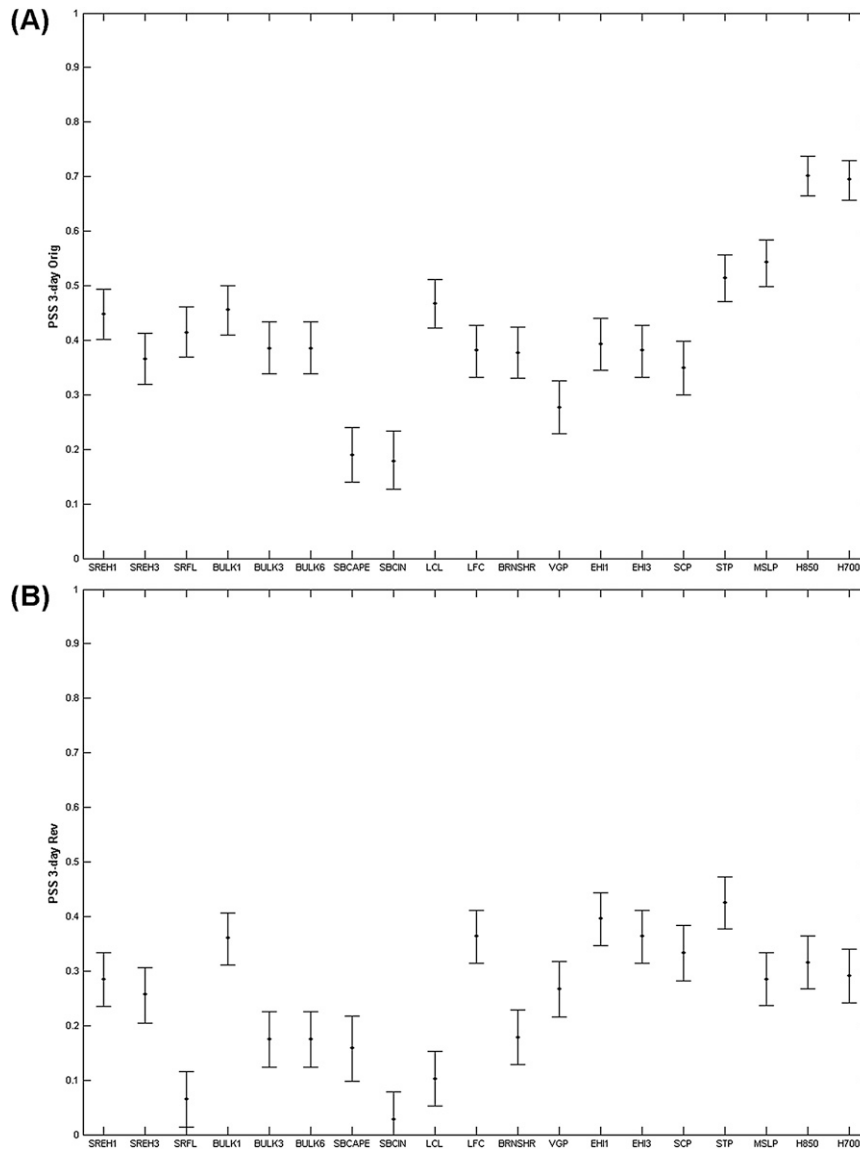


FIG. 9. As in Fig. 7, but for 3-day simulations.

The skill of LCL also declines considerably (cf. Figs. 7a,b), with its relative utility in distinguishing outbreak type for the original set of cases all but eliminated when revising the list of cases to those not occurring in the summer season. The skill of 0–1-km EHI also decreases substantially, perhaps in part because of the very poor skill observed with thermodynamic instability parameters (see section 3a). Finally, the observed skill of several severe weather indices, including the EHI, supercell composite parameter (SCP), and STP, in distinguishing outbreak type does not appear to be significantly better or worse than using other severe weather parameters (including bulk shear, BRNSHR, etc.) with 1-day simulations.

For the 2- and 3-day simulations, the PSS CIs for the original set of cases (Figs. 8a and 9a) and the revised set of cases (Figs. 8b and 9b) feature degradation in the scores with increased forecast length, with the best PSS CIs for the original (revised) set of cases decreasing from 0.8 to 0.9 (0.5 to 0.6) for 1-day simulations to approximately 0.65 to 0.75 (0.4 to 0.5) for 3-day simulations. The most skillful parameters for the original and revised set of cases are low-level synoptic parameters (such as 850- and 700-hPa geopotential heights and MSLP) and storm-relative helicity parameters (such as 0–1- and 0–3-km SREH). The least skillful parameters are thermodynamic instability parameters, such as SBCAPE and SBCIN. The use of indices (such as EHI, SCP, and STP) appears to be

TABLE 5. Number of correct and incorrect classifications of outbreak type for each of the 1-day simulations of the revised set of cases. The numbers in each category represent the number of correctly or incorrectly classified outbreak types for the relevant cases. For example, one “error TO” indicates a TO that is incorrectly classified as a PNO once out of the 1425 tests. Tests included 95 combinations of meteorological covariates, including parameters individually ($95 \times 15 = 1425$).

Outbreak	Error TO	Correct	Error PNO	Outbreak	Error TO	Correct	Error PNO
21 Jan 1999	1411	14	N/A	21 May 1989	N/A	1207	218
27 May 1973	1383	42	N/A	15 Nov 1988	203	1222	N/A
20 Oct 1984	N/A	88	1337	12 Oct 1993	N/A	1222	203
19 Apr 2000	N/A	152	1273	13 Mar 1990	199	1226	N/A
18 Mar 1984	N/A	185	1240	12 Jun 1998	N/A	1231	194
24 Mar 1996	N/A	202	1223	7 Jun 1984	187	1238	N/A
25 Nov 1988	N/A	255	1170	26 Apr 1984	183	1242	N/A
31 May 1985	1153	272	N/A	5 Jun 1985	N/A	1251	174
10 Apr 1995	N/A	337	1088	21 Nov 1992	170	1255	N/A
6 Apr 2001	N/A	349	1076	7 Apr 1980	163	1262	N/A
3 May 1999	1015	410	N/A	2 Apr 1982	160	1265	N/A
17 Apr 1970	945	480	N/A	8 May 1981	N/A	1265	160
8 Apr 1999	933	492	N/A	16 Apr 1990	N/A	1268	157
6 May 2003	928	497	N/A	12 May 1985	N/A	1271	154
14 Feb 1987	N/A	526	899	29 Mar 1976	153	1272	N/A
23 Apr 2000	838	587	N/A	25 Sep 1973	147	1278	N/A
18 May 1996	N/A	606	819	19 Apr 1996	147	1278	N/A
10 Apr 1994	N/A	608	817	10 Nov 2002	130	1295	N/A
4 Nov 1994	N/A	609	816	7 Jun 1993	125	1300	N/A
31 Mar 1993	N/A	623	802	15 Mar 1982	122	1303	N/A
20 Nov 1989	N/A	661	764	24 Feb 2001	121	1304	N/A
14 Jun 2001	N/A	666	759	10 Jun 1992	N/A	1306	119
16 Apr 1998	754	671	N/A	2 Jun 2000	N/A	1309	116
7 May 1993	698	727	N/A	14 Dec 1971	114	1311	N/A
23 Nov 2001	667	758	N/A	10 May 2003	109	1316	N/A
14 Apr 2001	N/A	793	632	1 Mar 1997	106	1319	N/A
2 Jun 1990	628	797	N/A	6 May 1986	N/A	1328	97
26 May 1973	618	807	N/A	28 Apr 1981	N/A	1332	93
26 Apr 1994	591	834	N/A	17 May 1988	N/A	1335	90
13 May 2000	N/A	842	583	2 Jun 1988	N/A	1336	89
2 May 2002	N/A	858	567	2 May 2003	N/A	1339	86
18 May 1995	551	874	N/A	5 May 1996	N/A	1339	86
14 Apr 1984	N/A	888	537	17 May 2001	N/A	1342	83
19 Mar 1992	N/A	964	461	15 Jun 2002	N/A	1343	82
4 May 1977	425	1000	N/A	3 Apr 1974	72	1353	N/A
27 May 1995	409	1016	N/A	10 Apr 1979	71	1354	N/A
31 May 1983	399	1026	N/A	13 May 1999	N/A	1354	71
15 Jun 1992	397	1028	N/A	24 May 1988	N/A	1360	65
8 Jun 1982	N/A	1028	397	24 Nov 2001	62	1363	N/A
9 Apr 2001	N/A	1064	361	22 May 1999	N/A	1367	58
27 Mar 1994	351	1074	N/A	4 Jun 1985	N/A	1369	56
16 Jun 1992	333	1092	N/A	21 May 1987	N/A	1369	56
8 May 1988	302	1123	N/A	24 May 1986	N/A	1376	49
13 May 2003	N/A	1147	278	21 Feb 1971	46	1379	N/A
10 Apr 1981	N/A	1151	274	4 May 1999	46	1379	N/A
20 Mar 1976	259	1166	N/A	7 Jun 1985	N/A	1379	46
26 Apr 1991	254	1171	N/A	28 Mar 1984	43	1382	N/A
10 May 2000	N/A	1178	247	29 May 1998	N/A	1404	21
22 Nov 1992	239	1186	N/A	17 May 1993	N/A	1405	20
4 May 2003	218	1207	N/A	8 Jun 1974	17	1408	N/A

more beneficial with longer forecast period, especially for the revised set of cases, whereas the relative utility with using SREH degrades with increased forecast period. The utility of low-level geopotential heights for the

original set of cases degrades significantly for the revised set of cases, indicating the sensitivity of these parameters to seasonal influences. Finally, the relative benefit of using LCL as a distinguishing parameter for the original

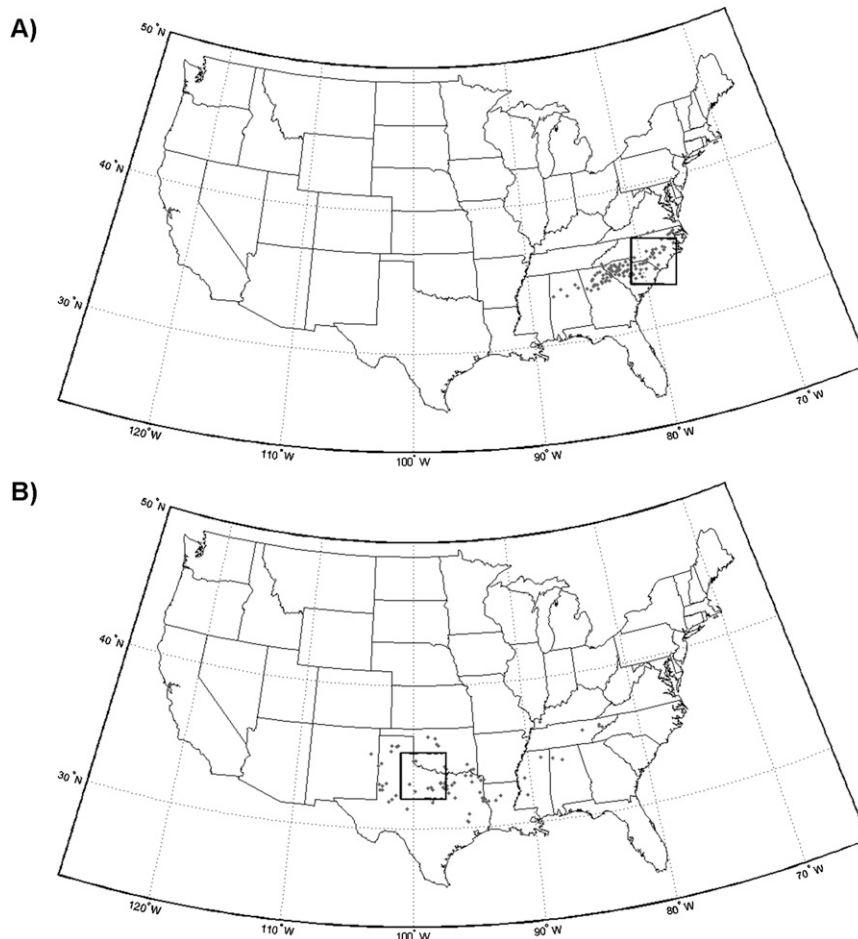


FIG. 10. Storm reports (gray dots) and the subdomain used for objective evaluation (black box) for the (a) 28 Mar 1984 TO and (b) 24 May 1986 PNO.

set of cases becomes a relative disadvantage when analyzing the revised set of cases, particularly in the 2- and 3-day simulations, confirming this parameter's sensitivity to seasonal influences.

4. Complementary subjective analysis

There are several ways to analyze the results subjectively to complement the objective analyses provided herein, including the relatively rigorous procedure developed by S09. Another suitable approach to this problem is to investigate which cases appear to be distinguished relatively well by the statistical models developed and which cases are not distinguished correctly most of the time. Analysis of these particular model simulations can provide insight into the parameters most suitable to distinguish outbreak type and possible sources of error in the prediction of these events (and relevant meteorological parameters).

As an example, the 1-day simulations for the revised set of cases will be used to illustrate this type of analysis.

Numerous meteorological parameters, including but not limited to those listed in Table 4, were tested individually and in combination to provide 95 sets of tests to determine the capability of the WRF to distinguish outbreak type. Because these tests are conducted so that each case is tested 15 times via cross validation, there were 1425 classifications for each case. A list of the correct and incorrect classifications for each case (Table 5) suggests that, as anticipated, a small number of cases were incorrectly classified a vast majority of the time, whereas most cases were classified correctly most of the time.

As indicated in Table 5, the 28 March 1984 TO (Fig. 10a) and 24 May 1986 PNO (Fig. 10b) were correctly classified a vast majority of the time. The WRF 1-day simulations of these outbreaks (Figs. 11 and 12) indicate correct discrimination should occur with these simulations. The 28 March 1984 TO (which occurred in Georgia, and North and South Carolina) is forecast to have very strong southwesterly midlevel flow ($>45 \text{ m s}^{-1}$ at 500 hPa), an intense surface cyclone ($<980 \text{ hPa}$ in middle

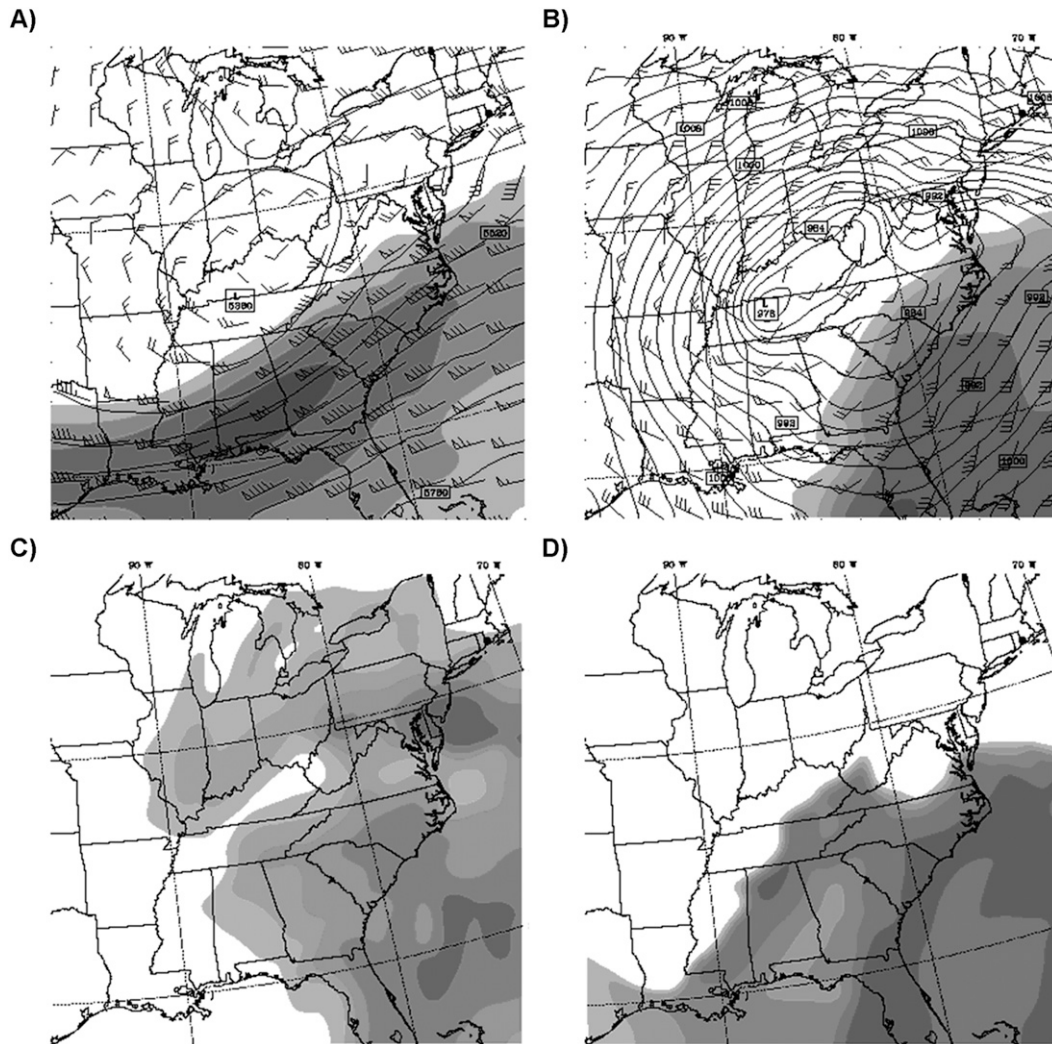


FIG. 11. WRF 1-day simulations of the (a) 500-hPa wind speeds (filled contours beginning at 25 m s^{-1} in 10 m s^{-1} increments), geopotential heights (contours in m), and winds (barbs in kt); (b) surface dewpoint temperature (filled contours beginning at 15°C in 2°C increments), MSLP (contours in hPa), and winds (barbs in kt); (c) 0–1-km SREH (filled contours beginning at $100 \text{ m}^2 \text{ s}^{-2}$ in increments of $100 \text{ m}^2 \text{ s}^{-2}$); and (d) LCL (filled contours beginning at 1500 m AGL in increments of -250 m AGL) for the 28 Mar 1984 TO.

Tennessee), very high 0–1-km SREH ($>400 \text{ m}^2 \text{ s}^{-2}$) in the eastern Carolinas, and low LCLs [$<1000 \text{ m}$ above ground level (AGL)] in the same region. However, the 24 May 1986 PNO (which occurred in much of Oklahoma and Texas) simulation features a very subtle shortwave trough with weak midlevel (500 hPa) flow in western Texas and Oklahoma, an approximately 1000-hPa surface cyclone in northeast Mexico and southwest Texas, relatively small values ($<100 \text{ m}^2 \text{ s}^{-2}$) of 0–1-km SREH, and LCLs generally above 1000 m AGL throughout the region. The subjective analysis indicates that TOs are distinguished from PNOs by forecasts of more intense synoptic-scale systems, higher low-level shear, and lower LCLs, in accord with expectations based on previous studies.

However, the 6 May 2003 TO (Fig. 13a) and the 10 April 1995 PNO (Fig. 13b) were classified incorrectly more than 50% of the time. The WRF 1-day simulations of these outbreaks (Figs. 14 and 15) explain why such misclassifications occurred with these cases. For the 6 May 2003 TO, a 997-hPa surface cyclone was forecast in northeast Oklahoma with 500-hPa southwesterly flow generally below 35 m s^{-1} . LCLs in the region were generally forecast above 1000 m AGL, and 0–1-km SREH was only locally high ($>300 \text{ m}^2 \text{ s}^{-2}$ in southeastern Missouri) and regionally below $200 \text{ m}^2 \text{ s}^{-2}$. With the 10 April 1995 PNO, very strong southerly to southwesterly 500-hPa flow was forecast ($>45 \text{ m s}^{-1}$) with a very strong ($\sim 991 \text{ hPa}$) surface cyclone forecast in northern

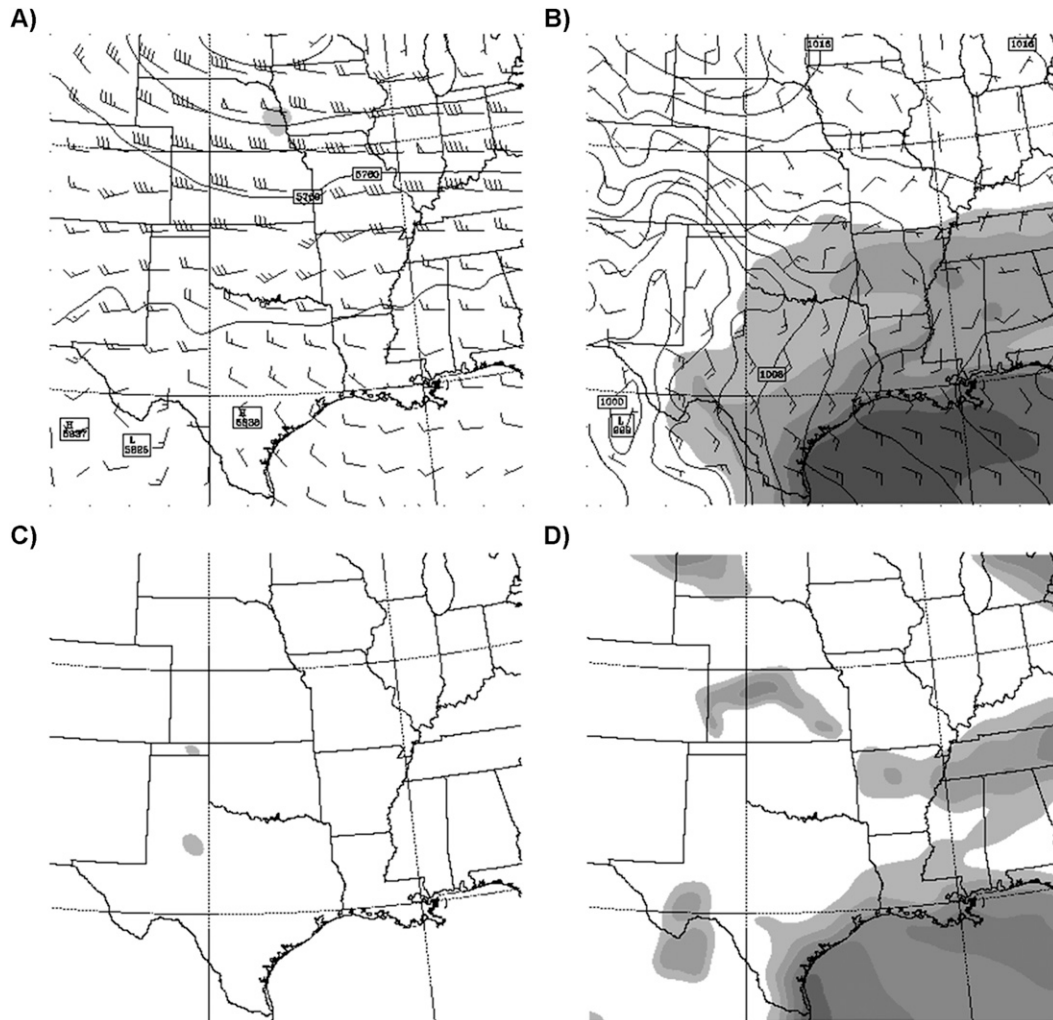


FIG. 12. As in Fig. 11, but for the 24 May 1986 PNO.

Oklahoma. LCLs were low (<1000 m AGL) in regions not affected by simulated convection occurring in Missouri and Arkansas, and 0–1-km SREH was very high (>300 m² s⁻²) in much of Arkansas and Louisiana. In general, cases more difficult to classify featured characteristics similar to those observed with these two cases. Possible sources of misclassification included spatial-temporal errors in the simulations (e.g., 17 April 1970, 27 May 1973, and 31 May 1985), substantial underprediction of strength of synoptic-scale systems for TOs (e.g., 7 May 1993 and 3 May 1999), overprediction of available instability (e.g., 6 April 2001), and convective modification. See S09 for more details.

5. Conclusions

S09 and M09 suggested that WRF simulations, using synoptic-scale input, were capable of distinguishing TOs

and PNOs up to three days in advance of the outbreak. However, subjective analysis of these cases (S09) indicated that PNOs occurring during the times of year in which TOs most often occurred were relatively difficult to distinguish. This study investigated the impacts of changing the PNO cases selected by S09 and M09 to cases that occurred exclusively during the period of time in which the sampled TOs were observed, to determine if outbreak classification was biased toward seasonal influences. WRF simulations for 1-, 2-, and 3-day simulations of TOs and PNOs were conducted for the set of cases used by S09 and M09 and for a revised set of PNO cases occurring outside of the time of year in which no sampled TOs were observed. Synoptic-scale data were initialized in the model, to determine the degree to which processes on these scales influence the occurrence or absence of tornado outbreaks. Because operational models cannot resolve tornadoes explicitly, analysis of

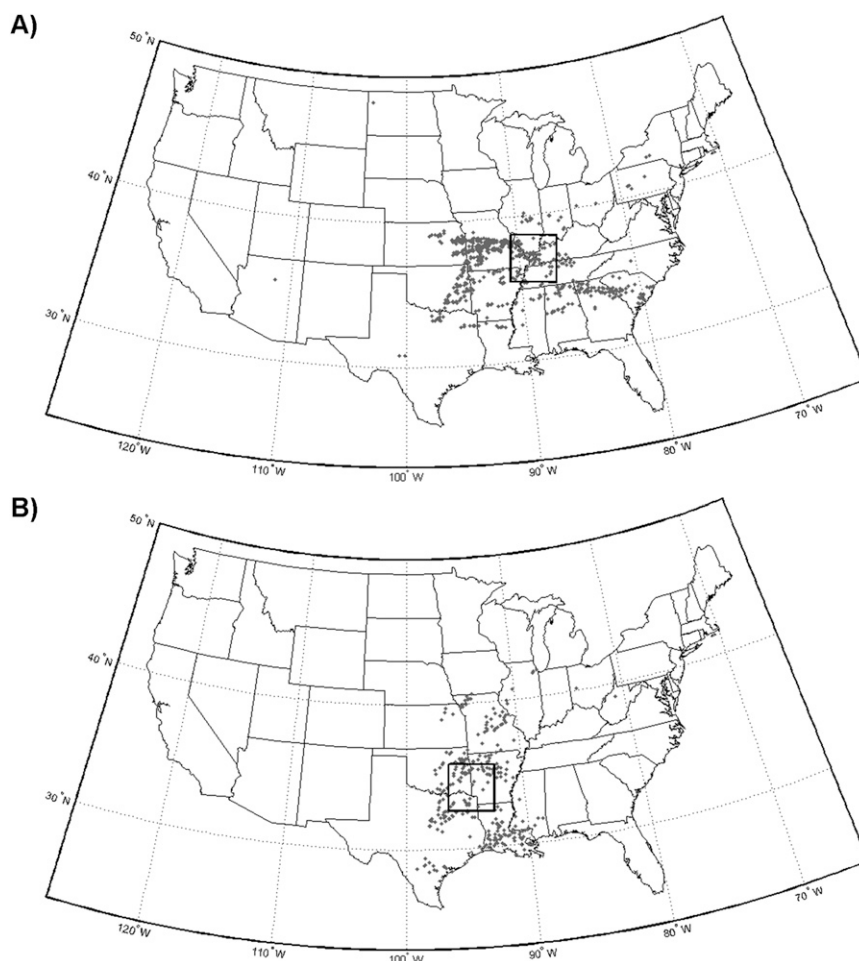


FIG. 13. As in Fig. 10, but for the (a) 6 May 2003 TO and (b) 10 Apr 1995 PNO.

the simulated mesoscale environments of the outbreaks using meteorological covariates was conducted. Numerous meteorological parameter fields were analyzed subjectively and objectively to compare the model's capability of distinguishing outbreak type using the two sets of cases, to pinpoint which parameters seemed to be most helpful in making correct discriminations, and to determine the degree to which these parameters were influenced by the modified sample of PNOs. In general, the WRF model simulations, when postprocessed through the algorithms presented herein, appeared capable of distinguishing outbreak type consistently when no time constraint was implemented on the outbreaks. The WRF's discrimination capability diminished but remained skillful when outbreaks were confined to days outside of the summer season. More specifically:

- Statistically significant degradation in skill was noted with most meteorological parameters when the PNOs were modified to include only those occurring outside

of the summer season. Accuracy generally dropped by at least 5%–10% and skill by 10%–15% for the most accurate and skillful parameters in distinguishing outbreak type. However, with accuracies greater than 70% and skill scores greater than 50% for 1-day simulations of TOs and PNOs outside of the summer season, synoptic-scale processes still appear pertinent in discriminating tornadic and primarily nontornadic outbreaks.

- Helicity parameters (0–1- and 0–3-km SREH) and synoptic parameters (such as low-level geopotential heights and MSLP) appeared to be most consistently helpful (highest accuracy and skill, in general) in distinguishing outbreak types.
- Thermodynamic instability parameters (e.g., SBCAPE and SBCIN) were not useful for discrimination.
- The LCL fields appeared to be especially susceptible to the time of year in which the outbreaks occurred. Overall, accuracy scores dropped from 15% to 20% when the PNOs were constrained to the same time of year as the TOs, and skill scores dropped by as much as 30%.

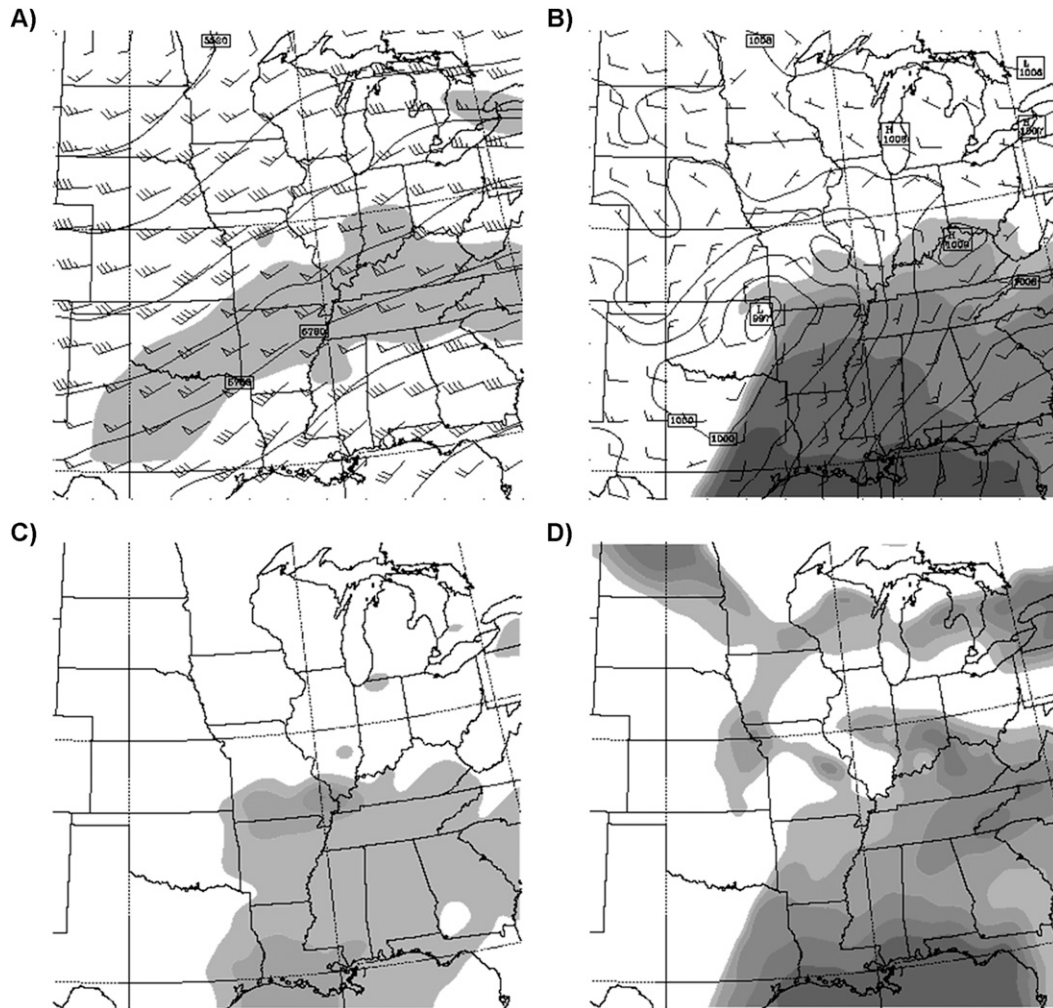


FIG. 14. As in Fig. 11, but for the 6 May 2003 TO.

- As observed in S09, the strength and spatial structure of synoptic-scale systems was observed to be crucial in distinguishing outbreak type for numerous cases. Miller's (1972) A, B, and D map types commonly favored TOs, whereas the C and E map types favored PNOs. Cases that tended to be misclassified often featured weaker (stronger) synoptic-scale systems for TOs (PNOs).
- Several index parameters (including EHI, SCP, and STP) exhibited comparable skill in distinguishing outbreak type to low-level storm-relative parameters and synoptic parameters. The value of indices is relatively higher with increasing forecast time.
- Forecast degradation was common from 1- to 3-day simulations. Whereas the most skillful parameters featured PSSs greater than 0.5 for 1-day simulations for the revised set of cases, few parameters scored values above 0.4 for 3-day simulations.

A key goal of this research is to implement the techniques developed by S09, M09, and this study in an operational setting; however, our study is diagnostic and should not be interpreted as a forecasting study. Four topic areas must be investigated before implementation in an operational environment can begin. The WRF simulations were analyzed where the outbreaks actually occurred, not where the model forecasted the outbreak to occur. Objective determination of the model's forecast outbreak location is difficult for a number of reasons. For example, it is possible that the model will not initiate deep convection for outbreak cases (see S09). Furthermore, severe convection does not always develop in the same location relative to particular synoptic-scale features (such as the location of the surface cyclone). An "objective" choice of model outbreak location will inevitably be subjective in nature because of these complications and may introduce biases into any type

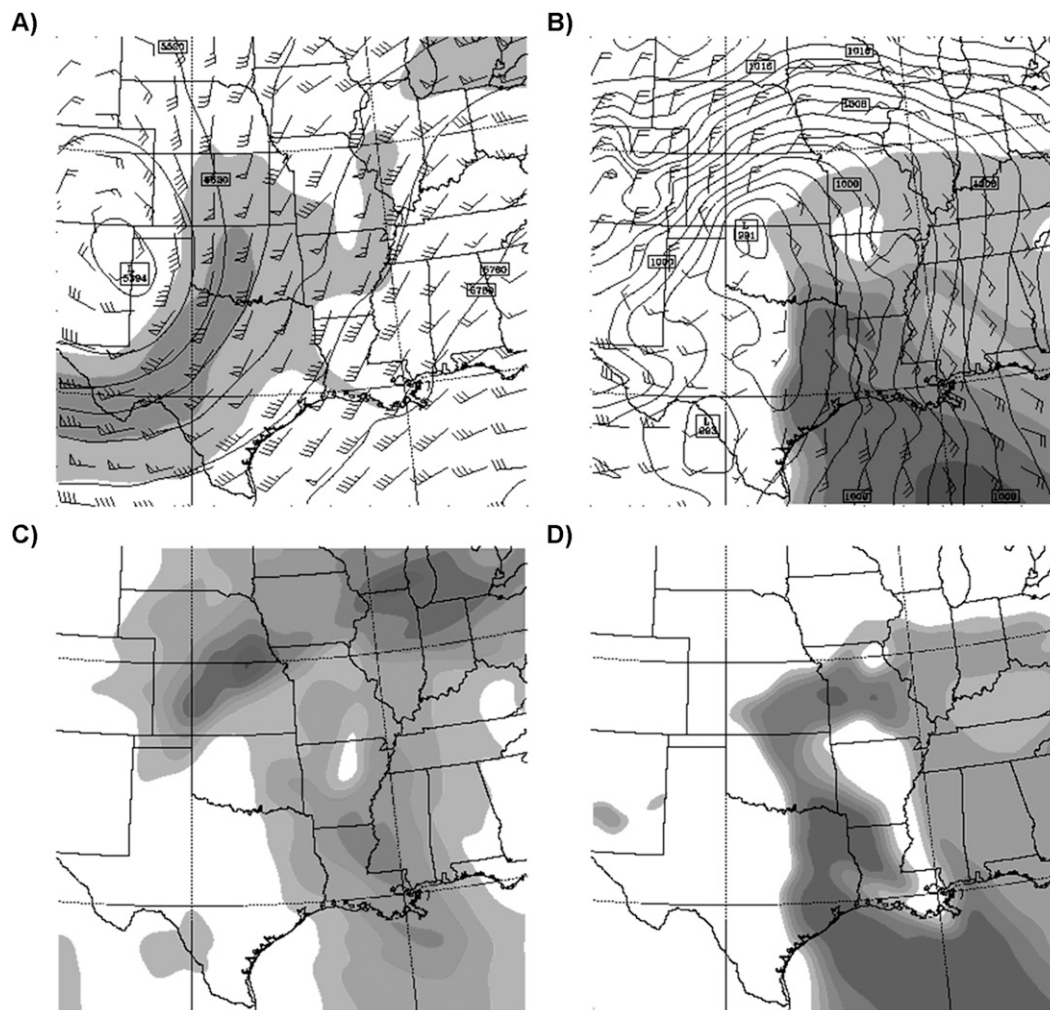


FIG. 15. As in Fig. 11, but for the 10 Apr 1995 PNO.

of objective classification. However, objective determination of a model simulation's outbreak location must be explored if the techniques developed in S09, M09, and in this study are to be utilized in an operational environment. Without this determination, it would be problematic to know the actual geographical center of a simulated outbreak. Of course, a simulated outbreak might not coincide in space and time with any actual severe weather outbreak that might occur.

Another complication is the “outbreak–no outbreak” question. This study only considers cases for which it is known a priori that an outbreak occurred. The determination of whether a particular synoptic-scale system will produce a severe weather outbreak is beyond the scope of this study. There is no guarantee, for example, that an outbreak will occur given a particular synoptic-scale system, even from those that appear “synoptically evident” (Doswell et al. 1993; Doswell and

Bosart 2001). Research investigating the differences in synoptically-evident environments that produce outbreaks versus those that do not would greatly benefit the severe weather forecasting community.

A recent study investigating the utility of SVMs in discriminating tornadic and nontornadic storms using radar data (Trafalis et al. 2009, manuscript submitted to *J. Commun. Network Syst. Sci.*) suggests the tacit assumption that a balanced training ratio (i.e., 50% of cases are tornadic and 50% are nontornadic) does not necessarily provide the best results when applying the statistical models to an independent testing set. Accordingly, there is a need to conduct experiments with an incremental range of percentages of tornadic storms (Adrianto et al. 2006). Analysis of the capability of the WRF model, using SVMs or other postprocessing algorithms, to distinguish TOs from PNOs should be conducted in such a manner for the two types of outbreaks to

determine if improved discrimination can be attained using unbalanced training sets.

Finally, most outbreaks of severe weather fall “between” the TO and PNO classifications. In order for forecasting to benefit from the research presented herein, accounting for and properly discriminating these events from TOs and PNOs will be necessary. Furthermore, some synoptic-scale systems may be forecast to produce an outbreak but do not do so. Investigating these cases will be necessary to develop a set of null events. Currently, research is ongoing to account for these cases.

Acknowledgments. This research was supported by NSF Grants ATM-0527934 and ATM-050867, and by Award NA06OASR4320264 06111039 to the Northern Gulf Institute by NOAA’s Office of Oceanic and Atmospheric Research, U. S. Department of Commerce. The University of Oklahoma Supercomputing Center for Education and Research (OSCER) administrators provided the computational facilities used for this project. All model simulations were performed and all model output was stored on the University of Oklahoma supercomputers known as Topdawg and Sooner (more information is available online at <http://www.oscer.ou.edu>). We also thank the three anonymous reviewers for their helpful suggestions.

REFERENCES

- Adrianto, I., M. B. Richman, J. Park, T. B. Trafalis, and S. Lakshmivarahan, 2006: Machine classifiers for tornado detection: Sensitivity analysis on tornado data sets. *Intell. Eng. Syst. Artificial Neural Networks*, **16**, 679–684.
- Blanchard, D. O., 1998: Assessing the vertical distribution of convective available potential energy. *Wea. Forecasting*, **13**, 870–877.
- Brooks, H. E., C. A. Doswell III, and J. Cooper, 1994a: On the environments of tornadic and nontornadic mesocyclones. *Wea. Forecasting*, **9**, 606–618.
- , —, and R. B. Wilhelmson, 1994b: On the role of mid-tropospheric winds in the evolution and maintenance of low-level mesocyclones. *Mon. Wea. Rev.*, **122**, 126–136.
- , —, 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 distributions of severe thunderstorm and tornado environments from global reanalysis data. *Atmos. Res.*, **67–68**, 73–94.
- Brown, B. G., and A. H. Murphy, 1996: Verification of aircraft icing forecasts: The use of standard measures and meteorological covariates. Preprints, *13th Conf. on Probability and Statistics in the Atmospheric Sciences*, San Francisco, CA, Amer. Meteor. Soc., 251–252.
- Bunkers, M. J., 2002: Vertical wind shear associated with left-moving supercells. *Wea. Forecasting*, **17**, 845–855.
- Colby, F. P., Jr., 1984: Convective inhibition as a predictor of convection during AVE-SESAME II. *Mon. Wea. Rev.*, **112**, 2239–2252.
- Craven, J. P., H. E. Brooks, and J. A. Hart, 2002a: Baseline climatology of sounding derived parameters associated with deep, moist convection. Preprints, *21st Conf. on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., 643–646.
- , R. E. Jewell, and H. E. Brooks, 2002b: Comparison between observed convective cloud-base heights and lifting condensation level for two different lifted parcels. *Wea. Forecasting*, **17**, 885–890.
- Cristianini, N., and J. Shawe-Taylor, 2000: *Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press, 189 pp.
- Davies, J. M., 1993: Hourly helicity, instability, and EHI in forecasting supercell tornadoes. Preprints, *17th Conf. on Severe Local Storms*, St. Louis, MO, Amer. Meteor. Soc., 107–111.
- , and R. Johns, 1993: Some wind and instability parameters associated with strong and violent tornadoes. Part I: Wind shear and helicity. *The Tornado: Its Structure, Dynamics, Prediction and Hazards, Geophys. Monogr.*, Vol. 79, Amer. Geophys. Union, 573–582.
- Davies-Jones, R., D. Burgess, and M. Foster, 1990: Test of helicity as a tornado forecast parameter. Preprints, *16th Conf. on Severe Local Storms*, Kananaskis Park, AB, Canada, Amer. Meteor. Soc., 588–592.
- Doolittle, M. H., 1888: Association ratios. *Bull. Philos. Soc. Washington*, **7**, 122–127.
- Doswell, C. A., III, 2007: Small sample size and data quality issues illustrated using tornado occurrence data. *Electron. J. Severe Storms Meteor.*, **2** (5), 1–16.
- , and L. F. Bosart, 2001: Extratropical synoptic-scale processes and severe convection. *Severe Convective Storms, Meteor. Monogr.*, No. 27, Amer. Meteor. Soc., 1–27.
- , and J. S. Evans, 2003: Proximity sounding analysis for derechos and supercells: An assessment of similarities and differences. *Atmos. Res.*, **67–68**, 117–133.
- , and D. M. Schultz, 2006: On the use of indices and parameters in forecasting severe storms. *Electron. J. Severe Storms Meteor.*, **1** (3), 1–14.
- , R. H. Johns, and S. J. Weiss, 1993: Tornado forecasting: A review. *The Tornado: Its Structure, Dynamics, Prediction and Hazards, Geophys. Monogr.*, Vol. 79, Amer. Geophys. Union, 557–571.
- , H. E. Brooks, and M. P. Kay, 2005: Climatological estimates of daily local nontornadic severe thunderstorm probability for the United States. *Wea. Forecasting*, **20**, 577–595.
- , R. Edwards, R. L. Thompson, J. A. Hart, and K. C. Crosbie, 2006: A simple and flexible method for ranking severe weather events. *Wea. Forecasting*, **21**, 939–951.
- Droegemeier, K. K., S. M. Lazarus, and R. Davies-Jones, 1993: The influence of helicity on numerically simulated convective storms. *Mon. Wea. Rev.*, **121**, 2005–2029.
- Dudhia, J., 1989: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107.
- Efron, B., and R. J. Tibshirani, 1993: *An Introduction to the Bootstrap*. Chapman and Hall/CRC, 436 pp.
- Grell, G. A., and D. Devenyi, 2002: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophys. Res. Lett.*, **29**, 1693, doi:10.1029/2002GL015311.
- Hart, J. A., 1993: SVR/PLOT: A new method of accessing and manipulating the NSSFC Severe Weather Database. Preprints, *17th Conf. on Severe Local Storms*, St. Louis, MO, Amer. Meteor. Soc., 40–41.

- , and W. Korotky, 1991: The SHARP workstation v1.50 users guide. NOAA/National Weather Service, 30 pp. [Available from NWS Eastern Region Headquarters, 630 Johnson Ave., Bohemia, NY 11716.]
- Haykin, S., 1999: *Neural Networks: A Comprehensive Foundation*. Pearson Education, 842 pp.
- Holton, J. R., 1992: *An Introduction to Dynamic Meteorology*. 3rd ed. Academic Press, 511 pp.
- Hong, S.-Y., and H.-L. Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon. Wea. Rev.*, **124**, 2322–2339.
- , H.-M. Juang, and Q. Zhao, 1998: Implementation of prognostic cloud scheme for a regional spectral model. *Mon. Wea. Rev.*, **126**, 2621–2639.
- Johns, R. H., and C. A. Doswell III, 1992: Severe local storms forecasting. *Wea. Forecasting*, **7**, 588–612.
- , J. Davies, and P. Leftwich, 1993: Some wind and instability parameters associated with strong and violent tornadoes. Part II: Variations in the combinations of wind and instability parameters. *The Tornado: Its Structure, Dynamics, Prediction and Hazards, Geophys. Monogr.*, Vol. 79, Amer. Geophys. Union, 583–590.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Kerr, B. W., and G. L. Darkow, 1996: Storm-relative winds and helicity in the tornadic thunderstorm environment. *Wea. Forecasting*, **11**, 489–505.
- Lee, Y., G. Wahba, and S. A. Ackerman, 2004: Cloud classification of satellite radiance data by multicategory support vector machines. *J. Atmos. Oceanic Technol.*, **21**, 159–169.
- Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Climate Appl. Meteor.*, **22**, 1065–1092.
- Maddox, R. A., 1976: An evaluation of tornado proximity wind and stability data. *Mon. Wea. Rev.*, **104**, 133–142.
- Markowski, P. M., C. Hannon, J. Frame, E. Lancaster, A. Pietrycha, R. Edwards, and R. L. Thompson, 2003: Characteristics of vertical wind profiles near supercells obtained from the Rapid Update Cycle. *Wea. Forecasting*, **18**, 1262–1272.
- Mercer, A. E., and M. B. Richman, 2007: Statistical differences of quasigeostrophic variables, stability, and moisture profiles in North American storm tracks. *Mon. Wea. Rev.*, **135**, 2312–2338.
- , —, H. B. Bluestein, and J. M. Brown, 2008: Statistical modeling of downslope windstorms in Boulder, Colorado. *Wea. Forecasting*, **23**, 1176–1194.
- , C. M. Shafer, C. A. Doswell III, L. M. Leslie, and M. B. Richman, 2009: Objective classification of tornadic and nontornadic outbreaks. *Mon. Wea. Rev.*, **137**, 4355–4368.
- Miller, R., 1972: Notes on analysis and severe-storm forecasting procedures of the Air Force Global Weather Center. Air Weather Service Tech. Rep. 200 (rev.), Air Weather Service, Scott Air Force Base, IL, 190 pp. [Available online at <http://chubasco.niu.edu/projects/miller/>]
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102** (D14), 16 663–16 682.
- Moller, A. R., 2001: Severe local storms forecasting. *Severe Convective Storms, Meteor. Monogr.*, No. 50, Amer. Meteor. Soc., 433–480.
- Monteverdi, J. P., C. A. Doswell III, and G. S. Lipari, 2003: Shear parameter thresholds for forecasting tornadic thunderstorms in northern and central California. *Wea. Forecasting*, **18**, 357–370.
- Orlanski, I., 1975: A rational subdivision of scales for atmospheric processes. *Bull. Amer. Meteor. Soc.*, **56**, 527–530.
- Peirce, C. S., 1884: The numerical measure of success of predictions. *Science*, **4**, 453–454.
- Rasmussen, E. N., and D. O. Blanchard, 1998: A baseline climatology of sounding-derived supercell and tornado forecast parameters. *Wea. Forecasting*, **13**, 1148–1164.
- Richman, M. B., 1986: Rotation of principal components. *J. Climatol.*, **6**, 293–335.
- , B. Santosa, and T. B. Trafalis, 2005: Feature selection of radar-derived tornado attributes with support vector machines. Preprints, *Fourth Conf. on Artificial Intelligence Applications to Environmental Sciences*, San Diego, CA, Amer. Meteor. Soc., J5.1. [Available online at <http://ams.confex.com/ams/pdfpapers/87991.pdf>.]
- Shafer, C. M., A. E. Mercer, C. A. Doswell III, M. B. Richman, and L. M. Leslie, 2009: Evaluation of WRF forecasts of tornadic and nontornadic outbreaks when initialized with synoptic-scale input. *Mon. Wea. Rev.*, **137**, 1250–1271.
- Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, 2005: A description of the Advanced Research WRF Version 2. NCAR Tech. Note NCAR/TN-468+STR, 88 pp. [Available from UCAR Communications, P.O. Box 3000, Boulder, CO 80307.]
- , and Coauthors, 2008: A description of the Advanced Research WRF Version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp. [Available from UCAR Communications, P.O. Box 3000, Boulder, CO 80307.]
- Stensrud, D. J., and J. M. Fritsch, 1994: Mesoscale convective systems in weakly forced large-scale environments. Part III: Numerical simulations and implications for operational forecasting. *Mon. Wea. Rev.*, **122**, 2084–2104.
- , J. V. Cortinas Jr., and H. E. Brooks, 1997: Discriminating between tornadic and nontornadic thunderstorms using mesoscale model output. *Wea. Forecasting*, **12**, 613–632.
- Thompson, R. L., R. Edwards, J. A. Hart, K. L. Elmore, and P. Markowski, 2003: Close proximity soundings with supercell environments obtained from the Rapid Update Cycle. *Wea. Forecasting*, **18**, 1243–1261.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.
- Woollen, J. S., 1991: New NMC operational OI quality control. Preprints, *Ninth Conf. on Numerical Weather Prediction*, Denver, CO, Amer. Meteor. Soc., 24–27.
- , E. Kalnay, L. Gandin, W. Collins, S. Saba, R. Kistler, M. Kanamitsu, and M. Chelliah, 1994: Quality control in the reanalysis system. Preprints, *10th Conf. on Numerical Weather Prediction*, Portland, OR, Amer. Meteor. Soc., 13–14.