4A.5 ASSOCIATING MAGNITUDES OF SEVERE WEATHER DIAGNOSTIC VARIABLES WITH SEVERE EVENT PROBABILITIES

Chad M. Shafer[1*], McKenna W. Stanford[1], Michael B. Richman[2,3], Lance M. Leslie[2,3], Charles A. Doswell III[3], and Andrew E. Mercer[4]

[1]University of South Alabama, Mobile, Alabama
[2]University of Oklahoma, Norman, Oklahoma
[3]Cooperative Institute for Mesoscale Meteorological Studies, Norman, Oklahoma
[4]Mississippi State University, Starkville, Mississippi

1. INTRODUCTION

Evaluation of severe weather diagnostic variables (SWDVs) has been conducted in numerous studies in the past few decades, to determine the accuracy and skill with which these variables identify and/or distinguish severe weather environments. Most of these studies have focused on discrimination of tornadic and nontornadic supercells (e.g., Rasmussen and Blanchard 1998; Rasmussen 2003; Thompson et al. 2003), convective mode (e.g., Doswell and Evans 2003; Dial et al. 2010), or outbreak type (e.g., Mercer et al. 2009; Shafer et al. 2010a,b; 2012). Several SWDVs have been proposed to distinguish these environments, including the bulk wind differential, convective available potential energy (CAPE), storm-relative environmental helicity (SREH; Davies-Jones et al. 1990), energy-helicity index (EHI; Hart and Korotky 1991), supercell composite parameter (SCP; Thompson et al. 2003), and significant tornado parameter (STP; Thompson et al. 2003), among many others.

These severe weather discrimination studies necessarily have used conditional datasets, owing to the difficulties in identifying meaningful null cases, classifying severe weather phenomena, archiving severe weather information, etc. Moreover, many severe weather discrimination studies do not assess the uncertainty associated with such diagnoses. As discussed by Doswell and Schultz (2006), this has led to the use of many SWDVs despite an incomplete assessment of these variables, particularly the uncertainty associated with their usage, and an unclear interpretation of their magnitudes, particularly those variables that have no obvious physical meaning (so-called indices, such as EHI, SCP, and STP).

This study proposes a method to evaluate SWDVs in the discrimination of severe weather by determining the skill with which preselected SWDV magnitudes identify regions with threshold probabilities of severe weather of a specific type or of any type. The probabilities that severe weather occurs at a given point are determined using the practically perfect (PP) technique introduced by Brooks et al. (1998). The PP technique attempts to account for the spatial and temporal errors and uncertainties associated with the observations of severe weather and the expectation of false alarms and misses of, for example, SPC convective outlooks and watches and NWS warnings. Additionally, this study attempts to quantify the sensitivity of these diagnoses to the time period of analysis and to various qualifying criteria for event consideration. The specifics regarding these sensitivities are discussed in the following section.

2. DATA AND METHODS

Each day from 1 January 2001 to 31 December 2010 is considered for evaluation of the SWDVs. North American Regional Reanalysis (NARR; Mesinger et al. 2006) data are used to determine the SWDV magnitudes for each grid point in a 40-km x 40-km Lambert conformal domain encompassing the conterminous United States. Only grid points in which severe weather was observed from the 1979–2010 period were assessed. Each SWDV

**Corresponding author address:* *Chad M. Shafer, Univ. of South Alabama, Dept. of Earth Sciences, 5871 USA Drive North Room 136, Mobile, AL, 36688-0002; e-mail cmshafer@usouthal.edu*

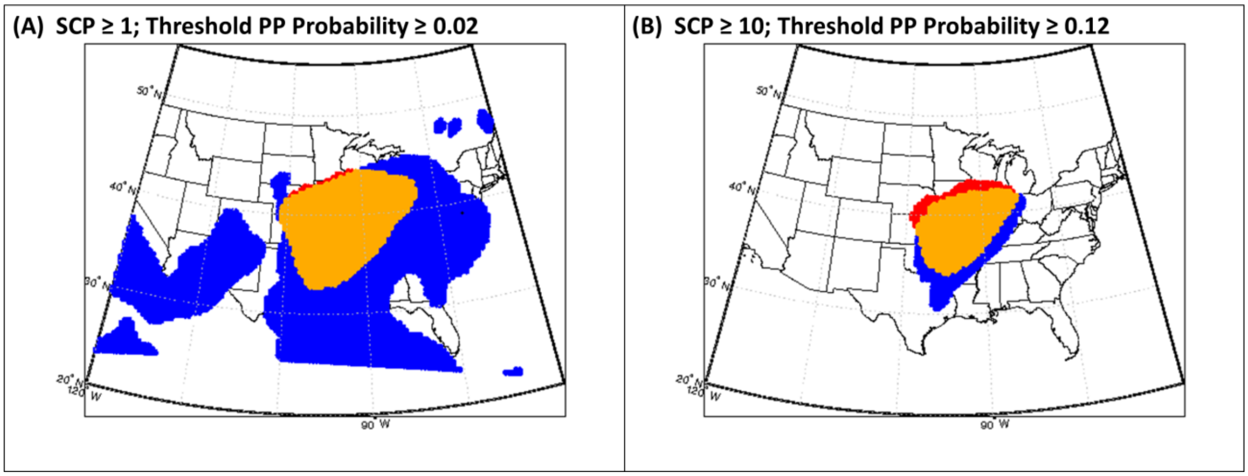

Figure 1: Grid points in which the SCP exceeds a preselected threshold and the threshold PP probability is not exceeded (blue), the SCP does not exceed a preselected threshold and the threshold PP probability is exceeded (red), and both thresholds are exceeded (orange) for the 12 March 2006 tornado outbreak. SCP and PP probability thresholds are (a) 1 and 0.02 and (b) 10 and 0.12.

is evaluated four times: (1) for each grid point on each day in the 10-y period (hereafter, *unconditional*), (2) for each grid point on days in which the type of severe weather being assessed is observed (hereafter, *conditional*), (3) for each grid point in which convection is observed on each day in the 10-y period (hereafter, *unconditional – convection required*), and (4) for each grid point in which convection is observed on days in which the type of severe weather being assessed is observed (hereafter, *conditional – convection required*). For this study, four SWDVs are evaluated: 0-1 km EHI, 0-3 km EHI, SCP, and STP.

The PP technique uses a nonparametric density function with a Gaussian kernel. This technique was conducted seven times: for any type of severe report, for any severe wind, hail, or tornado report, and for any significant wind (65+ kts), significant hail (5+ cm diameter), and significant tornado [(E)F2+] report. Reports are collected from the Storm Prediction Center severe weather database (Schaeffer and Edwards 1999). The kernel density function is as follows:

$$f = \sum_{i=1}^{N} \frac{1}{2\pi\sigma^2} \exp\left[-\frac{1}{2}\left(\frac{d_i}{\sigma}\right)^2\right] \quad (1),$$

where d_i is the distance of the *i*th report to the grid point of interest, N is the total number of reports in the given time period, and σ is the bandwidth. The bandwidth is associated with the confidence in the location of the report (see Brooks et al. 1998), which in this study is selected to be 120 km. This selection is consistent with past studies (Brooks et al. 1998; Brooks et al. 2003; Doswell et al. 2005; Shafer and Doswell 2011), where attempts are made to mimic characteristic regions outlined in SPC convective outlooks. The approximate probability density function (PDF) in (1) is converted to a probability of at least one report observed in the grid point of interest.

Preselected SWDV magnitudes and threshold PP probabilities are compared (see examples in Fig. 1). For points in which both values are exceeded (neither value is exceeded), the grid point is considered a hit (correct null). For points in which the threshold probability is exceeded but the SWDV magnitude is not, the grid point is considered a miss. For points in which the SWDV magnitude is exceeded but the threshold PP probability is not, the grid point is considered a false alarm. Binary contingency statistics (Wilks 2006) are computed based on these categorizations. *The PP probability for which the maximum Heidke skill score (HSS) is observed for a given magnitude of the evaluated SWDV is the probability assigned to that SWDV magnitude.*

Finally, the methods discussed above are evaluated for 24-, 12-, 6-, 3-, and 1-h time periods. For the 6+-h time periods, the maximum SWDV magnitude is used for evaluation; otherwise, the initial SWDV magnitude is used for evaluation. In summary, this study is attempting to quantify the sensitivities of the PP probabilities for a given SWDV magnitude to (1) the choice of SWDV, (2) the length of time of the diagnosis, (3) the type of severe weather, and (4) the predetermined qualifying conditions required for a grid point's evaluation. Other sensitivities not tested here include the selection of bandwidth and the grid spacing of the domain, which are currently being investigated (Section 4).

3. RESULTS

3.1 *Statistics for Severe Report Variables*

The maximum skill scores and the threshold PP probabilities unsurprisingly are sensitive to the severe weather report type being evaluated (Figs. 2 and 3). Using SCP as an example, the 12-h assessment indicates the threshold PP probabilities for which the maximum HSS was observed exceed 60% for SCP magnitudes near 50 for any report type, are approximately 50% for hail and wind reports, and range from 10-20% for tornadoes and all significant severe weather types. Thus, as expected, the probabilities for the rarest events are lowest. However, the maximum skill scores are highest for the rarest types of events for the largest magnitudes of SCP (Figs. 2c,d). Additionally, HSSs appear to be higher for hail reports than for wind reports, which is likely an indication that SCP is identifying supercell environments (which have a propensity for producing significant hail and tornadoes) versus linear convection (which can produce a relatively large number of significant wind reports).

For threshold PP probabilities as a function of SCP magnitude, there is very little difference between the unconditional and conditional data sets (cf. Figs. 2a,b). For maximum HSSs as a function of SCP magnitude, on the other hand, there are considerable differences. Notably, the maximum HSSs are higher and occur at lower SCP magnitudes for the conditional versus the unconditional data set (cf. Figs. 2c,d). That is, if it is known that severe weather of a particular type will occur on a given day, the identification of locations exceeding threshold PP probabilities of that report type is done with increased skill. The tendency for higher maximum HSSs at lower SCP magnitudes suggests that the largest improvement in HSS occurs at the lowest threshold PP probabilities, which is confirmed when comparing HSSs to threshold PP probabilities (cf. Figs. 2e,f). This is indicative of decreasing the number of false alarms by eliminating days in which severe weather of the relevant type is not observed.

The 3-h evaluation indicates very similar trends to that of the 12-h evaluation (see Fig. 3). The threshold PP probabilities and the maximum HSSs are slightly lower for the 3-h evaluation than for the 12-h evaluation (cf. Figs. 2a-d; 3a-d). The relatively high HSSs of the tornado and significant tornado reports to other types of severe weather are increasingly pronounced for 3-h forecasts. For the 12-h and 3-h evaluations, differences in the unconditional versus conditional data sets show up as increased threshold PP probabilities for high SCP magnitudes and increased maximum HSSs for low SCP magnitudes. Finally, all of the aforementioned trends as a function of severe weather report type are observed for EHI and STP as well (not shown).

3.2 *Statistics for Varying Time Periods*

The maximum skill scores and the threshold PP probabilities are also strongly sensitive to the time period of assessment (Figs. 4 and 5). The likelihood of an event increases with an increased time window. Thus, as anticipated, the threshold PP probabilities and maximum HSSs are highest for 24-h diagnoses and decrease considerably for 1-h diagnoses. Maximum HSSs decrease substantially for a given PP probability from the 24-h to 1-h time periods.

Comparing the assessment of any severe report type to that for tornadoes, the threshold PP probabilities are considerably lower, the maximum HSSs are somewhat lower, and the differences between the conditional and unconditional data sets are more pronounced for the tornadoes than when including any type of severe report. Other trends include the increase in 0-3 km EHI magnitude for which the maximum HSS is observed for decreasing time window (see Figs. 5c,d), particularly for the unconditional data set (where false alarms are more prevalent) and the narrowing of the range of probabilities and maximum HSSs for the conditional dataset compared to the unconditional dataset. Specifically, for tornado reports, maximum HSSs decrease by more than 0.15 from 24-h to 1-h diagnoses for the unconditional dataset (Fig. 5c)

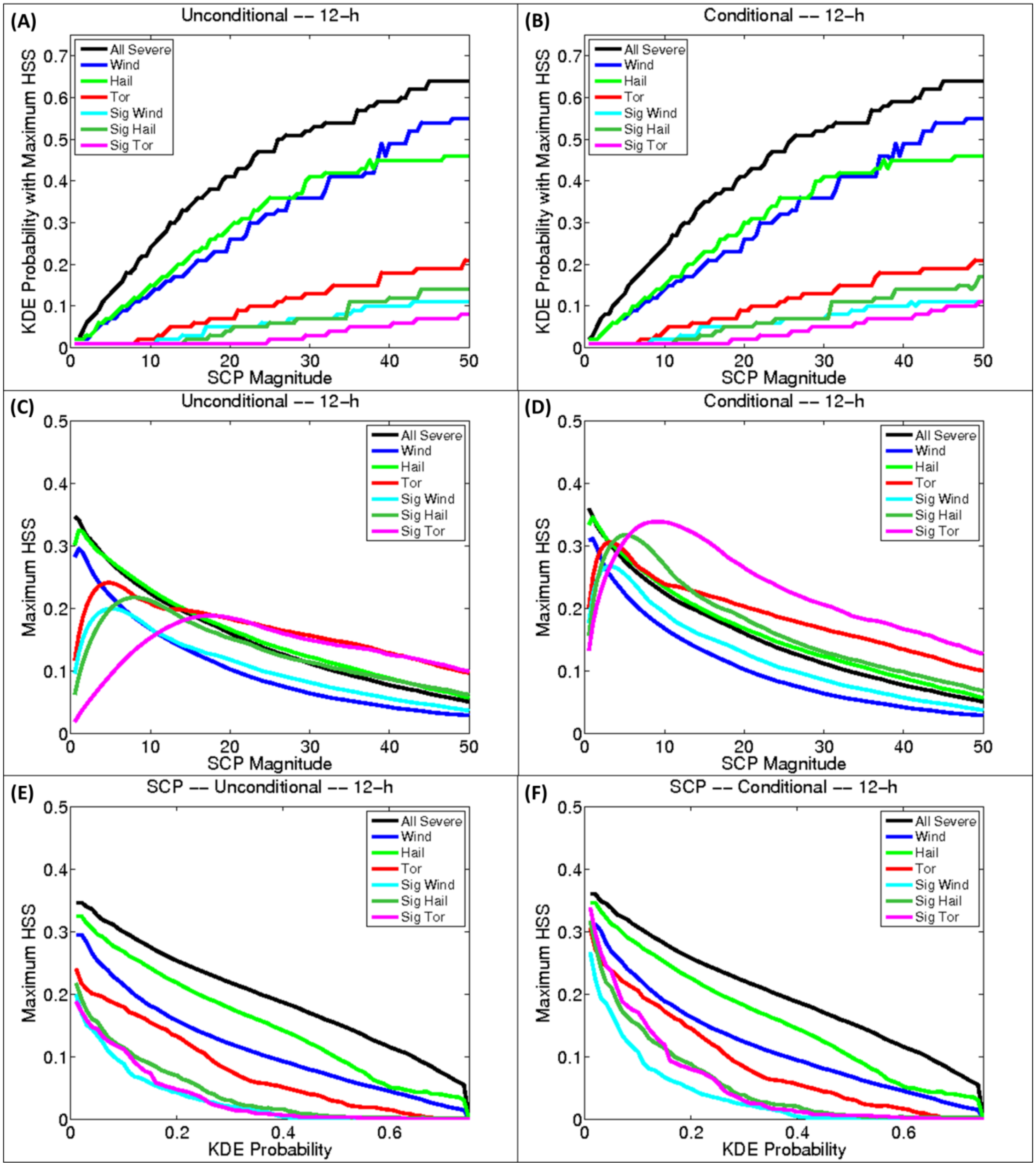

Figure 2: (a) Threshold 12-h PP probabilities determined using the maximum HSS for a given magnitude of SCP for all days from 1 January 2001 to 31 December 2010 for each type of severe report (legend). (b) As in (a), for all days from 1 January 2001 to 31 December 2010 in which the report type evaluated is observed. (c)-(d) As in (a)-(b), except the 12-h maximum HSSs as a function of SCP magnitude are shown. (e)-(f) As in (a)-(b), except the 12-h maximum HSSs as a function of the threshold PP probabilities (for SCP) are shown.

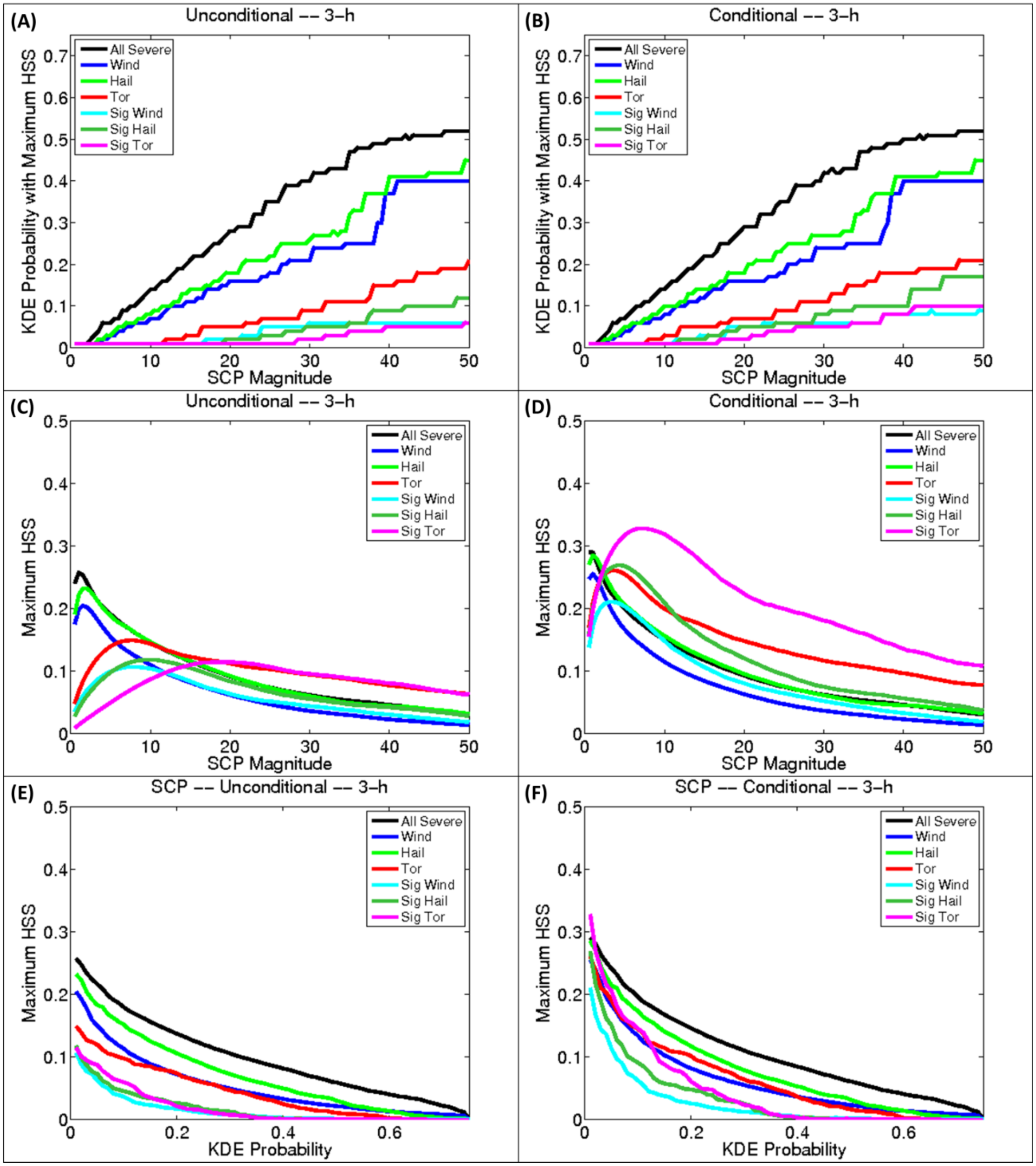

Figure 3: As in Fig. 2, for 3-h diagnoses.

and by less than 0.1 for the conditional dataset (Fig. 5d). Additionally, the most substantial improvement in the HSSs when comparing the conditional and unconditional datasets exists with the smallest time periods for assessment (i.e., 1-h diagnoses). Whereas the HSSs increase by < 0.05 for 24-h conditional diagnoses of tornadoes, the HSSs increase by > 0.10 for 1-h conditional diagnoses of tornadoes. When evaluating all types of severe weather, there appears to be a minor decrease in HSS when using the conditional dataset for 24-h diagnoses.

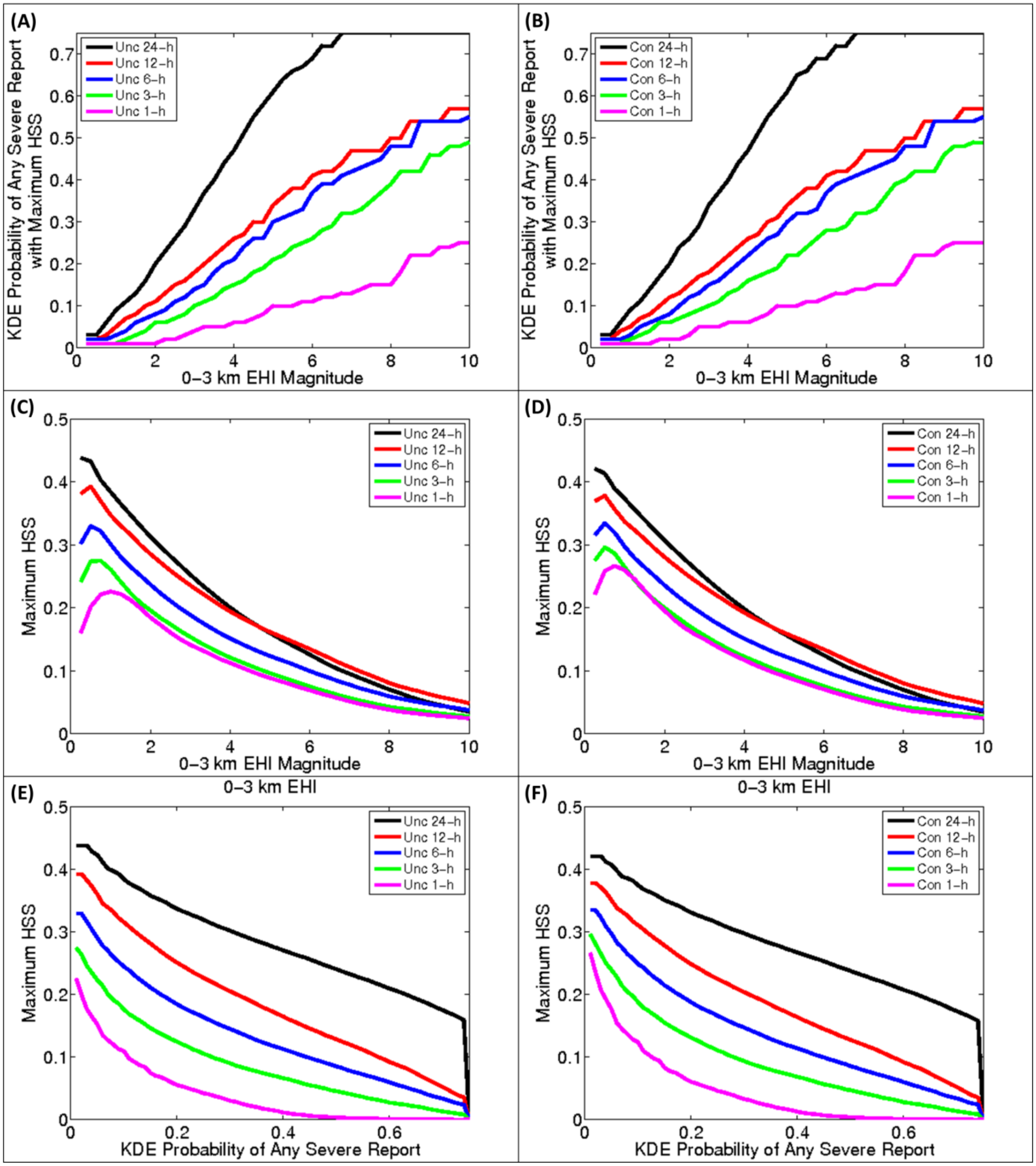

Figure 4: As in Fig. 2, except for differing time periods (legend) using 0-3 km EHI for any type of severe report.

3.3 *Statistics for Varying SWDVs*

Shafer et al. (2012) reported the skill with which SWDVs could diagnose the severity of convective outbreaks using areal coverage. In that study, they compared the areal coverage magnitudes of several SWDVs and noted strong (>0.8) correlations among SCP, STP, and EHI. The technique presented herein enables easy comparison among these SWDVs of the probabilities (and skill) of severe weather occurring in proximity to a given point. When

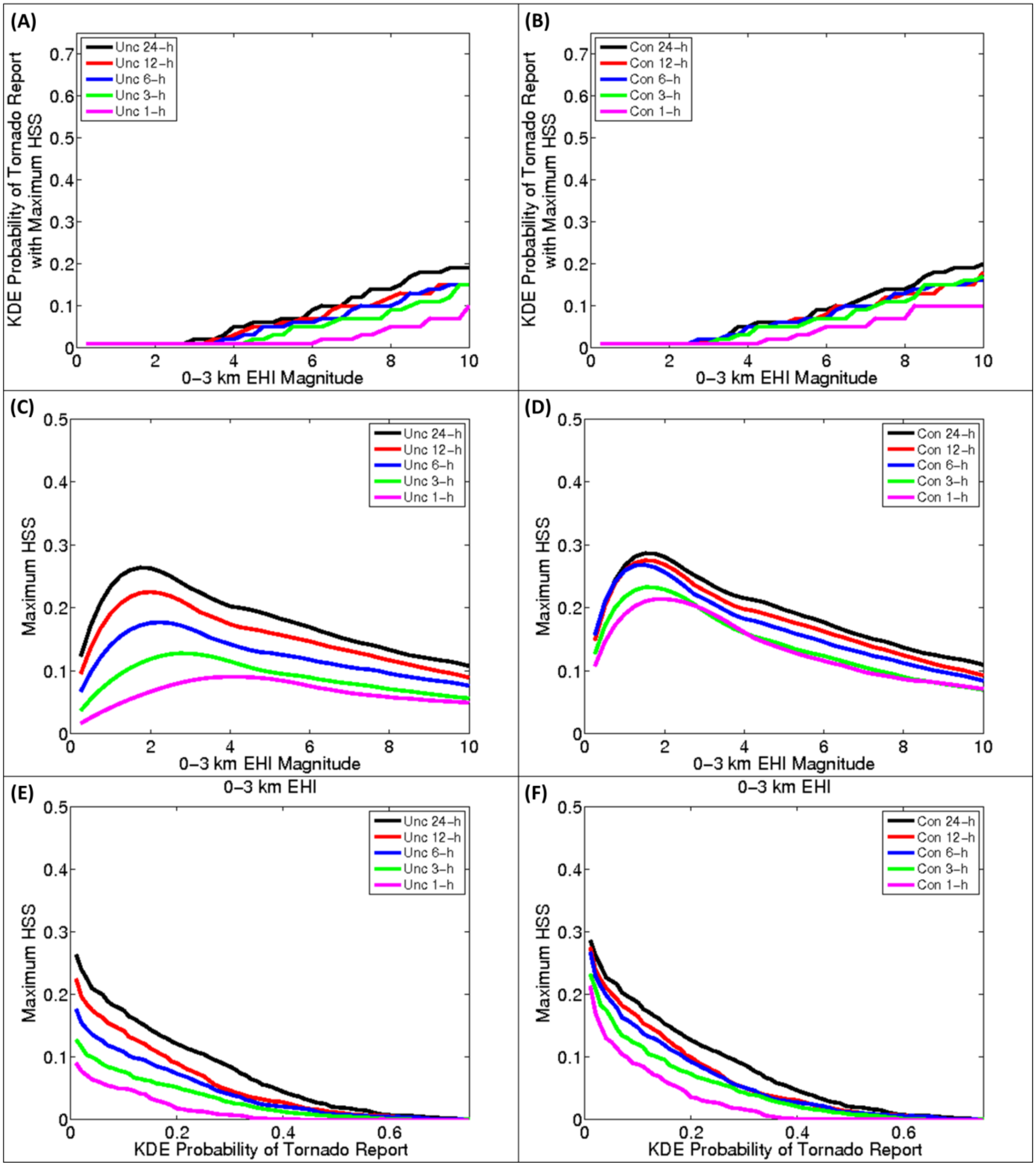

Figure 5: As in Fig. 4, except evaluating tornado reports only.

comparing 6-h diagnoses for any severe report (Fig. 6), differences in HSSs for PP probabilities among the SWDVs increase for decreasing probabilities. The *relative* skills of SCP and STP are highest for high PP probabilities, whereas 0-1 km EHI and 0-3 km EHI exhibit relatively high skills for low PP probabilities. The *absolute* skill is highest for SCP and 0-3 km EHI, in general. Differences in skill scores among the variables generally are < 0.05 for PP probabilities > 0.4, and are somewhat smaller for the conditional versus unconditional datasets. The relative value of EHI increases when convection is required for a grid point to be considered (cf.

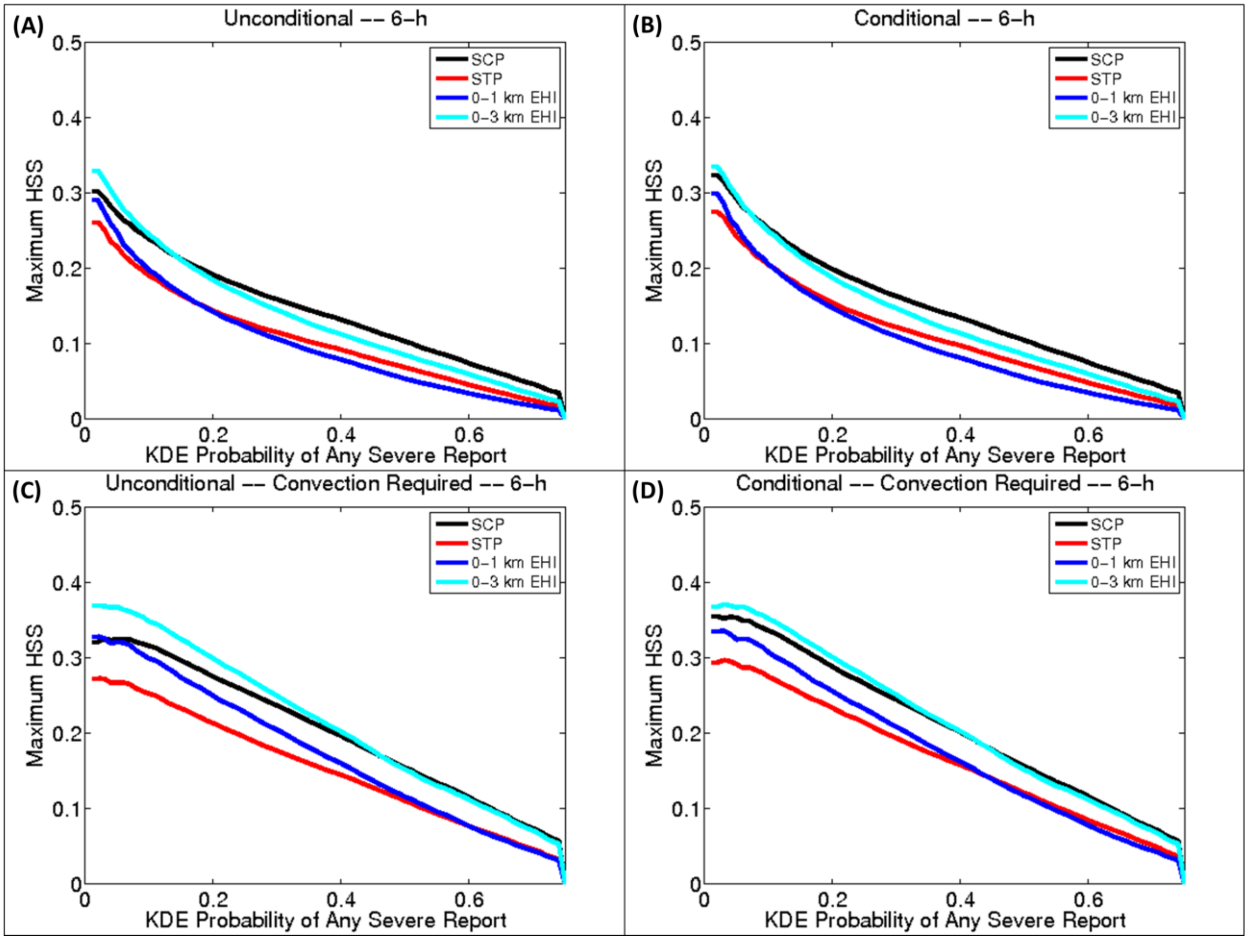

Figure 6: Maximum HSSs as a function of PP probabilities for each SWDV (legend), for 6-h diagnoses of any type of severe report, using (a) each grid point on each day from 1 January 2001 to 31 December 2010, (b) each grid point on each day in which at least one severe report is reported from 1 January 2001 to 31 December 2010, (c) as in (a), for grid points in which convection was reported, and (d) as in (b), for grid points in which convection was reported.

Figs. 6a,c; Figs. 6b,d), and skill scores for all SWDVs are slightly higher for low PP probabilities when considering only grid points where convection occurs.

When considering a rarer event (e.g., significant hail; Fig. 7), the tendencies observed in Fig. 6 become more pronounced. Improvements in skill occur both when considering grid points with convection only, and to a greater degree, when considering only days in which significant hail occurs. The improvement in skill for the same probability is substantial (between 0.1 and 0.2 for probabilities < 0.10) when considering only days in which significant hail occurs versus considering any day in the 10-y period. Evidently, variables with deeper-layer shear variables (i.e., 0-3 km EHI and SCP in Fig. 7) have somewhat higher skill than SWDVs with emphasis on low-level vertical wind shear (0-1 km EHI and STP), though these results are typically not statistically significant to 95% confidence (not shown). The lack of statistical significance, particularly for high PP probabilities, agrees with the findings by Shafer et al. (2012) regarding the similar performance in distinguishing major severe weather outbreaks from less significant events.

3.4 *Statistics for Preexisting Conditions*

The results in previous sections indicate strong sensitivity of skill to whether or not a severe report of the relevant type occurred on the day of interest and/or whether or not convection occurred at the grid point of interest

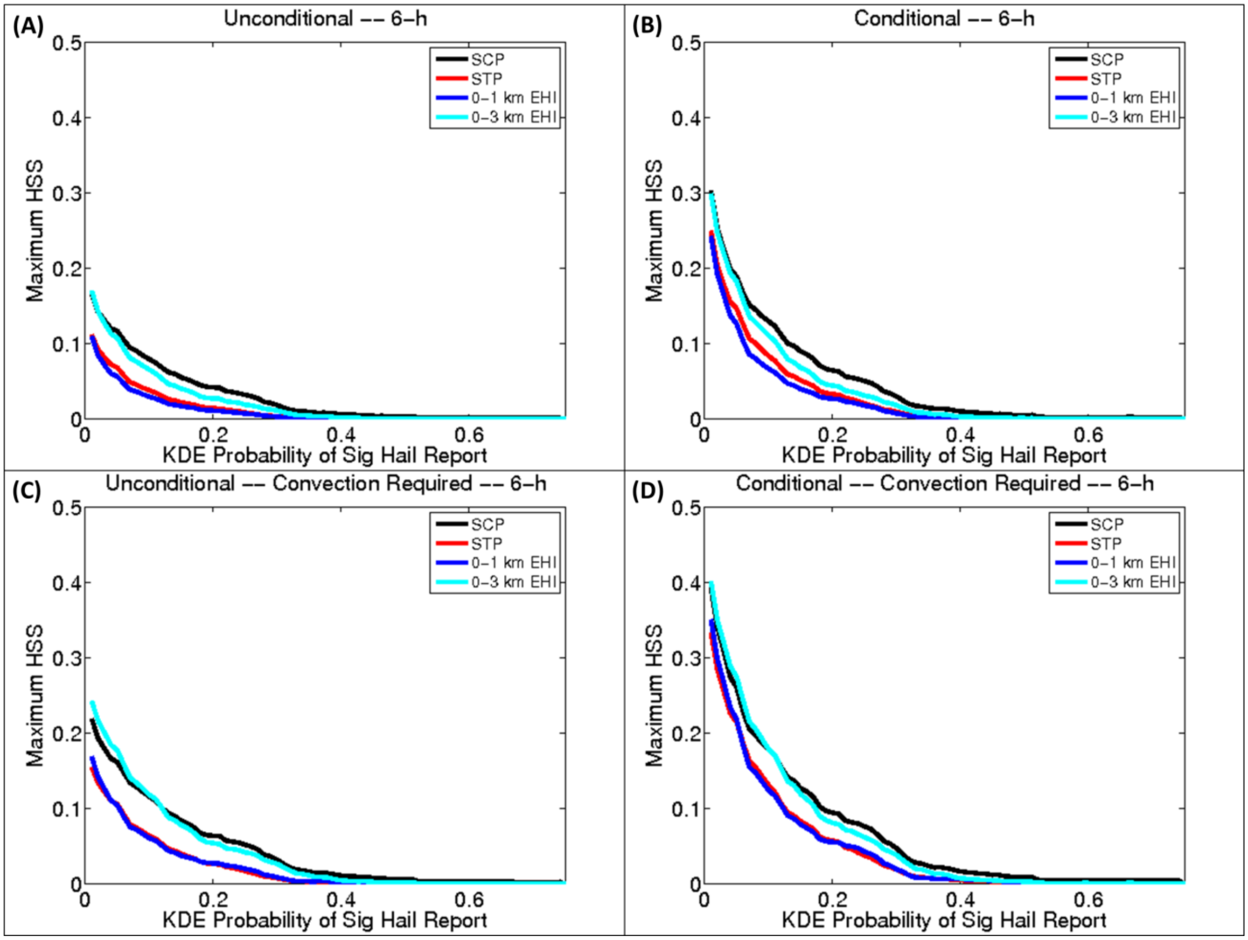

Figure 7: As in Fig. 6, except considering significant hail reports only.

during the time period of interest. Comparison of 24-h diagnoses of hail and significant hail reports (Fig. 8) confirm these sensitivities. However, these sensitivities are not consistent among the report variables. For example, for hail reports, considering days only in which severe hail occurs affects the skill scores as a function of SWDV magnitude more than whether or not convection is required to at a grid point (Fig. 8c). For significant hail reports, the opposite is observed (Fig. 8d). Conversely, for hail reports, considering grid points in which convection occurs affects skill scores as a function of PP probability more than considering only days in which severe hail reports are observed (Fig. 8e). Again, for significant hail reports, the opposite is true (Fig. 8f).

For hail reports, the consideration of only days in which severe hail is reported appears to reduce the skill observed for low SWDV magnitudes and appears to increase the skill observed for large magnitudes. Similarly, the consideration of only grid points in which convection occurs lowers skill for a given PP probability for low probabilities and increases skill for high probabilities. This is not observed for significant hail reports. The consideration of only days in which significant hail occurs and only grid points in which convection occurs generally increases skill for all PP probabilities and SWDV magnitudes, respectively. These tendencies are generally observed for any report type versus any significant report type, implying that it is the rarity of the event that is responsible for these tendencies.

4. DISCUSSION

4.1 *Sample size limitations*

The sensitivities reported in the previous section appear to be associated strongly with the nature of the dataset (notably, the rarity of

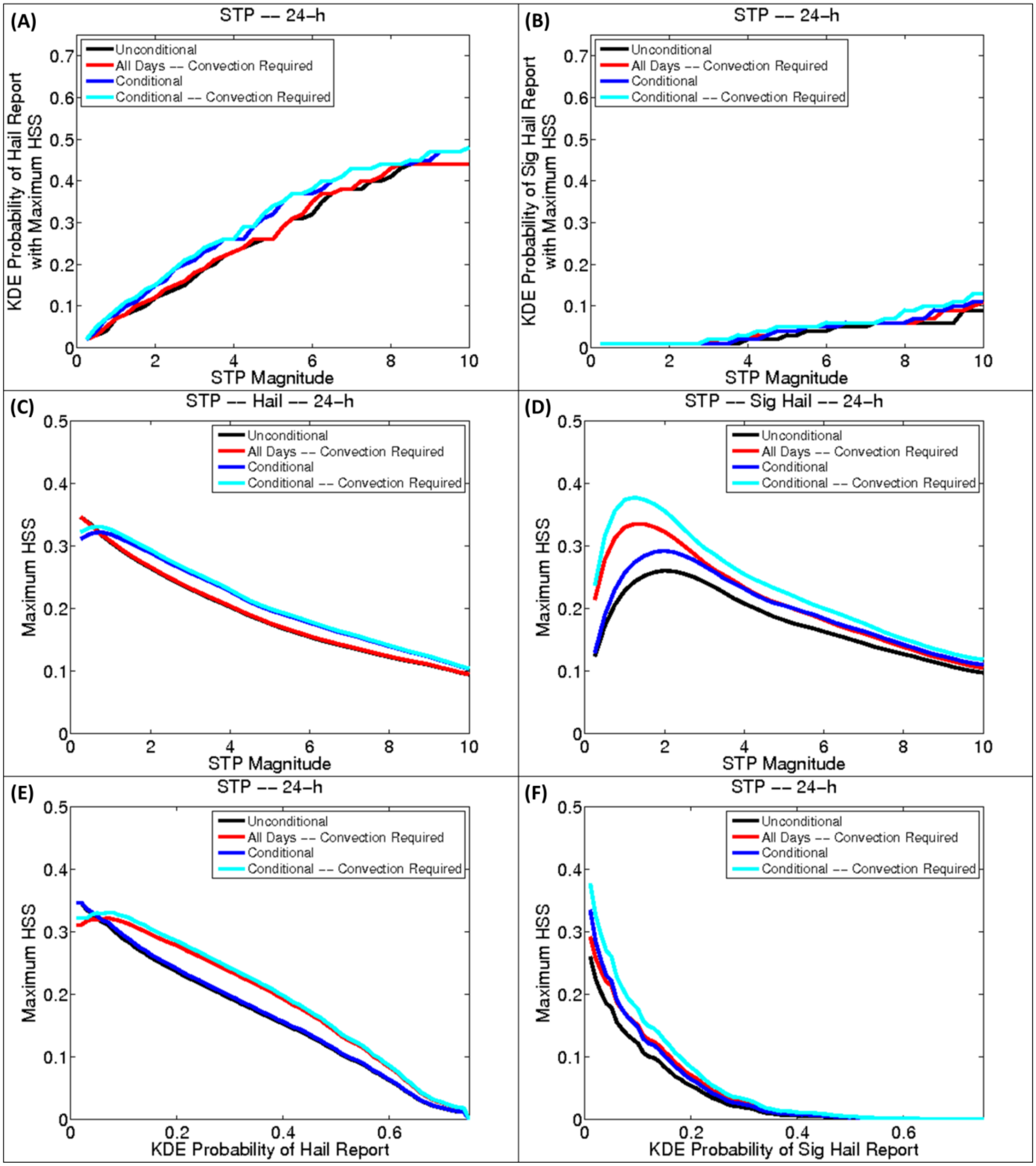

Figure 8: As in Fig. 2, except preexisting conditions associated with the dataset (legend) are considered for hail reports (a,c,e) and significant hail reports (b,d,f), using STP for 24-h diagnoses.

the event being considered), in addition to meteorological factors. As severe reports are exceedingly rare in general compared to the number of grid points considered (even when considering only grid points in which convection occurs), sample size sensitivities are likely affecting the results. An example of this sensitivity is the variability in the statistics and probabilities for high SWDV magnitudes versus low magnitudes (Fig. 9). The large and increasing variability of the probabilities for which the maximum HSS is observed for large

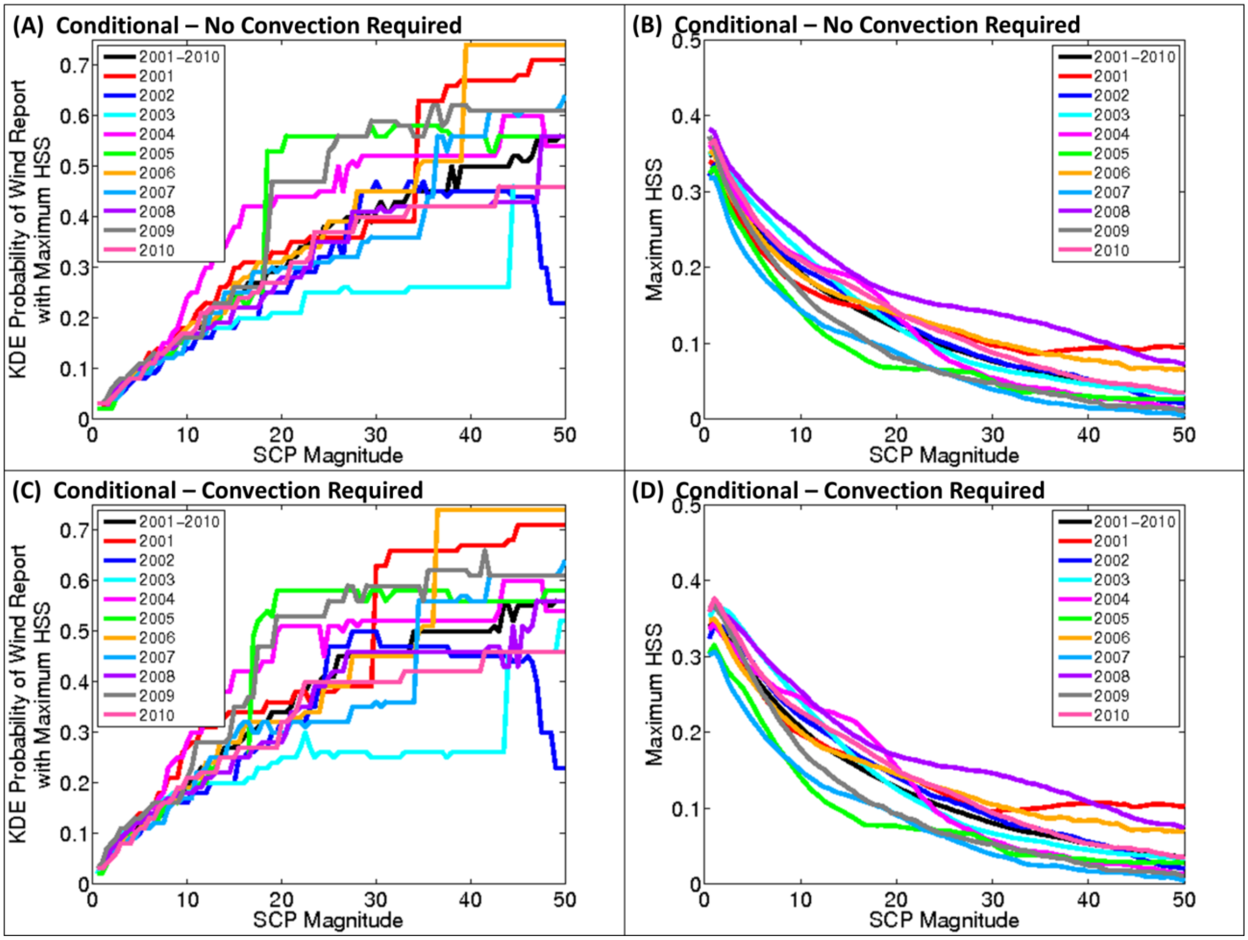

Figure 9: (a) The 24-h KDE-derived probabilities of a severe wind report for which the maximum HSS is observed for a given magnitude of SCP for each year or for all years in the 2001-2010 period (legend), for days in which a severe wind report was observed. (b) As in (a), except showing the maximum HSSs as a function of the SCP magnitude. (c)-(d) As in (a)-(b), except only evaluating grid points in which convection was observed.

SWDV magnitudes is undoubtedly a result of the increasingly rare occurrences in which these SWDV magnitudes are attained. Additionally, the years for which the HSSs are highest are the years in which large SWDV magnitudes occur more often (i.e., 2001, 2006, and 2008; Figs. 9b,d).

There is some evidence that seasonal sensitivities exist as well (e.g., Fig. 10). For example, the probabilities for which the maximum HSSs are observed for low SWDV magnitudes are noticeably higher in the summer (June-August) than in the other seasons. For high SWDV magnitudes, large variations season to season are observed (e.g., the large probabilities of wind reports in the winter for SCP magnitudes between 20 and 40 versus the low probabilities in the spring). A general trend is the considerably higher HSSs and somewhat lower probabilities for any SWDV magnitude in the spring. The relative dominance of wind reports in the summer (e.g., Doswell et al. 2005) versus the relative dominance of supercells in the spring (e.g., Brooks et al. 2003) may explain these two tendencies, given that SCP was introduced and identified as skillful in identifying supercell environments (Thompson et al. 2003). The tendency for probabilities to be higher for tornado and hail reports in the summer is not observed (not shown), providing further evidence that this sensitivity is associated with climatological frequencies. Also of note are the relatively limited differences in statistics when considering only grid points in which convection

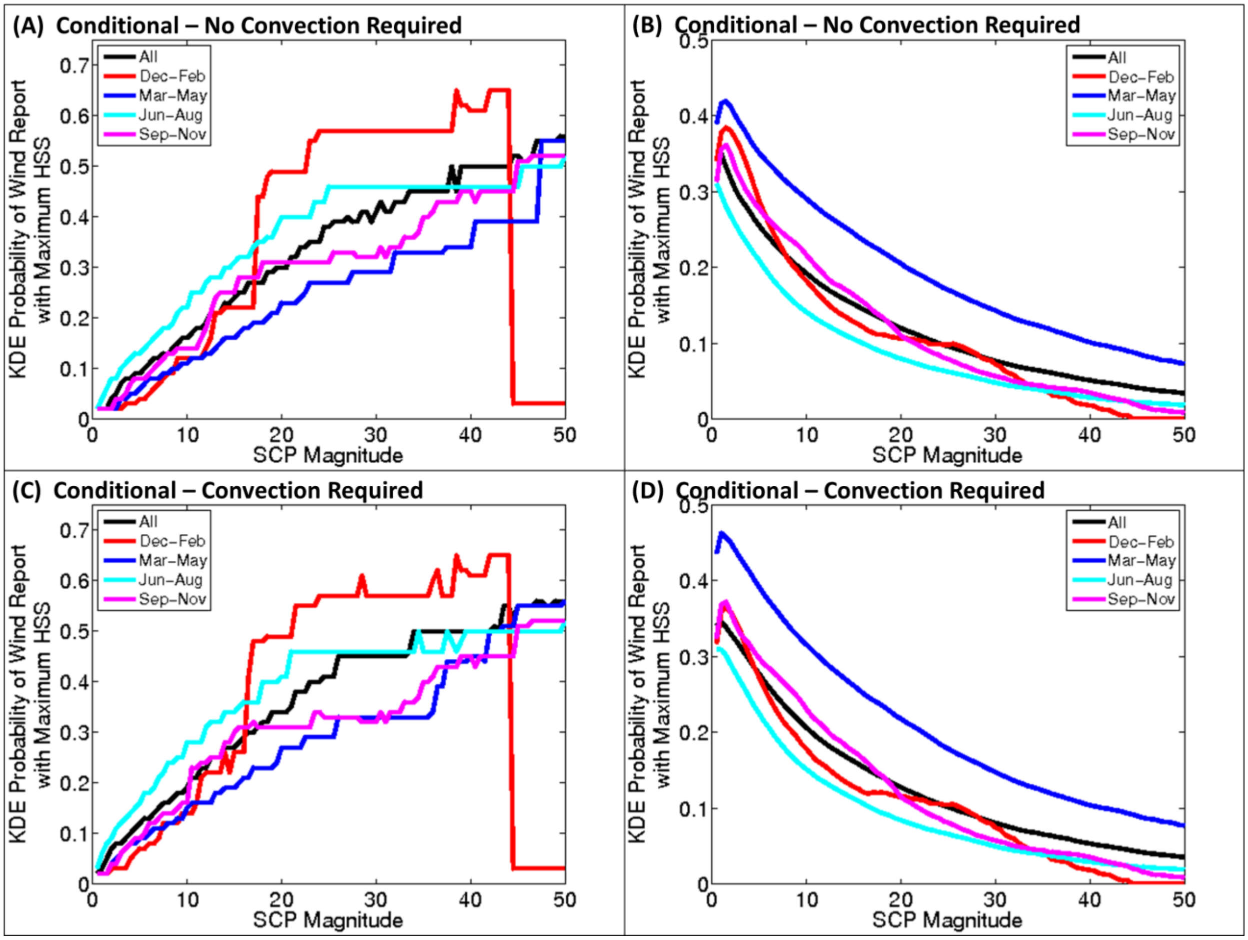

Figure 10: As in Fig. 9, except assessing seasonal sensitivities (see legend).

was observed in the 24-h periods, both yearly and seasonally.

4.2 *Diurnal sensitivities*

As sample size sensitivities appear to affect the yearly and seasonal statistics to some degree, it is appropriate to test diurnal sensitivities as well. The 3-h diagnoses (1200-1500 UTC, 1500-1800 UTC, etc.) and the 6-h diagnoses (1200-1800 UTC, 1800-0000 UTC, etc.) for any severe report indicate that such sensitivities are prevalent (Fig. 11). In particular, there is a clear indication that skill scores as a function of PP probabilities are higher for time periods just after 0000 UTC (see Figs. 11e,f). In addition, the probabilities of any severe report as a function of SWDV magnitude are considerably higher for time periods near 0000 UTC (Figs. 11a,b). HSSs are higher as a function of SWDV magnitude for time periods near 0600 UTC. Note that these relative maxima are consistent with the times near and just after peak heating, when severe weather is most common, supercellular convection is most frequent, and magnitudes of SWDVs tend to be largest (see Tables 1 and 2). The results for specific types of severe weather are consistent with the findings for any severe report (not shown) and with the frequencies/tendencies of the severe weather report variables and SWDVs (Tables 1 and 2).

4.3 *Overall conclusions*

Initial findings of the SWDV evaluation described in this study suggest that many important considerations are required to determine the utility of SWDVs in severe weather diagnosis and the interpretation of a particular magnitude of a SWDV in terms of severe weather probabilities. Many of these considerations are well-known, including the sensitivity of the probabilities and skill scores through which the probabilities are derived to the

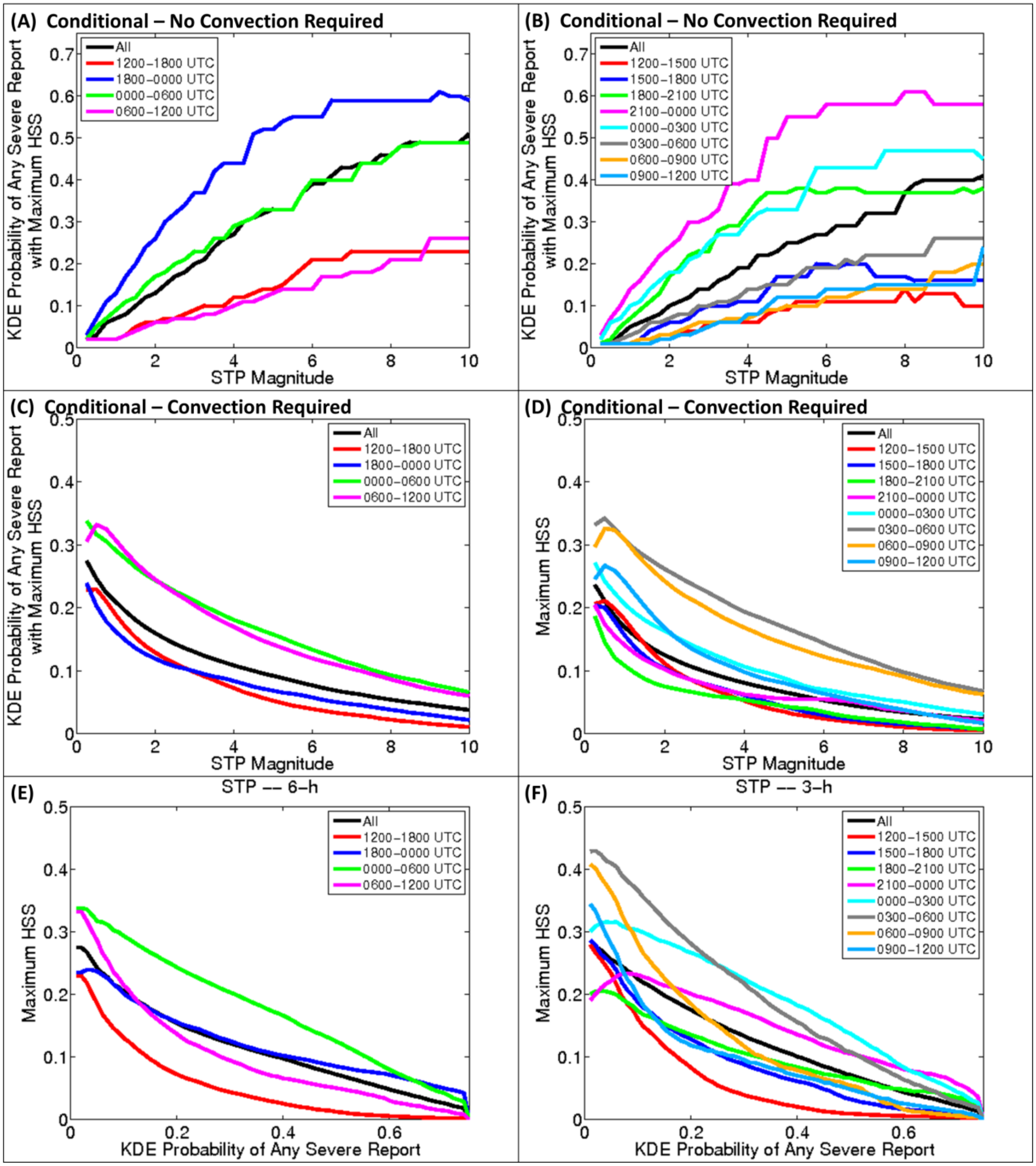

Figure 11: (a) KDE-derived 6-h probabilities of any severe report as a function of STP magnitude for each 6-h period and for all periods (legend) for any day in which a severe report is observed in the 2001 – 2010 period. (b) As in (a), except for 3-h probabilities. (c)-(d) As in (a)-(b), except for HSSs as a function of STP magnitude. (e)-(f) As in (a)-(b), except for HSSs as a function of KDE-derived probabilities.

Table 1: Number of reports and average SWDV magnitudes for all grid points in 3-h periods in which at least one severe report is observed in the domain, when considering all days from 2001–2010.

	1200-1500 UTC	1500-1800 UTC	1800-2100 UTC	2100-0000 UTC	0000-0300 UTC	0300-0600 UTC	0600-0900 UTC	0900-1200 UTC
Any Severe	8444	14590	53848	96828	71220	30037	14659	9219
Wind	3767	6402	24331	38923	27981	13725	7671	4837
Hail	3663	6202	23493	45831	33979	12425	5169	3221
Tornado	415	776	2165	4402	3357	950	478	387
Sig Wind	287	418	1101	2428	2362	1313	734	485
Sig Hail	119	156	711	2117	1876	508	163	87
Sig Tor	51	62	163	356	383	156	98	51
SCP	0.208	0.220	0.263	0.323	0.335	0.431	0.376	0.284
STP	0.082	0.079	0.064	0.057	0.067	0.110	0.121	0.102
0-1 km EHI	0.104	0.090	0.076	0.070	0.085	0.132	0.145	0.130
0-3 km EHI	0.128	0.128	0.148	0.176	0.176	0.212	0.194	0.163

Table 2: As in Table 1, for each year in the dataset.

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Any Severe	24379	24476	26400	26137	26386	30723	26153	35556	27013	27130
Wind	11288	11319	11459	11467	11728	13299	12631	16481	12807	15158
Hail	11899	12242	13596	12880	13401	16344	12441	17404	13071	10705
Tornado	1212	934	1376	1816	1265	1102	1098	1689	1156	1282
Sig Wind	734	727	1033	915	971	925	711	1215	873	1024
Sig Hail	457	452	623	576	527	660	417	691	676	658
Sig Tor	123	96	128	131	106	123	125	207	106	175
SCP	0.309	0.360	0.359	0.291	0.280	0.282	0.305	0.343	0.235	0.360
STP	0.078	0.092	0.087	0.088	0.074	0.064	0.084	0.090	0.062	0.097
0-1 km EHI	0.097	0.117	0.105	0.012	0.094	0.088	0.103	0.104	0.069	0.113
0-3 km EHI	0.174	0.191	0.186	0.163	0.163	0.152	0.175	0.168	0.123	0.190

length of time a SWDV is evaluated (and severe reports are considered), to the sample size of the report type and the SWDV magnitude (with increasing uncertainty associated with more significant severe weather and higher SWDV magnitudes), and to the qualifying conditions for evaluation. Regarding the last point, it is evident that a priori knowledge of the presence or absence of severe weather of a particular type provides a considerable increase in skill in SWDV magnitudes identifying regions with probability exceedances, particularly for small time windows. However, there is only a modest increase in the probabilities for which the highest skill is observed. Additionally, knowledge of locations in which convection occurs appears to improve skill scores for a given SWDV magnitude or PP probability; however, PP probabilities for which the maximum skill score is observed do not increase considerably. The improvement in skill appears to occur for low SWDV magnitudes and low PP probabilities. This suggests two scenarios where a priori knowledge of convection is helpful: (1) situations with considerable uncertainty regarding convective initiation, despite the presence of sufficient environmental conditions for severe weather, and (2) situations with potentially considerable convective coverage but marginal SWDV magnitudes. In the first scenario, if convection develops, convective coverage tends to remain sparse (low PP probabilities). In the second scenario, densities of severe reports remain limited owing to a marginally favorable severe weather environment. Thus, a priori knowledge of convection influences low-PP probability situations disproportionately. When SWDV magnitudes are large and cover large areas, a priori knowledge of convection tends to be of limited value, because convective initiation tends to be more certain and coverage tends to be considerable.

Evidently, datasets with conditional criteria affect rarer-event statistics considerably more. Interestingly, using datasets with conditional criteria may result in lower skill scores for low-PP probability scenarios, suggesting the relative lack of discriminating power SWDVs may be providing at these magnitudes. Beforehand knowledge of days in which any severe report (or hail or wind reports) occurs reduces skill scores as a function of SWDV magnitudes for relatively marginal values (e.g., STP < 0.5; Fig. 8c), whereas beforehand knowledge of where convection occurs reduces skill scores as a function of PP probabilities for marginal report densities (e.g., PP probabilities < 0.05; Fig. 8e).

The differences in the statistics/probabilities among the four SWDVs are minor. The results suggest the relative utility of deeper-layer shear variables (0-3 km EHI and SCP) to lower-layer shear variables (0-1 km EHI and STP) for most severe weather report variables, though results are generally not statistically significant to 95% confidence. There is some evidence to suggest that SCP and STP (0-1 km EHI and 0-3 km EHI) are *relatively* skillful at high (low) SWDV magnitudes and PP probabilities. Again, these discrepancies are not substantial, and overall, the results appear to confirm the findings of Shafer et al. (2012), who showed that these variables have statistically similar skill discriminating severe weather environments.

Lastly, diurnal, seasonal, and annual sensitivities are apparent, and are associated strongly with sample size. For example, probabilities associated with SWDV magnitudes for wind reports tend to be higher in the summer, when wind reports are at a climatological maximum. Additionally, probabilities of severe reports tend to be higher for times of the day in which severe weather is most common, and skill scores tend to be higher just after peak heating, when SWDV magnitudes tend to be somewhat larger (cf. Fig. 11 and Table 1). The latter may be associated with the nocturnal boundary layer wind maximum and low-level jet streams (Doswell and Bosart 2001), which are frequently present in severe weather events. Annual sensitivities appear to be associated with years in which the number of severe weather reports is high or low (e.g., the relatively high skill in 2006 and 2008; cf. Figs. 9b,d and Table 2).

Secular trends are known to exist in the dataset (Brooks et al. 2003; Doswell et al. 2005, 2006; Shafer and Dowell 2010, 2011). However, no effort was made in this study to ameliorate any secular trends that may exist, owing to the relatively short time period compared to the span of years in which these trends have been identified and the limited evidence of such trends in the 10-y period (see Table 2). Use of longer time periods will inevitably lead to nonstationarity of the annual report totals, and as a result, nonstationary relationships between SWDV magnitudes and PP probabilities. Detrending techniques (such as those employed by Doswell et al. 2006) are recommended for considerably larger datasets.

4.4 *Future work*

The evaluation of SWDVs in the discrimination of severe weather environments poses challenges, owing to the rare-events nature of severe weather, errors and uncertainty associated with observations of severe weather, uncertainties associated with sample size, secular trends and other nonmeteorological artifacts in the dataset, and the potentially rapidly evolving environments preceding and during severe weather events. Thus, SWDVs, many of which have not been evaluated rigorously, need to be evaluated for a variety of time periods on a variety of spatial scales. Additionally, it is not readily apparent whether using conditional criteria to evaluate SWDVs generalizes in null events. The results of this study suggest that many preceding studies using conditional datasets likely do not account for these null cases adequately, particularly in terms of the reported skill with which SWDVs discriminate various severe weather phenomena. *The evaluation of SWDVs with varying conditional criteria and with no conditional criteria is needed to describe their utility more accurately in operational environments.*

Although this study proposes a rigorous method of evaluating four SWDVs in the discrimination of severe weather in an attempt to provide a more accurate assessment of their utility, several challenges have not been investigated here, including the computation of PP probabilities using varying KDE bandwidths (which, according to Brooks et al. 1998, is analogous to the confidence with which one can pinpoint the locations of severe weather), varying horizontal grid spacing of the gridded field (e.g., the 80-km x 80-km domain used by

Brooks et al. 2003, which is similar to SPC probabilistic outlooks), the testing of alternative conditional criteria (e.g., grid points with nonzero CAPE, grid points with wet-bulb zero heights constraints, etc.), and the evaluation of diagnostic variables as forecast parameters (i.e., using a current value of a diagnostic variable to assess probabilities of severe weather at a future time; see Doswell and Schultz 2006). The main objectives of such work, in addition to that shown in the above sections, are to provide a comprehensive and objective method of describing a SWDV's utility in diagnosing severe weather and to propose these methods as a means of evaluating as yet untested SWDVs for operational implementation.

ACKNOWLEDGMENTS

This work is sponsored by NSF Grant AGS-0831359, the Undergraduate Committee for Undergraduate Research at the University of South Alabama, and the Alabama Space Grant Consortium. This research was funded in part by a Summer Profession Development Award provided by the University of South Alabama. We thank the University of Oklahoma Supercomputing Center for Education and Research (OSCER) and the Center for Hurricane Intensity and Landfall Investigation (CHILI) for providing the computing resources necessary for this project.

REFERENCES

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

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

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.

Dial, G. L., J. P. Racy, and R. L. Thompson, 2010: Short-term convective mode evolution along synoptic boundaries. *Wea. Forecasting*, **25**, 1430–1446.

Doswell, C. A. III, and J. S. Evans, 2003: Proximity sounding analysis for derechos and supercells: An assessment of similarities and differences. *Atmos. Res.*, **67-68**, 135–152.

——, and L. F. Bosart, 2001: Extratropical synoptic-scale processes and severe convection. *Severe Convective Storms, Meteor. Monogr.*, **28**, no. 50, Amer. Meteor. Soc., 27–69.

——, and D. M. Schultz, 2006: On the use of indices and parameters in forecasting severe storms. *Electronic J. Severe Storms Meteor.*, **1** (3), 1–22.

——, R. Davies-Jones, and D. L. Keller, 1990: On summary measures of skill in rare event forecasting based on contingency tables. *Wea. Forecasting*, **5**, 576–585.

——, H. E. Brooks, and M. P. Kay, 2005: Climatological estimates of daily 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.

Hart, J. A., 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.]

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

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

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

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

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

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

——, and ——, 2011: Using kernel density estimation to identify, rank, and classify severe weather outbreak events, 2011. *Electronic J. Severe Storms Meteor.*, **6** (2), 1–28.

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

——, 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.

——, ——, M. B. Richman, L. M. Leslie, and C. A. Doswell III, 2012: An assessment of areal coverage of severe weather parameters for severe weather diagnosis. *Wea. Forecasting*, **27**, 809–831.

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., 2006: *Statistical Methods in the Atmospheric Sciences*. 2d ed. Academic Press, 627 pp.