16.2 A DATABASE OF PROXIMITY SOUNDINGS FOR SIGNIFICANT SEVERE THUNDERSTORMS

Harold E. Brooks[1] and Jeffrey P. Craven[2]

1. Introduction

Analysis of upper-air soundings taken in proximity to severe thunderstorms has been a valuable tool in understanding the environmental conditions associated with severe weather(e.g., Darkow and Fowler 1971, Maddox 1976, Brooks, Doswell, and Cooper 1994, Rasmussen and Blanchard 1998). The proximity sounding dataset here grew out of the study of 1997-1999 soundings described in Craven et al. (2002), hereafter CBH. Previous studies necessarily have been limited in the number of soundings associated with thunderstorms that produced significant severe weather (hail at least 2 inches in diameter, winds of at least 65 kts, and/or an F2 or greater tornado), as a result of the rarity of the phenomena. Given that such events are the most hazardous, identification of the particular environments associated with these "high-end" weather events could be valuable in forecasting. In order to increase the sample size, every operational 0000 UTC sounding from 1 June 1957-31 December 1993 was considered. (Eventually, the period 1994-1996 will be added for completeness.) The criteria for proximity are a distance of 100 nm from the sounding location and occurrence from 2100-0300 UTC. Each sounding was examined by one of the authors (JPC) for gross considerations of representativeness. In addition, checks for a minimal amount of most unstable convective available potential energy (MUCAPE) of at least 150 J kg-1 removed some soundings. See CBH for more discussion of these issues. The final dataset contains 4012 soundings, 1500 of which are associated with tornadoes, 1979 with giant hail, and 1153 with strong winds. Note that some soundings are associated with more than one kind of severe weather event.

The interannual variability of the number of soundings is large (Fig. 1). There has been an increase in the number of non-tornadic soundings, relative to the number of tornadic soundings, through the years. This is particularly clear in the last few years of the dataset. Only 23% of the soundings in the last ten years in the dataset are tornadic with 38% in the ten years prior to that.

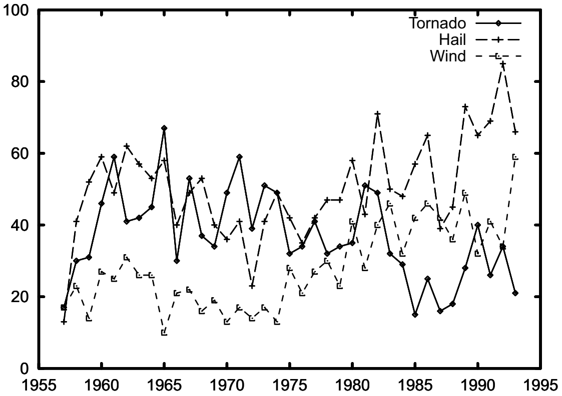

Figure 1. Number of proximity soundings per year by weather event. Note that some soundings (15% of total) are associated with more than one kind of event.

Using the results from CBH, we considered the historical performance of the 1997-1999 "best" discriminator between significant tornadic (F2 or stronger) and significant non-tornadic soundings. "Best" is measured by the performance of a single straight line in dividing a combination of two parameters into non-tornadic and tornadic regions, as measured by standard skill scores, such as the Heidke Skill Score (see Murphy 1996 and Doswell et al. 1990). This discriminator is a line defined by SH1=(MLLCL/40)-9, where MLLCL is the height, in m, of the lifted condensation level of a parcel mixed over the lowest 100 mb of the atmosphere, and SH1 is the vector wind difference between the surface and 1 km in knots. (We will refer to the vector wind difference over the lowest layers of the atmosphere as shear, even though it does not divide by the depth. Since we will only do so in the context of a standard depth, the only difference is in a multiplicative constant.) If the 0-1 km shear is above that line, then the environment is classified as "tornadic". We can then treat the discrimination problem as a 2x2 forecast problem where the forecast is "tornadic" or "non-tornadic" and the observation is "tornadic" or "non-tornadic." The fraction of the tornadoes that are in the "tornadic" environment region of the shear-MLLCL diagram averages 63% from 1973-1993, but a dramatic and sudden fall-off in performance is seen prior to that (Fig. 2). From 1958-1972, the average is only 44%. Clearly, there are a large number of tornadic soundings from 1958-1972, in the part of the parameter space associated with low tornado probability in the years 1973-1993 and 1997-1999. This implies that either atmospheric conditions in which strong thunderstorms form changed in 1973 or reporting changed. 1973 marks the change in tornado verification from the federal state climatologist program to the National Weather Service. In addition, tornadoes prior to the adoption of the Fujita scale in 1975 were rated after-the-fact. The change in performance of the discriminator is consistent with the hypothesis that the pre-1973 tornadoes were overrated on the Fujita scale, compared to post-1973 standards. As a result, we focus on the 1973-1993 period as a reasonably consistent period of record for consideration of environmental conditions. There are 2358 soundings in the 21 year period, including 741 tornadic soundings, 1185 hail soundings, and 786 wind soundings. We will look at differences between tornadic and non-tornadic environments in terms of single parameters and in probabilities of occurrence given combinations of two parameters.

---

[1] NOAA/National Severe Storms Laboratory. Corresponding author address: NOAA/NSSL, 1313 Halley Circle, Norman, OK 73069. E-mail: Harold.Brooks@noaa.gov

[2] NOAA/Storm Prediction Center, Norman,OK.

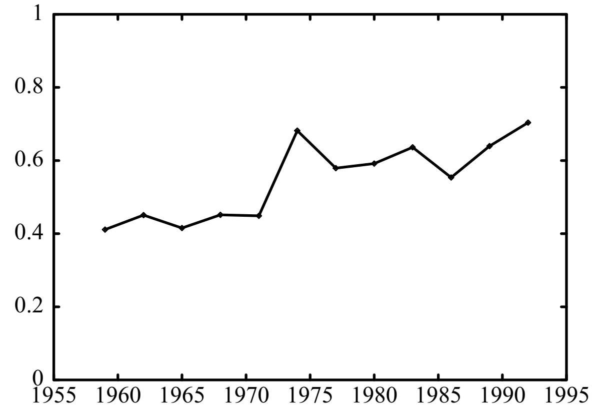

Figure 2. Fraction of significant tornadoes above “best” discrimination line based on 1997-9 study. Values over non-overlapping three-year periods are plotted at the middle year.

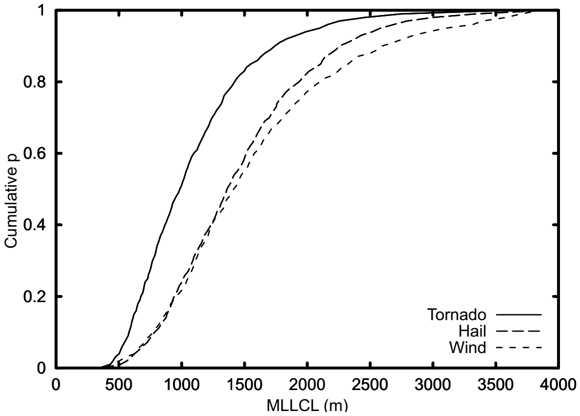

Figure 3. Cumulative density function of mean-layer lifted condensation level (MLLCL) (in meters) for tornadoes (solid line), hail (long dash), and wind (short dash).

2. Single Parameters

We have looked at a number of parameters that can be derived from the soundings, but will report results only on three of them here, the MLLCL, and the 0-1 km shear, as a measure of low-level shear, and 0-6 km shear, as a measure of deep shear. The process of parameter selection was guided in large part by the results of CBH.

A convenient way to show differences (or similarities) between values of parameters associated with different kinds of weather events is to plot the cumulative distribution function (CDF) of the parameter of interest for each weather type. The CDF is given by $DF(x) = \int_{-\infty}^{X} p(X)dX$, $p(X)$ is the probability density function. In other words, it’s the probability that the value will be less than the threshold, *X.* Obviously, it will vary from 0 to 1. As an example, consider the distribution of the MLLCL for tornadoes, wind and hail (Fig. 3). Separation in the CDFs indicates that the parameter discriminates between the different weather types. In the case of the MLLCL, the discrimination is good between tornadic and non-tornadic storms, but not between wind and hail. Tornadic environments are characterized by lower MLLCL heights. Over most of the range of the distribution, tornadic MLLCLs are about 400 m lower. Another way to consider this difference is that half of all tornadic soundings have a MLLCL less than 1000 m, but only 20% of non-tornadic soundings are.

Another parameter that the results of CBH indicated showed promise in discriminating between significant tornadic environments and non-tornadic environments is the 0-1 km shear. Tornadic environments have higher values of low-level shear than non-tornadic environments(Fig. 4). Throughout the distribution, the difference is almost 10 kts. Even more than in the case of the MLLCL, there is no discrimination between hail and wind. In contrast, the deep-layer shear (0-6 km) provides poorer discrimination between tornadic and non-tornadic environments. Wind soundings are somewhat lower in deep-layer wind shear, but the difference between hail and tornadic soundings is small (Fig. 5). We speculate that this is indicative of the stronger organization associated with storms producing giant hail (and tornadoes) than those associated with significant severe winds.

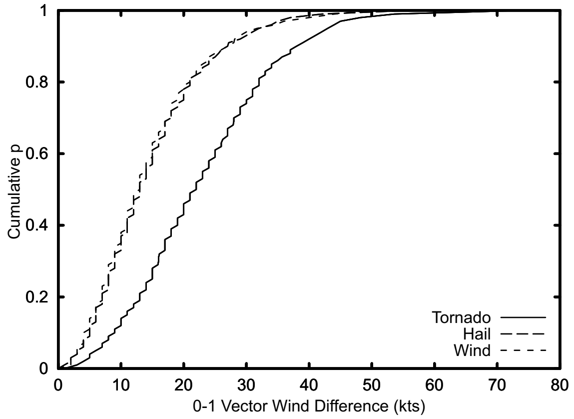

Figure 4. Same as Fig. 3 except for vector difference between surface wind and winds at 1 km (in knots).

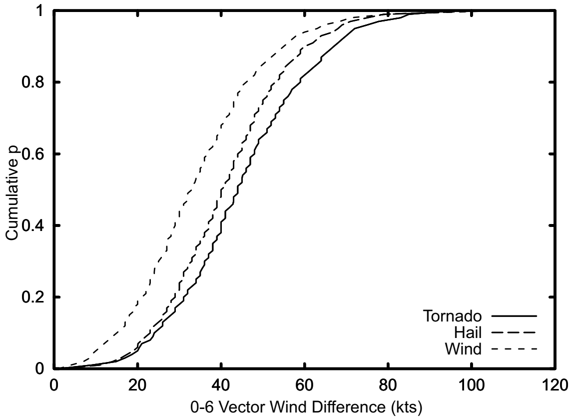

Figure 5. Same as Fig. 3 except for vector difference between surface wind and winds at 6 km (in knots).

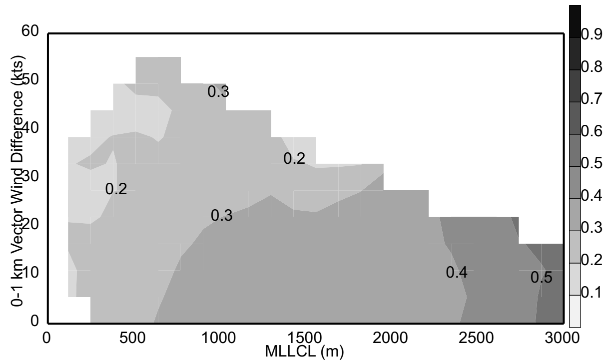

Figure 6. Probability of sounding being associated with convective wind gust of 65 kts or greater given combination MLLCL and vector difference between the surface wind and wind at 1 km above ground level (in knots)

3. Low-level Shear and MLLCL Height

Based on the results of CBH, we wanted to see how well a combination of low-level shear and MLLCL height discriminates between the various weather events. The larger dataset here allows us to carry out a more complete analysis than CBH. Rather than attempting to develop a single line of discrimination as in CBH, we want to create a field of probabilities of events, based on combinations of shear and MLLCL height.

To do this, we created a grid in shear-MLLCL space with 5 kt (shear) and 125 m (MLLCL) spacing and determined which soundings were “near” to the points in the grid. Our arbitrary definition of “near” is an ellipse with axes of 15 kts in the shear direction and 375 m in the MLLCL direction. If the shear and MLLCL of a sounding fell inside the ellipse, that sounding was associated with the point on the grid. Note that the ellipses overlap, so that some soundings are associated with more than one point on the grid. The effects of this are to increase the sample size at each point and to smooth the results.

Once soundings have been associated with the grid points, we calculated the fraction of the total soundings that had wind, hail, or tornadoes) associated with them. This provides an estimate of the conditional probability of the severe weather event, given that significant severe weather occurred. The calculations were carried out only for those grid points with at least thirty associated soundings.

The probability of a sounding being associated with 65 kt or greater winds increases with increasing MLLCL height, although the gradient is weak (Fig. 6). (Only those areas where the probability calculations were carried out are shaded. White regions indicate fewer than 30 soundings at that point on the grid.) Exactly 1/3 of the soundings in the sample are associated with high winds, so that a large area of the parameter space is nearly at the sample climatological frequency. There is some hint that, at the lowest MLLCL heights, increased shear decreases the probability of high winds. The MLLCL height dependence is consistent with a simple model of a dry boundary layer and an associated high MLLCL leading to enhancement of downdrafts by evaporation.

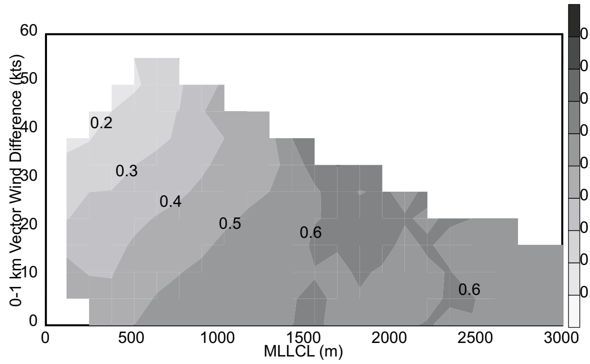

Figure 7. Same as Fig. 6 except for hail at least 2 inches in diameter.

The gradient of probabilities for hail is somewhat stronger than for wind, especially when the MLLCL is low and the shear is high (Fig. 7). In that region of the parameter space, increasing shear and decreasing the MLLCL lowers the probability of hail. Given that the frequency of hail in the sample climatology is 50.2%, again, as with the wind, there is a large region of the parameter space that is reasonably close to the climatological value. The gradient is concentrated below the climatology, suggesting that this pair of parameters may provide some guidance in identifying conditions in which hail is less likely than normal, but little guidance in identifying conditions that are much more likely than normal.

The strongest signal by far is for the tornado discrimination (Fig. 8). 31.4% of the soundings are associated with tornadoes, but the probabilities range from less than 10% to 90%. The strength of the gradient implies that this combination of parameters can help in forecasting those significant severe weather environments that are associated with the strongest tornadoes. Significant tornadoes are rare in high MLLCL environments, but are relatively likely when the MLLCL is low and low-level wind shear is high. It is possible that, with larger datasets, the gradient may actually continue in the low-MLLCL, high-shear direction.

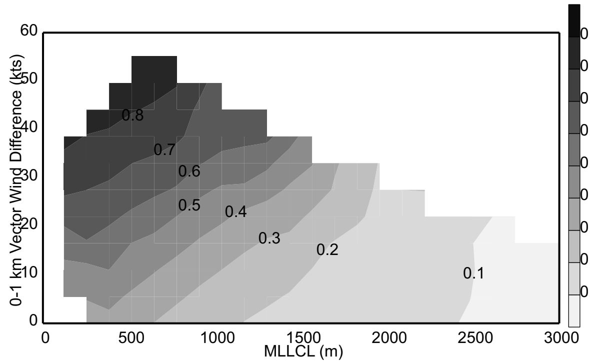

Figure 8. Same as Fig. 6 except for F2 or greater tornado.

4. Discussion

The preliminary conclusions of CBH about discrimination between significant tornadic (F2 or greater) and significant non-tornadic (65 kt winds or 2 inch diameter hail or greater) environments have been confirmed and extended. In particular, tornadic environments tend to be characterized by lower lifted condensation levels and larger values of 0-1 km shear. While the small dataset of CBH allowed for only a cursory examination of the tornado/no tornado problem, the probabilities in the shear-MLLCL parameter space here should provide a more robust guide for forecasters. We caution strongly against using the numerical values as exact estimates of the true probabilities, but the pattern should hold true (e.g., high 0-1 km shear/low MLLCL has a much greater probability of a significant tornado than high MLLCL/low 0-1 km shear).

The role of boundaries in tornadogenesis has been an important topic of research in recent years (Markowski et al. 1998). The relationship of boundaries to the two-dimensional parameter space shown in Fig. 8 is of interest. We offer some speculation on that problem. Boundaries possibly can be thought of as perturbations on the background parameters and “move” the atmospheric state around in the parametric space. Some perturbations may increase the probability of a tornado, while others decrease the probability. In some regions of the parameter space, perturbations may be very important, while in other regions, they may have little effect. It’s possible that, in large outbreak situations, the atmosphere is in the region of the parameter space that is associated with high probabilities over a large spatial area, so that perturbations are less important. In other events, perturbations on the background field may be critical.

Important operational questions arise from this work. The ability of numerical models to forecast these parameters is unknown. Preliminary studies by J. W. Lee of the University of Oklahoma indicate that the full NCAR/NCEP reanalysis shows roughly comparable discriminatory ability between tornadic and non-tornadic environments over the 1997-1999 period, using the same variables. The distributions are slightly different (less shear, lower MLLCL heights), but the analysis and initialization systems used in the reanalysis are clearly capable of preserving important features. That does not necessarily mean that NWP models can maintain that performance at long forecast lead times. Studies of the climatology of model-derived proximity soundings are needed. Nevertheless, it seems clear that the observational network, coupled with numerical models and skilled human forecasters, should be able to make use of this information in producing conditional probabilities of tornadoes that can be part of the convective outlook and watch decision-making process. It is also possible that the warning-decision process could benefit from considering the environment in which a supercell thunderstorm is occurring. If it is in a high MLLCL/low 0-1 km shear environment (cloud base observations and WSR-88D wind profile information can help with those parameters), then these results suggest that the storm is much less likely to produce a significant tornado than if the environment has a low MLLCL and high shear.

The database we have constructed of soundings associated with significant severe thunderstorms provides unique opportunities for studying severe thunderstorm environments because of its large size. We have only scratched the surface with the present study. We plan to investigate a larger number of parameters and to carry out studies of regional and temporal variability.

Finally, we believe that the changes in performance of simple discrimination measures from 1972-1973 provides objective evidence of changes in the nature of the official severe weather database. Unfortunately, this means that for some purposes, the useful length of the climatological record is on the order of 30 years and make interpretation of the decrease in number of F2 or greater tornadoes from the 1950s to the present difficult. The number of tornadoes in the proximity sounding set from 1957-1972 is approximately 44% more than the performance of a simple linear discriminator would suggest would occur, based on the 1973-1993 performance. We hesitate to suggest that this is an accurate estimate of the overestimate of the number of strong and violent tornadoes in the earlier time period, but we believe that the conclusion that the number of tornadoes is overestimated is quite robust.

5. Acknowledgments

The authors wish to thank John Hart for the use of the SHARP software package to analyze the soundings. Conversations over the years with a number of people interested in the proximity analysis problem, including Erik Rasmussen, Paul Markowski, Chuck Doswell, Rich Thompson, John Hart, and Roger Edwards have been invaluable.

6. References

Brooks, H. E., C. A. Doswell III, and J. Cooper, 1994: On the environments of tornadic and nontornadic mesocyclones. *Wea. Forecasting,* **9,** 606-618.

Craven, J. P., H. E. Brooks, and J. A. Hart, 2002: A baseline climatology of soundings, 1997-1999, this volume.

Darkow, G. L., and M. G. Fowler, 1971: Tornado proximity wind sounding analysis. Preprints, *Seventh Conf. on Severe Local Storms,* Kansas City, MO, Amer. Meteor. Soc., 148–151.

Doswell, C. A. III, 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.

Maddox, R. A., 1976: An evaluation of tornado proximity wind and stability data. *Mon. Wea. Rev.*, **104**, 133–142.

Markowski, P. M., E. N. Rasmussen, and J. M. Straka, 1998: Occurrence of tornadoes in supercells interacting with boundaries during VORTEX-95. *Wea. Forecasting*, **13**, 852-859.

Murphy, A. H., 1996: The Finley Affair: A signal event in the history of forecast verification. *Wea. Forecasting*, **11**, 3-20.

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