UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

TORNADO PROXIMITY SOUNDINGS FROM THE NCEP/NCAR REANALYSIS DATA

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

MASTER OF SCIENCE IN METEOROLOGY

Harold Brooks, chair

Kenneth Crawford

Michael Richman

By

JAMES W. LEE
Norman, Oklahoma
2002

TORNADO PROXIMITY SOUNDINGS FROM THE NCEP/NCAR REANALYSIS DATA

A THESIS APPROVED FOR THE
SCHOOL OF METEOROLOGY

BY

[redacted]

Harold Brooks, chair

[redacted]

Kenneth Crawford

[redacted]

Michael Richman

ACKNOWLEDGEMENTS

This work was sponsored by the NOAA Office of Global Programs (Grant #GC00-139). The author would like to thank Harold Brooks, Kenneth Crawford, and Michael Richman, the committee members. The programming skills of John Hart, Michael Baldwin, and James Clark made this project possible. Steve Fletcher and Doug Kennedy were instrumental in facilitating machine usage and storage space. Don Hooper provided extensive assistance in the acquisition and implementation of the reanalysis data files. Jeff Whitaker designed the Fortran90 module used with the Spherepack software, and was very helpful in the implementation and utilization of both programs. Wesley Ebisuzaki provided general information about the reanalysis data. Last, but certainly not least, the personal support of my brother, John, and my parents, James and Faith, was essential to the completion of this thesis.

CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

ABSTRACT

Thermodynamic soundings have been constructed from the reanalysis dataset produced by the National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR). This was done to expand the volume, resolution, and global coverage of sounding data available for research, particularly as it pertains to severe thunderstorm environments. The reanalysis describes global, three-dimensional atmospheric fields, every six hours, for a period of more than fifty years. Soundings were constructed using a series of software packages and programs designed for this project, and were compared with corresponding observational data. Some sounding-derived parameters had values in the reanalysis that were close to those calculated from observed data, while other parameters revealed less agreement. It was concluded that some reanalysis sounding-derived parameters were similar enough to observations to be useful analysis tools. The reanalysis soundings were also examined to assess their capability to differentiate between different types of severe weather environments. Statistical analyses indicate that the reanalysis is able to discriminate quite well in some instances. Thus, the reanalysis dataset represents a potentially significant development in the study of proximity soundings. It is reliable enough (compared to observations) for the increase in volume and resolution of data gained through its use to offset possible errors present in the reanalysis and sounding construction process. Further study of reanalysis soundings could yield more comprehensive results, and should represent a useful alternative (or supplement) to observations in many areas of research that utilize proximity soundings.

CHAPTER I

INTRODUCTION

Throughout the history of meteorological research, work has often focused on attempts to understand severe weather, and in particular, tornadoes. Studies in this discipline have made it clear that severe thunderstorms are very complex systems, and that their behavior is difficult to explain (Doswell and Burgess 1993). A primary method of storm research has been to assume that the nature and evolution of these phenomena can be partially explained by the characteristics of the environments in which they form. By studying the atmosphere in the immediate vicinity (in time and space) of a developing thunderstorm, useful information can be gained regarding the evolution and behavior of that storm. Because severe weather events are three-dimensional phenomena and result from atmospheric conditions at many vertical levels, thermodynamic soundings are a critical tool in understanding their environments.

Early studies of tornadic environments (Showalter 1943; Fawbush and Miller 1952 and 1954; Beebe 1955) explored the use of soundings from the same airmasses in which tornadoes developed. These early works analyzed individual (and sometimes mean) soundings and the similarities and differences between them. These authors focused on parameters that were observed directly from soundings, such as the temperatures at different levels and the heights of various features. Fawbush and Miller (1954) were the first to note that there were fundamental differences in the characteristics of some airmasses they studied (i.e.,

more than one kind of environment was being observed). They compared the airmass types and described the differences between them.

Beebe (1958) was the first to identify what would become a fundamental concept in the research of severe weather environments - the proximity sounding. This type of sounding is defined as being acquired within a specified time and distance of a severe weather event, so that the sounding is said to be "in proximity" to that event. The majority of proximity sounding studies have assumed that the atmosphere is isotropic – that its physical properties are independent of direction. While this assumption is not normally valid at a specific time and location, it is useful when a large dataset of proximity soundings is analyzed. In that case, it is assumed that the total number of events in each direction (relative to the soundings) will be the same. The construction of proximity sounding datasets involves a balance between the minimization of misrepresentative data and the maximization of dataset size. The assumption of isotropy increases the number of soundings in the dataset, but also increases the amount of misrepresentative data. These concepts are further discussed in Chapter III.

Soundings that previously had been studied as representative of tornadic environments were defined more generally by Beebe (1958) as precedent soundings – "those characteristic of the airmass but removed in time and/or space from the vicinity of tornado occurrence (p. 195)." That study also observed that a proximity sounding represented a change (on a smaller scale) in the environment previously characterized by the precedent sounding, so that the proximity

sounding may not have reflected accurately the previous (or larger) airmass. Instead, the proximity sounding more specifically resembled the immediate (and more localized) environment in which a tornadic thunderstorm formed.

The formal definition of a proximity sounding in Beebe (1958) included specific time and space requirements: A tornado must have occurred within 50 miles of and one hour following the release of the sounding instrument. In addition, the proximity definition also accounted for soundings that fit the space/time criteria yet did not represent the tornadic environment (e.g., post-frontal or post-storm soundings). Since the landmark work of Beebe, this general method of quality control has been employed in many similar studies. Although Beebe (1958) established the basis for proximity sounding research, his dataset was small (only 24 soundings). Also, his primary goal was simply to describe the characteristics of tornadic storm environments.

Darkow (1969) compared tornado proximity soundings to nearby "check" soundings (similar to the precedent soundings defined by Beebe [1958]). He noted several differences between the two (i.e., between the tornadic environment and the surrounding, large-scale airmass). The study by Darkow, along with Beebe (1963), represents the roots in the explanation of tornado formation and understanding the processes, which create tornadogenesis.

The initial goal of using proximity soundings was to describe tornado environments. However, these tools have proven useful in identifying potential predictors of storm characteristics. Many studies (e.g., Rasmussen and Wilhelmson 1983; Brooks and Doswell 1994; Rasmussen and Blanchard 1998;

Craven 2001) have explored the utility of using proximity soundings to discriminate between the atmospheric conditions in which different types of severe weather occur. This concept can be useful in storm prediction, because the characteristics of the observed atmosphere can be compared to results of proximity sounding studies, to see what types of severe weather are usually associated with those characteristics. Discrimination is also used to analyze the differences among severe storm environments, which can facilitate a better understanding of the processes responsible for storm formation.

The use of most sounding-derived parameters necessarily discards some information, because several pieces of data are combined into one quantity. However, these parameters have proven to be very useful tools in examining severe weather environments. Rasmussen and Wilhelmson (1983) demonstrated that some sounding-derived parameters (particularly potential buoyant energy [PBE] and the mean wind shear in the 0-4 km layer) could discriminate between the environments of tornadic, mesocyclonic, and non-severe thunderstorms. These authors also inferred the atmospheric processes that lead to the formation of different types of events.

Brooks and Doswell (1994) and Rasmussen and Blanchard (1998) performed studies similar to that of Rasmussen and Wilhelmson (1983), except they were progressively more comprehensive in terms of the amount of data and number of parameters analyzed. Brooks and Doswell (1994) used proximity soundings to investigate differences in the environments of tornadic and nontornadic mesocyclones. They also applied their results to support a conceptual

model of low-level cyclone development and maintenance, illustrating the utility of proximity soundings to facilitate a better understanding of atmospheric processes.

Craven (2001) is, to some extent, an observational counterpart to the current study. He used a large dataset (approximately 60,000 proximity soundings) to develop a climatology of sounding-derived parameters for different severe weather environments. A major finding was that sfc-1 km bulk shear (magnitude of vector difference of wind) and lifted condensation level height (LCLH) values demonstrated significant differences between significant tornado cases and all other severe weather types. Many methods and definitions used by Craven (2001) also were used in this study (see Chapter III).

While proximity sounding research has often yielded useful results, most studies have been hampered by the limited (in time and space) resolution of the rawinsonde network. The result is either a dataset so small that it leads to questions about the validity of conclusions drawn from it, or definitions of proximity so wide that they introduce the problem of whether the soundings are even representative of the storm environment. Even those few works (e.g., Rasmussen and Blanchard 1998; Craven 2001) that compiled large sounding datasets still produced results that were limited in their usefulness on a global scale, because the majority of observed sounding data represents a very small portion of our Earth.

The reanalysis dataset produced by the National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric

Research (NCAR) (Kalnay et al. 1996) presents a unique opportunity to greatly expand the amount of sounding data available for research. The reanalysis data represents a global, three-dimensional picture of the atmosphere, every 6 hours, for more than 50 years. Many output variables were produced in different spatial coordinate systems (both horizontally and vertically). The reanalysis has a much higher grid resolution than the rawinsonde network; the region analyzed in this study (as defined in Chapter III) contained 180 reanalysis gridpoints, but only 53 observed sounding locations. As a result, datasets can be constructed that are much larger in size and scope than those used in previous proximity sounding studies. Because the reanalysis contains global fields, high-resolution data can be studied across areas of the globe where a lack of observations would otherwise make a similar study impossible.

The primary goals of this project were to:

1. Construct thermodynamic soundings from the reanalysis data.

2. Determine whether or not the reanalysis soundings had characteristics consistent with those of observed data.

3. Assess whether or not the reanalysis soundings were able to discriminate between different kinds of storm environments, in a similar fashion to observed data.

If the reanalysis soundings proved to be reliable enough (as determined by comparisons with concurrent observations) to be useful as proximity soundings, the volume of data available for sounding research would be greatly expanded in space and time up to a global resolution.

This manuscript will document:

1. The characteristics of the reanalysis data.

2. The determination of proximity definitions and the ranges of data analyzed.

3. A methodology to construct, organize and manipulate proximity soundings.

4. The data analysis techniques employed, and interpretation of the results.

5. A summary and discussion of the results.

CHAPTER II

THE NCEP/NCAR REANALYSIS DATA

The reanalysis dataset was created through the cooperative efforts of NCEP and NCAR (Kalnay et al. 1996) to produce relatively high-resolution global analyses of atmospheric fields over a long time period. The reanalysis data record has since been extended to include January 1948 through July 2002. The basic concept of the reanalysis was to:

1. Recover all available observations from each time index and synthesize them with a static data assimilation system.

2. Use the observational fields to initialize a model for a six-hour forecast. The model used (hereafter referred to as the reanalysis model) was identical to the NCEP global operational model, except for the horizontal resolution. The reanalysis model is T62 (equivalent to a horizontal resolution of approximately 210 km), while the operational model is T126 (approximately 105 km).

3. Use the forecast as a first-guess, in conjunction with concurrent observational fields, to construct the reanalysis output. Reanalysis fields were generated with an optimal interpolation technique.

4. Repeat the process every six hours.

Thus, the reanalysis used model forecasts *and* observations to transport information from regions of high observational density to those with fewer observations. The state of the atmosphere could thus be estimated in areas that are relatively devoid of data. The result of the reanalysis process was a dataset

consisting of a global, three-dimensional picture of the atmosphere at six-hour intervals during a period of more than 50 years.

Output is available from the reanalysis on pressure and sigma levels (p/p_o, where p is pressure and p_o is surface pressure) in the vertical, and on spectral and grid coordinate systems in the quasi-horizontal. The sigma-level, spectral coefficient form of the data was determined to be the most useful for this study. The sigma-level data contains atmospheric fields on 28 vertical levels, while the pressure-level data contains only 10 vertical levels. In addition, the initial reanalysis model output was produced on sigma levels, so the data were interpolated in the vertical to obtain sounding information on pressure levels. Because the first goal of this project was to construct soundings, having almost three times as many vertical data points was desirable. Another advantage of using the sigma data is the elimination of a potential source of error (data interpolation). Approximately 10 sigma levels exist between the near-surface (the lowest having a sigma value of 0.995) and 700 mb, while in the pressure data, only four levels exist. Because many of the parameters that describe severe weather environments are heavily dependent on the behavior of the atmosphere at lower levels, more data close to the surface are critical.

The pressure-coordinate, gridded data in the reanalysis are on a grid with 144 x 73 data points on the globe. However, when the spectral coefficient data are translated onto an equally spaced (in latitude and longitude) grid, the result is 192 x 94 gridpoints. The spatial resolution is 1.875 degrees (112.5 nautical miles [nm]) longitudinally and 1.915 degrees (114.9 nm) latitudinally. These figures

represent a significant increase versus the quasi-horizontal resolution of the gridded data, a desirable characteristic in most research and when one is concerned with proximity. For these reasons, the spectral coefficient, sigma-level data were used in this study.

The reanalysis data (for spectral, sigma output) includes six atmospheric fields. Surface geopotential is constant over time. The other five fields are available every 6 hours. The natural log of surface pressure is the only one of these five variables not available above the surface. The other four (virtual temperature, specific humidity, divergence, and vorticity) are available at 28 vertical levels. Atmospheric parameters necessary for the construction of a sounding (i.e., temperature, dewpoint, wind speed and direction, heights, and pressure) were derived from the six initial fields. This methodology is described in Chapter IV.

In addition to the limitations generally associated with model data, caution should be used in analyzing some reanalysis fields. Specifically, the reanalysis model has problems with moisture (specific humidity), a common feature associated with model forecasts. These problems stem from, among other reasons, the poor accuracy of moisture observations in the original data and the small spatial scales associated with moisture in the atmosphere. The reanalysis may be even less reliable in data-poor regions (such as oceans), because atmospheric fields in these locations are largely derived from the reanalysis model itself; they are minimally influenced by the observational data.

The variables in the reanalysis data were classified by NCEP/NCAR (Kalnay et al. 1996) as either:

A: Strongly influenced by observed data (e.g., virtual temperature, vorticity).

B: Directly affected by observations, but also strongly influenced by the reanalysis model (e.g., natural log of surface pressure, divergence, specific humidity).

C: Not directly affected by observations, derived exclusively from the reanalysis model forecast fields (e.g., precipitation, surface fluxes).

D: Fixed from climatological values, independent of the reanalysis model (e.g., surface geopotential).

It is not clear that the classification system is easy to utilize. Most variables are differently influenced by the reanalysis model in some regions (e.g., oceans), because observational data there are sparser. Thus, the classifications of each field are not constant. The classifications of the reanalysis variables analyzed in this project were not used to interpret results. However, researchers that utilize the NCEP/NCAR reanalysis data should be aware that some variables are more reliable (i.e., more strongly based on observations) than others.

CHAPTER III

DEFINITIONS

Many studies have used the concept of proximity soundings to identify characteristics of the environments in which different kinds of severe weather form. Proximity is usually defined using specific requirements for the temporal and spatial distances between a sounding and an associated event - when they are met, the sounding is said to be "in proximity" to that event. Although these criteria are somewhat arbitrary (as discussed in Brooks and Doswell 1994), some logic can be used to redefine them for a particular study. Choosing the spatial distance requirement (between sounding and severe weather report) is a balance between expanding the size of the dataset at the expense of introducing misrepresentative data. As the distance is increased, more soundings are classified as being "in proximity". However, the likelihood increases of those soundings being representative of a different environment than the one in which the observed event formed.

The spatial criterion for proximity in this study was 100 nm. This distance was the same as that used in previous work, including Craven (2001). The fact that the "same proximity" was used was important when the reanalysis soundings were compared to Craven's observational dataset. The assumption of isotropy was employed; that is, the classification of events was not dependent on direction (relative to the associated sounding). The definition of proximity as a circle with a 100 nm radius introduced some misrepresentative data (due to soundings being located upstream from their associated events), but it also expanded the size of the

resulting dataset. Because the maximum distance (diagonal) between any two reanalysis soundings is 2.68 degrees or 160.8 nm, all severe weather reports are in spatial proximity to a sounding location (gridpoint). Because the distance between each severe weather event and the nearest reanalysis sounding is always available, the soundings can be sorted by distance. Thus, it is trivial to change the spatial proximity criterion for the purpose of data analysis.

Previous studies defined temporal proximity in only one direction; namely, that an event must have occurred *after* the associated sounding was acquired. This strategy was designed to eliminate soundings that were contaminated by the storms themselves (i.e., convective effects), or represented an environment different than the one in which the severe weather event formed (e.g., after frontal or dryline passage). However, when the reanalysis data is used, convective contamination is a minor issue. The majority of contamination cases occur on a spatial scale that is smaller than the resolution of the reanalysis model, and observations affected by contamination were smoothed by the optimal interpolation process. Thus, the reanalysis soundings do not exhibit contamination as sometimes occurs when observed soundings are used.

The criterion for temporal proximity was three hours before *or* after sounding time, the same definition used by Craven (2001). The temporal requirement introduced soundings representative of the post-storm environment. However, the criterion also greatly increased the number of soundings in the dataset that were representative of the storm environment. In previous studies (where a unidirectional temporal proximity criterion was used), the "extra"

soundings were excluded. The inclusion of events that occurred before sounding time doubled the number of severe weather proximity soundings.

Finally, previous studies (i.e., Beebe 1958; Darkow 1969; Brooks and Doswell 1994; Rasmussen and Blanchard 1998) introduced various measures of quality control and eliminated soundings that were incomplete, contained erroneous data, or were otherwise misrepresentative of the storm environment. For example, soundings with CAPE below a minimum value were eliminated, as were those with incomplete wind data, soundings that did not extend above a certain pressure level, or soundings with an inflow relative to the event to meet a research objective. Because this project utilized model data, incomplete soundings were not an issue. The "inflow sector" method of sounding elimination was not used, because the assumption of isotropy was employed. The inclusion of some misrepresentative data was acceptable, because it also increased the number of proximity soundings in the dataset. No "minimum CAPE" or other method of sounding elimination was used, to let the reanalysis data stand on its own. This simplified approach resulted in a more useful baseline to examine the utility of reanalysis data.

With the definition of proximity established (soundings within 100 nm and three hours before or after an event), the next step was to consider the spatial and temporal ranges of data to be analyzed. Although reanalysis data are available at 6-hour intervals for more than 50 years, soundings were constructed only for 0000 UTC data from a 3-year period (1997-1999). This smaller subset was necessary due to time constraints and limited data storage capacity; it was chosen to match

the time range used in Craven's (2001) study. Because his observational dataset is the first compared with reanalysis data, it was essential to use the same data range. Because these three years were characterized by an above average number of tornado occurrences and because the majority of tornadoes occur around 0000 UTC (versus 0600, 1200, or 1800 UTC), the reanalysis subset is considered of sufficient size to yield significant results.

The reanalysis data were produced at the same resolution across the global domain, but for this study, only the eastern two-thirds of the continental United States were analyzed. Verification data (including the record of significant tornadoes) within the United States were considered to be more complete and accurate than those found in most other regions. Furthermore, the majority of recorded tornadoes in the U.S. occur east of the Rocky Mountains. Thus, the spatial data consisted of all reanalysis gridpoints located over land, within the continental U.S. east of the Rocky Mountains (east of 106.88W longitude). The result was 180 sounding locations (reanalysis gridpoints). Based upon using data from the 0000 UTC analysis for this region of the United States for a 3-year period, a total of 197,100 soundings were constructed.

CHAPTER IV

METHODOLOGY

The reanalysis data files obtained for this study consisted of spectral coefficients in netCDF format. The first steps in the sounding construction process were to convert the data into a more useful file format and to translate it onto an equally spaced (in latitude and longitude) grid. The software package Spherepack 3.0 (Adams and Swarztrauber 1999) was essential to the achievement of this task. It was designed for use with reanalysis data, to translate from spectral space into grid space through spherical harmonic synthesis. The general form of the equation used for this transformation is:

$$A[i,j] = \sum [a_k * f(m,n) + b_k * g(m,n)]$$

where a_k and b_k are the spectral coefficients, m and n are longitudinal and latitudinal wave number, respectively, f and g are trigonometric functions, and A[i,j] is an array of gridded values, where i and j are equally spaced points in latitude and longitude.

A Fortran90 module used Spherepack functions to derive the wind field (zonal and meridional components) from divergence and vorticity fields. The program also derived temperature from virtual temperature and specific humidity, using the formula:

$$T = [T_v/(1+(0.608*q))]-273.15$$

where T = temperature (C), T_v = virtual temperature (K), and q = specific humidity (g/g). It should be noted that small negative values of specific humidity exist in the reanalysis data (not uncommon for model output). These values were

set to zero before the temperature calculation. Pressure was calculated at every gridpoint and vertical level using the natural log of surface pressure and the sigma level values:

$$p = \text{sigma}*[10*(\exp(\ln p_o))]$$

where $\ln(p_o)$ = natural log of surface pressure (Pa), sigma = the value of sigma at a given level, and p = pressure (mb) at that level. Thus, the output of the gridding program consisted of pressure (mb), virtual temperature (K), specific humidity (g/g), temperature (C), zonal wind (m/s), and meridional wind (m/s) at every quasi-horizontal gridpoint at every vertical level. Orography (surface geopotential) was output separately (due to its time invariant nature).

The gridded data was converted into GRIB format using the lats4d script (Christias 2002). Programs designed by Michael Baldwin at NSSL were used to extract vertical profiles from the GRIB data at the 180 locations matching reanalysis gridpoints in the eastern continental United States. These programs calculated variables necessary for sounding construction from those in the gridded data.

Heights were calculated using the hypsometric equation for a moist atmosphere (Bluestein 1992). Wind speed and direction were computed from the zonal and meridional components of the wind field, through trigonometric definitions. Dewpoint temperature was derived from pressure and specific humidity, using the formula (from Buck 1981):

$$T_d = [(C-Dz)/(B-z)]-273.15$$

where $z = \ln(e_s/A)$, e_s = saturation vapor pressure (Pa) = $(p*q)/(0.622+q)$, q = specific humidity (g/g), p = pressure (Pa), A = 613.3, B = 17.502, C = 4780.8, D = 32.19, and T_d = dewpoint temperature (C). Some relative humidity (RH) values that were greater than 100% were forced to 100% (i.e., T_d values greater than T were forced to T). While supersaturation is possible in the atmosphere, it should occur on scales smaller than the resolution of the reanalysis model. It was believed that the values of RH greater than 100% occurred because the formulas used to calculate dewpoint are empirical in nature and can produce small errors, especially in conditions near saturation. The profile extraction programs output data files from which soundings were constructed at all gridpoints within the specified temporal and spatial ranges.

The only changes made to the "raw" reanalysis data were to force limits on humidity (specific humidity to be non-negative and RH to be no greater than 100%) and to compute the variables necessary for sounding construction. The humidity limits could only have affected dewpoints and temperatures. Because the humidity values in question (specific humidity below 0 and RH above 100%) were only slightly modified, the effects on resulting temperature and dewpoint values were minimal. The formulas used to compute sounding variables were either valid by definition or represented standard meteorological relationships. The exception was the equation used to calculate dewpoint, which was partially empirical in nature.

A modified version of the NSHARP software (which was developed from SHARP [Hart and Korotky 1991]), called SHARPTAB (written by John Hart at

SPC), was used to calculate 96 parameters from each sounding (every gridpoint at every time index). Craven (2001) also used SHARPTAB, so the methods used to derive parameters from soundings were consistent between that work and the current study. Of the 96 parameters, 69 consisted of the same 23 calculations performed on three different parcels: Surface based, most unstable (the parcel in the lowest 300 mb with the highest wet bulb potential temperature) – hereafter MU, and mean layer (a parcel with the mean characteristics of the lowest 100 mb) – hereafter ML. These 23 quantities included standard sounding parameters (convective available potential energy [CAPE], convective inhibition [CINH], LCLH, level of free convection height [LFCH], lifted index [LI] at different levels, Cap, storm-relative [SR] wind at different levels, shear in the surface-LI_{max} layer, and mean relative humidity in the LCL-LFC layer) as well as combined parameters and indices. Of the 27 parameters not computed for surface, MU, or ML parcels, 25 were comprised of storm-relative helicity (SRH), bulk shear, lapse rate, mean RH, and storm-relative wind calculated for different vertical levels and layers. The final two parameters were mean mixing ratio in the lowest 100 mb and downdraft CAPE (DCAPE).

Although 96 parameters were output by SHARPTAB, many were not analyzed in this study. Of the 69 quantities calculated for surface, MU, and ML parcels (23 each), only 17 were included in the analysis. None of the surface parcel values were used, because it was believed (based on observations from previous studies) that these parameters would not provide useful results. Also, the reanalysis does not contain surface data (other than pressure and orography),

because the lowest sigma level (0.995) is approximately 5 mb above the surface at sea level. Combined parameters and indices computed for MU and ML parcels, with the exception of BRNShear, were eliminated. They were each calculated from several parameters, and it was believed that no more information would be gained from their use than through analysis of the parameters used to compute them. The value of BRNShear was equivalent for all three parcels, so only MU BRNShear was analyzed. Only the 500 mb level was included for LI (for both MU and ML parcels), because that is the most commonly used height. After these exclusions, 8 parameters were analyzed for both MU and ML parcels (16 total): CAPE, CINH, LCLH, LFCH, Cap, 3 km Cape, LI at 500 mb, and mean RH LCL-LFC.

Of the 25 quantities calculated for different vertical levels and layers, all SRH, bulk shear, lapse rate, and mean RH parameters were used, but the 7 SR wind values were not. This exclusion was made because SRH and bulk shear were sufficient to analyze the usefulness of parameters related to the wind field. Finally, both 100 mb mean mixing ratio and DCAPE were included. Thus, the analysis dataset consisted of a time, date, latitude and longitude, and 37 sounding-derived parameters (listed in Table 1) for each of the 197,100 soundings created.

The final step to prepare the data for analysis was the classification of each sounding according to the types of severe weather in proximity to it. Five categories initially were used for classification. "Significant tornado" (hereafter TOR) indicated the occurrence of an F2 or higher tornado, "significant hail" indicated 2-inch or larger hail, and "significant wind" indicated 65 kt or higher

wind. "Non-significant severe" (hereafter SVR) indicated any severe weather report that did not meet the criteria for any of the "significant" categories, and "non-severe" (hereafter NON) indicated no severe weather reports. The significant hail and significant wind categories were combined into "significant severe" (hereafter SIG), which thus indicated the occurrence of either 2-inch or larger hail *or* 65 kt or higher wind. The initial distinction between these two types of events was made so that future analyses that focus on only one of those categories (or the differences between the two) could be performed.

Severe weather reports were obtained from Severe Plot (Hart 2002) for 1997-1999. Only events that occurred between 2100 UTC and 0300 UTC (within three hours of 0000 UTC) were used. For each report, the distance to every reanalysis gridpoint was calculated, and the nearest sounding was classified as being in proximity to that event. Only "yes" and "no" classifications were used. Thus, if a sounding was in proximity to multiple events of the same type, it received only a "yes" in that category, rather than the number of reports (so that soundings in proximity to multiple events were not given greater weight).

Each sounding was analyzed only as a proximity sounding for one type of event – the highest of the four categories (in the order TOR, SIG, SVR, NON) in which it received a "yes". For example, if a significant tornado report and a significant wind report both occurred in proximity to the same sounding, that sounding was classified as "yes" for both TOR and SIG, but was only treated as TOR during data analysis. This was done to prevent soundings that were in proximity to multiple types of severe weather from being given greater weight.

CHAPTER V

RESULTS

Of the 197,100 soundings constructed in this study, 189,450 were classified as NON, 6,460 as SVR, 1,031 as SIG, and 159 as TOR. While more than 159 F2 or higher tornadoes occurred within the time-frame and geographical area analyzed, some events produced more than one such tornado in proximity to the same sounding. A sounding in proximity to more than one significant tornado was only classified once for all of those observations (rather than once for each).

A. Comparisons between reanalysis and observed sounding data

Before the soundings could be examined for their ability to discriminate between different storm environments, an important issue to address was whether or not the characteristics of the reanalysis data were consistent with those of observed soundings. The dataset from Craven (2001) was used to identify soundings that were classified as either TOR or SIG in both the observations and the reanalysis. There were 287 common soundings (taken at the same time, less than 100 nm apart, and both with the same classification). This comparison was simplified by the fact that both datasets used only 0000 UTC soundings, and the same spatial and temporal proximity criteria. There were 11 parameters (CAPE, CINH, and LCLH for both MU and ML parcels, sfc-3 km and 700-500 mb lapse rate, sfc-1 km and sfc-6 km bulk shear, and DCAPE), calculated in both studies, which were used to compare the characteristics of the two datasets.

The 287 common soundings were sorted by distance (between observation and nearest reanalysis sounding location) and divided into the lowest, middle, and highest thirds, so that each subset contained approximately the same number of soundings. Values of r^2 (square of correlation coefficient) and root-mean-square error (RMSE) were computed between each distance group of observed sounding-derived parameter values and the concurrent group of reanalysis values. RMSE was calculated with the following formula (from Wilks 1995):

$$RMSE = [(1/M)*\sum(y_m-o_m)^2]^{1/2}$$

where the sum is from m=1 to M, M is the number of values in each group, y_m is the m^{th} reanalysis value, and o_m is the m^{th} observed value. RMSE thus represented a typical error (reanalysis minus observed) magnitude for each distance group.

The distributions of observed versus reanalysis values for MU CAPE and sfc-6km bulk shear were plotted in Figures 1 and 2, respectively. The data points were colored to illustrate the distribution differences among the three distance groups. Values of r^2 for each group were also included. The 1:1 line on each plot represents perfect association (i.e., reanalysis value = observed value). The plots also contain three linear regression lines, colored to match the respective distance groups for which they were computed. The slope and intercept values used to plot these lines were obtained with Microsoft Excel. The linear regression lines represent a "best fit" for the distribution of each distance group. The spread of the points about each linear regression line is quantified by values of r^2. RMSE

represents the magnitude of the distance between the points and the "best fit" lines.

Table 3 contains the intercept and slope values for each of the linear regression lines plotted in Figures 1 and 2. These values indicate the bias of each distance group distribution, because they quantify the difference between each linear regression line and the 1:1 line (which has an intercept of zero and a slope of one). Intercept increased (away from zero) with distance for both parameters. Slope decreased (away from one) with increasing distance for both parameters, except that it slightly increased between the 0-37 nm and 37-50 nm groups for MU CAPE. These results indicated that the differences between observed and reanalysis values increased with distance. This was expected, because as distance increased between two sounding locations, there was a greater likelihood of environmental variation across that distance. Slope values for sfc-6 km bulk shear are closer to the 1:1 line than the MU CAPE slope values. This indicated that reanalysis values of sfc-6 km bulk shear match their concurrent observed values to a greater extent than MU CAPE values match their observational counterparts.

Table 2 displays r^2 and RMSE values for all 11 parameters. Low values of r^2 (below 0.07 for MU parcels and below 0.01 for ML parcels) indicated that the CINH parameters were unreliable (as compared to observations). The reanalysis values of these parameters disagreed with observations because the vertical resolution of the reanalysis soundings was not sufficient to detect the inversions that were critical to CINH calculations. MU LCLH also appeared to be unreliable

(with r^2 values from 0.193 to 0.308 and RMSE values from 743.8 to 884.2 m agl), which may be due to the fact that some reanalysis soundings indicated problems with low-level moisture (i.e., erroneous dewpoints). While the differences between reanalysis and observed dewpoints were usually relatively small (and in many cases almost nonexistent), even a minor error could have caused a significant change in LCLH values for the MU (usually low-level) parcel. However, the moisture problems were generally confined to the lower and upper (around and above the tropopause) levels of the atmosphere, and to only a few data points (if any) for most soundings. This was reflected in the r^2 values for ML LCLH (with r^2 values from 0.515 to 0.641 and RMSE values from 340.2 to 547.9 m agl), which were computed for 100 mb mean parcels. The lowest r^2 value for ML LCLH is still better (higher) than the highest r^2 value for MU LCLH, and the highest RMSE value for ML LCLH is better (lower) than the lowest RMSE value for MU LCLH. In general, it can be concluded that ML parcel parameters (which were more resistant to small numbers of erroneous data points) were more reliable in the reanalysis than were their MU parcel counterparts. An exception to this rule was CAPE.

MU and ML CAPE had almost identical r^2 values (with differences of less than 0.025) at each distance interval, although the RMSE values were lower for ML CAPE than MU CAPE at all distances. These r^2 values were unexpected, given the low-level moisture problems that contributed to the differences between the ML and MU LCLH values. However, CAPE was more resistant to such dewpoint discrepancies than LCLH (because it was computed using data from

many vertical levels). Both MU and ML CAPE were characterized by decreases in r^2 (from nearly 0.50 to around 0.25) and increases in RMSE (from 866.3 to 1387 J/kg for ML and from 1112 to 1886 J/kg for MU) as distance increased. This trend was expected, due to the increase in probability of environmental variation as distance increased.

DCAPE was different from the other two CAPE parameters. Its r^2 values were roughly constant between the lowest and highest distance groups (0.439 and 0.431, respectively), with only a small increase (to 0.482) in the middle distance group. Its RMSE values displayed a similar pattern. While DCAPE had a lower r^2 value than the other two CAPE parameters in the smallest distance group, it had larger values at larger distances. This implied that DCAPE had the same overall reliability (as compared to observations) as MU and ML CAPE, but that it did not suffer from environmental variations in the same way that the others did. This was due to the fact that DCAPE was calculated for a mid-level (rather than low-level) parcel. Atmospheric conditions at mid-levels are more accurately forecast by models than conditions at lower levels. Thus, forecast and verification differences are expected to be larger at lower levels. This derives from the fact that mid-level conditions vary on larger spatial and temporal scales. Also, low-level conditions are affected by many more factors (e.g., surface fluxes, vegetation, orography) than mid-levels, some of which are below the resolution of most models. The observation that r^2 values for DCAPE were nearly constant over distance supported that mid-level conditions vary on larger spatial scales.

Values of r^2 for 700-500 mb lapse rate increased, then decreased, as distance increased. Sfc-3 km lapse rate was characterized by r^2 values that increased as distance increased. Its r^2 values were much lower than those of 700-500 mb lapse rate in the smaller distance groups (as expected, given the more variable nature of low-level fields), but the two parameters had almost identical values (0.462 versus 0.464) in the largest distance group. RMSE values for both parameters behaved non-monotonically, just as the r^2 values did. It is possible that reanalysis lapse rates are unreliable in the reanalysis, because those parameters are each computed from only two temperature values, and are therefore highly susceptible to isolated erroneous data points. While their r^2 and RMSE values are non-monotonic, the r^2 values for 700-500 mb lapse rate are reasonable at smaller distances (0.589 for 0-37 nm and 0.678 for 37-50 nm). Nevertheless, it was concluded that lapse rates must be used with some measure of caution.

The two bulk shear parameters displayed better agreement between the reanalysis and observed values. Sfc-6 km bulk shear had the highest r^2 values (overall) of any parameter. The sfc-6 km r^2 values were higher than sfc-1 km because mid-level conditions varied on larger scales and were better predicted by the reanalysis model. Both shear parameters had r^2 values that decreased as distance increased. This is expected, due to environmental variation. The RMSE values for these parameters were non-monotonic. However, those values were roughly constant at all three distances for both parameters, so the small

differences between distance group RMSE values for each parameter could have been due to statistical anomaly.

Overall, most of the eleven parameters demonstrated only moderate agreement between reanalysis and observed values. The only r^2 values greater than 0.60 (generally characteristic of marginal agreement) corresponded to sfc-6 km bulk shear in the two smaller distance groups, ML LCLH in the smallest distance group, and 700-500 mb lapse rate in the middle distance group. MU and ML CINH were not reliable for this analysis, and any conclusions drawn from them were given little to no weight. MU LCLH was also determined to be generally unreliable. The other eight parameters displayed only marginal to below average agreement (in terms of r^2 values) between reanalysis and observed values. Many parameters exhibited non-monotonic values of r^2 and RMSE as distance was increased.

Despite the fact that the reanalysis soundings appeared to show only moderate agreement with observations, most parameters were not completely unreliable. Also, the results of the comparison analysis (summarized in Table 3) provided useful insight into the reliability of different kinds of reanalysis parameters. It was concluded that wind-related values agreed better with observations than thermodynamic parameters. Parameters related to mid-level conditions performed better than those calculated from near-surface data. Also, those parameters computed from data at many vertical levels (e.g., ML parcels and some CAPE values) appeared to be more reliable than those that were not (e.g., MU parcels and lapse rate values). Parameters used in the reanalysis dataset

that were not computed in the observed dataset could not be compared through correlation coefficients. However, their reliability can be inferred through the observations that were made of these 11 parameters.

B. Examination of the discrimination utility of the reanalysis soundings

All 37 parameters in the reanalysis sounding dataset were analyzed to see if they were capable of discrimination between different types of severe weather events (hereafter, "events" refers to proximity soundings associated with events). Single-parameter discrimination was investigated with cumulative distribution functions (CDFs). These plots displayed the values of a parameter against the percentage of all data points that were equal to or less than each value. Plots of CDFs for two distributions (corresponding to two types of events) revealed differences between those distributions. The dashed lines in Figure 7 indicate that 50% of the SIG sounding values of sfc-6 km bulk shear were less than or equal to 40 kt, while only a little more than 20% of TOR sounding values were less than 40 kt. Also, 51 kt of sfc-6 km bulk shear corresponded to the median (cumulative probability = 0.5) of the TOR distribution, but corresponded to the 75th percentile of the SIG distribution. Thus, it can be seen that the distance between the two curves (in this case, TOR and SIG) is proportional to the difference in the distribution of values for those two datasets. Also, the curve that is "below" the other (TOR in Figure 7) is characterized by higher values (in this case, of sfc-6 km bulk shear), since its median value is higher.

Multiple-parameter discrimination was investigated through the use of scatterplots. Initially, only qualitative observations were made of the differences between distributions of severe weather type subsets. Once those combinations were identified which showed substantial discrimination, a more quantitative analysis was performed, using 2x2 probability tables (and associated statistical quantities) calculated from each set of distributions.

Before TOR and SIG events could be compared, it was important to establish that the reanalysis data were capable of discrimination between "severe" (TOR+SIG+SVR) and NON events, and between "significant severe" (TOR+SIG) and SVR events. Many previous studies (e.g., Craven 2001) identified CAPE-shear combinations as good discriminators between these types of events. Figure 3 displays a plot of log(MU CAPE) versus log(sfc-6 km bulk shear) for TOR+SIG and SVR events. All log(MU CAPE) values were set to 0.0001 in cases where MU CAPE = 0. Upon inspection, TOR and SIG data points were clustered in the high-CAPE, high-shear region of the plot, whereas the distribution of SVR values was more diffuse. It was apparent that this combination of parameters demonstrated discrimination between TOR+SIG and SVR events.

To quantify this discrimination, a line was applied to the plot that transected the data distributions, so that the components of a 2x2 forecast table (see Table 4) could be calculated. All points above (below) the line were treated as yes (no) TOR+SIG forecasts. This line (also plotted in Figure 3) was given a fixed slope value of –0.25 (based upon inspection), and the intercept was varied

so that the best threshold line could be determined. A linear regression line might have resulted in a more accurate measure of discrimination, but it was believed that the difference would be marginal. Thus, slopes based upon inspection were used for all discrimination scatterplot analyses. The probability of detection (POD) and probability of false detection (POFD) were computed for each line (each combination of slope and intercept). The formulas used to compute these quantities were included in Table 4. These were also the two components of the Hanssen-Kuipers (H-K) skill score, which is simply POD-POFD, and defined the two axes of a relative operating characteristic (ROC) diagram (Mason 1982). A ROC curve represents the skill of a forecast system – in this case, the performance of two parameters (MU CAPE and sfc-6 km bulk shear) – in differentiating between two types of events (in this case, TOR+SIG and SVR). The area under a ROC curve (calculated in this study with the trapezoidal rule) represents the level of skill. An area of 0.5 (for a straight diagonal line on the diagram) indicates no skill.

A ROC curve (Figure 4) for the MU CAPE versus sfc-6 km bulk shear scatterplot of TOR+SIG and SVR distributions was constructed for values of intercept from 1 to 2.7 (at intervals of 0.1), with the slope held constant at –0.25. The POFD (x) and POD (y) values are listed in Table 5. The upper-right dashed line in Figure 4 corresponds to an intercept value of 2.0, for which the POFD was 0.762 and the POD was 0.903. This means that when a line with a slope of –0.25 and an intercept of 2.0 was applied to the scatterplot in Figure 3, 90.3% of SIG+TOR events were above the line, and 76.2% of SVR points were above the

line. Similarly, the lower-left dashed line plotted in Figure 4 corresponds to an intercept of 2.23, for which the POFD was 0.5 and the POD was 0.75. Thus, it is apparent that as intercept was increased (threshold line was moved "up" the scatterplot), POD and POFD both increased, because fewer events were above the line.

The ROC area of 0.585 for the curve plotted in Figure 4 indicated a marginal level of skill in discriminating between TOR+SIG and SVR events. The point furthest toward the upper-left on the diagram represented the highest H-K score (because the POD-POFD value was greatest there). Thus, it can be seen that a ROC with more points toward the upper-left on the diagram represented greater forecast skill. They resulted in higher ROC area values, and higher H-K (POD-POFD) scores. The intercept that corresponded to the most upper-left point in Figure 4 represented the line that best discriminated (as defined by the H-K score) between TOR+SIG and SVR events. In this case, the optimal intercept value was the one that corresponded to a POFD value of approximately 0.4. A plot (Figure 5) of intercept versus H-K skill score confirmed this, and illustrated the unimodal nature of discrimination (forecast skill) for the scatterplot in Figure 3. The highest score (0.269) resulted from an intercept of 2.3, which corresponded to a POFD value of 0.395 (in Table 5).

These results implied that the best line for discrimination on the plot of TOR+SIG and SVR MU CAPE versus sfc-6 km bulk shear (Figure 3), purely in terms of highest H-K score, had a slope of –0.25 and an intercept value of 2.3. However, that line resulted in a POD value of only 0.664 (see Table 5), so only

66.4% of TOR or SIG soundings were "correctly forecast". It is often desirable to have a higher POD, so an additional requirement was set that this value be at least 0.90. The intercept that resulted in the highest H-K skill score *and* a POD of at least 0.9 was 2.0. Thus, the line plotted in Figure 3 had a slope of –0.25 and an intercept of 2.0. This resulted in a POD of 0.903, a POFD of 0.762, and a H-K score of 0.140. If all points below the line were eliminated, 9.7% of TOR+SIG soundings and 23.8% of SVR soundings would be discarded.

The combination of MU CAPE and sfc-6 km bulk shear revealed some discrimination utility, and had been shown in previous studies to differentiate between significant (TOR+SIG) and non-significant severe (SVR) environments. However, it was still of interest to examine the results of the same analysis performed on other combinations of parameters. Several thermodynamic parameters were paired with deep shear, and the ROC area and maximum H-K skill score were calculated for each combination (with a slope value determined by inspection for each scatterplot). The results (displayed in Table 6) indicated that CAPE performed better with deep shear than the other parameters. This combination was also used to examine discrimination between "severe" (TOR+SIG+SVR) and NON events.

In the course of this analysis, the question arose of how much impact slope value actually had on a resulting ROC diagram. To investigate this, the MU CAPE versus sfc-6 km bulk shear scatterplot was reanalyzed; the slope was varied, and ROC curves were constructed for each slope value. The results (displayed in Figure 6 and Table 7) indicated that slope value had very little effect

on the ROC, except when it approached zero, because this was equivalent to the examination of the distribution of only one parameter.

The threshold of discrimination established for TOR+SIG versus SVR events (slope of –0.25 and intercept of 2.0 on a plot of log[MU CAPE] versus log [sfc-6 km bulk shear]) was applied to NON sounding data. The resulting POD was 0.242, which indicated that, if NON sounding values were plotted in Figure 3, only 24.2% would lie above the line. Thus, the elimination of points below the established threshold would discard 9.7% of TOR+SIG soundings, 23.8% of SVR soundings, and 75.8% of NON soundings. It was concluded that the reanalysis sounding data are capable of discrimination between severe and non-severe events *and* between significant severe and non-significant severe events. With this determination made, the more interesting issue of TOR versus SIG discrimination was examined.

The investigation of the utility of the reanalysis data in TOR versus SIG discrimination initially used CDFs. Examination of these plots (constructed for all 37 parameters) resulted in qualitative observations as to which parameters exhibited significant differences between the two event type distributions. Some revealed good discrimination (see Figure 7, sfc-6 km bulk shear, Figure 8, sfc-1 km bulk shear, Figure 9, ML LCLH, and Figure 10, sfc-1 km mean RH), while some did not (see Figure 11, MU CINH), and others showed moderate potential as discriminators (see Figure 12, MU CAPE). "Good" discriminators were generally defined as those whose CDFs were characterized by cumulative probability (y) differences of at least 0.2 over at least one-third of the range.

"Moderate" discriminators were generally defined as those whose CDFs were characterized by cumulative probability (y) differences of at least 0.1 over at least one-half of the range. CDFs that did not meet either of these standards were determined to be "poor".

Among thermodynamic parameters, MU and ML CINH, Cap, LI at 500 mb, MU LFCH, MU 3 km CAPE, ML CAPE, all six lapse rates based above the surface, 2-4 and 4-6 km mean RH, and 100 mb mean mixing ratio all displayed poor discrimination. DCAPE, all three surface-based lapse rates, MU CAPE, MU LCLH, and ML LFCH demonstrated moderate capability, and MU and ML mean RH LCL-LFC, ML LCLH, ML 3 km CAPE, and sfc-1 km and sfc-2 km mean RH indicated good discrimination. Because it had been observed that the reanalysis model does not predict humidity as well as most other fields, it was not unexpected that many thermodynamic parameters (most of which were dependent on dewpoint calculations) were virtually incapable of discrimination. In addition, many of the thermodynamic processes believed to be responsible for tornadogenesis are generally focused in the lower levels of the atmosphere, so it was not surprising that all mid-level thermodynamic parameters (MU and ML LI at 500 mb, six lapse rates based above the surface, and two mean RH values calculated above the surface) performed poorly. The surface-based lapse rates performed better because they were based on temperature rather than dewpoint. It was expected, based on the results of previous studies, that LCLH should be a good discriminator, and it was in the case of ML LCLH. MU LCLH performed more poorly because the reanalysis values of that parameter were less reliable

than its ML counterpart. Low-level moisture content should also have been a good discriminator, and this was reflected by the two surface-based mean RH parameters. Although the reanalysis model has some problems with humidity, these were mean values (thus resistant to isolated erroneous data points), so they were more accurately predicted than single points. The fact that 100 mb mean mixing ratio performed poorly was due to the fact that it was based solely on humidity, whereas the RH parameters were based upon both dewpoint and temperature (which was a more accurately forecast field).

All six wind-related parameters (sfc-1 km and sfc-3 km SRH, sfc-1 km, sfc-3 km, and sfc-6 km bulk shear, and BRNShear) indicated good discrimination. This was not unexpected, because many previous studies (e.g., Johns et al. 1993) have identified shear and SRH-related parameters as particularly good discriminators between storm types (due to the shear-induced rotational properties of tornadic thunderstorms). In addition, earlier analysis showed that the sfc-1 km and sfc-6 km bulk shear values in the reanalysis correlated well with those from observations, so it was concluded that the group characteristics of these parameters would have been maintained in the reanalysis.

A more useful method of discrimination between TOR and SIG events was the analysis of combinations of multiple parameters. These distributions were most easily viewed using scatterplots, as was earlier done for TOR+SIG versus SVR and TOR+SIG+SVR versus NON events. The combinations used for scatterplot construction each included one thermodynamic parameter and one wind-related value. Only those parameters that indicated at least moderate

discrimination utility (through CDF examination) were used. The exception was that the combination of MU CAPE and sfc-6 km bulk shear was examined, so that the results could be compared with those previously obtained for the first two levels of discrimination.

The scatterplot of log(MU CAPE) versus log(sfc-6 km bulk shear) for TOR and SIG events (Figure 13) was analyzed through the same methods as were used for TOR+SIG versus SVR and TOR+SIG+SVR versus NON discrimination. However, it was apparent that a slope of –0.25 was far from ideal for this distribution, and upon inspection, a slope of –0.10 resulted in better discrimination. Output statistics were computed for both slope values. The ROC curves for both slopes, along with those for the first two levels of discrimination, were plotted in Figure 14. It was apparent from the line with a –0.10 slope that a higher forecast skill resulted, because its ROC curve was further toward the upper-left of the plot than was its –0.25 counterpart. This result is consistent with the previously observed effect of varying slope on ROC area (see Figure 6), because one of the slope values (-0.10) was close to zero. The ROC area and maximum H-K for each curve in Figure 14 are displayed in Table 8. The combination of MU CAPE and sfc-6 km bulk shear revealed discrimination skill for TOR versus SIG events, at the same skill level as for discrimination between TOR+SIG versus SVR events. However, this skill level was below that of TOR+SIG+SVR versus NON events. Because the CDFs of MU CAPE for TOR and SIG events (Figure 12) showed only moderate discrimination utility, this result was not unexpected.

Several combinations of bulk shear (sfc-1 km and sfc-6 km) and thermodynamic parameters (sfc-1 km and sfc-2 km mean RH, ML LCLH, and ML 3 km CAPE) for TOR versus SIG distributions were analyzed using the same methods. The ROC area and maximum H-K score were computed for each scatterplot. The results were summarized in Table 9. Combinations that involved shallow (sfc-1 km) bulk shear performed better than those that involved deep (sfc-6 km) shear, regardless of the thermodynamic parameter. The ROC curves for the four shallow-shear scatterplots are displayed in Figure 15. All combinations (especially sfc-1 km bulk shear paired with low-level mean RH and ML LCLH) demonstrated significant discrimination. ROC areas exceeded 0.63 (and were as high as 0.669) and maximum H-K values were greater than 0.39 (and were as high as 0.526). The scatterplot of TOR and SIG data for sfc-1 km mean RH versus sfc-1 km bulk shear (along with the line corresponding to the maximum H-K score for that combination) is shown in Figure 16, and the corresponding ROC diagram is plotted in Figure 17. The discrimination between TOR and SIG events was evident in these figures, as there was a noticeable difference between the distribution of TOR and SIG data points, and the ROC diagram showed a good level of skill.

The results of this analysis indicated that the best discriminators between TOR and SIG events in the reanalysis soundings were combinations of sfc-1 km bulk shear with sfc-1 km mean RH, sfc-2 km mean RH, and ML LCLH. The ROC areas for these combinations had values between 0.658 and 0.669. Each had a maximum H-K score greater than 0.5 (in Table 9).

A linear discriminate analysis (LDA) was performed on the TOR and SIG event data. The analysis was conducted through the CLEAVER 1.0 website (Stanford 2002). A thorough description and methodology of LDA is provided in Wilks (1995). In the context of this study, LDAs were used to identify sounding-derived parameters in the reanalysis that could best distinguish between TOR and SIG events. Missing values were set to zero and the data were column normalized using the analysis options on the Cleaver website. Only the weight values output from the LDA (for each parameter) were investigated. These weights represented the relative importance of each parameter in discriminating between the two datasets (in this case, TOR and SIG events).

Initial LDA results indicated that all lapse rate parameters were significantly more important (had larger weights) than all other variables in TOR versus SIG discrimination. The weights were so much higher for lapse rates than other parameters, that the results indicated that lapse rate was the only piece of information needed to determine whether a sounding was TOR or SIG. It was concluded that the lapse rate distributions must have had some characteristic that resulted in an inaccurate representation of their discrimination utility. It is likely that the bimodal nature of these parameters was responsible for the substantially large weights. Also, lifted index parameters received weights higher than all other parameters (except lapse rates). Again, lifted index values were removed from the LDA, because they appeared to have distribution characteristics that led to unreliable results.

Another LDA was performed for TOR and SIG events, with all lapse rate and lifted index parameters excluded. The resulting weights for the top ten parameters (ranked by absolute value of weight) are listed in Table 10. LDA results indicated that ML LCLH and sfc-1 km mean RH were the most important parameters to discriminate between TOR and SIG events. Sfc-1 km bulk shear and sfc-2 km mean RH were also close to the top of the list (Table 10). These results agree very well with those previously obtained from scatterplot analysis.

The LDA helped to confirm the results of earlier analyses. However, scatterplots were considered more useful, because the LDA based its discrimination threshold on the minimization of total error. If the threshold is viewed as a yes/no forecast basis (as was done in the scatterplot analyses), the LDA determined the threshold that resulted in the fewest total number of incorrect forecasts. With scatterplot analyses, the results could be examined from a more operational point of view. For example, a threshold could be identified which resulted in the fewest total incorrect forecasts with a minimum POD, or in the highest H-K score. Nevertheless, the fact that the parameters determined to be the best discriminators through scatterplot analysis agreed with the results of the LDA demonstrates the reliability of these results.

CHAPTER VI

CONCLUSIONS

To expand the volume, resolution, and global coverage of thermodynamic sounding data available for research, soundings were constructed from the NCEP/NCAR reanalysis. The reanalysis dataset consists of a high-resolution, three-dimensional picture of the atmosphere at six-hour intervals during a period of more than 50 years. The reanalysis data were obtained in the form of spectral coefficients on sigma levels. Programs were designed and implemented to translate the data onto a grid and derive the necessary atmospheric variables from those in the initial data files. Soundings were constructed and classified according to the severe weather types they were in proximity to. The SHARPTAB software was used to derive parameters from each sounding.

Once a dataset of sounding-derived parameters had been created, it was compared to concurrent observations from Craven (2001). Some reanalysis parameters agreed relatively well with their observed counterparts. Wind-derived parameters, those computed using mean layer parcels, and those incorporating data from many levels (e.g., CAPE) were more reliable (compared to observations). CINH parameters, as well as some using the most unstable parcel (e.g., MU LCLH) were much less reliable; they did not agree well with observational data.

Other analyses examined the utility of reanalysis soundings to discriminate between different types of severe weather events. Figure 14 and Table 8 demonstrated that the combination of reanalysis MU CAPE and sfc-6 km

bulk shear was able to discriminate between non-severe and severe events, as well as between non-significant severe and significant severe events. The discrimination between severe and non-severe environments was particularly good, with a ROC area of 0.673 and a maximum H-K score of 0.543. Severe events were easier to distinguish from non-severe events than were significant events from non-significant severe events. This result derives from the fact that the "non-severe" subset included data for many cases in which convection did not occur. The difference between severe and significant severe environments was more difficult to identify, as evidenced by a lower level of skill in the reanalysis discrimination between TOR+SIG versus only SVR events. In this case, analysis of the combination of MU CAPE versus sfc-6 km bulk shear produced a ROC area of 0.585 and a maximum H-K score of 0.269.

Several parameters indicated that the reanalysis also discriminated well between TOR and SIG events. In particular, low-level thermodynamic parameters (e.g., sfc-1 km mean RH and ML LCLH) and those related to the wind field (i.e., bulk shear and SRH) performed very well. Several combinations of low-level shear and thermodynamic parameters were analyzed. All demonstrated a significant level of discrimination. The best performers were combinations of sfc-1 km bulk shear with sfc-1 km mean RH, sfc-2 km mean RH, or ML LCLH. These three distributions were characterized by ROC areas in excess of 0.65 and maximum H-K scores above 0.50. This result is a significant finding, because this high level of discrimination between TOR and SIG events is more challenging for models than differentiating between significant and non-

significant severe environments. LDA weights indicated that ML LCLH, sfc-1 km mean RH, sfc-1 km bulk shear, and sfc-2 km mean RH were important parameters for TOR versus SIG discrimination. These results agreed very well those obtained through scatterplot analysis.

In addition to the discrimination capabilities of the reanalysis data, parameters and combinations that performed the best in this analysis were consistent with those established by previous proximity sounding studies (e.g., Craven 2001) as being particularly useful. Several works identified CAPE-shear combinations as being the best discriminators between severe/non-severe and significant/non-significant environments. Craven (2001) also concluded that sfc-1 km bulk shear and ML LCLH were the most useful parameters for his observational dataset. His results agree well with the results of this study.

Issues that concern possible sources of error in this analysis include the process of sounding construction (including variable derivations), which could introduce interpolation error, particularly when dealing with moisture (i.e., specific humidity and dewpoint). Specific humidity values in the reanalysis output were occasionally negative; they were set equal to 0. In addition, the calculation of dewpoint is empirical in nature. Thus, relative humidity values greater than 100% were set equal to 100%. Both "forcing" measures were indicative of problems with dewpoint prediction and calculation. Therefore, moisture-related fields and the conclusions drawn from them must be used with caution.

While previous proximity sounding studies introduced methods of quality control to eliminate misrepresentative data, no such measures were used in this analysis. Thus, to increase the size of the dataset and examine the performance of reanalysis parameters, soundings were undoubtedly included which, although fitting the criteria for proximity, did not represent the environment in which associated severe weather events formed. It was impossible to quantify the extent to which these inclusions affected the results, without repeating the analysis with measures of quality control and comparing the output to that obtained in this study.

A large number of soundings (197,100) were constructed for the purposes of this study. Only 159 of these soundings were classified as TOR, which should be a large enough dataset to draw valid conclusions. However, it is possible that a small number of misrepresentative soundings could have had a measurable effect on the group characteristics of TOR soundings.

It is possible that some of the results of scatterplot discrimination analysis (e.g., ROC area and H-K scores) are unstable, that is, that the results obtained are not representative of the actual discrimination utility of each combination of parameters. Performing an identical analysis on subsets of the data obtained through resampling could reveal information about the stability of these results.

Despite the inherent risks of using model data, it was clear that the NCEP/NCAR reanalysis presented a significant opportunity to expand the volume and resolution of sounding data available for research. The goals of this study were achieved, and it was demonstrated that reanalysis proximity soundings are

capable of revealing information about severe weather environments. The reanalysis data will doubtless prove very useful in future endeavors.

There are many possible areas of future research. Only the eastern two-thirds of the continental United States was analyzed in this study. The reanalysis dataset makes possible, for the first time, high-resolution sounding analyses in remote regions (e.g., oceans or the poles) where observational data is limited or nonexistent. Analysis of sounding-derived parameters in these areas should yield useful results. Only three years of data (and only one time index per day) were analyzed. Thus, the behavior of the atmosphere at times other than near 0000 UTC or on very long (climatological) time-scales was not investigated. Parameter climatologies, the efficiency of the atmosphere, and temporal patterns (e.g., diurnal and annual cycles of parameters) can be investigated in future studies. Research in these areas should provide further insight into the atmospheric conditions that characterize the environments in which different types of severe weather occur.

Parameter (units)

MU CAPE (J/kg)
MU CINH (J/kg)
MU LCLH (m agl)
MU LFCH (m agl)
MU LI at 500 mb (dimensionless)
MU Cap (C)
MU 3 km CAPE (J/kg)
Mean RH MU LCL-LFC (%)
ML CAPE (J/kg)
ML CINH (J/kg)
ML LCLH (m agl)
ML LFCH (m agl)
ML LI at 500 mb (dimensionless)
ML Cap (C)
ML 3 km CAPE (J/kg)
Mean RH ML LCL-LFC (%)
Sfc-1 km SR helicity (m^2/s^2)
Sfc-3 km SR helicity (m^2/s^2)
Sfc-1 km bulk shear (kt)
Sfc-6 km bulk shear (kt)
Sfc-1 km lapse rate (C/km)
Sfc-2 km lapse rate (C/km)
Sfc-3 km lapse rate (C/km)
2 km-4 km lapse rate (C/km)
3 km-6 km lapse rate (C/km)
4 km-6 km lapse rate (C/km)
6 km-8 km lapse rate (C/km)
850-500 mb lapse rate (C/km)
700-500 mb lapse rate (C/km)
Sfc-1 km mean RH (%)
Sfc-2 km mean RH (%)
2 km-4 km mean RH (%)
4 km-6 km mean RH (%)
100 mb mean mixing ratio (g/kg)
Sfc-3 km bulk shear (kt)
MU BRNShear (J/kg)
DCAPE (J/kg)

Table 1. A list of the 37 sounding-derived parameters analyzed in this study.

Distance (nm)	r^2 0-37	RMSE 0-37	r^2 37-50	RMSE 37-50	r^2 50-82	RMSE 50-82
Number of soundings in distance group	95	95	95	95	97	97
MU CAPE	0.481	1112	0.445	1164	0.250	1886
MU CINH	0.069	63.05	0.009	108.8	0.047	62.23
ML CAPE	0.493	866.3	0.422	1009	0.234	1387
ML CINH	0.008	160.2	0.006	175.0	0.000	161.8
ML LCLH	0.641	340.2	0.515	498.0	0.522	547.9
MU LCLH	0.233	743.8	0.193	884.2	0.308	832.7
DCAPE	0.439	330.7	0.482	323.9	0.431	361.4
Sfc-3 km lapse rate	0.367	1.158	0.375	1.191	0.462	1.178
700-500 mb lapse rate	0.589	0.593	0.678	0.519	0.464	0.821
Sfc-1 km bulk shear	0.531	7.543	0.479	7.814	0.411	7.762
Sfc-6 km bulk shear	0.677	10.46	0.665	9.400	0.589	11.23

Table 2. The results of analysis of observed sounding-derived parameter values versus corresponding reanalysis values, for all three distance groups.

Parameter	Distance (nm)	Intercept	Slope
MU CAPE	0-37	372.8	0.586
MU CAPE	37-50	448.4	0.644
MU CAPE	50-82	983.7	0.286
Sfc-6 km bulk shear	0-37	8.824	0.832
Sfc-6 km bulk shear	37-50	11.64	0.747
Sfc-6 km bulk shear	50-82	15.64	0.661

Table 3. The intercept and slope values for the linear regression lines plotted in Figures 1 and 2 (scatterplots of observed versus reanalysis MU CAPE and sfc-6 km bulk shear, respectively), for all three distance groups.

	Observed Yes	Observed No
Forecast Yes	A	B
Forecast No	C	D

Probability of Detection (POD) = A/(A+C)

Probability of False Detection (POFD) = B/(B+D)

Hanssen-Kuipers (H-K) skill score = POD-POFD

Table 4. The 2x2 contingency table, and formulas for probability of detection, probability of false detection, and the Hanssen-Kuipers skill score.

Intercept	POFD	POD	H-K
1.0	0.996	0.999	0.00272
1.1	0.994	0.997	0.00321
1.2	0.991	0.997	0.00677
1.3	0.986	0.997	0.0103
1.4	0.979	0.995	0.0155
1.5	0.966	0.989	0.0227
1.6	0.950	0.982	0.0314
1.7	0.921	0.971	0.0498
1.8	0.881	0.956	0.0753
1.9	0.830	0.929	0.0991
2.0	0.762	0.903	0.140
2.1	0.673	0.859	0.186
2.2	0.546	0.790	0.244
2.3	0.395	0.664	0.269
2.4	0.228	0.469	0.241
2.5	0.0974	0.234	0.136
2.6	0.0255	0.0639	0.0383
2.7	0.00248	0.00588	0.00341

Table 5. The results of discrimination analysis of the scatterplot of MU CAPE versus sfc-6 km bulk shear for TOR+SIG versus SVR distributions.

Parameter 1	Parameter 2	ROC area	Maximum H-K
MUCAPE	Sfc-6 km bulk shear	0.585	0.269
MLCAPE	Sfc-6 km bulk shear	0.585	0.275
MULCLH	Sfc-6 km bulk shear	0.564	0.230
MLLCLH	Sfc-6 km bulk shear	0.566	0.232
Sfc-1 km mean RH	Sfc-6 km bulk shear	0.565	0.235
Sfc-3 km lapse rate	Sfc-6 km bulk shear	0.567	0.239
700-500 mb lapse rate	Sfc-6 km bulk shear	0.572	0.251

Table 6. The results of discrimination analysis of two-parameter scatterplots for TOR+SIG versus SVR distributions.

Slope	ROC area	Maximum H-K
-0.5	0.58	0.248
-0.3	0.584	0.269
-0.25	0.585	0.269
-0.2	0.585	0.264
-0.1	0.578	0.255
0.0	0.565	0.245

Table 7. The results of discrimination analysis of the scatterplot of MU CAPE versus sfc-6 km bulk shear for TOR+SIG versus SVR distributions, for various forecast threshold slope values.

Distributions	ROC area	Maximum H-K
TOR+SIG+SVR versus NON	0.673	0.543
TOR+SIG versus SVR	0.585	0.269
TOR versus SIG (slope = -0.25)	0.571	0.236
TOR versus SIG (slope = -0.10)	0.594	0.265

Table 8. The results of discrimination analysis of the scatterplot of MU CAPE versus sfc-6 km bulk shear for different levels of discrimination.

Parameter 1	Parameter 2	ROC area	Maximum H-K
Sfc-1 km mean RH	Sfc-1 km bulk shear	0.669	0.526
Sfc-2 km mean RH	Sfc-1 km bulk shear	0.661	0.513
ML LCLH	Sfc-1 km bulk shear	0.658	0.514
ML 3 km CAPE	Sfc-1 km bulk shear	0.658	0.489
Sfc-1 km mean RH	Sfc-6 km bulk shear	0.633	0.437
Sfc-2 km mean RH	Sfc-6 km bulk shear	0.633	0.449
ML LCLH	Sfc-6 km bulk shear	0.634	0.421
ML 3 km CAPE	Sfc-6 km bulk shear	0.632	0.398

Table 9. The results of discrimination analysis of two-parameter scatterplots for TOR versus SIG distributions.

Parameter	ABS(weight)
ML LCLH	1.607
Sfc-1 km mean RH	1.120
Sfc-1 km bulk shear	0.688
MU BRNShear	0.686
Sfc-2 km Mean RH	0.685
DCAPE	0.612
Sfc-3 km bulk shear	0.605
Sfc-1 km SR helicity	0.569
ML 3 km CAPE	0.565
Sfc-3 km SR helicity	0.560

Table 10. The results of linear discriminate analysis of sounding-derived parameters for TOR versus SIG distributions.

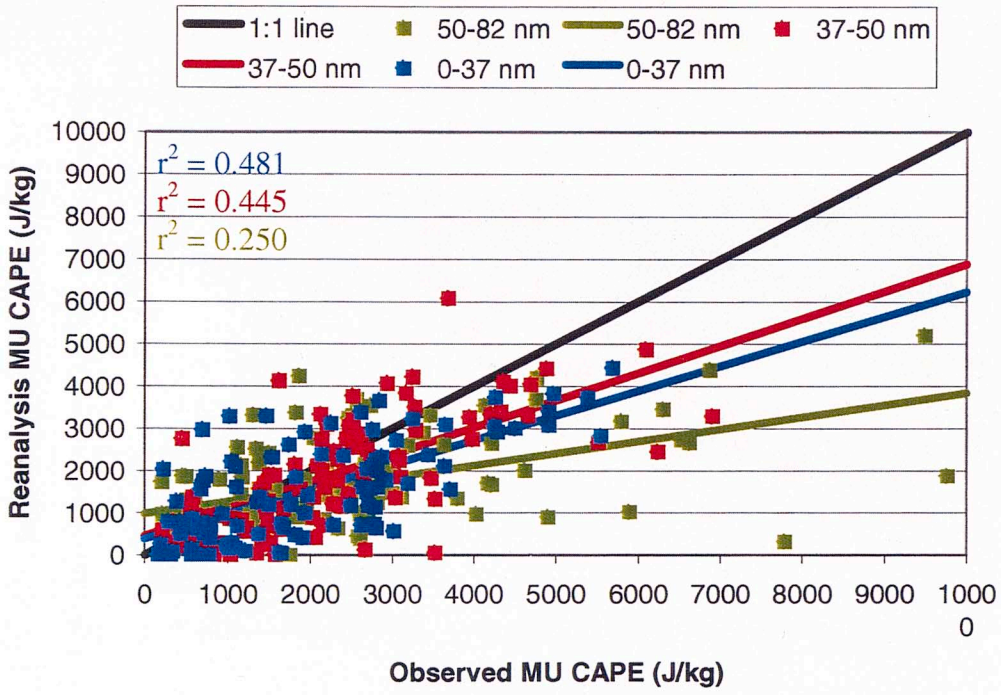

Figure 1. Observed versus reanalysis values of MU CAPE, with r^2 values (from Table 3) and linear regression lines (from Table 4) for all three distance groups.

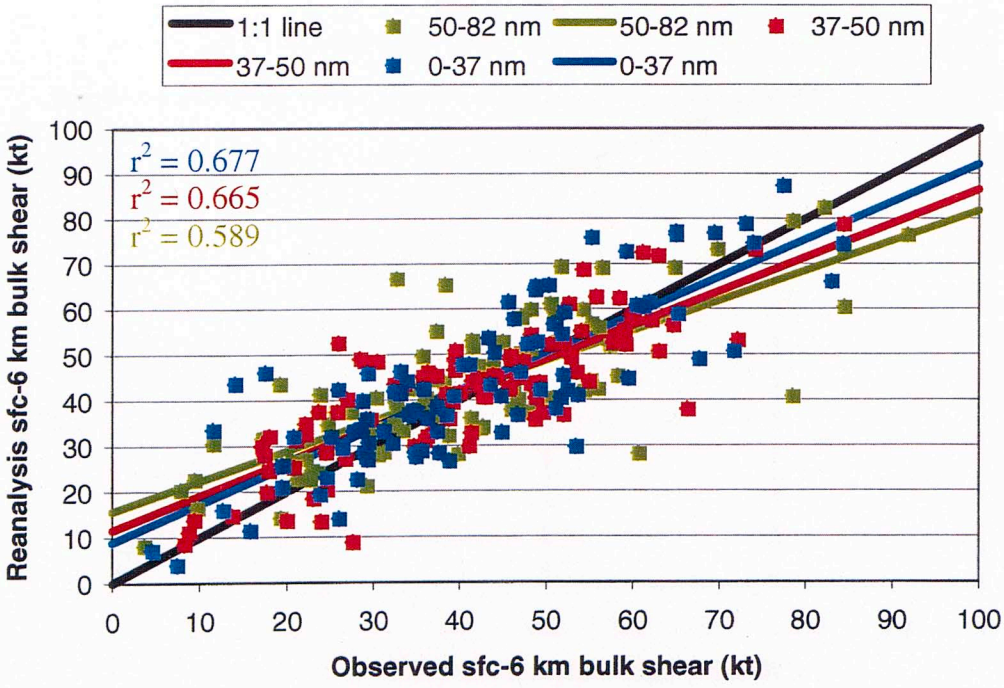

Figure 2. As in Figure 1, except for sfc-6 km bulk shear.

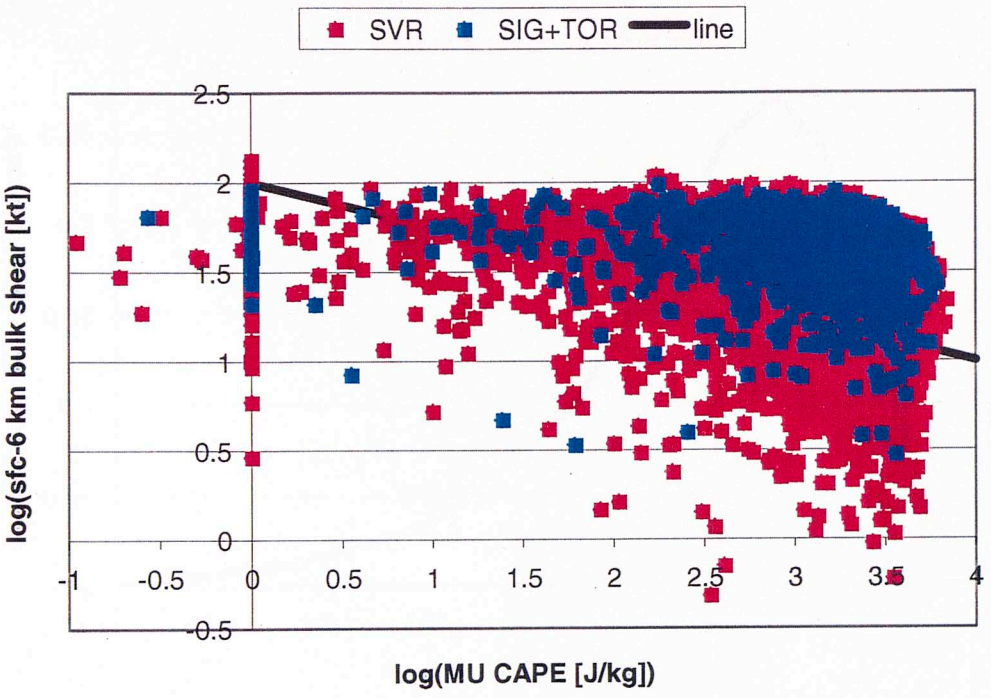

Figure 3. Log(MU CAPE) versus log(sfc-6 km bulk shear) for TOR+SIG and SVR distributions. The line is an example of a forecast threshold.

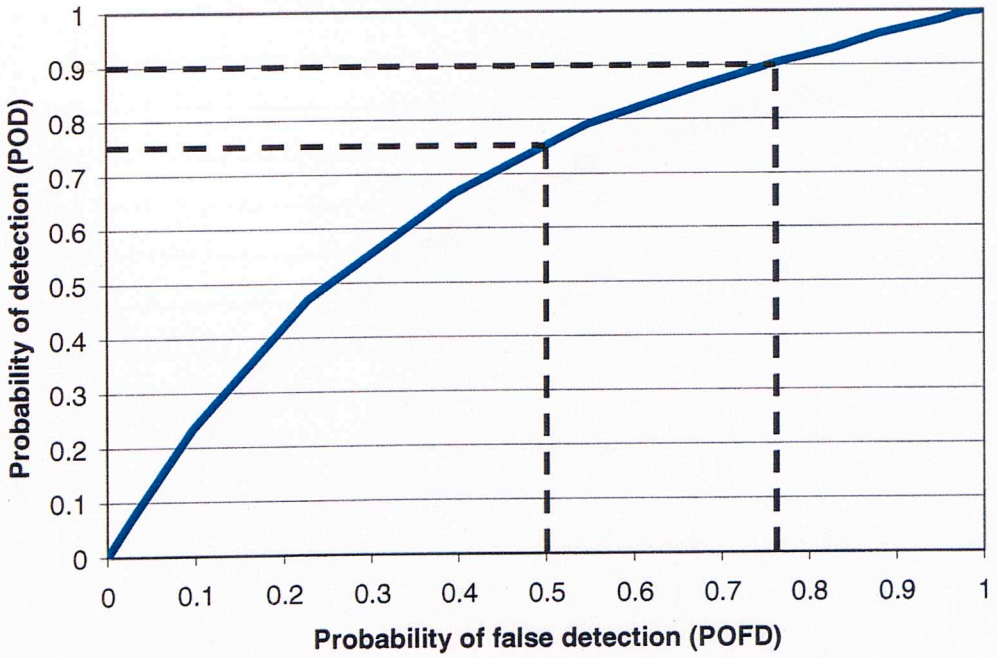

Figure 4. Relative operating characteristic diagram corresponding to Figure 3. The dashed lines are included solely for illustration of text examples.

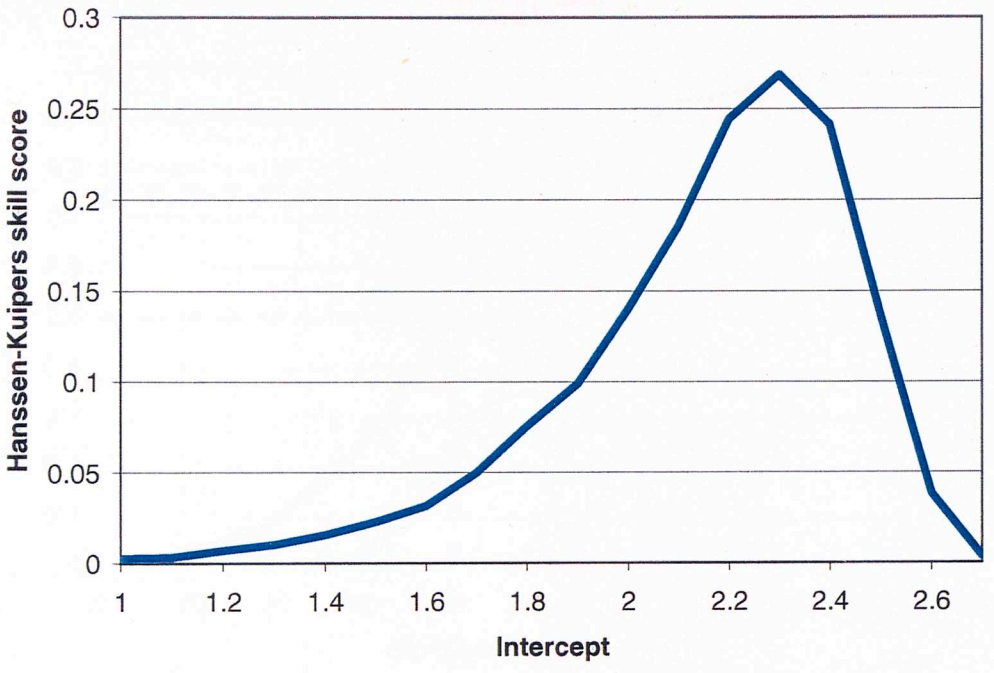

Figure 5. Intercept versus Hanssen-Kuipers score corresponding to Figure 4.

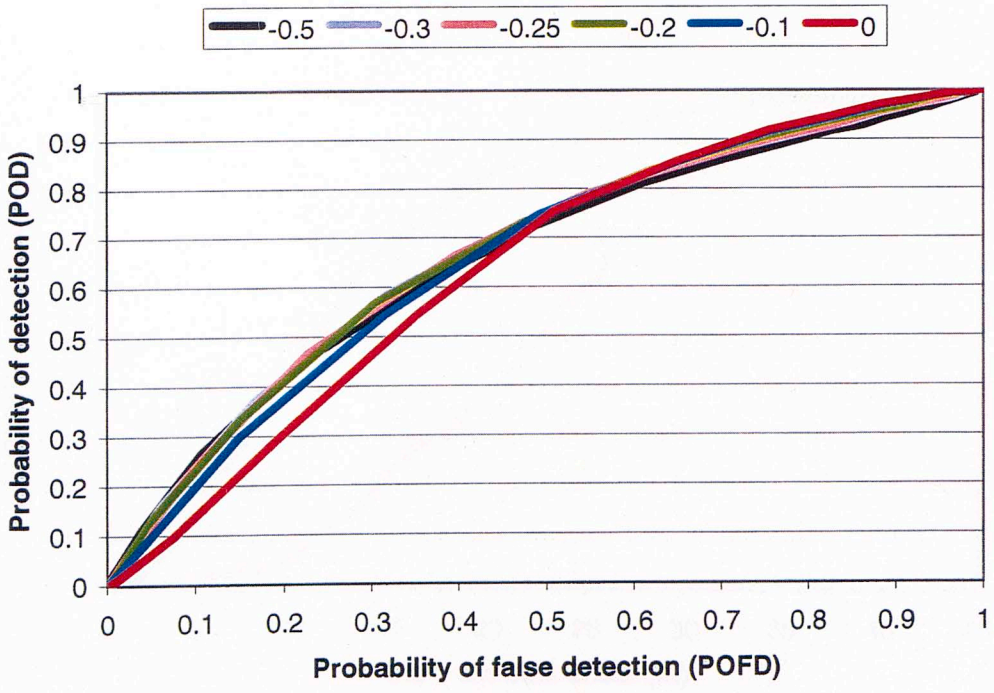

Figure 6. As in Figure 4, except for multiple values of forecast threshold slope.

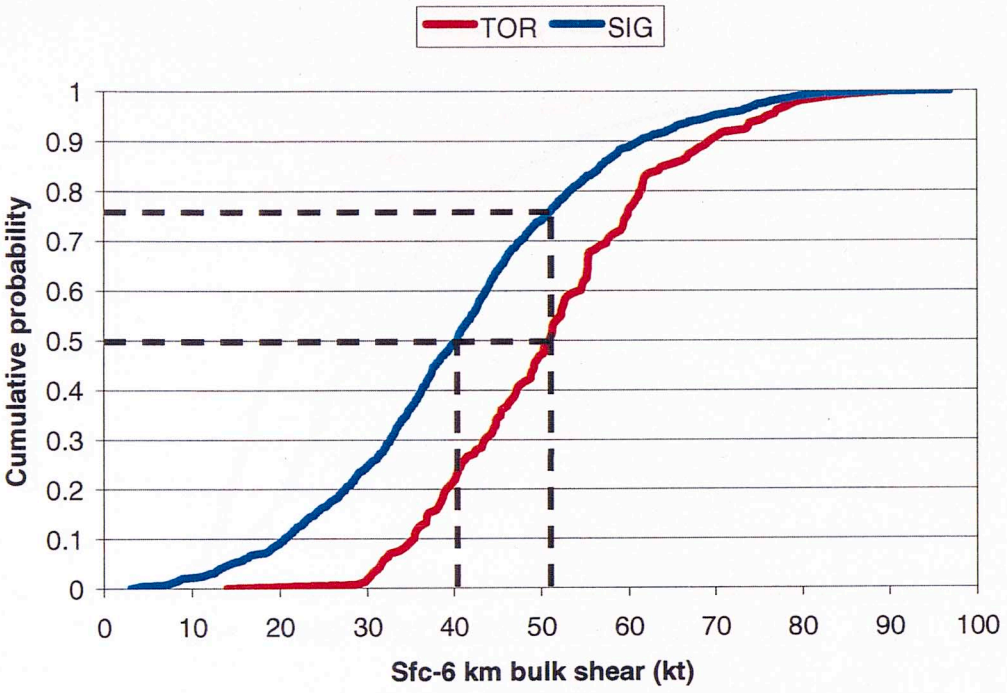

Figure 7. Cumulative distribution functions of sfc-6 km bulk shear for TOR and SIG soundings. The dashed lines are included solely for illustration of text examples.

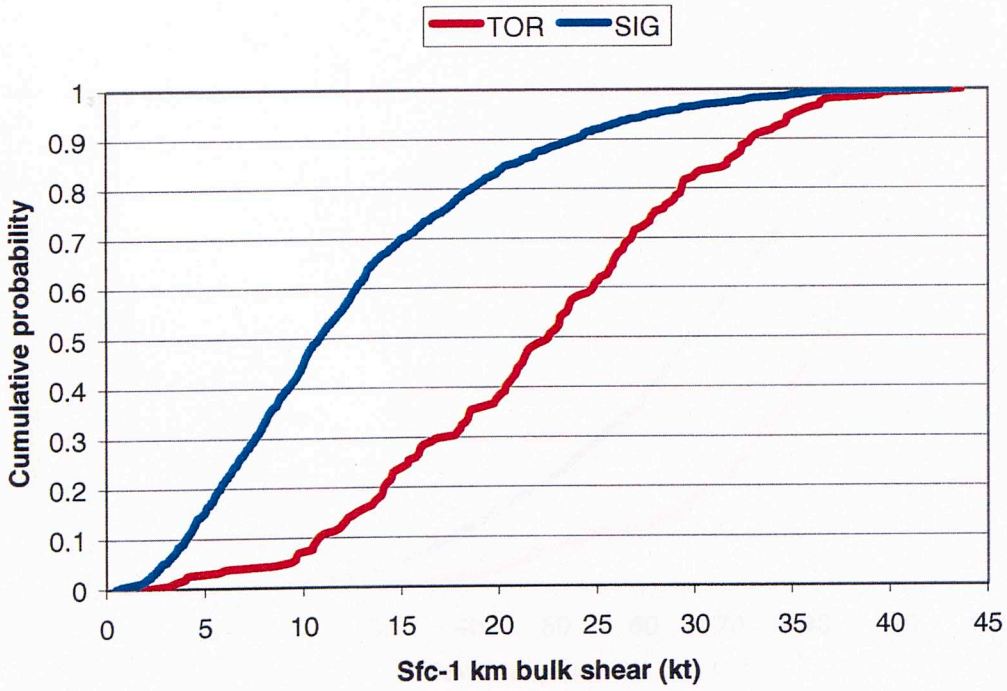

Figure 8. As in Figure 7, except for sfc-1 km bulk shear.

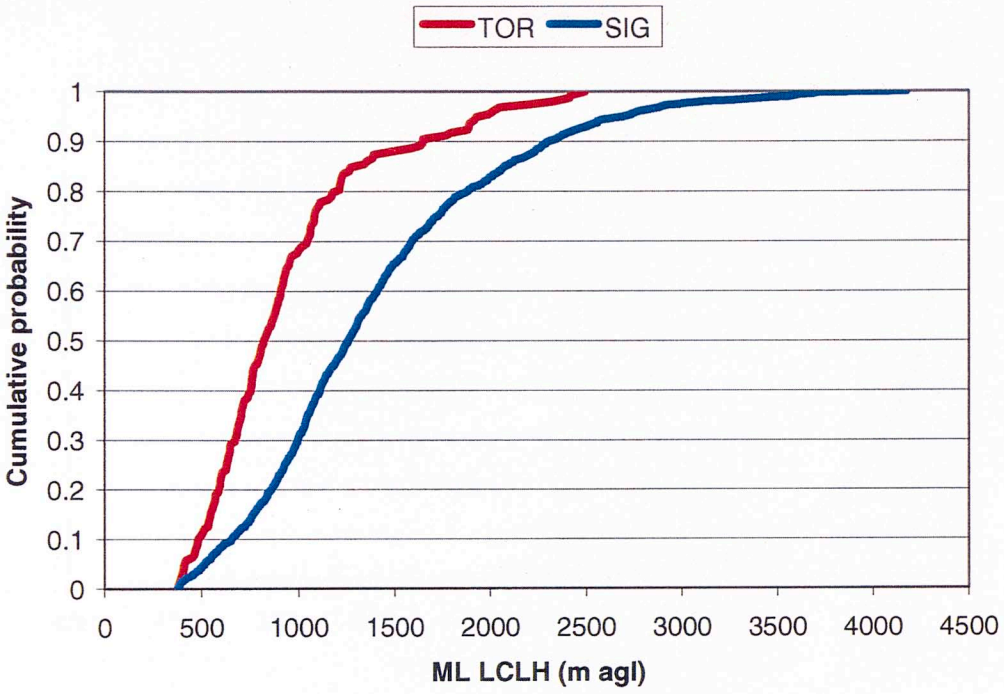

Figure 9. As in Figure 7, except for ML LCLH.

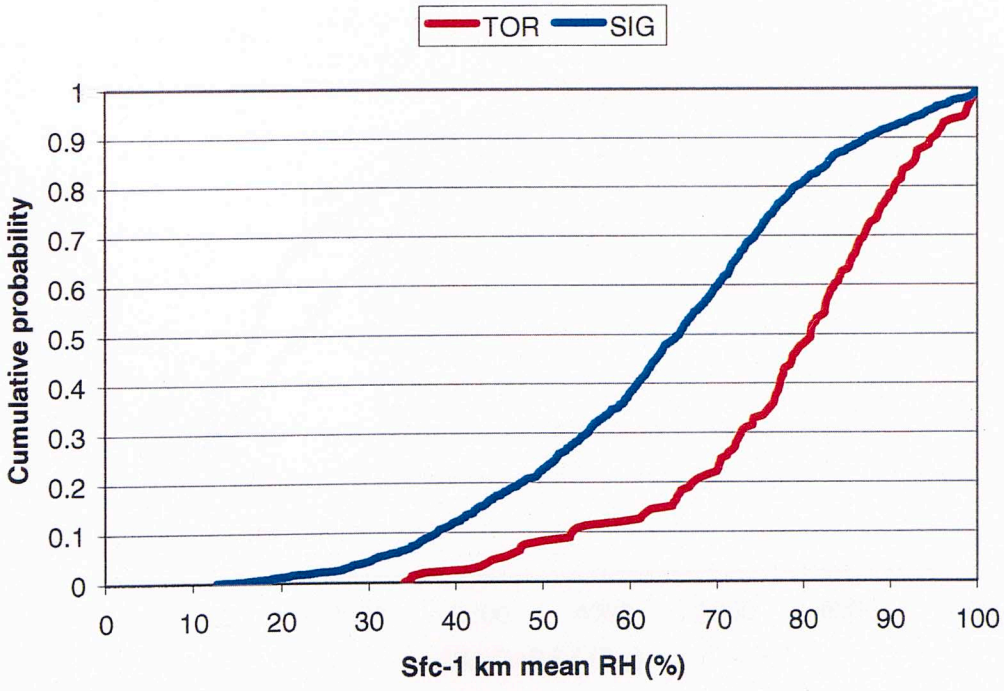

Figure 10. As in Figure 7, except for sfc-1 km mean RH.

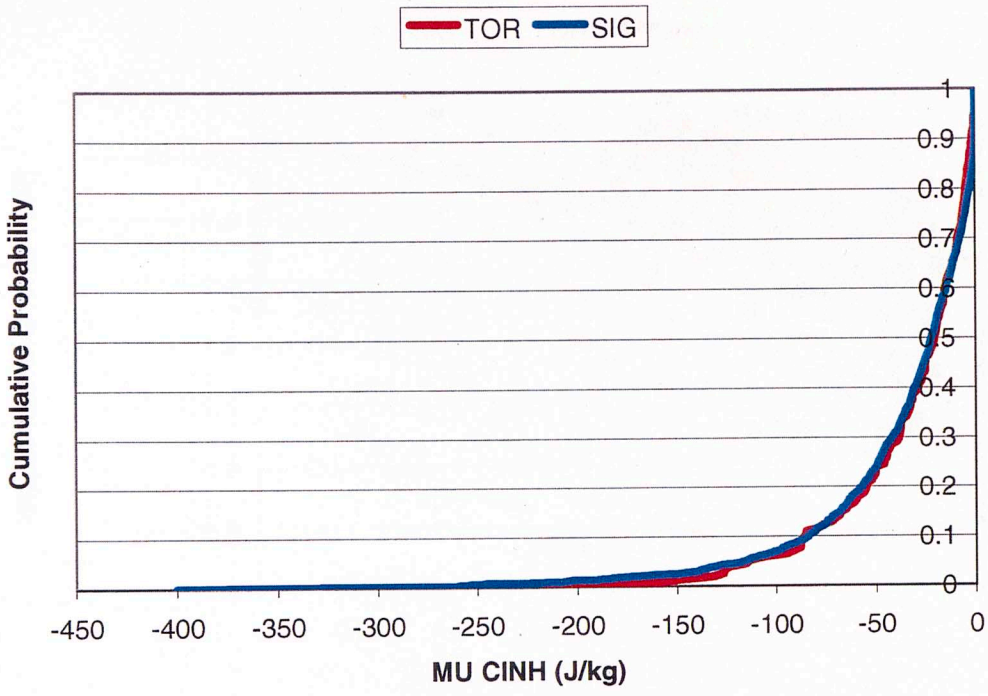

Figure 11. As in Figure 7, except for MU CINH.

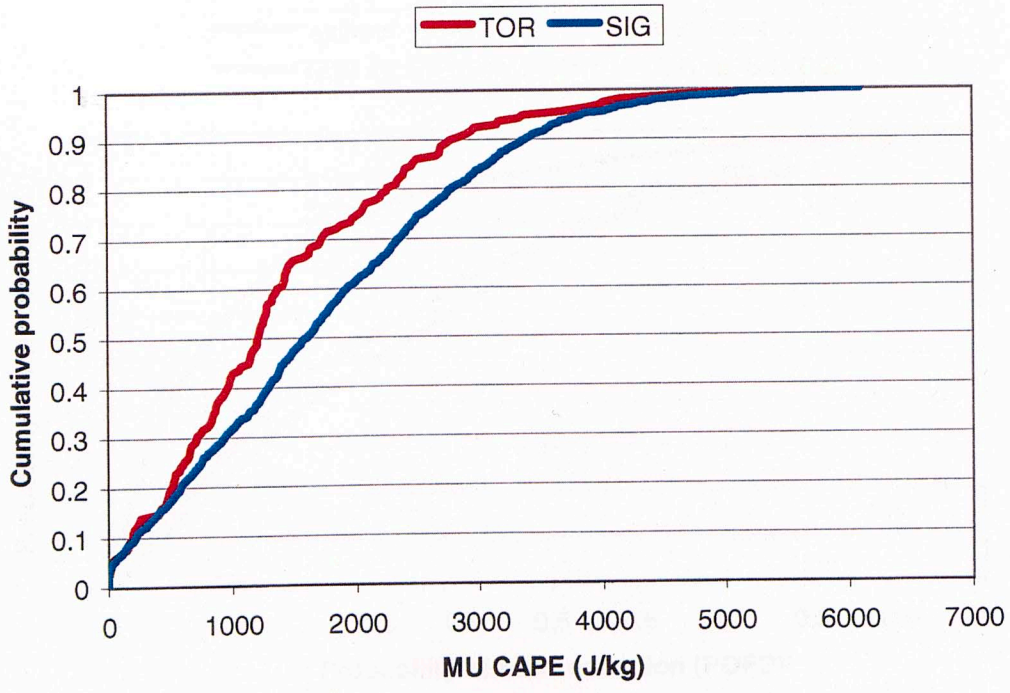

Figure 12. As in Figure 7, except for MU CAPE.

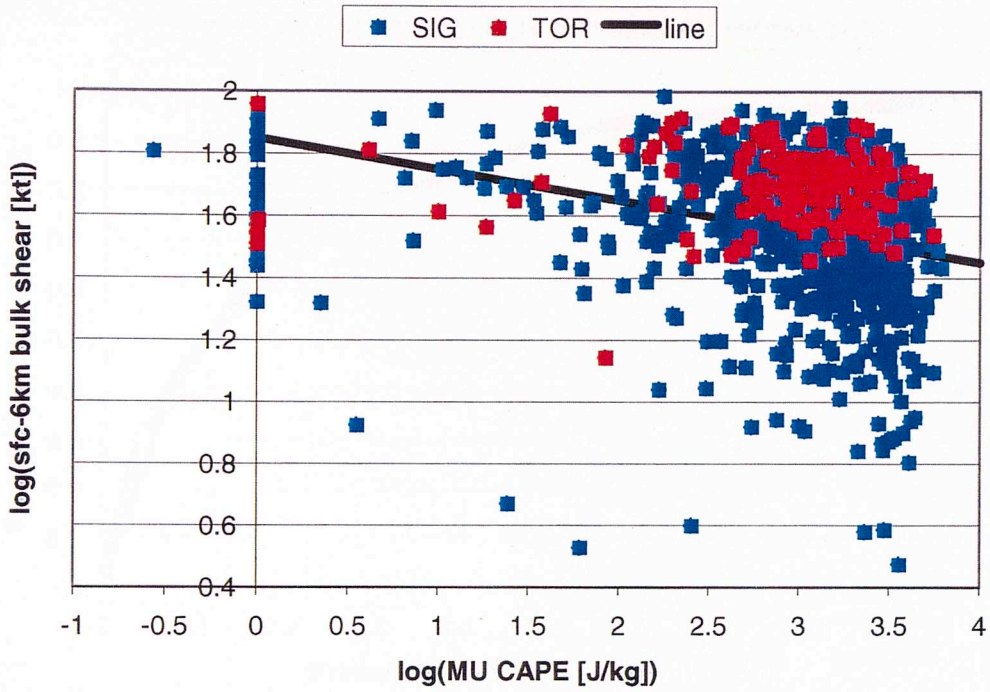

Figure 13. As in Figure 3, except for TOR and SIG distributions.

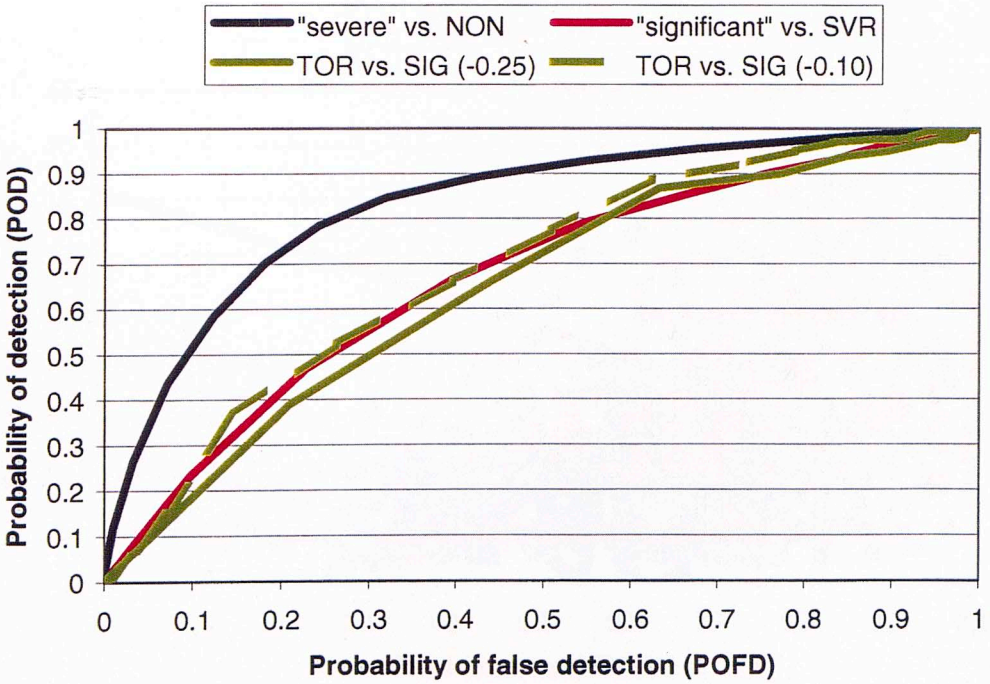

Figure 14. Relative operating characteristic diagram, with curves corresponding to the combination of MU CAPE and sfc-6 km bulk shear, for different levels of discrimination (with TOR versus SIG for two forecast threshold slope values).

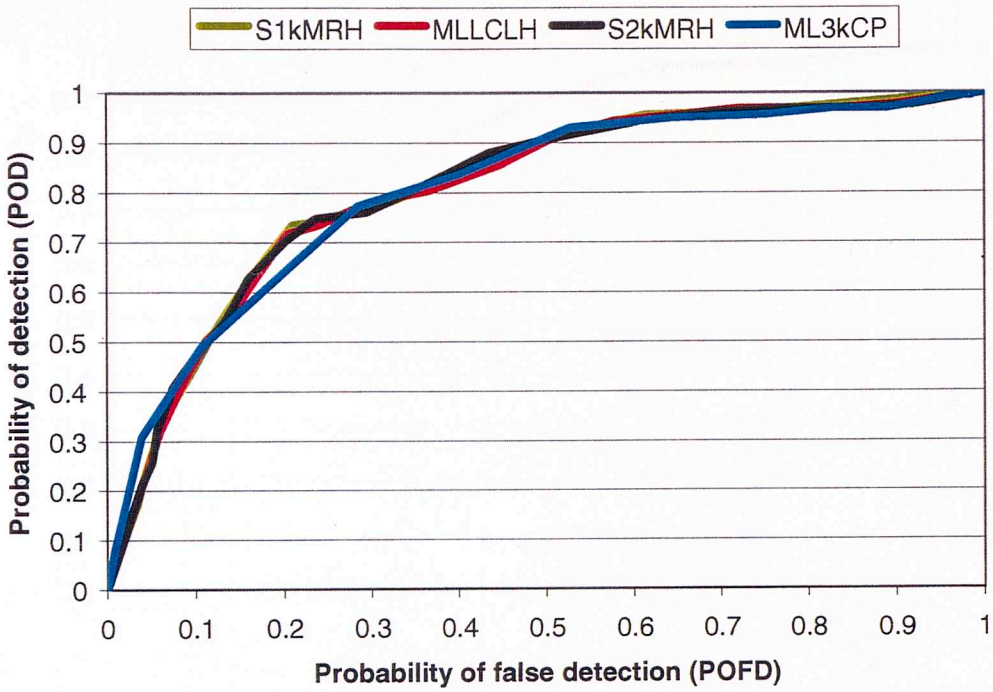

Figure 15. Relative operating characteristic diagram, with curves corresponding to combinations of sfc-1 km bulk shear with sfc-1 km mean RH, ML LCLH, sfc-2 km mean RH, and ML 3 km CAPE, for TOR versus SIG distributions.

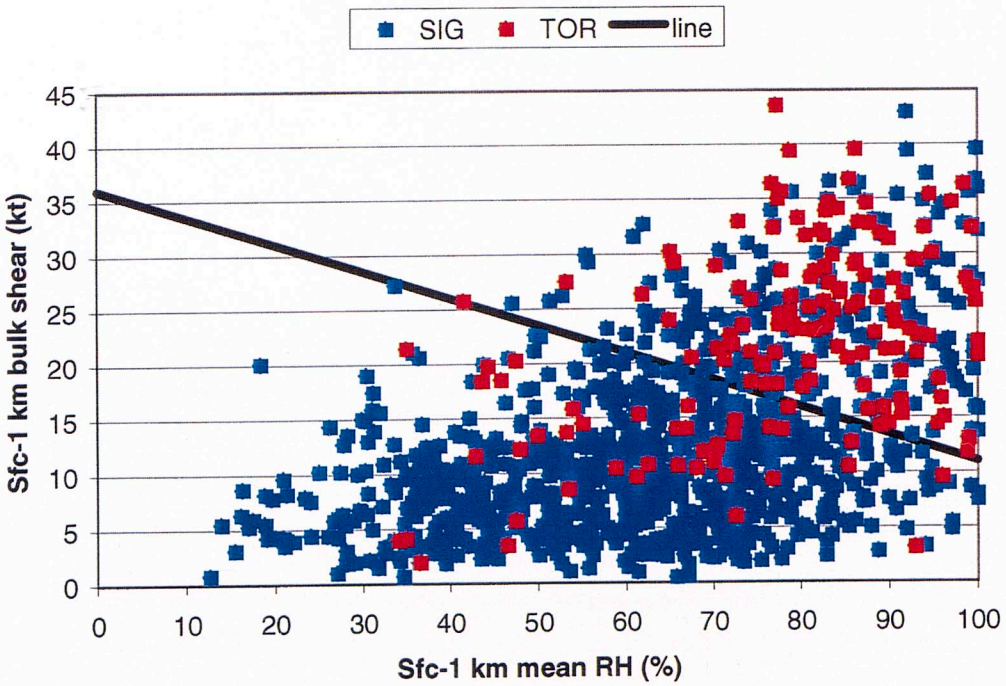

Figure 16. Sfc-1 km mean RH versus sfc-1 km bulk shear for TOR and SIG distributions. The line is an example of a forecast threshold.

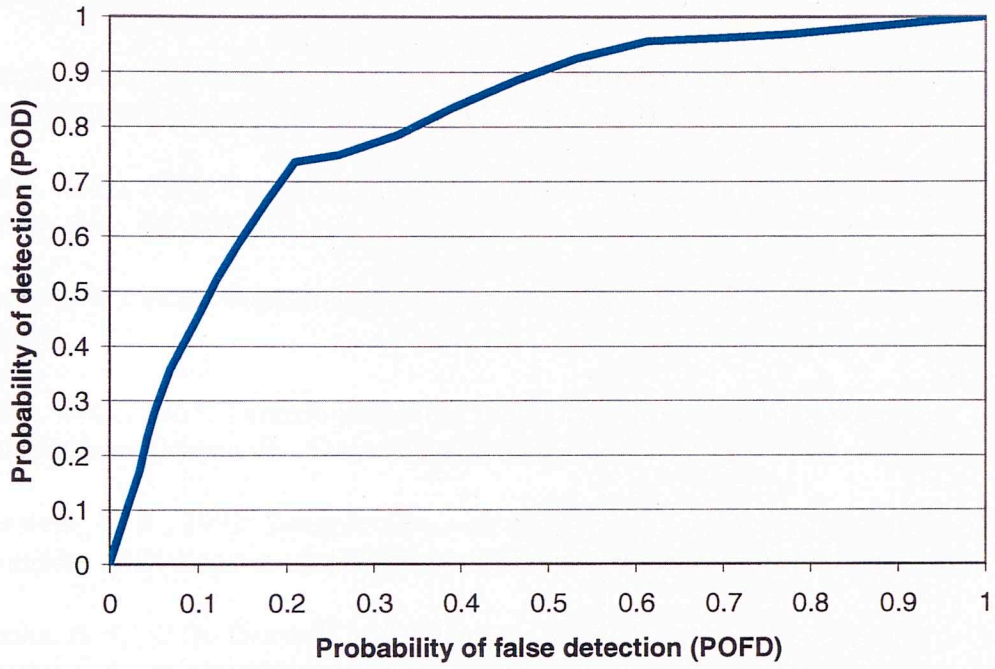

Figure 17. Relative operating characteristic diagram corresponding to Figure 16.

REFERENCES

Adams, J. C., and P. N. Swarztrauber, 1999: SPHEREPACK 3.0: A model development facility. *Mon. Wea. Rev.*, **127**, 1872-1878.

Beebe, R. G., 1955: Types of airmasses in which tornadoes occur. *Bull. Amer. Meteor. Soc.*, **36**, 349-350.

Beebe, R. G., 1958: Tornado proximity soundings. *Bull. Amer. Meteor. Soc.*, **39**, 195-201.

Beebe, R. G., 1963: Tornado proximity soundings. *Proc. 3rd Conf. on Severe Local Storms*, Urbana, IL, Amer. Meteor. Soc., 1-6.

Bluestein, H. B., 1992: *Synoptic-Dynamic Meteorology in Midlatitudes. Volume I: Principles of Kinematics and Dynamics*. Oxford University Press, 431 pp.

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

Buck, A. L., 1981: New equations for computing vapor pressure and enhancement factor. *J. Appl. Meteor.*, **20**, 1527-1532.

Christias, P., cited 2002: Information about the lats4d script. [Available online at http://www.nasa.proj.ac.il/guide/lats4d.htm.]

Craven, J. P., 2001: A baseline climatology of sounding derived parameters associated with deep, moist convection. Unpublished manuscript, 56 pp. [Available from Storm Prediction Center, 1313 Halley Circle, Norman, OK 73069 U.S.A.]

Darkow, G. L., 1969: An analysis of over sixty tornado proximity soundings. Preprints, *6th Conf. on Severe Local Storms*, Chicago, IL, Amer. Meteor. Soc., 218-221.

Doswell, C. A., III, and D. W. Burgess, 1993: Tornadoes and tornadic storms: A review of conceptual models. *The Tornado: Its Structure, Dynamics, Prediction and Hazards, Geophys. Monogr.*, No. 79, Amer. Geophys. Union, 161-172.

Fawbush, E. J., and R.C. Miller, 1952: A mean sounding representative of the tornadic airmass environment. *Bull. Amer. Meteor. Soc.*, **33**, 303-307.

Fawbush, E. J., and R.C. Miller, 1954: The types of airmasses in which North American tornadoes form. *Bull. Amer. Meteor. Soc.*, **35**, 154-165.

Hart, J. A., cited 2002: Severe Plot Version 2.0. [Available online at http://www.spc.noaa.gov/software/svrplot2/.]

Hart, J. A., and W. D. Korotky, 1991: The SHARP workstation - v1.50. A skew-t/hodograph analysis and research program for the IBM and compatible PC. User's manual, NOAA/NWS Forecast Office, Charleston, WV, 62 pp.

Johns, R. H., J. M. Davies, and P. W. Leftwich, 1993: Some wind and instability parameters associated with strong and violent tornadoes, 2. Variations in the combinations of wind and instability parameters. *The Tornado: Its Structure, Dynamics, Prediction and Hazards, Geophys. Monogr.*, No. 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.

Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Met. Mag.*, **30**, 291-303.

Rasmussen, E. N., and R. B. Wilhelmson, 1983: Relationships between storm characteristics and 1200 GMT hodographs, low-level shear, and stability. Preprints, *13th Conf. on Severe Local Storms*, Tulsa, OK, Amer. Meteor. Soc., J5-J8.

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

Showalter, A. K., and J.R. Fulks, 1943: Preliminary report on tornadoes. U.S. Department of Commerce, Weather Bureau, War Advisory Council on Meteorology, Restricted Rep. 1110, 162 pp.

Stanford Biomedical Informatics, cited 2002: Classification of Expression Arrays Version 1.0. [Available online at http://classify.stanford.edu/classify.html.]

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.