

Analysis of Stability Indices for Severe Thunderstorms in the Northeastern United States

Honors Thesis

Presented to the College of Agriculture and Life Sciences, Physical Sciences  
of Cornell University

in Partial Fulfillment of the Requirements for the  
Research Honors Program

by

Robert J. Gottlieb

May 2009

Mark W. Wysocki

## ABSTRACT

In operational forecasting, many indices are used to assess the stability of the atmosphere and predict the likelihood of severe thunderstorm development. One of the shortcomings of many of these indices is that they are mainly based on observations from the southern Plains. Severe thunderstorms can occur in the northeastern United States in conditions which significantly differ from those expected in the Plains. Few attempts have been made to modify these indices for thunderstorms in the Northeast. A new set of values specifically for use in the Northeast are computed. The stability indices examined are the Showalter index, lifted index, SWEAT, K index, total totals, CAPE, CIN, and equilibrium level pressure. Thunderstorms which occurred between the months of June through August during 1998-2007 are used in this analysis. Upper air data from radiosondes at eight sites in the Northeast are used to calculate the new index values. These data are analyzed with respect to the presence or absence of severe thunderstorms within 150 km and  $\pm 3$  hours of a sounding. Using these criteria, there are 423 soundings which contained severe thunderstorms and 13,012 soundings which do not. Forecast skill is calculated for each index. The best predictor in most cases is the LI, with CAPE a close second. For most indices, the threshold which result in the best forecast indicate less instability than what is typically required in the Plains. Probability density functions and scatter plots are created to visualize the data for all soundings.

### 1. Introduction

Although severe thunderstorms are not as common in the northeastern United States as they are in other parts of the country, they still caused over 100 deaths, 1000 injuries, and \$1 billion in combined property and crop damage between 1998 and 2007 (*Storm Data*). In light of these figures, producing accurate forecasts of severe thunderstorms in the Northeast is an important goal.

Generally speaking, instability, lift, and moisture are the three ingredients which must be present in the atmosphere in order to generate the deep moist convection necessary for severe thunderstorm development (Doswell 1987). This analysis focused on many of the parameters which indicate the stability (or lack thereof) of the atmosphere. These quantities should not be used to make predictions without regard to the state of the atmosphere (e.g. Doswell and Schultz 2006). However, they can be incorporated into decision trees which consider a large number of variables.

The cutoffs for these indices used in thunderstorm forecasts were developed based on data from the southern Plains. Previous experience indicates that these cutoffs have not been very useful for predicting severe thunderstorms in the northeastern United States. Therefore, the main goal of this study was to investigate how frequently severe thunderstorms occur in the Northeast under conditions which were generally thought to be too stable for their development.

#### *a. Previous Studies*

Studies of stability indices have been conducted in many different parts of the world. In addition to the southern Plains, these include the Florida Panhandle (Fuelberg and Biggar 1994), Switzerland (Huntrieser et al. 1997), the Netherlands (Haklander and van Delden 2003), and the Balearic Islands (Tudurí and Ramis 1997). Few studies have been conducted on the environments of thunderstorms in the northeastern United States. LaPenta et al. (2002) developed an equation for predicting the severity of thunderstorms in this region. Some recent studies have

averaged values for a variety of parameters but have not examined regional differences (Schneider et al., 2006).

Some of these studies have not used consistent categories for classifying thunderstorms. Some have subjectively labeled days as having “weak” or “strong” convection (Fuelberg and Biggar 1994). Others have counted the number of severe thunderstorms on a particular day in the studied area to assess the level of convection on that day (LaPenta et al. 2002). Due to the lack of uniformity of categorization in the past, comparisons of the results of this study to findings in previous studies are inexact.

## 2. Data Sources

### a. Upper Air Soundings

This study primarily used data from rawinsonde observations. The benefit of this approach is that it is based on direct observations of the atmosphere. A high resolution model, such as the RUC, could also be used (Thompson et al. 2003), but the data a model provides cannot be verified in places where no actual measurements are available. Archived upper air soundings from all of the sites in the Northeast were retrieved from the website of University of Wyoming’s Department of Atmospheric Science (<http://weather.uwyo.edu/upperair/sounding.html>). The website provides calculations for many different stability indices for each sounding. The states which were considered part of the Northeast were the same states that are monitored by the Northeast Regional Climate Center (Fig. 1). The main shortcoming of rawinsonde soundings is that they are only launched from eight locations in the Northeast, generally only at 0Z and 12Z. Even when using criteria that assume a sounding can represent a large section of the atmosphere, a large portion of the Northeast cannot be included (Fig. 1). In order to attain data for a larger number of thunderstorms, more years were examined than in most studies of this nature. The soundings were not altered in any way (e.g. by warming 12Z soundings to the temperature observed 12 hours later). While performing quality control on all of the soundings would be a very time consuming task, the large number of soundings included in the analysis should ensure that outliers and erroneous soundings would have a minimal effect on the results of statistical analyses (Craven et al. 2002).

The sounding site in Maniwaki, Quebec, could also have been included, since it is in a region which is climatologically similar to the Northeastern U.S., but the soundings range is entirely in Canada, and no storm reports were freely available from Environment Canada, which contains the Meteorological Service of Canada.

### b. Severe Thunderstorm Reports

Severe thunderstorm reports were obtained from the National Climatic Data Center’s *Storm Data*. The National Weather Service’s criteria for severe thunderstorms – winds greater than 58 mph, hail greater than 3/4” in diameter, or a tornado – were used. (It should be noted that the NWS increased the hail diameter threshold to 1” on April 1, 2009.) Thunderstorms with funnel clouds were also included, because they indicate the presence of rotation in the atmosphere. Each storm report was examined based on its starting location and beginning time.

### 3. Methodology

#### a. Requirements for Included Severe Thunderstorms

Soundings which were taken between the months of June and August between the years of 1998 and 2007 were examined. Soundings taken when a severe thunderstorm originated within 150 km of the location of the sounding and  $\pm 3$  hours of the timing of the sounding were categorized as 'severe.' In cases where a storm report was recorded within 150 km of two or more soundings, only the closest sounding was considered 'severe.' These criteria are most similar to those used by Craven et al (2002). Severe soundings were separated by the type of severe weather they produce. All soundings which did not fit these criteria were categorized as 'null.'

Under this classification, 423 storm reports were associated with 352 soundings. Of the storm reports studied, 238 were for wind, 139 were for hail, 31 were for tornadoes, and 15 were for funnel clouds. The vast majority of the severe soundings (296) were measured at 00Z, while 41 were obtained at 12Z. A total of 13,012 null soundings remained.

#### b. Examined Indices

Most of the indices analyzed describe the stability of the atmosphere, as opposed to shear or moisture. They can be determined by mathematical formulae or by plotting on a skew-T log-p diagram (Fig. 2). Indices analyzed include Showalter index, lifted index, K index, SWEAT, total totals, CAPE, CIN, and equilibrium level pressure. The lifted index, CAPE, and CIN are calculated using virtual temperature as well as normal air temperature. What follows is a list of the indices studied and a description of how each one was calculated. General cutoffs that forecasters use when predicting thunderstorms and their severity are also presented. The specific numbers that are used often vary depending on the source and local experience, and other conditions obviously must be considered. For example, Gordon and Albert (2000) provide a more complete summary of thresholds and other factors that forecasters at the NWS Weather Forecast Office in Springfield, MO, consider.

*Showalter Index:* The Showalter index (Showalter 1953) is calculated by adiabatically lifting a parcel of air at 850 hPa up to 500 hPa. The temperature (in  $^{\circ}\text{C}$ , as will be the case for all uses of temperature in calculating these indices) the actual environment at 500 hPa is then subtracted from the temperature of the theoretical parcel. Forecasters have generally looked for a small negative number, generally  $-3^{\circ}\text{C}$  or less as a sign of strong convection.

*Lifted Index:* The lifted index (Galway 1956) is calculated in a similar manner as the Showalter index, except that the theoretical parcel near the surface has different properties. There are several different ways of determining the characteristics of this parcel, but in this study temperature and the dewpoint were averaged over the lowest 500 m of the sounding. The expected values for the lifted index are the same as those of the Showalter index.

*K Index:* The K Index (George 1960) is determined by using a simple formula using the temperatures and dew points at different levels of the atmosphere.

$$KI = (T_{850} - T_{500}) + T_{D\ 850} - (T_{700} - T_{D\ 700})$$

where  $T$  is the temperature,  $T_D$  is the dewpoint, and the numbers represent the pressure level at which each variable is measured. A value  $\geq 30^\circ\text{C}$  is typically expected for severe thunderstorms.

*SWEAT:* The severe weather threat index, or SWEAT (Miller 1972) incorporates several variables in order to predict tornadoes.

$$\text{SWEAT} = 12(T_{D\ 850}) + 20(\text{TT} - 49) + 2(F_{850}) + F_{500} + 125(\text{SHEAR})$$

where  $\text{TT}$  is the total totals index (see below),  $F$  is the wind speed (in knots) at the level listed, and  $\text{SHEAR} = \sin(\text{WD}_{500} - \text{WD}_{850})$  where  $\text{WD}$  is the wind direction at a particular level. Forecasters often interpret a  $\text{SWEAT} \geq 250$  as a sign for severe thunderstorms or 400 as an indicator of the possibility of tornadoes.

*Total Totals:* The total totals index (Miller 1972) is simply the sum of the cross totals and vertical totals, which are quite simple themselves.

$$\begin{aligned} \text{CT} &= T_{D\ 850} - T_{500} \\ \text{VT} &= T_{850} - T_{500} \end{aligned}$$

A value  $\geq 45^\circ\text{C}$  indicates a higher possibility of severe thunderstorms.

*CAPE:* The convective available potential energy, or CAPE, (Moncrieff and Miller 1976) is an integrated value. Like the lifted index, a theoretical parcel is lifted adiabatically. In this case the parcel is lifted until its temperature becomes equal to the temperature of the surrounding environment (the equilibrium level, which was also analyzed). The positive area between the parcel's temperature and the environment's temperature is the CAPE. CAPE can also be calculated using the formula

$$\text{CAPE} = g \int_{\text{LFC}}^{\text{EL}} \frac{\theta - \theta'}{\theta'} dz$$

where  $\theta$  is the potential temperature of the parcel,  $\theta'$  is the potential temperature of the environment,  $\text{LFC}$  is the lifted condensation level, and the  $\text{EL}$  is the equilibrium level (Weisman and Klemp 1986). A value  $\geq 1000 \text{ J kg}^{-1}$  is typically seen as a strong indicator of an increased chance of thunderstorms activity.

While it is recognized that the CAPE, as well as other parameters in this study, hide some of the complexities in the state of the atmosphere and that adjustments such as normalization for depth of the convective layer can be performed, (e.g. Blanchard 1998) such values were not available for this study but can be examined in the future.

*CIN:* The convective inhibition (CIN) is determined using the same process that was used to calculate the CAPE, except the negative area is integrated.

It should be noted that some of these indices are best suited for forecasting during certain conditions. As already mentioned SWEAT is most useful for predicting tornadoes. As another example, the K index is optimal for predicting air mass

thunderstorms (George 1960). Since different regimes were not taken into account during this analysis, the accuracy of indices which were developed for a particular forecasting during particular scenarios may be masked by the averaging of a large number of soundings.

*c. Statistical Analyses*

*i. Skill Scores*

Skill scores for each index were calculated as in Huntrieser et al. (1997). These calculations involve the use of a 2x2 contingency table (Fig. 3). From the contingency table, POD, FAR, CSI, TSS, and S were calculated. These skill scores are defined in Table 1. Each index is assigned the threshold which gives the best TSS. In other words the cutoffs presented for each index are those which most successfully discriminated between soundings which were associated with a severe thunderstorm and those which were not. Most of the indices studied indicate increasing instability when they increase in value. Therefore, these indices must exceed their given threshold to predict a severe thunderstorm. Similarly, indices which indicate increasing instability with decreasing value (such as the lifted index and Showalter index) must be lower than their given threshold to predict a severe thunderstorm.

Skill scores for combinations of two indices were also computed. However, since all of the indices examined are used to estimate stability, and since the equations of some indices share terms, there is a high correlation between many of the indices (DeRubertis 2006). While combining indices did improve skill scores, the use of parameters to assess moisture, shear, and other factors will be meaningful for thunderstorm prediction (Maglaras and LaPenta 1997).

When all null soundings are included in skill score calculations, the FAR will be very high, and the CSI and Heidke skill score will be very low. This occurs because number of correct predictions of no severe thunderstorms is much higher than any other outcome. In order to offset this, skill score calculations were also made by randomly selecting a number of null soundings equal to the number of severe soundings. When this alteration is made, the Heidke skill score becomes equal to the TSS because  $A+B = C+D$ . However, in Table 3b, the TSS and S will not always be equal because not all soundings had an available value for each index. This is especially true for the equilibrium level pressure, since the potential temperature of the theoretical parcel will not always become equal to the potential temperature of the environment after passing the LFC. This also explains why  $A+B$  will not always equal the total number of severe soundings in a calculation and  $C+D$  will not always equal the total number of null soundings.

*ii. Probability Distribution Functions*

Kernel smoothed probability distribution functions (PDFs) for the distribution of an index for severe and null soundings are plotted on the same graph (Fig. 4). For a technical discussion of kernel density estimation see Botev (2006). The intersection of the functions represents the threshold this method selects. A graph in which the two PDFs have little overlap represents more distinction between severe and null

soundings. Generalized extreme value distributions were used for indices with irregularly shaped PDFs. These indices include CAPE and CIN, since a very high number of soundings produced values of CAPE and CIN near 0. Additionally, CAPE cannot be negative, and CIN cannot be positive. A Gaussian distribution cannot account for this restriction. Gaussian distributions are suitable for the remaining indices. The threshold which gives the highest skill score is determined by finding the intersection of the distribution for severe soundings and null soundings.

### *iii. Scatter Plots*

Scatter plots are used to visualize the relationship between two different indices (Fig. 5a, 5b, 6). Most of these graphs show linear relationships, but indices with skewed distributions produced graphs where power regressions were most appropriate. Glyph scatter plots (Wilks 2006) were used to show differences in the soundings for storms meeting different severe criteria. In some cases glyphs were plotted with different sizes according to the severity of each thunderstorm, but this did not produce significant results. Estimates of wind speed and hail size listed in *Storm Data* are often not reliable, (Trapp et al. 2006) so it is not surprising that this occurred. The inclusion of these degrees of severity was generally not very informative and was not factored into most calculations.

## **4. Results**

### *a. Statistical Analyses*

#### *i. Basic Statistics*

Before discussing more the results of the more advanced statistical methods outlined above, it is helpful to look at the basic computations of mean, mode, and standard deviation for all severe soundings (Table 2a) and null soundings (Table 2b). Unsurprisingly, the mean and median are much different for the severe soundings and null soundings. What is more pertinent is a comparison of the median values for the severe soundings and the expected values for severe thunderstorms listed earlier. For instance, the median Showalter index is almost zero, while the median lifted index is closer to typical expectations. The median CAPE is somewhat lower than  $500 \text{ J kg}^{-1}$ , while the median total totals is within the range conventional criteria. A cursory glance at the most basic statistics indicates that several of the indices for severe soundings may be different in the Northeast than they are in other parts of the United States.

#### *ii. Skill Scores*

Skill scores using all null soundings (Table 3a) and a number of null soundings equal to the number of severe soundings (Table 3b) are presented. The highest skill score for any single predictor was achieved when using an  $\text{LI} < 0.9^\circ \text{C}$ . This is surprising for two reasons. First, some forecasters have used an LI slightly less than  $0^\circ \text{C}$  as a loose threshold in the Northeast (Evans, personal communication, 2008), whereas this study finds that severe thunderstorms occur in even more stable conditions often enough to produce an LI threshold greater than 0. Second, CAPE

theoretically should be a more accurate representation of an environment's stability than LI. This should be true because the CAPE is calculated by integrating the majority of the sounding, while the LI is calculated simply by using temperatures at two different levels of the atmosphere. Still, LaPenta et al. (2002) found that LI had a better correlation with the thunderstorm severity categories they created than CAPE. Surface based lifted index yielded the best skill score in Switzerland (Huntrieser et al. 1997) and the Florida panhandle (Fuelberg and Biggar 1994). Additionally, the lifted index calculated by lifting a parcel with properties averaged over the layer 100 hPa above the surface resulted in the highest skill score with a threshold of  $\leq 3.0^{\circ}\text{C}$ , while the CAPE yielded much lower skill scores (Haklander and Van Delden 2003).

The thresholds for CAPE determined by skill scores were less than  $100 \text{ J kg}^{-1}$ . This is far lower than the values which are generally expected for severe thunderstorms in the Plains. However, values of CAPE which produced the highest skill scores for predicting thundery versus nonthundery days in Switzerland (Huntrieser et al. 1997) and the Netherlands (Haklander and Van Delden 2003) were similarly low. Additionally, the CAPE consistently had slightly worse skill when it was calculated using virtual temperature, even though that should be a more accurate representation of the CAPE (Doswell and Rasmussen 1994).

The combined usage of two indices increased skill scores (not shown). The highest TSS occurred when using LI and LI Tv, but this is not very meaningful since the two variables are so similar. The combination of LI and CAPE yielded the next highest TSS (.656), with  $\text{CAPE} > 10 \text{ J kg}^{-1}$  and  $\text{LI} < 1.8^{\circ}\text{C}$ . This TSS considerably higher than that for LI alone, and the best thresholds are very similar. Another interesting result came from the use of EL. For an  $\text{EL} < 550 \text{ hPa}$  and  $\text{LI} > 1.0^{\circ}\text{C}$ , the TSS was .619. This is greater than the TSS for LI alone. This result it is especially notable since the TSS for EL was .294, the worst of any index studied, but the use of EL and LI together is better than LI and some other indices. This indicates that EL can improve forecasts when combined with other parameters.

### *iii. Probability Distribution Functions*

The use of probability distribution functions yielded similar results to skill scores. For example, the Showalter index had a threshold of  $< 1.9631^{\circ}\text{C}$  when using PDFs and  $< 1.9^{\circ}\text{C}$  when using skill scores (Fig. 4) The results for all of the other examined indices were similarly close (Table 4). This increases confidence in the validity of the skill scores and thresholds.

### *iv. Scatter Plots*

Scatter plots quickly reveal a variety of interesting features in the data. The LI and CAPE have a power relationship (Fig. 5a) The CAPE remains very close to zero when the LI is greater than zero. However, the CAPE rapidly increases once the LI becomes slightly negative. Additionally, the severe soundings fall relatively neatly within the range of indices for the null soundings. The relationship between LI and CAPE for severe soundings can be approximated by the fit

$$\text{CAPE} = 1.363^{(18.39 - \text{LI})}$$

with  $r^2 = .7425$ . For the purposes of this fit, one outlier with an extremely low LI, which is likely the result of instrument error, was excluded from the data set (Fig. 5a)

The most unstable indices were found at the inland stations of Albany and Pittsburgh (Fig. 5b). The other stations in the Northeast are located near a large body of water (either the Atlantic Ocean or a Great Lake). However, Fuelberg and Biggar (1994) found that thunderstorms over the Florida panhandle occurred in environments with average parameter values similar to what is expected in the Plains. This information combined with comparisons to previously mentioned results in Switzerland and the Netherlands suggests that instability may be limited by the presence of a large body of cool water, but severe thunderstorms can still happen in these areas.

Another graph of note depicts the relationship between the lifted index and the equilibrium pressure (Fig. 6). While the severe soundings exhibit a reasonably strong linear correlation, the null soundings are clearly split into two clusters with a distinct node between the two. One group of null soundings has an EL pressure less than 500 hPa and an LI less than  $5^{\circ}\text{C}$ . This represents soundings of unstable environments which may not have had sufficient lift and/or moisture to produce severe thunderstorms. The majority of the severe soundings overlap this group. The other group has an EL pressure greater than 500 hPa and an LI greater than  $0^{\circ}\text{C}$ . These soundings could be said to be too stable for deep moist convection, but a substantial minority of severe soundings having slightly positive LI and EL between 500 and 600 hPa were located within this group. The graph is also notable for its unusual shape. There is a distinct node between the two clusters near  $\text{LI} = 0^{\circ}\text{C}$  and  $\text{EL} = 500$  hPa, and very few soundings are positioned outside of the two groups. The more unstable cluster is compact, while the more stable cluster is more dispersed. The more stable cluster also has a sharp boundary near  $\text{LI} = 0^{\circ}\text{C}$ . The significance of the peculiar shape of this graph is unknown.

#### *b. Comparison to Thunderstorms in Southern Plains*

The thresholds obtained for the Northeast were compared to thresholds found when using the same methods on thunderstorms in the Plains. Data from fifteen stations in the states extending from North Dakota to Texas were analyzed using the same criteria as the Northeast (Fig. 7). This is one way of checking the accuracy of the index calculations performed by Wyoming's website. This resulted in 1179 usable storm reports and 956 severe soundings. Of these reports, 578 were for hail, 454 were for wind, and 147 were for hail. Funnel clouds were not included in this set. A total of 29,747 null soundings remained. Both the median values for severe soundings (Table 5a) as well as null soundings (Table 5b) indicate higher instability in the Plains than in the Northeast.

The thresholds which resulted in the highest skill scores in the Plains were different from those in the Northeast, but not as much as one might expect given generally used forecasting guidelines (Table 6a, 6b). This suggests that the time and distance limits placed on soundings should be more restrictive in order to ensure that a more representative sounding is associated with each storm report. However, decreasing the time and distance between a sounding and severe thunderstorm report did not significantly change any results. Additionally, the indices had almost no

correlation between the indices and either time from the sounding or distance from the sounding. No significant differences were found between the distributions of indices for storms that occurred before or after the soundings to which they were matched. This suggests that contamination of soundings by other thunderstorms was not a major factor in the results. However, it is impossible to know this without access to information which would more strongly indicate the presence of thunderstorms, such as lightning data. Even with tighter limits on the time and distance between an observation and a thunderstorm, determining the data which best characterizes the environment around a thunderstorm can be quite difficult and uncertain (Thompson et al. 2003, Rasmussen and Blanchard 1998).

Additionally, skill scores were generally much higher for the Northeast than for the Plains. In particular, the CAPE performed especially poorly in the Plains, with a TSS of only 0.121. Interestingly, the indices that yielded the highest TSS in the Plains were not the same as those in the Northeast. The three best indices in the Plains were total totals (TSS = .377), SWEAT (.354), and the Showalter index (.352), while LI (.533), LI calculated using virtual temperature ( $T_v$ ) (.526), and CAPE (.502) were the three best indices in the Northeast.

When the scatter plots comparing LI and CAPE for the Northeast are superimposed on top of the same plot for the Plains, the two sets of data do not look very different (Fig. 8). The main difference between the two is that a higher degree of instability is achieved at times in the Plains, but the shape of the graphs is very similar. Furthermore, the LI at which the CAPE begins to increase sharply is slightly negative for both the Northeast and the Plains.

In order to more clearly determine whether or not these two sets of data could have come from the same distribution, a Wilcoxon-Mann-Whitney test was conducted for each variable. The general procedure for this test is to take two independent data sets, combine them, and rank each value of the tested variable from highest to lowest. The values are then separated into their original sets and the sum of the ranks for each set calculated. The null hypothesis for the test is that the two data sets were selected from the same distribution. If there is a significant difference between the two sums, then the null hypothesis is rejected. This test is more resistant to outliers than a two sample t-test because its evaluation is based the ranks of values, not the actual values. For more details about the Wilcoxon-Mann-Whitney test, see Wilks 2006. The p-value that is the result of this test has the same meaning as the t-value for a t-test. When the test was performed for LI and CAPE, the p-value was on the order of  $10^{-8}$  and  $10^{-11}$  respectively, and most other p-values were even smaller.

### *c. Verification using data from 2008*

Soundings and severe thunderstorm reports (without funnel clouds) were retrieved for the Northeast in June, July, and August of 2008 and skill scores were calculated. A total of 68 severe soundings and 1228 null soundings were obtained. The skill scores for 2008 were not as high as they had been in the previous decade. The indices which produced the highest TSS between 1998 and 2007 (LI, LI  $T_v$ , and CAPE) also resulted in the highest in 2008. During this year, an LI threshold equal to the cutoff which the best in TSS in the previous test ( $<0.9^{\circ}\text{C}$ ) resulted in a TSS of .385 in 2008, compared to .533 in the years 1998-2007. Some thresholds which produced the best

TSS in 2008 were different from those in 1998-2007. The best threshold for LI in 2008 was  $<-0.4^{\circ}\text{C}$ , and it produced a TSS of .464.

The overall decrease of skill scores in 2008 may have been a byproduct of the low sample size for a single year which results from using this study's data collection methods. Still, the ranking of the skill scores by TSS in 2008 using the thresholds computed from the larger data set was very similar to original ranking. Additionally, changes in the thresholds which yielded the best TSS in 2008 compared to the best cutoffs in 1998-2007 were small. These findings increase confidence in the validity of the cutoffs determined for 1998-2007.

## 5. Summary

### *a. Conclusions*

There is considerable evidence that severe thunderstorms can frequently occur in the northeastern United States in conditions which were generally thought to be too stable for strong convection. This information can be useful for forecasters as well as educators presenting these parameters to students for the first time. Skill scores, statistical tests, and graphing have been used to show that conditions for severe thunderstorms in the northeastern United States have some significant differences from conditions for severe thunderstorms in the Plains, though there are some ambiguities as well. The stability index that yielded the best skill scores was the lifted index, even though it only measures two temperature values while the CAPE is integrated over much of the sounding. Although forecasting based on one number is obviously inadvisable, many of the forecasts made using only one index studied perform remarkably well in the Northeast compared to other parts of the world.

### *b. Future Work*

This study focused on stability indices and did not analyze indices used to measure moisture and shear in the atmosphere. Additionally, synoptic conditions at the time of each thunderstorm have not been factored into the calculations. The same methods presented here can be used for thunderstorms occurring during different seasons to see if the indices change at different times of the year. More years can also be included to develop a more comprehensive climatology for the indices. However, the utility of such an undertaking is in doubt since DeRubertis (2006) found that the same frequency with which the LI and CAPE exceeded extreme values had a significant positive trend for the years 1973-97, while the K index and SWEAT had no such trend. Calculations of a greater number of indices at higher temporal and spatial resolution can be performed using the RUC model. This would give higher temporal and spatial resolution of index data and help ensure that the information which is most representative of the convective environment is analyzed. More statistical tests can also be performed to assess the significance of the results. A more detailed look at individual soundings will be necessary to learn more about cases which cause the indices to be misleading.

*Acknowledgements.* The author is grateful for the mentoring and support of Mark Wysocki. The author wishes to thank Mike Evans and Tom Niziol for their advice, as

well as Brian Belcher for his computing assistance. The author also wishes to thank the University of Wyoming and the Storm Prediction Center for providing easily accessible data.

### References

- Blanchard, D.O., 1998: Assessing the vertical distribution of convective available potential energy. *Wea. Forecasting*, **13**, 870-877.
- Botev, Z.I., 2006: A novel nonparametric density estimator. Postgraduate Seminar Series, Mathematics (School of Physical Sciences), The University of Queensland.
- Craven, J.P., H.E. Brooks and J.A. Hart, 2002: Baseline Climatology of Sounding Derived Parameters Associated with Deep, Moist Convection. *Preprints*, 21st Conf. Severe Local Storms, San Antonio.
- DeRubertis, D., 2006: Recent trends in four common stability indices derived from U.S. radiosonde observations. *J. Climate*, **19**, 309-323.
- Doswell, C.A. III, 1987: The distinction between large-scale and mesoscale contribution to severe convection: a case study example. *Wea. Forecasting*, **2**, 3-16.
- , and E. N. Rasmussen, 1994: The effect of neglecting the virtual temperature correction on CAPE calculations. *Wea. Forecasting*, **9**, 625-629.
- , and D.M. Schultz, 2006: On the use of indices and parameters in forecasting severe storms. *Electronic J. Severe Storms Meteor.*, 1(3), 1-22.
- Evans, M., Meteorologist in Charge, National Weather Service Forecast Office, Binghamton, NY. Personal communication. 2008.
- Fuelberg, H.E. and D.G. Biggar, 1994: The preconvective environment of summer thunderstorms over the Florida panhandle. *Wea. Forecasting*, **9**, 316-326.
- Galway, J.G., 1956: The lifted index as a predictor of latent instability. *Bull. Amer. Meteor. Soc.*, **43**, 528-529.
- George, J.J., 1960: *Weather Forecasting for Aeronautics*. Academic Press, 673 pp.
- Gordon, J. D., and D. Albert, 2000: A comprehensive severe weather checklist and forecast guide. National Weather Service Central Region, NWS Tech. Service Publication TSP-10.

- Haklander, A.J., and A. Van Delden, 2003: Thunderstorm predictors and their forecast skill for the Netherlands. *Atmospheric Research*, **67-68**, 273-299.
- Huntrieser, H., H.H. Schiesser, W. Schmid, and A. Waldvogel, 1997: Comparison of traditional and newly developed thunderstorm indices for Switzerland. *Wea. Forecasting*, **12**, 108-125.
- LaPenta, K.D., G.J. Malagras, J.W. Center, S.A. Munafo, and C.J. Alonge, 2002: An updated look at some severe weather forecast parameters. *Eastern Region Technical Attachment* No. 1. <http://www.erh.noaa.gov/er/hq/ssd/erps/ta/ta2002-01.pdf>
- Maglaras, G. J., and K. D. LaPenta, 1997: Development of a forecast equation to predict the severity of thunderstorm events in New York State. *National Weather Digest*, **21**, 3-9.
- Miller, R.C., 1972: Notes on analysis and severe storm forecasting procedures of the Air Force Global Weather Central. Tech. Report 200(R), Headquarters, Air Weather Service, Scott Air Force Base, IL 62225, 190 pp.
- Moncrieff, M.W. and M.J. Miller, 1976: The dynamics and simulation of tropical cumulonimbus and squall lines. *Q.J.R. Roy. Meteorol. Soc.*, **102**, 373-394.
- NCDC, 1998-2008: *Storm Data*. Vol. 37-47. [Data files available at <http://www.spc.noaa.gov/wcm/>].
- Rasmussen, E.N. and D.O. Blanchard, 1998. A baseline climatology of sounding-derived supercell and tornado forecast parameters. *Wea. Forecasting*, **13**, 1148-1164.
- Schneider, R.S., A.R. Dean, S.J. Weiss, and P.D. Bothwell, 2006: Analysis of Estimated Environments for 2004 and 2005 Severe Convective Storm Reports. *Preprints*, 23rd Conf. Severe Local Storms, St. Louis MO.
- Showalter, A.K., 1953: A stability index for thunderstorm forecasting. *Bull. Amer. Meteor. Soc.*, **34**, 250-252.
- Thompson, R.L., R. Edwards, J.A. Hart, K.L. Elmore, and P. Markowski, 2003: Close proximity soundings within supercell environments obtained from the rapid update cycle. *Wea. Forecasting*, **18**, 1243-1261.
- Trapp, R.J., D.M. Wheatley, N.T. Atkins, R.W. Przybylinski, R. Wolf, 2006: Buyer beware: some words of caution on the use of severe wind reports in postevent assessment and research. *Wea. Forecasting*, **21**, 408-415.

Tudurí, E., and C. Ramis, 1997: The environments of significant convective events in the western Mediterranean. *Wea. Forecasting*, **12**, 294-306.

Weisman, M. L., and J. B. Klemp, 1986: Characteristics of isolated convective storms. *Mesoscale Meteorology and Forecasting*, P.S. Ray, Ed., Amer. Meteor. Soc., 331–358.

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

### Figures and Tables

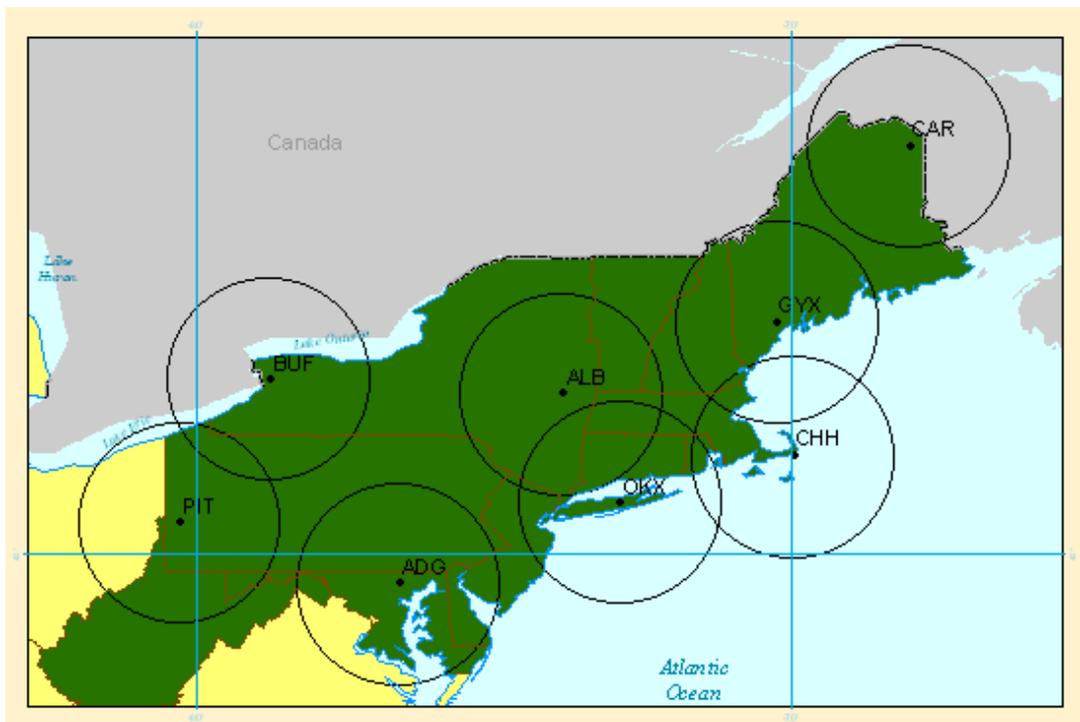
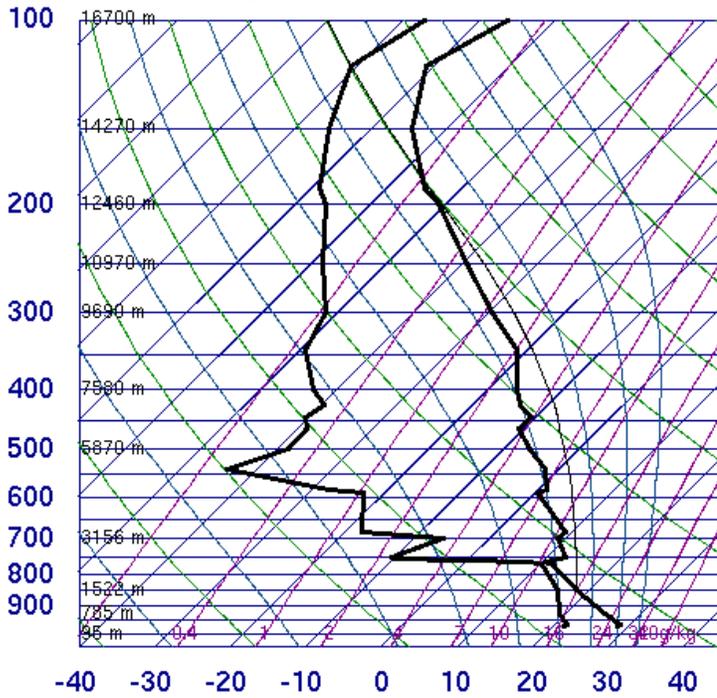


Fig. 1. The states defined as the Northeast for the purposes of this study and the locations of the sounding stations used. The states were selected because they are monitored by the Northeast Regional Climate Center. The rings around the stations have a radius of 150 km, one of the criteria which determined which severe thunderstorms would be analyzed. States shaded in green were included in the analysis.

**72520 PIT Pittsburgh**



- SLAT 40.53
- SLOX -80.23
- SELV 357.0
- SHOW -2.04
- LIFT -4.65
- LFTV -5.71
- SWET 280.1
- KINX 25.80
- CTOT 22.40
- VTOT 24.90
- TOTL 47.30
- CAPE 1165.
- CAPV 1454.
- CINS -6.55
- CINV -0.40
- EQLV 181.4
- EQTV 181.3
- LFCT 850.5
- LFCV 858.5
- BRCH 100.8
- BRCV 125.7
- LCLT 291.9
- LCLP 858.7
- MLTH 304.9
- MLMR 16.19
- THCK 5775.
- PWAT 33.76

**00Z 20 Aug 2005**

**University of Wyoming**

Fig. 2. An example of a sounding plotted on a skew-T log-p diagram from the University of Wyoming. The indices used in the analysis are located in the column on the right.

		Prediction		
		Event predicted +	Event not predicted -	
Observation	Event observed +	Correct information <b>40</b> <b>A</b>	No information <b>10</b> <b>B</b>	50
	Correct prediction	Surprise		
Event not observed -	False information <b>20</b> <b>C</b>	Correct information <b>30</b> <b>D</b>	50	
	False alarm	Correct prediction		
		<b>60</b>	<b>40</b>	<b>100</b>

Fig. 3. Example contingency square used in calculation of skill scores (from Huntrieser et al. 1997)

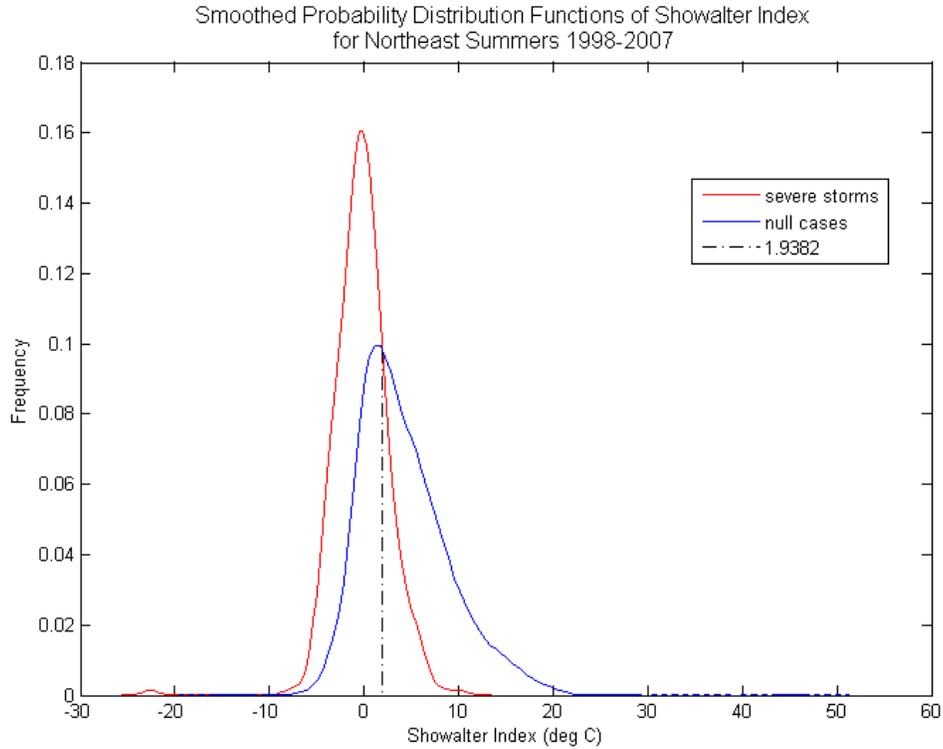


Fig. 4. The kernel smoothed PDFs of the distributions of severe soundings and null soundings for the Showalter Index ( $^{\circ}\text{C}$ ) for Northeast summers between 1998 and 2007. The SI where the distributions intersect (1.9631) is very close to the best threshold found by using skill scores (1.9).

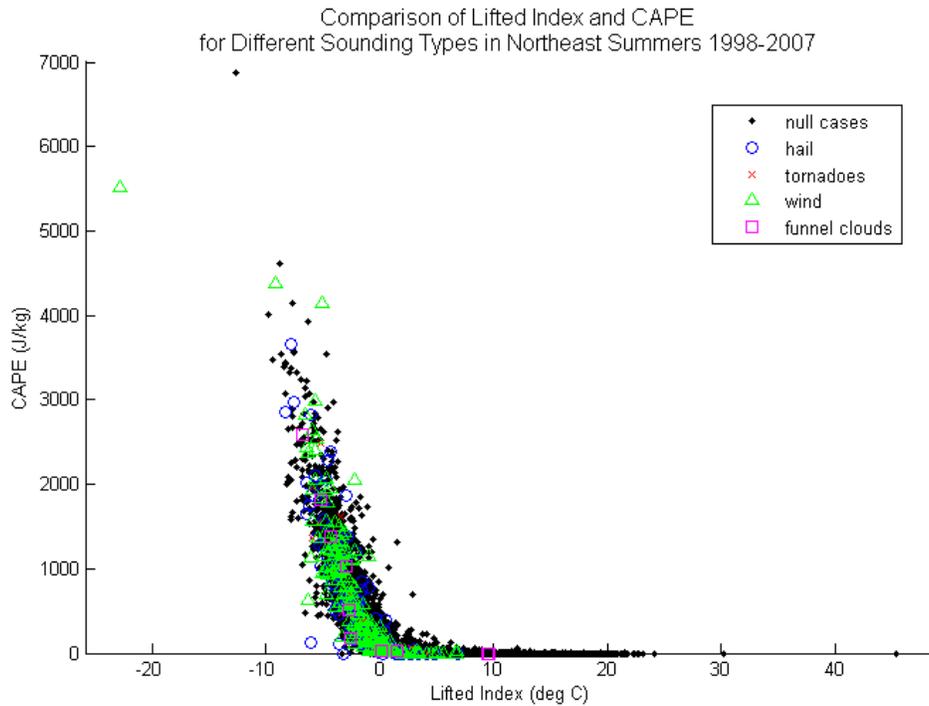


Fig. 5a. Lifted index ( $^{\circ}\text{C}$ ) plotted against CAPE ( $\text{J kg}^{-1}$ ) for all soundings in the Northeast in the months of June, July, and August for 1998-2007.

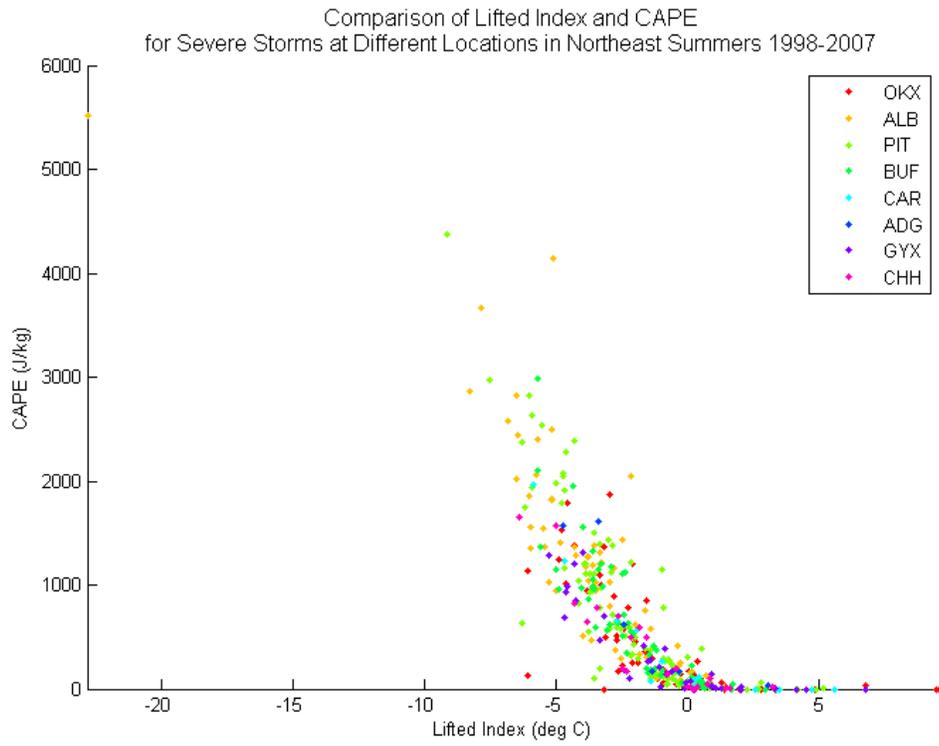


Fig. 5b. The same as Fig. 5a, except only severe soundings and are plotted and they have been color coded by station. The soundings with the highest CAPE and lowest LI were all taken at Albany and Pittsburgh.

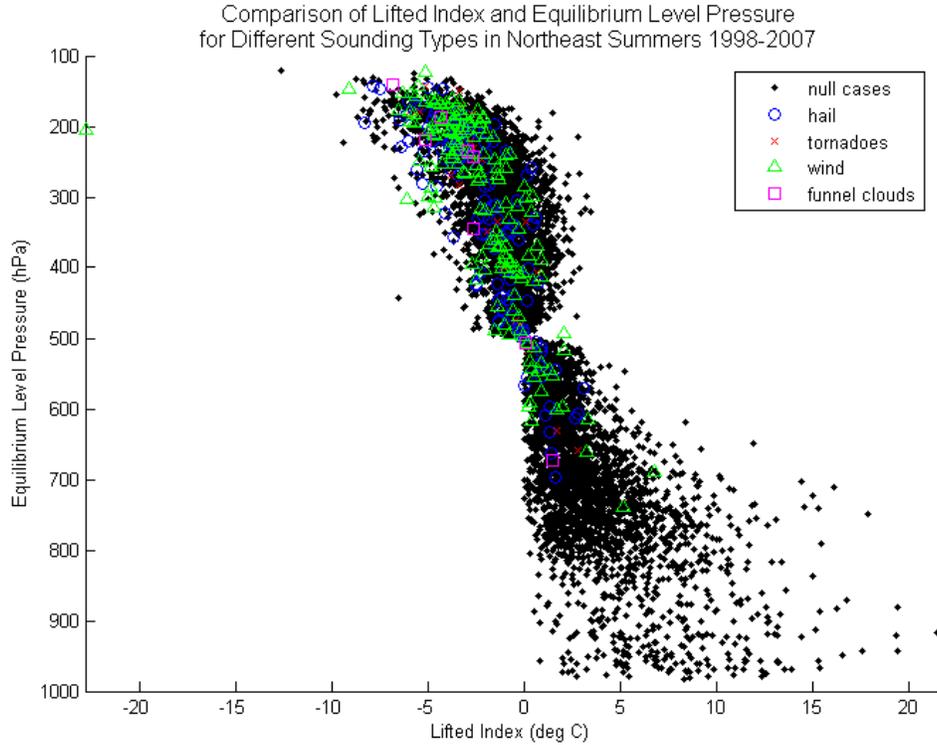


Fig. 6. Lifted index (°C) plotted against equilibrium level pressure (hPa) for all soundings in the Northeast in the months of June, July, and August for 1998-2007. Severe soundings are separated by the type of severe weather associated with them. The plot of the null soundings is notable for its unusual shape.

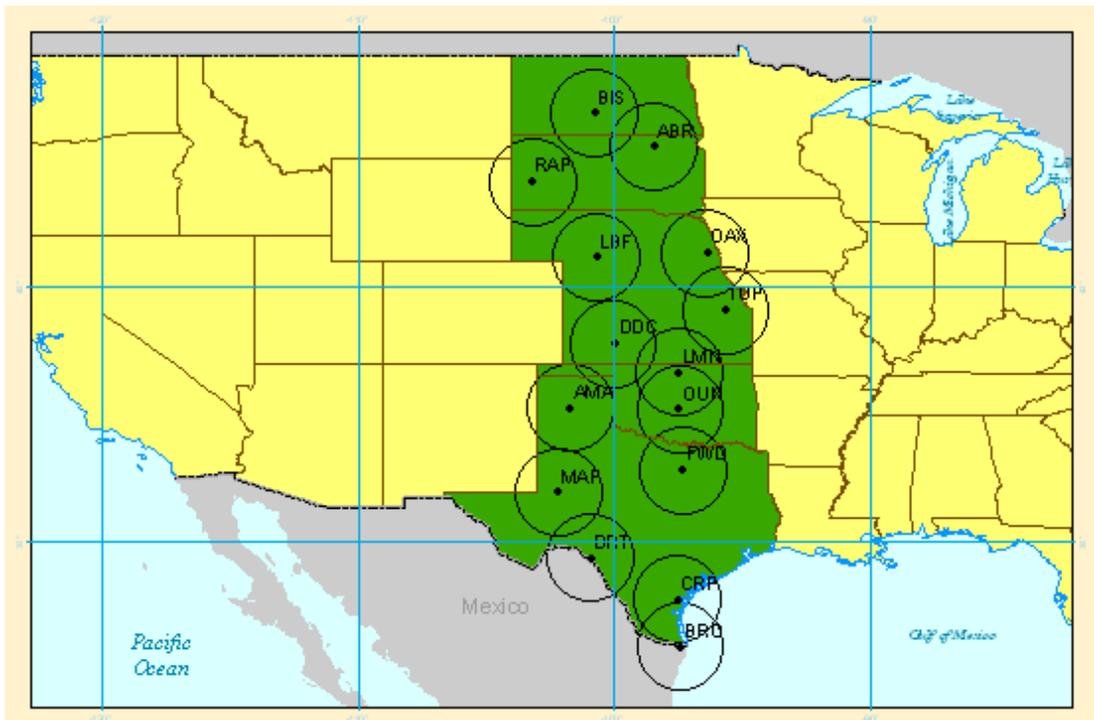


Fig. 7. The same as Fig. 1, except for the Plains.

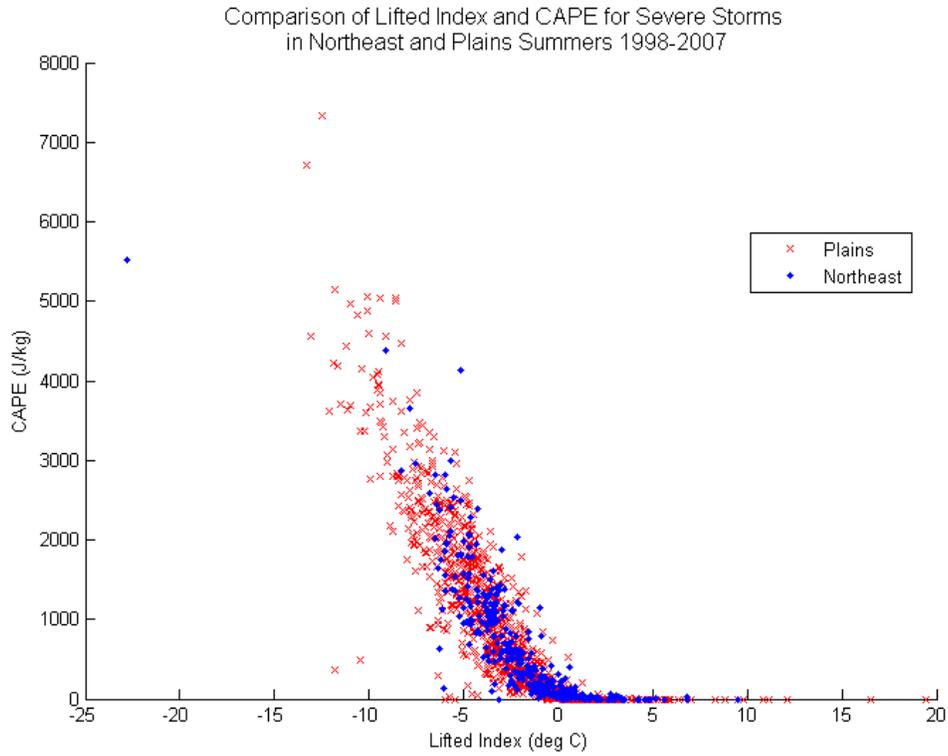


Fig. 8. Lifted Index ( $^{\circ}\text{C}$ ) and CAPE ( $\text{J kg}^{-1}$ ) for severe soundings in both the Northeast and the Plains for June, July, and August of 1998-2007.

Skill Score	Abbreviation	Formula	Limits
Probability of Detection	POD	$\text{POD} = \frac{A}{A+B}$	$0 \leq \text{POD} \leq 1$
False Alarm Ratio	FAR	$\text{FAR} = \frac{C}{A+C}$	$0 \leq \text{FAR} \leq 1$
Critical Score Index	CSI	$\text{CSI} = \frac{A}{A+B+C}$	$0 \leq \text{CSI} \leq 1$
True Skill Statistic	TSS	$\text{TSS} = \frac{A}{A+B} + \frac{C}{C+D}$	$-1 \leq \text{TSS} \leq 1$
Heidke Skill Score	S	$S = \frac{2(AD-BC)}{(A+B)(B+D) + (A+C)(C+D)}$	$-1 \leq S \leq 1$

Table 1. The definitions of the skill scores which are calculated for prediction of severe thunderstorms for each index (from Huntreser et al. 1997).

<b>Summary Statistics for Severe Soundings in the Northeast 1998-2007</b>					
<i>Index</i>	<i>Mean</i>	<i>Median</i>	<i>SD</i>	<i>Max</i>	<i>Min</i>
Showalter Index (°C)	-0.07	-0.14	2.84	10.18	-22.40
Lifted Index (°C)	-1.77	-1.96	3.02	9.48	-22.83
Lifted Index (Tv) (°C)	-2.17	-2.25	3.20	9.78	-23.58
SWEAT	235.72	228.11	89.39	1137.84	36.99
K Index (°C)	31.68	32.90	7.06	59.20	-15.30
Cross Totals (°C)	21.30	21.50	2.98	43.70	11.10
Vertical Totals (°C)	25.90	25.80	2.16	44.30	19.90
Total Totals (°C)	47.20	47.20	4.12	88.00	32.80
CAPE (J kg <sup>-1</sup> )	680.58	375.07	803.00	5515.04	0.00
CAPE (Tv) (J kg <sup>-1</sup> )	770.71	448.22	865.94	5653.00	0.00
CIN (J kg <sup>-1</sup> )	-82.71	-52.49	94.72	0.00	-611.78
CIN (Tv) (J kg <sup>-1</sup> )	-68.30	-38.13	84.57	0.00	-564.31
Equilibrium Level Pressure (hPa)	324.22	279.01	141.08	738.96	122.30

Table 2a. Mean, median, standard deviation, maximum, and minimum for the indices studied for all severe soundings in the Northeast in the months of June, July, and August 1998-2007. Indices followed by (Tv) were calculated using virtual temperature.

<b>Summary Statistics for Null Soundings in the Northeast 1998-2007</b>					
<i>Index</i>	<i>Mean</i>	<i>Median</i>	<i>SD</i>	<i>Max</i>	<i>Min</i>
Showalter Index	4.54	3.72	4.71	45.36	-13.53
Lifted Index	3.75	3.18	4.99	45.36	-12.63
Lifted (Tv.)	3.51	2.94	5.15	52.93	-14.50
SWEAT	161.61	155.67	75.44	666.62	8.00
K Index	18.05	22.80	15.97	54.41	-63.30
Cross Totals	17.39	18.70	6.10	30.73	-27.50
Vertical Totals	23.61	23.70	2.75	35.10	-14.80
Total Totals	41.00	42.60	7.53	62.21	-38.60
CAPE	154.46	0.00	383.01	6869.47	0.00
CAPE (Tv)	184.67	0.91	428.89	7372.40	0.00
CIN	-45.97	0.00	85.54	0.00	-999.94
CIN (Tv)	-39.40	-0.62	73.23	0.00	-900.11
Equilibrium Level Pressure	470.79	437.73	212.80	982.74	121.81

Table 2b. The same as Table 2a, except for all null soundings. Units for each index are the same as in Table 2a.

<b>Skill Scores Using All Soundings for Northeast Summers 1998-2007</b>										
<i>Index</i>	<i>Cutoff</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>POD</i>	<i>FAR</i>	<i>CSI</i>	<i>TSS</i>	<i>S</i>
SI (°C)	<1.9	278	65	4283	8642	0.811	0.939	0.060	0.479	0.069
LI (°C)	<0.9	301	50	4211	8758	0.858	0.933	0.066	0.533	0.079
LI Tv (°C)	<0.6	298	53	4196	8773	0.849	0.934	0.066	0.526	0.078
SWEAT	>189	252	85	4230	8630	0.748	0.944	0.055	0.419	0.060
K Index(°C)	>27	291	52	4891	8004	0.848	0.944	0.056	0.469	0.060
CT (°C)	>20	249	94	4763	8162	0.726	0.950	0.049	0.357	0.047
VT (°C)	>24	286	58	5868	7067	0.831	0.954	0.046	0.378	0.041
TT (°C)	>44	288	55	5051	7874	0.840	0.946	0.053	0.449	0.056
CAPE(J/kg)	>10	313	38	5052	7920	0.892	0.942	0.058	0.502	0.063
CAPE Tv	>40	294	57	4484	8488	0.838	0.939	0.061	0.492	0.069
CIN (J/kg)	<0	320	31	6321	6651	0.912	0.952	0.048	0.424	0.044
CIN Tv	<0	316	35	6679	6293	0.900	0.955	0.045	0.385	0.038
EL (hPa)	<520	276	34	3918	2649	0.890	0.934	0.065	0.294	0.042

Table 3a. The examined indices along with the values for 2x2 contingency tables and subsequent skill scores for analysis of all soundings in the Northeast during the months of June, July, and August during the years 1998-2007. Units for each index are the same as in Table 2a.

<b>Skill Scores Using a Sample of Null Soundings for Northeast Summers 1998-2007</b>										
<i>Index</i>	<i>Cutoff</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>POD</i>	<i>FAR</i>	<i>CSI</i>	<i>TSS</i>	<i>S</i>
SI	<1.9	278	65	117	231	0.811	0.296	0.604	0.474	0.474
LI	<0.9	301	50	127	223	0.858	0.297	0.630	0.495	0.495
LI Tv	<0.3	287	64	115	235	0.818	0.286	0.616	0.489	0.489
SWEAT	>192	246	91	114	232	0.730	0.317	0.546	0.401	0.400
K Index	>29	263	80	107	240	0.767	0.289	0.584	0.458	0.458
CT	>19	284	59	163	185	0.828	0.365	0.561	0.360	0.359
VT	>24	286	58	162	187	0.831	0.362	0.565	0.367	0.366
TT	>44	288	55	127	221	0.840	0.306	0.613	0.475	0.474
CAPE	>50	278	73	109	241	0.792	0.282	0.604	0.481	0.481
CAPE Tv	>60	283	68	117	233	0.806	0.293	0.605	0.472	0.472
CIN	<0	320	31	187	163	0.912	0.369	0.595	0.377	0.378
CIN Tv	<0	316	35	196	154	0.900	0.383	0.578	0.340	0.341
EL	<270	152	158	41	156	0.490	0.212	0.433	0.282	0.255

Table 3b. As Table 3a, except a random sample of null soundings with a size equal to the number of severe soundings is used to make each contingency table. This yields more realistic results by lowering the FAR while increasing the CSI and S. Units for each index are the same as in Table 2a.

<b>Comparison of Cutoffs for TSS and PDFs for Northeast Summers 1998-2007</b>		
<i>Index</i>	<i>Cutoff (from TSS)</i>	<i>Cutoff (from PDFs)</i>
Showalter Index	< 1.9	< 1.94
Lifted Index	< 0.9	< 0.93
Lifted (Tv)	< 0.6	< 0.73
SWEAT	> 189	> 187.5
K Index	> 27	> 26.9
Cross Totals	> 20	> 19.7
Vertical Totals	> 24	> 24.3
Total Totals	> 44	> 43.9
CAPE*	> 10	> 10.0
CAPE (Tv)*	> 40	> 27.1
CIN*	< 0	< -29.6
CIN (Tv)*	< 0	< -27.3
Equilibrium Level	< 620	< 281.3

Table 4. A comparison of the thresholds which give the best skill using the TSS and using PDFs. Indices marked with an asterisk had their cutoffs calculated using a generalized extreme value PDF, while all other cutoffs were calculated using a Gaussian PDF. Units for each index are the same as in Table 2a.

<b>Summary Statistics for Severe Soundings in the Plains 1998-2007</b>					
<i>Index</i>	<i>Mean</i>	<i>Median</i>	<i>SD</i>	<i>Max</i>	<i>Min</i>
Showalter Index	-2.44	-2.69	3.29	12.01	-17.07
Lifted Index	-3.03	-3.26	3.66	19.46	-13.30
Lifted (Tv)	-3.54	-3.82	3.89	19.65	-14.79
SWEAT	297.92	273.62	126.50	776.14	34.98
K Index	34.24	35.10	6.89	50.90	-7.30
Cross Totals	20.94	21.30	4.08	33.30	-8.30
Vertical Totals	30.16	29.90	3.81	43.50	17.70
Total Totals	51.10	51.40	4.97	71.60	23.40
CAPE	1091.53	797.94	1114.08	7333.01	0.00
CAPE (Tv)	1190.59	883.07	1186.00	7746.17	0.00
CIN	-140.61	-85.13	158.61	0.00	-858.63
CIN (Tv)	-118.24	-53.41	152.15	0.00	-813.82
Equilibrium Level	243.00	209.25	106.27	922.43	112.60

Table 5a. As Table 2a, except for the Plains. Units for each index are the same as in Table 2a.

<b>Summary Statistics for Null Soundings in the Plains 1998-2007</b>					
<i>Index</i>	<i>Mean</i>	<i>Median</i>	<i>SD</i>	<i>Max</i>	<i>Min</i>
Showalter Index	0.53	0.16	3.81	20.75	-25.28
Lifted Index	-1.09	-1.51	4.11	23.14	-49.25
Lifted (Tv)	-1.63	-2.06	4.33	23.14	-49.78
SWEAT	210.04	196.11	96.68	1261.56	5.83
K Index	27.14	29.30	10.27	60.00	-52.30
Cross Totals	18.37	18.90	4.65	50.73	-24.10
Vertical Totals	27.83	27.50	3.85	59.25	9.00
Total Totals	46.20	46.40	5.74	109.97	0.00
CAPE	835.53	418.52	1003.45	9329.38	-0.02
CAPE (Tv)	933.06	507.72	1086.51	9548.59	-0.02
CIN	-134.41	-54.40	181.20	0.00	-1489.12
CIN (Tv)	-104.68	-27.60	157.66	0.00	-1422.81
Equilibrium Level	258.54	210.25	137.55	940.86	105.37

Table 5b. As Table 2b, except for the Plains. Units for each index are the same as in Table 2a.

<b>Skill Scores Using All Soundings for Plains Summers 1998-2007</b>										
<i>Index</i>	<i>Cutoff</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>POD</i>	<i>FAR</i>	<i>CSI</i>	<i>TSS</i>	<i>S</i>
SI	<-1.5	597	328	7818	18779	0.645	0.929	0.068	0.352	0.072
LI	<-1.9	636	317	12325	14410	0.667	0.951	0.048	0.206	0.029
LI Tv	<-2.3	638	315	12815	13920	0.670	0.953	0.046	0.190	0.026
SWEAT	>250	540	379	6181	20294	0.588	0.920	0.076	0.354	0.088
K Index	>31	682	242	10870	15711	0.738	0.941	0.058	0.329	0.050
CT	>22	397	528	4971	21626	0.429	0.926	0.067	0.242	0.073
VT	>28	651	276	11694	14918	0.702	0.947	0.052	0.263	0.038
TT	>49	623	302	7888	18709	0.674	0.927	0.071	0.377	0.076
CAPE	>180	701	252	16018	10720	0.736	0.958	0.041	0.137	0.015
CAPE Tv	>220	701	252	16231	10507	0.736	0.959	0.041	0.129	0.014
CIN	<-10	752	201	17098	9641	0.789	0.958	0.042	0.150	0.016
CIN Tv	<-10	677	276	15291	11447	0.710	0.958	0.042	0.139	0.016
EL	<490	765	27	19281	1673	0.966	0.962	0.038	0.046	0.004

Table 6a. As Table 3a, except for soundings in the Plains. Units for each index are the same as in Table 2a.

<b>Skill Scores Using a Sample of Null Soundings for Plains Summers 1998-2007</b>										
<i>Index</i>	<i>Cutoff</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>POD</i>	<i>FAR</i>	<i>CSI</i>	<i>TSS</i>	<i>S</i>
SI	<-1.5	597	328	289	664	0.645	0.326	0.492	0.342	0.342
LI	<-1.8	647	306	468	488	0.679	0.420	0.455	0.189	0.189
LI Tv	<-2.1	655	298	488	468	0.687	0.427	0.455	0.177	0.177
SWEAT	>248	544	375	251	698	0.592	0.316	0.465	0.328	0.328
K Index	>30	727	197	441	510	0.787	0.378	0.533	0.323	0.322
CT	>20	580	345	397	556	0.627	0.406	0.439	0.210	0.210
VT	>28	651	276	401	552	0.702	0.381	0.490	0.282	0.281
TT	>49	623	302	297	656	0.674	0.323	0.510	0.362	0.362
CAPE	>130	720	233	607	349	0.756	0.457	0.462	0.121	0.121
CAPE Tv	>70	774	179	670	286	0.812	0.464	0.477	0.111	0.111
CIN	<-10	752	201	620	336	0.789	0.452	0.478	0.141	0.141
CIN Tv	<-10	677	276	551	405	0.710	0.449	0.450	0.134	0.134
EL	<490	765	27	702	61	0.966	0.479	0.512	0.046	0.047

Table 6b. As Table 3b, except for soundings in the Plains. Units for each index are the same as in Table 2a.