ANALYSIS OF THE FACTORS PROMOTING TORNADOGENESIS USING NCDC STORM EVENTS AND IGRA DATA

By

Adam Cavender

A THESIS

Submitted to the Faculty of the Rosenstiel School of Marine and Atmospheric Science

Coral Gables, Florida

May 2014
UNIVERSITY OF MIAMI

A thesis submitted in partial fulfillment of the requirements for
Departmental Honors in Marine and Atmospheric Science

ANALYSIS OF THE FACTORS PROMOTING TORNADOGENESIS USING NCDC
STORM EVENTS AND IGRA DATA

Adam Cavender

Approved:

Name of Thesis Advisor
Chair of Thesis Committee
Academic Rank and Department

Name of Committee Member
Academic Rank and Department

Name of Committee Member
Academic Rank and Department
Abstract

Official tornado reports from the National Climatic Data Center’s Storm Events Database and radiosonde sounding data from the Integrated Global Radiosonde Archive were gathered to analyze linear correlations between various sounding parameters and tornadogenesis. Proximity was defined based on previous studies as a tornado occurring within 200km and three hours of radiosonde release. For each sounding, a total of 35 parameters were computed, and counts of the number and intensity of tornadoes occurring in proximity to each observed sounding were taken. Only 00Z soundings taken in proximity to at least one tornado were considered. Low cloud bases, thermodynamic instability, and high helicity and shear values are shown to be correlated with tornadogenesis, in agreement with past literature. Helicity and shear were found to be the best discriminators between soundings taken in proximity to one or more strong tornadoes and those taken in proximity to only weak tornadoes. The results of this research suggest that the most-unstable parcel lifted condensation level is the most useful approximation to cloud base height in forecasting tornadoes. An experimental parameter that reflects the average instability of parcels within the effective inflow layer shows potential to be an effective tornado forecasting aid. Additionally, a new composite parameter was developed and verified using multivariate linear regression that appears to be strongly correlated with tornadogenesis.
1. Introduction

a. Motivation

Tornadogenesis has historically been one of the most difficult atmospheric phenomena for meteorologists to predict. Complex interactions of many dynamical processes and the innate danger of collecting data within the vicinity of tornadic supercell storms hinder improvements in tornado forecasting. Tornadoes are capable of immense property destruction and personal devastation, necessitating the need for timely warnings and accurate forecasts. While much about tornadogenesis has been revealed by years of study, many aspects of the process remain a mystery. An improvement in tornado forecasting has the potential to save many lives and millions of dollars in property damage. Thus, continued study in the area of tornadogenesis is imperative to those living in regions of the United States commonly affected by tornadoes - particularly, in Tornado Alley.

My motivation for doing this research is two-fold. First of all, I would like to improve my own understanding of the factors that lead to the formation of tornadoes; as a storm chaser, I have a deep interest in tornado forecasting and the physical processes behind both tornadoes and supercell thunderstorms. Secondly, I would like to contribute to the meteorological community’s understanding of the science of tornadic storms in any way that I can. By using a novel technique to analyze the factors leading to tornadogenesis and applying this technique to two experimental parameters, I hope to provide the meteorological community with further insights into both how tornadoes work and how to most effectively research these complex processes in the future.
b. Existing Literature

While plentiful research has been conducted in the area of tornadogenesis, many important questions remain unanswered. For many years, several factors have been known to support an increased likelihood of tornado formation; these factors include low cloud bases, thermodynamic instability, and vertical wind shear (both directional and speed shear).

Cloud base height is estimated using a parameter called the lifted condensation level, or LCL. Rasmussen and Blanchard (1998) found that the LCL was the most useful parameter for discriminating between supercell storms that produced strong tornadoes and those that produced weak or no tornadoes. LCLs were about 500m lower in the case of strong tornadoes. Similar results were found by Edwards and Thompson (2000). LCL heights are lower when the low-level environment is more humid; thus, moisture is another important factor to consider in tornado forecasting.

Thermodynamic instability is quantified in a variable called the convective available potential energy, or CAPE. High values of CAPE are indicative of a potential for strong storm updrafts (high vertical velocities). The calculated value for CAPE depends on many factors including the vertical distributions of temperature (i.e., lapse rates) and moisture. Steep mid-level lapse rates are related to higher CAPE values and hence greater instability. Doswell (1985) showed that the superposition of low-level moisture and steep lapse rates (equating to high CAPE values) was ideal for supercell storm and tornado formation. Craven (2000) found that steep lapse rates and greater CAPE values were associated with most major tornado outbreaks.
Vertical wind shear, both directional and speed, can be quantified in multiple ways. Bulk shear, which can be calculated across various atmospheric layers, is equal to the magnitude of the vector difference between the upper-level wind and the lower-level wind. Weisman (1996) showed that shear values over the lowest 4-6km above ground level of 20 m s\(^{-1}\) are sufficient to promote supercell formation. Another variable commonly used to assess the vertical wind field is storm-relative helicity, or SRH. This variable quantifies the extent of helical motion in the environmental wind distribution. Edwards and Thompson (2000) showed a statistically significant difference between the mean 0-1km SRH values for supercells with significant tornadoes and those with weak or no tornadoes.

Thompson (2003) analyzed all of the aforementioned variables and showed that CAPE increases, LCL decreases, 0-6km and 0-1km bulk shear values increase, and 0-1km and 0-3km SRH values increase as the intensity of tornadoes increases. Furthermore, because these variables represent different features of the thermodynamic and dynamic environments yet all increase the likelihood of tornadoes, Thompson (2003) examined the efficacy of “composite parameters” in distinguishing between storm/tornado types. While several composite parameters were evaluated, one of the most commonly used operationally (mainly by the National Weather Service’s Storm Prediction Center in Norman, Oklahoma) is the Significant Tornado Parameter, or SigTor. The SigTor parameter was shown to discriminate very strongly between storms that produced strong and only weak tornadoes.
Thompson, Edwards, and Mead (2005) updated the formulation of SigTor. They replaced the 0-6km bulk shear with a parameter called “effective bulk shear” (defined on pg. 31) and the 0-3km SRH with a parameter called “effective SRH.” The effective bulk shear is similar to the 0-6km bulk shear but is said to better reflect the shear relevant to supercells in the cases of exceptionally tall or short storms. The effective SRH is also better tailored to the environment in question; it is calculated across a layer referred to as the “effective inflow layer” (defined on pg. 30).

c. Goals

One goal of this research is to analyze, using a novel linear correlation technique, the efficacy of the various parameters mentioned above in predicting tornadogenesis. I hope that this linear correlation technique will provide a new method for either verifying or challenging the results of some of the studies mentioned above. Another goal is to examine the potential abilities of two new parameters, “effective CAPE” and “effective CIN” (defined on pg. 31), to assist in predicting the formation of tornadoes. The main goal of this project, however, is to use the results of correlation analysis to either improve upon the existing composite SigTor parameter or formulate a new tornado forecasting parameter altogether. While many of the independent factors leading to tornadogenesis have been examined in detail in previous studies, analysis of linear correlations has not yet been used to assist in this area of research, and it remains unclear what the ideal combination of parameters is to promote the formation of tornadoes.
2. Materials and Methods

a. Data Sources

The branch of the National Oceanic and Atmospheric Administration (NOAA) that deals with the study of climate and with maintaining archives of climatic data is called the National Climatic Data Center, or NCDC. The NCDC maintains a record of verified weather events in a database called the “Storm Events Database.” Within this archive are contained records of every confirmed tornado within the United States from January 1\textsuperscript{st}, 1996 through present. From the beginning of the database’s records through January 13\textsuperscript{th}, 2007, the F-scale, or the Fujita Scale, is used to record tornadoes’ intensities. Beyond this date, the EF-scale, or the Enhanced Fujita Scale, which is based on observed storm damage in addition to any measured wind speeds, is used to record intensities. Using a Java program, tornado data were assembled from the Storm Events Database. For each tornado, the following information was recorded: date, local time of tornado, (E)F-rating, state in which the tornado occurred, the time zone in which the tornado occurred, and the latitude and longitude coordinates at which the tornado was initially reported. The final compiled tornado list contained data for 24,929 tornadoes; this includes all confirmed tornadoes that occurred within the United States between January 1\textsuperscript{st}, 1996 and October 31\textsuperscript{st}, 2013.

After compiling the tornado data from the Storm Events Database, a list of all registered radiosonde release locations within the United States was gathered from the NCDC. These data are contained within the NCDC’s Integrated Global Radiosonde Archive, or IGRA (“Integrated”). Before radiosonde data are saved in the IGRA, they
undergo many quality control measures including “sanity” checks, internal consistency checks, climatologically-based checks, and data completeness checks. Due to these increasingly stringent quality control measures, no additional data “checks” were performed in this analysis. The IGRA contains not only the station IDs and latitude/longitude coordinates of every registered radiosonde release location, but also raw data from every sounding taken from these locations from 1996 to present (and well before these dates). After a list of 122 radiosonde release locations within the United States was compiled from the IGRA (for each location, the station ID and its latitude/longitude coordinates were recorded), soundings taken within proximity to each tornado event were located.

To determine the distance and time constraints defining “proximity,” a previous study by Craven (2004) was considered. In this study, proximity was defined as a tornado occurring within 185km of the radiosonde release location and within 3 hours before/after the release time. Additionally, in this study, only 00Z soundings were considered (vastly fewer tornadoes occur during the 6-hour span centered on 12Z). Two previous studies mentioned in Craven (2004) used distance constraints of 80km and 400km. In this study, proximity will be defined as a tornado occurring within 200km of the radiosonde release location and within the 6-hour span centered on the radiosonde’s release time. Since, as in Craven’s study, only 00Z soundings will be considered, the appropriate time span for tornadoes will be 2100-0300Z.

A program was used to go through the entire list of just under 25,000 tornadoes. For each tornado, its time zone was used to convert its local time to its UTC time; the
tornado needed to fall within the time range of 2100-0300Z to be considered. Then, the tornado’s straight-line distance from each of the 122 radiosonde release locations was determined using the Haversine formula below (“Calculate”):

\[
D(\phi_1, \lambda_1, \phi_2, \lambda_2) = r_{Earth} * c
\]

\[
c = 2 * a \tan 2(\sqrt{a}, \sqrt{1-a})
\]

\[
a = \sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + \cos(\phi_1)\cos(\phi_2)\sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)
\]

In the above formulae, \( \phi \) represents latitude in radians, \( \lambda \) represents longitude in radians, \( r_{Earth} \) is the mean Earth radius in kilometers (approximately 6,371km), \( \tan2 \) is a Java function that returns the angle \( \theta \) from the conversion of rectangular coordinates \((x, y)\) to polar coordinates \((r, \theta)\), and \( D \) represents straight-line distance in kilometers.

If the tornado was found to fall within the time range of 2100-0300Z and within 200km of one or more radiosonde release locations, these locations’ station IDs were recorded along with the date during which a 00Z sounding could have been taken at that location (not all soundings are available or were ever taken). After this procedure, a program was used to go through this list of possible soundings. For every unique sounding location/date, the total number of tornadoes that occurred within proximity to that sounding was recorded. The total number of strong \((E)F \geq 2\) tornadoes that occurred within proximity to the sounding was recorded as well. Finally, the value for a variable that will be referred to as the “tornado parameter” was recorded as well. This variable is equal to the sum of the \((E)F\)-ratings+1 of all of the tornadoes that occurred
within proximity to the sounding. For example, if 3 EF0 tornadoes occurred within proximity to a sounding, the “tornado parameter” for that sounding would be equal to 3. An additional EF2 tornado would give that sounding a tornado parameter value of 6. This value was recorded in order to combine, into one variable, both the number of proximal tornadoes and the strengths of these tornadoes.

Once this list of unique soundings was compiled (containing each 00Z sounding’s station ID, date, number of tornadoes, number of strong tornadoes, and tornado parameter), the raw data for each sounding (3085 soundings in total) were found within the IGRA. Next, from this raw data, 35 total parameters were computed for each sounding. The first step in computing these parameters is to interpolate the sounding data.

b. *Vertical Sounding Interpolation*

Radiosonde data retrieved from the NCDC’s IGRA is not vertically continuous. Measurements of atmospheric variables, such as temperature and pressure, are not taken continually; rather, these measurements are typically made at between 50 and 100 levels between the radiosonde’s release height and its maximum height. In order to both account for vertical variations in these variables between measurements and simplify computation of vertical integrals, interpolations of the sounding data were made at one-meter intervals. The necessity of such interpolations becomes evident when considering the computation of vertical wind shear. If wind vectors in a particular sounding were measured at heights (above ground level, or AGL) of 0m, 500m, and 1500m, then in the computation of 0-1km AGL wind shear, a vector intermediate to the
500m and 1500m wind vectors should be used. Use of either the 500m or 1500m wind
vector would neglect the vertical variation in both wind speed and direction between
these levels. Similar reasoning can be employed to justify the interpolation of scalar
quantities including temperature and pressure.

\textit{b1. Scalar Interpolation}

Five scalar variables were interpolated linearly at one-meter intervals: pressure,
temperature, virtual temperature, vapor pressure, and height. While vertical variations
of these variables are never perfectly linear, the values are measured frequently enough
(about every 100-200m) that the use of more computationally intensive interpolation
schemes, such as employing the hydrostatic approximation to estimate vertical pressure
variation, would likely result in meager improvements in accuracy. Therefore, the
following linear interpolation scheme was used:

\begin{equation}
\varphi_z = \varphi_{z_1} + \frac{z - z_1}{z_2 - z_1} (\varphi_{z_2} - \varphi_{z_1})
\end{equation}

In the above equation, \( \varphi \) represents either pressure, temperature, virtual
temperature, or vapor pressure, \( z \) represents the height at which the value of \( \varphi \) is
desired, \( z_1 \) represents the greatest height less than or equal to \( z \) at which a measurement
of \( \varphi \) has been taken, and \( z_2 \) represents the lowest height greater than \( z \) at which a
measurement of \( \varphi \) has been taken. In this equation, height is used as the vertical
coordinate. However, to interpolate heights of specified pressure levels (say, 500mb),
pressure was used as the vertical coordinate in the following interpolation scheme:
In the above equation, \( p \) is the pressure of the surface whose height is desired, \( p_1 \) is the lowest measured pressure greater than or equal to \( p \), and \( p_2 \) is the greatest measured pressure less than \( p \).

\[ z_p = z_{p_1} + \frac{p - p_1}{p_2 - p_1} (z_{p_2} - z_{p_1}) \]  

(b2. Vector (Wind) Interpolation)

While scalar variables can be interpolated linearly, interpolation of vectors necessarily requires a different approach. In atmospheric soundings, the only vector variable measured is wind. To see why a non-linear interpolation scheme is necessary, consider the case of a 500m above-ground-level (AGL) westerly 5 m s\(^{-1}\) wind and a 700m AGL southerly 5 m s\(^{-1}\) wind. In the NCDC’s IGRA, winds are recorded in terms of their \( u \) (horizontal) and \( v \) (vertical) components. Therefore, the 500m AGL wind would be recorded as \( u_{500} = 5 \text{ m s}^{-1}, v_{500} = 0 \text{ m s}^{-1} \) and the 700m AGL wind would similarly be recorded as \( u_{700} = 0 \text{ m s}^{-1}, v_{700} = 5 \text{ m s}^{-1} \). Linear interpolation of these winds would yield \( u_{600} = 2.5 \text{ m s}^{-1}, v_{600} = 2.5 \text{ m s}^{-1} \). While the direction of this interpolated 600m AGL wind appears valid, the speed of this wind would be \((2.5 \text{ m s}^{-1})^2 + (2.5 \text{ m s}^{-1})^2)^{0.5}\), or about 3.5 m s\(^{-1}\). This assumption is nonsensical; it is more likely that the 600m wind has a speed of about 5 m s\(^{-1}\).

To more accurately interpolate wind vectors, wind speed and wind direction are interpolated independently. This technique is described by Gorman (2009) as a “robust procedure” and as “the default method in many applications.” Wind speed, being a
 scalar quantity, was linearly interpolated according to the scheme described in “Scalar Interpolation” above. Wind speed was calculated at each level using the following equation:

\[
|\vec{V}_z| = \sqrt{u_z^2 + v_z^2}
\]  

(6)

Wind direction was interpolated according to the following scheme. First, wind direction was computed for the upper-level and lower-level winds as an angle greater than or equal to 0° and less than 360° using the following formula:

\[
\theta_z = \alpha \tan 2(v_z, u_z) \times \frac{180°}{\pi}
\]  

(7)

In the Java programming language as well as many other languages, the atan2 function “returns the angle theta from the conversion of rectangular coordinates (x,y) to polar coordinates (r, theta)” (“Class”). This angle is returned in units of radians, hence the conversion to degrees in the formula above. It is the angle between the positive x-axis and the specified point; it is positive for counter-clockwise angles and negative for clockwise angles. The range of this function is (-\(\pi\),\(\pi\)) (“Atan2”). Because an angle in the range [0°,360°) was desired, if the angle given by the formula above was less than 0°, 360° was added. Next, the angle between the upper-level and lower-level winds was computed using the dot product:

\[
\theta_{\text{between}} = \cos^{-1} \left( \frac{\vec{V}_{z1} \cdot \vec{V}_{z2}}{|\vec{V}_{z1}| \cdot |\vec{V}_{z2}|} \right) \times \frac{180°}{\pi}
\]  

(8)
The inverse cosine function has a range of $[0^\circ, 180^\circ]$, thereby giving the smaller of the two explementary (totaling $360^\circ$) angles between the winds. To determine whether the winds were veering or backing with increasing height, the angle between the wind vectors was added to the angle of the lower wind. If the sum was greater than or equal to $360^\circ$, then $360^\circ$ was subtracted from the sum. If this angle was equal to the angle of the upper wind, then the winds were rotating counter-clockwise with height. If this sum was not equal to the angle of the upper wind, then the winds were rotating clockwise. If the angle between the two winds was exactly $180^\circ$ (a rare occurrence), then the assumption was made that the rotation was clockwise. The following equation was then used to interpolate the direction of the wind at the desired height:

$$\theta_z = \theta_{z_1} \pm \frac{z - z_1}{z_2 - z_1} \cdot \theta_{\text{between}}$$ (9)

In this equation, the “+” was used if the winds were determined to rotate counter-clockwise, and the “-” was used if the winds were determined to rotate clockwise. If the resulting angle fell outside of the $[0^\circ, 360^\circ)$ range, $360^\circ$ was added or subtracted accordingly.

Finally, with the interpolated speed and direction computed, the $u$ and $v$ components of the wind were calculated using the following equations:

$$u_z = |\vec{V}_z| \cos(\theta_z)$$

$$v_z = |\vec{V}_z| \sin(\theta_z)$$ (10,11)
The wind vector interpolation scheme described above, while more complex than the scheme used for scalar variables, provides more reasonable estimates of wind speed and direction between levels at which the wind was measured directly.

c. Constants and Formulas

Accurate calculation of thermodynamic and dynamic parameters requires the use of many empirically determined atmospheric constants and derived relationships. Many of these can be found in a paper by David Bolton (1980) titled “The Computation of Equivalent Potential Temperature.” The constants and formulas listed below are taken from Bolton. The following formula is used to convert from degrees Celsius to Kelvins:

\[ T_K = T_C + 273.15 \quad (12) \]

Here, \( T_K \) is the absolute temperature (in Kelvins) and \( T_C \) is the temperature in degrees Celsius. Following are known as the dry air constants:

\[ R_d = 287.04 \quad (13,14) \]
\[ c_{pd} = 1005.7 \]

Here, \( R_d \) is the dry air gas constant, which has units of J kg\(^{-1}\) K\(^{-1}\). \( c_{pd} \) is the specific heat capacity at constant pressure for dry air, and it has units of J kg\(^{-1}\) K\(^{-1}\) as well. Below are constants for liquid water:
Above, c\textsubscript{w} is the specific heat capacity for liquid water, which has units of J kg\textsuperscript{-1} K\textsuperscript{-1}. L\textsubscript{v} is the latent heat of vaporization and has units of J kg\textsuperscript{-1}. This expression for L\textsubscript{v} implies that the latent heat of vaporization of water is temperature-dependent. In this equation, T\textsubscript{C} has units of degrees Celsius. Below are the constants for water vapor:

\[ R\textsubscript{v} = 461.50 \]
\[ c\textsubscript{pv} = 1875 \]  

(17,18)

Here, R\textsubscript{v} is the gas constant for water vapor and has units of J kg\textsuperscript{-1} K\textsuperscript{-1}. c\textsubscript{pv} is the specific heat capacity of water vapor at constant pressure and has units of J kg\textsuperscript{-1} K\textsuperscript{-1}. The gas constants for dry air and for water vapor are commonly combined in thermodynamic equations to yield the following unitless constant:

\[ \varepsilon = \frac{R\textsubscript{d}}{R\textsubscript{v}} = 0.62197 \]  

(19)

The gravitational constant near the Earth’s service is given approximately by the following:

\[ g = 9.81 \]  

(20)

This constant, g, is an acceleration and therefore has units of m s\textsuperscript{-2}. Finally, Bolton’s paper gives us an accurate approximation of equivalent potential temperature:
In these formulae, $\theta_e$ is the equivalent potential temperature measured in Kelvins, $T_K$ is the temperature in Kelvins, $p$ is the pressure in millibars (hPa), $w$ is the water vapor mixing ratio in g kg$^{-1}$, $e$ is the vapor pressure in millibars (hPa), and $T_L$ is the temperature at the lifted condensation level in Kelvins.

Equivalent potential temperature is defined as the final temperature of a parcel of air after it has undergone three processes: dry-adiabatic lifting to its lifted condensation level, pseudo-wet adiabatic lifting from the lifted condensation level to a great height (all condensed water is assumed to fall out immediately), and finally dry-adiabatic descent to the 1000mb level (Bolton 1980). Equivalent potential temperature, according to the National Weather Service (“Theta-e”), is useful in diagnosing atmospheric instability because it is directly related to the amount of heat energy contained in an air parcel. Simply put, parcels possessing high equivalent potential temperature can be viewed as warm, moist (and therefore unstable) parcels. For this reason, equivalent potential temperature will later be used to identify both the most unstable air parcels and the most negatively buoyant parcels in atmospheric soundings.

Saturation vapor pressure is another thermodynamic value of great importance. It represents the maximum amount of water vapor that a parcel can hold at a given
temperature. If the temperature of a parcel whose vapor pressure equals its saturation vapor pressure (a saturated parcel) is lowered, condensation will occur, provided that cloud condensation nuclei (CCN) are present. Several empirical formulas that estimate saturation vapor pressure at a given temperature are present in meteorological literature. One of the most commonly seen formulas, accurate to 0.1% within the temperature range of \(-30°C\) to \(35°C\), is given by Bolton (1980):

\[
e_s(T) = 6.112 \exp \left( \frac{17.67T}{T + 243.5} \right)
\]

(23)

In this formula, \(T\) is measured in degrees Celsius and \(e_s\) is given in hPa, or mb. While Bolton’s formula is quite accurate and sufficient for most purposes, to ensure the highest possible accuracy, a formula that accounts for saturation vapor pressure’s slight dependence on pressure was used. The saturation vapor pressure for pure water vapor over a plane surface of water is solely dependent on temperature; however, due to departures from the ideal gas law, the saturation vapor pressure for moist air over a plane surface of water shows a slight dependence on pressure. This can be accounted for with an “enhancement factor,” which is defined as the ratio of the saturation vapor pressure for moist air to that of pure water vapor over a plane surface of water. Alduchov and Eskridge (1996) give the following equation for calculating saturation vapor pressure, with the enhancement factor included:

\[
e_s(p,T) = 1.00071 \exp(0.0000045 p) \times 6.1094 \exp \left( \frac{17.625T}{T + 243.04} \right)
\]

(24)
In the above equation, \( p \) is in hPa or mb, \( T \) is in degrees Celsius, and \( e_s \) is given in hPa or mb. Since, in this analysis, saturation vapor pressure will be computed in a wide range of pressure conditions, it is important to use an approximation that takes pressure-dependence into account.

Vapor pressure and saturation vapor pressure are important variables because they collectively tell us how close a parcel is to a saturated state (vapor pressure tells us how much moisture is actually contained in a parcel, and saturation vapor pressure tells us the maximum amount of moisture the parcel could contain). Similar variables, referred to as mixing ratio \((w)\) and saturation mixing ratio \((w_s)\), also tell us how much water vapor a parcel contains and how much it could theoretically contain, respectively. Unlike vapor pressure, which is measured in units of Pascals, mixing ratio is unitless; it is defined as the ratio of the mass of water vapor in a parcel to the mass of dry air in the parcel. The water vapor mixing ratio can be computed from vapor pressure and pressure as follows (“Mixing”):

\[
w = \frac{\varepsilon e}{p - e}
\]  

(25)

Above, \( \varepsilon \) is the unitless ratio of gas constants defined previously, \( e \) is the vapor pressure measured in Pascals, \( p \) is pressure measured in Pascals, and \( w \) is the unitless mixing ratio. To convert the mixing ratio into the unit of g kg\(^{-1}\), as is commonly seen, simply multiply the unitless mixing ratio by 1000. The saturation mixing ratio is defined similarly (“Mixing”):

\[
w_s = \frac{\varepsilon e_s}{p - e_s}
\]  

(26)
Here, $e_s$ is the saturation vapor pressure measured in Pascals, and $w_s$ is the unitless saturation mixing ratio.

The quotient of mixing ratio and saturation mixing ratio gives us a fraction that represents how close the parcel is to saturation. When this fraction is converted into a percent, it is referred to as the relative humidity (RH). Thus, the relative humidity is given by the following equation:

$$RH = \frac{w}{w_s} \times 100\%$$  \hspace{1cm} (27)

Above, $w$ is the mixing ratio, $w_s$ is the saturation mixing ratio measured in the same units as the mixing ratio, and RH is the relative humidity given as a percent.

Quantification of the moisture content of an air parcel is important for several reasons. For example, we need to know the moisture content of a parcel to estimate cloud base height. Also, the amount of water vapor in a parcel affects that parcel’s density. Water vapor, being less dense than dry air, serves to lower the density of a parcel as moisture content increases. For this reason, moisture content must be taken into consideration when determining whether a parcel is buoyant in its environment. The ideal gas law is given below:

$$p = \rho RT$$  \hspace{1cm} (28)
Here, \( p \) is pressure, \( \rho \) is density, \( R \) is the gas constant, and \( T \) is absolute temperature. This equation, when rearranged and applied to a totally dry air parcel, shows us that temperature is inversely proportional to density. Therefore, for air containing no water vapor, warmer parcels will be less dense than cooler parcels; i.e., they will be positively buoyant when in a cooler environment. For moist parcels, however, water vapor content must be considered to determine buoyancy. For this reason, a variable called virtual temperature has been developed. The equation for virtual temperature is given below ("Virtual"):

\[
T_v = T \left( 1 + \frac{w}{1 + w} \right) \frac{\varepsilon}{1 + \varepsilon}
\]  

(29)

In this formula, \( w \) is the unitless mixing ratio, \( \varepsilon \) is the unitless ratio of gas constants defined previously, \( T \) is temperature in Kelvins, and \( T_v \) is the virtual temperature in Kelvins. The virtual temperature for a given parcel is defined as the temperature at which a totally dry parcel would have the same density as the parcel in question. The formula above shows that the virtual temperature is always greater than or equal to the parcel’s temperature. This is because at a given temperature, a moist parcel will always be less dense than a dry parcel. Therefore, to equate the parcel densities, the dry parcel’s temperature must be increased. Virtual temperature is useful in meteorology because comparison of two parcels’ virtual temperatures allows us to determine (regardless of each parcel’s moisture content) which parcel is less dense; a
parcel with a greater virtual temperature than its environment will always be positively buoyant. For this reason, virtual temperature is used in calculations of thermodynamic parameters related to buoyancy (e.g., CAPE).

As an unsaturated parcel is lifted upwards in the atmosphere, its temperature decreases at what is known as the dry adiabatic lapse rate ($\Gamma_d$). This temperature change occurs due to adiabatic cooling as the parcel’s pressure adjusts to that of the environment (this is assumed to occur instantaneously). The following formula gives the dry adiabatic lapse rate:

$$\Gamma_d = \frac{g}{c_{pd}} = 0.00975$$  \hspace{1cm} (30)$$

In the above formula, $g$ is the acceleration due to gravity, $c_{pd}$ is the specific heat capacity of dry air at constant pressure, and $\Gamma_d$ is given in units of Kelvins per meter (K m$^{-1}$). The dry adiabatic lapse rate is more commonly seen in units of Kelvins per kilometer. With these units, the dry adiabatic lapse rate is about equal to 9.8 K km$^{-1}$.

Once a lifted parcel reaches saturation, it no longer cools at the dry adiabatic lapse rate. Adiabatic cooling continues as the parcel rises, but it is partially offset by latent heat release as condensation occurs. As a result, the parcel begins to cool at a slower rate, called the moist adiabatic lapse rate ($\Gamma_m$). Different versions of the moist adiabatic lapse rate have been calculated; the reversible moist adiabatic lapse rate assumes that condensed water is carried with the parcel, while the irreversible moist adiabatic lapse rate (also called the pseudoadiabatic lapse rate) assumes that condensed
water is immediately removed from the parcel (it falls out). This distinction is important because the condensed water alters the heat capacity of the parcel and therefore affects the lapse rate. The Storm Prediction Center (SPC) in Norman, Oklahoma uses the pseudoadiabatic lapse rate assumption in its calculations of thermodynamic parameters (Thompson “Explanation”). Therefore, in this analysis, the pseudoadiabatic approximation will also be used. The following formula gives the pseudoadiabatic lapse rate (“Pseudoadiabatic”):

\[
\Gamma_m = g \frac{(1 + w) \left( 1 + \frac{L_v w}{RT} \right)}{c_{pd} + w c_{pv} + \frac{L_v^2 w (\epsilon + w)}{RT^2}}
\]

(31)

Above, \( g \) is the acceleration due to gravity, \( w \) is the unitless mixing ratio of water vapor, \( L_v \) is the latent heat of vaporization of water, \( R \) is the dry air gas constant, \( T \) is temperature in Kelvins, \( c_{pd} \) is the specific heat capacity of dry air at constant pressure, \( c_{pv} \) is the specific heat capacity of water vapor at constant pressure, \( \epsilon \) is the unitless ratio of gas constants defined previously, and \( \Gamma_m \) is the pseudoadiabatic moist lapse rate given in Kelvins per meter (K m\(^{-1}\)).

d. Calculation of Sounding Parameters

The parameters calculated in this analysis will be divided into two categories: thermodynamic parameters and dynamic parameters. Thermodynamic parameters are those involved with convective processes and instability, and dynamic parameters are those involved with the vertical distributions of wind speed and direction.
**d1. Thermodynamic Parameters**

The calculation of thermodynamic parameters from sounding data relies on a fundamental theory of meteorology called “parcel theory.” In this theory, a parcel is regarded as a volume of air with uniform qualities (temperature, pressure, vapor pressure) that remains isolated from its surroundings (its “environment”). However, it is assumed that a parcel’s pressure instantaneously equilibrates with the environmental pressure. When calculating thermodynamic parameters such as the lifted condensation level, it is important to first define the parcel in consideration.

Four different initial parcels were used in this analysis; each parcel is defined by its temperature, its pressure, and its vapor pressure. The “surface parcel” is defined by the temperature, pressure, and vapor pressure values measured at the lowest level in the sounding (ground level). This parcel originates at ground level. The two “mixed layer parcels” (variable and fixed) are defined by the mean temperature and vapor pressure values within the mixed layer. The “variable mixed layer” was defined by the value of the mixed layer height calculated by the National Climatic Data Center for each sounding in the IGRA. The “fixed mixed layer” was defined as the lowest 100mb of each sounding; this is the mixed layer that the Storm Prediction Center uses in its mixed layer calculations (Thompson “Explanation”). Each mixed layer parcel was assumed to originate at a height halfway from the surface to the top of the mixed layer; the pressure at this height was used to define the pressure of the parcel. Finally, the “most unstable parcel” was determined by analyzing the vertical equivalent potential temperature profile. Following the Storm Prediction Center’s method (Thompson “Explanation”),
equivalent potential temperature is calculated at each height in the lowest 300mb of the sounding. Once the level of the maximum equivalent potential temperature value is determined, the most unstable parcel is given values of temperature, pressure, and vapor pressure equal to those of the environment at that level. This parcel originates at the height of maximum equivalent potential temperature in the lowest 300mb of the sounding. Table 1, below, shows the characteristics of each of the four parcels.

*Table 1 - Temperature (T) and Vapor Pressure (e) Values and Heights of Origin for the Four Parcels Considered (sfc. = surface, $p_0$ = surface pressure)*

<table>
<thead>
<tr>
<th>Parcel</th>
<th>Characteristics</th>
<th>Height of Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface-Based</td>
<td>Measured T/e at surface</td>
<td>Surface</td>
</tr>
<tr>
<td>Fixed Mixed Layer</td>
<td>Mean T/e in lowest 100mb</td>
<td>Halfway from sfc. to $p_0$-100mb level</td>
</tr>
<tr>
<td>Variable Mixed Layer</td>
<td>Mean T/e between surface and NCDC mixed layer height</td>
<td>Halfway from sfc. to mixed layer height</td>
</tr>
<tr>
<td>Most Unstable</td>
<td>T/e of max. theta-e parcel in lowest 300mb</td>
<td>Height of max. theta-e parcel</td>
</tr>
</tbody>
</table>

For each of the four parcels described above, the following five parameters were calculated: lifted condensation level (LCL), level of free convection (LFC), equilibrium level (EL), convective available potential energy (CAPE), and convective inhibition (CIN).

The lifted condensation level, or LCL, is the height at which a lifted parcel becomes saturated; it is essentially an estimate of cloud base height. The LCL is the only one of the five aforementioned parameters whose value does not depend on the environmental temperature and moisture profile. To calculate the LCL, the parcel in question was raised in one-meter increments. After each iteration, the parcel’s pressure
was adjusted to the environmental pressure and its temperature was decreased according to the dry adiabatic lapse rate. The relative humidity was then calculated by dividing the parcel’s water vapor mixing ratio (which remains constant until the LCL is reached) by its calculated saturation mixing ratio and multiplying the quotient by 100. The first height at which the parcel’s relative humidity was greater than or equal to 100% was determined to be the parcel’s LCL.

The level of free convection, or LFC, is the height at which the parcel becomes positively buoyant. To determine the LFC, the parcel was raised in one-meter increments from the LCL. After each iteration, the parcel’s pressure was equilibrated with the environment and its temperature was decreased according to the calculated pseudoadiabatic moist lapse rate. The parcel’s virtual temperature was then calculated and compared with the environmental virtual temperature. The first height at which the parcel’s virtual temperature equaled or exceeded that of the environment was set as the LFC.

The equilibrium level, or EL, is the height above the LFC at which the parcel becomes negatively buoyant. This height approximates that at which the top of a thunderstorm (the “anvil”) is observed. In reality, rising parcels often rise beyond the equilibrium level to the maximum parcel level (MPL); this phenomenon is referred to as an overshooting top. The vertical distance between the EL and the MPL will be greater for stronger updrafts, as these updrafts possess greater kinetic energy. To determine the EL, a parcel with identical temperature, pressure, and vapor pressure to the parcel lifted to the LFC was created. This parcel was then lifted at one-meter increments in the
fashion described above until the parcel’s virtual temperature became less than that of the environment. This level was set as the parcel’s EL.

A complication to the method described above arises if there is a strong low- to mid-level temperature inversion. In such situations, a lifted parcel may rise to its apparent LFC, reach an apparent EL, and then (with further forcing or sufficient kinetic energy) reach another level at which it becomes positively buoyant. In other words, if a temperature inversion is present, a parcel can have two LFCs and two ELs. For this reason, each parcel was taken beyond its first EL to see whether there existed another LFC. If a second LFC for the parcel existed and was located within 6 kilometers of ground level, this was taken to be the parcel’s LFC, and the corresponding EL was determined. An additional complication arises if no LFC exists. In this situation, the LFC and EL were both set equal to the LCL.

Convective available potential energy, or CAPE, provides an estimate of updraft strength within a storm; high CAPE values suggest strong updrafts. CAPE represents the theoretical maximum kinetic energy that a rising parcel can attain. Therefore, the theoretical maximum updraft velocity of a parcel can be estimated as the square root of two times CAPE. CAPE can be estimated visually on a skewT-logP diagram as the area (from the LFC to the EL) between the parcel’s temperature trace and the environmental temperature trace. To compute CAPE, the following formula was used (“Definitions”):

\[ CAPE = g \int_{LFC}^{EL} \left( \frac{T_{vp} - T_{ve}}{T_{ve}} \right) dz \]  (32)
Above, \( g \) is the acceleration due to gravity, LFC is the level of free convection, EL is the equilibrium level, \( T_{vp} \) is the virtual temperature of the rising parcel, \( T_{ve} \) is the virtual temperature of the environment, and CAPE is the convective available potential energy (units of \( \text{m}^2 \text{s}^{-2} \) or \( \text{J kg}^{-1} \)). All integral formulas, including the above, given from this point forward were estimated as finite sums using \( dz = 1 \text{ meter} \).

Convective inhibition, or CIN, provides an estimate of the energy (per unit mass) that needs to be provided to an ascending parcel to lift it to its LFC; in other words, the energy needed to overcome the parcel’s negative buoyancy. Large (negative) CIN values represent a strongly “capped” environment. In this situation, strong vertical forcing is required before free convection can begin (if it begins at all). Temperature inversions are often responsible for a heavily capped environment. CIN can be estimated on a skewT-logP diagram as the negative area (from the surface to the LFC) between the parcel’s temperature trace and the environmental temperature trace. To compute CIN, the following formula was used (“Definitions”):

\[
CIN = g \int_{z_0}^{LFC} \left( \frac{T_{vp} - T_{ve}}{T_{ve}} \right) \, dz
\]

(33)

Above, the variables used are identical to those in the formula for CAPE, except that \( z_0 \) is the initial parcel height. The units of CIN are the same as those of CAPE.

The Storm Prediction Center has recently described new atmospheric layers that appear to offer improvements in predicting supercell thunderstorms and tornadoes. One of these layers is referred to as the “effective inflow layer.” The determination of
the lower and upper bounds of this layer is computationally intensive. Starting from the surface and moving upwards, every parcel’s CAPE and CIN values are computed. The base of the effective inflow layer is defined as the level closest to the surface at which a parcel has a CAPE value of greater than or equal to 100 J kg\(^{-1}\) and a CIN value of greater than -250 J kg\(^{-1}\). The top of the effective inflow layer is defined as the level closest to the layer’s base at which a parcel has a CAPE value of less than 100 J kg\(^{-1}\) or a CIN value of less than or equal to -250 J kg\(^{-1}\). These definitions are meant to confine a layer in which all parcels are both sufficiently buoyant and not strongly capped; thus, all inflowing parcels in this layer will contribute to the storm’s updraft (Thompson “Explanation”). In this analysis, parcel CAPE and CIN values were computed every 10 meters to limit computation time yet still provide accurate bounds for the inflow layer. If no effective inflow layer existed within a sounding, the base was set at zero meters AGL and the top was set at one meter AGL.

In the calculation of the bounds of the effective inflow layer, CAPE and CIN values are computed for every parcel (in this case, every parcel in 10-meter intervals) within the layer. During this process, values for two experimental parameters were additionally computed. These parameters will be referred to as “effective layer CAPE,” or eCAPE, and “effective layer CIN,” or eCIN. eCAPE is defined as the average CAPE value among parcels within the effective inflow layer, and eCIN is defined as the average CIN value among parcels in the layer. In essence, eCAPE represents the average kinetic energy per mass with which storm inflow will ascend in the updraft, and eCIN represents the average energy per mass that storm inflow will have to be
provided to contribute to the updraft via free convection. The following formulas can be used to define eCAPE and eCIN according to the description given above:

\[
eCAPE \equiv (z_t - z_b)^{-1} \int_{z_b}^{z_t} CAPE(z) \, dz
\]

\[
eCIN \equiv (z_t - z_b)^{-1} \int_{z_b}^{z_t} CIN(z) \, dz
\]

(34,35)

In these formulas, \( z_t \) represents the height of the top of the effective inflow layer, \( z_b \) represents the height of the base of the effective inflow layer, and \( CAPE(z)/CIN(z) \) represents the CAPE/CIN value for the parcel at height \( z \). The units of eCAPE/eCIN are the same as the units of CAPE/CIN.

In addition to the effective inflow layer, the Storm Prediction Center has also defined a layer called the “lower half of storm depth.” The lower bound of this layer is set as the base of the effective inflow layer. The upper bound of the layer is set as the height 50% of the way from the effective inflow layer base to the most unstable parcel’s equilibrium level (“Effective”). Essentially, the Storm Prediction Center is assuming that the base of the storm is located at the effective inflow layer base and that the top of the storm is located at the most unstable parcel’s equilibrium level. Therefore, this layer defines, approximately, the lower half of the storm’s depth. When bulk shear is calculated for this layer, the Storm Prediction Center refers to the value as the “effective bulk shear.” Recent studies suggest that the effective bulk shear is better at locating
environments conducive to supercell development than other fixed-layer shear values (Thompson “Explanation”).

Downdraft CAPE (DCAPE) is a thermodynamic parameter that attempts to predict the intensity of storm downdrafts. High DCAPE values are said to correspond to stronger downdrafts, though the effectiveness of DCAPE in predicting the strength of downdrafts has been questioned (Gilmore 1998). The questionable usefulness of DCAPE is likely due to the many drastic simplifications/assumptions used in its calculation. The downdraft parcel is assumed to originate at the “level of free sink,” or LFS. Various methods exist for determining this level (Rose, ”Various”). In this analysis, the height of the minimum equivalent potential temperature value in the lowest 400mb of the sounding is determined. This ensures that the most negatively buoyant parcel is selected. This height is the LFS. Finally, the saturated parcel is brought down to the surface along the moist adiabat, and DCAPE is computed via the following formula (Doswell 1997):

\[
DCAPE = -g \int_{SFC}^{LFS} \left( \frac{T_{vp} - T_{ve}}{T_{ve}} \right) dz
\]  

(36)

In this formula, SFC is the surface (ground level), LFS is the level of free sink, and the rest of the variables are the same as those used in the CAPE/CIN formulas. The units of DCAPE are J kg\(^{-1}\), and the sign of DCAPE is positive. Two significant assumptions made in this calculation are that the downdraft parcels always originate at the LFS and that the parcels are maintained at saturation via evaporation during their
descent to the surface. While these assumptions may severely limit the effectiveness of DCAPE in predicting the intensity of downdrafts, DCAPE was still calculated in this analysis in case it reveals an important thermodynamic feature of the sounding environment.

\textit{d2. Dynamic Parameters}

The dynamic parameters calculated in this analysis help to quantify important characteristics of the vertical distributions of wind speed and direction. One of the most computationally simple dynamic parameters is vertical wind shear. Vertical wind shear is the difference between winds at two levels in the atmosphere. The three most common atmospheric layers across which wind shear is calculated are 0-1 kilometers, 0-3 kilometers, and 0-6 kilometers AGL. More recently, the “lower half of storm depth” layer, described on page 31, has been used, and the shear across this layer is called the “effective bulk shear.” Additionally, in this analysis, shear was calculated across the effective inflow layer. The following formula is used to calculate vertical wind shear (Doswell 1997):

\[
SHEAR = \left| \vec{V}_{upper} - \vec{V}_{lower} \right|
\]  

(37)

Above, \( V_{upper} \) is the upper-level wind vector and \( V_{lower} \) is the lower-level wind vector. The shear is simply the magnitude of the vector difference between the two winds.

In addition to the five shear values mentioned above (0-1km, 0-3km, 0-6km, effective bulk, and effective inflow layer), a parameter referred to as the “Bulk
Richardson Number shear term” was calculated. This parameter is the denominator of another parameter called the Bulk Richardson Number, or BRN. From here forward, the Bulk Richardson Number shear term will be referred to as the BRN shear. To compute BRN shear, pressure-weighted mean winds must be calculated. The following formula is used to calculate a pressure-weighted vertical average (Holton 2013):

\[
\langle \rangle = \left( p_{upper} - p_{lower} \right)^{-1} \int_{p_{lower}}^{p_{upper}} (\ )dp
\]

(38)

Above, the angled brackets represent the pressure-weighted vertical average, \( p_{upper} \) is the pressure at the top of the layer, \( p_{lower} \) is the pressure at the bottom of the layer, and the parentheses represent the variable being averaged.

To compute a pressure-weighted mean wind, wind speed, \( u \) components, and \( v \) components were all averaged through the layer with pressure weighting. The final pressure-weighted mean wind through the layer was given the pressure-weighted speed. The mean wind’s direction was determined by the direction of the pressure-weighted \( u \) and \( v \) components. Using this method of calculating pressure-weighted mean winds, the following formula was used to compute the BRN shear (Doswell 1997):

\[
SHEAR_{BRN} = 0.5 \left| \langle \vec{V} \rangle_{0-6\text{km}} - \langle \vec{V} \rangle_{0-500\text{m}} \right|^2
\]

(39)

Above, the angled brackets denote a pressure-weighted vertical average, and the two layers considered are 0-6km and 0-500m AGL. Thus, the BRN shear is a form of
deep-layer shear. The shear magnitude is squared and multiplied by one half to give it units of energy per mass so that it can be compared with CAPE in the Bulk Richardson Number.

Storm-relative helicity, or SRH, is one of the best indicators of tornado potential. According to the Storm Prediction Center, storm-relative helicity is “a measure of the potential for cyclonic updraft rotation in right-moving supercells” (“Storm Relative”). It essentially represents both the degree of rotation of winds with increasing height and the strength of these winds. Clockwise rotation with increasing height gives a positive SRH value. Large, positive SRH values correspond to large, round hodograph plots and point to an increased risk for tornadoes. SRH is typically calculated across three atmospheric layers: 0-1km AGL, 0-3km AGL, and the effective inflow layer. Therefore, in this analysis, SRH values across these three layers were computed. The following formula is used to calculate storm-relative helicity (“Calculation”):

\[
SRH = -\int_{z_b}^{z_t} \hat{k} \cdot (\bar{V} - \bar{c}) \times \frac{d\bar{V}}{dz} \, dz
\]  \hspace{1cm} (40)

Above, \( z_b \) is the height of the bottom of the layer, \( z_t \) is the height of the top of the layer, \( \hat{k} \) is the vertical unit vector \([0,0,1]\), \( \bar{V} \) is the environmental wind vector, and \( \bar{c} \) is the storm motion vector.

From the equation for storm-relative helicity, it is evident that a storm motion vector must be estimated. Prediction of storm motion is a complex task and depends on the vertical wind field. The most accurate method of predicting storm motion appears
to be the “internal dynamics method,” or the ID method. Formulas for estimating storm motion for both right-moving and left-moving supercells were developed by Bunkers (2000). Since the majority of supercells are right-moving, and since the Storm Prediction Center defines storm-relative helicity in regard to right-moving supercells, the formula for right-moving storms was used in this analysis. Following is the formula for right-moving supercells (Bunkers 2000):

$$\bar{V}_{rm} = \bar{V}_{\text{mean}} + D \left[ \frac{\bar{V}_{\text{shear}} \times \hat{k}}{|\bar{V}_{\text{shear}}|} \right]$$

(41)

Above, $V_{rm}$ is the velocity vector for a right-moving supercell, $V_{\text{mean}}$ is a 0-6km non-pressure-weighted mean wind, $D$ is a constant (7.5 m s$^{-1}$), $k$ is the vertical unit vector [0,0,1], and $V_{\text{shear}}$ is the shear vector between a 5.5-6.0km non-pressure-weighted mean wind and a 0.0-0.5km non-pressure-weighted mean wind. The constant, $D$, was set at 7.5 m s$^{-1}$ to minimize storm motion error. In this analysis, the procedure for calculating a vertical non-pressure-weighted mean wind is similar to that described previously for calculating a pressure-weighted mean wind. The mean wind’s speed is the arithmetic mean of the observed wind speeds through the layer, and the mean wind’s direction is that given by the arithmetic means of the $u$ and $v$ wind components of the observed winds through the layer.
3. Results

a. Direct Results

After computing the thermodynamic and dynamic sounding parameters (35 in total) for each unique sounding and tallying each sounding’s number of proximal tornadoes, number of proximal strong tornadoes, and tornado parameter (as defined previously) under proximity constraints of 200km and 2100-0300 UTC, correlations were analyzed. In total, 3085 soundings were considered. Correlations were calculated between each sounding parameter and each of the three “tornado values” mentioned above; these correlations are shown in Table 2.

Table 2 - Sounding Parameter Correlations with Number of Tornadoes, Number of Strong Tornadoes, and Tornado Parameter (R=200km, 2100-0300Z)

<table>
<thead>
<tr>
<th>Sounding Parameter</th>
<th>Num. Tor.</th>
<th>Num. Str.</th>
<th>Tor. Par.</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sb LCL</td>
<td>-0.03</td>
<td>-0.04</td>
<td>-0.05</td>
<td>-0.04</td>
</tr>
<tr>
<td>Sb LFC</td>
<td>-0.01</td>
<td>0.00</td>
<td>-0.01</td>
<td>-0.01</td>
</tr>
<tr>
<td>Sb EL</td>
<td>0.05</td>
<td>0.03</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>Sb CAPE</td>
<td>0.10</td>
<td>0.06</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>Sb CIN</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Fml LCL</td>
<td>-0.03</td>
<td>-0.05</td>
<td>-0.05</td>
<td>-0.04</td>
</tr>
<tr>
<td>Fml LFC</td>
<td>-0.05</td>
<td>-0.04</td>
<td>-0.05</td>
<td>-0.05</td>
</tr>
<tr>
<td>Fml EL</td>
<td>0.09</td>
<td>0.06</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td>Fml CAPE</td>
<td>0.14</td>
<td>0.09</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>Fml CIN</td>
<td>0.03</td>
<td>0.02</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>Vml LCL</td>
<td>-0.02</td>
<td>-0.04</td>
<td>-0.04</td>
<td>-0.03</td>
</tr>
<tr>
<td>Vml LFC</td>
<td>-0.04</td>
<td>-0.03</td>
<td>-0.04</td>
<td>-0.04</td>
</tr>
<tr>
<td>Vml EL</td>
<td>0.08</td>
<td>0.06</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td>Vml CAPE</td>
<td>0.13</td>
<td>0.09</td>
<td>0.12</td>
<td>0.11</td>
</tr>
<tr>
<td>Vml CIN</td>
<td>0.03</td>
<td>0.02</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>Mu LCL</td>
<td>-0.05</td>
<td>-0.04</td>
<td>-0.05</td>
<td>-0.05</td>
</tr>
<tr>
<td>Mu LFC</td>
<td>-0.03</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.03</td>
</tr>
<tr>
<td>Mu EL</td>
<td>0.05</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>Mu CAPE</td>
<td>0.10</td>
<td>0.07</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td>Mu CIN</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>0-1km SRH</td>
<td>0.15</td>
<td>0.13</td>
<td>0.15</td>
<td>0.14</td>
</tr>
<tr>
<td>0-3km SRH</td>
<td>0.15</td>
<td>0.11</td>
<td>0.14</td>
<td>0.13</td>
</tr>
<tr>
<td>Eff. SRH</td>
<td>0.16</td>
<td>0.12</td>
<td>0.16</td>
<td>0.14</td>
</tr>
<tr>
<td>0-1km Shear</td>
<td>0.20</td>
<td>0.18</td>
<td>0.21</td>
<td>0.20</td>
</tr>
<tr>
<td>0-3km Shear</td>
<td>0.23</td>
<td>0.20</td>
<td>0.24</td>
<td>0.22</td>
</tr>
<tr>
<td>0-6km Shear</td>
<td>0.22</td>
<td>0.19</td>
<td>0.22</td>
<td>0.21</td>
</tr>
<tr>
<td>Eff. Inf. Shear</td>
<td>0.22</td>
<td>0.15</td>
<td>0.21</td>
<td>0.19</td>
</tr>
<tr>
<td>Eff. Bulk Shear</td>
<td>0.28</td>
<td>0.23</td>
<td>0.27</td>
<td>0.26</td>
</tr>
<tr>
<td>BRN Shear</td>
<td>0.15</td>
<td>0.11</td>
<td>0.14</td>
<td>0.14</td>
</tr>
<tr>
<td>Storm Motion</td>
<td>0.20</td>
<td>0.18</td>
<td>0.21</td>
<td>0.19</td>
</tr>
<tr>
<td>Eff. Base</td>
<td>-0.04</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.03</td>
</tr>
<tr>
<td>Eff. Top</td>
<td>0.09</td>
<td>0.05</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>Eff. CAPE</td>
<td>0.13</td>
<td>0.09</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td>Eff. CIN</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>DCAPE</td>
<td>0.08</td>
<td>0.06</td>
<td>0.07</td>
<td>0.07</td>
</tr>
</tbody>
</table>
The correlation analysis revealed a number of expected trends. First of all, it is worth noting that for a sample size of 3085, the critical value for the correlation coefficient ($\alpha=0.05$, non-directional) is only about 0.04; therefore, the majority of the correlation coefficients in Table 2 are statistically significant. All LCLs and LFCs were negatively correlated with increasing tornado parameter, and ELs and CAPEs were positively correlated. This supports the well-established idea that low cloud bases and atmospheric instability promote tornadogenesis. CIN values appear to be nearly uncorrelated with tornado formation; given that a subset of only tornadic soundings is being considered here, this is not an unexpected result. Provided that a certain threshold of capping is not exceeded, CIN likely has little effect on tornadogenesis. Albeit by insignificant margins, the most-unstable parcel’s LCL and the fixed mixed-layer parcel’s CAPE appear to be the most strongly correlated. Also worth noting is that eCAPE is nearly as strongly correlated as mlCAPE in this sample. The negative correlation with effective inflow layer base height and positive correlation with effective inflow layer top height suggest that a broader effective inflow layer is more supportive of tornado formation. DCAPE’s positive correlation may be due to a variety of factors. One possibility is that higher DCAPE values are associated with steep mid-level lapse rates, which are associated with higher CAPE values.

In terms of the dynamic sounding parameters, correlations are stronger than for the thermodynamic parameters. All types of storm-relative helicity are nearly equally correlated, and all forms of shear are strongly correlated. Among the shear variables, effective bulk shear was the most strongly correlated with tornado formation. Storm
motion’s strong positive correlation is likely due to the fact that fast storm motion is reflective of strong winds through the storm’s depth. Strong winds through the 0-6km layer are correlated with larger shear values and larger hodograph profiles, corresponding to higher helicity values. Thus, storm motion’s correlation with tornado formation is likely due to its correlation with shear and helicity, though this correlation may be worth further study.

Following is a plot of 0-1km AGL storm-relative helicity versus tornado parameter:

![0-1km Storm-Relative Helicity vs. Tornado Parameter](image)

*Figure 1 - 0-1km Storm-Relative Helicity vs. Tornado Parameter (R=200km, 2100-0300Z)*

*Data from all 3085 soundings are plotted*
Although Table 2 shows a relatively strong positive correlation between 0-1km SRH and tornado parameter, Figure 1 does not clearly depict such a relationship. Plots of every other sounding variable versus tornado parameter were equally cluttered. To de-clutter the plots and reduce the impact of non-representative sounding data, an averaging procedure was used. For each unique value of the tornado parameter, averages for each sounding parameter among all soundings possessing that tornado parameter value were computed. However, the average was only computed if at least three soundings possessed that tornado parameter value; this reduces the chances of plotting data from soundings that were not reflective of the tornadic environment. 45 of the 3085 total soundings were not considered due to this procedure. For example, say that 4 soundings have a tornado parameter of 15. If these soundings had respective values of mCAPE of 1500, 2000, 2500, and 2000 J kg$^{-1}$, only the point (15, 2000) was plotted. After calculating all of these averages, the following correlations were obtained:
Table 3 shows the same trends as Table 2, except the correlation coefficients obtained after the averaging procedure outlined above are notably greater in magnitude. For the new, averaged sample size of 33, the critical value for the correlation coefficient ($\alpha=0.05$, non-directional) is about 0.35; therefore, again, most of the correlations in Table 3 are significant. Worth noting is that the most-unstable parcel’s LCL is now more significantly correlated with tornado parameter than any other parcel’s LCL.
Additionally, eCAPE now appears to be better correlated than any single parcel’s CAPE. In terms of the dynamic parameters, storm-relative helicity now appears to be more strongly correlated than the shear parameters. Again, however, there is no significant difference between the various layers across which SRH is calculated. The most strongly correlated shear parameters are seen to be effective inflow layer shear and effective bulk shear. Storm motion’s positive correlation remains, although it is less significant than it was before the averaging procedure.

Below is a plot of 0-1km storm-relative helicity versus tornado parameter, after the averaging procedure:

![0-1km Storm-Relative Helicity vs. Tornado Parameter (R=200km, 2100-0300Z, Averaged Values)](attachment:figure2.png)

*Figure 2 - 0-1km Storm-Relative Helicity vs. Tornado Parameter (R=200km, 2100-0300Z, Averaged Values)*
As seen in Figure 2, the averaging procedure significantly de-cluttered the plots and allowed for a much easier visual identification of trends. Plots of all of the averaged sounding parameters will not be shown, as the data in Table 3 provide information about the trends that exist between all of the parameters and the tornado parameter.

Though nearly all of the correlations in Tables 2 and 3 reveal expected trends, the distance and time proximity constraints were altered to see whether different constraints may affect correlation strength. Distance constraints of 25km, 50km, 100km, 150km, and 200km were paired with time constraints of 2100-0300Z, 2200-0200Z, 2300-0100Z, 0000-0300Z, 0000-0200Z, and 0000-0100Z to see if correlations were significantly altered. The three time constraints beginning at 0000Z were considered in attempt to reduce the likelihood that tornadoes/thunderstorms occurring before radiosonde release stabilized the thermodynamic environment or altered the vertical wind field. Under each pair of distance and time constraints, the average correlation magnitude among all sounding parameters and the tornado parameter was calculated. These data are given in Table 4.

**Table 4 - Average Correlation Magnitudes under Various Distance/Time Proximity Constraints**

<table>
<thead>
<tr>
<th>Constraints</th>
<th>2100-0300Z</th>
<th>2200-0200Z</th>
<th>2300-0100Z</th>
<th>0000-0300Z</th>
<th>0000-0200Z</th>
<th>0000-0100Z</th>
<th>Averages</th>
</tr>
</thead>
<tbody>
<tr>
<td>25km</td>
<td>0.105</td>
<td>0.127</td>
<td>0.161</td>
<td>0.141</td>
<td>0.154</td>
<td>0.184</td>
<td>0.145</td>
</tr>
<tr>
<td>50km</td>
<td>0.096</td>
<td>0.104</td>
<td>0.1</td>
<td>0.12</td>
<td>0.134</td>
<td>0.135</td>
<td>0.115</td>
</tr>
<tr>
<td>100km</td>
<td>0.103</td>
<td>0.107</td>
<td>0.104</td>
<td>0.104</td>
<td>0.113</td>
<td>0.115</td>
<td>0.108</td>
</tr>
<tr>
<td>150km</td>
<td>0.098</td>
<td>0.104</td>
<td>0.102</td>
<td>0.12</td>
<td>0.13</td>
<td>0.131</td>
<td>0.114</td>
</tr>
<tr>
<td>200km</td>
<td>0.094</td>
<td>0.099</td>
<td>0.098</td>
<td>0.107</td>
<td>0.113</td>
<td>0.105</td>
<td>0.103</td>
</tr>
<tr>
<td>Averages</td>
<td>0.099</td>
<td>0.108</td>
<td>0.113</td>
<td>0.118</td>
<td>0.129</td>
<td>0.134</td>
<td>0.117</td>
</tr>
</tbody>
</table>
For reference, Table 5 displays the number of soundings in consideration under each set of distance and time proximity constraints.

*Table 5 - Number of Soundings Analyzed under Various Distance/Time Proximity Constraints*

<table>
<thead>
<tr>
<th></th>
<th>2100-0300Z</th>
<th>2200-0200Z</th>
<th>2300-0100Z</th>
<th>0000-0300Z</th>
<th>0000-0200Z</th>
<th>0000-0100Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>25km</td>
<td>176</td>
<td>131</td>
<td>74</td>
<td>77</td>
<td>66</td>
<td>41</td>
</tr>
<tr>
<td>50km</td>
<td>468</td>
<td>361</td>
<td>214</td>
<td>210</td>
<td>171</td>
<td>115</td>
</tr>
<tr>
<td>100km</td>
<td>1145</td>
<td>897</td>
<td>546</td>
<td>517</td>
<td>444</td>
<td>298</td>
</tr>
<tr>
<td>150km</td>
<td>1998</td>
<td>1598</td>
<td>1011</td>
<td>928</td>
<td>811</td>
<td>557</td>
</tr>
<tr>
<td>200km</td>
<td>3085</td>
<td>2479</td>
<td>1565</td>
<td>1500</td>
<td>1316</td>
<td>891</td>
</tr>
</tbody>
</table>

As Table 4 shows, the trend is for correlations to increase with more restrictive distance and time proximity constraints. However, Table 5 shows that the sample size decreases significantly under more restrictive constraints. Examination of correlation tables similar to Tables 2 and 3 reveals trends nearly identical to those observed under a distance constraint of 200km and a time constraint of 2100-0300Z. For example, the table below shows correlations under constraints of 50km and 2300-0100Z:
As can be seen when comparing Table 6 to Table 3, the same trends persist. These trends were also seen when analyzing correlations under different constraints. Making the distance and time constraints more restrictive served only to reveal the same trends and significantly decrease sample size. Therefore, only the original constraints of 200km and 2100-0300Z are considered from this point forward.
To examine these data in an alternate way, the 3085 soundings under consideration were divided into those that were taken within proximity of one or more strong ([E]F-rating ≥ 2) tornado and those that were not taken near a strong tornado. Then, averages were taken among the “strong” and “weak” tornado soundings for every sounding parameter and the sample means were compared using a Student’s T-test. 439 soundings were taken within proximity to one or more strong tornado and the remaining 2646 soundings were not taken near a strong tornado. The following table summarizes the results:

*Table 7 - Comparison of sounding parameter sample means between strong and weak tornado soundings. The far right column contains the (two-tailed) T-test probability that the population means are identical (R=200km, 2100-0300Z)*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Strong Mean</th>
<th>Weak Mean</th>
<th>P (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sb LCL</td>
<td>846</td>
<td>910</td>
<td>11.70</td>
</tr>
<tr>
<td>Sb LFC</td>
<td>1981</td>
<td>1864</td>
<td>9.12</td>
</tr>
<tr>
<td>Sb EL</td>
<td>9407</td>
<td>9231</td>
<td>43.31</td>
</tr>
<tr>
<td>Sb CAPE</td>
<td>1744</td>
<td>1446</td>
<td>0.05</td>
</tr>
<tr>
<td>Sb CIN</td>
<td>-59</td>
<td>-55</td>
<td>44.34</td>
</tr>
<tr>
<td>Fml LCL</td>
<td>1240</td>
<td>1305</td>
<td>10.03</td>
</tr>
<tr>
<td>Fml LFC</td>
<td>2150</td>
<td>2297</td>
<td>1.01</td>
</tr>
<tr>
<td>Fml EL</td>
<td>9044</td>
<td>8616</td>
<td>4.95</td>
</tr>
<tr>
<td>Fml CAPE</td>
<td>1373</td>
<td>1013</td>
<td>2.8E-05</td>
</tr>
<tr>
<td>Fml CIN</td>
<td>-47</td>
<td>-52</td>
<td>27.58</td>
</tr>
<tr>
<td>Vml LCL</td>
<td>1197</td>
<td>1242</td>
<td>27.91</td>
</tr>
<tr>
<td>Vml LFC</td>
<td>2144</td>
<td>2269</td>
<td>3.62</td>
</tr>
<tr>
<td>Vml EL</td>
<td>8976</td>
<td>8524</td>
<td>4.10</td>
</tr>
<tr>
<td>Vml CAPE</td>
<td>1360</td>
<td>1017</td>
<td>9.3E-05</td>
</tr>
<tr>
<td>Vml CIN</td>
<td>-51</td>
<td>-55</td>
<td>27.82</td>
</tr>
<tr>
<td>Mu LCL</td>
<td>1129</td>
<td>1199</td>
<td>9.62</td>
</tr>
<tr>
<td>Mu LFC</td>
<td>1859</td>
<td>1836</td>
<td>72.38</td>
</tr>
<tr>
<td>Mu EL</td>
<td>10593</td>
<td>10416</td>
<td>31.36</td>
</tr>
<tr>
<td>Mu CAPE</td>
<td>2111</td>
<td>1769</td>
<td>0.02</td>
</tr>
<tr>
<td>Mu CIN</td>
<td>-26</td>
<td>-26</td>
<td>98.43</td>
</tr>
<tr>
<td>0-1km SRH</td>
<td>165</td>
<td>92</td>
<td>6.6E-15</td>
</tr>
<tr>
<td>0-3km SRH</td>
<td>254</td>
<td>173</td>
<td>6.1E-15</td>
</tr>
<tr>
<td>Eff. SRH</td>
<td>148</td>
<td>81</td>
<td>1.7E-14</td>
</tr>
<tr>
<td>0-1km Shear</td>
<td>11</td>
<td>8</td>
<td>1.1E-23</td>
</tr>
<tr>
<td>0-3km Shear</td>
<td>18</td>
<td>14</td>
<td>2.3E-32</td>
</tr>
<tr>
<td>0-6km Shear</td>
<td>25</td>
<td>20</td>
<td>6.4E-30</td>
</tr>
<tr>
<td>Eff. Inf. Shear</td>
<td>10</td>
<td>6</td>
<td>2.5E-12</td>
</tr>
<tr>
<td>Eff. Bulk Shear</td>
<td>23</td>
<td>17</td>
<td>1.6E-32</td>
</tr>
<tr>
<td>BRN Shear</td>
<td>118</td>
<td>83</td>
<td>5.2E-15</td>
</tr>
<tr>
<td>Storm Motion</td>
<td>18</td>
<td>13</td>
<td>7.3E-26</td>
</tr>
<tr>
<td>Eff. Base</td>
<td>103</td>
<td>105</td>
<td>90.65</td>
</tr>
<tr>
<td>Eff. Top</td>
<td>1300</td>
<td>1171</td>
<td>0.94</td>
</tr>
<tr>
<td>Eff. CAPE</td>
<td>1217</td>
<td>924</td>
<td>1.2E-05</td>
</tr>
<tr>
<td>Eff. CIN</td>
<td>-41</td>
<td>-39</td>
<td>35.28</td>
</tr>
<tr>
<td>DCAPE</td>
<td>1250</td>
<td>1060</td>
<td>5.8E-05</td>
</tr>
</tbody>
</table>
Table 7 shows that all of the dynamic parameters (storm-relative helicity, shear, and storm motion) discriminate very strongly between strong and weak tornadoes. It is worth noting that P-values with large, negative exponents (such as the P-value for 0-3km shear), while still listed in Table 7, all occur at the extreme tail of the T-distribution and should be interpreted as equaling zero. The probability is near zero that the population means for any of the dynamic parameters are identical for soundings that produce strong tornadoes and soundings that do not produce strong tornadoes. Among the various forms of CAPE, these data suggest that eCAPE, by a very narrow margin, is the best discriminator between strong and weak tornado soundings. DCAPE also shows a statistically significant difference. Also, although the most-unstable parcel’s LCL appears to discriminate better than other parcels’ LCLs, it is worth noting that this result is borderline significant (P = 9.62%).

When the data in Table 7 were plotted, the most striking differences between strong and weak tornado-producing soundings were seen when comparing the various storm-relative helicity values:
Figure 3 - Average storm-relative helicity values for strong versus weak tornado-producing soundings (R=200km, 2100-0300Z)

b. Interpretive Results

The results of this analysis thus far have shown expected trends: lower cloud bases, increased thermodynamic instability, and increased shear and helicity values have all been shown to correlate with tornado formation. Of particular interest, though, is that eCAPE appears to be better correlated with tornadogenesis than any single parcel’s CAPE value. However, the above results also suggest that the formation of tornadoes depends on many factors other than just instability; thus, the use of composite parameters may be the most effective method of predicting tornadoes.
One of the most commonly used composite parameters by the Storm Prediction Center is the Significant Tornado Parameter, or SigTor. This parameter was not developed using correlation analysis, but rather by combining the parameters that were shown in various studies to be the best discriminators between strong, weak, and no tornadoes. Two variations of this parameter are still in use: one includes fixed-layer sounding parameters, and the other includes effective layer parameters. The “effective” formula for SigTor is given below (Thompson 2005):

\[
\text{SigTor}_{\text{eff}} = \frac{\text{mlCAPE} \times \text{EBS} \times \text{eSRH} \times \frac{2000 - \text{mlLCL}}{1000} \times 200 + \text{mlCIN}}{1500 \times 20 \times 150}
\]  

(42)

In this formula, mlCAPE is the 100mb mixed-layer CAPE in J kg\(^{-1}\), EBS is the effective bulk shear in m s\(^{-1}\), eSRH is the effective inflow layer storm-relative helicity in m\(^2\)s\(^{-2}\), mlLCL is the 100mb mixed-layer LCL in m, and mlCIN is the 100mb mixed-layer CIN in J kg\(^{-1}\).

To see whether this composite parameter could be improved, mlCAPE/CIN was replaced in the above formula for SigTor with eCAPE/CIN. Then, for each of the 3085 soundings considered in this analysis, values for the Storm Prediction Center’s “effective” SigTor were calculated along with values for the “modified” SigTor described above. Figure 4 is a plot of the SPC’s SigTor versus tornado parameter.
Figure 4 - Storm Prediction Center’s “Effective” SigTor vs. Tornado Parameter (R=200km, 2100-0300Z, Averaged Values)

For comparison, Figure 5 is a plot of the modified SigTor (with eCAPE/CIN instead of mlCAPE/CIN) versus tornado parameter:
As the above plots show, both versions of SigTor are strongly positively correlated with the tornado parameter. However, the correlation values are very similar; $R^2$ for the SPC’s version is 0.592, and $R^2$ for the modified version is 0.587. Thus, the substitution of mCAPE/CIN for eCAPE/CIN appears to make little difference in SigTor’s effectiveness.

To see whether the modified SigTor may be better at discriminating between strong and weak tornado soundings, a Student’s T-test was conducted for both versions of SigTor in a manner similar to that used to produce Table 7. Following are the results:
Table 8 - Comparison of sample means for the two versions of SigTor between strong and weak tornado soundings. The far right column contains the (two-tailed) T-test probability that the population means are identical (R=200km, 2100-0300Z)

<table>
<thead>
<tr>
<th></th>
<th>Strong Mean</th>
<th>Weak Mean</th>
<th>Probability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPC SigTor</td>
<td>1.18</td>
<td>0.38</td>
<td>8.1E-13</td>
</tr>
<tr>
<td>Mod. SigTor</td>
<td>0.92</td>
<td>0.29</td>
<td>2.7E-13</td>
</tr>
</tbody>
</table>

The above table shows that, once again, there is little difference between the SPC’s SigTor and the modified SigTor. The slightly lower T-test probability value for the modified SigTor may seem to suggest that replacing mlCAPE/CIN with eCAPE/CIN in the Significant Tornado Parameter slightly increases the parameter’s ability to discriminate between soundings that produce strong tornadoes and those that produce only weak ones; however, given that both P-values are extremely close to 0, the only conclusion that can be drawn from Table 8 is that both forms of SigTor discriminate strongly between strong and weak tornadoes.

Rather than simply swapping one parameter for another in an existing composite parameter, multivariate linear regression analysis was used to arrive at a new composite parameter. However, multivariate linear regression does not produce an equation of SigTor’s form; rather, it produces an equation of the following form, where N is the number of independent variables used in the regression, $x_i$ is the i’th independent variable, $a_i$ is $x_i$’s coefficient, b is a constant, and $y$ is the approximated value of the dependent variable:

$$y = b + \sum_{i=1}^{N} a_i x_i$$  \hspace{1cm} (43)
In order to produce a composite parameter of this form and have additional, independent data to verify the parameter, the soundings were divided into two groups: those taken in 2006 and prior (1438 soundings) and those taken after 2006 (1647 soundings). The 1438 soundings taken in 2006 and prior were used to develop the composite parameter. Multivariate regression was originally conducted using all 35 sounding variables; then, the least statistically significant coefficients and their associated independent variables were removed from consideration and the regression was conducted again. This procedure was repeated until all coefficients produced in the regression were statistically significant (P<0.05). The resulting composite parameter will be referred to as the tornadogenesis parameter (TGP). The TGP equation produced by the regression analysis is as follows:

\[
TGP = \frac{174}{10^3}(EBS) + \frac{104}{10^3}(ES) + \frac{315}{10^5}(SRH_{0-1}) - \frac{299}{10^6}(muEL) \\
+ \frac{816}{10^6}(mlCAPE) + \frac{389}{10^6}(sbLFC) - \frac{391}{10^6}(sbLCL) + 2.2245
\]  

(44)

In this equation, EBS is the effective bulk shear in m s\(^{-1}\), ES is the effective inflow layer shear in m s\(^{-1}\), SRH\(_{0-1}\) is the 0-1km storm-relative helicity in m\(^2\) s\(^{-2}\), muEL is the most-unstable parcel’s equilibrium level in m, mlCAPE is the 100mb mixed-layer CAPE in J kg\(^{-1}\), sbLFC is the surface-based LFC in m, and sbLCL is the surface-based LCL in m.

When TGP was plotted vs. tornado parameter for only those soundings taken in 2006 and earlier (the soundings used in the regression analysis), the R\(^2\) value was 0.428.
For comparison, when the SPC’s SigTor was plotted versus tornado parameter for this subset of data, the $R^2$ value obtained was only 0.341.

Next, the TGP had to be verified with the independent soundings taken after 2006. Following is a plot of TGP versus tornado parameter for the soundings taken after 2006:

![TGP vs. Tornado Parameter](image)

*Figure 6 - Tornadogenesis Parameter vs. Tornado Parameter (R=200km, 2100-0300Z, Averaged Values, Post-2006)*

For comparison, following is a plot of the SPC’s SigTor versus tornado parameter during the post-2006 period:
Figure 7 - Storm Prediction Center’s “Effective” SigTor vs. Tornado Parameter (R=200km, 2100-0300Z, Post-2006)

As Figures 6 and 7 show, the TGP shows extremely strong correlations with tornado parameter during the post-2006 period; in fact, the correlation between TGP and tornado parameter is substantially stronger during the post-2006 period than it is during the pre-2006 period whose data were used to develop it (R² is 0.846 post-2006 and 0.428 pre-2006). However, it must be noted that these correlations are the results of regression of the binned data rather than the raw data. When examining Figures 6 and 7, caution must be used, because the R² values (and the data) displayed are products of the binning method used in this study. The astonishingly large jump in R² for TGP between the pre-2006 and post-2006 periods, therefore, may simply be a product of the
averaging method employed. Alternatively (or additionally), the increase in correlation coefficients between the two time periods for both TGP and SigTor may be due in part to better reporting of tornadoes during the post-2006 period. More frequent and accurate tornado reports ensure that the “tornado parameter” calculated for each sounding actually accounts for all of the tornadoes that occurred within proximity to the sounding. Also, it is possible that a portion of this significant increase in correlation coefficient is due to improvements in radiosonde technology; this would allow for more accurate measurements of the sounding environment, which could increase correlations.

Nonetheless, within the limitations of the binning framework used, Figures 6 and 7 show that TGP has stronger correlation with tornado parameter than does SigTor; further study would be needed to determine whether this is due in large part to the use of the averaging procedure.

In addition to Figures 6 and 7, which show averaged TGP and SigTor values across various tornado parameter values (for the post-2006 period), Figures 8 and 9, below, show all TGP/SigTor values versus tornado parameter for the same time period.
Figure 8 - Storm Prediction Center’s “Effective” SigTor vs. Tornado Parameter (R=200km, 2100-0300Z, Post-2006)
Figures 8 and 9 show that, when considering all values for the predictors as opposed to solely the averaged values, the difference in correlation between TGP and SigTor is significantly smaller ($R^2$ for SigTor is 0.126, while $R^2$ for TGP is 0.112). For comparison, $R^2$ for TGP for the 2006-and-earlier period (when considering all TGP values) is 0.094. Therefore, the correlation improvement seen previously when comparing the 2006-and-earlier data to the post-2006 data is still present, albeit much less pronounced, when considering all TGP values. Figures 8 and 9 suggest that TGP is not a better predictor of tornadogenesis than SigTor; however, there are a couple of possible reasons for this. First of all, a major weakness of correlation analysis is that $R^2$
can be heavily swayed by the presence of outliers (it is not robust to them). Figures 8 and 9 contain outliers that severely weaken their utility; the averaging procedure used to produce Figures 6 and 7 removes these outliers. Additionally, the calculation of SigTor involves several “value checks,” including setting the entire parameter to 0 if the effective inflow layer base is above ground level and limiting the contributions of certain terms. Such value checks were not used in the calculation of TGP. Adding these types of checks to the TGP formulation (which could be done in future studies) would likely serve to improve TGP’s $R^2$ value by reducing variance. The averaging procedure used for Figures 6 and 7, though, likely makes $R^2$ more robust to the absence of value checks due to the variance reduction it provides. So, although the $R^2$ values given in Figures 6 and 7 depend heavily on the averaging procedure used to produce the plots, they are less influenced by outliers and by the presence/absence of value checks. Therefore, in assessing the potential utility of TGP in tornado forecasting, Figures 6 and 7 are likely of greater use than Figures 8 and 9.

Finally, the Student’s T-test was used to see whether strong and weak tornado soundings have statistically different mean values of TGP. To compare TGP to SigTor in terms of its ability to discriminate between strong and weak tornadoes, a row for TGP was simply added to Table 8:

<table>
<thead>
<tr>
<th></th>
<th>Strong Mean</th>
<th>Weak Mean</th>
<th>Probability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPC SigTor</td>
<td>1.18</td>
<td>0.38</td>
<td>8.1E-13</td>
</tr>
<tr>
<td>Mod. SigTor</td>
<td>0.92</td>
<td>0.29</td>
<td>2.7E-13</td>
</tr>
<tr>
<td>TGP</td>
<td>6.10</td>
<td>4.19</td>
<td>2.7E-43</td>
</tr>
</tbody>
</table>

*Table 9 - Comparison of sample means for the two versions of SigTor and TGP between strong and weak tornado soundings. The far right column contains the (two-tailed) T-test probability that the population means are identical (R=200km, 2100-0300Z)*
Table 9 shows that the P-value for TGP is about 30 orders of magnitude smaller than the P-value for SigTor; while this may sound significant, the P-values are all so close to zero that it is reasonable only to conclude that all three composite parameters discriminate very strongly between strong and weak tornadoes.

4. Discussion

a. Verification of Previous Studies

Low LCL heights were found in this analysis to favor tornado development. All four parcels’ LCLs were found to be negatively correlated with increased number/intensity of tornadoes. However, the T-test results suggested that LCLs are not significantly different between soundings that produced one or more strong tornadoes and soundings that produced only weak tornadoes (at the $\alpha=0.05$ level). Regardless, the results of previous studies were confirmed in that lower cloud base height tends to favor tornado development.

As previous studies have indicated, greater instability seems to be associated with a greater threat for tornadoes. All forms of CAPE were positively correlated with tornado parameter, and all forms also distinguished (significantly) between strong and weak tornadoes. The CAPE value that the Storm Prediction Center uses in its SigTor parameter, mixed-layer CAPE, was both most strongly correlated with tornadoes and the best discriminator between strong and weak tornadoes (excluding eCAPE).

All forms of storm-relative helicity (0-1km, 0-3km, and effective) and all forms of bulk shear (0-1km, 0-3km, 0-6km, effective inflow layer, and effective bulk) were both very strongly correlated with tornadoes (more so than CAPE/LCL) and very good at
discriminating (better than CAPE/LCL) between strong and weak tornadoes. Among
types of shear, low-level bulk shear appeared to have the best correlations; this result is
supported by past research.

Lastly, the Storm Prediction Center’s recently developed “effective” SigTor parameter appears to have very strong correlations with increased number/intensity of tornadoes, and it is very good at discriminating between strongly and weakly tornadic environments.

b. New Findings

Although previous studies have shown that LCL heights are an important factor to consider when forecasting the likelihood of tornadoes, there has been little research into which parcel’s LCL is the most important to consider. In the SigTor composite parameter, the 100mb mixed-layer parcel’s LCL is used. However, the results of this study suggest that not only is the most-unstable parcel’s LCL more strongly negatively correlated with tornadoes, but it is also a better discriminator between strong and weak tornadoes. In fact, muLCL is the only LCL that can statistically significantly discriminate between strong and weak tornadoes, albeit only at the $\alpha=0.10$ level. It is also worth noting that the temperature and vapor pressure values used to define the surface-based parcel may be affected by the conditions under which the radiosonde was stored. For example, if the radiosonde in question was stored in a cold, dry room, the temperature and vapor pressure values measured at ground level may be slightly lower than the actual environmental values.
In terms of the instability (CAPE) parameters, the results of this research suggest that effective CAPE (eCAPE) is better correlated with tornadogenesis than any single parcel’s CAPE value. Furthermore, it appears that eCAPE is a better discriminator between strong and weak tornadoes, but only by a narrow margin.

During the calculation of the 35 sounding parameters analyzed in this research, bulk shear was calculated across the effective inflow layer. Bulk shear is not typically calculated across this layer, and no research regarding this type of shear has been conducted. The results of this study are conflicting; effective inflow layer shear appears to have better correlations with tornado parameter than any fixed-layer shear variable, but it is worse at discriminating between strong and weak tornadoes. Despite this, the effective inflow layer shear variable appears in the TGP (tornadogenesis parameter) developed in this research through multivariate regression, so further analysis of the utility of this variable appears warranted.

An unexpected result was the strong correlation of both storm motion and downdraft CAPE with tornadogenesis. Storm motion is hardly ever discussed as a factor when forecasting the likelihood of tornadoes, and DCAPE has historically been questioned in terms of its significance in tornado forecasting. Storm motion’s positive correlation with tornadogenesis could be due to the fact that greater storm motion implies a faster mean wind through the depth of the storm; a faster mean wind suggests larger hodographs (greater SRH) and larger shear values, which are known to be correlated with tornadoes. DCAPE, whose value increases with increasing steepness of low- and mid-level lapse rates, may be correlated with tornadogenesis simply because it
is correlated with CAPE (steeper lapse rates cause CAPE to increase). Regardless of these possible explanations, the unexpectedly strong correlations of storm motion and DCAPE pose questions about their potential utility in tornado forecasting.

No significant support can be found in this research for replacing mlCAPE/CIN with eCAPE/CIN in SigTor. Despite eCAPE’s better correlations with tornadogenesis, replacing mlCAPE with eCAPE appeared to offer little-to-no improvement to SigTor. However, the tornadogenesis parameter (TGP) developed in this research through the use of multivariate linear regression appears to have great potential in tornado forecasting. It was developed based on tornado/sounding data from 2006 and prior, yet when it was tested on independent tornado/sounding data from 2007 to present, it was found to be even more strongly correlated with tornadogenesis. This better correlation, as explained previously, may be due to increased reporting of tornado events to the National Weather Service. Regardless of this effect, TGP’s utility verified with a set of entirely independent data, suggesting that it may be a useful tool for tornado forecasting. Additionally, it was shown that TGP discriminates better than SigTor between strong and weak tornadoes, further supporting its potential use in tornado prediction.

The formulation of TGP raises additional questions. For example, if the most-unstable parcel’s LCL was found to be the best correlated with tornadogenesis among all parcels’ LCLs, why does the surface-based LCL appear in TGP? The same question can be raised about other variables included in the formulation of TGP. However, tornadogenesis depends on the complex interaction of many different factors; while
muLCL may be the best stand-alone indicator of the likelihood of tornadoes, when all factors are included in a single parameter, perhaps it is justified to exclude it to obtain the best result. Another interesting feature of TGP is that its value decreases with increasing equilibrium level height and increases with increasing level of free convection height. This suggests that TGP (and hence tornadogenesis) favors high-CAPE environments in which the buoyant energy is contained within a shorter layer. If CAPE is held constant, TGP will be larger in the case of a shallower buoyant layer. Because CAPE is directly proportional to maximum updraft kinetic energy, it is also proportional to the maximum velocity that a parcel in the updraft could reach. Therefore, if the same amount of CAPE is contained within a shallower layer, the acceleration of a parcel in the updraft will be greater. This indicates that TGP favors updraft parcels with greater acceleration; this factor is not taken into account in other composite parameters such as SigTor.

c. Strengths and Limitations

One of the greatest strengths of this research is the large sample size used. Almost 25,000 tornadoes were considered, and over 3,000 soundings were analyzed in total. The large sample size, combined with the averaging procedure used to make plots such as Figure 6, helps to reduce the chance that soundings non-representative of the tornadic environment skewed the results. Furthermore, the linear regression and T-test procedures used in this research provide an effective, quantitative method of analyzing the factors leading to tornadogenesis. The robustness of the averaging procedure used in this study should be considered. Correlation coefficients are not robust to outliers,
and because the averaging procedure serves to greatly reduce (or entirely remove) the influence of outliers (such as the rare soundings with tornado parameters of over 100), it improves the robustness of this study’s results. Additionally, the averaging procedure reduces the impact of excessive variance on $R^2$ caused by the inclusion of non-representative soundings, further increasing the robustness of the final results.

However, there are several limitations to this type of research. First of all, as mentioned above, soundings that are not representative of the environments that produced their “proximal” tornadoes pose an issue. For example, some soundings in this analysis contained zero instability. In all likelihood, these soundings were not representative of the environments they were supposed to represent. A large sample size, though, combined with averaging can offset many of these concerns.

Another assumption worth noting is that all of the analyzed variables are linearly related to the number/intensity of tornadoes produced. While some of the 35 variables analyzed may display a linear relationship with tornadogenesis, others may display exponential relationships, for example. None of the plots examined, however, revealed an identifiable relationship other than a linear one.

A major factor that has been left out in this analysis is lifting. Without an initial source of lifting to bring a parcel to its LFC, no thunderstorm will initially develop, and hence, no tornadoes will occur. This presents a major problem for using parameters like SigTor or TGP to forecast tornadoes. Environments with high SigTor/TGP values may never develop storms to begin with if no source of lifting is present. Unfortunately, analysis of individual soundings can provide no indication of a source of lifting. Lifting
can be provided by various synoptic and mesoscale features including long/shortwave troughs, fronts, drylines, and outflow boundaries, but radiosonde soundings provide no way to determine whether a source of lifting is present. Therefore, plots of SigTor/TGP values must always be compared with a surface/upper-air analysis to determine whether a source of lifting is present.

Lastly, only tornadic soundings were considered in this analysis. Any interpretation of the results of this research must take this consideration into account. For example, this research seems to suggest that CIN (convective inhibition) has little-to-no influence on tornado development. This is not the case; a very strong cap can sometimes prevent even a volatile environment from producing any convection. Therefore, most of the soundings used in this analysis have a weak to moderate cap; it is important to recognize that some soundings with a high TGP value may never produce storms due to very negative CIN (or due to lack of lifting).

do. Suggestions for Future Research

The results of this analysis suggest that effective CAPE (eCAPE), the average CAPE value for all parcels within the effective inflow layer, could be used as an effective aid in tornado forecasting and/or as a replacement for mCAPE in composite parameters such as SigTor. eCAPE’s ability to discriminate between strong and weak tornadoes has been shown, and it has also been shown to correlate strongly with the number/intensity of tornadoes produced in a particular environment, but its ability to discriminate between tornadic and non-tornadic soundings should be evaluated. Furthermore, the most-unstable parcel’s LCL should be considered as potentially more
relevant in tornado prediction than either the surface-based LCL or the mixed-layer LCL. Again, comparison of muLCL heights between tornadic and non-tornadic soundings should be evaluated. Storm motion (speed) and downdraft CAPE also warrant further consideration as potential tornado forecasting parameters.

The tornadogenesis parameter (TGP) developed in this research shows promise for use as a tornado forecasting parameter when compared with the Storm Prediction Center’s “effective” SigTor. However, its utility in distinguishing between tornadic supercells, non-tornadic supercells, and non-supercell storms should be evaluated further before its use is considered. Additional modifications to TGP could be made by expanding the data set used in this research or by using model data. Improvements to TGP, as well as to other composite parameters, could potentially be made if a term is included that represents lifting (perhaps, a measure of convergence).

Finally, the methods of using linear correlation coefficients and multivariate regression appear, based on this research, to be effective ways to both verify and modify our current understanding of the atmospheric conditions that promote tornadogenesis.
Acknowledgments

I would like to thank Dr. David Nolan, Dr. Paquita Zuidema, and Dr. Sharan Majumdar for their valuable input. I would also like to thank Matt Onderlinde for his assistance and for his suggestions regarding multivariate linear regression. Lastly, I would like to thank Dr. Gary Hitchcock for overseeing the Departmental Honors/senior thesis program for marine and atmospheric science at the University of Miami.


