5-12-2023

Tornado outbreak false alarm probabilistic forecasts with machine learning

Kirsten Reed Snodgrass
Mississippi State University, kirstensnodgrass13@gmail.com

Tornado outbreak false alarm probabilistic forecasts with machine learning

By

Kirsten Reed Snodgrass

Approved by:

Andrew E. Mercer (Major Professor/Graduate Coordinator)
Jamie L. Dyer
Michael E. Brown
Rick Travis (Dean, College of Arts & Sciences)

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Geosciences
in the Department of Geosciences

Mississippi State, Mississippi

May 2023

Name: Kirsten Reed Snodgrass

Date of Degree: May 12, 2023

Institution: Mississippi State University

Major Field: Geosciences

Major Professor: Andrew E. Mercer

Title of Study: Tornado outbreak false alarm probabilistic forecasts with machine learning

Pages in Study: 66

Candidate for Degree of Master of Science

Tornadic outbreaks occur annually, causing fatalities and millions of dollars in damage. By improving forecasts, the public can be better equipped to act prior to an event. False alarms (FAs) can hinder the public's ability (or willingness) to act. As such, a probabilistic FA forecasting scheme would be beneficial to improving public response to outbreaks.

Here, a machine learning approach is employed to predict FA likelihood from Storm Prediction Center (SPC) tornado outbreak forecasts. A database of hit and FA outbreak forecasts spanning 2010 – 2020 was developed using historical SPC convective outlooks and the SPC Storm Reports database. Weather Research and Forecasting (WRF) model simulations were done for each outbreak to characterize the underlying meteorological environments. Parameters from these simulations were used to train a support vector machine (SVM) to forecast FAs. Results were encouraging and may result in further applications in severe weather operations.

DEDICATION

This thesis is wholeheartedly dedicated to my beloved parents, who have been my source of inspiration, who have given me strength when I thought of giving up, and who continually provide their moral, spiritual, and emotional support. I would also like to dedicate this work to my sister, my partner, and my dear friends who shared their words of advice and encouragement to finish this study. And lastly, I dedicate this thesis to God. Thank you for the guidance, strength, and wisdom to surpass all trials encountered and for giving me determination to pursue this study and for ultimately making this opportunity possible.

ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my committee chair, Dr. Andrew Mercer, as without his guidance, patience, and persistent help this thesis would not have been possible. I would also like to thank my committee members Dr. Jamie Dyer and Dr. Mike Brown whose assistance, perspective, and helpful suggestions were paramount in this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Tornadic outbreaks occur in the United States every year, causing numerous fatalities and costing millions of dollars in damage, thus highlighting the importance of mitigating outbreak impacts. For example, the April 3-4, 1974 "Super Outbreak" spawned 148 tornadoes, and it is estimated that if the event were to happen today, the insured losses could reach as much as $3.5 billion from wind damage (i.e., tornadoes and derechos) alone (Beatty 2004). In more recent history, the April 27, 2011 "Super Outbreak" generated 199 tornadoes over a 24-hour period which resulted in 316 fatalities (Sanders et al. 2020), more than 2,700 injuries, and approximately $4 billion dollars in damage (Knupp et al. 2014). Another recent impactful event was the 2020 Easter Outbreak, which was responsible for 145 tornadoes and 30 tornado-related fatalities during the two-day period (Guyer 2021). These major tornado outbreaks are troubling to those living in regions prone to severe weather, as well as forecasters. By improving tornado outbreak forecasts, the public can be better equipped to act prior to an impactful event. However, false alarms (FAs), which are forecasted outbreaks that do not materialize, can hinder the public's ability (or willingness) to take action. Many studies have examined the effects of individual tornado FAs (Barnes et al. 2007, Brotzge et al. 2011, and Walters et al. 2020), but the scope and impact of FA forecasts with tornado outbreaks is mostly unknown.

Anecdotal evidence implies that FAs, in the context of individual tornadoes, reduce public response to subsequent warnings related to severe weather, which prompted research into

this issue (Barnes et al. 2007, Simmons and Sutter 2009). The Simmons and Sutter (2009) study showed an increased number of fatalities and injuries sustained from tornadoes that occur in areas with a higher false alarm ratio (FAR). With a 1 standard deviation increase in the FAR, expected fatalities increased by 12%–29% and expected injuries by 13%–32% (Simmons and Sutter 2009), providing evidence for the assumed relationship between FA tornadoes and casualties. A FA diminishes the reliability of a warning and forecasting system, and the loss of credibility has significant implications in how people respond (e.g., less likely to interrupt what they are doing to take shelter, replanning activities the day an outbreak is forecasted to occur, etc.). It is reasonable to assume that the same applies to FAs in the context of tornado outbreaks, though at a broader scale. Though purely speculative, FAs and the overuse of certain words such as "bust" can desensitize the public to future events no matter the magnitude. Outbreak forecasts are challenging, and our culture frowns upon planning for the worst and nothing occurring. Occasionally, there appears to be frustration if nothing happens and there was preparation involved.

These concerns suggest the need for research and education to systemically improve outbreak forecasting by reducing this FA problem, which would improve public response, and potentially reduce future casualties. To address this issue, this project will quantify the likelihood that an outbreak forecast is a FA forecast, alerting forecasters of FA potential, based on atmospheric predictors that best separate FA and actual outbreak environments.

CHAPTER II

BACKGROUND

Overview of Tornado Outbreaks

Tornado outbreaks are rather rare, occurring about 20–30 times each year in the United States (Schneider et al. 2004). While rare, a formal definition for a tornado outbreak does not exist. However, the American Meteorological Society's Glossary of Meteorology defines an outbreak as "multiple tornado occurrences associated with a single synoptic-scale system" (Glickman 2000). Past work has expanded on this description using different variables and varying criteria to create a definition specific to their areas of research. Some studies (Pautz 1969, Galway 1977) define an outbreak by the number of tornadoes that occurred, while other studies include additional parameters in their definition (Doswell et al. 2006, Fuhrmann et al. 2014, Shafer and Doswell 2010). Doswell et al. (2006) avoided specifically defining an outbreak, but instead produced a ranking of different outbreak cases using the previously stated American Meteorological Society's Glossary of Meteorology definition. This ranking index was implemented on tornadic outbreaks and nontornadic outbreaks, which were classified as six or fewer tornado reports. The present study will define outbreaks and FAs loosely based on this criterion, which will be outlined in a later section.

Tornado Outbreak Meteorological Characteristics

Mercer et al. (2012) address the synoptic-scale patterns of tornado outbreaks and nontornadic outbreaks. Through this same study, many case studies (Fujita et al. 1970, Ferguson

et al. 1987, Corfidi et al. 2010, Roebber et al. 2002, Agee et al. 1975, Lee et al. 2006) were found to have comparable synoptic conditions prior to the development of each type of outbreak. Tornado outbreaks had three key features:

- Upper-level trough west of the outbreak location that prompted the intensification of a surface cyclone through synoptic-scale processes.
- Upper-level jet streak coupled with the center of rapid surface pressure falls leading to the deepening of a synoptic cyclone and an increase in vertical shear.
- Development of an intense surface cyclone that includes an area of greater moisture that allows for the generation of large amounts of instability in the warm sector.

As for nontornadic outbreaks, Johns and Hirt (1987) provide environmental characteristics preceding the development of derechos, one of the most common types of nontornadic outbreaks:

- Rather shallow upper-level trough west of the location of the outbreak.
- Propagation of the trough toward the region of derecho development (characterized as 500 mb height falls of 60 meters in 12 hours before the derecho).
- Ample instability due to moisture located in the warm sector of the surface cyclone that results from the approaching upper-level trough.

To analyze the patterns of the two outbreak types, Mercer et al. (2012) initialized Weather Research and Forecasting (WRF) model simulations for each outbreak type (tornadic and nontornadic). Tornado outbreak fields were characterized by a robust 500-mb trough with an associated vorticity maximum. The trough composites associated with tornado outbreaks showed all mid-level trough tilt types (i.e., positive, negative, and neutral tilts). Tornado outbreaks also revealed stronger low-level temperature advection than the nontornadic outbreaks, which likely enhanced both vertical forcing within the outbreak, a larger magnitude of moist, unstable air into the outbreak region, and veering wind profiles that enhanced vertical wind shear. Atmospheric

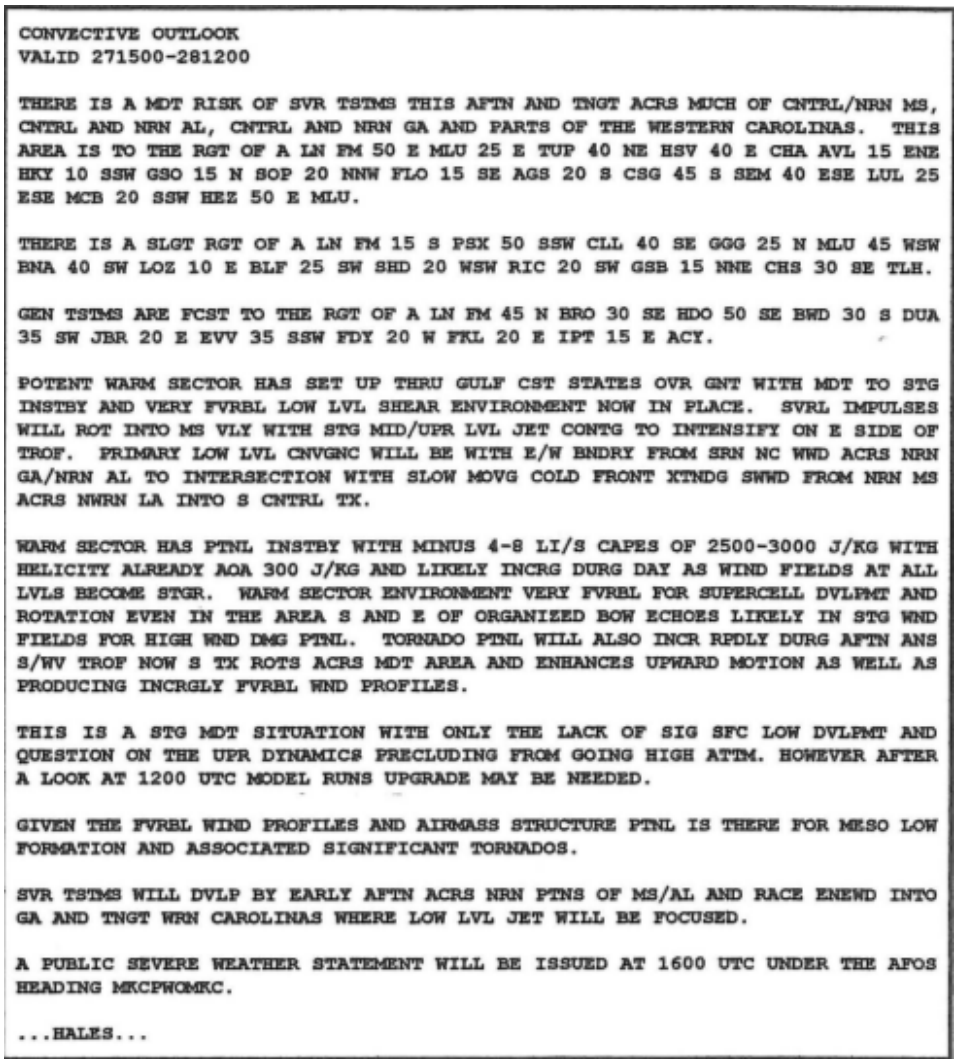
soundings in their study also unveiled statistically significant larger helicity and bulk shear values for tornado outbreaks than for the nontornadic outbreaks. Overall, their study revealed the significance of helicity and shear, particularly directional shear with veering wind profiles, for an environment associated with a tornado outbreak. These findings, specifically shear, are also consistent with other past work (Mercer et al. 2009, Shafer et al. 2010). The characteristics mentioned here should also be considered when determining variables to quantify differences in accurate and FA tornado outbreak forecasts for this research.

Tornado Outbreak Forecasts

Though literature focusing on tornado outbreak forecasting is sparse, some studies have assessed Storm Predication Center (SPC) convective outlooks as a proxy for an outbreak forecast. Metrics such as probability of detection (POD) and false alarm ratio (FAR) can highlight the quality of these outlooks when assessing forecast accuracy. For the slight risk outlook category provided by SPC, Hitchens and Brooks (2012) found an increase in the POD during 1973-1993 followed by the frequency of hits (FOH) increasing, indicating a growth in areal coverage of slight risk followed by smaller, better positioned areas of risk.

While literature pertaining specifically to FA outbreak forecasts is sparse, several studies have addressed outbreak forecasting issues. Hales et al. (1997) discussed the 27 March 1994 outbreak which occurred with an abnormal synoptic setup (i.e., absence of a deepening surface low, strong upper-level trough, and jet streak dynamics). Forecasters at the Storm Prediction Center (SPC) knew there was an unstable air mass present across the outbreak area after analyzing the upper air data and the sounding taken at 1200 UTC on 27 March, as mentioned in their Day One Convective Outlook issued at 1500 UTC (Figures 2.1 and 2.2; Hales et al. 1997). Convective instability increased during the overnight hours due to low-level temperature and

moisture advection. Wind fields also became favorable for supercell development (i.e., increased vertical speed and directional shear). SPC forecasters were able to anticipate the formation of supercells with some nonclassical atmospheric characteristics and they effectively updated their forecast outlooks accordingly. This tornado outbreak event would be considered a correct outbreak forecast (COF) as the outbreak occurred where forecasters anticipated it to happen (Figure 2.3; Hales et al. 1997).



CONVECTIVE OUTLOOK
VALID 271500-281200

THERE IS A MDT RISK OF SVR TSTMS THIS AFTN AND TNGT ACRS MUCH OF CNTRL/NRN MS, CNTRL AND NRN AL, CNTRL AND NRN GA AND PARTS OF THE WESTERN CAROLINAS. THIS AREA IS TO THE RGT OF A LN FM 50 E MLU 25 E TUP 40 NE HSV 40 E CHA AVL 15 ENE HKY 10 SSW GSO 15 N SOP 20 NNW FLO 15 SE AGS 20 S CSG 45 S SEM 40 ESE LUL 25 ESE MCB 20 SSW HEZ 50 E MLU.

THERE IS A SLGT RGT OF A LN FM 15 S PSX 50 SSW CLL 40 SE GGG 25 N MLU 45 WSW BNA 40 SW LOZ 10 E BLF 25 SW SHD 20 WSW RIC 20 SW GSB 15 NNE CHS 30 SE TLH.

GEN TSTMS ARE FCST TO THE RGT OF A LN FM 45 N BRO 30 SE HDO 50 SE BWD 30 S DUA 35 SW JBR 20 E EVV 35 SSW FDY 20 W FKL 20 E IPT 15 E ACY.

POTENT WARM SECTOR HAS SET UP THRU GULF CST STATES OVR GNT WITH MDT TO STG INSTBY AND VERY FVRBL LOW LVL SHEAR ENVIRONMENT NOW IN PLACE. SVRL IMPULSES WILL ROT INTO MS VLY WITH STG MID/UPR LVL JET CONTG TO INTENSIFY ON E SIDE OF TROF. PRIMARY LOW LVL CNVGNC WILL BE WITH E/W BNDRY FROM SRN NC WND ACRS NRN GA/NRN AL TO INTERSECTION WITH SLOW MOVG COLD FRONT XTNDG SWWD FROM NRN MS ACRS NWRN LA INTO S CNTRL TX.

WARM SECTOR HAS PTNL INSTBY WITH MINUS 4-8 LI/S CAPES OF 2500-3000 J/KG WITH HELICITY ALREADY AOA 300 J/KG AND LIKELY INCRG DURG DAY AS WIND FIELDS AT ALL LVLS BECOME STGR. WARM SECTOR ENVIRONMENT VERY FVRBL FOR SUPERCELL DVLPMT AND ROTATION EVEN IN THE AREA S AND E OF ORGANIZED BOW ECHOES LIKELY IN STG WND FIELDS FOR HIGH WND DMG PTNL. TORNADO PTNL WILL ALSO INCR RPDLY DURG AFTN ANS S/WV TROF NOW S TX ROTS ACRS MDT AREA AND ENHANCES UPWARD MOTION AS WELL AS PRODUCING INCRGLY FVRBL WND PROFILES.

THIS IS A STG MDT SITUATION WITH ONLY THE LACK OF SIG SFC LOW DVLPMT AND QUESTION ON THE UPR DYNAMICS PRECLUDING FROM GOING HIGH ATTM. HOWEVER AFTER A LOOK AT 1200 UTC MODEL RUNS UPGRADE MAY BE NEEDED.

GIVEN THE FVRBL WIND PROFILES AND AIRMASS STRUCTURE PTNL IS THERE FOR MESO LOW FORMATION AND ASSOCIATED SIGNIFICANT TORNADOS.

SVR TSTMS WILL DVLP BY EARLY AFTN ACRS NRN PTNS OF MS/AL AND RACE ENEWD INTO GA AND TNGT WRN CAROLINAS WHERE LOW LVL JET WILL BE FOCUSED.

A PUBLIC SEVERE WEATHER STATEMENT WILL BE ISSUED AT 1600 UTC UNDER THE AFOS HEADING MKCPWOMKC.

...HALES...

Figure 2.1 Day One Convective Outlook issued at 1500 UTC 27 March 1994.

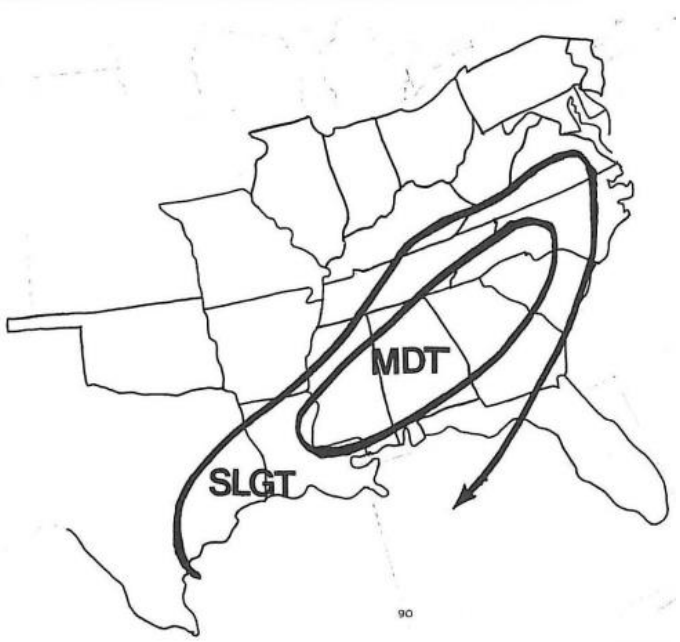


Figure 2.2 Graphic of the Day One Convective Outlook referenced in Figure 2.1.

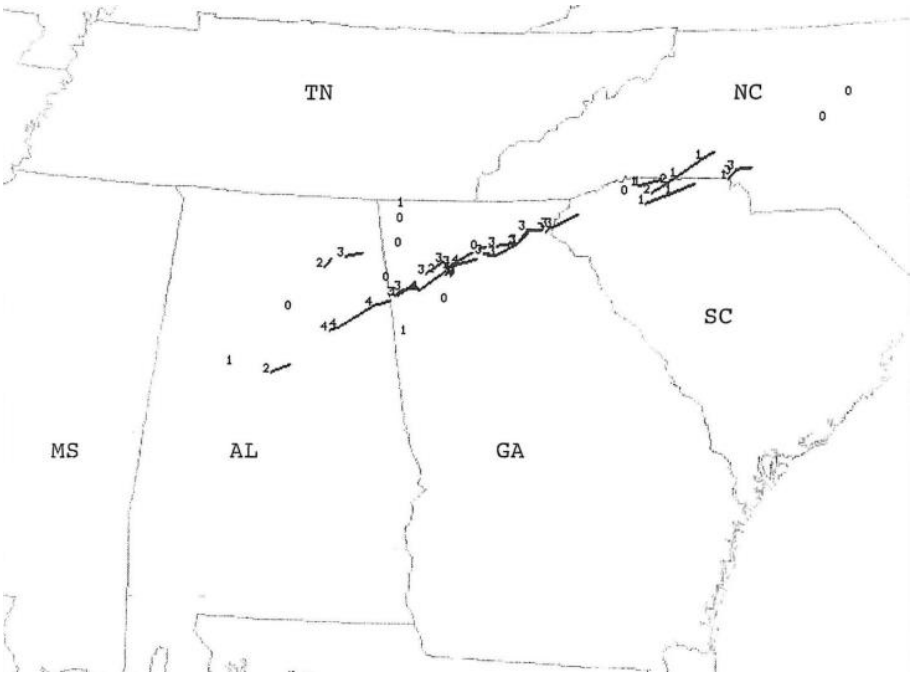


Figure 2.3 Tracks of the tornadoes during the 27 March 1994 outbreak with their F-scales.

A more difficult forecasting challenge arose for the tornado outbreak on 3 May 1999, as documented by Thompson and Edwards (2000) and Edwards et al. (2002). With this event, the buoyancy and shear were favorable and synoptic conditions seemed ideal for an outbreak, but the lack of convergence in the dryline regions and a thick cirrus cloud deck that lowered the potential for heating and mixing in the boundary layer near these drylines caused models to underrepresent the outbreak threat until a few hours prior to its initiation. At that time, the initial supercell developed during a break in the thick cirrus deck. Surface temperatures were also sufficiently high enough to minimize convective inhibition and so the convective outlook forecasts were updated to reflect the more favorable storm environment. This tornado outbreak event is considered more of an initial "miss" scenario, since there was a consensus that severe weather would happen with some supercell thunderstorms, but the exact location was unclear (Figures 2.4 a-e).

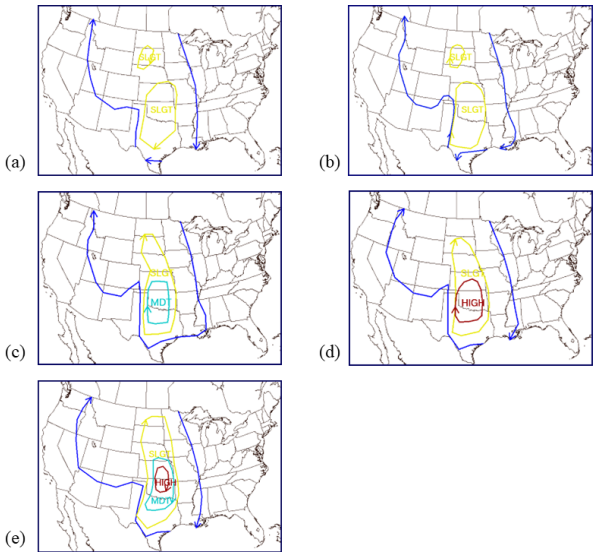


Figure 2.4 Maps of SPC Day One categorical convective outlook risk areas with ending valid times at 1200 UTC 4 May 1999. The beginning valid times are on 3 May 1999 at (a) 0600 UTC, (b) 1300 UTC, (c) 1630 UTC, (d) 2000 UTC, and (e) 4 May 1999 at 0100 UTC.

Opposite of a COF or a miss, FAs are forecasted outbreaks that do not occur. While there are many studies that focus on FAs in the context of individual tornadoes (Barnes et al. 2007, Brotzge et al. 2011, and Walters et al. 2020), there are not many that concentrate on FAs in the context of tornado outbreaks. Mercer and Bates (2014) touches on FAs within SPC's outbreak forecasts (as estimated by convective outlooks). They assessed differences in FA and COF forecast environments at the synoptic scale through the generation of outbreak composites. The study found that FAs had weak synoptic scale setups, and while there were not many differences between high- and low-accuracy forecasts, there was evidence that some meteorological variables were good indicators of FAs. These variables included limited vertical shear, thermal advection, differential vorticity advection, and jet streak magnitudes, which resulted in significant forecast uncertainty that ultimately resulted in a FA forecast (Mercer and Bates 2014). Higher accuracy forecasts were characterized by high amounts of warm air advection (WAA), while low accuracy forecasts had weaker WAA. This implies inadequate amounts of energy and therefore uplift to create tornado outbreaks, leading to a FA case. Their FA composites also exhibited less organized differential vorticity advection distant from a surface low as well as weaker jet streaks and limited vertical wind shear. These fields are feature candidates in the current study. Pressure, wind speeds, moisture aloft and at the surface, lapse rates, and surface temperature were nearly identical amongst the composites, indicating these factors did not strongly differ among the high and low accuracy forecasts.

The Mercer and Bates (2014) study also noted patterns of seasonality for FAs, where most occurred in the summer and fall months. Most tornadoes and tornado outbreaks occur during the spring, and these seasonal observations show that tornado outbreaks that do not occur during the "regular season" have higher forecast uncertainty, suggesting research needs to be

done to establish the synoptic scale parameters connected with off-season outbreaks. This suggests seasonality could be useful in predicting FA outbreak forecasts as well. Specifically, seasonal differences in convective availability potential energy (CAPE) may play a role in FAs. CAPE is usually higher during the warm season and lower during the cool season. As such, fall events may require enhanced synoptic-scale vertical forcing for convective initiation to occur. Additionally, wind shear supportive of supercell development and tornadoes are more frequent during the spring and fall seasons than during the summer. It is understandable why most tornadoes and tornado outbreaks occur during the ”regular season”, as it is more common for favorable instability and wind shear to be collocated. In the summer, wind profiles are normally less favorable for mid-level cyclone development. The lifted condensation level (LCL) is also particularly affected by season. Higher LCLs are common with many primarily nontornadic outbreaks during the warm season, due to the existence of hot, dry, well-mixed boundary layers (Shafer et al. 2010).

Past work has been focused on synoptic-scale discrimination of tornadic and nontornadic outbreaks as well as ranking and identifying outbreak types. Many studies (Mercer et al. 2009 and Shafer et al. 2010) used a ranking scheme (Doswell et al. 2006) to select the top 50 tornadic and nontornadic outbreaks to evaluate the synoptic-scale variables best used to differentiate between the outbreaks. These studies found that kinematic parameters (i.e., bulk shear) were the most effective at differentiating between the tornadic and nontornadic outbreaks while thermodynamic variables were the least useful. Mercer et al. (2012) also agreed with these results and stated the Weather Research and Forecasting (WRF) model has great skill at discriminating between the two outbreak types. This may be the case as the WRF model is a mesoscale model that uses a “compressible, nonhydrostatic dynamical core” (Shafer et al. 2009).

Studies have expressed the need for more research into the false alarm outbreaks and operational classification techniques (Mercer et al. 2009, Mercer et al. 2012, Mercer and Bates 2014). The present study takes a step toward accomplishing those goals, as it seeks to determine whether a statistical classifier, specifically logistic regression and support vector machines (SVM), can correctly identify COFs and forecasts of events that resulted in a FA. This will be done by first identifying variables that show the largest differences among FA and COF environments, working under a null hypothesis that these environments are indistinguishable. It is expected that by identifying these environmental differences, an improved ability to identify FA outbreak environments will be possible, which will help to improve outbreak forecasts in the future.

CHAPTER III

DATA & METHODS

Datasets

This study evaluated a 10-year period (2010-2019) using data supplied by the SPC's U.S. Tornadoes (1950-2021) dataset (https://www.spc.noaa.gov/wcm/). Data from the year 2020 were also included for model verification such that the study spans 11 years. This database includes information regarding tornado intensity, path length, timing, etc., as well as information about severe hail and wind reports. Outbreaks were limited to a single day (defined as 6 AM to 6 AM CST the next day), such that individual days within a multi-day event were treated as independent events. The valid time for an outbreak was set as a three-hour window during which the outbreak reached its peak tornado production. If multiple three-hour windows produced the same peak outbreak tornado rate, the first of these windows was used.

As the ultimate goal of this work was to create a probabilistic classifier that will identify upcoming outbreaks as "likely" or "unlikely" to yield a FA, a defined an outcome of 1 represents a 100% chance of a FA and an outcome of 0 represents a 100% chance of a COF. To classify outbreaks into these two categories, a list of verified tornadic and nontornadic outbreaks were needed. This project followed Doswell et al. (2006) by examining severe weather reports in 24-hour periods as stated above.

To formulate these lists, the SPC tornado database was used to isolate tornado outbreaks with above-average and below-average tornado counts. Importantly, past research (Shafer and

Doswell 2010) showed that the annual mean tornado counts within outbreaks had an upward trend prior to 2010. However, recent updated storm report information post-2010 showed this upward trend largely flattened. To verify this shift, the yearly means were calculated and plotted for this study's 11-year period (Figure 3.1). The slope of the resulting trend line (m = -0.530, p = 0.245) showed a non-significant negative slope, which is counter to the results from Shafer and Doswell (2010)'s study period. As a result, the global mean over the 11-year study period was used to categorize outbreaks as above-average and below-average. With a global mean of 5.630 tornadoes per outbreak, an overall mean of six (a value also used in Doswell et al. 2006) was the value chosen to establish COF and FA cases.

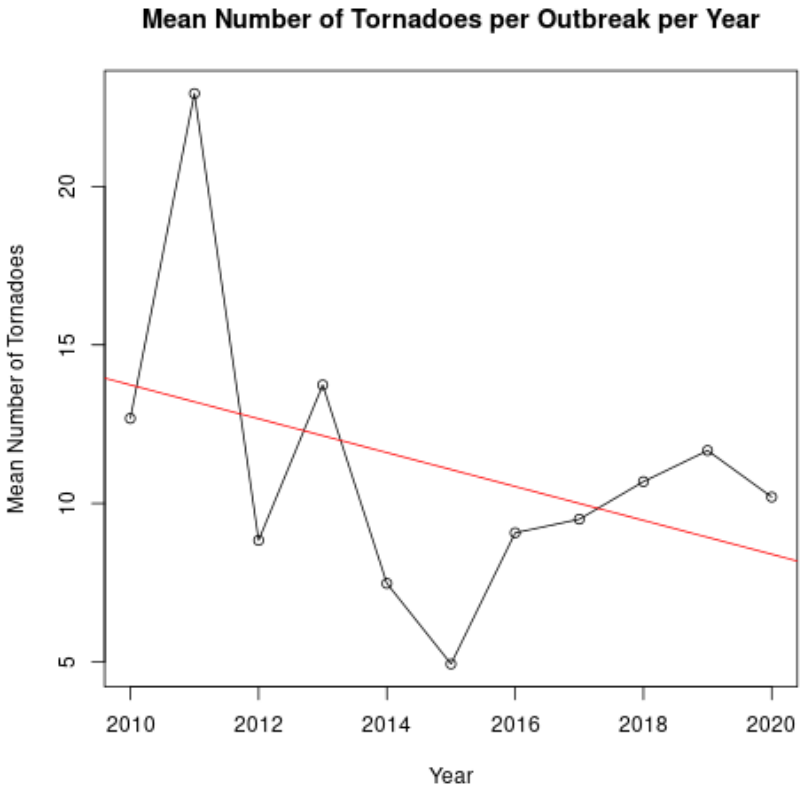


Figure 3.1 The yearly means calculated for the study period, and it's fitted trend line.

The SPC's Day One tornado probability forecasts include 2%, 5%, 10%, 15%, 30%, 45%, and 60% neighborhood probability contours in addition to probabilistic graphics that depict severe and general thunderstorm threats across the continental United States (CONUS). For each event, the SPC's Convective Outlook Archive (https://www.spc.noaa.gov/archive/) at 1200 UTC was explored to identify events that occurred on days with at least a 10% tornado probability percentage present (Figure 3.2).

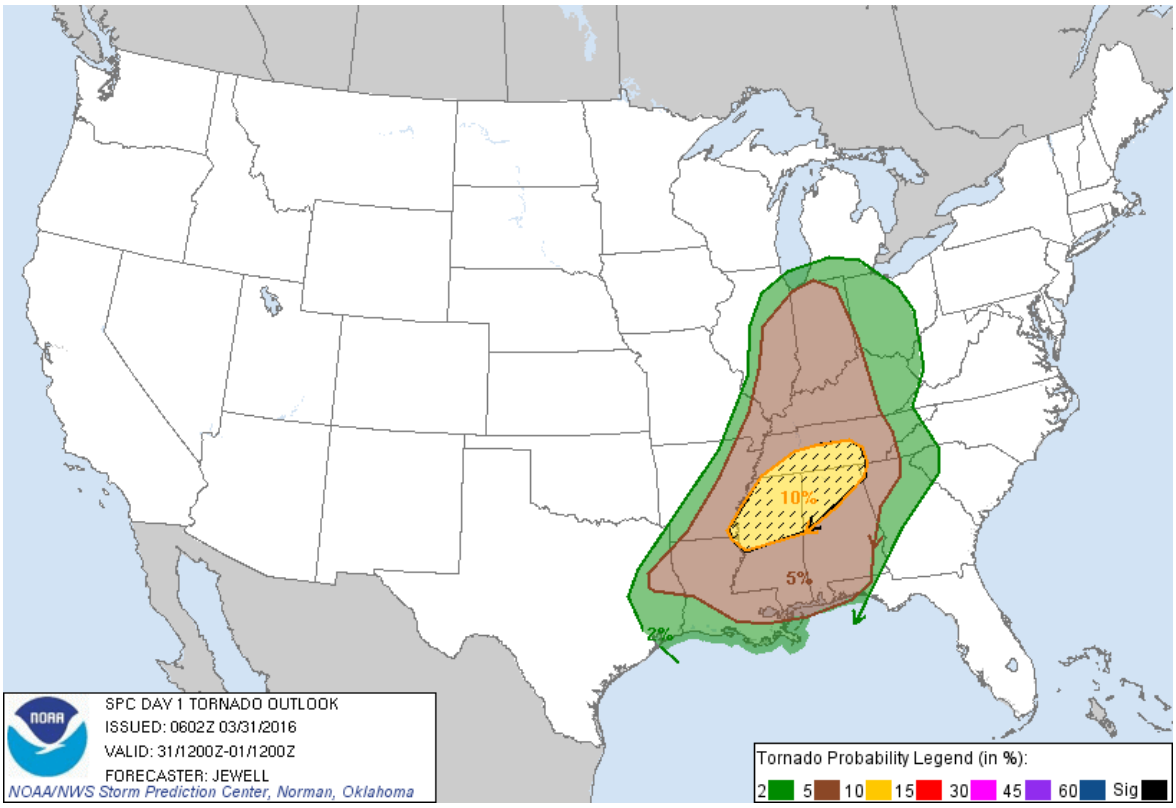


Figure 3.2 Example event from 31 March 2016, with a 10% tornado probability.

This approach assumes that a SPC tornado probability forecast of 10% or higher suggests the SPC is expecting an outbreak of tornadoes over the specified 10% region. With this in mind, 10% was chosen to formulate the final list of cases. Thus, six or more tornadoes within the forecasted 10% outbreak region was defined as a COF, while a 10% region that had fewer than six tornadoes (and thus was below-average) within it was counted as a FA. While the number of 10% cases has slightly declined over the period of this study (Figure 3.3), the 10% tornado probability region yielded a reasonable sample size for both FAs and COFs, with 93 FA cases and 109 COF cases being found over the total 11-year period (Figure 3.4 and 3.5).

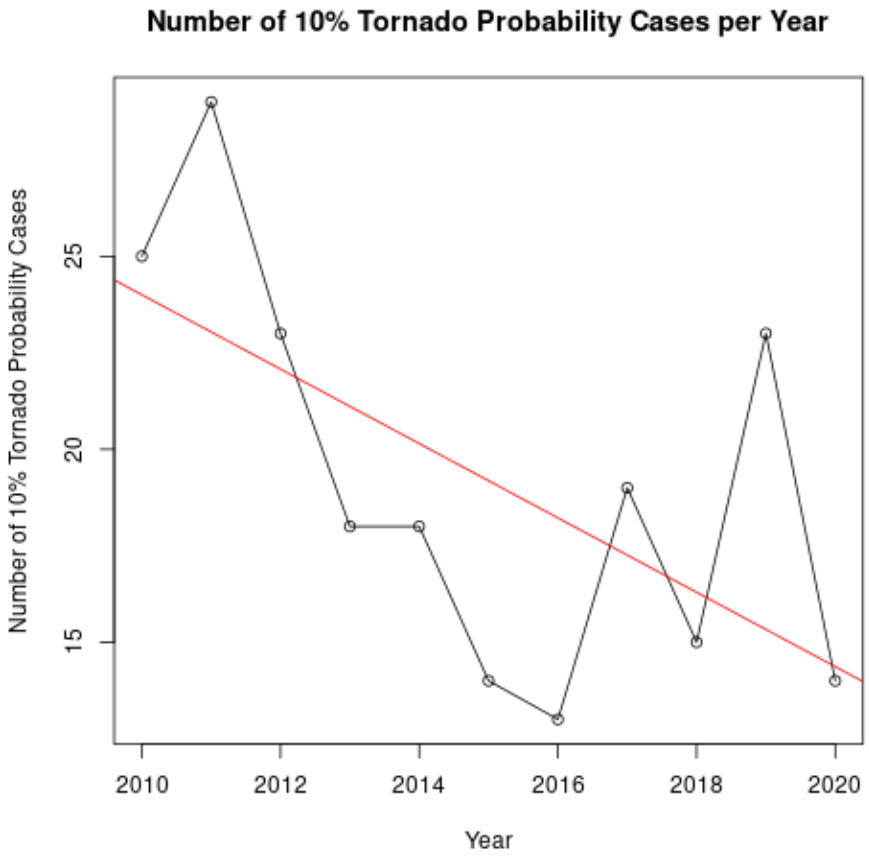


Figure 3.3 The number of cases with a 10% tornado probability over the study period, with a fitted trend line.

2010 (n=16)	2011 (n=8)	2012 (n=7)	2013 (n=15)	2014 (n=6)	2015 (n=6)	2016 (n=9)	2017 (n=13)	2018 (n=8)	2019 (n=14)	2020 (n=7)
20100310	20110224	20120122	20130129	20140403	20150508	20160121	20170122	20180224	20190223	20200110
20100424	20110228	20120302	20130210	20140427	20150509	20160202	20170214	20180319	20190303	20200328
20100430	20110522	20120414	20130417	20140428	20150510	20160223	20170228	20180413	20190413	20200412
20100501	20110524	20120601	20130418	20140511	20150516	20160224	20170402	20180414	20190418	20200413
20100510	20110525	20120617	20130518	20140616	20151223	20160331	20170405	20180501	20190419	20200419
20100518	20110530	20121017	20130519	20140630	20151227	20160426	20170516	20180502	20190430	20200422
20100519	20110620	20121225	20130520			20160509	20170517	20180920	20190507	20200423
20100522	20111107		20130527			20160622	20170518	20181031	20190508	
20100524			20130529			20161129	20170527		20190517	
20100605			20130530				20170612		20190520	
20100617			20130531				20170613		20190522	
20100623			20130612				20170910		20190523	
20100713			20131004				20171105		20190527	
20100714			20131117						20190529	
20101026			20131221							
20101129										

Figure 3.4 COF Case Dates (YYYYMMDD).

2010 (n=8)	2011 (n=7)	2012 (n=17)	2013 (n=4)	2014 (n=13)	2015 (n=8)	2016 (n=4)	2017 (n=7)	2018 (n=8)	2019 (n=9)	2020 (n=8)
20100406	20110511	20120224	20130408	20140406	20150408	20160330	20170306	20180320	20190309	20200312
20100423	20110519	20120228	20130409	20140413	20150424	20160429	20170328	20180415	20190314	20200319
20100512	20110619	20120229	20130521	20140429	20150603	20160508	20170329	20180503	20190521	20200424
20100620	20110626	20120303	20130613	20140430	20150622	20160526	20170330	20181003	20190526	20200522
20100717	20110726	20120311		20140508	20150713		20170404	20181105	20190528	20200607
20100719	20110801	20120319		20140528	20150822		20170426	20181130	20190719	20200903
20100723	20110905	20120412		20140603	20151116		20170510	20181220	20190924	20201224
20100830		20120413		20140614	20151117			20181226	20191020	20201231
		20120415		20140629					20191216	
		20120427		20140727						
		20120505		20140831						
		20120524		20141013						
		20120530		20141123						
		20120614								
		20120908								
		20120918								
		20121226								

Figure 3.5 FA Case Dates (YYYYMMDD).

Methodology

To emulate an operational forecasting environment, a nonhydrostatic mesoscale numerical weather prediction model was needed to predict meteorological conditions underlying each outbreak. Given the desire for a true forecast mode, initial conditions for atmospheric and soil fields were obtained from the NCEP GFS 0.5° analysis valid at 0000 UTC for each outbreak date, and lateral boundary conditions were updated every three hours using the corresponding GFS forecast fields. COF and FA cases from April 2011 were excluded as this period had date and time discrepancies in the data. These GFS model data were used to initialize a Weather Research and Forecasting (WRF) model simulation of each outbreak to characterize the underlying outbreak environment on the synoptic scale and mesoscale. WRF output was used to build a predictor matrix of severe weather parameters that have been shown useful in past research in characterizing tornado outbreak environments. The WRF simulations employed the same physics parameterizations as the National Severe Storms Laboratory version of the WRF Model (WRF-NSSL; Kain et al. 2008, Skamarock et al. 2008). However, this study used 12-km grid-spacing with a model initialization at 0000 UTC, a model period of 36 hours, and a cumulus scheme, which contrasts with the WRF-NSSL. The use of a 12-km grid rather than the WRF-NSSL 4-km was done to reduce simulation run time, and a cumulus scheme was included to determine when to generate a convective column and how fast to make the convection act (Dudhia 2014). WRF version 4.0 was used with a domain encompassing most of the United States except for portions of the west (Figure 3.6). Other model specifications during the 11-year analysis period were unchanged and are listed in Table 3.2. Initially, 132 predictors were chosen for the analysis and were based on spatial and temporal statistics among 22 unique meteorological predictors (Table 3.3).

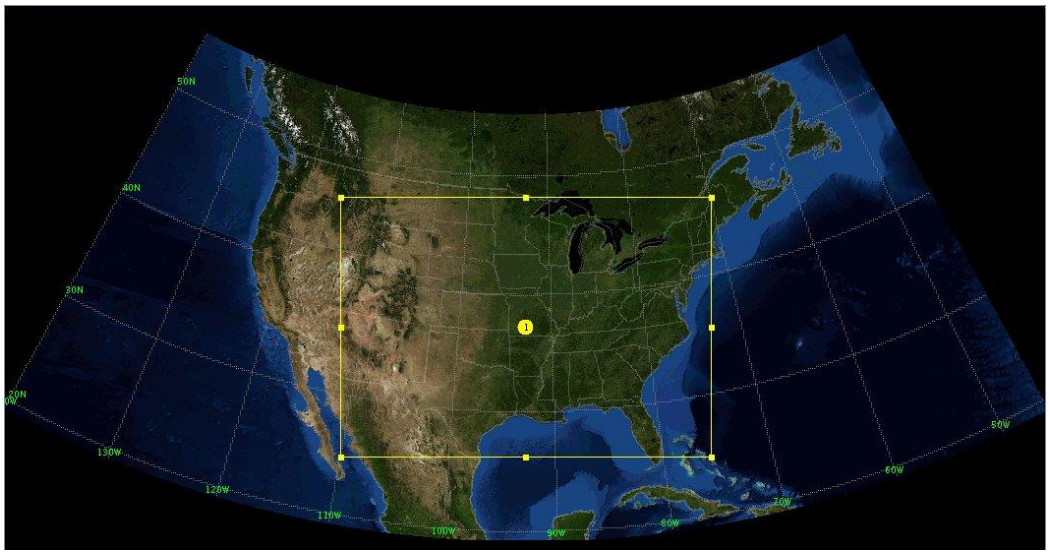


Figure 3.6 Domain of the GFS ensemble simulations generated with WRF version 4.0.

Table 3.1 Model configuration and physic parameterizations used in the GFS ensemble simulations generated with WRF version 4.0. (**deviates from WRF-NSSL)

Grid Configuration	
Initial Conditions	0.5° GFS analysis from 0000 UTC
Lateral Boundary Conditions	3-h 0.5° GFS analyses
Horizontal Grid Spacing	12 km**
Number of Grid Points	343 x 240
Number of Vertical Levels	35
Model Top	50 mb
Time Step	24 s
Physics Parameterizations	
Cumulus	Kain-Fritsch scheme**
PBL	Mellor-Yamada-Janjic scheme (Mellor and Yamada 1982; Janjić 2002)
Surface Layer	Eta similarity
Land Surface Model	Noah Land Surface Model (Chen and Dudhia 2001)
Microphysics	WRF Single-Moment 6-class scheme (Hong and Lim 2006)
Shortwave Radiation	Dudhia scheme (Dudhia 1989)
Longwave Radiation	RRTM scheme (Mlawer et al. 1997)

Table 3.2 22 overall predictors based on spatial and temporal statistics.

Overall Predictors	**Reference**
Surface CAPE (J/kg)	Stensrud et al. (1997)
Surface CIN (J/kg)	Markowski (2002)
Surface LCL (m)	Rasmussen and Blanchard (1998)
Most Unstable CAPE (J/kg)	Ukkonen and Mäkelä (2019)
Most Unstable CIN (J/kg)	
Most Unstable LCL (m)	
LI (K)	
SRH (0-1 km) (m^2/s^2)	Colquhoun and Riley (1996)
SRH (0-3 km) (m^2/s^2)	
Effective SRH (m^2/s^2)	
Bulk Wind Difference (0-1 km) (m/s)	Weisman and Klemp (1984)
Bulk Wind Difference (0-6 km) (m/s)	
Effective Bulk Wind Difference (m/s)	
SCP	Grams et al. (2012)
STP	
Specific Humidity (g/kg)	Mercer and Bates (2014)
Temperature (K)	
Temperature Advection (850 mb) (K/s)	
Wind Magnitude (300 mb) (m/s)	
Wind Magnitude (850 mb) (m/s)	
Vorticity Advection (500 mb)	
Lapse Rate (700-500 mb) (K/m)	

As each of the 22 predictors had both spatial variability within the outbreak region and temporal variability at each gridpoint, dimension reduction was completed to represent the spatial outbreak environment as individual predictors. This was done in two ways. First, the outbreak's spatial maximum and mean within the 10% outbreak region for each of the 22 predictors was computed for each date for each WRF timestep (yielding a 24-hour time series of spatial maximum/mean for each outbreak) (Figure 3.7). Next, the temporal minimum, maximum,

and mean were computed from these spatial time series'. Using these dimension reduction methods, each of the 22 predictors had 6 possible subsets, yielding 132 total WRF-derived predictors for the machine learning models.

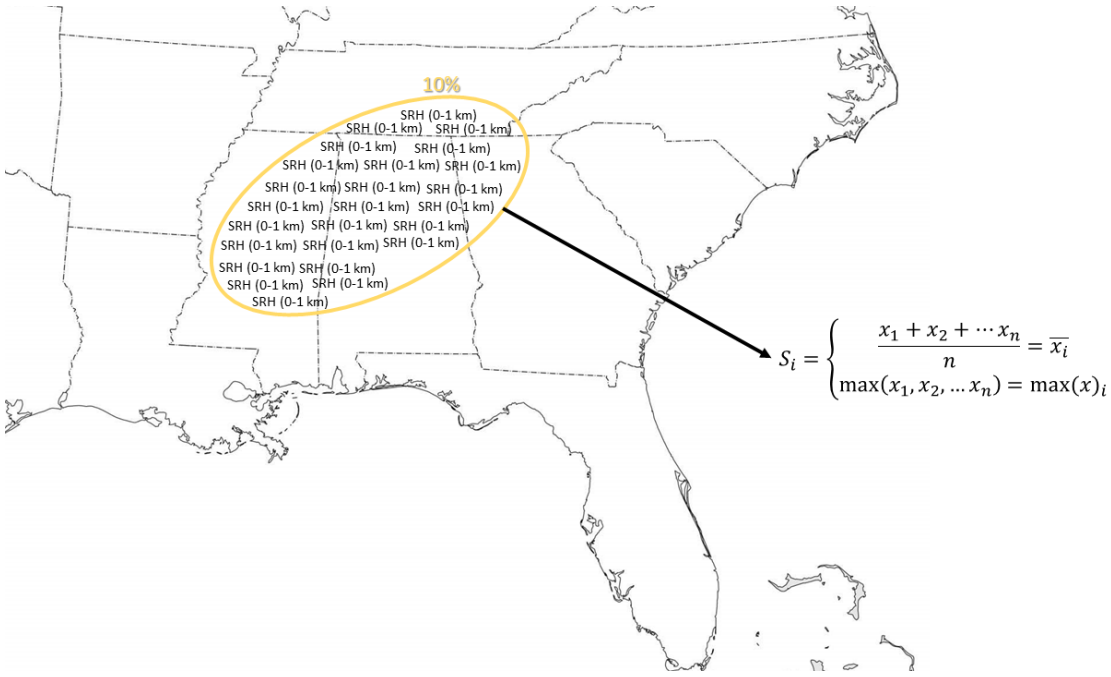


Figure 3.7 Example of a spatial calculation for one (SRH 0-1 km) of the 22 meteorological predictors. Every predictor's (x) spatial mean and maximum are computed within the 10% outbreak region for each date for each WRF timestep (i = 0-23 hours), yielding a 24-hour time series of spatial mean/maximum for each outbreak.

As the machine learning methods are sensitive to predictor magnitude, normalization of the predictors was done to ensure each predictor had equal predictive weight in the models. The normalization was done as follows:

$$X'_i = \frac{(X_i - X_{min})}{(X_{max} - X_{min})} \tag{3.1}$$

where X'_i is the normalized data for parameter i; X_i is the original data for i; X_{max} is the maximum of i; and X_{min} is the minimum of i (Zhang et al. 2018). This results in scaling all predictors from 0 to 1.

As this number of predictors is near the size of the number of outbreaks spanning 2010 – 2019 (187), it was necessary to reduce the predictor set to a more manageable size to reduce the likelihood of overfitting or an ill-posed statistical model. Three feature selection methodologies were utilized to reduce the predictor set and determine which yielded the best predictor set and, therefore, the best performing model. The first method employed was a forward selection stepwise methodology where the goal was to maximize the model skill in detecting FAs via a logistic regression. Stepwise was selected due to its ability to handle numerous amounts of potential predictors and quickly choose the most beneficial predictors from the available options. The forward stepwise procedure works by adding predictors cumulatively based on the amount of skill they provide. The resulting analysis yields a list of predictors sorted in order of increasing skill offered to the model. Skill was measured by the Heidke Skill Score (HSS) which is explained shortly.

The next method that was utilized was permutation testing. This was completed to assess which predictors have the best overall separation (that is statistically significant), between FA and COF environments, which logically would prove useful in distinguishing FAs and COFs. A permutation test is a data resampling method that determines if the means of two distributions are different (Efron and Tibshirani 1993, Mercer and Richman 2007, Mercer et al. 2009, Potvin et al. 2010). Permutation tests with 2000 replications were conducted on each of the 132 predictors and the 2010-2019 outbreaks, where the FA and COF cases are treated as the separate testing entities. The resulting *p* values from those tests were compared against rejection criteria of 0.05, 0.025, 0.01, and 0.001 (corresponding to the 95%, 97.5%, 99%, and 99.9% confidence limits). While these criteria led to an increasingly high probability of committing a type II error,

the distinctness of the predictors was greatest when this p-value threshold was smallest (which should suggest the best classification performance).

The final method employed to better assist with the feature selection process was principal component analysis (PCA). PCA is a linear analysis procedure which decreases the dimensionality of a dataset by transforming the dataset into a new dataset. This new dataset is comprised of linear combinations of the original data which are known as principal components (PCs). PCA begins with the primary equation:

$$\boldsymbol{Z} = \boldsymbol{F}\boldsymbol{A}^T \tag{3.2}$$

where $\boldsymbol{Z}$ is the original matrix in standard anomaly form, $\boldsymbol{F}$ is the PC score matrix, and $\boldsymbol{A}$ is the matrix of PC loadings used to transform $\boldsymbol{Z}$ into $\boldsymbol{F}$ (Wilks 2019, Mercer et al. 2012, Mercer and Bates 2014). The first step to solving this equation for $\boldsymbol{F}$ and $\boldsymbol{A}$ is to obtain a correlation matrix for $\boldsymbol{Z}$. Once $\boldsymbol{R}$ is computed (the correlation matrix), the eigenanalysis is performed:

$$\boldsymbol{R} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^T \tag{3.3}$$

where $\boldsymbol{R}$ is diagonalized into an eigenvalue matrix $\boldsymbol{D}$ with an associated eigenvector matrix $\boldsymbol{V}$. Once we have the eigenvalues and eigenvectors, we can compute the loading matrix $\boldsymbol{A}$ as:

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{D}^{1/2} \tag{3.4}$$

After this is completed, the score matrix $\boldsymbol{F}$ can be solved:

$$\boldsymbol{F} = \boldsymbol{Z}(\boldsymbol{A}^T)^{-1} \tag{3.5}$$

and the variance explained by each PC can be computed:

$$Variance\ Explained = \frac{\lambda_n}{\sum \lambda_n}(100\%) \tag{3.6}$$

where λ_n represents the eigenvalues derived from the PCA. The PC scores were used as predictors in both models. They were thought to be beneficial as they provide unique, uncorrelated variability within the predictor space that can summarize covariability within the predictors as a small subset. This makes the model more concise at the expense of losing some variance explained.

The resulting predictors from each of the feature selection methods were then used in a logistic regression and a support vector machine (SVM; Haykin 1999, Cristianini and Shawe-Taylor 2000) to assess the classification skill. Both methods are statistical analysis techniques that are used to predict a binary outcome (i.e., the probability of a FA versus a COF in this present study). This probability can be utilized as a forecast application. A threshold cutoff probability (typically based on climatology) can then be used to categorize each probabilistic forecast as FA or COF.

While logistic regression is well known (Wilks 2019), SVM is a more novel machine learning technique. A SVM defines a multidimensional hyperplane for classification between binary classes (here FAs and COFs). This technique has been used in previous meteorological studies (Richman et al. 2005, Trafalis et al. 2005, Mercer et al. 2008, Mercer et al. 2009), but its presence in literature pertaining to tornado outbreaks is limited. However, the SVM model was selected for this study due to its ability to model nonlinearly separable data. SVMs attempt to solve for the hyperplane surface, which, according to Haykin (1999), is mathematically given as:

$$\boldsymbol{w}^T\boldsymbol{x} + b = 0 \tag{3.7}$$

where **w** is a vector of weights, **x** represents the covariates, and b is an intercept. Since SVMs are classification machines designed to distinguish between classes, modifications to the hyperplane equation for our two classes can be represented as:

$$\boldsymbol{w}^T\boldsymbol{x} + b \geq 0 \quad for \;\; d_i \; = 1 \tag{3.8}$$

$$\boldsymbol{w}^T\boldsymbol{x} + b \leq 0 \quad for \;\; d_i = 0 \tag{3.9}$$

The goal of SVM is to distinguish best between the two classes (here 1 for FA and 0 for COF). To ensure the best result, the distance between the points closest to the separating hyperplane must be maximized. Figure 3.8 shows an example of binary data that are optimally separated into two separate classes, represented by 1 and -1 (Mercer et al. 2008).

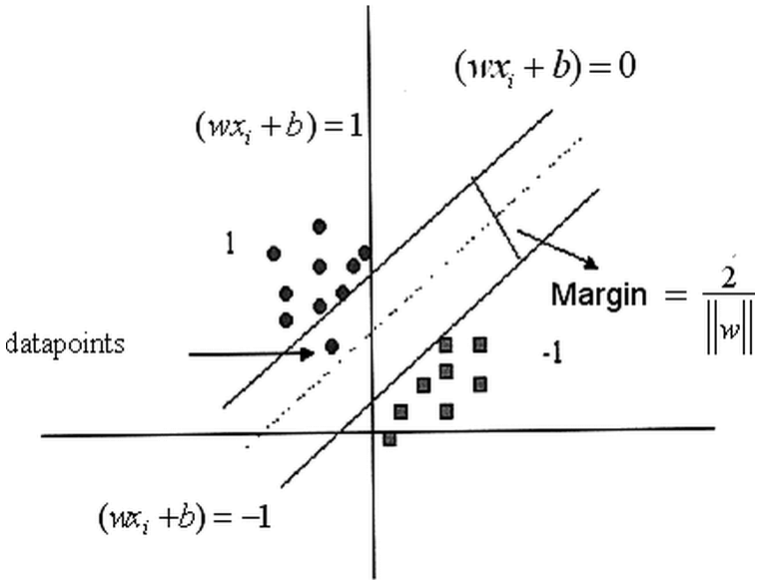


Figure 3.8 A representation of an SVM model where the dashed line is the best solution.

Many datasets that utilize SVMs are not linearly separable. In cases such as these (i.e., this present study), the data are input into a kernel function so a separating hyperplane can be found (Cristianini and Shawe-Taylor 2000, Schölkopf and Smola 2002). The kernel function projects the data into a nonlinear hyperspace where they may retain this nonlinear separability. For this study, the SVM was tuned using the following kernel functions:

1. Gaussian radial basis kernel

$$K(\boldsymbol{x}, x_i) = e^{\left(-\frac{1}{2\sigma^2}\|\boldsymbol{x}-x_i\|^2\right)} \tag{3.10}$$

2. Polynomial kernel

$$K(\boldsymbol{x}, x_i) = (\boldsymbol{x}^T x_i)^p \tag{3.11}$$

In addition to kernel functions, the SVM also requires the cost coefficient C and the associated parameters of the kernel functions, which are tuned through cross validation. This cross validation was conducted by withholding 80% of the dataset for training and using the remaining 20% for subsequent validation. This was repeated 1000 times using a bootstrapping method (i.e., random sampling of training/testing) to obtain confidence intervals on the performance statistics. Numerous kernel functions, configurations, and cost coefficients were tested using this cross-validation dataset to determine the optimal values of these SVM parameters for our dataset to improve the classification skill. The greater values of cost heighten the influence of non-separable points and decrease the complexity of the problem. Higher degree polynomial kernels allow a more flexible decision boundary. Gamma regulates the impact of new features. The lower the gamma value, the less influence the new features will have on the

decision boundary. The tuning parameters for the Gaussian radial basis kernel and polynomial kernel included (Adriano et al. 2009):

- Cost: 10^0 - 10^3 (by factors of 10)
- Gamma: 0.01, 0.05, 0.1, 0.2, 0.5
- Degrees: 1 - 5 (for polynomial kernel)

To evaluate the performance of the classification, contingency statistics were calculated (Wilks 2019) on the results of both the logistic regression and the SVM. The contingency statistics require the creation of a contingency table (Table 3.4).

Table 3.3 Contingency table, where a represents the number of correctly classified yes's (1s, FAs), b represents predicting yes (FA) but observing no (COF), c represents predicting no (COF) but observing yes (FA), and d represents the number of correctly classified no's (0s, COFs).

	Obs	
Forecast	Yes (1, FA)	No (0, COF)
Yes (1, FA)	a	b
No (0, COF)	c	d

Four contingency statistics are then computed from the contingency table. Hit rate (HR), also known as the probability of detection (POD), is given as:

$$POD = \frac{a}{a + c} \tag{3.12}$$

This statistic is the ratio of correct yes (FA) forecasts to the total number of yes (FA) observations. Higher values indicate better classification. The false alarm ratio (FAR) is given as:

$$FAR = \frac{b}{a + b} \tag{3.13}$$

This statistic is the number of misclassified COFs versus the total number of times a yes (FA) is predicted. A lower value is indicative of lower false alarms, which is desirable. Heidke skill score (HSS) is given as:

$$HSS = \frac{2(ad - bc)}{(a + c)(c + d) + (a + b)(b + d)} \tag{3.14}$$

HSS is the measure of success of the forecast relative to what it would be by chance without the underlying assumption of the distributions being the same. Values close to 1 are better. The best SVM configuration was found using the median bootstrap replicate HSS values of each configuration, as HSS is an objective measure of discrimination and increases with increasing POD concurrent with decreasing FAR. Doswell et al. (1990) demonstrated that the HSS was superior to the critical success index (CSI) for evaluating forecasts of rare events because it gave credit for a correct forecast of a nonevent. The final contingency statistic considered herein is bias (BIAS):

$$BIAS = \frac{a + b}{a + c} \tag{3.15}$$

which is the ratio of the number of yes (FA) predictions to the number of no (COF) predictions. An unbiased forecast has a bias value of 1, while values less than 1 predict too many 0s (COFs) and values greater than 1 predict too many 1s (FAs).

Finally, these contingency statistics were used to assess the performance of the feature selection methodologies and ultimately the logistic regression and SVM models, determine the optimal configuration of the SVM, and evaluate the performance of the optimal model configurations on the 2020 case data.

Feature Selection Methodologies

Of the three feature selection methodologies employed (i.e., stepwise, permutation testing, and PCA), stepwise proved to yield the best predictor set and therefore the better logistic and SVM model results. Continency statistics were computed for both the SVM and logistic models after each feature selection method was utilized (Tables 3.5, 3.6, and 3.7). These results show that the HSS values for the logistic and SVM models utilizing the stepwise predictors had the highest values when compared to the best performing versions using predictors from PCA and permutation testing. However, the SVM HSS value of the best performing permutation test was close to that of the SVM which utilized predictors reduced via stepwise.

Table 3.4 Model results utilizing the 10 optimal predictors that were computed using stepwise.

Logistic Regression	
HSS	0.353
BIAS	1.063
POD	0.684
FAR	0.357
SVM	
HSS	0.323
BIAS	1.133
POD	0.667
FAR	0.389

Table 3.5 Model results utilizing PCA with different numbers of kept PCs. Variance explained is also included.

Logistic Regression						
	3 PCs Kept	**4 PCs Kept**	**5 PCs Kept**	**6 PCs Kept**	**7 PCs Kept**	**8 PCs Kept**
HSS	0.250	0.240	0.248	0.239	0.255	0.248
BIAS	1.061	1.071	1.177	1.154	1.156	1.133
POD	0.619	0.625	0.667	0.667	0.667	0.667
FAR	0.412	0.421	0.429	0.429	0.417	0.421
SVM						
	3 PCs Kept	**4 PCs Kept**	**5 PCs Kept**	**6 PCs Kept**	**7 PCs Kept**	**8 PCs Kept**
HSS	0.197	0.197	0.234	0.180	0.202	0.202
BIAS	1.143, 1.154	1.056	1.118	1.231	1.222	1.235
POD	0.632	0.600	0.625	0.667	0.647	0.667
FAR	0.000, 0.000	0.000	0.421	0.000	0.444	0.447
Variance Explained						
	3 PCs Kept	**4 PCs Kept**	**5 PCs Kept**	**6 PCs Kept**	**7 PCs Kept**	**8 PCs Kept**
	0.522	0.580	0.629	0.671	0.701	0.731

Table 3.6 Model results utilizing permutation testing of the 10 optimal predictors.

Logistic Regression				
	$p < 0.001$	**$p < 0.01$**	**$p < 0.025$**	**$p < 0.05$**
HSS	0.205	0.198	0.138	0.138
BIAS	1.000	1.053	1.053	1.000
POD	0.583	0.583	0.546	0.539
FAR	0.429	0.436	0.471	0.469
SVM				
	$p < 0.001$	**$p < 0.01$**	**$p < 0.025$**	**$p < 0.05$**
HSS	0.308	0.281	0.271	0.264
BIAS	1.053	1.063	1.077	1.077
POD	0.611	0.611	0.625	0.625
FAR	0.375	0.389	0.400	0.409

It is not a surprise that models utilizing predictors chosen by stepwise regression performed well as stepwise regression can handle significant amounts of potential predictors and

select the top predictor variables from the available options. As for PCA, it should be used mainly for variables that are correlated and in linear modeling contexts. In other words, it forces a linear relationship, so a nonlinear modeling approach won't work well with linearly separable data. If there is weak correlation between variables, PCA does not work well to reduce data. Correlation values always range from -1 for a negative relationship to 1 for a positive correlation. Values at, or close to, zero indicate no linear relationship or a very weak correlation. The correlation matrix values in this case determined that most of the correlation coefficients are smaller than 0.5 (Figure 3.7). This shows why PCA may not have been the most helpful in reducing the number of predictors and producing a successful SVM and logistic model.

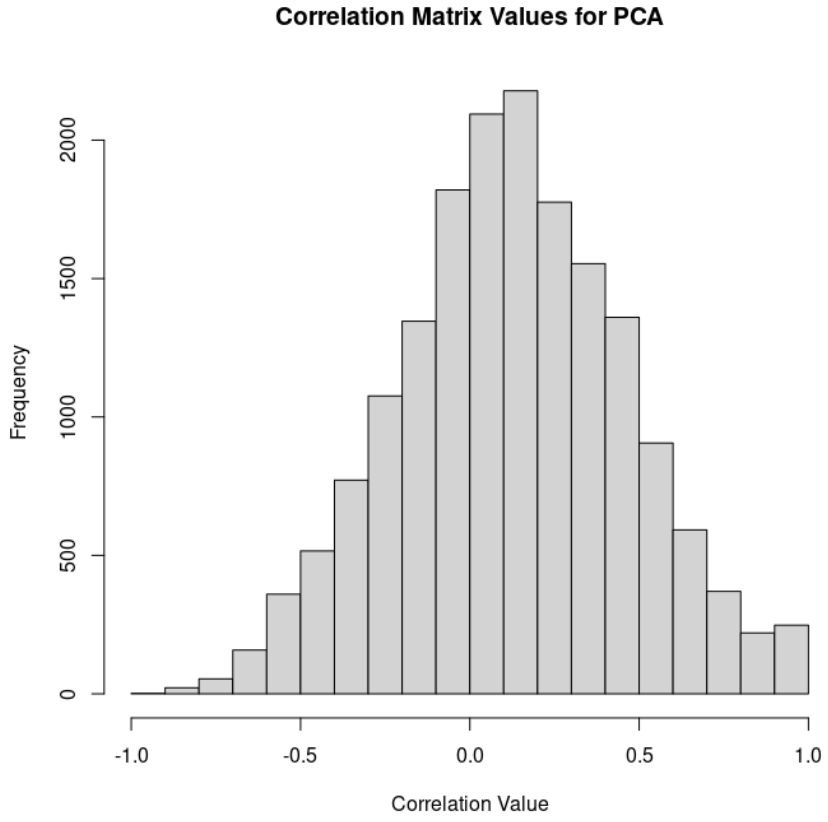


Figure 3.9 Histogram of the best performing PCA correlation matrix values.

Lastly, for permutation testing, the null hypothesis is more specific. You must assume that the samples used for testing have identical distributions (i.e., shape, center, and spread). Permutation testing presumes that the only difference between the samples is the random assignment. A reason why the best performing SVM utilizing predictors from permutation testing is so close to that of the SVM utilizing predictors from stepwise may be due to the small sample size. The best performing model with permutation testing predictors was the model utilizing p-values less than 0.001. Only 34 predictors had p values less than 0.001. Permutation tests are also effective when parametric assumptions are not met. As permutations only require exchangeability, they are a very robust test.

Optimal SVM Configuration

As stated previously, the stepwise feature selection method proved to yield the best predictor set and therefore the better logistic and SVM model results. Because of this, confidence intervals were completed for each contingency statistic on the versions of both the logistic regression and the SVM (Figures 3.10a-d) that utilize the predictors chosen by the stepwise selection. Regarding the comparison of models that has been done in this study, it has been found that the SVM and the logistic regression have close overall performance measures. The contingency statistics show that the two models are statistically indistinguishable. While the overall numbers for logistic were slightly better, the uncertainty in both models suggests that either method yields similar predictability, meaning linear separability works decently for this problem. The SVM parameters for the top ten configurations are given in Table 3.8.

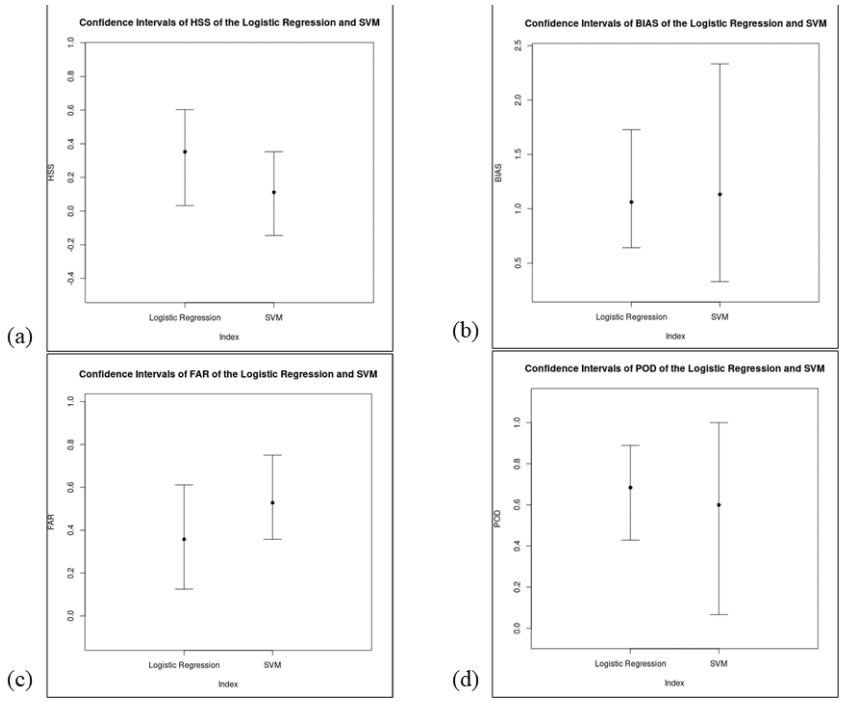


Figure 3.10 Confidence intervals on the contingency statistics of the best performing versions of the logistic regression and SVM models ($\alpha = 0.05$), including: (a) HSS, (b) BIAS, (c) FAR, and (d) POD.

Table 3.7 Top ten performing SVM configurations in order of decreasing HSS.

HSS	0.323	0.320	0.317	0.316	0.316	0.316	0.316	0.316	0.316	0.315
Kernel	Poly	Poly	Poly	Poly	Poly	Poly	Poly	Poly	Poly	Poly
Degree	1	1	1	1	1	1	1	1	1	1
Gamma	0.5	0.05	0.5	0.2	0.01	0.01	0.1	0.2	0.5	0.5
Cost	1	10	10	1	100	1000	1000	1000	1000	100

As stated above, a kernel function is regularly applied in SVMs if the data are not linearly separable (which was hypothesized here), as the kernel can occasionally find a decision hyperplane that separates the classes better than a linear classifier. The SVM performance must also be optimized based on the tuning parameter listed above. The optimal configuration for this study revealed a linear polynomial kernel, which yields a solution very similar to a logistic

regression (a linear separability within the data). Further, the lower cost suggests points far from the decision boundary should be used in its weighting, which ultimately suggests the SVM, despite being linear, is a complex classifier for this problem. The gamma value also increases the influence of additional features on the hyperplane boundary, which is likely the reason for the small subset of features performing best (as seen in the feature selection discussion above).

CHAPTER IV

RESULTS & DISCUSSION

Interpretation of Predictors

The predictors chosen by the stepwise model (Table 4.1) showed that relatively greater discrimination between outbreak types occurs with thermodynamic variables rather than kinematic. This is consistent with the findings from studies like Shafer et al. (2010) and Grams et al. (2012). Shafer et al. (2010) found that storm-relative helicity parameters supplied better discrimination power between WRF model composites of tornado outbreaks and primarily nontornadic outbreaks, compared to CAPE. Grams et al. (2012) found that convective mode, composite parameters, and kinematic variables (i.e., 0-1 km and 0-6km bulk wind difference) provided greater discrimination between tornadic events than thermodynamic variables (i.e., mixed layer CAPE, mixed layer CIN, mixed layer LCL).

Table 4.1 The 10 optimal predictors utilized in the log regression and SVM models in order of increasing HSS values. The results are cumulative, such that HSS on the last row represents HSS when retaining all 10 predictors listed here.

Variable	Spatial Component	POD	FAR	BIAS	HSS
Min Daily Temperature Advection (850 mb) (K s^{-1})	Max	0.612	0.402	1.024	0.268
Mean Daily Surface CIN (J kg^{-1})	Mean	0.659	0.341	1.000	0.375
Min Daily Vorticity Advection (500 mb) (s^{-2})	Mean	0.647	0.267	0.882	0.455
Max Daily Vorticity Advection (500 mb) (s^{-2})	Max	0.671	0.250	0.894	0.487
Min Daily Surface MUCAPE (J kg^{-1})	Mean	0.682	0.247	0.906	0.500
Mean Daily Vorticity Advection (500 mb) (s^{-2})	Max	0.694	0.244	0.918	0.511
Mean Daily SCP	Mean	0.706	0.241	0.929	0.523
Min Daily Surface CIN (J kg^{-1})	Max	0.694	0.213	0.882	0.543
Mean Daily Surface MUCAPE (J kg^{-1})	Mean	0.706	0.211	0.894	0.554
Min Daily Surface MUCIN (J kg^{-1})	Max	0.706	0.189	0.871	0.575

Nevertheless, Grams et al. (2012) does stress that certain thermodynamic parameters may provide relatively high levels of discrimination when comparing a significant tornado outbreak and a FA environment where significant severe storms did not occur. Mercer and Bates (2014) supported this result as well, showing warm air advection (WAA) was weaker with cold air advection comparatively stronger and more widespread in FA outbreak cases. This suggests that forecasters may have overlooked the lack of strong WAA in FA outbreak cases when considering other factors (e.g., kinematic variables) favorable for tornado outbreaks, leading to an incorrect outbreak expectation. Bootstrap confidence intervals on the mean temperature advection for both FAs and hits show that there is a significant difference between the two

outbreak types (Figure 4.1). The mean for COFs is statistically greater than that for FAs, signifying that past research such as Mercer and Bates (2014) was correct in its findings.

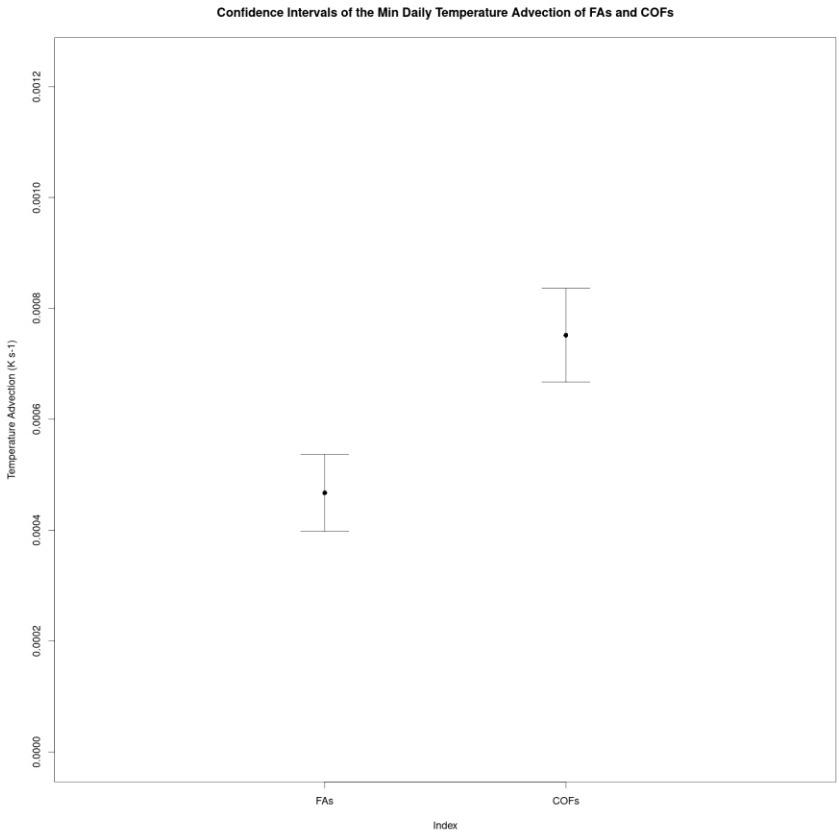


Figure 4.1 Confidence intervals on the means of the minimum daily temperature advection of FAs and COFs ($\alpha = 0.05$).

Most unstable CAPE (MUCAPE) is a measure of instability that indicates the amount of potential energy available to the most unstable parcel of air found within the lowest 300mb of the atmosphere while being lifted to the level of free convection (LFC). According to Dean and Schneider (2008), the probability of severe thunderstorms and tornadoes increased when MUCAPE and 0-6 km bulk shear values were greater. The relationship of these environments to the performance of watches issued by the SPC were then analyzed to find that, the good area

percentage (the percentage of tornadoes in the issued watch versus outside of the watch) for tornadoes in tornado watches had a propensity to decrease as MUCAPE values fell below 1000 J kg-1 (Dean and Schneider 2008). This suggests forecasters do a better job of forecasting when MUCAPE values are high and larger uncertainty exists for tornado development in low CAPE environments, which are more common in the cool season. This implies that FA forecasts may have commonly occurred due to the interpretation of levels of MUCAPE during cooler seasons. Weaker tornadoes are common in low-CAPE environments during these cooler seasons across the Southeast United States (Childs et al. 2018). During meteorological fall and winter, 22 FAs occurred during this study period, with the FA outbreaks having lower average MUCAPE values than the COFs. Bootstrap confidence intervals on the mean MUCAPE for both FAs and COFs show there is not a statistical difference between cases (Figures 4.2 and 4.3).

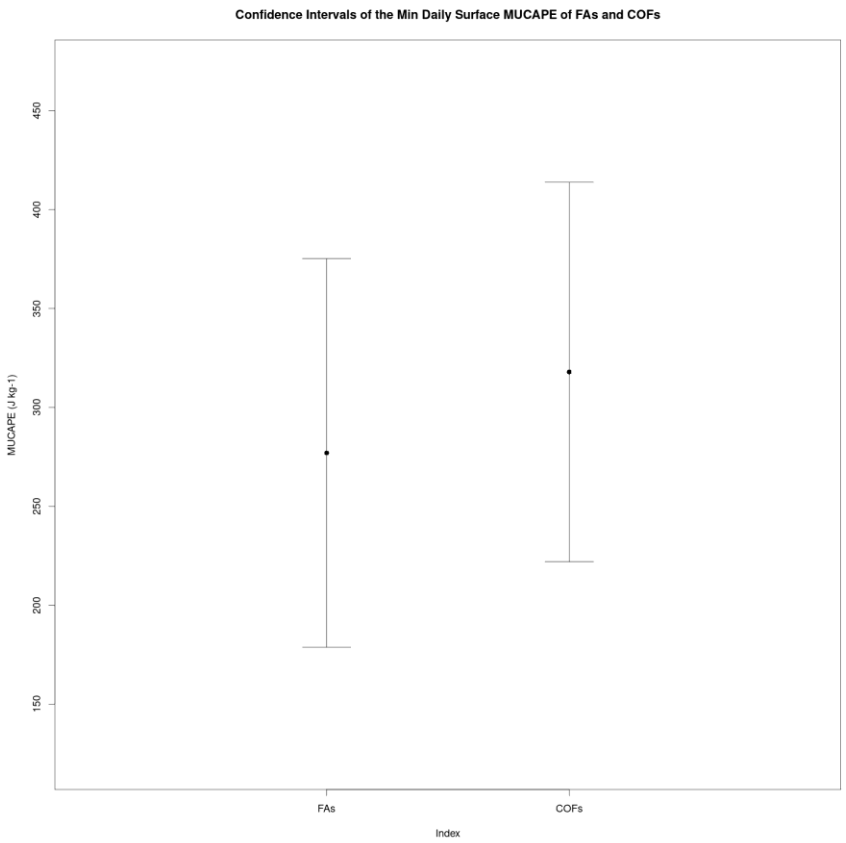


Figure 4.2 Confidence intervals on the means of the minimum daily surface MUCAPE of FAs and COFs ($\alpha = 0.05$).

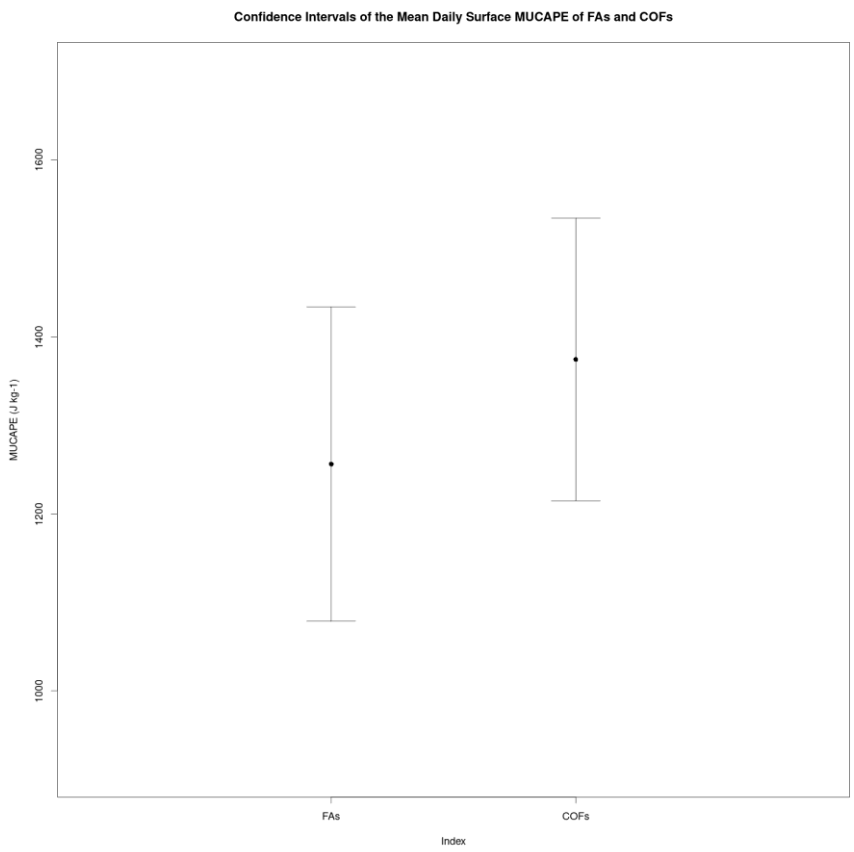


Figure 4.3 Confidence intervals on the means of the mean daily surface MUCAPE of FAs and COFs ($\alpha = 0.05$).

In addition to the apparent importance of CAPE from the feature selection methodology, CIN also was revealed as an important distinguishing characteristic of FA environments. While CIN, in general, is thought to inhibit convective activity and thus reduce the probability of severe storms, in some cases, enhanced moisture and/or diurnal heating overcome the CIN, and result in a greater than baseline probability of a storm reaching severe strength (Davies 2004). Thus, in a FA context, elevated CIN coupled with marginal synoptic-scale vertical forcing may have led a forecaster to anticipate convection that never materialized (hence a FA). According to the confidence intervals completed on the means of the FAs and COFs for each of the CIN

predictors, those intervals overlap, which concludes that the difference in means between groups is not statistically significant (Figures 4.4 – 4.6). Since there is no significant difference in the average values between outbreak types, this confirms that CIN values may have led a forecaster to predict convective storms that never occurred.

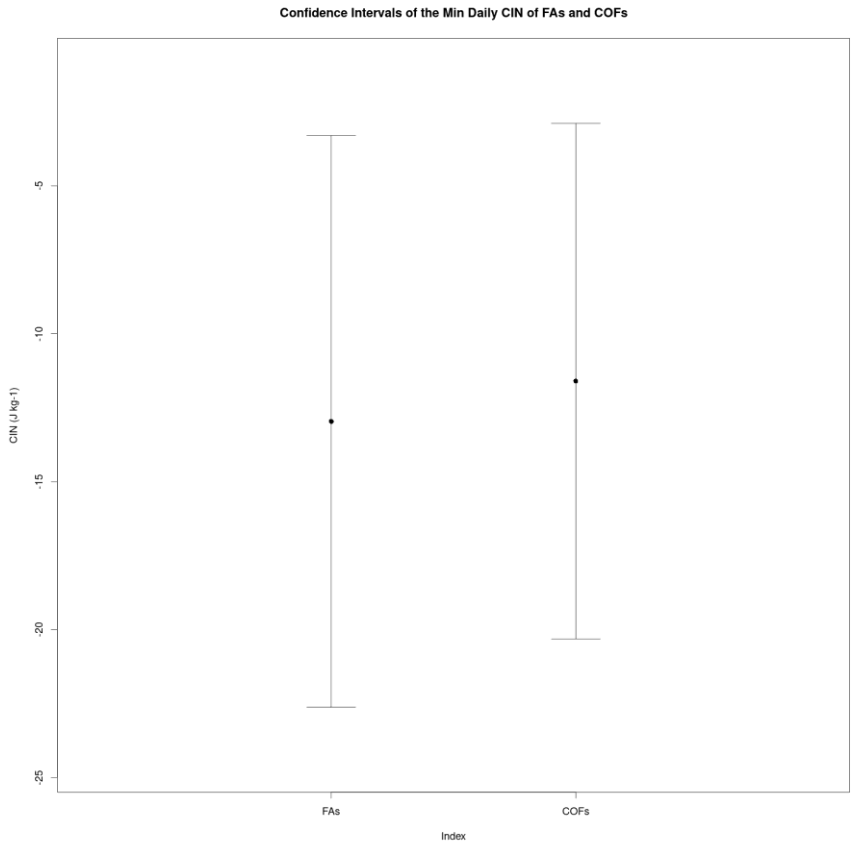


Figure 4.4 Confidence intervals on the means of the minimum daily CIN of FAs and COFs ($\alpha = 0.05$).

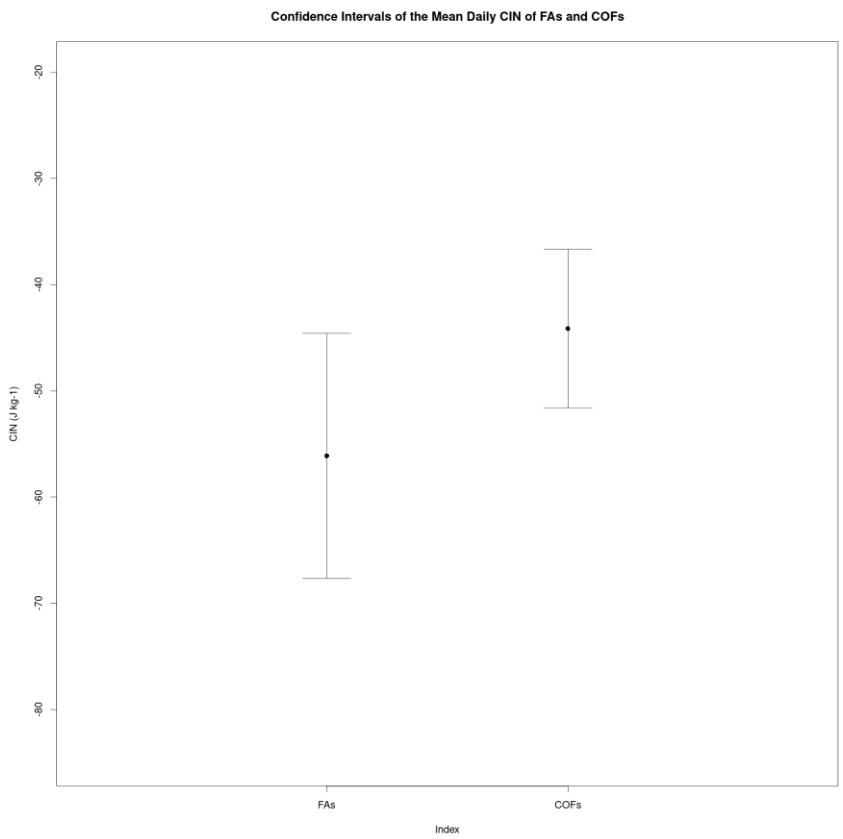


Figure 4.5 Confidence intervals on the means of the mean daily CIN of FAs and COFs (α = 0.05).

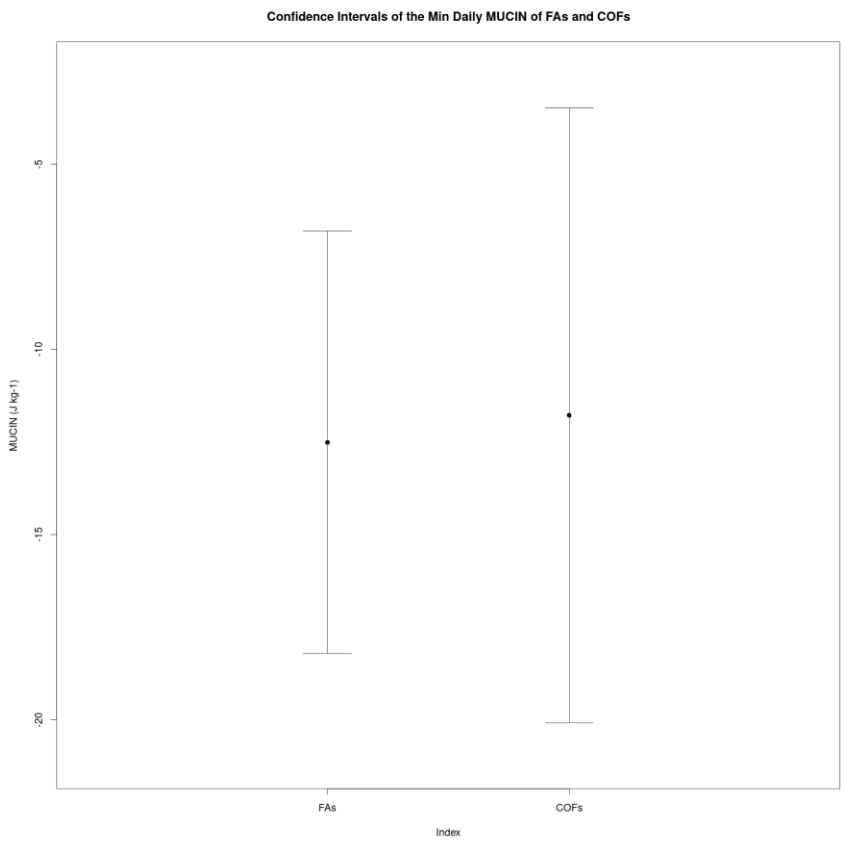


Figure 4.6 Confidence intervals on the means of the minimum daily MUCIN of FAs and COFs ($\alpha = 0.05$).

The supercell composite parameter (SCP) is an index that includes MUCAPE, effective storm relative helicity, and effective bulk wind difference (Thompson et al. 2004):

$$SCP = \left(\frac{MUCAPE}{1000\ Jkg^{-1}}\right) \times \left(\frac{ESRH}{50\ m^2s^{-2}}\right) \times \left(\frac{EBWD}{10\ ms^{-1}}\right) \quad (4.1)$$

While each of the optimal predictors so far have been thermodynamic in nature, this composite index includes both thermodynamic (i.e., MUCAPE) and kinematic variables (i.e., ESRH and EBWD). As such, it makes sense as to why this predictor had a higher HSS value and was therefore one of the better predictors in differentiating between outbreak type. The three parameters embody combinations of thermodynamic and vertical wind shear potential to best

separate supercells from non-supercells. Each of the three ingredients are normalized to supercell threshold values following the work done by Thompson et al. (2003) and Thompson et al. (2007), where larger values of SCP indicate greater overlap of the three ingredients. Based on past research by Thompson et al. (2003), SCP has a strong ability to distinguish between tornadic and other supercell groups, where the differences in the means were statistically significant across all categories. According to Krocak et al. (2021), as SCP increased across a range of more frequently observed values (e.g., 0–10), the percentage of warned tornadoes increased. The same can be said for Thompson et al. (2012), where larger values of the effective-layer SCP more often saw right-moving supercells that produced significant tornadoes, and smaller values of SCP saw weaker tornadoes or supercells that generated hail and/or damaging winds. Average SCP values were slightly higher for the COF cases than for FAs; however, when comparing the confidence intervals for those groups, those intervals overlap (Figure 4.7). This concludes that the difference in means between groups is not statistically significant.

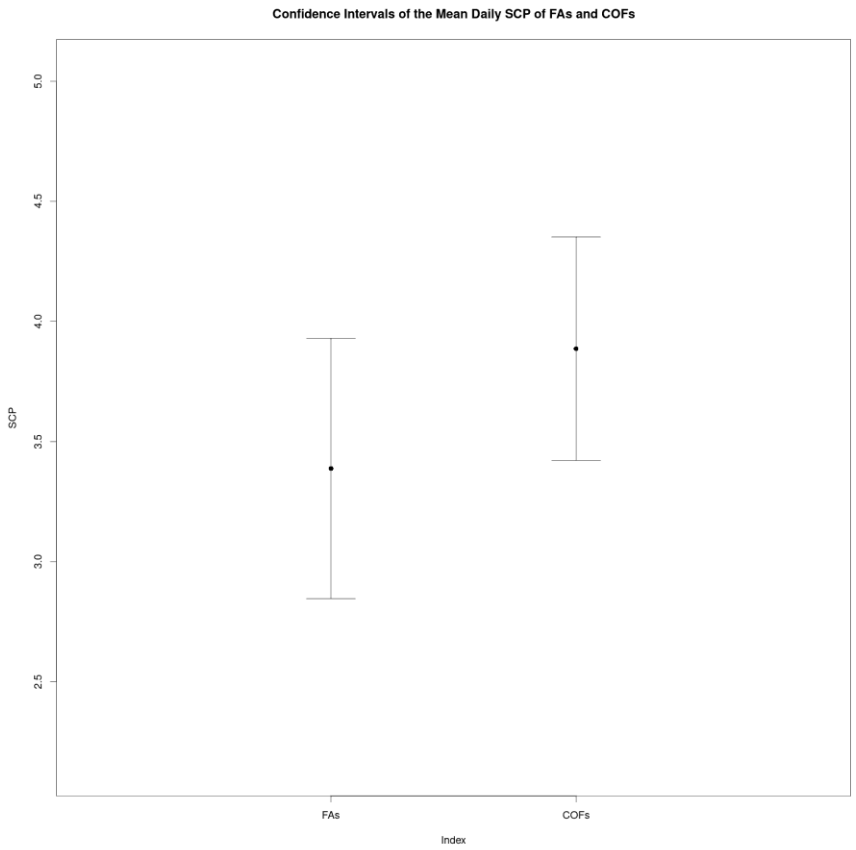


Figure 4.7 Confidence intervals on the means of the mean daily SCP of FAs and COFs ($\alpha = 0.05$).

The final optimal predictor is vorticity advection, which was the only individual kinematic variable (though it is not a typical severe weather diagnostic variable). In past research, such as Mercer and Bates (2014), the FA group has the weakest relative vorticity maxima, which in turn resulted in smaller differential positive vorticity advection (DPVA), an indicator of mid-level rising motion. In Mercer and Bates (2014), FAs had the weakest DPVA six hours before the outbreak. This was to be expected, as nontornadic outbreaks are known to have weaker DPVA than tornado outbreaks (Mercer et al. 2012). In this study, FAs had weaker vorticity advection than the COF cases (Figures 4.8 – 4.10) apart from the maximum daily

vorticity advection, which had a slightly higher mean. Vorticity advection helps diagnose areas of rising or sinking motion. Positive vorticity advection (PVA) at 500 mb implies divergence is occurring in the upper troposphere, and therefore, rising motion is occurring. Areas of PVA at 500 mb suggest that DPVA is also occurring, meaning positive vorticity increasing from the surface to the upper levels is being advected causing the troposphere to become more dynamically unstable. PVA values are seen for both FAs and COFs for maximum daily vorticity advection and mean daily vorticity advection (Figures 4.9 and 4.10, respectfully). Oppositely, negative vorticity advection (NVA) values are seen for minimum daily vorticity advection (Figure 4.8) for both FAs and COFs, meaning convergence is occurring in the upper troposphere contributing to sinking air. Thus, precipitation and severe weather will be less likely to occur. It can be inferred that differential negative vorticity advection (DNVA) is also occurring, meaning negative vorticity increasing from the surface to the upper levels is being advected causing the troposphere to become more stable. NVA normally occurs upstream of an area of maximum vorticity. Figure 4.8 shows NVA values are smaller for FAs than for COFs. This could signify the potential for a ridge building in robustly in FAs than COFs.

As shown by the quasigeostrophic omega equation (Equation 4.2):

$$\left[\nabla_p{}^2 + \frac{f_o{}^2}{\sigma}\frac{\partial^2}{\partial p^2}\right]\omega = -\frac{f_o}{\sigma}\frac{\partial}{\partial p}\left[-\vec{V}_g \cdot \nabla_p\left(\zeta_g + f\right)\right] - \frac{R}{\sigma p}\nabla_p{}^2\left[-\vec{V}_g \cdot \nabla_p T\right] \tag{4.2}$$

both DPVA and WAA can provide uplift when working in tandem. In this study, temperature advection has a lower HSS value than DPVA, so it provides the most significant distinction between cases. Therefore, it is possible it could have been the main cyclogenetic factor in these cases compared to DPVA (Table 4.1). This is consistent with the findings in Mercer and Bates (2014).

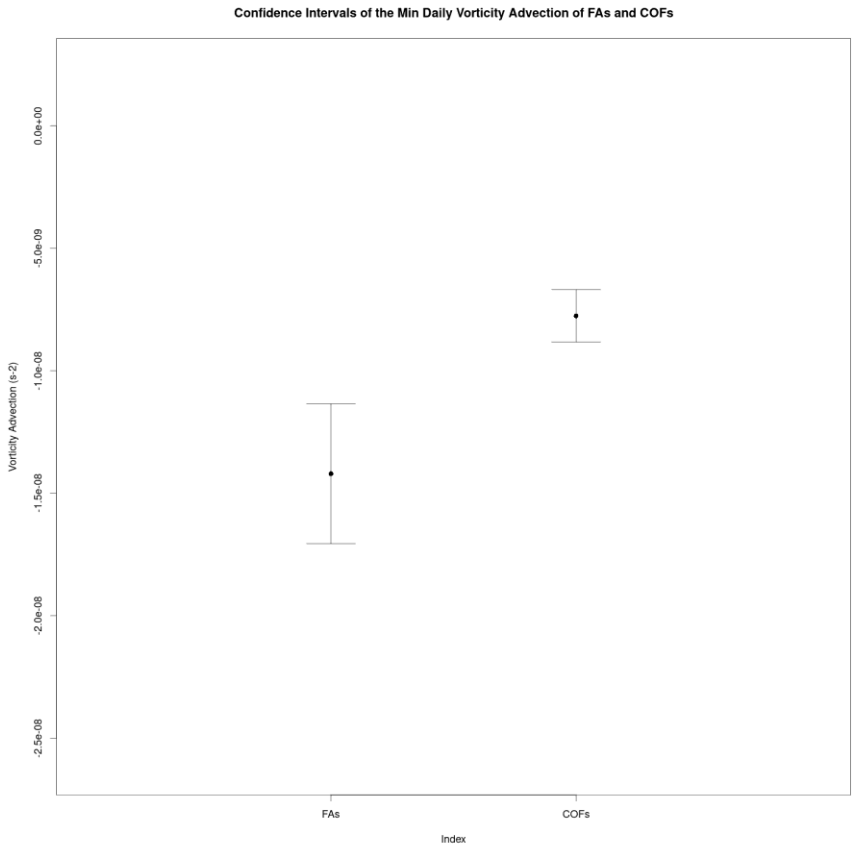


Figure 4.8 Confidence intervals on the means of the minimum daily vorticity advection of FAs and COFs ($\alpha = 0.05$).

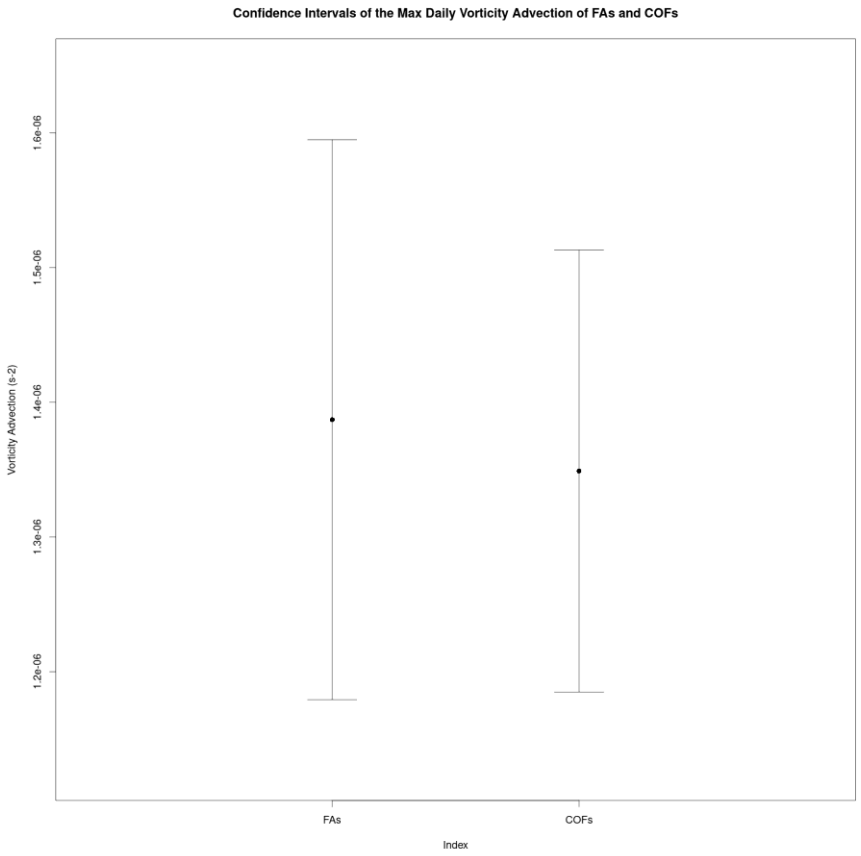


Figure 4.9 Confidence intervals on the means of the maximum daily vorticity advection of FAs and COFs ($\alpha = 0.05$).

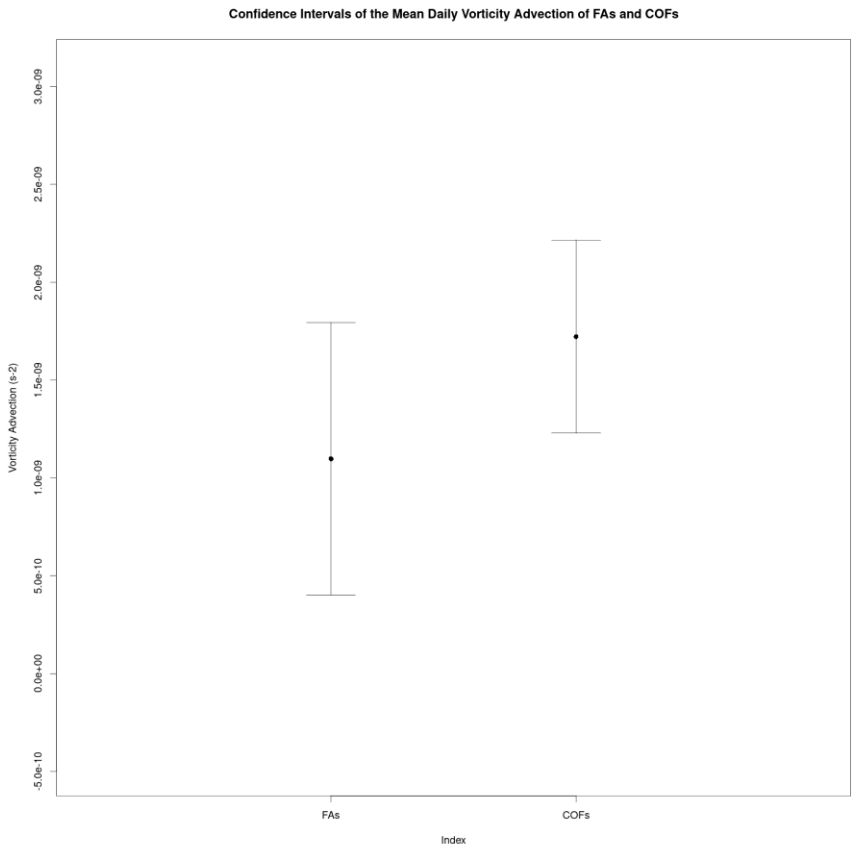


Figure 4.10 Confidence intervals on the means of the mean daily vorticity advection of FAs and COFs ($\alpha = 0.05$).

Performance of the SVM vs. the Logistic Regression

There are some possible reasons as to why the performance of the logistic regression and the SVM are not significantly different. Logistic regression and SVMs are similar. Logistic regression aims at maximizing the probability of the data. The farther the data lies from the hyperplane on the proper side, the better the logistic regression performs. An SVM attempts to find the hyperplane that maximizes the distance of the support vectors, and this optimization theoretically minimizes classification error in the given statistical space. The linear kernel was the optimal configuration of the SVM, and logistic regression performs well when the data is

linearly separable. This reinforces why the results are similar. The risk of overfitting is also less in SVM, while logistic regression is vulnerable to overfitting. This is particularly true in cases with many predictors, such as with this study. While a stepwise feature selection methodology was utilized to lessen the number of predictors to those that are most optimal, it is possible that the number was still too much and therefore caused some overfitting to occur leading to a slightly better performance by the logistic regression.

Testing of the 2020 Data

After the optimization and training phases were completed, both the SVM and logistic classifiers were tested on the 2020 outbreak case set (Table 4.2). Compared to both model performances in the optimization and training phase with the 2010-2019 data, the testing of model performance on the 2020 case data yielded HSS values which increased by approximately 0.249 between the logistic regression models and 0.279 between the SVM models. Again, there is no notable benefit in the SVM compared to the logistic regression as the results from testing were identical for both models.

Table 4.2 Contingency statistics of the 2020 data using best configuration of the logistic regression and SVM models.

Logistic Regression & SVM	
HSS	0. 602
BIAS	0.875
POD	0.750
FAR	0. 143

Both the logistic regression and the SVM models misclassified the same cases. Of the 15 case dates in the 2020 testing data, six of the seven COFs and six of the eight FAs were classified correctly by both the logistic regression and the SVM models. Two cases were predicted to be

COFs but were actually FAs (December 24, 2020, December 31, 2020), and one case was forecasted as a FA but was actually a COF (April 13, 2020; Table 4.3).

Table 4.3 Contingency table of the 2020 testing data.

	Obs	
Forecast	Yes (1, FA)	No (0, COF)
Yes (1, FA)	6	1
No (0, COF)	2	6

The April 13, 2020 case had a minimum temperature advection value that fell within the confidence interval of FAs for that variable (Table 4.4). The minimum vorticity advection value was also within the confidence interval for FAs for that variable. The values of the other eight variables did not fall within their respective confidence intervals for either FAs or COFs but were on the lower end nearer to the FA range. This may be why the logistic regression and the SVM misclassified this case as a FA. The Day One Convective Outlook at 1200 UTC on this day stated that warm, moist air from the Gulf of Mexico strengthened an already extensive air mass at the surface with warm, high dew point air over the southeast United States. However, robust heating was likely hindered by prevalent cloudiness ahead of severe thunderstorms that were occurring that morning. Nevertheless, forecasters correctly predicted narrow areas of moderate instability that were conducive for tornadic storms in addition to favorable kinematic conditions.

For the December 24, 2020 case, the minimum temperature advection value falls within the confidence intervals of that variable for FAs while minimum vorticity advection is close to the confidence intervals of COFs (Table 4.4). The values of the other eight variables did not fall within their respective confidence intervals for either FAs or COFs. According to the Day One Convective Outlook at 1200 UTC for this case, forecasters mentioned the existence of mid-level height falls. Positive vorticity advection (PVA) is accompanied by height falls as it promotes

upper-level divergence and rising air. The convective outlook also stated that pre-frontal WAA and the advancement of the warm sector would allow for reasonable destabilization that could sustain strong updrafts in advance of the cold front. Forecast soundings also depicted MUCAPE values greater than 1000 J/kg. It is likely that forecasters as well as the logistic and SVM models found sufficient amounts of each of the previously mentioned variables to constitute COF conditions. Four tornadoes did occur within the 10% tornado probability region in this case, which is less than the established global mean of six. These correlated with the tornado threat that was noted to occur in the convective outlook with embedded circulations along the frontal squall line. It is likely that while the variables existed for tornadoes to occur, the squall line passed quickly leading to conditions that were not robust enough to warrant a COF outcome; therefore, the case was misclassified as a COF.

As for December 31, 2020, minimum temperature advection, minimum vorticity advection, and mean vorticity advection values fall within the confidence intervals for COFs of the respective variables (Table 4.4). The values of the other seven variables did not fall within their respective confidence intervals for either FAs or COFs. When referencing the 1200 UTC Day One Convective Outlook on this case day, forecasters highlighted that extensive clouds and feeble lapse rates were only conducive to a moderate thermodynamic environment for severe storms. However, they did mention showers and thunderstorms were expected to be more prevalent throughout the day and develop northward due to warm air advection. A majority of the outlook discussed favorable kinematic variables, such as storm relative helicity and effective bulk shear. Their analysis concluded that most of the storms that could develop would likely be slightly elevated (which are associated with large hail and strong, damaging winds) but that the overall environment would support a tornado threat if decent updrafts could occur in the

boundary layer. It is likely that while the previously mentioned variables fell within COF constraints, conditions in the boundary layer were not conducive to support sufficient updrafts to form an outbreak of tornadoes. Therefore, the case was misclassified by both models as a COF.

Table 4.4 Values of the optimal predictors for the misclassified cases.

Variable	**Spatial Component**	**April 13, 2020 (misclassified as FA)**	**December 24, 2020 (misclassified as outbreak)**	**December 31, 2020 (misclassified as outbreak)**
Min Daily Temperature Advection (850 mb) (K s^{-1})	Max	3.840e-04	3.813e-04	7.586e-04
Mean Daily Surface CIN (J kg^{-1})	Mean	-24.816	-4.628	-16.208
Min Daily Vorticity Advection (500 mb) (s^{-2})	Mean	-1.364e-08	-4.157e-09	-5.759e-09
Max Daily Vorticity Advection (500 mb) (s^{-2})	Max	2.595e-06	9.596e-07	2.131e-06
Min Daily Surface MUCAPE (J kg^{-1})	Mean	6.238	0.455	3.860
Mean Daily Vorticity Advection (500 mb) (s^{-2})	Max	1.149e-11	3.588e-09	2.237e-09
Mean Daily SCP	Mean	1.266	1.543	1.791
Min Daily Surface CIN (J kg^{-1})	Max	0.000	0.000	0.000
Mean Daily Surface MUCAPE (J kg^{-1})	Mean	356.561	222.396	358.927
Min Daily Surface MUCIN (J kg^{-1})	Max	0.000	0.000	0.000

An example of a FA case that the models correctly classified is March 19, 2020. Eight tornadoes occurred on this date. However, only two tornadoes occurred within the 10% tornado probability percentage region, hence the FA categorization. Minimum daily temperature

advection and minimum and maximum daily vorticity advection values were within the range of confidence intervals for FAs (Table 4.5). The SPC's 1200 UTC Day One Convective Outlook stated that significant convection was already occurring in the southern Plains which was southwest of the region of interest. This activity was forecasted to propagate into eastern Oklahoma and western Missouri by sunrise and into the Ohio Valley later that day. There was uncertainty stated regarding the extent of the severe potential and how severe this activity would be. However, forecasters were confident that strong shear and adequate moisture/buoyancy would lead to damaging winds and some tornado threat by that afternoon as the strongest areas of forcing moved into the region of interest. Lapse rates were forecasted to not be steep, but strong shear and higher precipitable water content were thought to be ample enough for damaging winds and a tornado threat through a linear storm mode with embedded supercells. It is probable that the deeply buoyant profiles that forecasters anticipated were inhibited by the milder lapse rates aloft, causing a lower probability of tornadic storms.

Table 4.5 Values of the optimal predictors for 19 March 2020.

Variable	**Spatial Component**	**Value**
Min Daily Temperature Advection (850 mb) (K s^{-1})	Max	4.212e-04
Mean Daily Surface CIN (J kg^{-1})	Mean	-15.796
Min Daily Vorticity Advection (500 mb) (s^{-2})	Mean	-1.004e-08
Max Daily Vorticity Advection (500 mb) (s^{-2})	Max	1.595e-06
Min Daily Surface MUCAPE (J kg^{-1})	Mean	1.799
Mean Daily Vorticity Advection (500 mb) (s^{-2})	Max	3.856e-09
Mean Daily SCP	Mean	1.531
Min Daily Surface CIN (J kg^{-1})	Max	0.000
Mean Daily Surface MUCAPE (J kg^{-1})	Mean	524.690
Min Daily Surface MUCIN (J kg^{-1})	Max	0.000

CHAPTER V

SUMMARY & CONCLUSIONS

Machine learning techniques were utilized to predict the probability that an SPC tornado outbreak forecast would result in an FA. Based on a global mean of 5.630 tornadoes per outbreak spanning 2010 – 2020, an overall mean of six was chosen to discern between COF and FA outbreak forecasts. This definition was applied to develop the COF and FA databases of cases using historical convective outlooks from the SPC that included a 10% tornado probability region. WRF model simulations were then completed for a set of 202 outbreaks (93 FAs and 109 COFs) to illustrate the meteorological environment producing each outbreak. Thermodynamic and kinematic variables from the WRF simulations were then used to train both a logistic regression model and an SVM model that forecast the probability of a FA.

Apart from two predictors, all the predictors were thermodynamic. Past research (Grams et al. 2012) of tornadic events and nontornadic hail events and wind events indicated the importance of kinematic variables (i.e., low-level and mid-level winds) to discriminate between these types of severe events. The stronger the low-level and mid-level winds, the occurrence of significant tornadic events was deemed more likely than significant nontornadic hail and nontornadic wind events. Wind shear aids in tilting the storm, allowing the updraft to be sustained for longer. This allows for the development of a mesocyclone. While moderate to large amounts of shear are needed in storms that produce hail, even greater wind shear can enhance rotation within the thunderstorm updraft, which can lead to tornado development. Wind shear for

nontornadic wind events is similar to nontornadic hail events. The stronger the low-level and mid-level winds, the likelihood for significant tornadic events increases versus significant nontornadic wind events. Thermodynamic variables were not of great importance in distinguishing between storm types in the Grams et al. (2012) study as kinematic variables were, as similar, large amounts of instability (i.e. CAPE) are needed for each storm type. These results correspond to Shafer et al. 2010 where kinematic variables (i.e., storm-relative helicity (SRH)) provided greater discrimination between tornado outbreaks and primarily nontornadic outbreaks.

These results show the importance of thermodynamic parameters when comparing a significant tornado event and a FA environment where significant tornadic storms did not occur where they were forecasted. Of those predictors that are thermodynamic, MUCAPE, surface CIN, and MUCIN remain largely unexplored in past research related to differentiating between storm or outbreak type. Expanding research on these three variables and how they impact outbreak environments and, therefore, outbreak type, would allow for the creation of machine learning models that more accurately represent the impact of thermodynamic variables and result in improved model performance.

The optimal configuration of the SVM and the logistic regression model were tested on case data from 2020. The initial optimization and training of the model yielded that the logistic regression performed slightly better than the SVM, though for the testing of the 2020 outbreak sets both methods performed identically. This study sought to verify whether an SVM could properly distinguish between COF and FA forecasts. An improved ability to identify false alarm outbreak environments did result from the work completed in this study as HSS values were greater than 0.

These conclusions are contingent upon the limitations associated with this study. The biggest drawback was the exclusion of hit and FA cases in April 2011. This period had date and time discrepancies in the data that compiled from the NCEP GFS Ensemble Forecast System archive yet was likely the most outbreak-active month of the study period. The exclusion of these cases may have played a small role on hindering model performance. The April 2011 "Super Outbreak" was the largest, costliest, and one of the deadliest tornado outbreaks ever recorded since formal record keeping began in 1950 (NWS 2011), and its inclusion would have been useful in this study (though it was well forecast and would have qualified as a COF).

Another limitation to this study was the use of a cumulus parameter in the WRF simulations. The use of 12-km grid spacing rather than the WRF-NSSL 4-km grid spacing inhibits the cumulus scheme as it cannot resolve CAPE at larger resolutions. Instability and convection are meteorological processes that are too small and complex to model at larger grid sizes, however smaller grid sizes can explicitly represent convective clouds, although they need to parameterize cloud microphysics which occur at a smaller scale. So, the WRF simulations that aided in building both the logistic regression model the SVM cannot fully simulate instability.

Additionally, had there been more time to complete this study, the use of a 30-year training period would have been ideal. The World Meteorological Organization recommends that each member country recompute their 30-year climate normals every 10 years (Arguez and Vose 2011), the most recent series of decadal normals being from 1991-2020. Adding more well-fit data to the analysis could assist in producing better accuracy and, therefore, a more trusted product. The use of other machine learning methodologies, such as Bayesian neural networks and random forests, would be ideal in this situation, as they can efficiently handle large amounts of data. They would also be helpful in training and testing the period of this current study.

Furthermore, it would be ideal to include a seasonality parameter when further developing the model. As previously mentioned, Mercer and Bates (2014) observed seasonal pattens of FAs in their study, where nearly all occurred in the summer and fall. Most tornadoes and tornado outbreaks happen throughout the spring, and this finding demonstrates that those that do not occur during the "regular season" have higher forecast uncertainty. This implies seasonality may be useful parameter to include when predicting FA forecasts. In the period of this study, most of the FAs occurred during the summer and fall months, which is in agreement with the results from Mercer and Bates (2014) (Figure 5.1). As previously explained, seasonal differences in CAPE may play a role, as values of CAPE are generally higher during the warm season and lower during the cool season. For that reason, events in the fall may require more synoptic-scale vertical forcing to initiate convection. Furthermore, favorable wind shear for supercell development and tornadoes is more common during the spring and fall than the summer. Lastly, LCL height depends on the season. Higher LCLs are typical with many primarily nontornadic outbreaks during the summer. It is likely that forecasting and interpretation of these variables led to higher instances of FAs during the summer and fall.

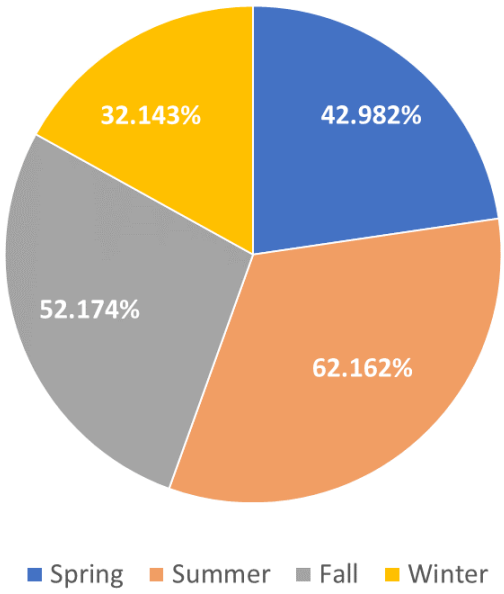


Figure 5.1 Percent of cases that were FAs from 2010-2020 by season.

There is still work to be done, but the work in this study is laying the groundwork for machine learning models to become more integral in tornado outbreak forecasting. Overall, models show promise in detecting FAs within outbreak forecasts, but more work is needed to ensure these benefits have operational value. A critical challenge encountered in this study is how little difference there was between the performance of the SVM model and the logistic regression. Future work will require updating the features for the SVM specifically (e.g., a new feature set for SVM versus the logistic features). Once more adjustments, enhancements and testing are done on the model to heighten its overall efficacy, the results could be shared with the SPC and National Weather Service weather forecasting offices to assist in their forecasting procedures. The final configuration will hopefully yield an algorithm that can be used to inform future tornado outbreak forecasts and lessen the burden of current challenges related to FAs.

REFERENCES

Adrianto, I., Trafalis, T. B., & Lakshmanan, V. (2009). Support vector machines for spatiotemporal tornado prediction. International Journal of General Systems, 38(7), 759-776.

Agee, E., C. Church, C. Morris, and J. Snow, 1975: Some synoptic aspects and dynamic features of vortices associated with the tornado outbreak of 3 April 1974. Mon. Wea. Rev., 103, 318–333.

Arguez, A., & Vose, R. S. (2011). The Definition of the Standard WMO Climate Normal: The Key to Deriving Alternative Climate Normals. Bulletin of the American Meteorological Society, 92(6), 699–704. http://www.jstor.org/stable/26218540

Barnes, L. R., Gruntfest, E. C., Hayden, M. H., Schultz, D. M., & Benight, C. (2007). False Alarms and Close Calls: A Conceptual Model of Warning Accuracy, Weather and Forecasting, 22(5), 1140-1147. Retrieved Oct 24, 2021, from https://journals.ametsoc.org/view/journals/wefo/22/5/waf1031_1.xml

Beatty, K. A. (2004). American Meteorological Society 22nd Conference on Severe Local Storms. In Potential Insurance Losses from a Major Tornado Outbreak: The 1974 Super Outbreak Example. Newark, CA; Risk Assessment Solutions Inc. Retrieved March 26, 2022, from https://ams.confex.com/ams/pdfpapers/81713.pdf.

Brotzge, J., Erickson, S., & Brooks, H. (2011). A 5-yr Climatology of Tornado False Alarms, Weather and Forecasting, 26(4), 534-544. Retrieved Oct 24, 2021, from https://journals.ametsoc.org/view/journals/wefo/26/4/waf-d-10-05004_1.xml

Chen, F., and Dudhia J. , 2001: Coupling an advanced land-surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model description and implementation. Mon. Wea. Rev., 129, 569–585.

Childs, S. J., Schumacher, R. S., & Allen, J. T. (2018). Cold-Season Tornadoes: Climatological and Meteorological Insights, Weather and Forecasting, 33(3), 671-691. Retrieved Feb 23, 2023, from https://journals.ametsoc.org/view/journals/wefo/33/3/waf-d-17-0120_1.xml

Colquhoun, J. R., and P. A. Riley, 1996: Relationships between tornado intensity and various wind and thermodynamic variables. Wea. Forecasting, 11, 360–371.

Corfidi, S., S. Weiss, J. Kain, S. Corfidi, R. Rabin, and J. Levit, 2010: Revisiting the 3–4 April 1974 Super Outbreak of tornadoes. Wea. Forecasting, 25, 465–510.

Cristianini, N., and J. Shawe-Taylor, 2000: Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press

Dean, A. R., & Schneider, R. S. (2008). 9A. 2 FORECAST CHALLENGES AT THE NWS STORM PREDICTION CENTER RELATING TO THE FREQUENCY OF FAVORABLE SEVERE STORM ENVIRONMENTS.

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

Doswell, C. A., III, Edwards, R., Thompson, R. L., Hart, J. A., & Crosbie, K. C. (2006). A Simple and Flexible Method for Ranking Severe Weather Events, Weather and Forecasting, 21(6), 939-951. Retrieved Sep 16, 2021, from https://journals.ametsoc.org/view/journals/wefo/21/6/waf959_1.xml

Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. J. Atmos. Sci., 46, 3077–3107.

Dudhia, J. (2014). Overview of WRF physics. University Corporation for Atmospheric Research, Boulder, CO, http://www2.mmm.ucar.edu/wrf/users/tutorial/201401/Physics_full.pdf.

Edwards, R., Corfidi, S. F., Thompson, R. L., Evans, J. S., Craven, J. P., Racy, J. P., McCarthy, D. W., & Vescio, M. D. (2002). Storm Prediction Center Forecasting Issues Related to the 3 May 1999 Tornado Outbreak, Weather and Forecasting, 17(3), 544-558. Retrieved Apr 1, 2022, from https://journals.ametsoc.org/view/journals/wefo/17/3/1520-0434_2002_017_0544_spcfir_2_0_co_2.xml

Efron, B., and R. J. Tibshirani, 1993: An Introduction to the Bootstrap. Chapman and Hall/CRC, 436 pp.

Ferguson, E. W., F. Ostby, and P. Leftwich, 1987: The tornado season of 1985. Mon. Wea. Rev., 115, 1437–1445.

Fujita, T., 1974: Jumbo tornado outbreak of 3 April 1974. Weatherwise, 27, 116–126.

Fuhrmann, C. M., Konrad , C. E., II, Kovach, M. M., McLeod, J. T., Schmitz, W. G., & Dixon, P. G. (2014). Ranking of Tornado Outbreaks across the United States and Their Climatological Characteristics, Weather and Forecasting, 29(3), 684-701. Retrieved Oct 22, 2021, from https://journals.ametsoc.org/view/journals/wefo/29/3/waf-d-13-00128_1.xml

Galway, J. G. (1977). Some Climatological Aspects of Tornado Outbreaks, Monthly Weather Review, 105(4), 477-484. Retrieved Oct 20, 2021, from https://journals.ametsoc.org/view/journals/mwre/105/4/1520-0493_1977_105_0477_scaoto_2_0_co_2.xml

Glickman, T., Ed., 2000: Glossary of Meteorology. 2nd ed. Amer.Meteor. Soc., 855 pp.

Grams, J. S., Thompson, R. L., Snively, D. V., Prentice, J. A., Hodges, G. M., & Reames, L. J. (2012). A Climatology and Comparison of Parameters for Significant Tornado Events in the United States, Weather and Forecasting, 27(1), 106-123. Retrieved Feb 7, 2023, from https://journals.ametsoc.org/view/journals/wefo/27/1/waf-d-11-00008_1.xml

Guyer, J., (2021) The 2020 Tornado Season: Active and Quiet in the Same Year, Weatherwise, 74:3, 36-43, DOI: 10.1080/00431672.2021.1896937

Hales, J., Jr., 1988: Improving the watch/warning program through use of significant event data. Preprints, 15th Conf. on Severe Local Storms, Baltimore, MD, Amer. Meteor. Soc., 165–168.

Hales, J. E., Vescio, M. D., & Koch, S. E. (1997). The 27 March 1994 Tornado Outbreak in the Southeast U.S.: The Forecast Process from a Storm Prediction Perspective. National Weather Digest, 21(4), 3–17.

Haykin, S., 1999: Neural Networks: A Comprehensive Foundation. Pearson, 842 pp.

Hitchens, N. M., & Brooks, H. E. (2012). Evaluation of the Storm Prediction Center's Day 1 Convective Outlooks, Weather and Forecasting, 27(6), 1580-1585. Retrieved Oct 24, 2021, from https://journals.ametsoc.org/view/journals/wefo/27/6/waf-d-12-00061_1.xml

Hong, S.-Y., and Lim J.-O. J. , 2006: The WRF single-moment 6-class microphysics scheme (WSM6). J. Korean Meteor. Soc., 42, 129–151.

Janjić, Z. I., 2002: Nonsingular implementation of the Mellor–Yamada level 2.5 scheme in the NCEP Meso Model. NCEP Office Note 437, 61 pp.

Johns, R., and W. Hirt, 1987: Derechos: Widespread convectively induced windstorms. Wea. Forecasting, 2, 32–49.

Kain, J. S., Weiss, S. J., Dembek, S. R., Levit, J. J., Bright, D. R., Case, J. L., ... & Schwartz, C. S. (2008, October). Severe-Weather Forecast Guidance from the First Generation of Large Domain Convection-Allowing Models: Challenges and Opportunities. In 24th Conference on Severe Local Storms.

Knupp, K. R., Murphy, T. A., Coleman, T. A., Wade, R. A., Mullins, S. A., Schultz, C. J., Schultz, E. V., Carey, L., Sherrer, A., McCaul, E. W., Jr., Carcione, B., Latimer, S., Kula, A., Laws, K., Marsh, P. T., & Klockow, K. (2014). Meteorological Overview of the Devastating 27 April 2011 Tornado Outbreak, Bulletin of the American Meteorological Society, 95(7), 1041-1062. Retrieved Mar 26, 2022, from https://journals.ametsoc.org/view/journals/bams/95/7/bams-d-11-00229.1.xml

Krocak, M. J., Flournoy, M. D., & Brooks, H. E. (2021). Examining Subdaily Tornado Warning Performance and Associated Environmental Characteristics, Weather and Forecasting, 36(5), 1779-1784. Retrieved Feb 13, 2023, from https://journals.ametsoc.org/view/journals/wefo/36/5/WAF-D-21-0097.1.xml

Lee, B. D., B. F. Jewett, and R. B. Wilhelmson, 2006: The 19 April 1996 Illinois Tornado outbreak. Part II: Cell mergers and associated tornado incidence. Wea. Forecasting, 21, 449–464.

Malakar, P., Kesarkar, A. P., Bhate, J. N., Singh, V., & Deshamukhya, A. (2020). Comparison of reanalysis data sets to comprehend the evolution of tropical cyclones over North Indian Ocean. Earth and Space Science, 7, e2019EA000978. https://doi.org/10.1029/2019EA000978

Markowski, P. M., 2002: Mobile mesonet observations on 3 May 1999. Wea. Forecasting, 17, 430–444.

Mellor, G. L., and Yamada T. , 1982: Development of a turbulence closure model for geophysical fluid problems. Rev. Geophys., 20, 851–875.

Mercer, A. E., & Richman, M. B. (2007). Statistical Differences of Quasigeostrophic Variables, Stability, and Moisture Profiles in North American Storm Tracks, Monthly Weather Review, 135(6), 2312-2338. Retrieved Jan 24, 2023, from https://journals.ametsoc.org/view/journals/mwre/135/6/mwr3395.1.x

Mercer, A. E., M. B. Richman, H. B. Bluestein, and J. M. Brown, 2008: Statistical modeling of downslope windstorms in Boulder, Colorado. Wea. Forecasting, 23, 1176–1194.

Mercer, A. E., Shafer, C. M., Doswell, C. A., III, Leslie, L. M., & Richman, M. B. (2009). Objective Classification of Tornadic and Nontornadic Severe Weather Outbreaks, Monthly Weather Review, 137(12), 4355-4368. Retrieved Dec 2, 2021, from https://journals.ametsoc.org/view/journals/mwre/137/12/2009mwr2897.1.xml

Mercer, A. E., Shafer, C. M., Doswell, C. A., III, Leslie, L. M., & Richman, M. B. (2012). Synoptic Composites of Tornadic and Nontornadic Outbreaks, Monthly Weather Review, 140(8), 2590-2608. Retrieved Mar 29, 2022, from https://journals.ametsoc.org/view/journals/mwre/140/8/mwr-d-12-00029.1.xml

Mercer, A.E. & Bates, A.V., "Contrasting Environments Associated with Storm Prediction Center Tornado Outbreak Forecasts using Synoptic-Scale Composite Analysis" (2014). Theses and Dissertations. 1173. https://scholarsjunction.msstate.edu/td/1173

Mlawer, E. J., Taubman S. J., Brown P. D., Iacono M. J., and Clough S. A., 1997: Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the long-wave. J. Geophys. Res., 102 (D14), 16 663-16 682.

NOAA / National Weather Service, National Centers for Environmental Prediction, & Storm Prediction Center. (2021). Storm Prediction Center Product & Report Archives. map, Norman, OK; Storm Prediction Center. Retrieved December 2, 2021, from https://www.spc.noaa.gov/archive/.

NWS, 2011: Service assessment: The historic tornadoes of April 2011. NOAA Rep., 77 pp.

Pautz, M. E., 1969: Severe local storm occurrences, 1955–1967. ESSA Tech. Memo. WBTM FCST12, Washington, DC, 3–4.

Potvin, C. K., Elmore, K. L., & Weiss, S. J. (2010). Assessing the Impacts of Proximity Sounding Criteria on the Climatology of Significant Tornado Environments, Weather and Forecasting, 25(3), 921-930. Retrieved Jan 24, 2023, from https://journals.ametsoc.org/view/journals/wefo/25/3/2010waf2222368_1.xml

Raschka S. (2015). Python Machine Learning. Birmingham: Packt Publishing Ltd.

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

Richman, M. B., Santosa B. , and Trafalis T. B. , 2005: Feature selection of radar-derived tornado attributes with support vector machines. Preprints, Fourth Conf. on Artificial Intelligence Applications to Environmental Sciences, San Diego, CA, Amer. Meteor. Soc., J5.1. [Available online at http://ams.confex.com/ams/pdfpapers/87991.pdf.].

Roebber, P., D. Schultz, and R. Romero, 2002: Synoptic regulation of the 3 May 1999 tornado outbreak. Wea. Forecasting, 17, 399–429.

Sanders, S., Adams, T., & Joseph, E. (2020). Severe Weather Forecasts and Public Perceptions: An Analysis of the 2011 Super Outbreak in Tuscaloosa, Alabama, Weather, Climate, and Society, 12(3), 473-485. Retrieved Mar 26, 2022, from https://journals.ametsoc.org/view/journals/wcas/12/3/wcasD180090.xml

Schneider, R. S., J. T. Schaefer, and H. E. Brooks, 2004: Tornado outbreak days: An updated and expanded climatology (1875–2003). Preprints, 22nd Conf. on Severe Local Storms, Hyannis, MA, Amer. Meteor. Soc., P5.1.[Available online at http://ams.confex.com/ams/11aram22sls/techprogram/paper_82031.htm.]

Schölkopf, B., and A. Smola, 2002: Learning with Kernels. MIT Press, 650 pp.

Shafer, C. M., A. E. Mercer, C. A. Doswell, M. B. Richman, and L. M. Leslie, 2009: Evaluation of WRF Forecasts of Tornadic and Nontornadic Outbreaks When Initialized with Synoptic-Scale Input. Mon. Wea. Rev., 137, 1250–1271, https://doi.org/10.1175/2008MWR2597.1.

Shafer, C. M., & Doswell, C. A. (2010). A multivariate index for ranking and classifying severe weather outbreaks. E-Journal of Severe Storms Meteorology, 5(1).

Shafer, C. M., Mercer, A. E., Leslie, L. M., Richman, M. B., & Doswell, C. A., III. (2010). Evaluation of WRF Model Simulations of Tornadic and Nontornadic Outbreaks Occurring in the Spring and Fall, Monthly Weather Review, 138(11), 4098-4119. Retrieved Feb 8, 2023, from https://journals.ametsoc.org/view/journals/mwre/138/11/2010mwr3269.1.

Shafer, C. M., and C. A. Doswell III, 2011: Using kernel density estimation to identify, rank, and classify severe weather outbreak events. Electronic J. Severe Storms Meteor., 6 (2), 1–28

Simmons, K. M., & Sutter, D. (2009). False Alarms, Tornado Warnings, and Tornado Casualties, Weather, Climate, and Society, 1(1), 38-53. Retrieved Oct 24, 2021, from https://journals.ametsoc.org/view/journals/wcas/1/1/2009wcas1005_1.xml

Skamarock, W., J. Klemp, J. Dudhia, D. Gill, M. Barker, W. Wang, and J. Powers, 2008: A Description of the Advanced Research WRF Version 3: NCAR Technical Note. National Center for Atmospheric Research.

Stensrud, D. J., J. V. Cortinas, and H. E. Brooks, 1997: Discriminating between tornadic and nontornadic thunderstorms using mesoscale model output. Wea. Forecasting, 12, 613–632.

Thompson, R. L., & Edwards, R. (2000). An Overview of Environmental Conditions and Forecast Implications of the 3 May 1999 Tornado Outbreak, Weather and Forecasting, 15(6), 682-699. Retrieved Apr 1, 2022, from https://journals.ametsoc.org/view/journals/wefo/15/6/1520-0434_2000_015_0682_aooeca_2_0_co_2.xml

Thompson, R. L., Edwards, R., Hart, J. A., Elmore, K. L., & Markowski, P. (2003). Close Proximity Soundings within Supercell Environments Obtained from the Rapid Update Cycle, Weather and Forecasting, 18(6), 1243-1261. Retrieved Feb 10, 2023, from https://journals.ametsoc.org/view/journals/wefo/18/6/1520-0434_2003_018_1243_cpswse_2_0_co_2.xml

Thompson, R. L., Edwards R. , and Mead C. M. , 2004: An update to the supercell composite and significant tornado parameters. Preprints, 22nd Conf. Severe Local Storms, Hyannis, MA, Amer. Meteor. Soc., P8.1. [Available online at https://ams.confex.com/ams/11aram22sls/techprogram/paper_82100.htm.]

Thompson, R. L., Mead C. M. , and Edwards R. , 2007: Effective storm-relative helicity and bulk shear in supercell thunderstorm environments. Wea. Forecasting, 22, 102–115.

Thompson, R. L., Smith, B. T., Grams, J. S., Dean, A. R., & Broyles, C. (2012). Convective Modes for Significant Severe Thunderstorms in the Contiguous United States. Part II: Supercell and QLCS Tornado Environments, Weather and Forecasting, 27(5), 1136-1154. Retrieved Feb 13, 2023, from https://journals.ametsoc.org/view/journals/wefo/27/5/waf-d-11-00116_1.xml

Trafalis, T. B., B. Santosa, and M. B. Richman, 2005: Learning networks for tornado forecasting: A Bayesian perspective. Proc. Sixth Int. Conf. on Data Mining, Skiathos, Greece.

Ukkonen, P., & Mäkelä, A. (2019). Evaluation of machine learning classifiers for predicting deep convection. Journal of Advances in Modeling Earth Systems, 11, 1784– 1802. https://doi.org/10.1029/2018MS001561

Walters, J. E., Mason, L. R., Ellis, K., & Winchester, B. (2020). Staying Safe in a Tornado: A Qualitative Inquiry into Public Knowledge, Access, and Response to Tornado Warnings, Weather and Forecasting, 35(1), 67-81. Retrieved Oct 24, 2021, from https://journals.ametsoc.org/view/journals/wefo/35/1/WAF-D-19-0090.1.xml

Weisman, M. L., and J. B. Klemp, 1984: The structure and classification of numerically simulated convective storms in directionally varying wind shears. Mon. Wea. Rev., 112,2479–2498.

Wilks, D. S., (2019). Statistical Methods in the Atmospheric Sciences (4th ed.). Elsevier.

Zhang, Y., Wang, X., Li, C. et al. NDVI dynamics under changing meteorological factors in a shallow lake in future metropolitan, semiarid area in North China. Sci Rep 8, 15971 (2018). https://doi.org/10.1038/s41598-018-33968-w