UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

IDENTIFICATION OF MID-TROPOSPHERIC PATTERNS ASSOCIATED WITH TORNADO OUTBREAKS IN THE UNITED STATES

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfilment of the requirements for the

Degree of

MASTER OF SCIENCE IN GEOGRAPHY

By

PAULINA ĆWIK
Norman, Oklahoma
2019

IDENTIFICATION OF MID-TROPOSPHERIC PATTERNS ASSOCIATED WITH TORNADO OUTBREAKS IN THE UNITED STATES

A THESIS APPROVED FOR THE
DEPARTMENT OF GEOGRAPHY AND ENVIRONMENTAL SUSTAINABILITY

BY THE COMMITTEE CONSISTING OF

Dr. Renee A. McPherson, Chair

Dr. Harold Brooks

Dr. J. Scott Greene

Acknowledgments

I would like to express my deepest gratitude to my advisor and committee chair, Dr. Renee McPherson, who supported me since our first meeting in my journey to pursue a Master of Science degree in Geography at the University of Oklahoma. I sincerely appreciate her boundless kindness, great passion, knowledge, and enthusiasm for this work. Dr. Renee, thank you for your patience, countless productive discussions, and your trust in me, which helped shape my work and overcame all obstacles. You are an incredibly powerful and positive spirit and it is an honor to work with you.

I would also like to sincerely thank my committee member, Dr. Harold Brooks. His great expertise and undeniable passion for topics associated with severe weather sparked countless motivating discussions that enriched my learning expertise and generated new ideas. Dr. Brooks, your help and positive encouragement during past years consolidated the foundation of this work and allowed me to explore interesting aspects of tornado outbreaks nomenclature without getting overwhelmed and lost, for which I am extremely grateful.

Dr. Greene, thank you so much for being a valuable member of my committee and taking the time to review my work. I appreciate your guidance and ideas you shared with me to improve my research.

I would like to especially thank Prof. Michael Richman for his incredible support and extremely important assistance in this work. His expert skills, scientific experience, and knowledge of the complex method of Principal Component Analysis are invaluable. Prof. Richman, without you this work wouldn't be possible, and I sincerely appreciate the countless hours, days, and weeks of discussions and analysis. Thank you from the bottom of my heart.

Furthermore, I would like to acknowledge the work of Dr. Andrew Mercer, who supported this research by developing parts of the R coding for the tornado outbreak classification, and to Chad Shafer for helpful discussion on tornado outbreak definition approach. I would like to thank everyone in the South Central Climate Adaptation Science Center and in the Department of Geography and Environmental Sustainability for all the guidance and help that I received. It has been a great adventure and I appreciate each and every person who supported me throughout the way.

I want to acknowledge my funding source, the Office of the OU Vice President for Research, for supporting my work, and the University of Oklahoma for providing me with the opportunity to contribute to the Academy.

Finally, I thank my wonderful, sweet, and loving parents, sister, and Eric Jacobsen. I am grateful beyond Moon and back for all the love, faith, and support you all gave me. Your encouragement, constant uplifting thoughts, and every single smile helped me overcome all the troubles. Thanks to you, I find myself today ready to share this work with the World. THANK YOU!!

Table of Contents

List of Tables

List of Figures

Abstract

The identification of large-scale atmospheric patterns associated with tornado outbreaks poses a great challenge. It involves analysis of physical processes occurring at different time and space scales that, in the right configuration, result in environmental conditions favoring tornado outbreak formation. Over the years, there have been numerous studies that utilize the notion of 'tornado outbreaks.' The term has been used to define severe weather events where the occurrence of multiple tornados has been determined. The exact meaning of 'tornado outbreak,' however, has been repeatedly redefined and has evolved throughout the years. Depending on the availability of scientific data, technological advancements, and purpose of these definitions, different authors offered diverse approaches to shape the perception and applications of 'tornado outbreak.' This work provides an extensive review of the evolving nature of the 'tornado outbreak' definition. Each decade contains multiple examples of manuscripts that contributed to either changes in the tornado outbreak definition or its perception.

This work also offers a statistical approach that can be used to define tornado outbreak events and identify historic cases from the tornado report database of the National Weather Service's Storm Prediction Center. The approach was informed by the review of tornado outbreak definitions and used kernel density estimation to identify 4,991 outbreaks from 1950 to 2017 — an average of 73 outbreaks per year. Applying a data-driven threshold of seven or more tornados in a cluster, 333 major tornado outbreaks were found to occur east of the Rocky Mountains throughout the 68 years of the analysis. The highest count of tornado outbreaks by month was found in May.

Finally, to support efforts directed towards research on large-scale atmospheric patterns and subseasonal forecasting of tornado outbreaks, this study applies principal component analysis

(PCA), hierarchical clustering, and silhouette analysis to identify synoptic-scale patterns of 500-hPa geopotential height associated with tornado outbreaks in the United States. The PCA was performed on a similarity matrix derived from monthly 500-hPa geopotential height anomalies from the 20th Century Reanalysis during times when tornado outbreaks initiated. The analysis was performed using T-mode decomposition for observations during May from 1950 to 2014. To determine the number of PC patterns to retain, congruence coefficient analysis on loadings from Promax and Varimax transformations was performed. The PC analysis identified two major atmospheric patterns for the month of May. To validate these results, two additional statistical methods were used: hierarchical clustering and silhouette analysis. Both methods identified the same patterns as the PC analysis, and thus validated our results.

Chapter I: Introduction

Multi-tornado events that cause large damage to property, infrastructure, and people are common in the United States (U.S.). According to Schneider et al. (2006) [1], the majority of tornado-related fatalities result from tornado outbreaks. In the U.S., typically the atmospheric conditions that are most conducive to the formation of tornados, either as isolated events or as a part of a major outbreak, occur east of the Rocky Mountains. For example, one of the biggest tornado outbreaks in recent history occurred on 25-28 April 2011 in the Southeastern U.S. and caused 316 fatalities, 2700 injuries, and over $4.2 billion dollars of economic losses across five states [2-3]. With a continued increase in population density and expansion of urban areas, communities and businesses located in areas prone to tornado-outbreak impacts remain vulnerable. Considering the significance of such damaging events, research on the climatology of tornado outbreaks is of great importance.

Numerous studies over past decades utilized the notion of 'tornado outbreak;' however, there were many different definitions and criteria by which tornado outbreaks were classified, including some that did not include any specific characteristics that could be measured or quantified, such as 'group' or 'family' of tornados. As noted by Galway (1977) [4], "A tornado 'outbreak' can mean many things to many people." The diversity of approaches regarding the use of the term often reflected individual research goals of investigators. For instance, Pautz (1969) [5] counted the number of tornados in a day and grouped tornado outbreaks into three categories: a 'small family outbreak' with 6 to 10 tornados, a 'moderate family outbreak' with 11 to 20 tornados, and a 'large family outbreak' with more than 20 tornados. A more specific definition, including magnitude of the tornados and its spatial characteristics, was provided by Grazulis (1990) [6]. In his work,

tornado outbreaks were considered in the context of outbreaks of significant tornados, defined as F-2 or greater on the Fujita scale, and with total path lengths exceeding 100 miles [6].

Further, tornado outbreaks also were studied in association to specific geographic regions [7-8]. One such region was Florida due to its particular geographic location. For example, Hagemeyer (1994) [9] defined tornado outbreak as the "occurrence of four or more tornados in four hours or less at, or south of, 30° latitude" and added that the tornados needed to be related to the same synoptic forcing mechanisms. Numerous research examples used definitions that focused on multiple variables or statistical modeling [10-12]. Consequently, the term was attributed in many contexts to events that had a different number of reported tornados as well as associated magnitudes, durations, or spatial distribution. In addition, the term was applied in the context of famous events, such as the "jumbo outbreak" of 3 April 1974 [13-15], the "Super outbreak" of 27 April 2011 [16-18], or the "the Super Tuesday outbreak" on 5-6 February 2008 [19-21]. Additionally, the term 'outbreak' was used to describe other types of severe weather events, such as convective storms, regardless of whether the event included tornado reports or not [12]. Therefore, there was not any general agreement or acceptance of what constitutes a tornado outbreak and, depending on the requirements of the research project or scientific application, the definition remains adaptable.

Owing to the potential impacts of tornados on life and property, studies of tornadic storms and tornadic storm systems are of great interest among scientists around the world. Accurate forecasts of potential tornado outbreaks are essential, considering the significant potential for loss of property and life. One key feature to improvements in tornado outbreak forecasts is understanding what and how environmental conditions favor the formation of tornados and tornado outbreaks. Characterization of the atmospheric conditions that are necessary for tornado development, such

as high instability, vertical wind shear, and a lifting mechanism to name few, has been the subject of much research [22-26]. Typically, tornado outbreaks are forced by large-scale drivers, such as synoptic-scale atmospheric patterns [27]. An example of such patterns is the shift in the jet stream associated with different phases of El Niño-Southern Oscillation [28-29]. The relationship of the major tornado outbreaks and associated atmospheric patterns, such as high- and low-pressure systems, are of interest in this research. This study will focus on an analysis of patterns that occur in the mid-troposphere (at 500 hPa), a level associated with energy and momentum for tornado outbreaks. The primary goal of this work is to identify and describe mid-tropospheric, geopotential height patterns that occur most often in association with historic tornado outbreaks.

Chapter 2 of this manuscript provides an extensive review — by decade — of the evolving nature of the term 'tornado outbreak,' with multiple examples of manuscripts that contributed to the changes in tornado outbreak definition. Additionally, Chapter 2 outlines an updated statistical approach to identify tornado outbreaks from a historic record of tornado reports. Chapter 3 identifies and characterizes the atmospheric patterns associated with major tornado outbreaks, and Chapter 4 summarizes the main results.

References

1. Schneider, R. S., A. R. Dean, S. J. Weiss, P. D. Bothwell, 2006: Analysis of estimated environments for 2004 and 2005 severe convective storm reports. *Preprints, 23rd Conf. on Severe Local Storms*, St. Louis, MO, American Meteorological Society, 3-5.
2. Knupp, K.R., T.A. Murphy, T.A. Coleman, R.A. Wade, S.A. Mullins, C.J. Schultz, E.V. Schultz, L. Carey, A. Sherrer, E.W. McCaul, B. Carcione, S. Latimer, A. Kula, K. Laws, P.T. Marsh, and K. Klockow, 2014: Meteorological overview of the devastating 27 April 2011 tornado outbreak. *Bulletin of American Meteorological Society*, **95**, 1041–1062.
3. Hayes J.L., 2011: The historic tornadoes of April 2011. U.S. Department of Commerce, NOAA, CreateSpace Independent Publishing Platform, Silver Spring, Maryland, 76.
4. Galway, J. G., 1977: Some climatological aspects of tornado outbreaks. *Monthly Weather Review*, **105**, 477–484.
5. Pautz, M.E., ed., 1969. Severe local storm occurrences 1955-1967 (Vol. 12). Environmental Science Services Administration, Weather Bureau, U.S. Department of Commerce; Silver Spring, MD, 77 pp.
6. Grazulis, T.P., 1990: Significant tornadoes…..107 year perspective. *Journal of Wind Engineering and Industrial Aerodynamics*, **36**, 131-151.
7. Hales, J.E., and M.D. Vescuo, 1997: The 27 March 1994 tornado outbreak in the southeast US: The forecast process from a storm prediction center perspective. *National Weather Digest*, **21**, 1-15.
8. Hagemeyer, B.C., and G.K. Schmocker, 1991: Characteristics of east-central Florida tornado environments. *Weather and Forecasting*, **6**, 499-514.

9. Hagemeyer, B.C., and G. Matney, 1994: Peninsular Florida tornado outbreaks (1950-1993). NOAA Tech. Memo. SR-151, 108 pp. [Available from National Technical Information Service, 5285 Port Royal Rd., Springfield, VA 22161.].

10. Mercer, A. E., C. M. Shafer, C. A. Doswell III, L. M. Leslie, and M. B. Richman, 2009: Objective classification of tornadic and nontornadic severe weather outbreaks. *Monthly Weather Review,* **137**, 4355–4368.

11. Lewellen, D.C., 2012: Effects of topography on tornado dynamics: A simulation study. In *Preprints, 26th Conf. on Severe Local Storms,* Nashville, TN, American Meteorological Society, Vol. 4.

12. Doswell III, C.A., R. Edwards, R.L. Thompson, J.A. Hart, and K.C. Crosbie, K.C., 2006: A simple and flexible method for ranking severe weather events. *Weather and Forecasting,* **21**, 939-951.

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

14. Corfidi, S.F., S.J. Weiss, J.S. Kain, S.J. Corfidi, R.M. Rabin, R.M. and J.S. Levit, 2010: Revisiting the 3–4 April 1974 super outbreak of tornadoes. *Weather and Forecasting*, **25**, 465-510.

15. Locatelli, J.D., M.T. Stoelinga, and P.V. Hobbs, 2002: A new look at the super outbreak of tornadoes on 3–4 April 1974. *Monthly Weather Review*, **130**, 1633-1651.

16. Knupp, K.R., T.A. Murphy, T.A. Coleman, R.A. Wade, S.A. Mullins, C.J. Schultz, E.V. Schultz, L. Carey, A. Sherrer, E.W. McCaul Jr, and B. Carcione, 2014: Meteorological overview of the devastating 27 April 2011 tornado outbreak. *Bulletin of the American Meteorological Society,* **95**, 1041-1062.

17. Knox, J.A., J.A. Rackley, A.W. Black, V.A. Gensini, M. Butler, C. Dunn, T. Gallo, M.R. Hunter, L. Lindsey, M. Phan, and R. Scroggs, 2013: Tornado debris characteristics and trajectories during the 27 April 2011 super outbreak as determined using social media data. *Bulletin of the American Meteorological Society,* **94**, 1371-1380.

18. Gaffin, D.M., 2012: The influence of terrain during the 27 April 2011 super tornado outbreak and 5 July 2012 derecho around the great Smoky Mountains National Park. In *Preprints 26th Conference on Severe Local Storms,* Nashville, TN. American Meteorological Society.

19. McMillan, A., B.J. Reynolds, T. Brown, D. Liang, and J.A. Womble, 2006: Advanced Technology for Rapid Tornado Damage Assessment Following the "Super Tuesday" Tornado Outbreak of February 2008. *Multidisciplinary Center for Earthquake Engineering Research (MCEER),* MCEER-08-SP01

20. Chaney, P.L., and G.S. Weaver, 2010: The vulnerability of mobile home residents in tornado disasters: The 2008 Super Tuesday tornado in Macon County, Tennessee. *Weather, Climate, and Society,* **2**, 190-199.

21. Carbin, G.W., and J.T. Schaefer, 2008: The "Super Tuesday Tornado Outbreak" of February 5-6, 2008, SPC forecasts and historical perspective. *In Preprints, 36th Conference on Broadcast Meteorology,* Denver, CO, American Meteorological Society, P4, Vol. 5.

22. Doswell, C. A., III, R. Edwards, R.L. Thompson, J.A. Hart, and K.C. Crosbie, 2006: A simple and flexible method for ranking severe weather events. *Weather and Forecasting,* **21**, 939-951.

23. Rasmussen, E.N., 2003: Refined supercell and tornado forecast parameters. *Weather and Forecasting,* **18**, 530–535.

24. C.A. Doswell III, 1987: The distinction between large-scale and mesoscale contribution to severe convection: a case study example. *Weather and Forecasting,* **2**, 3-168.

25. Sherburn, K.D., and M.D. Parker, 2014: Climatology and ingredients of significant severe convection in high shear, low CAPE environments. *Weather and Forecasting,* **29**, 854-877.

26. Grams, J. S., R. L. Thompson, D. V. Snively, J. A. Prentice, G. M. Hodges, and L. J. Reames, 2012: A climatology and comparison of parameters for significant tornado events in the United States. *Weather and Forecasting*, **27**, 106–123.

27. Lee, S.-K., R. Atlas, D. Enfield, C. Wang, and H. Liu, 2013: Is there an optimal ENSO pattern that enhances large-scale atmospheric processes conducive to tornado outbreaks in the United States? *Journal of Climate,* **26**, 1626–1642.

28. Knowles, J. B., and R. A. Pielke, 2005: The Southern Oscillation and its effects on tornadic activity in the United States. *Atmospheric Sciences Paper 755,* Colorado State University, 15.

29. Bove, M. C., 1998: Impacts of ENSO on United States tornado activity. *Preprints, 19th Conf. on Severe Local Storms,* Minneapolis, MN, American Meteorology Society, 313–316.

Chapter II: A systematic way of tornado outbreak classification

1. Introduction

Each year, the United States is affected by at least a few tornado outbreaks [1] that cause large damage to property and infrastructure and pose a great threat to people. According to Schneider et al. (2006), the majority of tornado-related fatalities in the United States result from tornado outbreaks [2]. Major tornado outbreaks are particularly dangerous because they consist of multiple, often violent (EF4-EF5), long-track tornados that generally persist on the ground for tens of minutes and, thus, are likely to affect populated areas [3]. Considering the significance of such damaging events to communities and businesses located in areas prone to severe weather, it is essential to study the climatology of tornado outbreaks.

Over the years, numerous studies have utilized the notion of 'tornado outbreaks,' and the concept of an outbreak has become common. However, the term 'tornado outbreak' has not been universally agreed upon [4]. Some definitions fail to describe specific characteristics that could be measured or quantified. For example, the *Glossary of Meteorology* from American Meteorological Society (AMS) defines a tornado outbreak as "multiple tornado occurrences associated with a particular synoptic-scale system," but the AMS does not mention how many tornados of what magnitude meet the requirements of "multiple" [5]. Other definitions focus on selective characteristics, such as a threshold of a specific number of tornados [6], or multiple variables and statistical modeling techniques to distinguish tornado outbreaks [7]. Consequently, the term has been attributed to events that had a different number of reported tornados, magnitudes, durations, or spatial distribution, and it has been applied in the context of famous events, such as the "jumbo outbreak" of 3 April 1974 [8]. Additionally, the term 'outbreak' has been used interchangeably to

describe other types of severe weather events, such as convective storms, regardless of whether the event included tornado reports or not [9]. As a result, there is no universal acceptance of what constitutes a tornado outbreak, and the definition remains adaptable depending on the requirements of a given research project.

The evolution of tornado outbreak definitions and classifications throughout the decades will be presented. This review will build a context for a recommendation of a data-driven, statistical approach to define tornado outbreaks proposed in this research. Using this proposed definition, historic cases of major tornado outbreaks will be identified and later used in the analysis of atmospheric patterns.

2. Tornado outbreak definitions in historic literature

As research into tornado development and impacts grew, various definitions of 'tornado outbreaks' arose. This diversity results from several reasons. First, over the past 70 years, the availability and quality of tornado-related meteorological data have changed tremendously. It is a collective result of (1) an increase in sample size with each new year in the historical record, (2) changes in tornado detection techniques, such as the introduction of Weather Surveillance Radar-1988 Doppler (WSR-88D; [10]), and (3) the modification of methods that guided the process of collecting reports (e.g., establishment of the Fujita [11] and Enhanced Fujita scales [12]). Second, a significant increase in tornado reporting statistics, especially in the number of weak tornado reports [13], has resulted from the expansion of the observational network. Namely, an increase in population density, urbanization, and better public severe weather awareness have caused more people to be attentive to weather conditions [10]. Furthermore, the recent proliferation of recording devices in

smartphones with an internet connection has enabled almost instant confirmation of tornado existence and, in some recent cases, live streaming of tornado development, even from remote places in the United States. This evolution of tornado reporting has been discussed by many authors [14-17] who described non-meteorological factors influencing individual tornados in the historical record. Naturally, by extension, changes in the tornado database have influenced how to define a tornado outbreak simply by offering a longer, more specific record of individual tornados used in new types of analyses. Finally, improvements in computational power, as well as new software and programming tools, over the past few decades have drastically increased the speed of data processing. Consequently, mining large datasets has become more approachable and time efficient, thus offering new ways to explore tornado datasets with sophisticated statistical techniques and adding new perspectives to tornado outbreak definition.

Considering these many changes, a comprehensive review of tornado outbreak definitions and underlying their approaches over the decades has been presented below.

2.1 1950-1960

The term ‘tornado outbreak’ started wide appearance in the scientific literature in the middle of the past century. The first publication that mentioned tornado outbreak can be traced to 1951. Major Fawbush, Captain Miller, and Captain Starrett, from the United States (U.S.) Air Force and the Air Weather Service of Tinker’s Air Force Base in Oklahoma City, presented empirical methods of tornado development forecasting [18]. In this work, the forecasters described the synoptic situation of a frontal system where “seven individual tornados, with paths up to 30 miles in length” (p. 8) were recorded by the U.S. Weather Bureau. In their manuscript, they used ‘outbreak of tornados’ interchangeably with ‘outbreak of storms;’ thus, it is unclear if the authors

used the term ‘outbreak’ in association with tornados only, or only to emphasize the unusual scale of a storm event. Carr (1952) published a report on the tornados that occurred on 21-22 March 1952 and affected the lower Mississippi and Tennessee Valleys [19]. Despite the fact that ‘outbreak’ in his publication meant only “a series of tornados” (p. 50), this paper was one of the first to focus on a specific, individual event of multiple tornados and included a description of atmospheric conditions leading to the event.

The first extensive listing of tornado outbreaks appears in the book titled, “Tornados of the United States” from 1953 [20]. Here, Flora defined ‘outbreak’ as a “family or series of tornados […] that occur in groups that break out approximately the same hour or within a few hours of each other” (p. 207). Other contemporary publications used the term ‘outbreak’ to describe the severity of a tornado event. For instance, Van Tassel (1955) published “The North Platte Valley tornado outbreak of June 27, 1955” [21], wherein ‘tornado outbreak’ referred to one particularly strong tornado, specifically “the largest and most devastating tornado in the history of western Nebraska” (p. 255).

Beebe (1956), a researcher and operational meteorologist from the U.S. Weather Bureau’s Severe Local Storms Forecast Center in Kansas City, was the first author to define a tornado outbreak. In his 1956 publication [22], an outbreak was defined as “one in which 3 or more tornadoes occurred within a specific area and at least 2 of these tornadoes were separated by a distance of 100 miles or more” (p. 140). By the end of the decade, the term ‘tornado outbreak’ had been used loosely in publications mostly referring to severe weather events with multiple tornados but with no details on the number of tornados or other threshold conditions [23-27].

2.2 1960-1970

Literature in the 1960s included a little more detail to the definition of a tornado outbreak. In that decade, Wolford (1960) proposed a tornado outbreak definition in "Tornado occurrences in the United States" [28]: a "family-type series of tornadoes, that travel in the same direction, following parallel paths that are rather close together, within a space of few hours in the same State or section of the country" (p.12). The term "family-type" associated with tornado outbreaks was commonly used in the literature of this decade [29-32]. It is not entirely clear, however, if this term referred to Wolford's definition that suggested parallel tornado paths or to the previous definition from Flora (1953) that named any group of tornados (including multiple tornados from a single storm) as a family of tornados. Judging from the characterization of the events included in the publications, most authors likely were distinguishing between multiple independent events as "family-like tornado outbreaks," and instances where multiple tornados occurred from a single storm as a "multiple tornado outbreak" [33].

O'Connor (1965) was the first author to describe the Palm Sunday tornado outbreak, which at the time was "the second greatest tornado disaster in the Nation's history in terms of the number dead (271)" [34, p. 465]. Only a few tornado outbreaks in modern history were remembered by a specific name due to the magnitude of destruction that they caused. The Palm Sunday tornado outbreak was one of them, documented and analyzed by many authors [35-40].

Finally, Pautz (1969) grouped tornado outbreaks into three categories: a 'small family outbreak' with 6 to 10 tornados, a 'moderate family outbreak' with 11 to 20 tornados, and a 'large family outbreak' with more than 20 tornados [41]. He also defined the term 'family outbreak' as the occurrence of six or more tornados on one tornado day over a relatively small area, with

no clarification what a ‘tornado day’ or ‘small area’ were. This publication marked the end of the decade with the first definition of tornado outbreak based on a number of observed tornados and some general guidance for classifying the sizes of outbreaks.

2.3 1970-1980

New understanding of tornados that arose in the 1970s modernized the approaches to characterizing individual tornados and, consequently, helped to generate new tornado outbreak classifications. Fujita (1971) documented results from the Tornado Watch Experiment project [42], which investigated satellite-viewed cloud characteristics in relation to tornado occurrences. This report suggested that the characterization of individual tornados should be based on two measures: tornado intensity and area of impact. To address this idea, Fujita [42] proposed a new six-level scale based on damaging wind ranges. The Fujita scale, or F-scale, assigned a tornado intensity using estimates of wind speeds from structural or tree damage and classified tornados as Gale (F0), Weak (F1), Strong (F2), Severe (F3), Devastating (F4), and Incredible (F5) [42]. In addition, tornado-affected areas also were categorized by size as Trace (TR), Decimicro (DM), Micro (MI), Meso (ME), Macro (MA), Giant (GI), and Decagiant (DG).

The adoption of the Fujita scale in tornado outbreak research allowed scientists to compare one aspect of the damage magnitude among tornados within an outbreak. For example, the scale was used to examine individual tornados in one of the most famous outbreaks in U.S. history — the “Jumbo” tornado outbreak of 3 April 1974 [43-46]. The new classification based on storm damage transformed the way in which tornados were reported and gave a possibility for exploration of new tornado outbreak definitions that now could include the magnitude of an event.

A quite different approach to tornado outbreak classification was introduced by Maddox and Gray (1973). In their publication, tornado outbreaks were grouped into 'tornado outbreak days' to analyze atmospheric conditions closely associated with tornados using respective proximity soundings. A 'tornado outbreak day' was defined as "a day on which an unusually large number (roughly twenty or more) of destructive tornadoes occurred over a contiguous region of radius approximately 200 nautical miles" [47, p. 2]. This work was the first one to incorporate the number of tornados and to add specific temporal and spatial characteristics to the tornado outbreak definition.

Two years later, Galway (1975) reintroduced the tornado outbreak definition from 1969 of six tornados or more [41], with a slight modification to number of tornados in each category [48]. The definition classified small outbreaks with 6 to 9 tornados, moderate with 10 to 19 tornados, and large with 20 or more. Subsequently, Galway (1977) presented yet another approach [49] whereby an 'outbreak' was defined as 10 or more tornados from a single organized weather system. Further, tornado outbreaks were divided into three categories: local, progressive, and line. The 'local outbreak' was defined as an outbreak with a maximum duration of seven hours and area of activity confined to a circular envelope of $\sim 1.0*10^4$ nm^2 (i.e., square nautical miles). A 'progressive outbreak' was characterized as an outbreak that progressed (or advanced) from west to east within a duration of 9.5 h on average, where the distance between first and last tornado report was greater than 350 nm, and the activity envelope was at least $5.4*10^4$ nm^2. Finally, a 'line outbreak' was defined as an outbreak "with a limited eastward progression that forms on an axis, generally oriented north-south" (p. 478) with a duration of about 8 h, and an area of $5.9*10^4$ nm^2. Galway's tornado outbreak definition from 1977, conceptually close to that of Maddox and Gray (1973), was first to portion spatial and temporal classifications into smaller bins to capture more

detailed character of tornado outbreaks. His work showed an emerging pattern of utilizing the outbreak definition as a tool in a quest to define precursor conditions leading to tornado outbreak development.

2.4 1980-1990

At the turn of the 1980s, comprehensive studies of the statistical properties of tornados, such as the average number of tornados within a certain distance of given point or statistical center of tornadic activity, became popular [50-51]. Scientists conducted detailed analyses of information gathered in various tornado databases to acquire new information about U.S. tornado probabilities, climatologies, and risk assessments [52-56]. By the end of the 1970s, researchers noticed that there were multiple inconsistencies present in current tornado databases [57-58]. Namely, some of the early reports of tornados contained spatial and temporal inconsistencies in the reports of the same events, such as, for example, one long-track tornado passing simultaneously through two different cities separated by a distance of almost 70 miles [59]. Another matter in question, causing heated discussions among scholars, was the use of the Fujita scale rating and viability of the assigned categories [60-62]. Consequently, there emerged an increased need for verification of tornado databases and improvement in quality of the tornado record [63-67].

In addition, awareness of data inconsistencies in the databases raised questions about the definition of a tornado itself and, by extension, tornado outbreaks. The current definition, formulated in 1959, defined a tornado as “a violently rotating column of air, pendant from a cumulonimbus cloud, and nearly always observable as a funnel cloud or tuba” [68, p. 638]. Many tornados documented in contemporary (late 1970’s) literature failed to meet those criteria [69-73]. Therefore, Forbes and Wakimoto (1983) offered a new, “more pragmatic” tornado definition, where a “vortex is

classified as tornado if (i) it produces at least F0 damage or exhibits wind speeds capable of producing such damage and (ii) if it forms in association with the wind field of a thunderstorm or its accompanying mesoscale features, such as the gust front and flanking line" [74, p. 232]. Following this definition, they analyzed a case study of 18 tornados resulting from a bow-echo structure on 6 August 1977 and compared the density of tornado occurrences in that event with that of the "super outbreak" of 3 April 1974. The spatial separation of tornados from the 1977 event was roughly a factor of 10 greater than that of the 1974 event. Such a high density of tornados in a single event encouraged the idea of a 'concentrated tornado outbreak' [74], where an outbreak was defined by many weak tornados that did not have to be associated with cumulonimbus clouds. As may be expected, however, many researchers continued with the definitions of tornado outbreaks offered by previous authors [75-77] or applied the term to major events characterized by multiple strong tornados, fatalities, and property damage [78-82].

Statistical analysis of different weather events that produced tornados was not the only measure by which scientists investigated tornado outbreaks in the 1980s. Some focused on the corresponding environmental atmospheric conditions, such as humidity, wind, pressure, or kinetic energy, at different atmospheric levels that occurred prior to and during a specific tornado outbreak event. The often-studied Red River Valley tornado outbreak from 10–11 April 1979 was one example [82-90]. Many of these studies were possible to undertake because of the availability of new, detailed data resulting from routine observations and measurements taken systematically from radiosonde, satellite, mobile soundings, and radar observations [91-94]. The abundance of observational datasets and techniques employed individually or collectively [73] in the 1980s allowed progressive development of research aiming not only to improve the quality of the tornado

record, but also analyses of individual components of atmospheric conditions and their roles in the process of tornado outbreak development.

2.5 1990-2000

In the early 1990s, the study of tornado outbreaks gravitated towards concepts inspired by the research of the preceding decade. While analyzing past major tornado outbreaks, Grazulis (1990) noted limitations in the tornado database [95] such as a“ bias in both tornado documentation and the Fujita scale rating process” (p. 131), leading to a conclusion that the distribution of tornado risk still was not well understood. Major tornado outbreaks were defined in his work as outbreaks of significant tornados (F2-F5, later expanded to include any tornado that caused death) with total path lengths exceeding 100 miles [95-96]. In subsequent work, Grazulis (1993) based a tornado outbreak on Galway (1975) [48] and defined it as “a group or a family of six or more tornados which are spawned by the same general weather system” [97, p. 13]. In this definition, a small outbreak could include six tornados from two different thunderstorms in advance of a cold front; however, if the six tornados were scattered across different weather systems, they would not be treated as an outbreak. Simultaneously, Grazulis (1993) extended the definition to include information on the gap between the end of one outbreak and the start of another. He marked this gap as a threshold of a “six-hour lull in tornado activity” and underscored that an outbreak did not have to be confined to a single calendar day [97]. Previous tornado outbreak definitions (e.g., Galway (1977)) continued to be in use during this time period [98-99].

Meanwhile, research focusing on individual cases of tornado outbreaks, such as the 15 June 1988 Colorado outbreak [100-101], continued to apply radar and satellite technologies to test new hypotheses and to compare different analysis tools [102]. Additionally, given the proliferation of

popular and inexpensive video cameras [103], photogrammetric analyses of tornados were reestablished. New or evolving observational capabilities and enhanced computational resources led to the advent of improved computer modeling, resulting in more accurate numerical weather simulations. That advancement was beneficial for research on tornado outbreaks and contributed to improvements of prediction [104], enhancement of the understanding of the environmental conditions favoring the evolution of tornado outbreaks [105], and acceleration of the development of tornado outbreak simulations for comparison and analysis of damage paths [106-107]. Analyses of the damage potential of tornado outbreaks also were conducted. For example, Doswell et al. (1993) found that outbreak-related tornados were capable of the highest damage potential of any tornados and that they were usually produced by supercells [108].

While research on synoptic-scale environmental conditions for outbreaks had been performed frequently [82-90, 108-110], it was only since the 1990s that scale-dependent research on tornado outbreaks began to increase [111-113]. Particular attention in this decade was dedicated to mesoscale processes associated with tornado outbreaks, such as kinematic and thermodynamic conditions supporting tornado outbreaks [114-117]. Researchers found that certain mesoscale features contributed not only to the dynamics of tornado outbreaks but also to the longevity of the supercell storms that produced tornados [118]. In some research, tornado outbreaks were analyzed through the lens of particular geographic region [119-120], with a main focus on the improvement of regional knowledge, forecasting, and warnings of those events. An example of such regionally centered research was the study of tornado outbreaks occurring in distinctive geographic locations, such as Florida [121-125]. Hagemeyer (1994) noted that because peninsular Florida was isolated from surrounding states, a unique regional definition of tornado outbreak was needed [126]. In this work, a tornado outbreak was defined as the “occurrence of four or more tornados in four hours or

less at, or south of, 30° latitude" (p. 3), adding that tornados within an outbreak needed to be related to the same synoptic forcing mechanisms [126]. He used this definition later to analyze the tornado outbreak climatology for peninsular Florida, distinguishing three tornado outbreak environments: extratropical, tropical, and a hybrid of those two [127]. In the following years, tornado outbreaks associated with tropical environments and hurricanes continued to be researched, contributing to a broader understanding of the influence of different environmental factors on the evolution of tornado outbreaks [128-133].

Finally, by the end of the decade, scientific interests expanded towards the analysis and estimation of the total threat associated with tornado outbreaks. For instance, Thompson and Vescio (1998) observed that it was difficult to compare different tornado outbreaks with each other [134]. The existing Fujita scale relied on damage to manmade structures and thus tended to "preclude high ratings for tornados occurring in sparsely populated areas." To address this issue, the authors proposed an improved method of tornado categorization, called the Destruction Potential Index (DPI). The DPI measured the potential for damages and casualties for any specific time period in a particular tornado outbreak. It was calculated as the total tornado damage area multiplied by the weighted mean F-scale for all tornados that occurred during the time period of interest. DPI values were calculated for several historical tornado outbreaks, and the authors found that tornado outbreaks with many short-track tornados often received small DPI ratings. On the other hand, tornado outbreaks that produced long-track, violent tornados typically had higher potential for more destruction and, thus, received higher DPI ratings, even when there were fewer tornados overall. Thompson and Vescio's method not only incorporated two different measures (i.e., tornado intensity and damage area) but also provided a consistent way to compare different tornado outbreaks. This research aimed to improve the definition and categorization of tornados by

adopting improvements in observational technologies (e.g., WSR-88 Doppler radars) and numerical models. In general, this work represented an example of a broader, comprehensive effort undertaken by researchers in the 1990s to aid tornado forecasting and provide greater accuracy in the prediction of tornado outbreaks across the United States.

2.6 2000-2010

Work since the early 1990s substantially influenced tornado outbreak research of the first decade of the 21st century, as researchers sought a more robust understanding of meteorological environments associated with tornado outbreaks. The analyses of environmental characteristics (such as instability, shear, and storm-relative flow) and their associated role in tornado outbreak development were investigated frequently in the 2000s [135-140]. Researchers sought to find what environmental factors controlled tornado outbreaks, leading to revisions in conceptual ideas about the evolution of the outbreaks [141]. Consequently, the research results supported both high-resolution modeling [142-143] and numerical weather prediction [143-145] in operational forecasting and, thus, provided potentially useful information for forecasters to better diagnose environmental settings that supported tornado outbreaks. Scientists investigated a range of historic tornado outbreaks that spanned different spatial ranges: mesoscale [146-150], synoptic-scale [151-158], or a mix of both scales [159-161].

Further, some research was focused on high-resolution, ground-based profiling of temperature, wind, and moisture to monitor the trends in the thermodynamic state of the atmosphere during tornado outbreaks [162-164]. To improve tornado detection in real-time during outbreaks, scientists utilized diverse techniques and resources in their analyses, such as three-dimensional (3D) visualization software [165], satellite imagery [166], remote sensing [167], and enhanced-

resolution radar data [168-170]. Accelerated technological innovation led also to the development of a tornado-debris recognition method (using polarimetric radars) [171] as well as various tornado-damage estimation methods [172-174], resulting in enhanced tornado detection and assessment capabilities.

The considerable changes in prediction and reporting practices, however, contributed to a substantial increase in the number of reported tornados, especially in the lower end of the F-rating scale (i.e., F0, F1) [175]. This increase, caused by non-meteorological factors, became important to consider when defining tornado outbreaks based only on thresholds associated with the damage classification. Scientists found additional sources of potential biases and errors in the data. For instance, Verbout (2004) observed inaccuracies in the tornado reporting system, where reports for tornados prior to the mid-1970s underestimated the fraction of F1 tornados and overestimated the fraction of F2 and higher tornados [176]. Consequently, a search for another approach to identify tornado outbreaks continued.

Following the methodology of Brooks et al. (2003), who analyzed the climatology of tornado days [177], Schneider et al. (2004) investigated the climatology of 'tornado outbreak days' for the first time [178]. A 'tornado outbreak day' was defined as a calendar day when a tornado outbreak occurred. To determine if a day qualified as a 'tornado outbreak day,' a summary database with severity categories based on tornado counts was established. The database included information on the F-scale ratings, tornado counts, total path length, and 'destructive potential index,' and it identified both major and minor tornado outbreak events. Simultaneously, the same authors also analyzed 'tornado outbreak day sequences,' defined as a continuous or near-continuous sequence of tornado outbreaks [179], such as during the outbreaks from 3-11 May 2003 (also described as an extended tornado outbreak [180]).

Given these results and the absence of a means to measure "density," "importance," or "quality" (p. 1) of tornado outbreaks on a nationwide basis, Edwards et al. (2004) adopted a new strategy to assess and define tornado outbreaks [181]. As part of a broader project, the team proposed a set of criteria for tornado outbreaks, including the number of tornados, number of violent (F4+) tornados, number of significant (F2+) tornados, DPI, cumulative path length (km), and number of deaths to develop a 'tornado outbreak index' (O-index). The instances of severe weather days that were characterized by positive O-index values (i.e., O-index > 0) were classified as tornado outbreak days and ranked. The higher the ranked position, the more significant an outbreak event was. The new approach provided a useful means to compare different tornado outbreaks with each other. Despite this, the authors recognized that some elements in the analysis were subjective and a certain amount of arbitrariness in the classification was unavoidable.

Verbout (2006) also expressed that the criteria defining tornado outbreaks was dependent upon decisions of a user [10]. She analyzed tornado outbreaks through the lens of 'big tornado days,' defined as "a single day when numerous tornados and/or many tornados exceeding a specified intensity threshold were reported anywhere in the country" (p. 1). To identify a 'big tornado day,' Verbout used (1) a minimum number of reported tornados determined from a fraction of the annual value associated with a simple least-squares linear regression (i.e., the minimum depended on the year under investigation), and (2) a minimum number of F1 tornados or higher. The author used a simple least-squares linear regression to account for almost 50% of the increase in the number of reported tornados from 1950 to 2000. Many threshold combinations were possible, depending on the user's research objectives, and hence, the definition of 'big tornado days' never was completely objective [10].

Another example of an approach to define and rank tornado outbreaks was the 'Forbes impact index' [182], developed by Forbes (2006) to rank tornado outbreaks depending on their impact. In this work, a tornado outbreak was defined as an event with at least 45 tornado reports occurring in adjoining states and with no tornado-free gap of six hours or longer during the outbreak [182]. Next, 11 attributes (e.g., tornado count, number of fatalities and injuries, number of killer tornados, number of significant tornados, and amount of total damage) were used to calculate the Forbes impact index. Each attribute was assigned an integer value from 0 to 10 points (with two attributes of 5 points maximum); the sum of all points for attributes determined the ranked position of a tornado outbreak. Forbes (2006) analyzed the 15 highest-ranked outbreaks and determined the large-scale meteorological patterns for these events.

The aforementioned tornado outbreak rankings focused either on societal impacts or on meteorological significance. Yet, not all meteorologically significant events will strongly affect society; conversely, not all tornado outbreaks of large societal impact will be exceptionally unusual from a meteorological perspective. As noted by Doswell et al. (2006), meteorological and societal significance were not mutually exclusive and there was no convincing rationale of attaching greater importance just to a single criterion [9]. Hence, Doswell et al. (2006) developed a multivariate index to account for *both* meteorological and societal impact variables for tornado outbreaks, yielding a ranking that would be robust to any parameter choices. First, they identified days when seven tornados or more occurred during the period 1970-2003. Then, they ranked these days according to the linearly weighted average of eight variables (index I) that were converted to a standard normal form (with a mean of zero and a standard deviation of unity). As a result, the team produced the final 20 highest-ranked tornado outbreaks events with associated weighted variables [9]. Mercer et al. (2009) applied this ranking of tornado outbreaks in combination with

statistical modeling techniques and numerical weather prediction on the synoptic scale to test if model-predicted covariates could determine the type of severe weather outbreak that occurred [7]. This research exhibited promising results of a high probability of outbreak detection, even several days prior to the event.

The various approaches to define and classify tornado outbreaks in the literature at the beginning of the 21st century demonstrated the desire for more accurate long-term forecasts while preserving considerable flexibility depending on the user's research goals. These approaches, while distinct, still led towards further development in the understanding and prediction of tornado outbreaks.

2.7 2010-2019

In the decade of the 2010s, the literature defining tornado outbreaks mainly focused on the development of classification and ranking methods that used different statistical approaches. For example, Shafer et al. (2010) ranked all types of severe weather outbreaks (during 24-h periods) for the period 1960-2006 [183]. The authors created a group of 26 linear-weighted indices from 14 environmental and societal variables, and then, using k-means cluster analysis of the outbreak days, created a four-dimensional representation of outbreak cases. This procedure allowed them to determine five types of outbreak classifications — major tornado, wind-dominated, hail-dominated, multi-modal, and days with considerable spatial scatter of the severe reports [183] — and then ranked them across years of occurrence. This study offered a standardized approach to identify the relative severity of any type of severe weather outbreak. However, a major disadvantage to this method was the presence of "geographic scatter" (p. 1) in the tornado reports considered for given day [184]. For instance, some outbreak days included multiple reports of severe weather widely dispersed across the United States (rather than being in

a geographically compact region). To account for this issue, Shafer et al. (2011) presented a new method that examined all severe weather events based on clusters of reports using the technique of kernel density estimation [184]. This method identified groups of reports that were associated with a single weather event within a particular geographic region. The clusters obtained from the analysis were ranked using a multivariate linear-weighting method and classified into four main groups: major tornado, hail-dominant, wind-dominant, and minor mixed-mode events.

Numerous other studies focused on the quantification of the severity of tornado outbreaks. For instance, Malamud and Turcotte (2014) suggested that the statistics of tornado touchdown path lengths could serve as a quantitative measure of tornado intensity [185]. In their research, the strength of a tornado outbreak was calculated from the total path length of all "severe tornados" (i.e., individual tornados with a path length equal to or greater than 10 km) during a "convective day" (i.e., 24-h period starting at 1200 CST). Based on cumulative frequencies of the path-length data for 1952-2011, it was estimated that, on average for a single convective day, one severe tornado outbreak would reach a total path length equal or greater than 480 km per year and length equal or greater than 1200 km per decade.

Clark et al. (2013) examined tornado path lengths as related to simulated rotation (represented by updraft helicity) in convection-allowing numerical models [186]. The authors analyzed updraft helicity in relation to total tornado path lengths based on a 3D object-identification algorithm and the Pearson correlation coefficient. Results suggested that a strong relationship between both measures could be used as a predictor of the severity of a tornado outbreak. Moreover, follow-up research [187] proposed that the cumulative updraft-helicity path length could serve as an indicator of the frequency of tornado outbreaks on seasonal time scales; thus, this measure had a likely application in seasonal forecasting and regional climate modeling.

Fuhrmann et al. (2014) proposed yet another approach to assess the strength, or physical magnitude, of tornado outbreaks [188]. The team proposed a new metric that focused solely on the strength or "physical magnitude" (p. 1) of tornado outbreaks, defined as the sum of the wind force across the area of impact (i.e., the work, in Joules, generated by all tornados in the outbreak). In their context, a tornado outbreak was defined as a sequence of at least six tornados of F1 or greater intensity that had a maximum gap of 6 h between consecutive tornados in the sequence. For each tornado in the outbreak, two metrics were developed: 'hectopascal miles' (i.e., the estimated pressure exerted by the median wind speed multiplied by the tornado's path length) and 'Fujita miles*'* (i.e., the F-scale rating multiplied by the tornado's path length). The researchers summed 'hectopascal miles' and 'Fujita miles' to estimate the strength of each tornado outbreak. Then, the ratio of summed 'Fujita miles' (Fujita miles summed across each segment of the tornado track) to the number of 'Fujita miles' was multiplied by the number of 'Fujita miles' [188] to determine the 'adjusted Fujita miles' (or AFMs) for each tornado. The AFMs were used to rank the tornado outbreaks and to analyze their climatological and geographical characteristics. Because AFMs correlated with the number of fatalities (correlation coefficient = 0.80) and injuries (correlation coefficient = 0.81), the authors suggested that AFMs could be used to assess potential lethality of tornado outbreaks and to estimate the relationship between tornado outbreak exposure and the number of injuries or fatalities. Tippett and Cohen (2016) applied this definition of a tornado outbreak proposed by Fuhrmann et al. (2014) to document a significant increase in the annual variance and a gradual increase in the mean number of tornados during 1954–2014 [189]. They found that the number of tornados per outbreak was increasing (discussed also in Tippett et al. (2016) [190]) and that, as a result, tornado outbreaks might become more frequent than originally anticipated.

Gensini and Marinaro (2016) suggested that the tornado outbreak frequency might be related to the global relative angular momentum, particularly the global wind oscillation (GWO) phase [191]. Here, they defined 'tornado outbreak' as a single day when 15 tornados or more occurred east of the Rocky Mountains. Using 285 (90th percentile) out of a possible 2440 outbreak days, they showed that the frequency of tornado outbreak occurrence was significantly enhanced during periods when the time tendency of atmospheric angular momentum was negative (i.e., GWO phases 1, 2, and 8; see [191] for more information on GWO phases). Thus, the team indicated that the GWO framework might be useful in subseasonal forecasts of tornado outbreaks. Gensini et al. (2019) documented that signals from the GWO could be used to predict an extended period with favorable severe weather conditions 3-4 weeks prior to any event [192]. This article was the first to document a successfully performed forecast of an unusually active period for tornadic thunderstorms, including tornado outbreaks, at subseasonal lead times.

Finally, many researchers studied environmental conditions associated with tornado outbreaks during the 2010s [193-198]. Much of this research focused on the relationship between the tornado-outbreak occurrence and particular environmental indices [199-200] or their associated global-scale weather patterns [201-206]. For example, Mercer and Bates (2019) produced synoptic and mesoscale composite analyses to help identify major atmospheric features associated with tornado outbreak forecasts of varying quality [206]. In their research, a tornado was part of a tornado outbreak if it occurred between 1200 CST and 1159 CST (the following day) and it was located within the SPC's convective forecast region (calculated using kernel density estimation). Tornado outbreaks were combined into synoptic-scale composites using k-means cluster analysis and hierarchical clustering. Then, the differences between the outbreak environments was assessed using Weather Research and Forecasting Model [207] simulations for

three levels of tornado outbreak forecasts: high, medium, and low quality. Results indicated that the medium- and low-quality forecasts lacked key kinematic characteristics that were present in the high-quality forecasts. This work provided valuable feedback on important characteristics of meteorological fields to inform forecasters and to improve the quality of tornado outbreak prediction.

This final decade demonstrated examples of dynamical and innovative approaches to define tornado outbreaks. The changes to techniques and methods across the decades were driven by improvements in both knowledge and technologies that helped to overcome initial barriers and challenges associated with the concept of a tornado outbreak.

3. An example of a statistical approach to a tornado outbreak definition

Recently published work [192, 206] suggests that, despite successful progress in research and various attempts to define and rank tornado outbreaks, our understanding of those severe events can and should be improved. Accordingly, we propose a statistical approach to define a ‘tornado outbreak.’ This approach will help to identify historic cases from the tornado-report database of the Storm Prediction Center that can be used in any type of analysis that requires tornado outbreak quantification or identification. Our method represents a data driven approach that utilizes statistical method of kernel density estimation (KDE). The goal of this work is to develop a simple method to create a tornado-outbreak database for use in future studies on these outbreaks.

3.1 Datasets

To define 'tornado outbreaks,' we used the severe weather database of the Storm Prediction Center (SPC), an operational unit of the National Oceanic and Atmospheric Administration's National Weather Service [208]. This database includes both tornado reports as well as other non-tornadic reports related to severe thunderstorms. It specifies the type of event, number of events reported each day, associated damage rating (Fujita scale), and the event's geographic characteristics and societal impacts, such as the event location, number of casualties, and estimated cost of damages [12]. We extracted the following information from the database: year, month, day, and hour of each tornado report as well as the latitude and longitude of the tornado track. We used tornado reports from 1 January 1950 to 31 December 2017 that were kept in SPC's original time format of Central Standard Time (CST).

The domain for this research is the contiguous United States (northwest corner at 55°N, 140°W to the southeast corner at 20°N, 60°W). For mapping purposes, we represent the geographical center of a tornado outbreak region as a single point, calculated as a mathematical mean of longitudes and latitudes of all tornados in the cluster. The clusters are groups of tornado reports associated with a single tornado outbreak and identified based on the nonparametric method of kernel density estimation (see *Methods* section). U.S. boundaries are from the Census Bureau's 2017 Cartographic Boundary Files [209] at 20-m (1:20,000,000) resolution, and we plot the data using the Albers equal-area conic map projection.

3.2 Methods

To accomplish the goal of tornado outbreak identification, our analysis focused on clustering of historic tornados that occurred within a 24-hour time period within a constrained region. Tornado outbreaks that lasted beyond 24 h were classified into independent days, rather than considered as one event.

First, we limited possible outbreaks to those with significant tornados — those greater or equal to F2 that occurred in the contiguous 48 states of the U.S. [1]. Significant tornados have been shown to have the greatest societal impact, such as structure damage or causalities [210], and to be associated with a more stable record in the SPC database over time [175].

Next, we assigned each significant tornado an associated tornado day — a 24-hour period from 0600 CST to 0559 CST — based on the method of Shafer and Doswell (2011) [188]. Then, to create a database of tornado outbreaks based on geographical locations of tornados in coherent clusters, we applied kernel density estimation (KDE) on each tornado day [188, 211]. In the one-dimensional KDE technique, the probability density function at a particular point can be represented as the following:

$$f(x) = \frac{1}{n}\sum_{i=1}^{n} K_h\left(x - x_i\right) \tag{1}$$

Here, n represents the number of significant tornado reports on a given day, K_h (a kernel function) is a non-negative function that describes a continuous probability distribution, and parameter h represents a tunable smoothing parameter (bandwidth) and is a measure of uncertainty associated

with the distance from a given point Also, in equation (1), x represents the grid point position and x_i represents the location of an individual tornado report. Accordingly, the ($x-x_i$) represents the distance (D_i in equation 2) between the observed tornado report and a given grid points. In this study, using latitude and longitude, we apply a multivariate version of KDE, shown in Equation 2. As is typical for two-dimensional analyses, the kernel function implemented is Gaussian [212].

$$f(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2\pi h^2}\exp\left[\frac{-D_i^2}{2h^2}\right]. \quad (2)$$

To calculate the particular geographic region for each outbreak cluster, we applied the smoothing parameter h = 1 and the PDF threshold probability = 0.001 (i.e., the outermost contour) following Shafer and Doswell (2011) [184]. They tested and selected these particular values to include the most tornado reports within the cluster while allowing regionally distinct events (for more information, see [184]). Fig. 1 presents an example of a single outbreak, clustered using the method described.

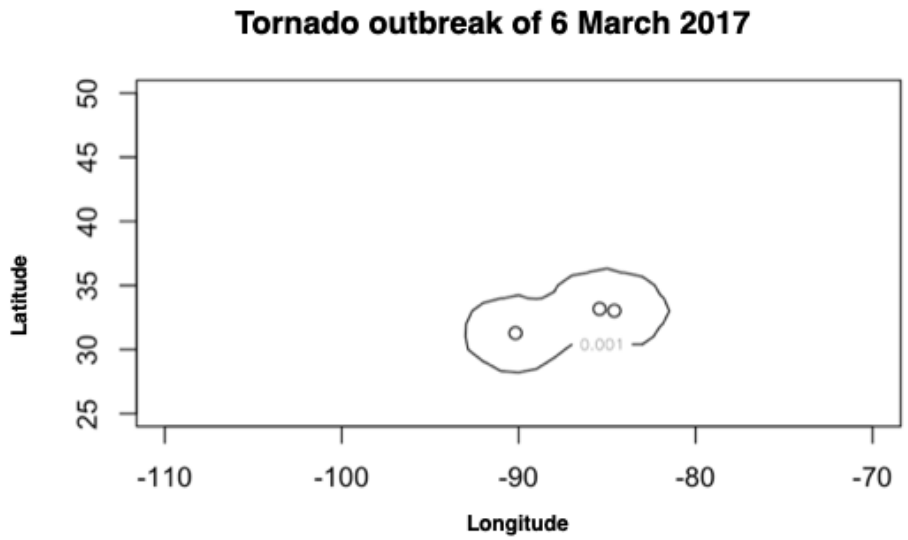

Fig. 1. Example of a computed cluster of significant tornadoes using multivariate KDE with h=1 and PDF threshold=0.001 (after Shafer and Doswell (2011)). The cluster represents the outbreak from 6 March 2017, starting at 1800 CST and ending at 0000 CST.

Additionally, we calculated the start and end times of each tornado outbreak (cluster). To reduce computational processing, times were calculated in 3-h increments centered on (i.e., 1.5 h before and after) 0000, 0300, 0600, 0900, 1200, 1500, 1800, and 2100 CST. In that sense, the outbreak start time was designated as the 3-h block that contained the first tornado report in the cluster, and the outbreak end time was specified as the 3-hour block that contained the last report in the cluster. If there were multiple tornado outbreaks on any given day, multiple outbreak start and end times were recorded in the database.

Finally, we summed the total number of tornado reports for each cluster on each tornado day, then computed an annual mean of tornados per cluster. Following Shafer and Doswell (2011) [184], we detrended the outbreaks over time by removing the clusters with fewer reports than their associated

annual mean. The result filtered out clusters with relatively scattered coverage or with a small number of significant tornados.

4. Results

From 1950 to 2017, 12,210 significant tornados occurred, with most of them east of the Rocky Mountains and the highest density in the Great Plains and Southeast United States (Fig. 2).

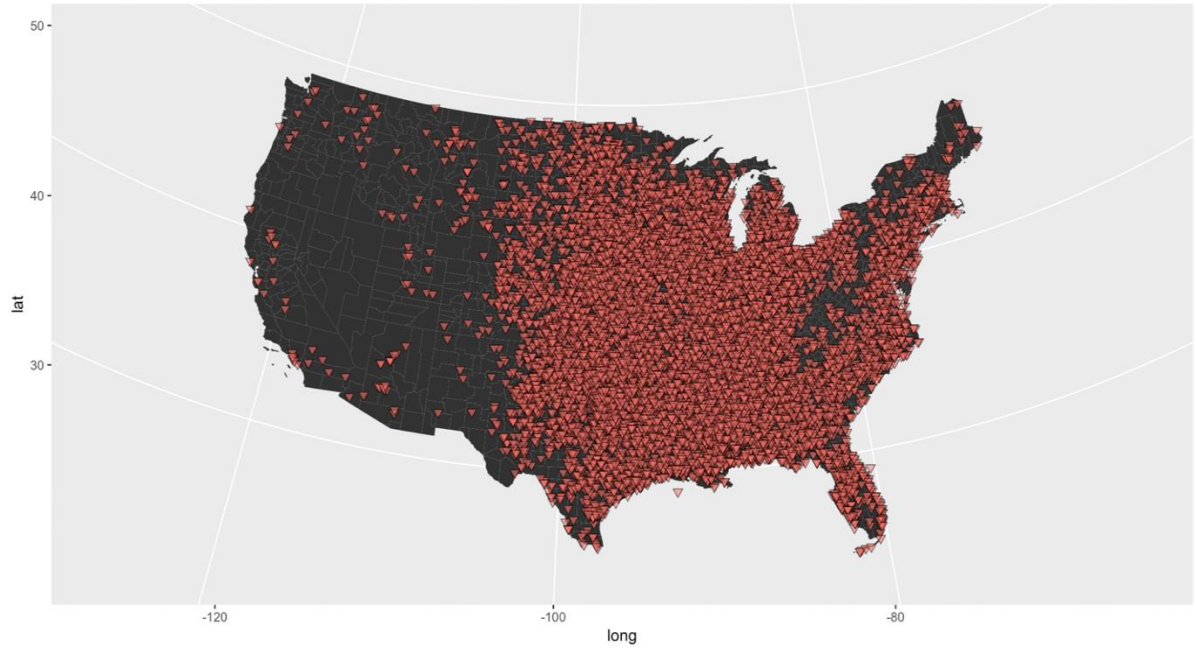

Fig. 2. Distribution of significant tornados (F2 and higher on the Fujita scale) from 1950 to 2017 using the severe weather database of the Storm Prediction Center.

The most common type of outbreak was one with three (336 outbreaks) or four (241) significant tornados (Fig. 3). This technique verified the two largest outbreaks of significant tornadoes — the

1974 Super Outbreak (3 April 1974), with 96 tornados of F2 and larger, and the 2011 Super Outbreak (27 April 2011), with 51 significant tornados during a single 24-hour period.

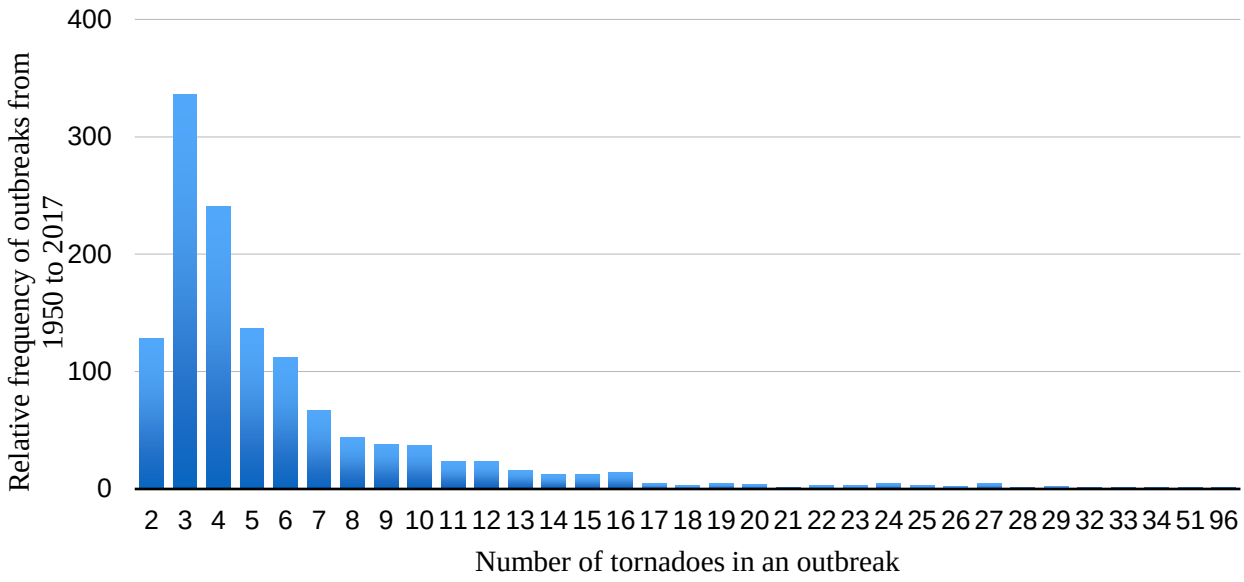

Fig. 3. Frequency distribution of tornado outbreaks from 1950-2017. Outbreaks were clustered using kernel density estimation.

Following this method, we identified 4,991 clusters (tornado outbreaks) over the 68-year record, or an average of 73 tornado outbreaks per year. Some clusters contained only a few significant tornados (Fig. 3), which many would argue should not be considered an outbreak (see Section 2). Doswell et al. (2006) [9] used a threshold of seven tornados to define major tornado outbreaks, resulting in the top five outbreaks, on average, each year. Appendix A displays the annual average number of tornado outbreaks by number of tornados per cluster. For instance, a threshold of 5 tornados or more in an outbreak resulted in an average of 9 outbreaks per year; a threshold of 11 tornados yielded an average of 2 tornado outbreaks per year. Although subjective, we followed Doswell et al. (2006) [9], with a threshold of seven tornados in an outbreak to define our ‘major

tornado outbreak,' and we focused on these outbreaks for the remaining research. Using this threshold, 333 major tornado outbreaks spanned 1950-2017 (Fig. 4).

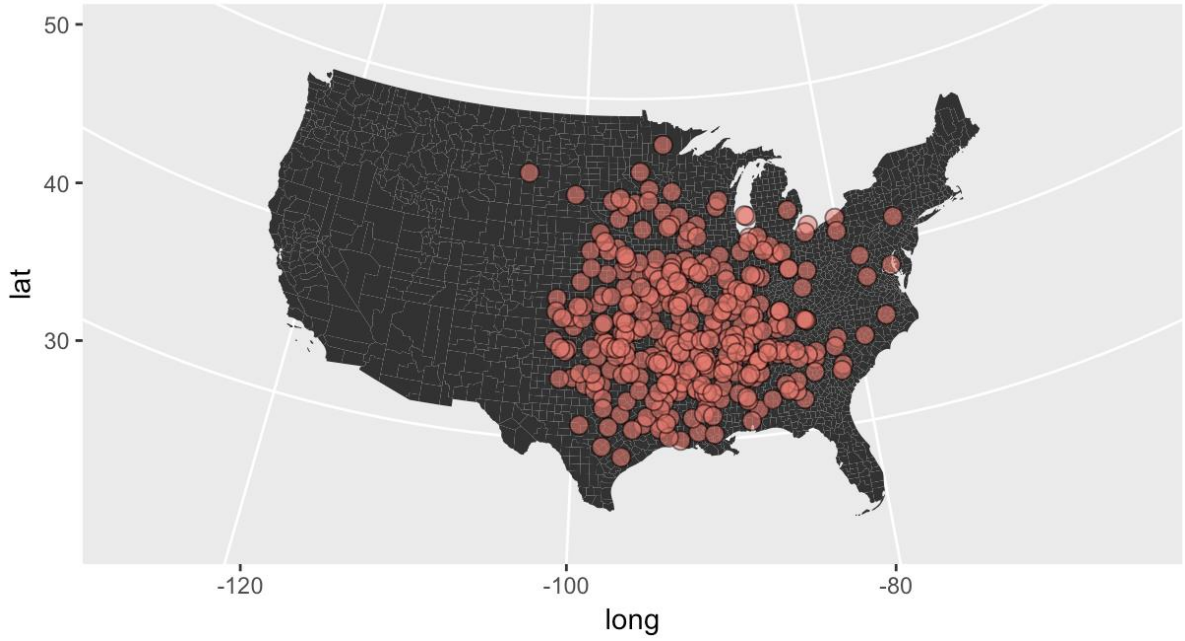

Fig. 4. Spatial distribution of the cluster centers of 333 major tornado outbreaks that spanned 1950-2017, based on KDE analysis.

Most of the major tornado outbreaks occurred in the south-central and southeastern U.S. There were no major outbreaks west of the Rocky Mountains, as tornadic storms can be affected by topography [213-219]. To support the large buoyant instability necessary for the development of tornadic storms, boundary-layer moisture has to be present. The Gulf of Mexico provides an abundant source of warm water that can significantly modify air masses over the eastern half of the U.S., also contributing to the formation of deep surface cyclones (one of key features for tornado outbreak development) [220]. Additionally, there are other synoptic conditions that are key to the tornado outbreak formation, such as an upper-level jet streak collocated with the center

of rapid surface pressure or upper-level trough axis located west of the outbreak, but such considerations are beyond the scope of this research (for more information, see [139, 151, 221]).

Figure 5 displays a total count of major tornado outbreaks by month, with two peaks clearly distinguished. The largest peak in the distribution occurred in May, with 89 major tornado outbreaks, and was encompassed by April, March, and June, with 79, 35, and 33 major tornado outbreaks, respectively. On average, at least one outbreak occurred in May and in April each year. A second peak was observed in November, December, and October, with 26, 18, and 12 outbreaks, respectively.

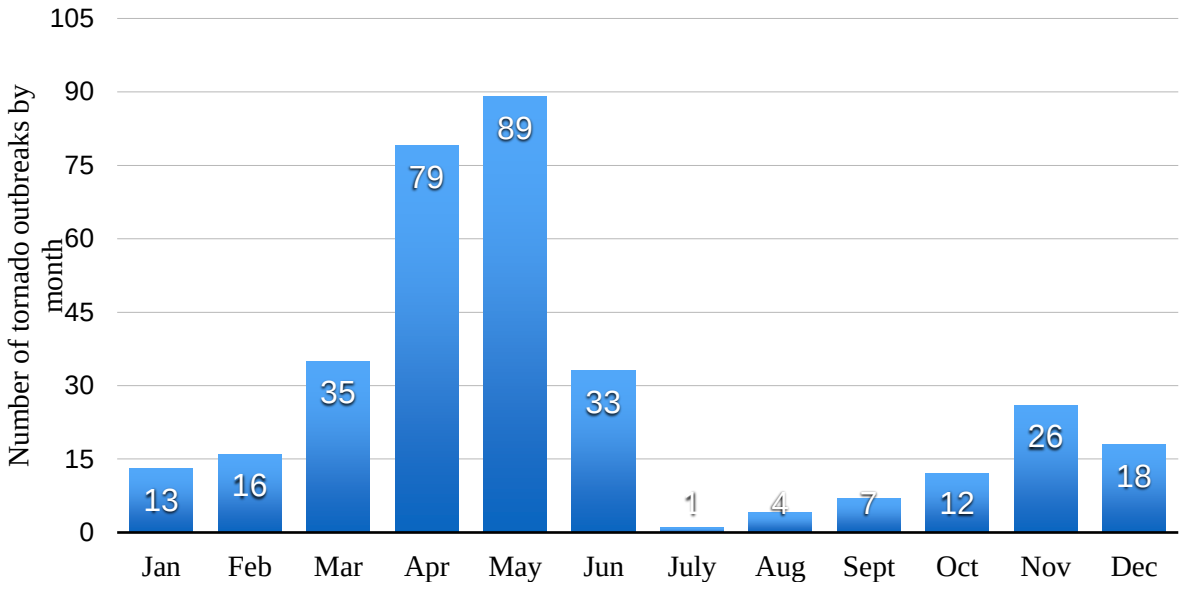

Fig. 5. Distribution of major tornado outbreaks by month from 1950 to 2017.

5. Conclusion and Discussion

The primary goal of this work was to identify historic cases of major tornado outbreaks. The process of identification, however, was not straightforward, as there had been considerable ambiguity in the term 'tornado outbreak.' Over the years, the term was interchangeably attributed to many types of severe weather events, sometimes regardless of whether the event included tornado reports. As noted by Galway (1977), "A tornado 'outbreak' can mean many things to many people" [49, p. 477]. Our extensive review of the evolving nature of tornado outbreak definitions and classifications, present in the literature since the early 1950s, served as contextual background for our data-driven approach to identify major tornado outbreaks. It also served to demonstrate the challenges of defining tornado outbreaks as technologies change.

To determine groups of tornado outbreaks, tornado reports from the National Weather Service's Storm Prediction Center database were used. All significant tornados (F2 and higher) from 1950 to 2017 were grouped into 24-hour periods, from 0600 CST to 0559 CST and then assigned into clusters using kernel density estimation analysis. We removed clusters that had fewer reports than their associate annual mean, resulting in 4991 clusters of significant tornados. We defined a 'major tornado outbreak' as one with seven or more tornados, based on our analysis of tornado frequency distribution and the method of Shafer and Doswell (2011) [184]. As a result, we found 333 major tornado outbreaks, all east of the Rocky Mountains, during the 68-year period. Thus, there was an average of 5 major tornado outbreaks per year, typically with one each in April and May somewhere in the United States.

This research uses kernel density estimation to create a tornado-outbreak database for use in future studies. Like any study, there are some limitations. First, any available database that documents quantitative information about tornados, including one used in this research, has inherent errors that affect the accuracy and reliability of the data [10, 67]. The problems with these databases arise from changes in tornado reporting methods, changes to rating and damage survey procedures, introduction of new technologies to detect and record tornados, and changes in population or population density, to name few. Additional errors may be associated with uncertainties in applying the Fujita scale to a tornado damage path, especially in natural and rural environments, resulting in inconsistency in ratings across the historic record [12]. Despite the fact that none of these issues can be solved completely, some measures to reduce the bias, such as detrending by year (see Section 3.2), were applied.

Second, there are limitations to applying the clustering technique to events that lasted beyond 24 h. For example, if an event began at 0500 CST and lasted until 0900 CST, it would be classified as two separate tornado outbreaks, leading to an underestimation of the total number of tornado outbreaks in the final database. Such events are rare, however, and most likely would not significantly influence our analysis results. Finally, the choice of the threshold for the number of significant tornados in the outbreak was subjective, thus affecting our results. However, as noted by Verbout (2006) [10, p. 93], "all outbreak definitions require some degree of subjectivity." This method, although not the only way to identify historic tornado outbreaks, offers a generally data-driven approach that can be modified depending on investigator's research goals.

It would be helpful to conduct a sensitivity analysis on the number of reports within outbreaks, the magnitudes of the individual tornados within outbreaks, and the threshold values for the yearly frequency distribution of tornado outbreak clusters. If the number of major tornado outbreaks were

to decrease over time, will the corresponding number of reports within those outbreaks also decrease? What are the characteristics of the magnitude of individual reports within those outbreaks? Meanwhile, results presented in this work will be used for the analysis of synoptic-scale atmospheric patterns which are associated with major tornado outbreaks.

6. References

1. Doswell, C. A., C.A Doswell III, and D.M. Schultz, 2006: On the use of indices and parameters in forecasting severe storms. *Electronic J. Severe Storms Meteor*., **1**, 1–22.
2. Schneider, R S., A.R. Dean, S.J. Weiss, and P.D. Bothwell, 2006: Analysis of estimated environments for 2004 and 2005 severe convective storm reports. *Preprints, 23rd Conf. Severe Local Storms*, Vol. 3 of, St Louis, MO, Amer. Meteor. Soc., 3.5.
3. Brooks, H. E., 2004: On the relationship of tornado path length and width to intensity. *Weather Forecast.*, **19**, 310–319.
4. Doswell III, C.A., and D. Schultz, 2006: On the use of indices and parameters in forecasting severe storms. *Electronic. J. Severe Storms Meteor.*, **1**, 1–22.
5. *American Meteorological Society cited 2019: Tornado outbreak. Glossary of Meteorology. [Available online at http://glossary.ametsoc.org/wiki/Tornado_outbreak]*.
6. Galway, J.G., 1977: Some Climatological Aspects of Tornado Outbreaks. *Mon. Weather Rev.*, **105**, 477–484.
7. Mercer, A.E., C.M. Shafer, C.A. Doswell, L.M. Leslie, and M.B. Richman, 2009: Objective classification of tornadic and nontornadic severe weather outbreaks. *Mon. Weather Rev.*, **137**, 4355–4368.
8. Fujita, T.T., 1974: Jumbo tornado outbreak of 3 April 1974. *Weatherwise*, **27**, 116–126.

9. Doswell, C.A., R. Edwards, R.L. Thompson, J.A. Hart, and K.C. Crosbie, 2006: A simple and flexible method for ranking severe weather events. *Weather Forecast.,* **21**, 939–951.

10. Verbout, S. M., H. E. Brooks, L. M. Leslie, and D. M. Schultz, 2006: Evolution of the U.S. Tornado Database: 1954–2003. *Weather Forecast.*, **21**, 86–93.

11. Fujita, T. T., 1981: Tornadoes and Downbursts in the Context of Generalized Planetary Scales. *J. Atmos. Sci.*, **38**, 1511–1534.

12. Doswell, C. A., H. E. Brooks, and N. Dotzek, 2009: On the implementation of the enhanced Fujita scale in the USA. *Atmos. Res.*, **93**, 554–563.

13. Brooks, H. E., and N. Dotzek, 2008: The spatial distribution of severe convective storms and an analysis of their secular changes. *Climate Extremes and Society*, H.F. Diaz and R.J. Murnane, Eds., Cambridge University Press, 35–53.

14. Brooks, H. E., G. W. Carbin, and P. T. Marsh, 2014: Increased variability of tornado occurrence in the United States. *Science (80)*, **346**, 349–352.

15. Diffenbaugh, N. S., R. J. Trapp, and H. Brooks, 2008: Does Global Warming Influence Tornado Activity? *Eos, Transactions American Geophysical Union*, *89*(53), 553-554.

16. Doswell III, C. A., G. W. Carbin, and H. E. Brooks, 2012: The tornadoes of spring 2011 in the USA: an historical perspective. *Weather*, **67**, 88–94.

17. Tippett, M. K., 2014: Changing volatility of U.S. annual tornado reports. *Geophys. Res. Lett.*, **41**, 6956–6961.

18. Fawbush, E. J., R. C. Miller, and L. G. Starrett, 1951: An Empirical Method of Forecasting Tornado Development. *Bull. Am. Meteorol. Soc.*, **32**, 1–9.

19. Carr, J. A., 1952: A Preliminary Report on the Tornadoes of March 21-22, 1952. *Mon. Weather Rev.*, **80**, 50–58.

20. Flora, S. D., 1953: *Tornadoes of the United States.* [1st ed.]. Norman, University of Oklahoma Press, 194

21. Van Tassel, E. L., 1955: The North Platte Valley Tornado Outbreak of June 27, 1955. *Mon. Weather Rev.*, **83**, 255–264.

22. Beebe, R. G., 1956: Tornado Composite Charts. *Mon. Weather Rev.*, **84**, 127–142.

23. Whiting, R. M., and R. E. Bailey, 1957: Some Meteorological Relationships in the Prediction of Tornadoes. *Mon. Weather Rev.*, **85**, 141–150.

24. Beebe, R. G., 1958: Tornadoes During 1957. *Weatherwise*, **11**, 13–38.

25. Ludlum, D. M., 1959: Drought Features Early Spring. *Weatherwise*, **12**, 126–132.

26. Beebe, R. G., 1959: Tornadoes During 1958. *Weatherwise*, **12**, 13–16.

27. Smith, W., 1959: Mesoanalysis of a Tornado-Producing Situation in Texas, 25–26 May 1955. *Bull. Am. Meteorol. Soc.*, **40**, 134–145.

28. Wolford, L. V, 1960*: Tornado occurrences in the United States. Tech Paper No 20*. US Weather Bureau, 12.

29. Beebe, R. G., 1961: Tornadoes During 1960. *Weatherwise*, **14**, 16–29.

30. Ludlum, D. M., 1961: Some Noteworthy Snows of 1959–60. *Weatherwise*, **14**, 24–29.

31. Hardy, W. E., 1962: Tornadoes During 1961. *Weatherwise*, **15**, 16–41.

32. Hardy, W. E., 1963: Tornadoes during 1962. *Weatherwise*, **16**, 22–48.

33. Galway, J. G., 1966: The Topeka Tornado of 8 June 1966. *Weatherwise*, **19**, 144–149.

34. O'Connor, J. F., 1965: The Weather and Circulation of April 1965. *Mon. Weather Rev.*, **93**, 465–473.

35. Bradbury, D.L. and T.T. Fujita, T.T., 1966: Features and Motions of Radar Echoes on Palm Sunday, 1965. *Satellite and Mesometeorology Research Project*, Department of the Geophysical Sciences, Chicago University, Illinois.

36. Andrews, J. F., 1966: The Circulation and Weather of 1965. *Weatherwise*, **19**, 4–23.

37. Pearson, A. D., 1968: Tornado Warnings Over Radio and Television. *Bull. Am. Meteorol. Soc.*, **49**, 361–363.

38. Fujita, T., 1967: Mesoscale aspects of orographic influences on flow and precipitation patterns (No. SMRP-RP-67), *Satellite and Mesometeorology research Project*, Chicago University, Illinois.

39. Agee, E. M., 1969: The climatology of Indiana tornadoes. *Proceedings of the Indiana Academy of Science*, Vol. 79 of, 299–308.

40. Fujita, T. T., D. L. Bradbury, and C. F. Van Thullenar, 1970: Palm Sunday Tornadoes of April 11, 1965. *Mon. Weather Rev.*, **98**, 29–69.

41. Pautz, M.E., 1969: Severe local storm occurrences 1955-1967 (Vol. 12). Environmental Science Services Administration, Weather Bureau.

42. Fujita, T.T., 1971: Proposed characterization of tornadoes and hurricanes by area and intensity. *University of Chicago SMRP Research Paper*, 42.

43. Fujita, T. T., 1974: Jumbo Tornado Outbreak of 3 April 1974. *Weatherwise*, **27**, 116–126.

44. 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. Weather Rev.*, **103**, 318–333.

45. Hoxit, L. R., and C. F. Chappell, 1974: *Tornado outbreak of April 3–4, 1974: Synoptic analysis. NOAA Tech. Rep. ERL 338-APCL 37*. 48

46. Purdom, J. F. W., 1974: From Above: The 3 April 1974 Tornado Outbreak. *Weatherwise*, **27**, 120–121.

47. Maddox, R. A., and W. M. Gray, 1973: Study of tornado proximity data and an observationally derived model of tornado genesis, A. *Atmos. Sci. Pap. no. 212*.

48. Galway, J. G., 1975: Relationship of Tornado Deaths to Severe Weather Watch Areas. *Mon. Weather Rev.*, **103**, 737–741.

49. Galway, J. G., 1977: Some Climatological Aspects of Tornado Outbreaks. *Mon. Weather Rev.*, **105**, 477–484.

50. McNulty, R.P., D.L. Kelly, and J.T. Schaefer, 1979: January. Frequency of tornado occurrence. *In* Preprints, 11th Conf. on Severe Local Storms, Kansas City, MO, Amer. Meteor. Soc., 222-226.

51. Schaefer, J. T., D. L. Kelley, C. A. Doswell III, J. G. Galway, R. J. Williams, R. P. McNulty, L. R. Lemon, and B. D. Lambert, 1980: Tornadoes When Where How Often. *Weatherwise*, **33**, 52–59.

52. Schaefer, J.T., Kelly, D.L. and Abbey Jr, R.F., 1980: Tornado track characteristics and hazard probabilities. *In Wind Engineering*, Pergamon, 95-109.

53. Reinhold, T.A. and B. Ellingwood, 1982: Tornado damage risk assessment (No. NUREG/CR--2944). National Bureau of Standards.

54. Twisdale, L. A., and W. L. Dunn, 1983: Probabilistic Analysis of Tornado Wind Risks. *J. Struct. Eng.*, **109**, 468–488.

55. Grazulis, T.P., 1984: Violent tornado climatography, 1880-1982 (No. NUREG/CR-3670; PNL-5006). *Environmental Films*, Inc., St. Johnsbury, VT (USA).

56. Grazulis, T. P., and R. F. Abbey Jr, 1983: 103 years of violent tornadoes: Patterns of serendipity, population, and mesoscale topography. *Preprints, 13th Conf. on Severe Local Storms*, Tulsa, OK, Amer. Meteor. Soc., 124–127.

57. Kelly, D. L., J. T. Schaefer, R. P. McNulty, C. A. Doswell, and R. F. Abbey, 1978: An Augmented Tornado Climatology. *Mon. Weather Rev.*, **106**, 1172–1183.

58. Tecson, J.J., T.T. Fujita, and R.F. Abbey, 1979: Statistics of US tornadoes based on the DAPPLE (Damage Area Per Path Length) tornado tape. Amer. Meteor. Soc.

59. Galway, J. G., 1977: Some Climatological Aspects of Tornado Outbreaks. *Mon. Weather Rev.*, **105**, 477–484.

60. Minor, J.E., J.R. McDonald, and K.C. Mehta, 1977: January. The tornado: An engineering-oriented perspective. NOAA Tech. Memo. ERL NSSL-82, NTIS Accession No. PB-281860/AS., 196.

61. Schaefer, J. T., and J. G. Galway, 1982: Population biases in tornado climatology. 12th Conf. on *Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc.

62. Colquhoun, J. R., and D. J. Shepherd, 1985: The relationship between tornado intensity and the environment of its parent severe thunderstorm. *Preprints, 14th Conf. Severe Local Storms*, Indianapolis, IN, Amer. Meteor. Soc., 1–4.

63. Schaefer, J. T., D. L. Kelly, and R. F. Abbey, 1986: A Minimum Assumption Tornado-Hazard Probability Model. *J. Clim. Appl. Meteorol.*, **25**, 1934–1945.

64. Tecson, J. J., T. T. Fujita, and R. F. Abbey Jr, 1982: Climatological mapping of US tornadoes during 1916–1980. *Preprints, 11th Conf. on Severe Local Storms*, Kansas City, MO, Amer. Meteor. Soc., 38–41.

65. Tescon, J. J., T. T. Fujita, and R. F. Abbey Jr, 1983: Statistical analyses of US tornadoes based on the geographic distribution of population, community, and other parameters. *Preprints, 13th Conf. on Severe Local Storms*, Tulsa, OK, Amer. Meteor. Soc., 120–123.

66. Fujita, T. T., 1981: Tornadoes and Downbursts in the Context of Generalized Planetary Scales. *J. Atmos. Sci.*, **38**, 1511–1534.

67. Doswell, C. A., and D. W. Burgess, 1988: On Some Issues of United States Tornado Climatology. *Mon. Weather Rev.*, **116**, 495–501.

68. Huschke, R. E., 1959: *Glossary of meteorology*. American Meteorological Society, 638

69. Moller, A. R., 1978: The Improved NWS Storm Spotters' Training Program at Ft. Worth, Tex. *Bull. Am. Meteorol. Soc.*, **59**, 1574–1582.

70. Burgess, D. W., and R. J. Donaldson, 1979: Contrasting tornadic storm types. *Preprints, 11th Conf. on Severe Local Storms*, Kansas City, MO, Amer. Meteor. Soc., 189–192.

71. Burgess, D. W., and R. P. Davies-Jones, 1979: Unusual Tornadic Storms in Eastern Oklahoma on 5 December 1975. *Mon. Weather Rev.*, **107**, 451–457.

72. Bluestein, H. B., 1980: The University of Oklahoma Severe Storms Intercept Project—1979. *Bull. Am. Meteorol. Soc.*, **61**, 560–567.

73. Wilson, J., R. Carbone, H. Baynton, and R. Serafin, 1980: Operational Application of Meteorological Doppler Radar. *Bull. Am. Meteorol. Soc.*, **61**, 1154–1168.

74. Forbes, G. S., and R. M. Wakimoto, 1983: A Concentrated Outbreak of Tornadoes, Downbursts and Microbursts, and Implications Regarding Vortex Classification. *Mon. Weather Rev.*, **111**, 220–236.

75. Schaefer, J. T., and I. Doswell, Charles A., 1984: Empirical Orthogonal Function Expansion Applied to Progressive Tornado Outbreaks. *J. Meteorol. Soc. Japan. Ser. II*, **62**, 929–936.

76. Galway, J. G., and A. Pearson, 1981: Winter Tornado Outbreaks. *Mon. Weather Rev.*, **109**, 1072–1080.

77. Galway, J. G., 1981: Ten Famous Tornado Outbreaks. *Weatherwise*, **34**, 100–109.

78. Witten, D. E., 1985: May 31, 1985 a Deadly Tornado Outbreak. *Weatherwise*, **38**, 193–198.

79. Myers, C.H., 1987: An overview of the June 7, 1984 Iowa tornado outbreak. *NOAA Technical Memorandum, NWS CR-84*, Des Moines, Iowa.

80. Ostby, F. P., and L. F. Wilson, 1981: Tornado! *Weatherwise*, **34**, 26–32.

81. Ostby, F. P., and L. F. Wilson, 1980: Tornado! *Weatherwise*, **33**, 31–35.

82. Moore, J. T., and M. F. Squires, 1982: Ageostrophic winds and vertical motion fields accompanying upper level jet streak propagation during the Red River Valley tornado outbreak. *Preprints, 9th Conf. of Weather Analysis and Forecasting*, Seattle, WA, Amer. Meteor. Soc., 424–429.

83. Homan, J. H., and D. G. Vincent, 1983: Mesoscale Analysis of Surface Variables During the Severe Storm Outbreak of 10–11 April 1979. *Mon. Weather Rev.*, **111**, 1122–1130.

84. Ferretti, R., F. Einaudi, and L. W. Uccellini, 1988: Wave disturbances associated with the red river valley severe weather outbreak of 10-11 April 1979. *Meteorol. Atmos. Phys.*, **39**, 132–168.

85. Doswell III, C.A., 1985. The Operational Meteorology of Convective Weather. Volume 2. Storm Scale Analysis (No. AWS/TN-85/001). Air Weather Service Scott AFB, Illinois.

86. Moore, J. T., and H. E. Fuelberg, 1981: A Synoptic Analysis of the First AVE-SESAME ’79 Period. *Bull. Am. Meteorol. Soc.*, **62**, 1577–1590.

87. Alberty, R., D. Burgess, and T. Fujita, 1980: Severe weather events of 10 April 1979. *Bull. Am. Meteorol. Soc.*, **61**, 1033–1034.

88. Kocin, P. J., L. W. Uccellini, and R. A. Petersen, 1986: Rapid evolution of a jet streak circulation in a pre-convective environment. *Meteorol. Atmos. Phys.*, **35**, 103–138.

89. Belt, C. L., and H. E. Fuelberg, 1982: The Effects of Random Errors in Rawinsonde Data on Derived Kinematic Quantities. *Mon. Weather Rev.*, **110**, 91–101.

90. Vincent, D. G., and J. H. Homan, 1983: Mesoscale Analysis of Pressure and Precipitation Patterns during AVE-SESAME 1979, 10–11 April. *Bull. Am. Meteorol. Soc.*, **64**, 23–28.

91. Fuelberg, H. E., and G. J. Jedlovec, 1982: A Subsynoptic-Scale Kinetic Energy Analysis of the Red River Valley Tornado Outbreak (AVE-SESAME I). *Mon. Weather Rev.*, **110**, 2005–2024.

92. Bluestein, H. B., E. W. McCaul, G. P. Byrd, and G. R. Woodall, 1988: Mobile Sounding Observations of a Tornadic Storm near the Dryline: The Canadian, Texas Storm of 7 May 1986. *Mon. Weather Rev.*, **116**, 1790–1804.

93. Bluestein, H. B., and K. W. Thomas, 1984: Diagnosis of a Jet Streak in the Vicinity of a Severe Weather Outbreak in the Texas Panhandle. *Mon. Weather Rev.*, **112**, 2499–2520.

94. Hung, R. J., and Y. D. Tsao, 1987: Study of pre-storm environment by using rawinsonde and satellite observations. *Int. J. Remote Sens.*, **8**, 1123–1150.

95. Grazulis, T. P., 1990: Significant tornadoes.....A 107 year perspective. *J. Wind Eng. Ind. Aerodyn.*, **36**, 131–151.

96. Grazulis, T.P., 1993: A 110-year perspective of significant tornadoes. *Washington DC American Geophysical Union Geophysical Monograph Series*, 79, 467-474.

97. Grazulis, T.P., 1993: Significant Tornadoes, 1680–1991, Environmental Films, St. Johnsbury, Vermont, 13

98. Spector, D.A., 1993: Investigation of large tornado outbreaks in the United States (Doctoral dissertation), Texas Tech University, [online] Available: http://hdl.handle.net/2346/60885

99. Johns, R. H., and C. A. Doswell, 1992: Severe Local Storms Forecasting. *Weather Forecast.*, **7**, 588–612.

100. Purdom, J. F. W., and J. F. Weaver, 1990: A satellite perspective of the June 15, 1988 tornado outbreak in Denver, Colorado. *Preprints, I6th Conf on Severe Local Storms*, Kananaskis Park, Alberta, Canada, Amer. Meteor. Soc., 167–170.

101. Szoke, E.J. and R. Rotunno, 1993: A comparison of surface observations and visual tornado characteristics for the June 15, 1988, Denver tornado outbreak. *Washington DC American Geophysical Union Geophysical Monograph Series*, 79, 353-366.

102. Guerrero, H., and W. Read, 1993: Operational use of the WSR-88D during the November 21, 1992 southeast Texas tornado outbreak. *Preprints, 17th Conf. on Severe Local Storms*, St Louis, MO, Amer. Meteor. Soc., 399–402.

103. Davies, J. M., C. A. Doswell, D. W. Burgess, and J. F. Weaver, 1994: Some Noteworthy Aspects of the Hesston, Kansas, Tornado Family of 13 March 1990. *Bull. Am. Meteorol. Soc.*, **75**, 1007–1017.

104. Kaplan, M. L., Y. L. Lin, D. W. Hamilton, and R. A. Rozumalski, 1998: The Numerical Simulation of an Unbalanced Jetlet and Its Role in the Palm Sunday 1994 Tornado Outbreak in Alabama and Georgia. *Mon. Weather Rev.*, **126**, 2133–2165.

105. Hamilton, D. W., Y. L. Lin, R. P. Weglarz, and M. L. Kaplan, 1998: Jetlet Formation from Diabatic Forcing with Applications to the 1994 Palm Sunday Tornado Outbreak. *Mon. Weather Rev.*, **126**, 2061–2089.

106. Lee, B. D., and R. B. Wilhelmson, 1997: The Numerical Simulation of Nonsupercell Tornadogenesis. Part II: Evolution of a Family of Tornadoes along a Weak Outflow Boundary. *J. Atmos. Sci.*, **54**, 2387–2415.

107. Adlerman, E. J., K. K. Droegemeier, and R. Davies-Jones, 1999: A Numerical Simulation of Cyclic Mesocyclogenesis. *J. Atmos. Sci.*, **56**, 2045–2069.

108. Doswell III, C. A., S. J. Weiss, and R. H. Johns, 1993: ornado forecasting: A review. *The Tornado: Its Structure, Dynamics, Prediction, and Hazards, Geophys. Monogr, 79*, 557-571

109. Davies-Jones, R. P., 1993: Helicity trends in tornado outbreaks. *Preprints, 17th Conf. on Severe Local Storms*, St Louis, MO, Amer. Meteor. Soc., 56–60.

110. Uccellini, L., 1990: The relationship between jet streaks and severe convective storm systems. *16th Conference on Severe Local Storms/Conference on Atmospheric Electricity*, Kananaskis Park, Alberta, Canada, American Meteorological Society, 121–130.

111. July, M., 1990: Forcing factors in the violent tornado outbreak of May 5, 1989: A study of scale interaction. *Preprints, 16th Conf. on Severe Local Storms*, Kananaskis Park, Alberta, Canada, Amer. Meteor. Soc., 72–77.

112. LaPenta, K. D., R. J. Kane, and J. S. Waldstreicher, 1990: A multiscale examination of the 10 July 1989 northeast tornado outbreak. *Preprints, 16th Conf. Severe Local Storms*, Kananaskis Park, Alberta, Canada, Amer. Meteor. Soc., 548–553.

113. Manuel, P., and M. Delisi, 1993: Multiscale Evaluation of the 2 June 1990 Tornado Outbreak. *Preprints, 17th Conf. on Severe Local Storms*, St Louis, MO, Amer. Meteor. Soc., 215–219.

114. Saito, A., 1992: Mesoscale Analysis of Typhoon-Associated Tornado Outbreaks in Kyushu Island on 13 October 1980. *J. Meteorol. Soc. Japan. Ser. II*, **70**, 43–55.

115. Koch, S. E., D. Hamilton, D. Kramer, and A. Langmaid, 1998: Mesoscale Dynamics in the Palm Sunday Tornado Outbreak. *Mon. Weather Rev.*, **126**, 2031–2060.

116. Langmaid, A. H., and A. J. Riordan, 1998: Surface Mesoscale Processes during the 1994 Palm Sunday Tornado Outbreak. *Mon. Weather Rev.*, **126**, 2117–2132.

117. Corfidi, S. F., and S. P. Center, 1998: Some thoughts on the role mesoscale features played in the 27 May 1997 central Texas tornado outbreak. *Preprints, 19th Conf. on Severe Local Storms*, Minneapolis, MN, Amer. Meteor. Soc., 177–180.

118. Wood, A., 2006: Mesoscale Supercell Dynamics of the Comfrey/St. Peter Tornado Outbreak March 29, 1998. *J. Univ. Wisconsin Atmos. Ocean. Sci. Dep.*, **1**, 1–13.

119. Johns, R. H., and R. A. Dorr Jr, 1996: Some meteorological aspects of strong and violent tornado episodes in New England and eastern New York. *Natl. Weather Dig.*, **20**, 2–12.

120. Hales, J. E., and M. D. Vescuo, 1997: The 27 March 1994 tornado outbreak in the southeast US: The forecast process from a storm prediction center perspective. *Natl. Weather Dig.*, **21**, 1–15.

121. Hagemeyer, B. C., and G. K. Schmocker, 1991: Characteristics of East-Central Florida Tornado Environments. *Weather Forecast.*, **6**, 499–514.

122. Hagemeyer, B. and G.K. Schmocker, 1992: A study of central Florida tornado outbreaks. *In Symposium on Weather and Forecasting*, Atlanta, GA, 148-154.

123. Hagemeyer, B. C., and D. A. Matney, 1993: Relationship of twenty upper air indices to central Florida tornado outbreaks. *Preprints, 13th Conf. of Weather Analysis and Forecasting*, Vienna, VA, Amer. Meteor. Soc., 574–577.

124. Schmocker, G.K., D.W. Sharp, and B.C. Hagemeyer, 1990: Three initial climatological studies for WFO Melbourne, Florida: A first step in the preparation for future operations. *NOAA Technical Memoradnum NWS SR-132*, Fort Worth, Texas

125. Sharp, D. W., A. J. Cristaldi, S. M. Spratt, and B. C. Hagemeyer, 1998: Multifaceted General Overview of the East Central Florida Tornado Outbreak of 22-23 February 1998. *Preprints, 19th Conf. on Severe Local Storms*, Minneapolis, MN, Amer. Meteor. Soc., 140–143.

126. Hagemeyer, B.C.,and D.A. Matney, 1994: Peninsular Florida tornado outbreaks (1950-1993), *NOAA Technical Memorandum NWS SR-151*, Fort Worth, Texas

127. Hagemeyer, B. C., 1997: Peninsular Florida Tornado Outbreaks. *Weather Forecast.*, **12**, 399–427.

128. McCaul, E. W., 1991: Buoyancy and Shear Characteristics of Hurricane-Tornado Environments. *Mon. Weather Rev.*, **119**, 1954–1978.

129. Hagemeyer, B. C., and S. J. Hodanish, 1995: Florida tornado outbreaks associated with tropical cyclones. *Preprints, 17th Conf. on Hurricanes and Tropical Meteorology*, Miami, FL, Amer. Meteor. Soc., 312–314.

130. Cammarata, M., E. W. McCaul, and D. Buechler, 1996: Observations of shallow supercells during a major tornado outbreak spawned by Tropical Storm Beryl. *Preprints, 18th Conf. on Severe Local Storms*, San Francisco, CA, Amer. Meteor. Soc., 340–343.

131. Vescio, M. D., S. J. Weiss, and F. P. Ostby, 1996: Tornadoes associated with tropical storm Beryl. *Natl. Weather Dig.*, **21**, 2–10.

132. Hagemeyer, B.C., 1998: 1.2 Significant tornado events associated with tropical and hybrid cyclones in Florida. NOAA/National Weather Service Melbourne, Florida.

133. Edwards, R., 1998: Tornado production by exiting tropical cyclones. *Preprints, 23rd Conf. on Hurricanes and Tropical Meteorology*, Dallas, TX, Amer. Meteor. Soc., 485–488.

134. Thompson, R. L., and M. D. Vescio, 1998: The destruction potential index—A method for comparing tornado days. *Preprints, 19th Conf. on Severe Local Storms*, Minneapolis, MN, Amer. Meteor. Soc., 280–282.

135. Thompson, R. L., and R. Edwards, 2000: An Overview of Environmental Conditions and Forecast Implications of the 3 May 1999 Tornado Outbreak. *Weather Forecast.*, **15**, 682–699.

136. Markowski, P. M., 2002: Mobile Mesonet Observations on 3 May 1999. *Weather Forecast.*, **17**, 430–444.

137. *Pietrycha, A.E., J.M. Davies, M. Ratzer, and P. Merzlock, 2004: P1. 3 Tornadoes in a Deceptively Small CAPE Environment: The 4/20/04 Outbreak in Illinois and Indiana.*

138. Lee, B. D., B. F. Jewett, and R. B. Wilhelmson, 2006: The 19 April 1996 Illinois Tornado Outbreak. Part I: Cell Evolution and Supercell Isolation. *Weather Forecast.*, **21**, 433–448.

139. 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. *Weather Forecast.*, **21**, 449–464.

140. Hamill, T. M., R. S. Schneider, H. E. Brooks, G. S. Forbes, H. B. Bluestein, M. Steinberg, D. Meléndez, and R. M. Dole, 2005: The May 2003 Extended Tornado Outbreak. *Bull. Am. Meteorol. Soc.*, **86**, 531–542.

141. Miller, D. J., 2006: Observations of low level thermodynamic and wind shear profiles on significant tornado days. *Preprints, 23rd Conf. on Severe Local Storms, St. Louis, MO, Amer. Meteor. Soc*, Vol. 3 of, St Louis, MO, Amer. Meteor. Soc., 3.1.

142. Gold, D. A., and J. W. Nielsen-Gammon, 2008: Potential Vorticity Diagnosis of the Severe Convective Regime. Part IV: Comparison with Modeling Simulations of the Moore Tornado Outbreak. *Mon. Weather Rev.*, **136**, 1612–1629.

143. Zupanski, D., M. Zupanski, E. Rogers, D. F. Parrish, and G. J. DiMego, 2002: Fine-Resolution 4DVAR Data Assimilation for the Great Plains Tornado Outbreak of 3 May 1999. *Weather Forecast.*, **17**, 506–525.

144. Roebber, P. J., 2004: The risks and rewards of high resolution and ensemble numerical weather prediction. *20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., 24.5.

145. Xue, M., D. Wang, J. Gao, K. Brewster, and K. K. Droegemeier, 2003: The Advanced Regional Prediction System (ARPS), storm-scale numerical weather prediction and data assimilation. *Meteorol. Atmos. Phys.*, **82**, 139–170.

146. Egentowich, J. M., M. L. Kaplan, Y.-L. Lin, and A. J. Riordan, 2000: Mesoscale simulations of dynamical factors discriminating between a tornado outbreak and non-event over the southeast US Part I: 84-48 hour precursors. *Meteorol. Atmos. Phys.*, **74**, 129–157.

147. Egentowich, J. M., M. L. Kaplan, Y.-L. Lin, and A. J. Riordan, 2000: Mesoscale simulations of dynamical factors discriminating between a tornado outbreak and non-event over the southeast US Part II: 48-6 hour precursors. *Meteorol. Atmos. Phys.*, **74**, 159–187.

148. Egentowich, J. M., M. L. Kaplan, Y.-L. Lin, and A. J. Riordan, 2000: Mesoscale simulations of dynamical factors discriminating between a tornado outbreak and non-event over the southeast US Part III: 6 hour precursors. *Meteorol. Atmos. Phys.*, **74**, 189–215.

149. Stensrud, D. J., and S. J. Weiss, 2002: Mesoscale Model Ensemble Forecasts of the 3 May 1999 Tornado Outbreak. *Weather Forecast.*, **17**, 526–543.

150. Seko, H., K. Saito, M. Kunii, and M. Kyouda, 2009: Mesoscale Ensemble Experiments on Potential Parameters for Tornado Outbreak. *SOLA*, **5**, 57–60.

151. Roebber, P. J., D. M. Schultz, and R. Romero, 2002: Synoptic Regulation of the 3 May 1999 Tornado Outbreak. *Weather Forecast.*, **17**, 399–429.

152. Rose, S. F., P. V. Hobbs, J. D. Locatelli, and M. T. Stoelinga, 2004: A 10-Yr Climatology Relating the Locations of Reported Tornadoes to the Quadrants of Upper-Level Jet Streaks. *Weather Forecast.*, **19**, 301–309.

153. McCaul, E. W., D. E. Buechler, S. J. Goodman, and M. Cammarata, 2004: Doppler Radar and Lightning Network Observations of a Severe Outbreak of Tropical Cyclone Tornadoes. *Mon. Weather Rev.*, **132**, 1747–1763.

154. Curtis, L., 2004: Midlevel Dry Intrusions as a Factor in Tornado Outbreaks Associated with Landfalling Tropical Cyclones from the Atlantic and Gulf of Mexico. *Weather Forecast.*, **19**, 411–427.

155. Stephanie, M. V., 2005: Tornado outbreaks associated with land-falling tropical cyclones in the Atlantic Basin. *Sixth Conference on Coastal Atmospheric and Oceanic Prediction and Processes*, San Diego, CA, Amer. Meteor. Soc., 7.1.

156. Watson, A. I., M. A. Jamski, T. J. Turnage, J. R. Bowen, and J. C. Kelley, 2005: The tornado outbreak across the north Florida panhandle in association with hurricane Ivan. *Preprints 32nd Radar Conf.*, Albuquerque, NM, Amer. Meteor. Soc., 10R.3.

157. Verbout, S. M., D. M. Schultz, L. M. Leslie, H. E. Brooks, D. J. Karoly, and K. L. Elmore, 2007: Tornado outbreaks associated with landfalling hurricanes in the North Atlantic Basin: 1954–2004. *Meteorol. Atmos. Phys.*, **97**, 255–271.

158. Belanger, J. I., J. A. Curry, and C. D. Hoyos, 2009: Variability in tornado frequency associated with U.S. landfalling tropical cyclones. *Geophys. Res. Lett.*, **36**, L17805.

159. Rogash, J. A., and R. D. Smith, 2000: Multiscale Overview of a Violent Tornado Outbreak with Attendant Flash Flooding. *Weather Forecast.*, **15**, 416–431.

160. Stuart, N. A., 2004: The Anatomy of the Big Event That Never Happened—The Grand Finale of the May 2003 Tornado Outbreak. *22nd Conference on Severe Local Storms*, Hyannis, MA, Amer. Meteor. Soc., 12.8.

161. Darbe, D., and J. Medlin, 2005: Multi-scale analysis of the 13 October 2001 central Gulf Coast shallow supercell tornado outbreak. *Electron. J. Oper. Meteor*, **6**, 1–7.

162. Feltz, W. F., and J. R. Mecikalski, 2002: Monitoring High-Temporal-Resolution Convective Stability Indices Using the Ground-Based Atmospheric Emitted Radiance

Interferometer (AERI) during the 3 May 1999 Oklahoma–Kansas Tornado Outbreak. *Weather Forecast.*, **17**, 445–455.

163. Feltz, W. F., W. L. Smith, H. B. Howell, R. O. Knuteson, H. Woolf, and H. E. Revercomb, 2003: Near-Continuous Profiling of Temperature, Moisture, and Atmospheric Stability Using the Atmospheric Emitted Radiance Interferometer (AERI). *J. Appl. Meteorol.*, **42**, 584–597.

164. Benjamin, S. G., B. E. Schwartz, E. J. Szoke, and S. E. Koch, 2004: The Value of Wind Profiler Data in U.S. Weather Forecasting. *Bull. Am. Meteorol. Soc.*, **85**, 1871–1886.

165. Nietfeld, D. D., 2003: The synoptic environment of the 11 April 2001 central plains tornado outbreak viewed in three dimensions. *Proceedings of the 19th IIPS Conference*, Long Beach, CA, Amer. Meteor. Soc., P1.1.

166. Bikos, D., J. Weaver, and B. Motta, 2002: A Satellite Perspective of the 3 May 1999 Great Plains Tornado Outbreak within Oklahoma. *Weather Forecast.*, **17**, 635–646.

167. Myint, S., M. Yuan, R. Cerveny, and C. Giri, 2008: Comparison of Remote Sensing Image Processing Techniques to Identify Tornado Damage Areas from Landsat TM Data. *Sensors*, **8**, 1128–1156.

168. Brown, R. A., V. T. Wood, and D. Sirmans, 2002: Improved Tornado Detection Using Simulated and Actual WSR-88D Data with Enhanced Resolution. *J. Atmos. Ocean. Technol.*, **19**, 1759–1771.

169. Lee, R. R., and R. M. Steadham, 2004: WSR-88D algorithm comparisons of VCP 11 and new VCP 12. *Preprints, 20th Int. Conf. on Interactive and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, Vol. 12 of, Seattle, WA, Amer. Meteor. Soc., 12.7.

170. Hu, M., 2007: Analysis and prediction of 8 May 2003 Oklahoma City tornadic thunderstorm and embedded tornado using ARPS with assimilation of WSR-88D radar data. *22nd Conference on Weather Analysis and Forecasting/18th Conference on Numerical Weather Prediction*, Park City, UT, Amer. Meteor. Soc., 1B.4.

171. Ryzhkov, A. V., T. J. Schuur, D. W. Burgess, and D. S. Zrnic, 2005: Polarimetric Tornado Detection. *J. Appl. Meteorol.*, **44**, 557–570.

172. Yuan, M., M. Dickens-Micozzi, and M. A. Magsig, 2002: Analysis of Tornado Damage Tracks from the 3 May Tornado Outbreak Using Multispectral Satellite Imagery. *Weather Forecast.*, **17**, 382–398.

173. Yuan, M., 2005: Beyond Mapping in GIS Applications to Environmental Analysis. *Bull. Am. Meteorol. Soc.*, **86**, 169–170.

174. Camp, P. J., 2008: Integrating a geographical information system into storm assessment: The southeast Alabama tornado outbreak of 1 March 2007. *Preprints, 24th Conf. on IIPS*, Vol. 1 of, New Orleans, LA, Amer. Meteor. Soc., P1.4.

175. Brooks, H., and C. A. Doswell, 2001: Some aspects of the international climatology of tornadoes by damage classification. *Atmos. Res.*, **56**, 191–201.

176. Stephanie, M. V., 2004: Leveling the field for tornado reports through time: Inflation-adjustment of annual tornado reports and objective identification of extreme tornado reports. *22nd Conference on Severe Local Storms*, Hyannis, MA, Amer. Meteor. Soc., 7B.3.

177. Brooks, H. E., C. A. Doswell, and M. P. Kay, 2003: Climatological Estimates of Local Daily Tornado Probability for the United States. *Weather Forecast.*, **18**, 626–640.

178. 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*, Vol. 5 of, Hyannis, MA, Amer. Meteor. Soc., P5.1.

179. Schneider, R. S., H. E. Brooks, and J. T. Schaefer, 2004: Tornado outbreak day sequences: Historic events and climatology (1875-2003). *Preprints, 22nd Conf. on Severe Local Storms*, Hyannis, MA, Amer. Meteor. Soc., 12.1.

180. Hamill, T. M., R. S. Schneider, H. E. Brooks, G. S. Forbes, H. B. Bluestein, M. Steinberg, D. Meléndez, and R. M. Dole, 2005: The May 2003 Extended Tornado Outbreak. *Bull. Am. Meteorol. Soc.*, **86**, 531–542.

181. Edwards, R., R. L. Thompson, K. C. Crosbie, J. A. Hart, and C. A. Doswell III, 2004: A proposal for modernized definitions of tornado and severe thunderstorm outbreaks. *Preprints, 22nd Conf. on Severe Local Storms*, Vol. 7 of, Hyannis, MA, Amer. Meteor. Soc., 7B.2.

182. Forbes, G., 2006: Meteorological aspects of high-impact tornado outbreaks. *Symposium on the Challenges of Severe Convective Storms*, Vol. 28 of, Atlanta, GA, American Meteorological Society, P1.12.

183. Shafer, C., and C. Doswell, 2010: A Multivariate Index for Ranking and Classifying Severe Weather Outbreaks. *Electron. J. Sev. Storms Meteorol.*, **5**(1).

184. Shafer, C. M., and C. A. Doswell, 2011: Using kernel density estimation to identify, rank, and classify severe weather outbreak events. *Electron. J. Sev. Storms Meteorol.*, **6**, 1–28.

185. Malamud, B. D., and D. L. Turcotte, 2012: Statistics of severe tornadoes and severe tornado outbreaks. *Atmos. Chem. Phys.*, **12**, 8459–8473.

186. Clark, A. J., J. S. Kain, P. T. Marsh, J. Correia, M. Xue, and F. Kong, 2012: Forecasting Tornado Pathlengths Using a Three-Dimensional Object Identification Algorithm Applied to Convection-Allowing Forecasts. *Weather Forecast.*, **27**, 1090–1113.

187. Clark, A. J., J. Gao, P. T. Marsh, T. Smith, J. S. Kain, J. Correia, M. Xue, and F. Kong, 2013: Tornado Pathlength Forecasts from 2010 to 2011 Using Ensemble Updraft Helicity. *Weather Forecast.*, **28**, 387–407.

188. Fuhrmann, C. M., C. E. Konrad, M. M. Kovach, J. T. McLeod, W. G. Schmitz, and P. G. Dixon, 2014: Ranking of Tornado Outbreaks across the United States and Their Climatological Characteristics. *Weather Forecast.*, **29**, 684–701.

189. Tippett, M. K., and J. E. Cohen, 2016: Tornado outbreak variability follows Taylor's power law of fluctuation scaling and increases dramatically with severity. *Nat. Commun.*, **7**, 10668.

190. Tippett, M. K., C. Lepore, and J. E. Cohen, 2016: More tornadoes in the most extreme U.S. tornado outbreaks. *Science (80-.).*, **354**, 1419–1423.

191. Gensini, V. A., and A. Marinaro, 2016: Tornado Frequency in the United States Related to Global Relative Angular Momentum. *Mon. Weather Rev.*, **144**, 801–810.

192. Gensini, V. A., D. Gold, J. T. Allen, and B. S. Barrett, 2019: Extended U.S. Tornado Outbreak During Late May 2019: A Forecast of Opportunity. *Geophys. Res. Lett.*, **46**, 10150–10158.

193. Knupp, K. R., and Coauthors, 2014: Meteorological Overview of the Devastating 27 April 2011 Tornado Outbreak. *Bull. Am. Meteorol. Soc.*, **95**, 1041–1062.

194. Yussouf, N., D. C. Dowell, L. J. Wicker, K. H. Knopfmeier, and D. M. Wheatley, 2015: Storm-Scale Data Assimilation and Ensemble Forecasts for the 27 April 2011 Severe Weather Outbreak in Alabama. *Mon. Weather Rev.*, **143**, 3044–3066.

195. Tochimoto, E., and H. Niino, 2016: Structural and Environmental Characteristics of Extratropical Cyclones that Cause Tornado Outbreaks in the Warm Sector: A Composite Study. *Mon. Weather Rev.*, **144**, 945–969.

196. Anderson-Frey, A. K., Y. P. Richardson, A. R. Dean, R. L. Thompson, and B. T. Smith, 2018: Near-Storm Environments of Outbreak and Isolated Tornadoes. *Weather Forecast.*, **33**, 1397–1412.

197. Schumacher, P. N., and J. M. Boustead, 2011: Mesocyclone Evolution Associated with Varying Shear Profiles during the 24 June 2003 Tornado Outbreak. *Weather Forecast.*, **26**, 808–827.

198. Megnia, R. W., T. W. Humphrey, and J. A. Rackley, 2019: The Historic 2 April 2017 Louisiana Tornado Outbreak. *J. Oper. Meteorol.*, **7**, 27–39.

199. Saide, P. E., and Coauthors, 2015: Central American biomass burning smoke can increase tornado severity in the U.S. *Geophys. Res. Lett.*, **42**, 956–965.

200. Thompson, D. B., and P. E. Roundy, 2013: The Relationship between the Madden–Julian Oscillation and U.S. Violent Tornado Outbreaks in the Spring. *Mon. Weather Rev.*, **141**, 2087–2095.

201. Lee, S.-K., R. Atlas, D. Enfield, C. Wang, and H. Liu, 2013: Is There an Optimal ENSO Pattern That Enhances Large-Scale Atmospheric Processes Conducive to Tornado Outbreaks in the United States? *J. Clim.*, **26**, 1626–1642.

202. Lee, S.-K., A. T. Wittenberg, D. B. Enfield, S. J. Weaver, C. Wang, and R. Atlas, 2016: US regional tornado outbreaks and their links to spring ENSO phases and North Atlantic SST variability. *Environ. Res. Lett.*, **11**, 44008.

203. Sparrow, K. H., and A. E. Mercer, 2016: Predictability of US tornado outbreak seasons using ENSO and northern hemisphere geopotential height variability. *Geosci. Front.*, **7**, 21–31.

204. Cook, A. R., L. M. Leslie, D. B. Parsons, and J. T. Schaefer, 2017: The Impact of El Niño–Southern Oscillation (ENSO) on Winter and Early Spring U.S. Tornado Outbreaks. *J. Appl. Meteorol. Climatol.*, **56**, 2455–2478.

205. Tippett, M. K., 2018: Robustness of Relations between the MJO and U.S. Tornado Occurrence. *Mon. Weather Rev.*, **146**, 3873–3884.

206. Mercer, A., and A. Bates, 2019: Meteorological Differences Characterizing Tornado Outbreak Forecasts of Varying Quality. *Atmosphere (Basel).*, **10**, 16.

207. Skamarock, W.C. and Coauthors, 2008: A Description of the Advanced Research WRF Version 3; *NCAR Technik Note*; NCAR/TN-475+STR: Boulder, CO, USA, 125

208. Schaefer, J. T., and R. Edwards, 1999: The SPC tornado/severe thunderstorm database. *Preprints, 11th Conf. on Applied Climatology*, Vol. 6 of, Dallas, TX, Amer. Meteor. Soc

209. U.S. Census Bureau (2017). Cartographic Boundary Files: County level. Retrieved from https://www.census.gov/geographies/mapping-files/2017/geo/kml-cartographic-boundary-files.html. [Accessed: 9-April-2019].

210. Concannon, P.R., H.E. Brooks, and C. A. Doswell III, 2000: Climatological risk of strong and violent tornadoes in the United States. *In Preprints, 2nd Symposium on Environmental Applications*, Long Beach, CA, Amer. Meteor. Soc., 212-219.

211. Bowman, A. W., and A. Azzalini, 1997: Applied Smoothing Techniques for Data Analysis:the Kernel Approach Using S-Plus Illustrations. Oxford University Press, 208.

212. Brooks, H. E., M. Kay, and J. A. Hart, 1998: Objective limits on forecasting skill of rare events. *Preprints, 19th Conf. on Severe Local Storms*, Minneapolis, MN, Amer. Meteor. Soc, 552–555.

213. Coleman, T. A., 2010: The effects of topography and friction on mesocyclones and tornadoes. Preprints, 25th Conf. on Severe Local Storms,Denver, CO, Amer. Meteor. Soc., P8.12.

214. Cannon, J. B., J. Hepinstall-Cymerman, C. M. Godfrey, and C. J. Peterson, 2016: Landscape-scale characteristics of forest tornado damage in mountainous terrain. *Landsc. Ecol.*, **31**, 2097–2114.

215. Matsangouras, I. T., I. Pytharoulis, and P. T. Nastos, 2013: Numerical Investigation of the Role of Topography in Tornado Events in Greece. *Advances in Meteorology, Climatology and Atmospheric Physics*, C.G. Helmis and P.T. Nastos, Eds., Springer Berlin Heidelberg, 209–215.

216. Lewellen, D. C., 2012: Effects of topography on tornado dynamics: A simulation study. *26th Conference on Severe Local Storms*, Nashville, TN, American Meteorological Society, 4B.1.

217. Bosart, L. F., A. Seimon, K. D. LaPenta, and M. J. Dickinson, 2006: Supercell Tornadogenesis over Complex Terrain: The Great Barrington, Massachusetts, Tornado on 29 May 1995. *Weather Forecast.*, **21**, 897–922.

218. Razavi, A., and P. P. Sarkar, 2018: Laboratory Study of Topographic Effects on the Near-surface Tornado Flow Field. *Boundary-Layer Meteorol.*, **168**, 189–212.

219. Brooks, H. E., J. W. Lee, and J. P. Craven, 2003: The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data. *Atmos. Res.*, **67**–**68**, 73–94.

220. Mercer, A. E., C. M. Shafer, C. A. Doswell III, L. M. Leslie, and M. B. Richman, 2012: Synoptic Composites of Tornadic and Nontornadic Outbreaks. *Mon. Weather Rev.*, **140**, 2590–2608.

221. Ferguson, E. W., F. P. Ostby, and P. W. Leftwich, 1987: The Tornado Season of 1985. *Mon. Weather Rev.*, **115**, 1437–1445.

Chapter III: Identification of mid-tropospheric patterns associated with tornado outbreaks in the United States

1. Introduction

Owing to the impacts of tornados on life and property, studies of tornadic storms and tornadic storm systems are of great interest among scientists around the world. Tornado outbreaks, in particular, pose a great threat to society because they consist of multiple, often violent, long-track tornadoes, and therefore are likely to affect populated areas [1]. In the USA, regions east of the Rocky Mountains are most susceptible to the atmospheric conditions that are conducive to the formation of tornadoes, either as isolated events or as a part of a major outbreak. Each year, a few major tornado outbreaks affect the United States [2]. One of the biggest in recent history, the 25-28 April 2011 Southeastern U.S. tornado outbreak, caused 316 fatalities, 2700 injuries, and over $4.2 billion dollars of economic losses across five states [3-4]. According to Schneider et al. (2004), over 80% of all tornado-related fatalities in the U.S. arise as a result of tornado outbreaks [5]. With a continued increase in population density and expansion of urban areas, exposure of many people to future tornado outbreak impacts is increasing.

Understanding what and how environmental conditions favor the formation of tornadoes is a key to weather prediction and risk mitigation. Many studies have focused on the description of atmospheric conditions that are necessary for tornado development [6-12]. These conditions include high instability, vertical wind shear, an abundant amount of low-level humidity, and some sort of lifting mechanism, such as a front or convergence zone [13-14]. Most tornado outbreaks are forced by large-scale drivers, such as synoptic-scale systems [15-16]. Some researchers have examined the relationship of tornados with specific weather patterns, for instance, shifts in the jet

stream during different phases of El Niño-Southern Oscillation [17-21]. Others describe changes in low-level moisture patterns [22-23] and their influence on tornado frequencies and intensity.

Despite such examples, there are few studies that focus entirely on finding commonalities among atmospheric patterns that are associated with the biggest historic tornado outbreaks. Of those, Mercer et. al (2012) identified synoptic-scale differences between tornado outbreaks and non-tornadic outbreaks. The research compared composite maps of each outbreak type formulated on five meteorological variables: geopotential height, relative humidity, temperature, and u and υ wind components [10]. In their research, the synoptic-scale characteristics of tornado outbreaks were determined as composite patterns of the outbreaks. They did not explore specific characteristics of common synoptic patterns for an individual variable (such as just in the geopotential height) and its relationship with tornado outbreaks.

The primary goal of our work is to identify and describe atmospheric patterns that occur most often in association with historic tornado outbreaks. Several studies have revealed the presence of a 500-mb trough west of the outbreak location as one of few key features prior to the tornado outbreak formation [24-26]. Based on this prior work, we hypothesize that there are relatively common, large-scale patterns in the 500-hPa geopotential height field that are associated with major tornado outbreaks. We test this hypothesis for major tornado outbreaks in May from 1950 to 2017 using three independent statistical techniques for pattern identification: principal component analysis, hierarchical clustering, and silhouette clustering.

2. Data and methods

2.1 Data sources

The Storm Prediction Center's (SPC) severe report database [27] served as our dataset for tornado reports from 1 January 1950 to 31 December 2017, with each 24-h period specified from 0600 Central Standard Time to 0559 CST, consistent with the SPC database. These data retained some temporal and spatial inconsistencies that were not caused by meteorological factors [28-31], including changes in methods of collecting reports [32], increases in tornado detection due to Doppler radar and National Weather Service vigilance, and heightened public awareness of severe weather [32]. Still, Brooks and Doswell (2001) noted that the SPC database was the most reliable tornado dataset currently available for the United States [28]. In this study, we restricted tornado reports to significant tornados — those F2 or greater on the Fujita scale — as they are responsible for the majority of deaths caused by tornados in the US and cause millions of dollars in damages every year [33]. We also used only those outbreaks with seven or more significant tornados, defining them as 'major tornado outbreaks.'

To examine atmospheric patterns and assess any synoptic-scale signal associated with major tornado outbreaks, a reanalysis dataset was required. Reanalyses combine observations and numerical models that simulate Earth systems to produce an estimate of the state of the atmosphere for long-term monitoring and research purposes. Several reanalysis datasets exist, such as those created by National Centers for Environmental Prediction (NCEP) and National Center for Atmospheric Research (NCAR); however, they begin only in the mid-twentieth century. The Twentieth Century Reanalysis (20CR), however, offers a continuous dataset from 1851, currently providing the longest record of a representative state of the atmosphere every 6 hours [34]. Because

our future research will require a large sample size of events (for climate change research), we chose the 20CR (20CR-version 2c) for this study.

The 20CR is a high-resolution global atmospheric dataset created to validate climate model simulations of the Twentieth Century [35]. As boundary conditions, the 20CR uses surface pressure reports, monthly observed sea-surface temperature, and sea-ice concentration fields, and the reanalysis dataset spans 1851 to 2014 [35]. This retrospective analysis provides full a three-dimensional characterization of the tropospheric state for 24 pressure levels on a 2° latitude x 2° longitude grid (corresponding to 220-225 km). Additionally, this reanalysis uses the Ensemble Kalman Filter data assimilation system [35], which provides direct computations on the uncertainty of that reanalysis. Moreover, comparison with independent radiosonde data indicates that the reanalysis dataset is high quality [35]. The 20CR dataset enables diagnostic studies and numerical model validations; therefore, these data are considered one of the most valuable sources for climate and pattern recognition research, such as the research conducted herein.

2.2 Methods of analysis

After filtering the tornado reports to retain only significant tornados (in 24-hour periods from 0600 CST to 0559 CST), we used kernel density estimation to cluster tornados into outbreaks and further filtered to those outbreaks with seven or more significant tornados (see Chapter 2, section 3.2 Methods). We then computed the following information for each major tornado outbreak: geographic location, including latitude and longitude of outbreak center point; and time of occurrence, including outbreak start date/time and end date/time (see Chapter 2 for more details).

Using this technique, we found 333 major tornado outbreaks occurred from 1950-2017, all east of the Rocky Mountains.

For each major tornado outbreak, we selected an associated 500-hPa geopotential height field to represent the outbreak by extracting this field from the 20CR at the first available time prior to the initiation of the outbreak. For instance, if tornado outbreak number 325 started at 1500 CST on 16 June 2014, our filtering algorithm selected the 500-hPa geopotential height field at 1200 CST on 16 June 2014 to represent the large-scale atmospheric pattern associated with the development of this outbreak. We limited our analysis to one atmospheric field at one time (i.e., a snapshot) to capture the state of mid-troposphere dynamics prior to the tornado outbreak occurrence and to build a database of snapshots that were independent (for statistical analyses).

To aid the pattern-recognition techniques and to test the robustness of our results, we used four different domains (Figure 6): the "wide CONUS domain" encompassed 55°N,140°W to 20°N, 60°W (779 grid points on a 2° latitude x 2° longitude grid); the "CONUS domain" from 50°N, 126°W to 25°N, 66°W (434 grid points); the "eastern U.S. + Rocky Mountains domain" from NW 50°N, 108°W to 25°N, 70°W (280 grid points); and the "Great Plains domain" from 46°N, 104°W to 28°N, 82°W (120 grid points).

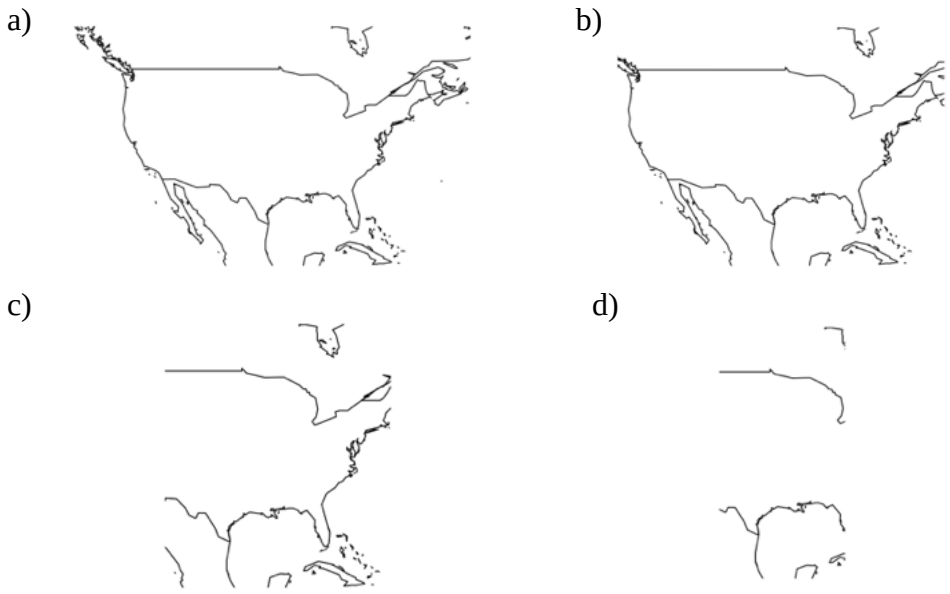

Fig. 6. Four domain sizes tested in this research: a) the "wide CONUS domain", b) the" CONUS domain", c) the "eastern U.S.+ Rocky Mountains domain", d) the "Great Plains domain".

Results presented herein focus only on the "CONUS domain." Given limitations of the temporal coverage of 20CR, our analysis of patterns spanned 1950 to 2014. Accordingly, the number of major tornado outbreaks was reduced from 333 to 325. Figure 7 displays the frequency distribution of major tornado outbreaks, by month, for that time period.

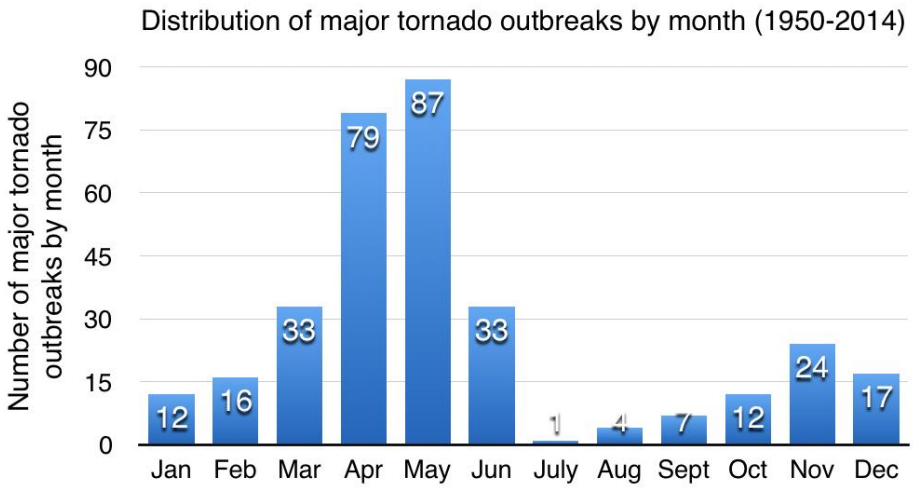

Fig. 7. Monthly counts of major tornado outbreaks from 1950 to 2014. The highest number of major tornado outbreaks by month occurs in May.

Finally, to reduce the effects of seasonality, our analyses examined the 500-hPa geopotential height *anomalies*, defined as deviations in the 500-hPa geopotential height field for each outbreak from the 30-year average values of all geopotential heights for the corresponding month. For that purpose, the average was computed based on the 6-hour reanalysis intervals for the consecutive period of 30 years from 1 January 1981 to 31 December 2010 (Fig. 8).

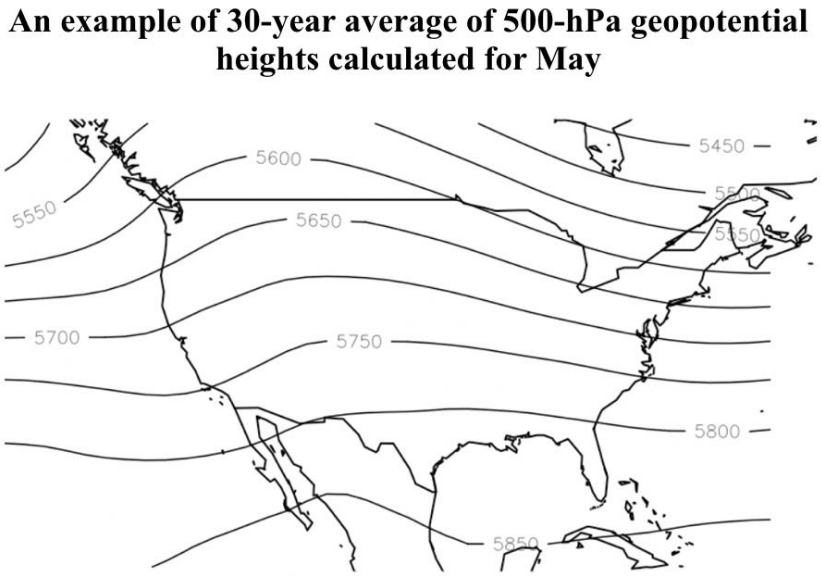

Fig. 8. The 30-year (1981-2010) average of 500-hPa geopotential heights (in meters) calculated for May.

Principal component analysis (PCA) has been used successfully to explore and analyze large, multivariate data sets [36-39], such as geopotential height fields in atmospheric studies. PCA is a statistical technique that transforms a data set of correlated data into a new set of orthogonal variables. Because the leading modes account for most of the variation in highly correlated datasets, PCA can offer a more compact representation of variability contained in a multivariate dataset and reduce the degrees of freedom to the leading low-order modes of the dataset [37]. Moreover, post-processing of the resulting patterns can be used to explore a general structure and trends in the dataset, as well as explain spatial and temporal variations in the data matrix.

For the purpose of this work, the major tornado outbreak matrix consisted of two elements: gridpoint observations of 500-hPa geopotential heights, representing the spatial dimension, and major tornado outbreaks cases from 1950 to 2014, representing the temporal dimension. Thus, each row of the matrix represented a major tornado outbreak, with 87 total outbreaks for May (Fig. 7). Each column contained values of 500-hPa geopotential height anomalies for the CONUS domain on a grid with 2 degrees latitude-longitude spacing.

The PC model could be defined from a data matrix $\mathbf{Z} = \{z_{ij} : i=1, \ldots N; j=1,\ldots, n\}$, where i represented the individual outbreaks and j represented the values for 500-hPa geopotential height anomalies. The z_{ij} was defined as follows:

$$z_{ij} = \sum_{m=1}^{r} f_{im} a_{mj}^{\mathrm{T}} \tag{3}$$

with f_{im} representing the principal component score for each individual outbreak (i) and a_{mj}^{T} representing the principal component loading for each variable transposed (j). In the matrix form, Equation (3) becomes:

$$\boldsymbol{Z = FA^{T}} \tag{4}$$

where **Z** is a matrix of standardized anomalies, **F** is a matrix of uncorrelated PC scores, and **A** is a matrix with columns representing PC loadings [37]. Depending on the choice of decomposition mode (Equation 3) and similarity matrix (Equation 4), the results from and interpretations of the PCA will differ. First, because we are interested in those subgroups of outbreaks with similar

spatial patterns, we applied the T-mode approach. T-mode transposed the data matrix **Z** and treated the spatial field of grid points (or *n* locations) at each of the times of tornado outbreaks (*m* times) as variables (Fig. 9). For our work, T-mode transformation was expressed as:

$$\boldsymbol{Z_T = F_T A_T^T} \tag{5}$$

where PC loadings for each column of the matrix ($\boldsymbol{A_T^T}$) represented the time series of 500-hPa height anomalies associated with major tornado outbreaks, and the columns of the PC scores ($\boldsymbol{F_T}$) corresponded to a geographical distribution of 500-hPa height anomalies associated with tornado outbreaks.

Second, the PC analysis was directly derived from a parent similarity matrix that described the variability along either rows or columns. The similarity matrix could consist of a correlation, covariance, or cross-products matrix. Depending on the choice of the similarity matrix, the type of relationship desired from the PC analysis varied, perhaps leading to different physical interpretations of the data. Considering the goal of this analysis, we chose to apply a covariance matrix to portray insights into how much 500-hPa geopotential height anomalies time series change with respect to each other. Figure 9 displays the covariance matrix structure under T-mode decomposition.

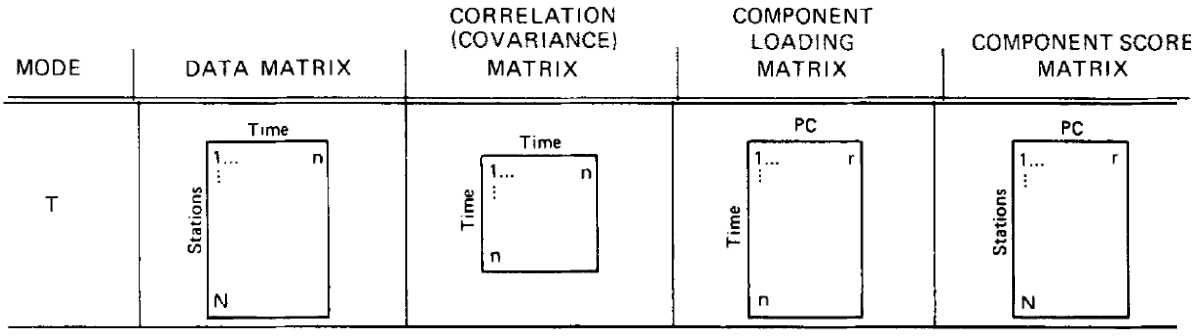

Fig. 9. T-mode decomposition matrix structures from Richman (1986) [37, p. 315]: "Data matrix (with columns treated as variables), dispersion matrix, PC loading matrix and PC score matrix under T-mode."

Next, the covariance similarity matrix was diagonalized into eigenvalues (**λ**) and associated eigenvectors (**V**). The eigenvectors were scaled by the square root of the associated eigenvalues to define the PC loadings (**A**) as:

$$\boldsymbol{A} = \boldsymbol{V} * \boldsymbol{\lambda}^{1/2} \tag{6}$$

In spite of the fact that unrotated PCA extracts the maximum variance from a data set, its simple structure might not separate individual modes of variation adequately because of domain shape dependence, instability of the subdomain, sampling errors, or inaccuracy in the depiction of physical relationships [37]. An alternative method to examine those individual modes as well as to uncover meaningful physical relationships in the data was to apply a linear transformation, such as the Varimax orthogonal rotation [37] or the Promax oblique simple structure rotation [37]. The transformation process included rotation of the PC loading matrix (**A**) so that the new rotated loadings (**B**) were defined as **B**=**AT**, where **T** was a transformation matrix that served to find a

configuration of 500-hPa height anomalies that was maximally simplified, agreeing with the signal embedded in the similarity matrix.

Often the primary goal of conducting a PCA is data reduction; hence, only the leading, most significant principal components (PCs) are retained in further analysis. This filtering of PCs also is helpful in finding major patterns in a complex dataset, such as ours. The smaller number of PCs retained are considered to contain the signal, rather than noise, and are used for further computation of $\boldsymbol{A}$ [37]. There are multiple ways to establish how many PCs should be retained, such as the scree test [40] or methods based on the amount of variance explained [41]. In this study, we computed a congruence coefficient (g) [42-43] as a measure of similarity of the PC loadings to the corresponding flow patterns, as expressed by [44]:

$$\rho_{AB} = \frac{\sum_{j=1}^{n} b_{jA} b_{jB}}{[(\sum_{j=1}^{n} b_{jA}^2)(\sum_{j=1}^{n} b_{jB}^2)]^{\frac{1}{2}}} \tag{7}$$

where A and B represented any two chosen factors (e.g., PC loadings and 500-hPa anomalies), b represented the component loading or pattern coefficients, n represented the number of tornado outbreaks (i.e., 87 for May), and j represented the sample of individuals (i.e., a 500-hPa anomaly with maximum absolute value for given number of loadings). Because the congruence coefficient does not remove the mean of the variable, it is an optimal method to measure similarity of PC components. Values of the congruence coefficient range from –1.0 to +1.0, where +1.0 represents total agreement, –1.0 for total inverse agreement, and 0 for no relationship. The goodness of match

is important for component identification and is based on specific ranges of absolute congruence coefficients [41]. An excellent match occurs when congruence coefficients range from |0.98| to |1.00|, a good match for values from |0.92| to |0.98|, a borderline match from |0.82| to |0.92|, a poor match from |0.68| to |0.82|, and finally a terrible match when congruent coefficient is less than |0.68| [37-38]. Richman (1986) also noted that any results with a congruence coefficient less than |0.7| could be random [37]. For this research, we aimed to meet the borderline match of |0.82| or higher.

Additionally, we applied two clustering methods to compare with results obtained from the PCA congruence coefficient analysis and to validate any identified patterns: *hierarchical clustering* and *silhouette clustering*. Hierarchical clustering is an approach to identify groups of datasets without the need to pre-specify the number of clusters [45]. We implemented an agglomerative (bottom-up) type, where each data point was initially considered as an individual cluster [46]. Similar points (i.e., patterns in the 500-hPa geopotential height anomalies of major tornado outbreaks) were merged into a cluster according to an increasing dissimilarity measure, and the process continued until all points were clustered [47]. For this study, the dissimilarity structure was obtained by computing distances between the rows of the data matrix using the Euclidean distance measure. Then, to merge the data points into clusters, an average linkage clustering criterion was used. This method allowed computation of all pairwise dissimilarities between the elements in cluster 1 and the elements in cluster 2, and it considered the average of these dissimilarities as the distance between the two clusters. The average linkage technique was beneficial because it minimized the within-cluster variance while maximizing the between-cluster variance [47]. The output was displayed as dendrograms, tree-like diagrams, that illustrated the hierarchical relationship between the clusters produced [48].

Finally, silhouette clustering [49] was used in this research because, unlike the PCA and hierarchical methods, it could group data into clusters and then validated the cohesion within these clusters (i.e., evaluated how similar each object was to its own cluster). Each cluster was represented by a silhouette value that measures the tightness (cohesion) and separation between clusters, expressed as:

$$Silhouette\ (a) = \frac{Separtaion(a) - Cohesion(a)}{\max[Separation(a), Cohesion(a)]} \quad (8)$$

where *Cohesion* was the average distance between points within the same cluster and the cluster *Separation* was the average distance between a point and the nearest cluster of which it was not a member [41]. The silhouette method is advantageous because it does not depend on the clustering algorithm (e.g., k-means clustering) used to obtain the clusters, rather on the actual partition of the objects (e.g., 500-hPa geopotential height anomalies) [41].

In this research, the silhouettes were calculated based on Euclidean distance. The construction of silhouettes required application of an algorithm for partitioning a set of objects into a number of clusters. We utilized the k-means clustering algorithm, which partitioned the data into *k* clusters through minimizing the sum of squares in a cluster [50], keeping cluster centroids as small as possible and allocating every data point in the nearest cluster. Silhouette values range from −1 to +1, where +1 indicates perfect clustering (i.e., the object is very well matched to its own cluster), while a negative value indicates that the element is in the wrong cluster. Because the number of clusters in this research was unknown a priori, it was essential to maximize cluster separation and minimize cluster cohesion. To find the optimal number of clusters, we tested a set of 2 to 10 cluster

centers (input into k-means clustering) and chose the cluster number with the highest average silhouette score as the final cluster number.

3. Results

Table 1 (below) presents the congruence coefficients for the unrotated PC analysis of the 500-hPa geopotential height anomaly fields for all four domains noted in Chapter 3 section 2.2. Here, the congruence coefficient values for unrotated PCs did not meet our threshold value of |0.82| in most of domains; thus, we eliminated unrotated PCs from further analysis. This result was not surprising because, as mentioned in Chapter 3 section 2.2, a simple structure (unrotated PCA) might not be suitable for a particular data set [41].

	Wide CONUS (Domain 1)		**CONUS (Domain 2)**		**Eastern U.S. + Rocky Mountains (Domain 3)**		**Great Plains (Domain 4)**	
Number of loadings	**min**	**median**	**min**	**median**	**min**	**median**	**min**	**median**
2	0.834	0.903	0.620	0.800	0.684	0.840	0.399	0.697
3	0.611	0.833	0.620	0.675	0.117	0.684	0.393	0.398
4	0.611	0.734	0.620	0.647	0.117	0.508	0.083	0.395s

Table 1. Minimum and median absolute values of the congruence coefficients for the first four, unrotated principal component loadings for all domains.

Accordingly, we calculated the minimum and median congruence coefficients for both orthogonal Varimax and oblique Promax transformations for every domain (Appendix B). Across all four domains, Domain 3 (i.e., Eastern U.S. + Rocky Mountains) exhibited the most optimal outcome,

with the highest congruence coefficient values across the largest spatial extent possible. Thus, our results focus on this domain hereafter.

Figure 10 displays the first 10 eigenvalues for Domain 3, indicating a distinct separation between PCs 1, 2, and 3, and, albeit small, separation between PCs 3 and 4 (indicated by a horizontal line on Fig. 10).This lack of overlap indicated that the first three principal components were independent. Additionally, the first three eigenvalues explained 83% of the variance (calculated as the sum of first three eigenvalues divided by the sum of all eigenvalues; Table 2), and thus three PCs were retained initially for further analysis.

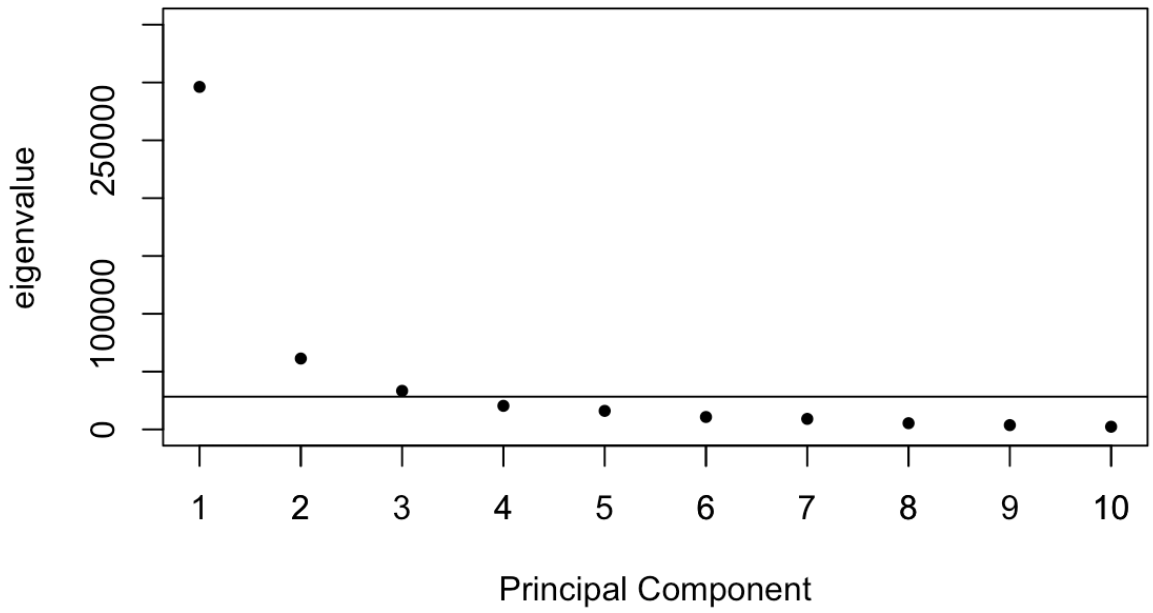

Fig. 10. Plot of first 10 eigenvalues for Domain 3 (Eastern U.S. + Rocky Mountains; *Figure 6).* The dots represent eigenvalue for each principal component.

We also examined the number of PCs to retain by computing the congruence coefficients for the first 10 loadings for both the Varimax and Promax rotations (Table 2).

Principal Component	1	2	3	4	5	6	7	8
Eigenvalue	296194.6	61303.2	33378.2	20412.8	16058.7	10760.4	9126.2	5395.7

Table 2. Example of the first 8 (of 87) eigenvalues for Domain 3. The sum of all eigenvalues equaled 469669.2.

The method used during the process of congruence coefficient testing was supervised, meaning that each increment in the number of loadings required visual inspection of the resulting loadings and incorporation of corresponding outbreaks to the congruence coefficient calculations.

Number of loadings	Varimax min	Varimax median	Promax min	Promax median
2	0.978	0.987	0.908	0.948
3	0.968	0.969	0.926	0.927
4	0.499	0.934	0.430	0.826
5	0.510	0.913	0.460	0.770
6	0.443	0.834	0.430	0.751
7	0.198	0.640	0.279	0.597
8	0.255	0.549	0.335	0.503
9	0.021	0.624	0.174	0.451
10	0.038	0.531	0.028	0.439

Table 3. Absolute values of the minimum (min) and median congruence coefficients for first 10 loadings from Domain 3. Values were calculated using both Varimax and Promax transformations.

The highest congruence coefficient values (highlighted in yellow in Table 3) across both Varimax and Promax transformations were associated with two or three loadings. Given slightly better results for two loadings with the Varimax transformation, we present the spatial (scores) and temporal (loadings) representation of patterns for the first two loadings along with a major tornado outbreak associated with each pattern.

Note that the interpretation of the PC score map depended on the sign of the PC loading. For those outbreaks that were characterized by positive PC loadings, the signs of values in the PC score map were interpreted as displayed. However, for those outbreaks that had negative PC loading values, the signs on the PC score map were interpreted as opposite to what was displayed. For example, the PC loading value of tornado outbreak number 39 was positive (Fig. 11), and thus the positive values on the PC score map corresponded to positive anomalies in the 500-hPa geopotential height field. Similarly, the areas with negative PC score values corresponded to negative anomalies. On the contrary, the loading for tornado outbreak number 1 was negative, and therefore, the areas with positive PC score values on the map corresponded to negative anomalies in the 500-hPa geopotential height field. Similarly, the negative score values on the map corresponded to positive geopotential height anomalies.

For this work, the largest magnitude of the PC loading value corresponded to the largest covariance between the outbreak and the principal component, and thus the stronger the signal. The magnitudes of the PC loading then could be used to identify examples of actual tornado outbreaks that best represent that PC score pattern, as we have below.

In our analysis, the majority of loadings (roughly 90%) in the Varimax and Promax transformations for the first PC pattern were characterized by negative PC loading values, with the greatest magnitude indicated by loading 27 (red circle in Fig. 11). The first PC score for both transformations captured a long-wave trough of negative score values centered over the north-central part of the domain and extending southward to Mexico (Fig. 11, left map). The long-wave trough was accompanied by a distinct region of contrasting score values positioned over New England.

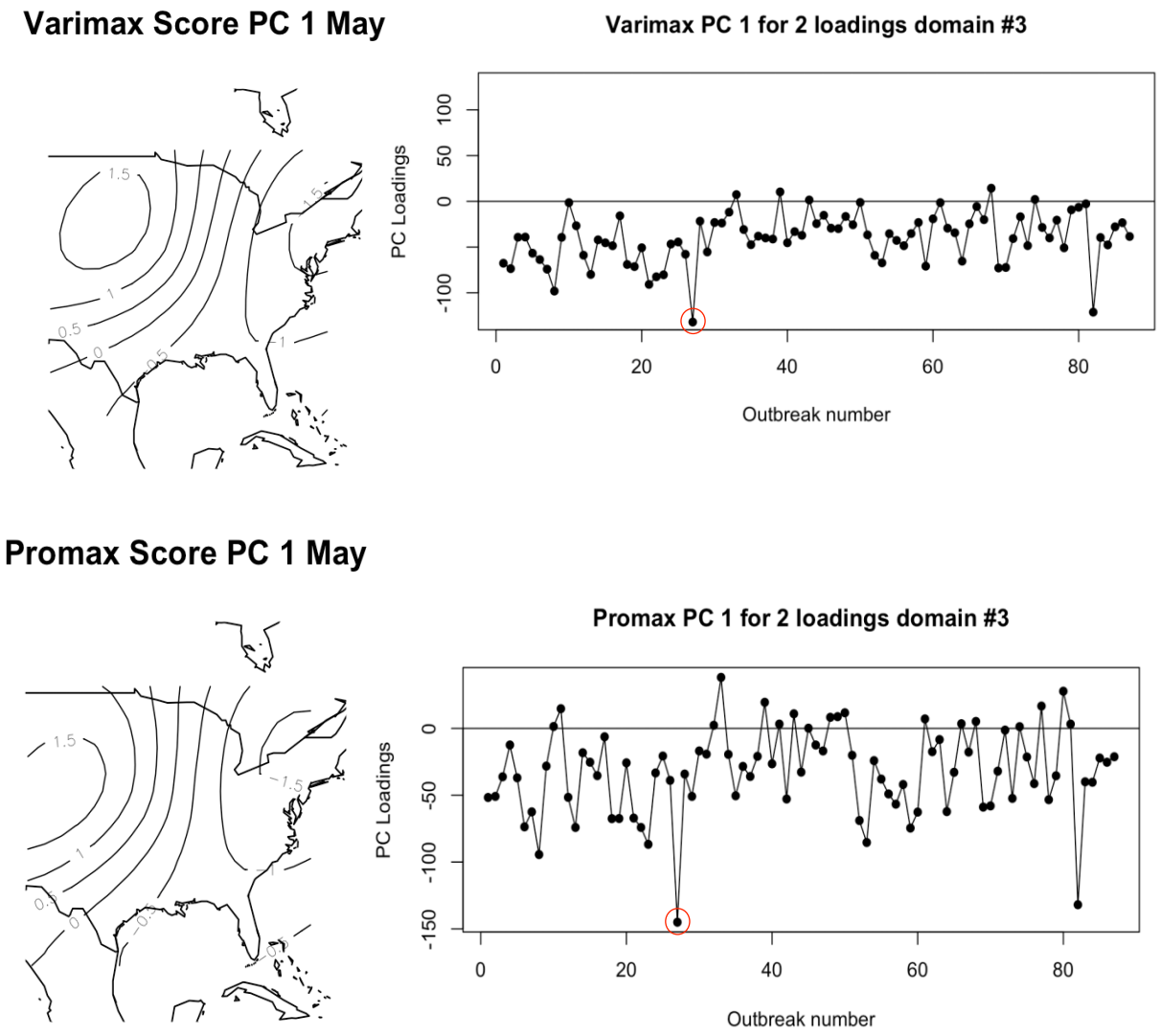

Fig. 11. The scores (left) for the first PC pattern and corresponding loadings (right) for Varimax (top) and Promax (bottom) transformations when retaining two loadings in Domain 3 for May from 1950 to 2011. The red circle indicates the outbreak most representative of this PC loading for all May tornado outbreaks.

The most representative case for the first PC pattern was tornado outbreak 27. This outbreak occurred on 1 May 1967 and included 10 tornadoes (nine F2, one F3) that caused damage in three

states: Texas, Louisiana, and Mississippi. The outbreak started at 0600 CST in Texas and ended at 0000 CST in Louisiana, injuring 9 people. Figure 12 displays the 500-hPa geopotential height anomalies for that event. (For better visualization of the mid-tropospheric patterns, the anomalies maps for actual outbreak events are displayed on the whole CONUS domain.) In this outbreak, the 500-hPa geopotential height anomalies form a long-wave trough of negative values centered over north-central part of the domain, north-northwest of the tornado outbreak centroid location (Fig. 12.) The negative values indicated that prior to the tornado outbreak occurrence, the geopotential heights were much below average for this time of year.

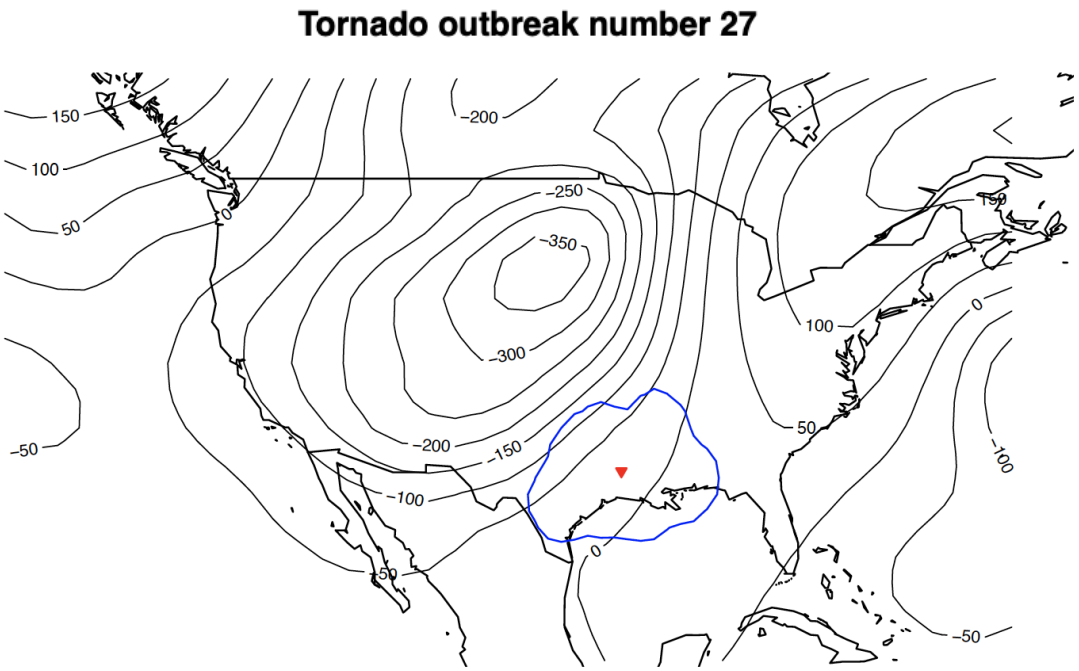

Fig. 12. The most representative tornado outbreak for the first PC pattern. Displayed are the 500-hPa height anomalies (contours, in meters), the KDE cluster that represented the geographical extent of this major tornado outbreak (blue outline), and the center of the tornado outbreak (red triangle).

The majority of loadings in both Varimax and Promax transformations for the second PC pattern also were characterized by negative PC loading values, with the greatest magnitude indicated by loading number 11 (Fig. 13).

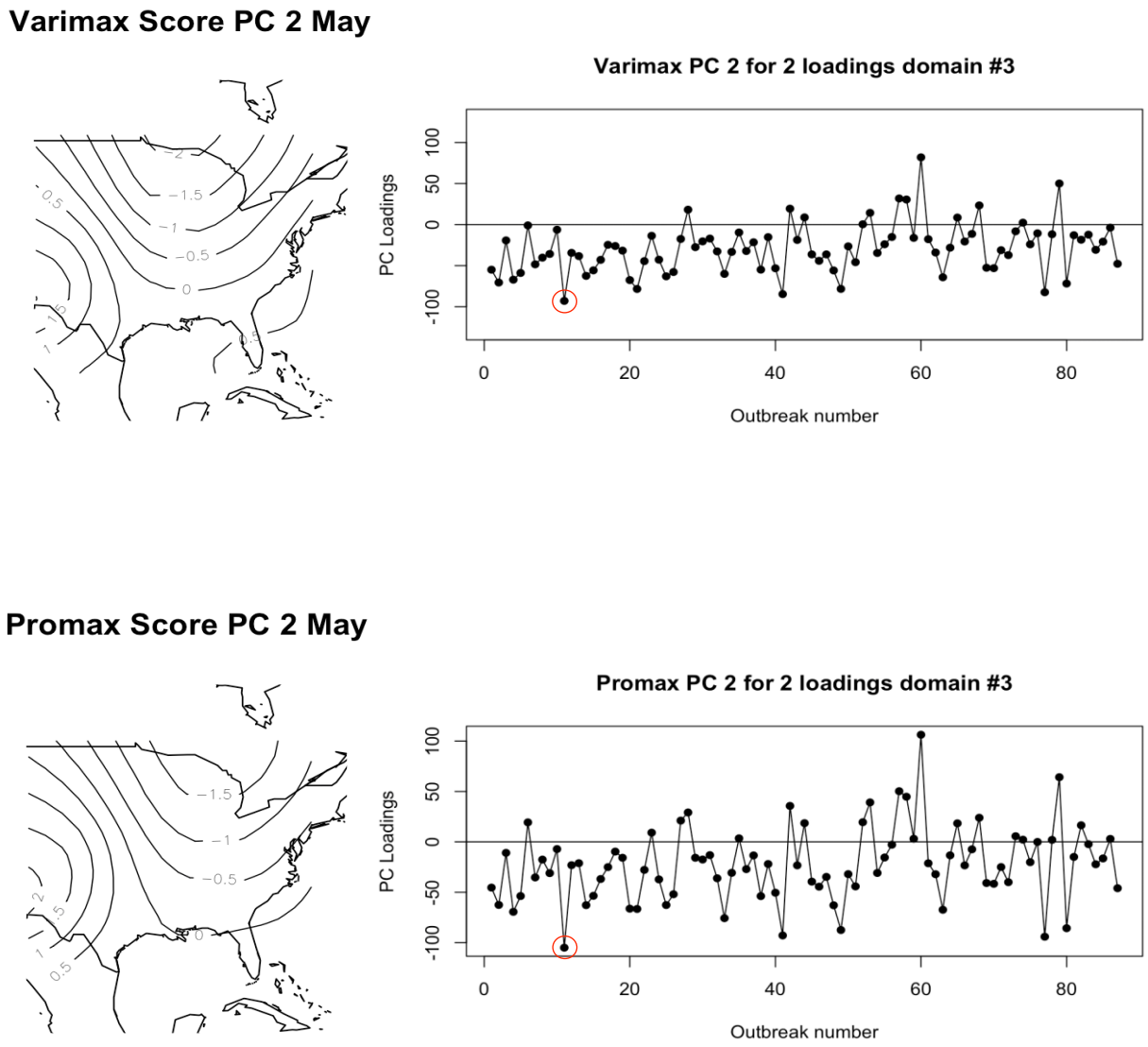

Fig. 13. The scores (left) and loadings (right) for Varimax and Promax transformation for two loadings in 3rd domain for the second PC pattern. The red circle indicates the most representative PC loading case for all May tornado outbreaks.

The second PC score in Varimax and Promax transformation showed a pronounced long-wave trough with positive score values (see Chapter 3, section 2.2), centered over northern Minnesota and spreading southward across the U.S. That trough was accompanied by a partially visible negative-tilt[1] PC trough spreading from the U.S. West through the central part of the country, with the strongest magnitude over the southern Rocky Mountains.

The most representative case for the second PC pattern was tornado outbreak number 11 (based on PC loading analysis). This outbreak occurred on 4 May 1959 and included 13 F2 tornados that occurred across four states: Oklahoma, Kansas, Iowa, and Wisconsin. The outbreak started around 1200 CST and ended about 0000 CST. There were no fatalities, but one injury was reported. Figure 14 presents the 500-hPa geopotential height anomalies for that event.

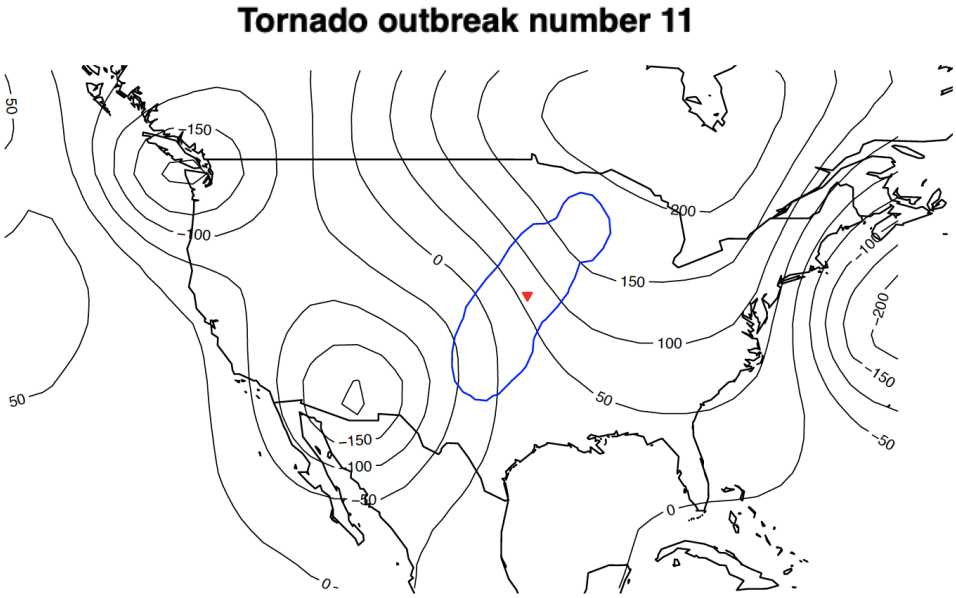

Fig. 14. The most representative tornado outbreak for the second PC pattern. Displayed are the 500-hPa height anomalies (contours, in meters), the KDE cluster that represented the geographical extent of this major tornado outbreak (blue outline), and the center of the tornado outbreak (red triangle).

[1] The 'negative-tilt' describes a trough that tilts from the northwest toward the southeast (here, the nomenclature is used only to describe the shape of PC score spatial pattern.)

In that event, a pronounced long-wave trough with positive values of 500-hPa geopotential height anomalies (200 m higher than the May average for the same area) extended from northern Michigan to the south across the U.S. Simultaneously, another pronounced long-wave trough with negative anomalies extended from the northwest through the central part of the country, with a strong center over Arizona and New Mexico. The geopotential heights for that area were 200 m lower than the average May values.

Our results indicated that there were two PC patterns that were most common for tornado outbreaks in May. To test the robustness of this outcome, we performed the same PC analysis and congruence coefficient testing for all four domains. We identified the same two patterns across all domains, though with different values for the congruence coefficients (Appendix B). In addition, to help validate the results from the PC analysis, we applied two clustering methods: hierarchical clustering and silhouette clustering. Figure 15 displays results from the agglomerative hierarchical clustering method using a cluster dendrogram.

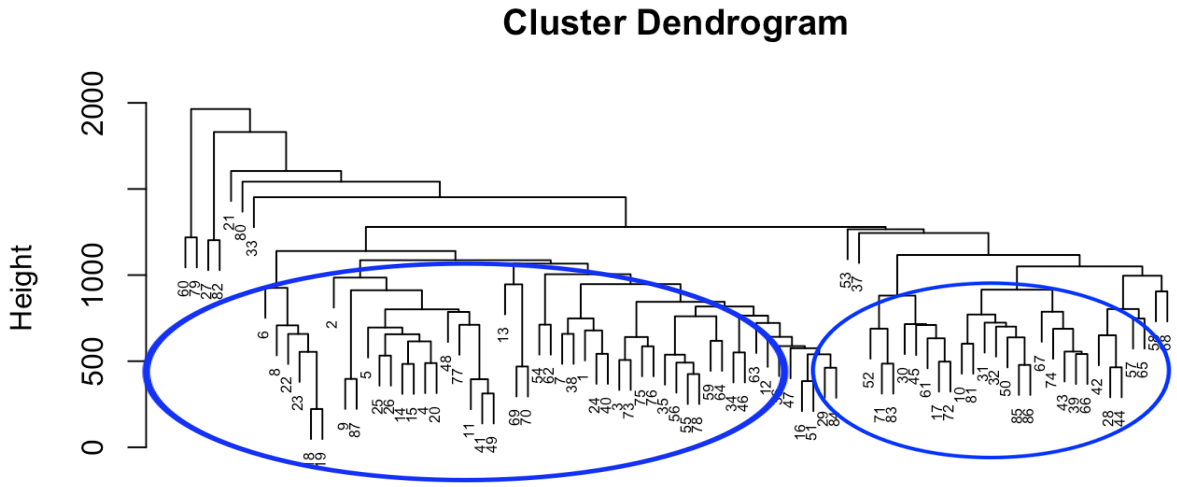

Fig. 15. Plot of 87 tornado outbreaks for May clustered using hierarchical clustering method for Domain 3. Circles depict two distinct groups. The y-axis refers to the Euclidean distance calculated using method of average linkage.

The hierarchical clustering identified two distinct groups. One cluster consisted of 29 tornado outbreaks (Fig. 15, smaller circle), with the smallest dissimilarity between outbreaks 28 and 44. Outbreak 28 (Fig. 16, left) occurred on 6 May 1967, only 5 days after the occurrence of the outbreak 27 (described in the first PC pattern above). Outbreak 28 affected Alabama, Georgia, Louisiana, Arkansas, Tennessee, and Kentucky with seven tornados: six F2 and one F3. The outbreak lasted from about 1800 CST to about 0600 CST the following day. There was one fatality and 31 people were injured. Outbreak 44 (Fig. 16, right) occurred on 9 May 1981 in Texas. It started at 1200 CST and finished 6 hours later, initiating 9 tornados: seven F2 and two F3. There were no reports of fatalities or injuries. The 500-hPa geopotential height anomalies patterns for both outbreaks were characterized by a long-wave trough with negative anomalies extending southward from south-central Canada to the central U.S. In both cases, the tornado outbreaks occurred southeast of the trough center. Additionally, the trough was accompanied by two centers of positive anomalies — over New England and the northwestern part of the domain.

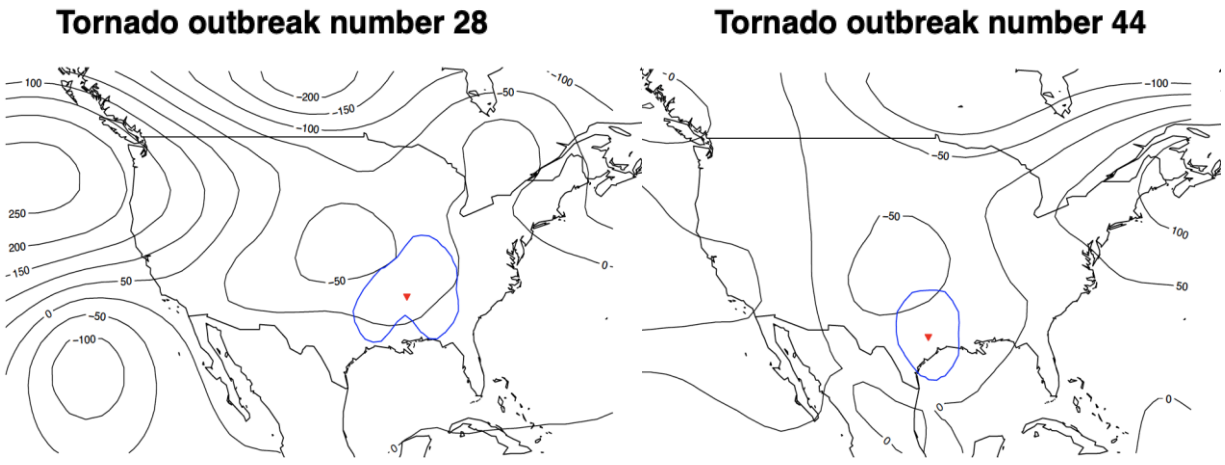

Fig. 16. Tornado outbreaks number 28 (left) and number 44 (right), identified as those most representative (least dissimilar) within the first group in the hierarchical clustering. Displayed are the 500-hPa height anomalies

(contours, in meters), the KDE cluster that represented the geographical extent of this major tornado outbreak (blue outline), and the center of the tornado outbreak (red triangle).

The second group identified by the hierarchical method was composed of 51 tornado outbreaks (Fig. 15, bigger circle). The smallest dissimilarity was for tornado outbreaks 18 and 19. These outbreaks, however, occurred on two consecutive days (6 and 7 May 1961, respectively). The 500-hPa geopotential heights anomaly pattern for both days look almost the same (Fig. 17).

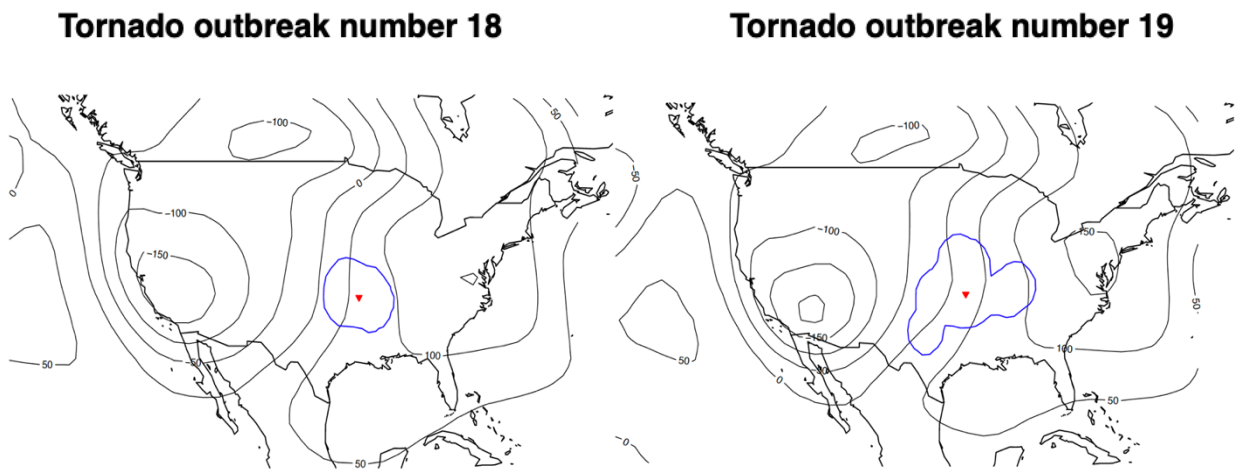

Fig. 17. Tornado outbreak number 18 (left) and number 19 (right) on two consecutive days. Displayed are the 500-hPa height anomalies (contours, in meters), the KDE cluster that represented the geographical extent of this major tornado outbreak (blue outline), and the center of the tornado outbreak (red triangle).

Therefore, effectively both outbreaks depict only one pattern that progressed in 24-hour time frame. Accordingly, the next smallest dissimilarity was for tornado outbreaks 41 and 49 (Fig. 18), selected as most representative. Tornado outbreak number 41 occurred on 20 May 1977 from 1500 CST to 0000 CST. There were 10 tornadoes (seven F2, three F3) that affected Oklahoma and Texas, with three people injured. Tornado outbreak 49 occurred in Texas on 12 May 1982 and

injured 13 people. The outbreak started around 1500 CST and ended about 0000 CST. Both outbreaks were characterized by long-wave positive anomalies spreading from Hudson Bay to the Gulf of Mexico and a long-wave trough with negative anomalies from the northwest U.S. through the central part of the country. Both outbreaks occurred southeast of the trough center. Similarly, this pattern was identified by the PC analysis as the second pattern.

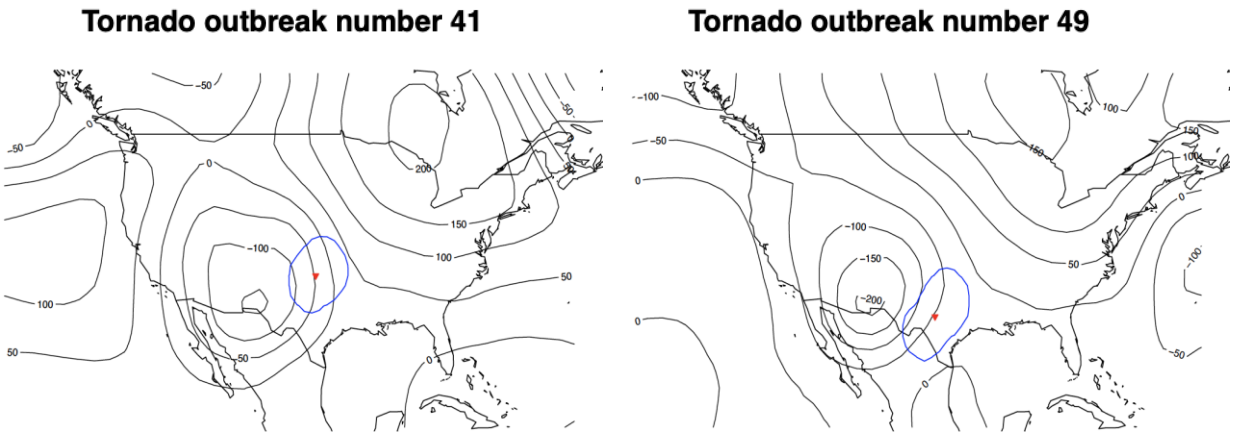

Fig. 18. Tornado outbreak number 41 (left) and number 49 (right) identified as most representative (least dissimilarity) for the second group in the hierarchical clustering. Displayed are the 500-hPa height anomalies (contours, in meters), the KDE cluster that represented the geographical extent of this major tornado outbreak (blue outline), and the center of the tornado outbreak (red triangle).

Finally, the last method used to identify mid-tropospheric patterns associate with major tornado outbreaks was silhouette clustering. As noted above, the novelty of that method is that, unlike PCA and hierarchical clustering, it could identify cohesion within each cluster. All 87 May tornado outbreaks were grouped into clusters based on the average silhouette scores (Fig. 19). Then, we calculated a silhouette value for each outbreak in the cluster. The higher the silhouette value, the more cohesion there was between the outbreak and the whole cluster.

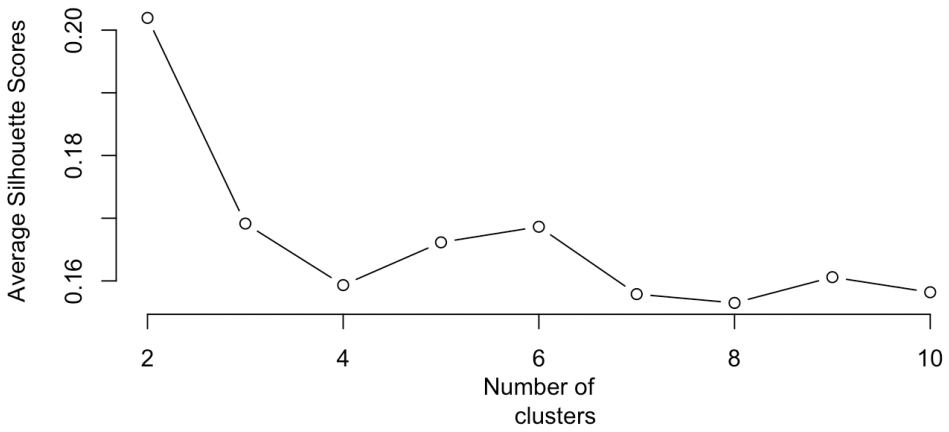

Fig. 19. Average silhouette score plot used to determine the number of clusters for the silhouette clustering.

As depicted in Fig. 20, silhouette clustering also identified two distinct groups of similar tornado outbreaks. The first group consisted of 47 tornado outbreaks, with the highest silhouette score (0.358) represented by outbreak 28. (The same tornado outbreak was identified and described in the hierarchical clustering method.)

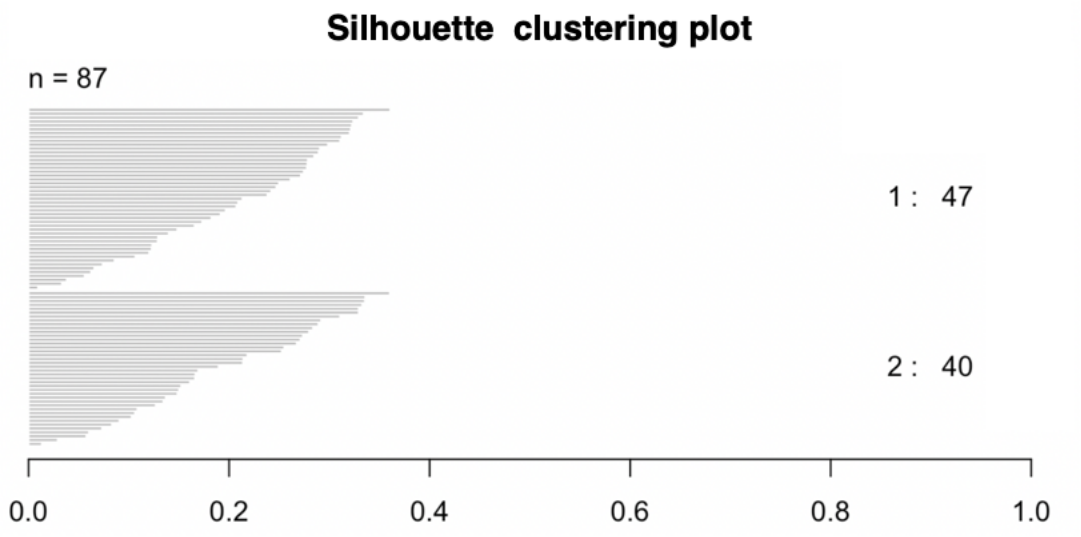

Fig. 20. The silhouette plot with two separate clusters: the top comprised 47 tornado outbreaks, the bottom comprised 40 tornado outbreaks. The average silhouette score for cluster 1 was 0.21 and for cluster 2 was 0.20.

The second group consisted of 40 tornado outbreaks. The most representative was outbreak 20 (silhouette score of 0.357). This outbreak started at 1500 CST on 28 May 1962 and ended at 3000 CST the following day. It included 11 F2 tornadoes across Texas, Oklahoma, Kansas, Iowa, Illinois, and Missouri. According to the SPC database, five people were injured. The 500-hPa geopotential height anomaly pattern for the outbreak 20 was similar to those of outbreaks 11, 41, and 49 (previously described). Namely, it was characterized by a long-wave trough with negative anomalies from the northwest U.S. through the central part of the country, with a closed contour over Arizona and New Mexico. Both of the patterns resulting from silhouette analysis were consistent with results obtained from the PC and hierarchical analyses, where tornado outbreaks 28 and 20 represented the first and second patterns, respectively (Fig. 21).

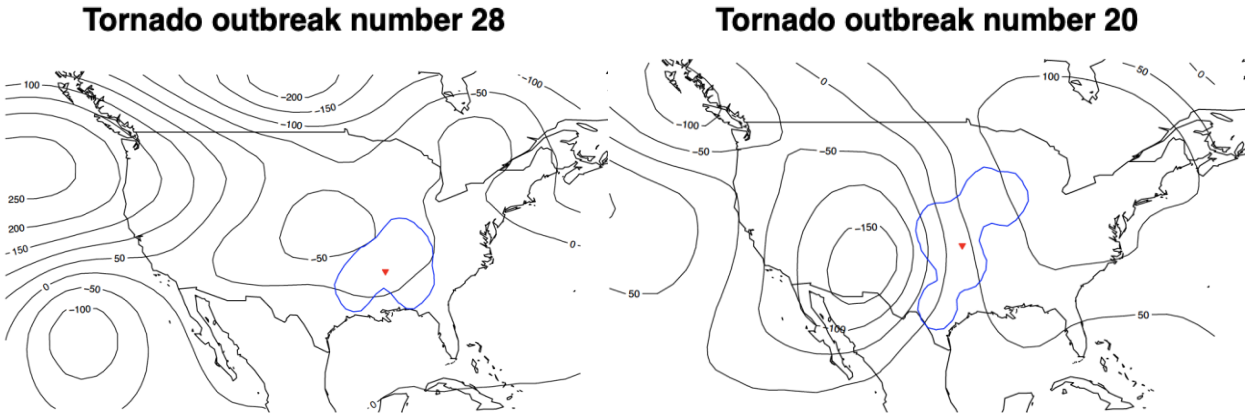

Fig. 21. Tornado outbreaks 28 (left) and 20 (right), identified as most representative (highest silhouette values) for the first and second group, respectively, in the silhouette clustering analysis. Displayed are the 500-hPa height anomalies (contours, in meters), the KDE cluster that represented the geographical extent of this major tornado outbreak (blue outline), and the center of the tornado outbreak (red triangle).

4. Conclusion and Discussion

This research identified large-scale patterns in 500-hPa geopotential height anomalies associated with major tornado outbreaks in the United States for May from 1950 to 2011. Tornadoes rated F2 and higher (i.e., significant tornados) were grouped into 24-hour periods and clustered into tornado outbreaks using kernel density estimation. Major tornado outbreaks were defined as outbreaks with seven or more significant tornados, resulting in 87 major tornado outbreaks for May. Based on temporal characteristics of those outbreaks (i.e., start time and location obtained from the KDE analysis), corresponding 500-hPa geopotential heights were obtained from the NOAA-CIRES Twentieth Century Reanalysis.

Through several statistical tests (principal component analysis, hierarchical clustering, and silhouette clustering), we concluded that there were two main atmospheric patterns associated with major tornado outbreaks in May. The first pattern was characterized by a long-wave, positive-tilt trough of strong negative anomalies in 500-hPa geopotential heights extending southward from south-central Canada through the north-central U.S. (Fig. 22). Additionally, two regions of high anomalies were localized over New England and in the far northwestern part of the domain. The second pattern was characterized by a long-wave, strong negative-tilt trough of negative anomalies extending from southwestern parts of British Columbia and Washington to north-central Mexico, with a closed low over Arizona and New Mexico. It was accompanied by a long-wave trough of positive anomalies spreading from Hudson Bay and the Great Lakes southward towards Florida and the Gulf of Mexico (Fig. 22). Both patterns were consistently identified through all statistical methods used in this research.

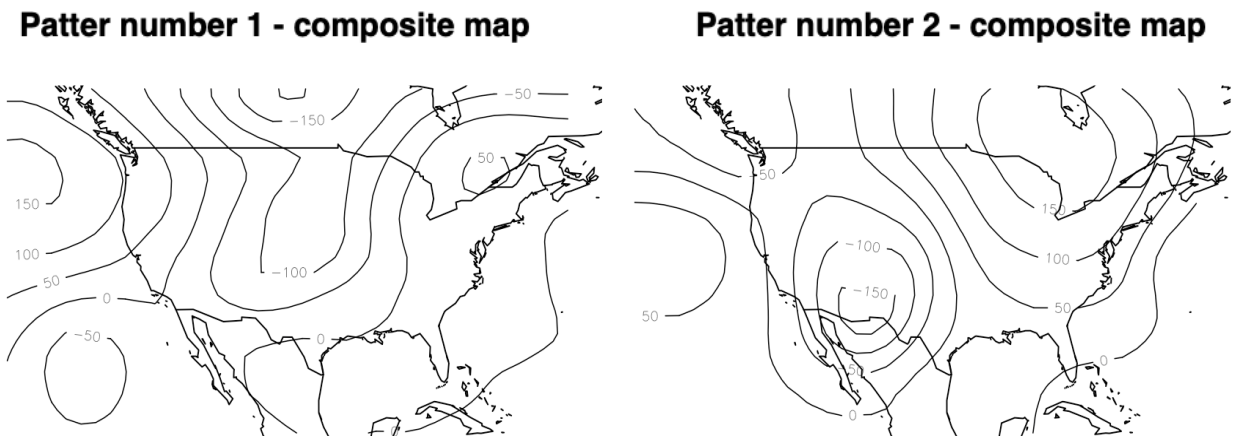

Fig. 22. The 500-hPa geopotential height anomalies (in meters) with tornado outbreaks in May: composite map of tornado outbreaks specified as most representative for the first (left) and second (right) pattern across all statistical methods.

All tornado outbreaks mentioned in this research were found to occur east-southeast from the mid-level trough axis, which was consistent with previous case studies on synoptic characteristics associated with the tornado outbreak occurrence [51-53]. The negative anomalies associated with the trough indicated much lower geopotential heights than the average values for May (see Chapter 3, section 3: Results). On the contrary, areas with the positive anomalies indicated higher than average values of geopotential heights for May. Because geopotential-height surfaces are low in colder air masses and high in warmer ones, similar patterns to the geopotential height anomalies may be found in 500-hPa temperature fields. To verify this hypothesis, however, an additional analysis of the mid-tropospheric temperature fields for the same tornado outbreaks would be required. This step might serve as a future point of departure for pattern identification and comparison of results between various atmospheric fields.

This research is novel not only because it is a successful case of tornado outbreak pattern identification, but more importantly, it obtained consistent and robust results across three independent statistical methods. However, there are few limitations to this work. First, although two patterns were found that were associated with major tornado outbreaks, we did not examine whether they also occur without association with severe weather (null cases). Future work needs to address this issue to determine the uniqueness of these two patterns to major tornado outbreaks.

Another limitation of the study was that the 6-hourly 500-hPa geopotential height fields represented one 'frozen' snapshot of an evolving convective situation. It is possible that the chosen snapshots do not depict the state of the atmosphere at the best time during the evolution of each individual outbreak. Similarly, a different reanalysis dataset with higher temporal or spatial resolution may be more suitable to represent the 500-hPa geopotential height fields. Datasets with hourly resolution, however, will be first available for use in 2020 (ERA5 from European Centre for Medium-Range Weather Forecasts [54]).

Following from this study, future research will include the identification of patterns for other months, perhaps resulting in more pattern types. Results possibly may open a pathway for improvements in tornado outbreak prediction across a portion of the sub-seasonal timescale. In addition, we intend to determine if these patterns are well represented in the historical dataset of a global climate model and, if so, study how these patterns may change with future climate change.

5. References

1. Brooks, H. E., 2004: On the relationship of Tornado path length and width to intensity. *Weather Forecast.*, **19**, 310–319.

2. Doswell III, C. A., and D. Schultz, 2006: On the use of indices and parameters in forecasting severe storms. *Electron. J. Sev. Storms Meteorol.*, **1**, 1–22.

3. Knupp, K. R., and Coauthors, 2014: Meteorological Overview of the Devastating 27 April 2011 Tornado Outbreak. *Bull. Am. Meteorol. Soc.*, **95**, 1041–1062.

4. Hayes J.L.,2011: The historic Tornadoes of April 2011. U.S. Department of Commerce, NOAA, *CreateSpace Independent Publishing Platform*, Silver Spring, Maryland.

5. Schneider, R. S., A. R. Dean, S. J. Weiss, and P. D. Bothwell, 2006: Analysis of estimated environments for 2004 and 2005 severe convective storm reports. *23rd AMS Conference on Severe Local Storms*, St Louis, MO, Amer. Meteor. Soc., 3.5.

6. Rasmussen, E. N., 2003: Refined Supercell and Tornado Forecast Parameters. *Weather Forecast.*, **18**, 530–535.

7. Doswell III, C. A., 1987: The Distinction between Large-Scale and Mesoscale Contribution to Severe Convection: A Case Study Example. *Weather Forecast.*, **2**, 3–16.

8. Sherburn, K. D., and M. D. Parker, 2014: Climatology and Ingredients of Significant Severe Convection in High-Shear, Low-CAPE Environments. *Weather Forecast.*, **29**, 854–877.

9. Grams, J. S., R. L. Thompson, D. V. Snively, J. A. Prentice, G. M. Hodges, and L. J. Reames, 2012: A Climatology and Comparison of Parameters for Significant Tornado Events in the United States. *Weather Forecast.*, **27**, 106–123.

10. Mercer, A. E., C. M. Shafer, C. A. Doswell III, L. M. Leslie, and M. B. Richman, 2012: Synoptic Composites of Tornadic and Nontornadic Outbreaks. *Mon. Weather Rev.*, **140**, 2590–2608.

11. Doswell III, C. A., R. Edwards, R. L. Thompson, J. A. Hart, and K. C. Crosbie, 2006: A Simple and Flexible Method for Ranking Severe Weather Events. *Weather Forecast.*, **21**, 939–951.

12. Thompson, R. L., R. Edwards, and C. M. Mead, 2004: An update to the supercell composite and significant tornado parameters. *Preprints, 22nd Conf. on Severe Local Storms*, Vol. 8 of, Hyannis, MA, Amer. Meteor. Soc., P8.1.

13. Moore, T. W., 2018: Annual and seasonal tornado trends in the contiguous United States and its regions. *Int. J. Climatol.*, **38**, 1582–1594.

14. 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. Weather Rev.*, **137**, 1250–1271.

15. Mercer, A. E., C. M. Shafer, C. A. Doswell, L. M. Leslie, and M. B. Richman, 2009: Objective Classification of Tornadic and Nontornadic Severe Weather Outbreaks. *Mon. Weather Rev.*, **137**, 4355–4368.

16. Lee, S.-K., R. Atlas, D. Enfield, C. Wang, and H. Liu, 2013: Is There an Optimal ENSO Pattern That Enhances Large-Scale Atmospheric Processes Conducive to Tornado Outbreaks in the United States? *J. Clim.*, **26**, 1626–1642.

17. Knowles, J. B., and R. A. Pielke, 2005: The Southern Oscillation and its effects on tornadic activity in the United States. *Atmospheric Sciences Paper 755*, Colorado State University, 15.

18. Bove, M. C., 1998: Impacts of ENSO on United States tornadic activity. *Preprints, 19th Conf. on Severe Local Storms,* Minneapolis, MN, Amer. Meteor. Soc., 313–316.
19. Hagemeyer, B. C., 1998: Significant extratropical tornado occurrences in Florida during strong El Niño and strong La Niña events. *Preprints, 19th Conf. on Severe Local Storms*, Minneapolis, MN, Amer. Meteor. Soc., 412–415.
20. Browning, P., 1998: ENSO related severe thunderstorm climatology of northwest Missouri. *Preprints, 19th Conf. on Severe Local Storms*, Minneapolis, MN, Amer. Meteor. Soc., 291–292.
21. Cook, A. R., and J. T. Schaefer, 2008: The Relation of El Niño–Southern Oscillation (ENSO) to Winter Tornado Outbreaks. *Mon. Weather Rev.*, **136**, 3121–3137.
22. Molina, M. J., J. T. Allen, and V. A. Gensini, 2018: The Gulf of Mexico and ENSO Influence on Subseasonal and Seasonal CONUS Winter Tornado Variability. *J. Appl. Meteorol. Climatol.*, **57**, 2439–2463.
23. Raddatz, R. L., and J. D. Cummine, 2003: Inter-annual Variability of Moisture Flux from the Prairie Agro-ecosystem: Impact of Crop Phenology on the Seasonal Pattern of Tornado Days. *Boundary-Layer Meteorol.*, **106**, 283–295.
24. Fujita, T. T., D. L. Bradbury, and C. F. Van Thullenar, 1970: Palm Sunday Tornadoes of April 11, 1965. *Mon. Weather Rev.*, **98**, 29–69.
25. Corfidi, S. F., S. J. Weiss, J. S. Kain, S. J. Corfidi, R. M. Rabin, and J. J. Levit, 2010: Revisiting the 3–4 April 1974 Super Outbreak of Tornadoes. *Weather Forecast.*, **25**, 465–510.
26. Roebber, P. J., D. M. Schultz, and R. Romero, 2002: Synoptic Regulation of the 3 May 1999 Tornado Outbreak. *Weather Forecast.*, **17**, 399–429.

27. Schaefer, J. T., and R. Edwards, 1999: The SPC tornado/severe thunderstorm database. *Preprints, 11th Conf. on Applied Climatology*, Vol. 6 of, Dallas, TX, Amer. Meteor. Soc.

28. Brooks, H., and C. A. Doswell, 2001: Some aspects of the international climatology of tornadoes by damage classification. *Atmos. Res.*, **56**, 191–201.

29. Stephanie, M. V., 2004: Leveling the field for tornado reports through time: Inflation-adjustment of annual tornado reports and objective identification of extreme tornado reports. *22nd Conference on Severe Local Storms*, Hyannis, MA, Amer. Meteor. Soc., 7B.3.

30. Doswell, C. A., H. E. Brooks, and N. Dotzek, 2009: On the implementation of the enhanced Fujita scale in the USA. *Atmos. Res.*, **93**, 554–563.

31. Doswell, C. A., 1980: Synoptic-Scale Environments Associated with High Plains Severe Thunderstorms. *Bull. Am. Meteorol. Soc.*, **61**, 1388–1400.

32. Verbout, S. M., H. E. Brooks, L. M. Leslie, and D. M. Schultz, 2006: Evolution of the U.S. Tornado Database: 1954–2003. *Weather Forecast.*, **21**, 86–93.

33. , P. R., H. E. Brooks, and C. A. Doswell III, 2000: Climatological risk of strong and violent tornadoes in the United States. *Preprints, 2nd Symp. on Environmental Applications*, Long Beach, CA, Amer. Meteor. Soc., 9.4.

34. 20th Century Reanalysis V2c data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, Available [online]: https://www.esrl.noaa.gov/psd/data/gridded/data.20thC_ReanV2c.html.

35. Compo, G. P., and Coauthors, 2011: The Twentieth Century Reanalysis Project. *Q. J. R. Meteorol. Soc.*, **137**, 1–28.

36. Hotelling, H., 1933: Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.*, **24**, 498–520.

37. Richman, M. B., 1986: Rotation of principal components. *J. Climatol.*, **6**, 293–335.

38. Preseindorfer, R.W. and C. Mobley, 1988: Principal component analysis in meteorology and oceanography. *Elsevier*, New York, 425.

39. Preisendorfer, R. W., and C. D. Mobley, 1982: Data intercomparison theory. III. S-phase and T-phase tests for spatial pattern and temporal evolution, NOAA Technical Memorandum, ERL PMEL-40, Seattle, Washington

40. Wilks, D. S., 2006: Statistical Methods in the Atmospheric Sciences. *Academic Press*, 627.

41. Mercer, A. E., M. B. Richman, and L. M. Leslie, 2011: Identification of severe weather outbreaks using kernel principal component analysis. *Procedia Comput. Sci.*, **6**, 231–236.

42. Richman, M. B., and P. J. Lamb, 1985: Climatic Pattern Analysis of Three- and Seven-Day Summer Rainfall in the Central United States: Some Methodological Considerations and a Regionalization. *J. Clim. Appl. Meteorol.*, **24**, 1325–1343.

43. Compagnucci, R. H., and M. B. Richman, 2008: Can principal component analysis provide atmospheric circulation or teleconnection patterns? *Int. J. Climatol.*, **28**, 703–726.

44. Harman, H. 1976: Modern Factor Analysis. *The University of Chicago Press*, Chicago, Illinois, 344-347.

45. Kalkstein, L. S., G. Tan, and J. A. Skindlov, 1987: An Evaluation of Three Clustering Procedures for Use in Synoptic Climatological Classification. *J. Clim. Appl. Meteorol.*, **26**, 717–730.

46. Dubes, R.C. and Jain, A.K., 1988: Algorithms for clustering data, *Prentice hall Englewood Cliffs*, p. 59.

47. Huth, R., I. Nemesova, and N. Klimperová, 1993: Weather categorization based on the average linkage clustering technique: An application to European mid-latitudes. *Int. J. Climatol.*, **13**, 817–835.

48. Zhao, Y., G. Karypis, and U. Fayyad, 2005: Hierarchical Clustering Algorithms for Document Datasets. *Data Min. Knowl. Discov.*, **10**, 141–168.

49. , P. J., 1987: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.*, **20**, 53–65.

50. Hartigan, J. A., and M. A. Wong, 1979: Algorithm AS 136: A K-Means Clustering Algorithm. *Appl. Stat.*, **28**, 100.

51. 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. *Weather Forecast.*, **21**, 449–464.

52. Ferguson, E. W., F. P. Ostby, and P. W. Leftwich, 1987: The Tornado Season of 1985. *Mon. Weather Rev.*, **115**, 1437–1445.

53. 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. Weather Rev.*, **103**, 318–333.

54. Hersbach, H., 2016. The ERA5 Atmospheric Reanalysis. *In AGU Fall Meeting Abstracts.*

Chapter IV: Conclusion

Major tornado outbreaks have created billions of dollars in damages and caused death, injury, or distress to thousands of people. Throughout the historic record, numerous studies have been conducted to help society mitigate risks associated with the destructive power of tornado outbreaks. Accordingly, this study aimed to improve understanding of some atmospheric characteristics associated with tornado outbreaks. Considering the ambiguity of the term 'tornado outbreak,' the initial process of tornado outbreak identification required extensive review of the literature. The diversity of definitions and classifications found in research conducted over past few decades was extensive. The choice of an approach to specify a 'tornado outbreak' was found to be dependent upon the investigator's research purpose and technologies that were available for a particular time.

Our study offered a data-driven approach to define tornado outbreaks. Groups of tornado outbreaks were determined based on the Storm Prediction Center severe weather database, which contains tornado reports from 1950 to 2017. All significant tornados (F2 and higher) were grouped into 24-h periods, from 0600 CST to 0559 CST, and then assigned into clusters using kernel density estimation (KDE) analysis. Our research resulted in 4,991 clusters of significant tornados. Then, a threshold of seven or more tornados in a tornado outbreak was applied to obtain major tornado outbreaks. As a result, 333 major tornado outbreaks were found to occur east of the Rocky Mountains over the 68 years of the analysis. The total count of major tornado outbreaks by month indicated that the most outbreaks occurred in May (89 outbreaks) and April (79 outbreaks). Not surprisingly, the analysis indicated a second tornado outbreak season, with the seasonal peak of tornado outbreaks in November (26 outbreaks) and December (18 outbreaks). Further, according

to the results, it can be expected to see on average 4 to 5 major tornado outbreaks a year, one of which would possibly occur in May.

This study offered one of many possible statistical approaches that can be applied to quantify tornado outbreaks. The advantage of our approach is that it can be adapted to different purposes of tornado outbreak quantification by a change in either KDE parameters or the threshold value indicating the number of tornados to retain. An example of such research would be the statistical comparison of different threshold values for the yearly frequency distribution of tornado outbreak clusters, providing new information regarding changes in the number of major tornado outbreaks over time.

The results of our KDE analysis were used to initiate an analysis of 500-hPa atmospheric patterns which are associated with major tornado outbreaks. Based on temporal characteristics of major tornado outbreaks, corresponding 500-hPa geopotential heights were obtained from NOAA-CIRES Twentieth Century Reanalysis. Applying three independent statistical tests — PCA, hierarchical clustering, and silhouette clustering, we concluded that there are two main atmospheric patterns associated with major tornado outbreaks in May. Both patterns were characterized by the occurrence of strong negative anomalies in 500-hPa geopotential heights.

The first pattern had a long wave of strong negative anomalies that extended over the central part of the U.S. The second pattern extended from southwestern parts of British Columbia and Washington to north-central Mexico. In both cases, tornado outbreaks were found to occur east southeast of the mid-level trough axis. The negative anomalies associated with the trough indicated much lower geopotential heights than the average values for May.

One limitation of this study is that the atmospheric fields were selected according to a specific group of selected events, namely tornado outbreaks. It is possible that the identified

patterns occurred also during times with no severe weather. Further, the ’frozen’ snapshot of 500-hPa geopotential height fields used in this study may not be adequately capturing the evolving convective situation associated with tornado outbreaks. A higher temporal resolution reanalysis dataset, such as an hourly estimate of geopotential heights, may be more suitable to determine the accurate snapshots.

Finally, future research related to this study can focus on the identification of patterns for the remaining months, especially April, November, and December. In addition, it would be important to compare the April results with the patterns found for May and, ultimately, with patterns found for the second tornado outbreak season (Oct-Dec). This analysis potentially could result in more pattern types, depending on season, and possibly could open a pathway for improvements in tornado outbreak prediction across a portion of the sub-seasonal timescale.

Appendix A:

Table A.1. Frequency of tornado outbreaks from 1950 to 2017 using KDE analysis.

# of tornadoes in an outbreak	# of outbreaks that happened from 1950 to 2017	number of tornadoes in outbreak: 7 and up	number of tornadoes in outbreak: 11 and up	number of tornadoes in outbreak: 15 and up	number of tornadoes in outbreak: 20 and up
2	128				
3	336				
4	241				
5	137				
6	112				
7	67	67			
8	44	44			
9	38	38			
10	37	37			
11	23	23	23		
12	23	23	23		
13	16	16	16		
14	12	12	12		
15	12	12	12	12	
16	14	14	14	14	
17	5	5	5	5	
18	3	3	3	3	
19	5	5	5	5	
20	4	4	4	4	4
21	1	1	1	1	1
22	3	3	3	3	3
23	3	3	3	3	3
24	5	5	5	5	5
25	3	3	3	3	3
26	2	2	2	2	2
27	5	5	5	5	5
28	1	1	1	1	1
29	2	2	2	2	2
32	1	1	1	1	1
33	1	1	1	1	1
34	1	1	1	1	1
51	1	1	1	1	1
96	1	1	1	1	1
680	**1287**	**333**	**147**	**73**	**34**
Average # of tornado outbreaks per year	**18.93**	**4.90**	**2.16**	**1.07**	**0.50**

Appendix B:

1. Domain #1: "wide CONUS domain" (west =-140, east=-60, north=55, south=20)

Eigenvalues: 189554.609, 82591.278, 55912.567, 40734.653, 33934.277, 30900.772, 19705.997, 13183.753, 11753.130, 8070.31

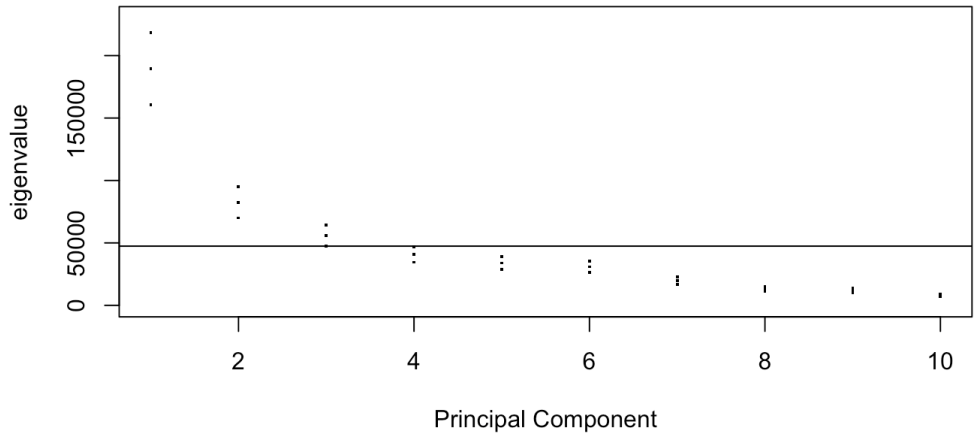

Fig. B.1.1. Plot of first 10 eigenvalues for the "wide CONUS domain."

- lower limit of 3rd PC: 47435.13
- upper limit of 4th PC: 46910.83
- variance explained with first 3 eigenvalues: 61%

Number of loadings	Varimax min	Varimax mean	Promax min	Promax mean
2	0.862	0.908	0.819	0.876
3	0.906	0.933	0.890	0.912
4	0.749	0.865	0.782	0.862
5	0.467	0.814	0.572	0.787
6	0.576	0.810	0.542	0.751
7	0.574	0.810	0.538	0.743
8	0.720	0.829	0.663	0.817
9	0.739	0.834	0.674	0.793
10	0.708	0.817	0.674	0.786

Tab. B.1.2. Absolute value of congruence coefficients for the first 10 loadings for the "wide CONUS domain."

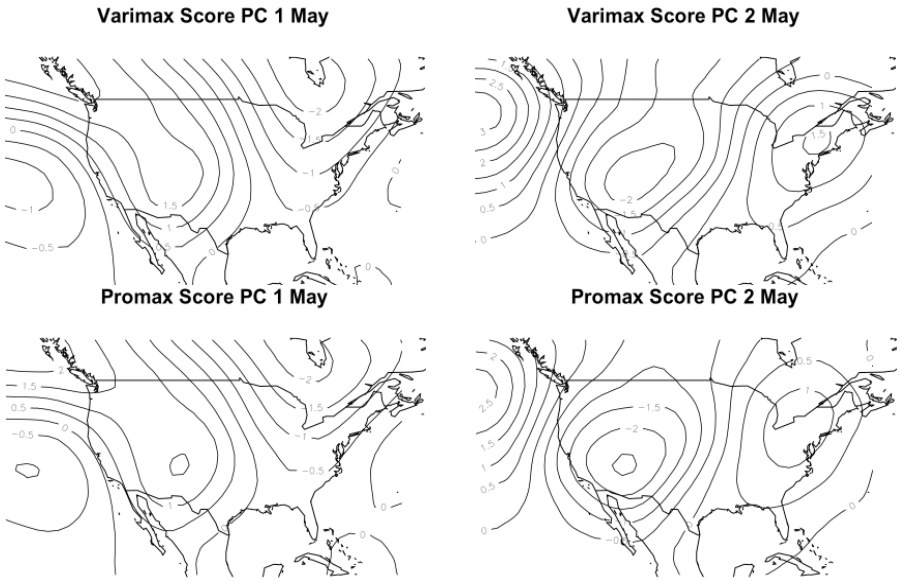

Fig. B.1.3. The Varimax and Promax transformation for two loadings in the "wide CONUS domain."

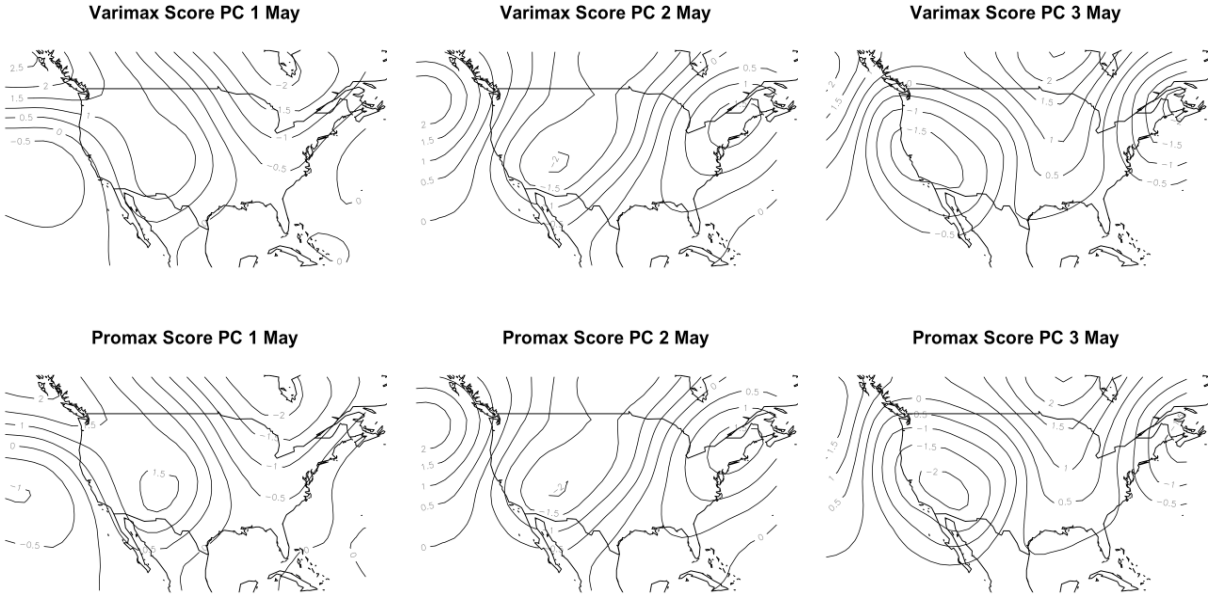

Fig. B.1.4. The Varimax and Promax transformation for three loadings in the "wide CONUS domain."

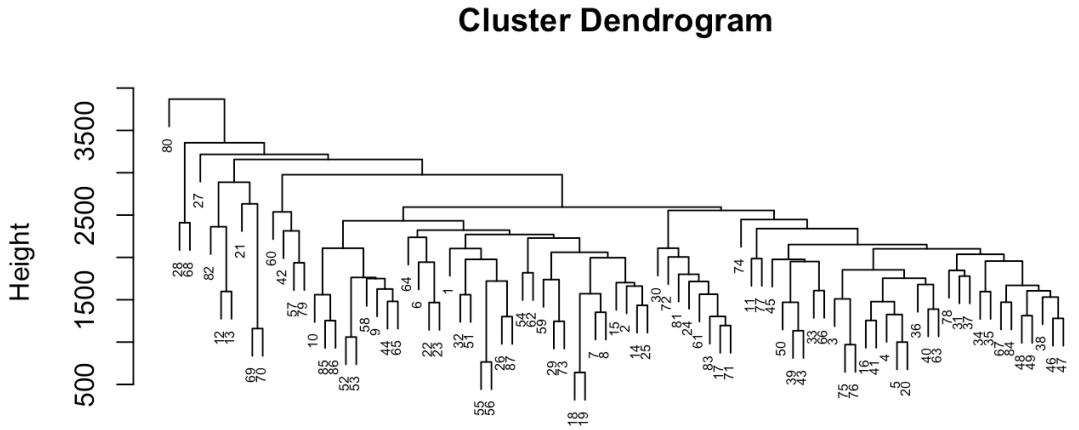

Fig. B.1.5. Dendrogram from Hierarchical clustering for the "wide CONUS domain."

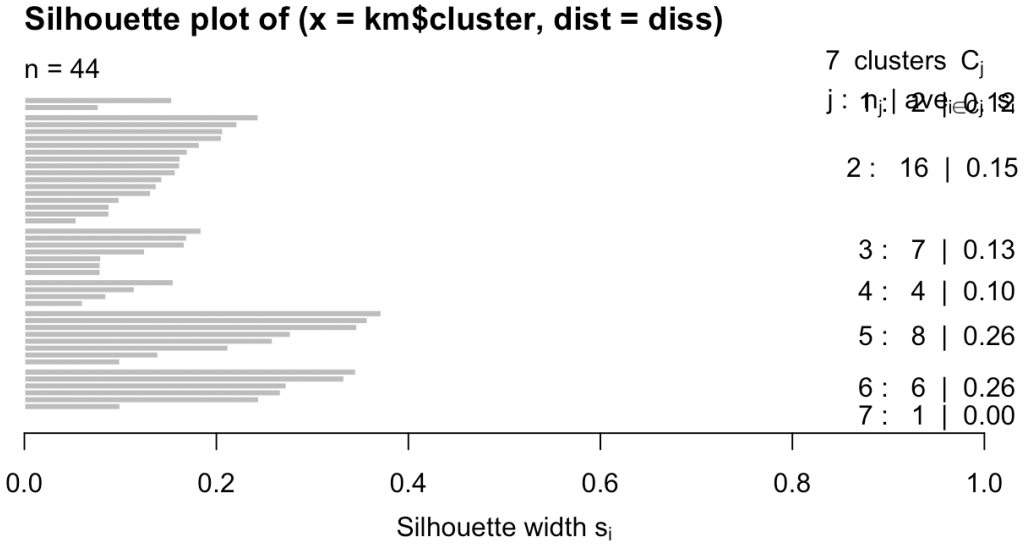

Fig. B.1.6. Silhouette clustering for the "wide CONUS domain." Clustering process removed 43 tornado outbreak cases.

2. Domain #2: “CONUS domain” (west =-126, east=-66, north=50, south=25)

Eigenvalues: 249697.584 , 55737.826, 40726.019, 37122.382, 27109.314, 15595.254, 12542.208, 9426.251, 5777.527, 4863.220

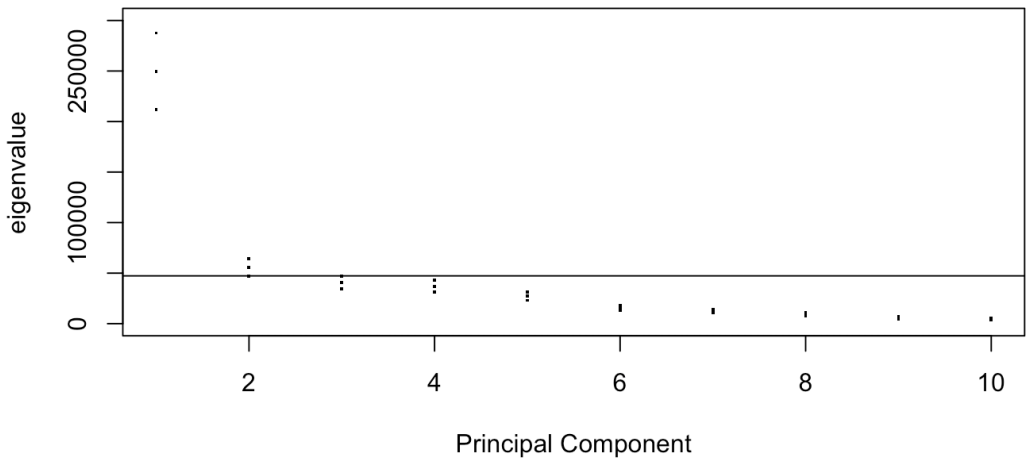

Fig. B.2.1. Plot of first 10 eigenvalues for the “CONUS domain.”

- lower limit of 2nd PC: 47286.88
- upper limit of 3rd PC: 46900.88
- variance explained with first 2 eigenvalues: 63 %

Number of loadings	Varimax min	Varimax mean	Promax min	Promax mean
2	0.956	0.968	0.881	0.885
3	0.860	0.916	0.797	0.844
4	0.688	0.852	0.700	0.810
5	0.571	0.807	0.673	0.757
6	0.498	0.761	0.506	0.714
7	0.467	0.787	0.429	0.692
8	0.432	0.758	0.380	0.664
9	0.261	0.686	0.403	0.645
10	0.379	0.689	0.431	0.642

Tab. B.2.2. Absolute value of congruence coefficients for 10 loadings for the “CONUS domain.”

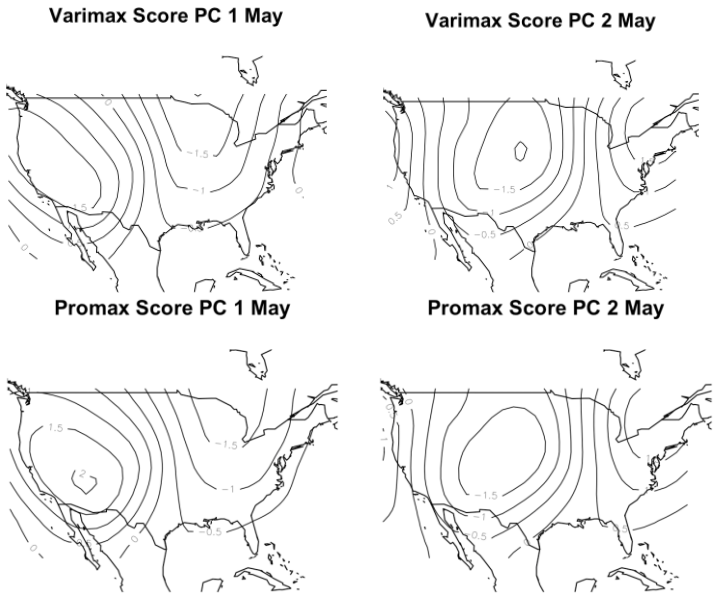

Fig. B.2.3. The Varimax and Promax transformation for two loadings in the “CONUS domain.”

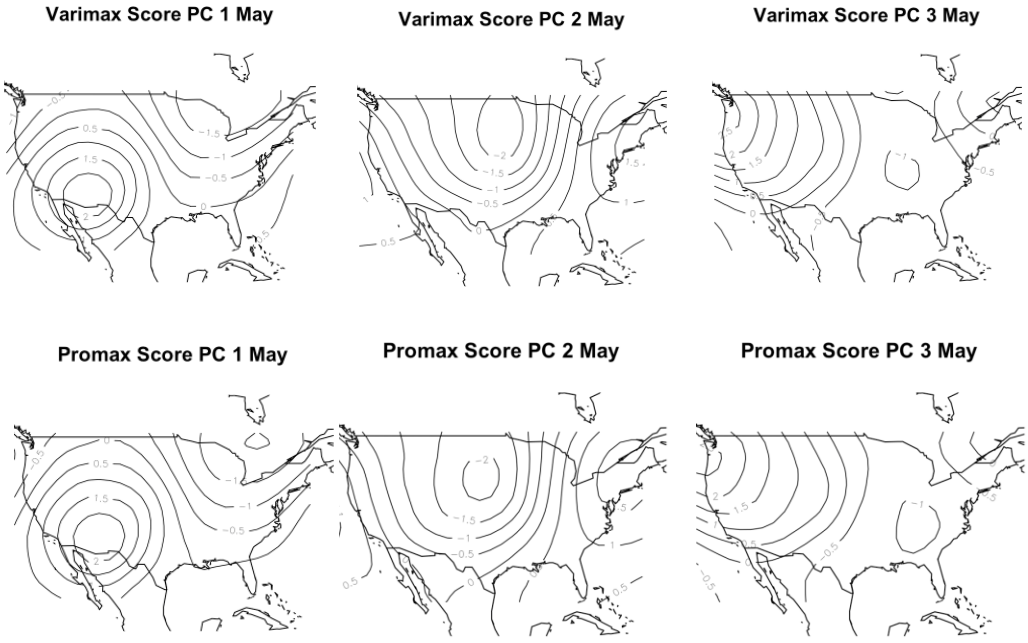

Fig. B.2.4. The Varimax and Promax transformation for three loadings in the “CONUS domain.”

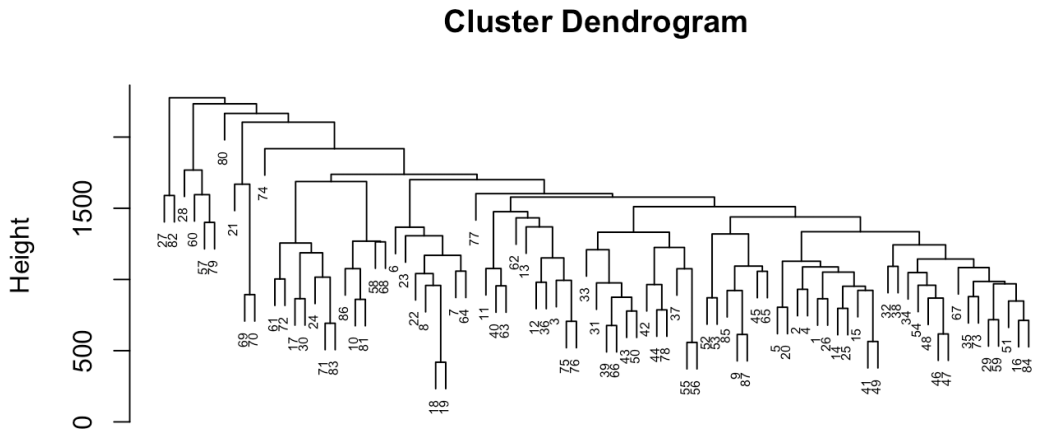

Fig. B.2.5. Dendrogram from Hierarchical clustering for the "CONUS domain."

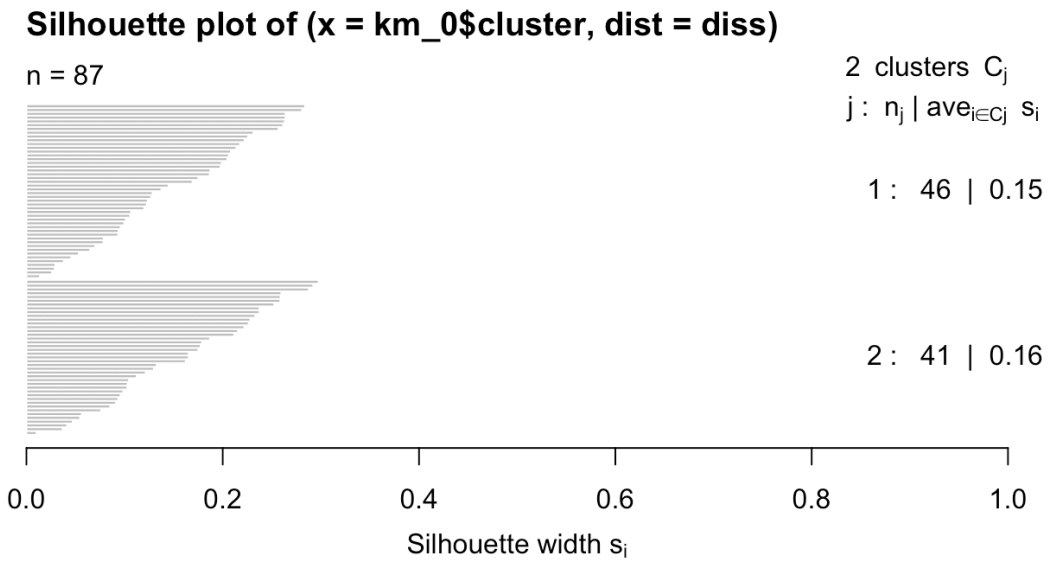

Fig. B.2.6. Silhouette clustering for the "CONUS domain." Clustering process didn't remove any of tornado outbreak cases.

3. Domain #3: “eastern U.S. + Rocky Mountains domain” (west =-108, east=-70, north=50, south=25)

Eigenvalues: 296194.628, 61303.225, 33378.155, 20412.775, 16058.685, 10760.443, 9126.230, 5395.720, 3718.753, 2351.944

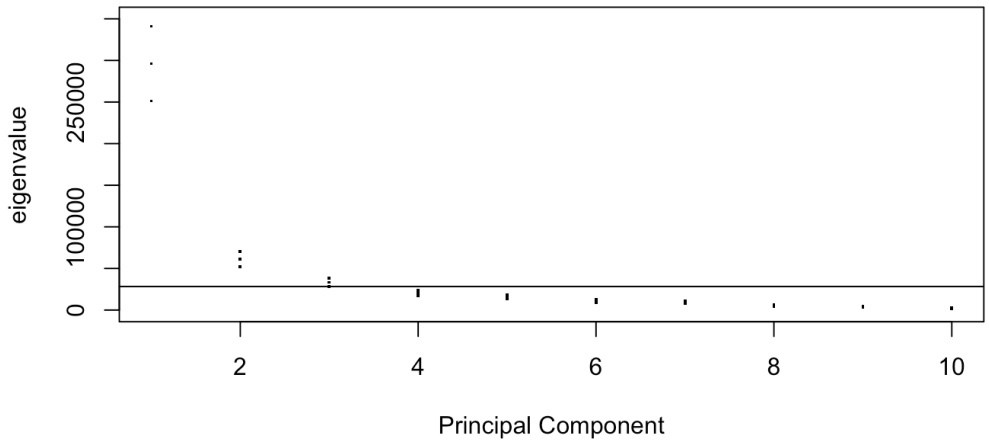

Fig. B.3.1. Plot of first 10 eigenvalues for the “eastern U.S. + Rocky Mountains domain.”

- lower limit of 3rd PC: 28317.37
- upper limit of 4th PC: 23507.75
- variance explained with first 3 eigenvalues: 83 %

Number of loadings	Varimax min	Varimax mean	Promax min	Promax mean
2	0.978	0.987	0.908	0.948
3	0.968	0.977	0.926	0.932
4	0.499	0.839	0.430	0.745
5	0.510	0.812	0.460	0.729
6	0.443	0.742	0.430	0.670
7	0.198	0.661	0.279	0.611
8	0.255	0.617	0.335	0.573
9	0.021	0.577	0.174	0.514
10	0.038	0.534	0.028	0.473

Tab. B.3.2. Absolute value of congruence coefficients for 10 loadings for the “eastern U.S. + Rocky Mountains domain.”

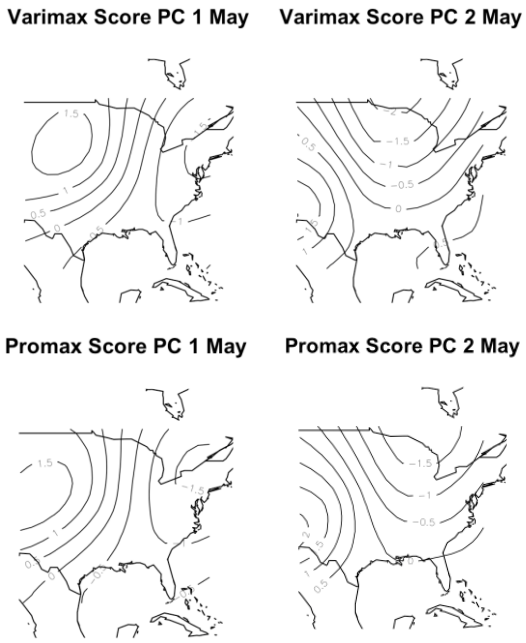

Fig. B.3.3. The Varimax and Promax transformation for two loadings in the "eastern U.S. + Rocky Mountains domain."

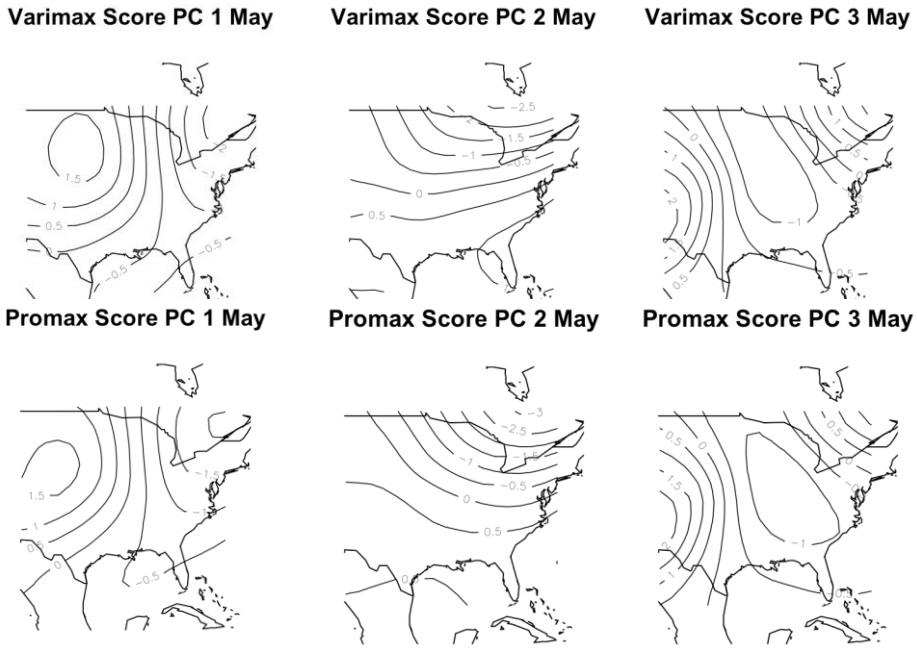

Fig. B.3.4. The Varimax and Promax transformation for three loadings in the "eastern U.S. + Rocky Mountains domain."

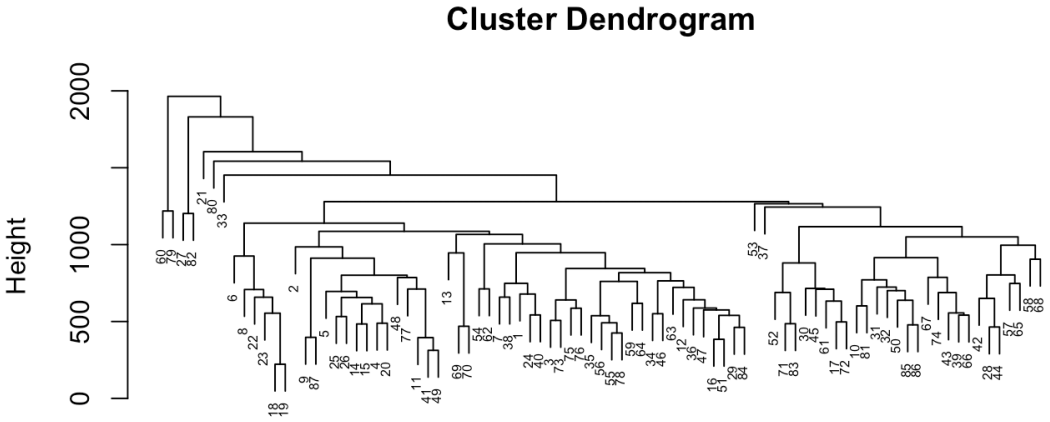

Fig. B.3.5 Dendrogram from Hierarchical clustering for the "eastern U.S. + Rocky Mountains domain."

Fig. B.3.6. Silhouette clustering for the "eastern U.S. + Rocky Mountains domain." Clustering process didn't remove any of tornado outbreak cases.

4. Domain #4: “Great Plains domain” (west =-104, east=-82, north=46, south=28)

Eigenvalues: 256469.3606, 65726.8735, 19631.2635, 12140.2611, 8721.3181, 3465.4789, 2609.0898, 1837.9418, 1596.3105, 915.8781

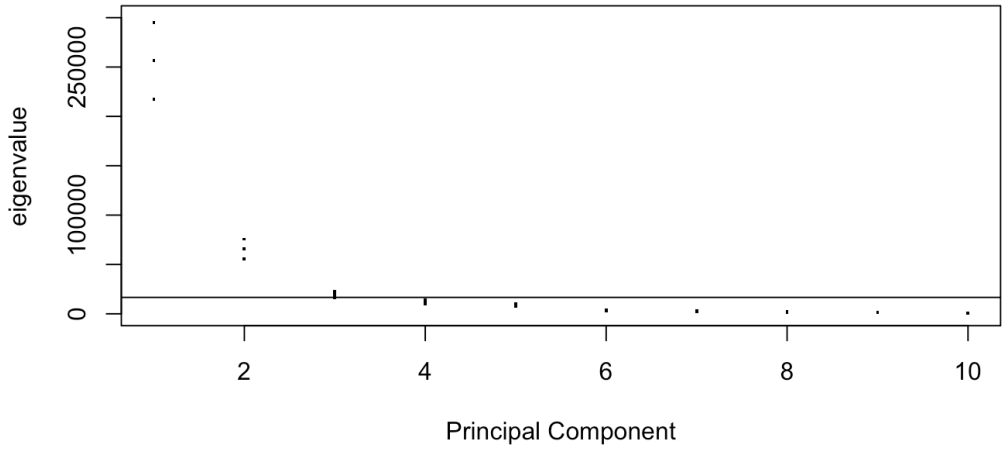

Fig. B.4.1. Plot of first 10 eigenvalues for the “Great Plains domain.”

- lower limit of 3rd PC: 16654.78
- upper limit of 4th PC: 13980.96
- variance explained with first 3 eigenvalues: 91 %

Number of loadings	Varimax min	Varimax mean	Promax min	Promax mean
2	0.994	0.995	0.947	0.952
3	0.815	0.876	0.729	0.832
4	0.384	0.793	0.491	0.743
5	0.422	0.751	0.511	0.748
6	0.309	0.657	0.446	0.683
7	0.067	0.583	0.314	0.627
8	0.074	0.534	0.109	0.569
9	0.041	0.481	0.056	0.516
10	0.104	0.495	0.299	0.560

Tab. B.4.2. Absolute value of congruence coefficients for 10 loadings for the “Great Plains domain.”

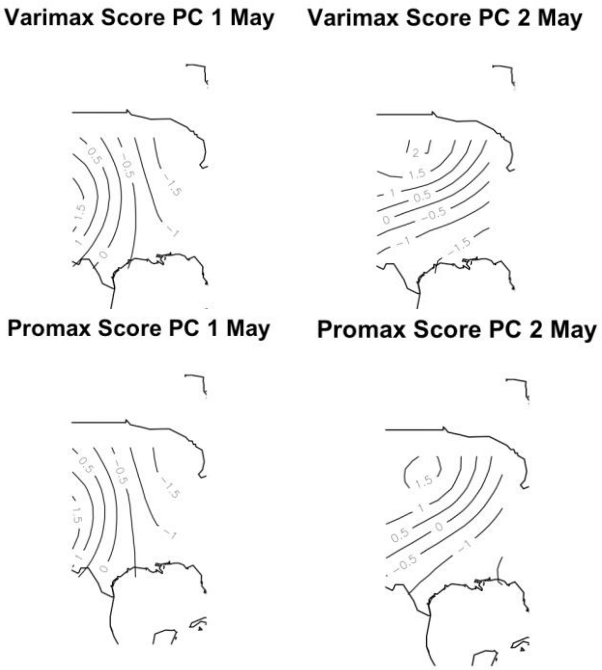

Fig. B.4.3. The Varimax and Promax transformation for two loadings in the "Great Plains domain."

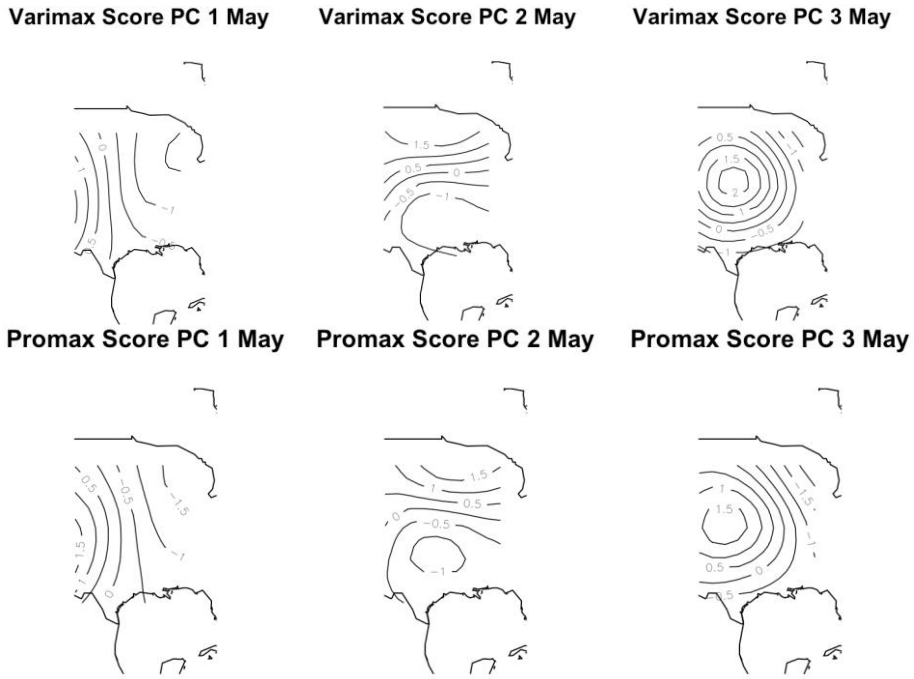

Fig. B.4.4. The Varimax and Promax transformation for three loadings in the "Great Plains domain."

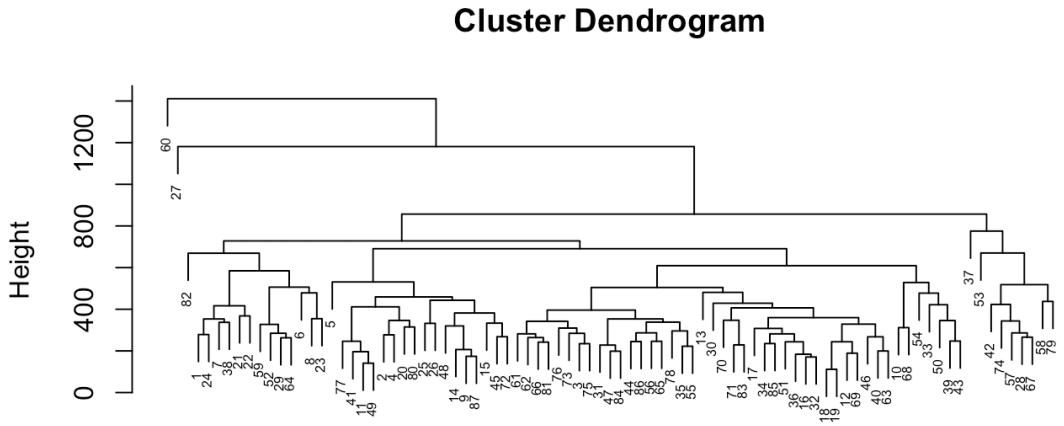

Fig. B.4.5. Dendrogram from Hierarchical clustering for the “Great Plains domain.”

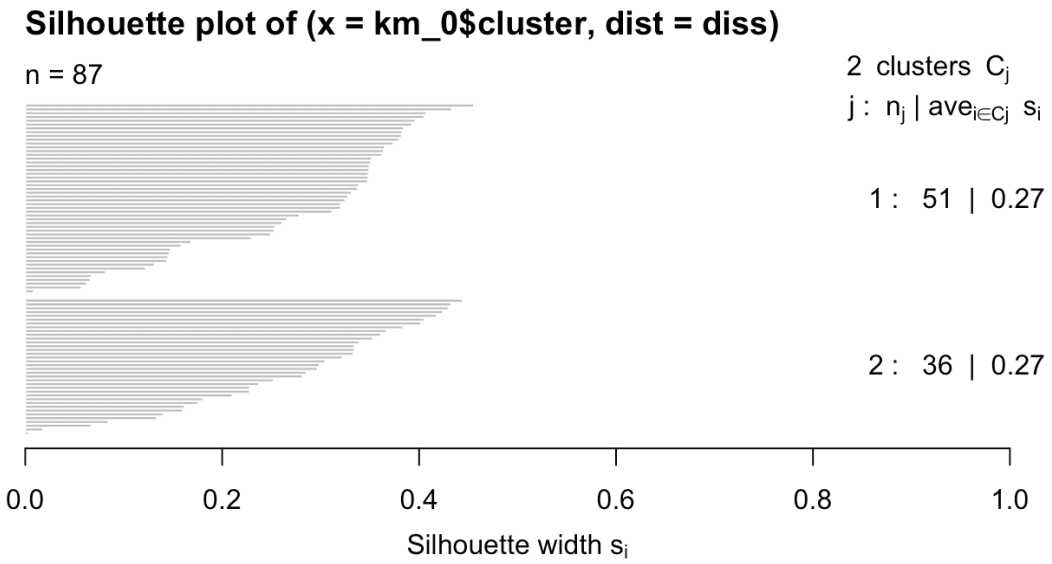

Fig. B.4.6. Silhouette clustering for the “Great Plains domain.” Clustering process didn’t remove any of tornado outbreak cases.