UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

COLORFUL LANGUAGE: INVESTIGATING THE INTERPRETATION OF THE STORM PREDICTION CENTER'S CONVECTIVE OUTLOOK BY BROADCAST METEOROLOGISTS AND THE US PUBLIC

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE IN METEOROLOGY

By

SEAN R. ERNST
Norman, Oklahoma
2020

COLORFUL LANGUAGE: INVESTIGATING THE INTERPRETATION OF THE STORM PREDICTION CENTER'S CONVECTIVE OUTLOOK BY BROADCAST METEOROLOGISTS AND THE US PUBLIC

A THESIS APPROVED FOR THE
SCHOOL OF METEOROLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Harold Brooks, Chair

Dr. Elinor Martin

Dr. Scott Salesky

Dr. Naoko Sakaeda

Dr. Joseph Ripberger

Dr. Kodi Berry

Acknowledgements

I would first like to thank my lead advisors, Dr. Joe Ripberger and Dr. Harold Brooks, for providing me guidance, opportunity, and funding to perform this research, without them I would not have been able to perform any of the research presented in this paper. I would also like to thank Dr. Kodi Berry and Holly Obermeier, who welcomed me into their CTA broadcaster study with open arms and helped me organize and record interviews with broadcasters from across the nation. As part of that effort, I would like to thank Joseph Trujillo and Emma Landeros for assisting with note-taking and moderation during the focus group and CTA interviews, both helped me keep track of all the information flying around so that I could focus on grilling the broadcasters! I'd be remiss without thanking Dr. Patrick Marsh for his assistance in outlining the development of the 5-tier SPC outlook and for encouraging my work on this project, having the ability to share my results and learn about future SPC product development was invaluable to this project. I'd like to thank Dr. Carol Silva and Dr. Hank Jenkins-Smith, as well as the rest of the students and professors at the Center for Risk and Crisis Management and the National Institute for Risk and Resilience, for being supportive of me as I have grown as a researcher, and for helping build the survey infrastructure that allowed me to collect an incredible amount of data on public SPC outlook comprehension. I am also incredibly grateful to my master's thesis committee, whose School of Meteorology members included Dr. Elinor Martin, Dr. Naoko Sakaeda, and Dr. Scott Salesky, for supporting me on this journey as well, and for challenging me to do the best research I could! I'm also grateful to have received help and support from Dr. Makenzie Krocak and Jinan Allan at the CRCM, whose expertise in statistics, numeracy, and just surviving grad school allowed me to reach the research goals I set for myself. Finally, I'd like to thank my Dad for always supporting my love of weather, my Mom for sitting with me during scary nighttime thunderstorms, my sister for keeping me in line, my Nana and Papa for taking me on "storm chases" back home in MA, and my Meme and Grampy for always encouraging me to be passionate about storms even if they can be scary!

Table of Contents

List of Tables

Table of Figures

Abstract

Though severe weather forecast products, such as the Storm Prediction Center (SPC) convective outlook, have shown to be significantly more accurate than climatology at day-to-week time scales, tornadoes and severe thunderstorms still claim dozens of lives and cause billions of dollars in damage every year. While the accuracy of this outlook has been well documented, less work has been done to explore the value of the product for non-governmental users like broadcast meteorologists and the general public. This study seeks to fill this key knowledge gap by interviewing a set of broadcasters from regions affected by severe convective weather, as well as collecting data from a representative survey of U.S. adults in the lower 48 states, about their use and interpretation of the SPC convective outlook. Data from broadcasters, collected through a combination of Cognitive Task Analyses and focus group interviews, are analyzed through thematic coding schemes, while survey data is processed through statistical tests and into visualizations. Results suggest that both broadcasters and the public take issue with the words that define each level of risk in the outlook, though overall the outlook is considered a valuable product that has a meaningful impact on users' decisions. Multiple linear regression tests also reveal that younger, White, and numerate individuals that objectively understand the difference between tornado warnings and watches are better able to interpret the SPC outlook words and colors. Overall these findings suggest that the words used in the convective outlook may confuse users as they try to derive meaning from the outlook, and that future work should use input from broadcasters and data from public surveys to develop potential replacements for the words. Using more easily understood words may help to increase the outlook's decision-making value and potentially reduce the harm caused by severe weather events.

Keywords: Severe Weather Communication, SPC Convective Outlook, Weather Broadcasters, General Public

Chapter 1 – Introduction

Motivation

The Storm Prediction Center (SPC) convective outlook is one of the oldest continuous severe weather forecasts, having existed in one form or another since 1955 (Corfidi 1999, Hitchens and Brooks 2012). The outlook forecasts the likelihood of severe weather, including tornadoes, convective wind, and hail, within 25 miles of a point across the US for the 1-8 day period, presenting those probabilities both in their natural form as well as translating them into a five-tier scale with five words and matched colors. The outlook product was originally designed for internal government use, but in the last decade has become widely referenced across social media and television (Cappucci 2020). Some of this increased visibility may originate from the increasing value individuals place on advance warning of severe weather, as the cost of severe weather disasters has soared from a 5-year average of 1.3 billion in 1984 to 15 billion dollars in 2020 (NCDC 2020). Forecasts like the convective outlook do not inherently have value, however, as Murphy (1993) describes that value is generated through the decisions forecast users make using the information contained within the forecast product. Though many studies have investigated the forecast quality and accuracy of the outlook (Hitchens and Brooks 2012, 2014, 2017; Hitchens et al. 2013), and some studies have investigated the value that emergency managers generate from the outlook (Ernst et al. 2018), there has been a lack of research into the value that non-governmental users like broadcast meteorologists and the general public are able to generate from the convective outlook. This study seeks to take the first steps towards bridging this knowledge gap by simultaneously investigating how broadcast meteorologists use the convective outlook, as well as how well the general public are able to interpret the outlook for their own use. The results of this study may further indicate ways to improve the design of the outlook to better communicate threats to the public, potentially increasing the lead time that they use to prepare for a severe weather event to a timescale of days instead of the hours or minutes offered by tornado watches and warnings.

Origins of the SPC convective outlook

The origins of the SPC convective outlook can be traced back to the efforts of Sgt. John P. Finley in the 1880s, who organized the first tornado spotter network. Finley used data from these tornado spotters and national weather observations to build a conceptual model for conditions favorable to tornado events, which he applied to issue 57 experimental tornado alerts in 1884 and 1885 (Finley 1884). Though these are considered the first ever tornado forecasts, a ban on the use of the word “tornado” prevented any of Finley’s alerts from reaching the public after 1886. For decades after, few advances into the science of tornado forecast issuance were made, until Major E. J. Fawbush and Captain R. C. Miller of Oklahoma’s Tinker Air Force Base (AFB) were tasked with developing a tornado forecast in the wake of a damaging tornado that hit the base on 20 March, 1948 (Maddox and Crisp 1999; Sandlin 2013). The pair spent the next five days furiously reviewing the existing literature on tornadoes and severe weather forecasting while the base commander developed a severe weather plan for protecting the base from tornadoes. On 25 March, a storm system with greatly more favorable conditions for severe weather than that on 20 March moved into the Southern Plains, prompting Fawbush and Miller to issue a tornado forecast for Tinker that afternoon (Maddox and Crisp 1999). A widespread severe weather outbreak did unfold that evening, with multiple tornadoes reported across Oklahoma – including one that struck Tinker AFB, where damages were significantly reduced due to the base commander activating his severe weather plan in response to Fawbush and Miller’s forecast (Maddox and Crisp 1999; Corfidi 1999). Attention to this remarkably accurate forecast led to the creation of the Severe Weather Warning Center (SWWC) at Tinker AFB under Fawbush and Miller’s leadership, which forecast tornadoes, severe winds, and extreme turbulence for all Air Force bases in the United States.

After SWWC forecasts were prohibited from being released to the public due to an event where forecasts were rumored to have been issued at “the exclusion of the public,” the Weather Bureau was pressured to form a severe weather analysis and forecasting team. This team issued their first tornado “bulletin” in March 1952, eventually settling under the name of the Severe Local Storms Warning Service (SELS) in 1953 (Corfidi 1999). The unit came under pressure for missing the deadly 1953 Worcester tornado in Massachusetts but was returned to success by a new director, D. C. House, and a move to Kansas City, Missouri in 1954. The year after, SELS renamed their daily “Severe Weather Discussions” to “Convective Outlooks,” and began regular transmission of the product (Corfidi 1999, Hitchens and Brooks 2012). Plain language public forecasts were soon initiated alongside aviation weather shorthand forecasts beginning in 1957, and tornado/severe thunderstorm forecasts were renamed "watches" in 1966, separating them from the convective outlook product. The traditional 1200 UTC convective outlook debuted in 1973 (Hitchens and Brooks 2012), and the forecasting arm of SELS was renamed the Storm Prediction Center in 1995 (Corfidi 1999).

This first version of the SPC convective outlook highlighted areas of Moderate and High Risks for severe weather, with the third category "Slight" only added after 1974 (Hitchens and Brooks 2012). Though originally only issued for the day 1 period, outlooks for day 2, day 3, and day 4-8 periods were introduced in 1986, 2000, and 2005, respectively (Edwards and Ostby 2015). The three risk level system was originally based on subjective forecaster expectations of the coverage and intensity of a forecast severe event, but in response to work by Dr. Allan Murphy on the usefulness of probabilistic forecasts, was married to probabilistic forecasts of storm coverage in the early 2000s (P. Marsh 2020, personal communication). This was achieved by Brooks and Kay (1998), who developed the Practically Perfect forecast. The Practically Perfect forecast generates an estimate of the most accurate SPC outlook that a forecaster could be expected to make given the reports of severe weather for a given period. Using this framework, the probability of a severe weather report occurring within 25 miles of a point was matched to the existing three categories in the outlook, as seen in Figures 1 and 2 (P. Marsh 2020, personal communication). For wind and hail, thresholds of 5, 15, 30, 45, and 60% probability of a report within 25 miles of a point were contoured, with 2 and 10% thresholds added for tornadoes (Imy and Edwards 2013, Grams et al. 2014). An additional hatched area could be added that suggested the chance of a significant severe weather report (defined as tornadoes of EF2 strength or greater, hail larger than 2 inches in diameter, or winds in excess of 74 mph) within 25 miles of a point was greater than 10%. Threats were not broken down for day 2 and beyond, with probabilities referencing the chance of any severe report occurring within 25 miles of a point. These new probabilistic thresholds were introduced to the outlook product in 2003 and have been issued daily since (SPC 2020).

Day 1 Probability to Categorical Outlook Conversion

(SIGNIFICANT SEVERE area needed where denoted by hatching - otherwise default to next lower category)

Outlook Probability	TORN	WIND	HAIL
2%	SEE TEXT	NOT USED	NOT USED
5%	SLGT	SEE TEXT	SEE TEXT
10%	SLGT	NOT USED	NOT USED
15%	MDT	SLGT	SLGT
30%	HIGH	SLGT	SLGT
45%	HIGH	MDT	MDT
60%	HIGH	HIGH	MDT

Figure 1: The probabilistic breakdown of the original Day 1 Convective Outlook for tornado, wind, and hail threats.

Day 2 Probability to Categorical Outlook Conversion

(SIGNIFICANT SEVERE area needed where denoted by hatching - otherwise default to next lower category)

Outlook Probability	Combined TORN, WIND, and HAIL
5%	SEE TEXT
15%	SLGT
30%	SLGT
45%	MDT
60%	HIGH

Figure 2: The probabilistic breakdown of the original Day 2 Convective Outlook, with probabilities linked to the likelihood of any severe report occurring within 25 miles of a point.

In the late 2000s and early 2010s, Hitchens and Brooks (2012) returned to the convective outlook and practically perfect forecast verification, this time to verify the accuracy of the outlook over the previous decades. Initially, they found that total outlook area had decreased over time, and outlooks were better placed with regards to storm reports, reducing the number of false alarms while maintaining consistent probability of detection (Hitchens and Brooks 2012).

The next year, Hitchens et al. (2013) identified that SPC forecaster skill had grown over time, which explained the increases in outlook precision and accuracy identified by Hitchens and Brooks (2012) after the mid-1990s. These improvements over time were also found in the more recently developed day 2 and 3 outlooks, suggesting this forecaster skill increase had led to all-around better SPC forecasts (Hitchens and Brooks 2014). More recently, Hitchens and Brooks (2017) found that missed events for significant severe weather greatly impacted estimates of forecaster skill, with significant wind events most frequently missed by outlooks. Further, Herman et al. (2018) found that SPC probabilistic outlooks are more skillful in areas with frequent severe weather events when forecast conditions involve the high-instability and high-shear conditions typically associated with severe convection. These studies combined suggest that the SPC evolved and improved the science of issuing convective outlooks with time. However, more marginal events still pose a challenge for forecasters.

Though the probabilities that defined the Moderate and High categorical outlooks have not changed since their inception, in 2014 two new categories were added to the outlook, increasing its detail at lower risk levels (Edwards and Ostby 2015). This decision was the result of pressure from emergency managers, who were unhappy with the broad definition of the Slight Risk category and wanted clarity on the difference between "normal Slight Risk and bad Slight Risk" days (P. Marsh 2020, personal communication). Emergency managers and the head of FEMA at the time, Craig Fugate, were also adamant that any changes to the outlook should not involve the Moderate or High categories, due to these terms being heavily used in weather safety plans in severe weather prone regions. As a result, the SPC sought to identify terms they could use to describe the lowest tier of probabilities currently described as "see text" (2% for tornadoes, 5% for wind and hail) and the higher probabilities in the Slight category (10% for tornadoes and 30% for wind and hail) through a customer survey administered on their website (P. Marsh 2020, personal communication). This effort, unfortunately, did not yield conclusive results, and the SPC decided to move forward with naming these two new categories "Marginal" and "Enhanced Slight." Enhanced Slight was then shortened to "Enhanced" in the final design, which was applied to the day 1, 2, and 3 outlooks in 2014 (see Figs. 3 and 4, P. Marsh 2020, personal communication).

and Marginal, tiers.

Day 1 Outlook Probability	TORN	WIND	HAIL
2%	MRGL	Not Used	Not Used
5%	SLGT	MRGL	MRGL
10%	ENH	Not Used	Not Used
10% with Significant Severe	ENH	Not Used	Not Used
15%	ENH	SLGT	SLGT
15% with Significant Severe	MDT	SLGT	SLGT
30%	MDT	ENH	ENH
30% with Significant Severe	HIGH	ENH	ENH
45%	HIGH	ENH	ENH
45% with Significant Severe	HIGH	MDT	MDT
60%	HIGH	MDT	MDT
60% with Significant Severe	HIGH	HIGH	MDT

Figure 3: The probabilistic breakdown of the current Day 1 and 2 Convective Outlooks for tornado, wind, and hail threats. Note the inclusion of the ENH and MRGL, or Enhanced and Marginal, tiers.

Day 3 Outlook Probability	Combined TOR, WIND, HAIL
5%	MRGL
15%	SLGT
15% with Significant Severe	SLGT
30%	ENH
30% with Significant Severe	ENH
45%	ENH
45% with Significant Severe	MDT

Figure 4: The probabilistic breakdown of the current Day 3 Convective Outlook with probabilities linked to the likelihood of any severe report occurring within 25 miles of a point. Note SEE TEXT and the second SLGT segments have been replaced by MRGL and ENH.

Though they moved forward operationally with the new categorical outlook design, SPC leadership anticipated controversy would emerge over the new names in the outlook (P. Marsh 2020, personal communication). In anticipation of concerns over the wording, the SPC worked to wed colors and numbers to each of the categorical levels, and developed a graphic describing the expected outcomes that each category forecast (see Fig. 5). The SPC color and word scale has since become the de-facto scale used by multiple government organizations, such as in the Weather Prediction Center's excessive rainfall outlook (WPC 2020). Since the words and colors used for the outlook were decided in 2014, no changes have been made to the design of the operational product, beyond the addition of the probabilistic breakdown between tornado, wind, and hail threats for the day 2 outlook in 2020 (P. Marsh 2020, personal communication, Grams et al. 2014).

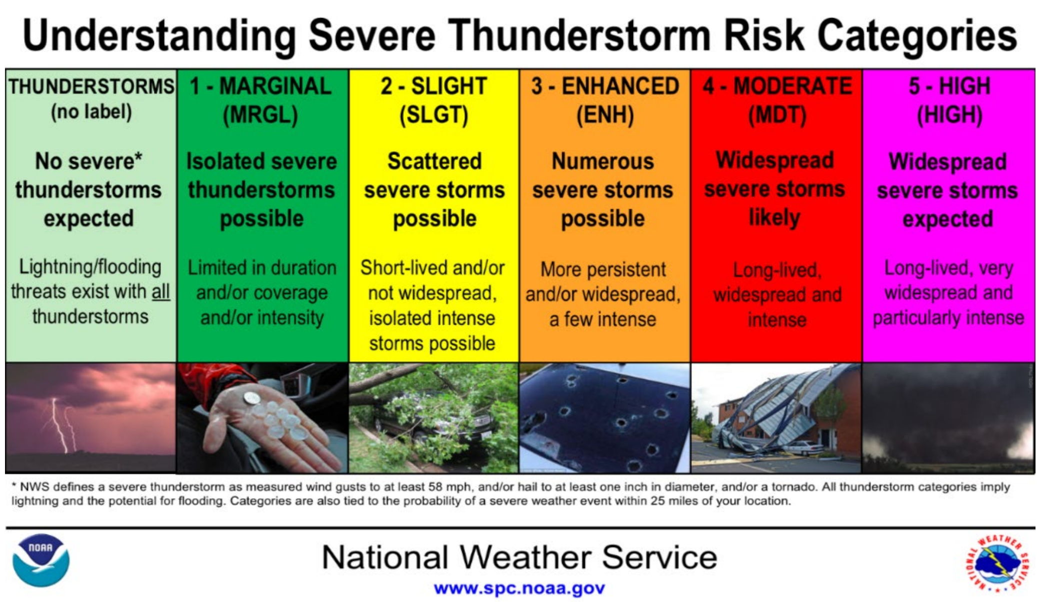


Figure 5: The SPC's online guide to the meaning of each tier of the convective outlook. Note that the image closely links the words and colors of each outlook tier, also adding numbers and coverage word descriptions to each level.

Significant changes to the convective outlook may not be far in the future, however, as the SPC is working on updating their products as part of the Forecasting A Continuum of Environmental Threats (FACETs; Rothfusz et al. 2018) initiative. At National Oceanic and

Atmospheric Administration's (NOAA) Hazardous Weather Testbed (HWT) in Norman, Oklahoma, the SPC has been testing a new product that communicates the forecast severity of wind, hail, and tornado reports in an outlook area. This new product is expected to diverge in some way from the current coverage-based convective outlook, existing as its own product that adds information to the severe weather forecast (P. Marsh 2020, personal communication). Though the forecast accuracy of this product is the focus of current HWT testing, the value of the outlook to users across the weather enterprise is still uncertain and need additional research beyond post-event service assessments (Pietrycha and Fox 2004; NOAA 2011; Stough et al. 2012; Ernst et al. 2018).

The Crucial Role of Broadcast Meteorologists

Few events in recent history reveal the influence that broadcasters can have on weather risk communication and response than the 31 May 2013 severe weather in central Oklahoma. On that day, a powerful supercell thunderstorm formed over El Reno, Oklahoma, a suburb to the west of Oklahoma City (OKC). This supercell storm produced a record-breaking 2.6 mile wide tornado that killed eight motorists, some of whom were storm chasers (NWS 2020b). Further storms then repeatedly tracked over the OKC metropolitan area and deposited over six inches of rain in a matter of hours, killing 13 people (NWS 2020b). However, the tornado dissipated before reaching the metro, a blessing considering the situation that occurred there that day.

The El Reno tornado and flash flooding event occurred 11 days after an EF-5 tornado struck Moore, Oklahoma, a suburb directly to the south of OKC that has been hit multiple times by violent tornadoes in the last two decades. Twenty-four people died in this storm, including seven children at the Plaza Towers elementary school (NWS 2020a). In their post-analysis of the event, the NWS identified that "the public felt the fear induced by the May 20 event led people to take actions on May 31 they would not normally take" (NOAA 2014). This base state of fear set in motion a series of events that led to widespread panic in the region. As the supercell and its attendant tornado moved towards OKC on May 31, during the middle of rush hour on a Friday, broadcasters in the region issued unusual advice, telling viewers and listeners that they would not be able to survive the tornado aboveground and should evacuate to the south (NOAA 2014; Ripberger et al. 2015). Thousands of Oklahomans evacuated their homes via their cars, snarling traffic across the metropolitan area while leaving evacuees stranded without shelter beneath a storm producing deadly flooding rains and several brief tornadoes (NOAA 2014; Ripberger et al. 2015; NWS 2020a). Worse still, OKC's Hispanic population is underserved by the media market in the region. Tragically a Guatemalan family who only spoke Spanish was killed in a culvert by floodwaters while trying to seek shelter from the El Reno tornado. Further, hospitals in the region noted large increases in the number of Hispanics attempting to shelter at their facilities (NOAA 2014).

Members of the public interviewed after this event often reported feeling ashamed by their actions during this storm and knew that they had made mistakes. A tornado survivor interviewed as part of NOAA's service assessment said "people knew not to get in cars… No. Shelter in place. But the fear took over. Especially if you had a weatherman you've trusted all your life tell you to get in your car and drive" (NOAA 2014). Messages from trusted sources, particularly the media, conflicted with individuals' preexisting tornado plans, and led them to change their plans in a way that ended up putting them in danger. Ripberger et al. (2015) noted that providing individuals with higher consequence language describing potential tornado impacts resulted in a decrease in sheltering action and an increase in individuals leaving their homes, suggesting that messaging by the media that individuals could die if they did not leave their homes may have decreased sheltering behaviors. Media and broadcast meteorologist messaging had a major impact on individual behaviors during this tornado event, and though this event had a negative outcome, different messaging tactics may be able to improve rates of positive response behaviors instead.

However, the El Reno tornado and flash flood event was not the only one of its kind to reveal how critical broadcasters are to the weather communication process. Hammer and Schmidlin (2002), in the aftermath of the infamous 3 May 1999 tornado in Moore, Oklahoma, found that 89% of the residents who lived near the path of the storm received warning of it from the television. Further, 47% of respondents reported that they evacuated their homes upon receiving the warning, and 35% of those who evacuated cited television reports as their reason for doing so. Hammer and Schmidlin's results mirror that of Sherman-Morris (2005), who found in a survey of the public that respondents who trusted their local broadcaster were more likely to take shelter in severe weather if their broadcaster suggested they do so. Broadcasters influence the crisis decision making of their audience to a significant degree and are often the main source members of the public receive severe weather information from. Though in the El Reno tornado case that severe weather information was in the form of tornado and flash flood warnings, the same can be said in the days ahead of a severe weather event for products like the SPC convective outlook.

Recognizing the important role that broadcast meteorologists play in the communication of weather information from NWS products, the SPC itself strongly encourages broadcasters to use their products. In an outreach document aimed at broadcasters, the organization declares that "the media are important partners helping the Storm Prediction Center, National Weather Service, and NOAA achieve their core missions" (McCarthy and Pirtle 2020). This document suggests that the SPC wants to develop a dialogue with their media partners, which is a vital part of creating shared understandings between these two groups of professionals. These shared understandings, such as what forecasts can be made for severe weather and the confidence that can be placed in those forecasts, are crucial to efficient dissemination of information before and during a weather event (Morss et al. 2015). Indeed, broadcasters today report that they appreciate NWS products and services and find themselves relying on NWS output more and more (Morrow et al. 2008: Demuth et al. 2009).

This partnership remains crucial for weather communication, as television, especially local news, is the primary source of weather information for a majority of Americans. Lazo et al. (2009) performed one of the first investigations of weather information sources for the overall U.S. public, identifying that over 70% of participants reported using local television to get forecasts at least once a day, followed by cable television stations and radio. These results were correlated with earlier studies, such as Lazo and Chestnut (2002), and are consistent over time apart from significant increases in internet forecast use with time (Lazo et al. 2009). More recent studies, such as Drobot et al. (2014), continue to see this trend, with local television the most used weather source, and internet, local radio, and cable television ranking second through fourth. Much of this change is driven by adults ages 18-29, as the percent of respondents watching local news has dropped 14% in Pew Research Center from 2006 to 2012. In 2012, only 29% of respondents age 18-29 years old claimed to watch local news regularly, with smartphone apps and internet searches the primary method they used to get weather information

(Meiners and Lusuriello 2015). Though television forecasts may reach this younger group less on the air, a plurality of Americans still turn to their favorite local television station for their daily weather information, cementing the role of broadcasters in weather communication for the near future.

With this sort of public attention, multiple studies have sought to understand how TV weather broadcasters' presentations can impact their audiences. Most of these studies focus on climate and science communication, and how broadcaster presentations of science information can elicit changed outlooks in their audience. Broadcast meteorologists are trusted to deliver climate change information by 60% of Americans, behind only climate scientists and general scientists (Leiserowitz et al. 2012). An overwhelming majority of weather broadcasters are the only on-air staff with science training and report frequently presenting science-based stories, visiting communities to speak about science, and seeking credentials from organizations like the American Meteorological Society to bolster their scientific authority (Wilson 2008). Many of these weather broadcasters have begun making presentations about climate science, which Zhao et al. (2014) found can improve audiences' climate literacy overall. Other studies suggest some nuance to this effect, however, as audience members' positive or negative feelings towards the weather broadcasters speaking about climate change were correlated with positive and negative changes in concern about climate change, respectively (Anderson et al. 2013). These audience impacts were studied with regards to climate change, however, which is perceived to be a slower and less imminent threat than severe convective storms by most, and thus the impacts of broadcaster coverage on individuals' concern may not be the same.

Study into how broadcasters impact their audiences' extreme weather decision-making has focused less on individual products, like the SPC convective outlook, and more on broadcasters' general coverage of weather events. Keul and Holzer (2013) identified that the Austrian public regards TV and radio weather coverage as important, with warnings for dangerous weather receiving the highest public interest. However, the authors noted that these warning situations were particularly impactful compared to fair weather broadcasting. Audience members showed negative emotions towards commands and reported feeling that they were overprotected from weather, even as the consequences for errors in interpretations of forecasts become more significant (Keul and Holzer 2013). Continuous coverage during intense weather events may help broadcasters present critical information to their audiences more effectively, with updates on the location and severity of storms on radar and updates on NWS watches and warnings rated as the most important services by viewers (Daniels and Loggins 2007; Drost et al. 2016). Other products, such as the ubiquitous 7 and 10-day extended forecasts, may lead to confusion among viewers as they fail to provide timing and severity information that users seek (Reed and Senkbeil 2019).

Further research generally focuses on hurricane and tornado messaging by broadcasters, and what the public gains from their coverage of these weather disasters. In the aftermath of tornado events, multiple studies have found that, even though other technologies are being developed to reach those in imminent danger, broadcast media and weather radio are still relied upon as the pillars of tornado warning reception (Pietrycha and Fox 2004; Schumacher et al. 2010). Interviews with the public and broadcasters in hurricane-prone regions generally find similar statements, with broadcast meteorology seen as a primary source of hurricane information (Demuth et al. 2012; Bostrom et al. 2016, 2019). The National Hurricane Center's (NHC) Tropical Cyclone Track Forecast Cone product, a forecast that shows the NHC's best guess at a hurricane's future track surrounded by a series of circles with their radii determined by the past five years of NHC forecast track error, is particularly prominent in these broadcaster and forecaster testimonies. Forecasters at the NHC reported feeling concerned that broadcast meteorologists focus too much on displaying the cone, versus their other wind and surge forecast products (Bostrom et al. 2016). In both tornadoes and hurricanes, emphasis is given to wall-to-wall coverage, with broadcasters seeking more updates from the NWS to share while members of the public can feel that the constant coverage is overwhelming (Schumacher et al. 2010; Demuth et al. 2012; Bostrom et al. 2018). For hurricanes continuous coverage can often start days in advance of a hurricane landfall, while for tornadoes coverage usually starts while storms are already ongoing, with both events requiring significant advance planning and staffing to maintain coverage.

Otherwise, research investigating the individual products presented by broadcasters generally seeks to understand how broadcasters communicate forecast uncertainty, with precipitation forecasts as well as the hurricane forecast track cone (also known as the cone of uncertainty) receiving particular focus (Broad et al. 2007; Morrow et al. 2008; Demuth et al.

2009). In interviews with broadcasters, Broad et al. (2007) report that many broadcasters edit the cone of uncertainty, as some noted that they remove the forecast track from the center of the cone from their on-air presentations. These broadcasters feared that the track centerline creates a perception of greater certainty in the track forecast than what the forecaster estimates there is. Broadcasters also appear to use a great deal of experimentation and leverage the advantages of the on-air communication medium to explain uncertainty. Morrow et al. (2008) and Demuth et al. (2009) identified that broadcasters used a variety of methods to communicate probability of precipitation. Excerpts from broadcaster interviews in these two studies also mention adding value to NWS forecasts in severe weather events, as they felt that describing the potential timing of severe weather is "more useful than Slight, Moderate, or High" (Demuth et al. 2009). Broadcasters did occasionally struggle with how to best communicate this information, however, as they reported feeling uncertain about whether their viewers preferred categorical descriptors or numbers as their ideal method of understanding forecast uncertainty (Morrow et al. 2008). Overall, broadcasters seek to use their television medium to communicate uncertainty to their viewers, often adapting NWS products and information to do so.

Though there is a relative dearth of research into the value, or economic and social benefit of decisions made using forecast information (Murphy 1993), generated by broadcasters using the SPC outlook, there are known limits to broadcasters' ability to present some weather information. Consultants are a driver of many station-level changes to the newscast, as they suggest changes to station management based on their market research. Henson (2010) provides a notable example, as "consultants traditionally steer weathercasters away from using the skinny lines… 'maps should not have isobars or perhaps even fronts.'" These decisions are often outside the broadcaster's control, as consultant's market research is generally proprietary and not shared with broadcasters beyond guidance from station management (Henson 2010). Indeed, many stations have recently adapted language describing "Code Red" days based on consultant guidance, with management playing a key role in deciding when a day is "Code Red," even going as far as to overwrite the broadcast meteorologist's opinion (Stelter 2019). Thus, many broadcasters' ability to present meteorological information, like the SPC outlook, is partially constrained by their station's guidelines as dictated by these outside consultants. Overall, broadcasters have a crucial role in the generation of forecast value for NWS products and can greatly impact viewer decisions, though their presentations can be constrained by time, popularity, and outside management guidelines.

Public SPC Outlook Interpretation

A significant published research gap exists with regards to the general public's interpretation of the SPC convective outlook. In recent years the SPC sought to work with its partners on how the outlook is interpreted (P. Marsh 2020, personal communication), and a suite of papers investigated the verification of the accuracy of the outlook (Hitchens and Brooks 2012, 2014, 2017; Hitchens et al. 2013; Herman et al. 2018), but these investigations did not seek to define how the outlook is interpreted by the general public. Williams et al. (2020) shifts towards this facet of forecast evaluation, identifying whether consistency across SPC, NWS, and broadcaster presentations of the convective outlook impact the general public's understanding of the graphic. However, this research sought to understand how inconsistent messaging impacts individual behavior, not how the current form of the convective outlook is understood by individuals.

Though public comprehension of the convective outlook is less well understood, a great deal of research exists into how members of the public understand of tornado warnings. Individuals process threat information, such as a tornado warning, by first receiving, then comprehending, and finally reacting to threat information (Lindell and Perry 2012). This comprehension step is crucial to enabling an effective reaction, as failed comprehension can derail the entire process. Studies of tornado warning comprehension vary greatly in their findings for participants in the U.S., ranging from as high as the 90% range (Schultz et al. 2010) to as low as 47% of participants in some regions and racial groups (Powell and O'Hair 2008, Mason and Senkbeil 2015). Additionally, significant interactions between comprehension and demographics have been identified in multiple studies, with region, race, age, gender, ethnicity, and education all being found to have varying effects on warning comprehension (Powell and O'Hair 2008; Jauernic and Van den Broeke 2017; Allan et al. 2019; Ripberger et al. 2019). White, highly educated, middle aged individuals living in the tornado-prone Great Plains region had the greatest levels of tornado warning comprehension in these studies. However, as Ripberger et al. (2020) show, notable differences exist between predicted and observed tornado warning comprehension when modeled by NWS county warning area. The researchers suggest these errors are likely due to demographic differences not captured by their model, suggesting research into further untested demographic links is crucial to understanding warning comprehension. Overall, these findings reveal that tornado warning comprehension is generally high but also linked to demographic factors, which may mean that the SPC outlook shows patterns of comprehension across demographic groups with moderate to high overall understanding.

General weather product comprehension can also be impacted by psychological states, for example the level of weather salience an individual feels. Weather salience is described by Stewart (2009) as the extent to which an individual pays attention to and is emotionally affected by the weather, manifesting in the form of greater interest in weather information. Measures of overall weather salience in the US vary, though overall it appears that salience in the US is normally distributed with a mean higher than the midpoint of commonly used salience scoring measures (Stewart et al. 2012; Williams et al. 2017). Greater salience is positively correlated with tornado watch and warning comprehension, as well as trust in forecasts made by the NWS (Stewart et al. 2012). More weather salient individuals also seek and use weather forecasts more frequently, increasing the chances that they would encounter and use the SPC convective outlook (Stewart et al. 2012). This trend has also been found in individuals' use of NWS wind products. Individuals surveyed by Williams et al. (2017) defined high wind warnings and wind advisories more often in terms of the impacts wind events could have on their daily lives when they scored higher in weather salience. Combined, these studies suggest that weather salience could be another demographic factor that explains individuals' understanding of SPC products.

A final demographic variable that may also impact individual SPC outlook comprehension is numeracy, defined by Cokely et al. (2012) as the ability to use mathematical skills to understand and reason with probabilities. Numeracy is strongly correlated with maximum education level, but individuals in highly educated professional groups were found to vary widely in their levels of numeracy (Cokely et al. 2018). Related to SPC outlook comprehension, numeracy was linked to risk judgements and general decision-making efficacy in the literature, likely due to the importance of statistical literacy to the kind of inductive reasoning tested in decision-making experiments (Cokely et al. 2012; Allan 2018). As interpreting the SPC convective outlook product involves linking words and colors to levels of risk, this suggests that numeracy may play a role in how well individuals interpret the outlook product.

Though these demographic factors will likely influence SPC outlook comprehension, other research reveals limitations in the currently used outlook design. Trujillo et al. (2020), in identifying a Spanish translation for the SPC outlook words, suggested that the relevance of the Spanish words to the threat posed by each level is more useful to Spanish-language users than direct translations of the English version of the outlook. This is due to the cultural differences between English and Spanish uses of translated words, as well as the multitude of different Spanish dialects spoken in the US, though the method of focusing on cultural meaning and implied threat of chosen words could be applied in English as well (Trujillo et al. 2020). Pennesi (2007) comes to a similar conclusion, finding that the communication of a Brazilian climate forecast is hampered by language that is not culturally relevant to subsistence farmers in rural areas. Both studies, though not in the English language, suggest that working with linguistic experts and the population that the forecast product is intended for is crucial to improving user understanding of forecast products.

Further studies investigate the use of color in forecast products, though the findings of this research are generally less consistent than that of the linguistics investigations. Lipkus and Hollands (1999) suggested that risk information presented as numbers alone is more difficult for individuals to process, while presenting that information in graphical form can hold a viewer's attention better and improve their ability to process information versus using numbers alone. However, individual interpretations can vary greatly across products and presentations. While rainbow-colored storm surge forecasts were preferred by members of the public over monochromatic graphics, rainbow-colored radar precipitation graphics are correctly interpreted less often than monochromatic graphics (Morrow et al. 2014; Bryant et al. 2014). Respondents in Bryant et al. (2014) also suggested that the rainbow color scale had too many colors, while they also wanted more colors in the monochromatic scales to enunciate areas of higher precipitation. Ash et al. (2014) also found that monochromatic tornado warning plumes generally performed better than rainbow spectral plumes, with respondents more accurately identifying the area of greatest risk and the direction of travel of a tornado, as well as a more definite warning edge, than the spectral warnings. Combined, these studies seem to suggest that the spectral scale used in the SPC outlook may not map well onto individual's perceptions of risk. It is possible, however, that the lower number of color tiers in the outlook product, similar to the storm surge graphic in Morrow et al. (2014) and the complaints about over-complicated rainbow output in Bryant et al. (2014), will actually improve an individual's ability to discern the risk level communicated by each color.

The Current Study

The goal of this work is to better understand the value (as defined by Murphy 1993) generated by broadcast meteorologists using the SPC convective outlook, as well as to link broadcaster concerns and observations to data collected on the public's ability to interpret the outlook product. It is known that understanding how broadcasters present the outlook to their viewers is important, because of the wealth of studies that suggest broadcasters are the primary conduit for weather information for the public (Pietrycha and Fox 2004; Schumacher et al. 2010; Drobot et al. 2014). As part of this effort to inform the public, I expect that broadcasters will describe using the convective outlook in their shows as well as to initiate discussions about upcoming severe weather potential with their station management. I can also hypothesize based on Reed and Senkbeil (2019) that the lack of timing and severity data provided by the coverage forecast of the SPC outlook frustrates broadcasters as they try to present a threat to their viewers. Finally, evidence from Broad et al. (2007), Morrow et al. (2008), and Demuth et al. (2009) suggests that broadcasters will mention reinterpreting the SPC outlook for their public similarly to how they change the hurricane cone of uncertainty before sharing it with viewers.

As for the general public, based on the studies of tornado warning comprehension (Jauernic and Van den Broeke 2017; Allan et al. 2019; Ripberger et al. 2019), I hypothesize that white, educated, middle-aged men living in tornado-prone regions will best comprehend the SPC outlook. Data from studies of weather salience and numeracy also suggest that more salient and more numerate individuals should interpret the convective outlook product with greater ease (Allan 2018, Cokely et al. 2012, Stuart et al. 2012). Additionally, as the public gets most of their weather information from broadcasters, I hypothesize that broadcaster concerns about the SPC outlook and how the public interprets it will be reflected in the data on how the public interprets the outlook words and colors. Thus, the major hypotheses of this study are as follows:

- **H1:** Broadcasters use the SPC Outlook as a visual aid to describe severe weather threats in their shows, as well as to drive conversations with station management to prepare for severe weather television coverage.
- **H2**: Broadcasters will voice frustrations with the lack of forecast detail contained within the convective outlook, with additional information on timing and intensity most needed to improve the outlook.
- **H3:** Broadcasters will mention adding information or reinterpreting the convective outlook before presenting it to their viewers based on their concerns about what information the outlook lacks.
- **H4:** White, educated, middle-aged men living in tornado-prone regions will best comprehend the SPC outlook when asked to rank the outlook colors and words from least to greatest risk.
- **H5:** More numerate and weather salient individuals will best comprehend the SPC outlook when asked to perform the outlook ranking task.
- **H6:** Broadcaster concerns about public interpretation of the SPC outlook words and colors will be reflected in the analysis of the public survey data.

Chapter 2 – Data and Methods

Interview Data Collection

To collect data for this investigation, I used three separate data collection techniques, each with a different goal in mind. First, a set of 15 Cognitive Task Analyses (CTAs) were collected from a convenience sample of broadcast meteorologists that participated in the HWT experiment in Norman, Oklahoma, from 2017 to 2019 (see Figure 6 for participant locations of origin). As this was a convenience sample where broadcasters interviewed were those that volunteered to participate, this sample is not necessarily representative of all broadcasters. CTAs interviews like these are commonly used to explore the process that weather forecasters use to develop forecasts and warnings (Hahn et al. 2003; Kirschenbaum 2003), as they allow researchers to generate new understandings by studying the work that participants perform and the insights they use to perform that work. Some studies suggest that CTA success is measured by whether the analysis reveals completely unexpected elements (Militello et al. 2011). These analyses are especially useful for identifying new knowledge from complex decision-making chains, as studies of forecaster methodology and ambulance dispatchers have shown (Hahn et al. 2003, Militello et al. 2011).

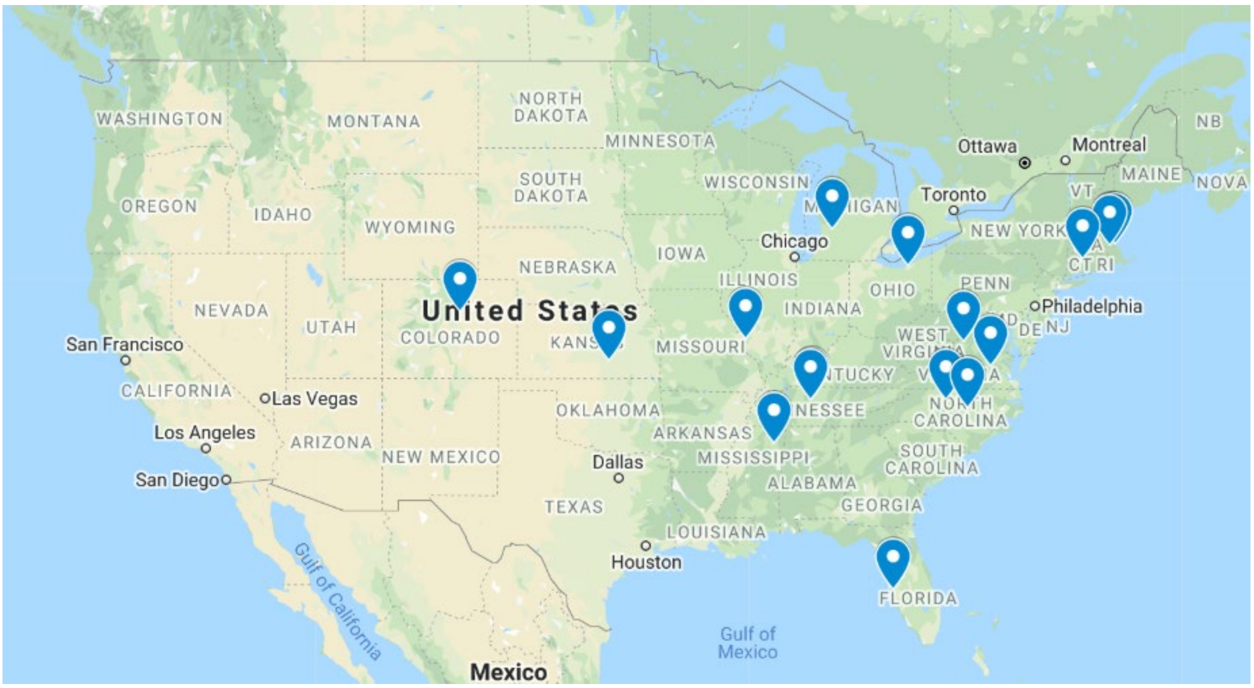


Figure 6: Map detailing where broadcast meteorologists who participated in the study work on air. Note three participants came from the Boston market, while two were from Nashville.

Due to their usefulness in studying work habits, the CTAs were used to identify how broadcasters change their activities based on SPC convective outlook information. This allows us to evaluate the value of the forecast, as value is generated by forecast user decisions (Murphy 1993). For this study a form of workspace analysis, called Activities Observations, was adapted for use from Hoffman (2005). This type of CTA generally asks researchers to observe participants at work, but due to constraints on time and travel, the activity observation format was developed into an interview and a questionnaire for participants. To complete the data collection, participants were first emailed a Word document containing the questionnaire (Appendix 1), asking participants to fill in a timetable for their daily routine given a day with No

Risk for severe weather, a day with a Slight Risk of severe, and a day with a Moderate or High Risk of severe weather. When participants arrived in Oklahoma for the HWT experiment, they were interviewed about their answers to the online survey, with interviewers seeking to add depth to their initial answers. Though this method differs from that in the template given by Hoffman (2005), one of the strengths of CTA is its ability to be modified to fit researcher needs, such as the limitations faced by this study (Militello et al. 2011).

The second data collection method consisted of four focus group interviews with nine broadcast meteorologists, conducted in June and October of 2019. The first of these focus groups was conducted with a separate convenience sample of three broadcasters at the AMS 47th Conference on Broadcast Meteorology in San Diego, California, while the other three interviews were conducted with broadcasters who were participating in the October HWT experiment and had also completed CTA interviews. These focus groups were based on guidelines from Breen (2006), Krueger (2002), and the Center for the Promotion of Health in the New England Workspace (CPH-NEW 2020). Focus group interviews seek to generate ideas about a topic, with a moderator guiding a conversation centered on questions related to the research topic (Breen 2006). Participants should first be permitted to give their consent to participate, as well as ground rules focused on creating a safe and participatory environment for the participants (Breen 2006, CPH-NEW 2020). For the interview itself, Krueger (2002) recommends an outline that moves from a broad opening question, to an introductory question that elicits recall of the topic of interest, and then to a transition question that pilots the research question. This is followed by the key questions, which effectively ask participants to discuss the research question, and finally the interview is concluded with a wrap-up question that asks participants for any final thoughts. This wrap-up can also be accompanied by the moderator's summary of the points made in the interview, which participants can then offer their thoughts on (CPH-NEW 2020).

Based on these guidelines, I developed an interview outline that was used for the four focus groups. The interview outline (Appendix 2) began with several pages of introductory information, welcoming the participants to the study, briefly explaining the motivation of the study, and setting ground rules and assuring participant confidentiality. The introductory information was followed by the discussion during which the moderator asked questions listed in the guide and ensured the conversation remained on topic. Next, an opening question asked participants to speak about where they were from and how they became interested in weather, to make participants comfortable with each other and the moderator. Participants were then asked what kinds of SPC products they currently used, whether convective outlooks were useful to them, and whether they liked the way the outlook presented information. This led to the transitional question, which asked participants to recount the most significant severe weather event they had covered, and what they learned about forecasting and broadcasting from that event. As I sought to identify the value of the outlook product as described by Murphy (1993), my key questions asked participants to describe how they would change their messaging and daily routine for three convective outlook sequences of decreasing risk level. This question was designed to highlight differences in the actions participants would take when shown each different outlook, to reveal how the outlook information allowed them to generate value through informed decisions. The interviews were then concluded with a moderator summary and any final thoughts from participants about the interview. Interview audio was recorded and transcribed for analysis.

Survey Data Collection

Individual data were collected through the 2019 Severe Weather and Society Survey (WX19), an annual survey of contiguous United States adults over age 18. This survey was developed at the University of Oklahoma by the Center for Risk and Crisis Management (CRCM) and administered to a demographically representative sample by Survey Sampling International (SSI). Participants were contacted through email and dynamically sampled to generate a representative sample based on U.S. Census data (Table 1, Silva et al. 2019). WX19 sampled 3006 adults, of which 51.3% were male and 48.7% were female.

The WX19 survey consisted of three major types of questions. The first type were demographic questions that polled details including participant age, race, and gender. The second were a set of annually repeated questions designed to capture data on national tornado warning comprehension or weather information sources, which can then be compared year-to-year to identify national trends. Finally, there was a series of single-issue questions, which sought answers to questions deemed important at survey issuance by the CRCM team. In the

WX19 survey, a pair of single-issue questions on SPC outlook interpretation asked participants to order the words and the colors used in the SPC convective outlook from least to greatest risk (Table 2). These questions sought to answer how well individuals interpret the SPC outlook, as well as identify which words or colors were the most problematic in individual's interpretations.

Table 1: The demographic breakdown of the 2019 WXSurvey as compared to the most recent US Census. This survey is a good representation of the overall breakdown of the US population.

	U.S. Adult Population* (%)	Participants (%)
Gender		
Female	51.3	51.3
Male	48.7	48.7
Age		
18 to 24	12.0	12.0
25 to 34	18.0	18.2
35 to 44	16.3	16.3
45 to 54	16.4	16.3
55 to 64	16.7	16.7
65 and up	20.6	20.5
Ethnicity		
Hispanic	16.3	16.4
Non-Hispanic	83.7	83.6
Race		
White	77.9	77.9
Black or African American	13.0	12.8
Asian	5.9	5.9
Other Race	3.2	3.4
NWS Region		
Eastern	31.6	32.0
Southern	27.1	26.5
Central	20.7	20.9
Western	20.6	20.6

Data was also collected from a series of questions that measured each participant's ability to interpret forecast products like the SPC convective outlook (Table 2). First, numeracy is a measure of participants probabilistic and statistical understanding, which was estimated using the Berlin Numeracy Test (BNT, Cokely et al. 2012). This test first uses four questions to measure a respondent's statistical numeracy skills. The version I used for this study, the BNT-S, included additional items adapted from Schwartz et al. (1997) that increased the sensitivity of the measure for less skilled or educated individuals. This version of the BNT was also used in previous studies of extreme weather decision-making, such as Allan et al. (2017).

Table 2: Questions used to collect data on the demographic variables compared in this study.

Question Group	*Question Wording*
SPC Word Ranking	The National Weather Service Storm Prediction Center uses the following phrases to describe the risk of severe thunderstorms and tornadoes. We want to know what these phrases mean to you. Can you rank them from one (lowest risk) to five (highest risk)? (*Words are Marginal, Slight, Enhanced, Moderate, and High, randomly assigned across the five ranks.*)
SPC Color Ranking	The Storm Prediction Center also uses colors to describe the risk of severe thunderstorms and tornadoes. We want to know what these colors mean to you. Can you rank these colors from one (lowest risk) to five (highest risk)? (*Colors are Green, Yellow, Orange, Red, and Magenta, randomly assigned across the five ranks.*)
Age	How old are you? (*numeric response*)
Gender	Are you male or female? (*multiple choice*)
Ethnicity	Do you consider yourself to be Hispanic, Latino, or Spanish or to have Hispanic, Latino, or Spanish origins? (*Yes or No*)
Race	Which of the following best describes your race? (*Choice of White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Pacific Islander, Two or more races, or Some other race*)
State	Please select the state or district where your primary residence is located (*drop down list*)
Education	What is the highest level of education you have COMPLETED? (*Choose from Less than high school, High school/GED, Vocational or Technical Training, Some College; NO degree, 2-year College/Associate's degree, Bachelor's Degree, Master's Degree, or PhD/JD (law)*)
Salience	I follow the weather very closely. (*5-point Likert scale, Strongly Disagree to Strongly Agree*)
	I plan my daily routine around the weather. (*5-point Likert scale, Strongly Disagree to Strongly Agree*)
	I don't understand what causes extreme weather events like thunderstorms, tornadoes, and hurricanes. (*5-point Likert scale, Strongly Disagree to Strongly Agree*)
Subjective Tornado Watch/Warning Comprehension	In general, do you understand the difference between watches and warnings? (*5-point Likert scale, Definitely No to Definitely Yes*)
Objective Tornado Watch/Warning Comprehension	This alert is issued when severe thunderstorms and tornadoes are possible in and near the area. It does not mean that they will occur. It only means they are possible. (*Select Tornado WATCH (correct), Tornado WARNING, or Don't know*)
	This alert is used when a tornado is imminent. When this alert is issued, seek safe shelter immediately. (*Select Tornado WATCH, Tornado WARNING (correct), or Don't know*)
Tornado Responsiveness	Sometimes I ignore tornado warnings that are issued for my area. (*5-point Likert scale, Strongly Disagree to Strongly Agree*)
	I always take protective action when tornado warnings are issued for my area. (*5-point Likert scale, Strongly Disagree to Strongly Agree*)
	Sometimes I am too busy to take protective action when tornado warnings are issued for my area. (*5-point Likert scale, Strongly Disagree to Strongly Agree*)

	I am not sure what to do when tornado warnings are issued for my area. (*5-point Likert scale, Strongly Disagree to Strongly Agree*)
Disaster Preparedness	In the last year, have you prepared a Disaster Supply Kit with emergency supplies like water, food and medicine that is kept in a designated place in your home? (*Choose No, Yes, Not sure*)
	In the last year, have you prepared a small kit with emergency supplies that you keep at home, in your car or where you work to take with you if you had to leave quickly? (*Choose No, Yes, Not sure*)
	In the last year, have you made a specific plan for how you and your family would communicate in an emergency situation if you were separated? (*Choose No, Yes, Not sure*)

Weather salience was determined using three measures, which asked participants on a 5-point Likert scale to recount whether they followed the weather closely, planned their day around the weather, and did not understand why weather events occur. Participant answers to the understanding question were reversed to match the orientation of the other two questions, and the three answers were averaged to generate a participant salience score.

Participants' tornado responsiveness and disaster preparedness were similarly estimated by combining answers to multiple questions. For tornado responsiveness participants were asked on a 5-point Likert scale, if they sometimes ignore, always take protective action during, are sometimes too busy to act during, or are unsure what to do during tornado warnings. Answers for the last three questions were inverted to match the protective action question's orientation and then averaged to produce a tornado responsiveness score. Disaster preparedness was measured using three questions, which asked whether participants kept a disaster supply kit, a small emergency kit, and had a disaster communication plan for their family. Participant answers of "yes" were given a value of one, and the sum of each participant's three answers was recorded as their preparedness score.

Finally, subjective and objective measures of participants' understanding of tornado watches and warnings were collected in the survey. For the subjective measure, participants were asked whether they understood the difference between a tornado watch and a tornado warning, on a 5-point Likert scale, from definitely no to definitely yes. They were then tested with one of two questions, which contained a prompt describing either a tornado warning or a tornado watch. Participants who correctly chose what the prompt described were then given a dummy variable value of one.

Thematic Analysis

After collecting the interview data, a thematic analysis was conducted on the focus group and CTA datasets to analyze the participants' answers to the interview questions. Defined by Braun and Clarke (2006) as "a method for identifying, analyzing, and reporting patterns (themes) in data," thematic analysis is a commonplace, though sometimes poorly defined, qualitative analysis technique (Mohammed 2012; Nowell et al. 2017). Thematic analysis was formally defined by Boyatzis (1998), though it has been adapted and defined by a multitude of researchers since then (Nowell et al. 2017). This method of analysis is frequently used with qualitative data sets, due to the flexibility, ease of use and training, and ability to identify connections between datasets collected in different spaces and times it offers (Mohammed 2012; Nowell et al. 2017). However, unless data analysis is properly defined, recorded, and reported by researchers, thematic analysis can struggle with trustworthiness, replicability, and academic rigor (Fereday and Muir-Cochrane 2006; Braun and Clarke 2006; Nowell et al. 2017). Careful attention to the planning and process of thematic analysis is thus crucial to producing generalizable findings.

Prior to data collection, Braun and Clarke (2006) suggest researchers address a series of five decisions about their planned data collection and analysis process. To perform a replicable, trustworthy analysis, they recommend that researchers define 1) what counts as a theme; 2) whether an overall description or focused investigation of the data is sought; 3) whether analysis is inductive or deductive; 4) whether the themes sought are semantic or latent; 5) and what epistemology the investigation will adhere to. To determine what ideas qualify as a theme researchers must answer the following questions: Must ideas occur a set number of times across the dataset to qualify? Does their relation to the key research questions matter more? Defining what counts as a theme is crucial to replicability, as it defines the unit of measurement for the analysis (Braun and Clarke 2006). Researchers must then choose between a rich description of the breadth of themes identified in the analysis, or a deep dive into a theme or series of themes that are defined by a predetermined research question (Braun and Clarke 2006). The third decision is whether to use a top-down deductive coding scheme, where the analyst's theories and research questions guide data analysis, or a bottom-up inductive analysis, where codes naturally emerge from the data and may not relate to the initial research question (Braun and Clarke 2006; Fereday and Muir-Cochrane 2006). The researcher must then determine the type of theme sought by analysis, which can range from semantic, first-level themes that describe actions and participant perspectives to latent second-layer themes that suggest underlying motivations and guiding viewpoints. Finally, researchers must decide on an epistemology – to analyze their data from an essentialist/realist perspective, where a clear path is assumed to exist between meaning and experience and language, or from a constructionist perspective where the latent influences of social context define meaning and experience (Braun and Clarke 2006). It is important to note that these decisions, while crucial, are not binary, as Fereday and Muir-Cochrane (2006) describe their use of a mixed deductive/inductive coding scheme for data analysis. These decisions instead act to frame the process of data analysis, and help the reader understand the thought process and goals of the analyst.

With these five decisions in mind, I defined a theme for this study as an idea the participants discuss that relates to the key research question, which is "how do convective outlooks change broadcasters' work behaviors and weather coverage?" Given this more precise research question, I decided to take a deep dive into this subset of themes in the data using a deductive coding scheme while incorporating inductively developed themes that related to the research question. Finally, the coding strategy focused on identifying semantic, or first-level, themes from an essentialist perspective, meaning I assumed that what broadcasters said in their interviews directly reflected their experiences and opinions on the outlook. This final definition of constitutes a theme for this study helps to determine which method of thematic analysis to use, which can develop findings from the transcribed interview data.

While the five decisions suggested by Braun and Clarke (2006) define the scope and goals of thematic analyses for a given study, the analysis process must also be rigorous enough for conclusions to be drawn. To define coding rigor, Fereday and Muir-Cochrane (2006) discuss using Schutz's three postulates, which are logical consistency, subjective interpretation, and adequacy (Schutz 1958). For thematic analysis to be logically consistent, it should have a defined method for developing themes that is documented throughout the analysis process. Subjective interpretation requires analysis to preserve the participant's point of view, with excerpts from raw data used to support themes. Finally, adequacy requires that the participants espouse the findings of a thematic analysis, and can be checked through summarizing participant thoughts at the end of an interview or studying the way that the findings of a thematic analysis are applied within the field to which they apply (Fereday and Muir-Cochrane 2006).

More recently, Nowell et al. (2017) sought a more definitive set of guiding principles to ensure rigor in thematic analysis, developing a set of six criteria that should be established to ensure trustworthiness of the analysis. First, credibility is described by Nowell et al. (2017) as "the fit between respondents views and the researcher's representation of them," and is maintained through debriefing or checking with participants about analysis results. The second criterion, transferability, refers to whether findings are generalizable to the larger world and can be transferred from the analysis into practice. Third, dependability asks if "the research process [is] logical, traceable, and clearly documented," similar to the Schutz's postulate of logical consistency (Schutz 1958; Nowell et al. 2017). These first three criteria fold into the fourth criterion, confirmability, which is met when a researcher can demonstrate that the findings of their research were derived directly from the raw data. The fifth criterion regards the process of creating an audit trail, as Nowell et al. (2017) recommend researchers carefully document their path from data to findings to the point where a third-party researcher with a shared perspective and background could reach similar findings. Finally, researchers should maintain reflexivity in auditing by maintaining a research journal that documents daily logistics, methodological decisions and changes, and personal reflections on insights from the data, to ensure transparency in the research process. By meeting all six criteria, Nowell et al. (2017) suggest that thematic analysis can establish trustworthiness, ensuring academic rigor is applied to the research method.

With all these considerations in mind, there are also multiple theories on what the general blueprint for the thematic analysis process should look like. Mohammed (2012) suggests working from data collection to data reduction, data display, and then conclusions, while others espouse a more detailed process (Braun and Clarke 2006; Fereday and Muir-Cochrane 2006; Nowell et al. 2017). Nowell et al. (2017) and Braun and Clarke (2006) suggest a six-step process that reaches from initial familiarization with the data to reporting findings. The first step of the process is to become familiar with the data by reading and re-reading all items in the dataset, while documenting ideas for codes and other thoughts. The transcription process, as well as archiving raw data in a de-identified format, is crucial to this step. Second, researchers should develop their initial codes, which are defined by Boyatzis (1998) as "the most basic segment, or element, of the raw data or information that can be assessed in a meaningful way regarding the phenomenon". This process organizes the data into meaningful groups, exposing patterns in the data that can be extracted and analyzed in context with other excerpts. Codes used in analysis can be generated deductively or inductively and should be recorded using a codebook and reflexive journaling (Nowell et al. 2017).

The excerpts generated by codes are then used to find themes, by searching for links between codes. As in Mohammed (2012), Nowell et al. (2017) and Braun and Clarke (2006) suggest the use of visualizations such as tables or web diagrams to help organize these connections. These themes are reviewed by re-reading the excerpts that support them, to identify whether they form a coherent pattern. Themes that do not meet this standard should be reworked, which narrows down the number of themes and builds a stronger thematic map that describes the data. For the final phase of the review step, this map should be compared to the overall dataset, to ensure it is an accurate representation of the participants' statements. The finished thematic map is improved by defining and naming themes, which requires the researcher to write detailed analyses for each theme while considering how the excerpts define the theme and why this theme is interesting. Themes should also be renamed to match the words of the participants, to ensure a connection exists to the original data (Nowell et al. 2017). The final step of thematic analysis is to report the results, using quotations and the coded extracts to argue why the resulting themes are supported by the data and related to the research questions. As part of the report, analyses completed as part of creating the thematic map and describing each theme should be included in at least an appendix in the report (Nowell et al. 2017; Braun and Clarke 2006). Following this six-step process, while considering the criteria for trustworthiness and rigor, can ensure that the researcher develops findings that are well-supported by the dataset, and worthy of addition to the scientific discourse.

To ensure rigor in my analysis, I utilized the six-steps defined by Nowell et al. (2017) in my thematic analysis process. I first transcribed the audio from the focus group interviews, then read both the focus group transcriptions and the CTA interview data to be immersed in the data. These data were anonymized, and participant names were removed and replaced with an identifying number consistent across both the CTA and focus group interviews (e.g. MET7, MET12). Though most broadcasters participated in either the CTA or the focus group interviews, there were five broadcasters that participated in both. To keep respondent identifiers consistent, labels were identical across the two groups, meaning that MET1-9 were interviewed in the focus groups and MET4-18 were interviewed in the CTA interviews. These identifiers are used throughout the remainder of the paper to notate which broadcaster quotes are sourced from. Next, a list of initial impressions of the data and a list of codes were generated to organize the data. Reflexive journaling, where the coder listed their impressions and defined new codes, was also performed during the coding process. Once the data were coded, excerpts related to codes from the data were extracted and compiled into a series of Word documents organized by theme. From these documents, I developed themes by linking ideas in the excerpts, placing these themes on a visual web and linking them together. Themes in this thematic map were then reviewed and renamed using participants' words to summarize each theme's contents. The entire thematic map was finally reviewed before results were drawn from the data. Separate thematic maps were developed from both the CTA data and the focus groups, due to their different data formats.

Statistical Inference

To interrogate the survey data, I used the statistical programming language *R* to generate useful output statistics and graphics. The first step of the analysis was to characterize the data by displaying, in order, how many respondents chose each of the five risk words for the five possible risk levels. Next, the sizes of each ordering group (e.g., how many respondents ranked the risk words as "Marginal, Slight, Enhanced, Moderate, High") were compared. The results of these grouping efforts were developed into a series of visualizations, to aid in interpretation. This process was then repeated for the color ranking data.

After this initial investigation of the data, I developed a score which could quantify participant ranking accuracy for the SPC words and colors. Scores allow statistical tests to be applied to data that otherwise would not be quantitative in format, which in this case allows for statistical analysis of the qualitative SPC word and color ranking data. I created a squared error score based on the difference between each participant's chosen ranking order and the actual ranking of the SPC words and colors. A number value was then assigned to each response based on its correct rank (Marginal = 1, etc), and each response was recoded to match these values. The following equation was then applied to the recoded data:

$$SPCScore = 20 - \frac{\sum_{i=1}^{n}(n - nthRank)^2}{2}$$

where the SPCScore is the final score value for a participant, *n* is the rank in question (from 1 to 5), and *nthRank* is the value that the participant applied to that rank level (e.g., if a participant set "slight" as the first rank word, *n = 1* and *nthRank = 2*). The resulting score is divided by two to address the way errors occur in a ranking task, as there are always two misplaced ranks for any single mistake, which results in all scores being even numbers. This value was then subtracted from 20 to invert the score. As a result, a score of 20 represented a ranking that matched the SPC format, while a score of 0 represented a ranking that was the exact opposite of the SPC format. This scoring method was chosen because of how it weights errors, as interchanging "high" with "enhanced" could result in a potentially more significant real-life consequence than interchanging "marginal" and "slight". This method also provides greater granularity than a score that sums the differences without a square.

Using these scores, I developed a pair of multiple linear regression models to explain participant SPC word and color ranking accuracy across the other demographic factors and survey measures. The equation for these models included 12 independent variables: Age, gender, ethnicity, race, education, numeracy, NWS region, weather salience, tornado responsiveness, preparedness, subjective tornado watch and warning understanding, and objective tornado watch and warning understanding. The final term in the equation is the error or residual term, which accounts for the prediction errors between the model prediction and observed values of the dependent variable.

$$
\begin{aligned}
SPCWordScore \\
&= A + B_1(Age)_i + B_2(Gender)_i + B_3(Ethnicity)_i + B_4(Race)_i \\
&+ B_5(Education)_i + B_6(Numeracy)_i + B_7(NWSRegion)_i \\
&+ B_8(WeatherSalience)_i + B_9(TornadoResponsiveness)_i \\
&+ B_{10}(Preparedness)_i + B_{11}(SubjectiveTor.Watch/Warn)_i \\
&+ B_{12}(ObjectiveTor.Watch/Warn)_i + E_i
\end{aligned}
$$

The equation for the SPC color ranking score model is identical to the one used for the word ranking scores, save for the dependent variable which was changed to the SPC color score. A combination of continuous (i.e. age, numeracy) and discrete (i.e. gender, NWS region) variables were used in this study, which is facilitated using dummy variables applied to the discrete variables. These dummy variables allowed the model to identify changes in the dependent variable across the different values of the discrete variable (Jenkins-Smith et al. 2017). For example, Central Region was used as the reference value for the NWS region variable, which was then compared individually to Eastern Region, Southern Region, and Western Region. Linear regression can be used in these cases as scores for the two different discrete values are compared, creating two points from which the slope of a regression line can be identified.

Multiple linear regression uses ordinary least squares estimation to generate a linear model of the effects of multiple independent variables on a dependent variable of interest. This method is a more complex version of ordinary least squares (OLS) estimation. The process of OLS estimation uses matched values of the dependent and independent variable across a sample to find the slope and intercept of a line that has the smallest possible value of the residual term E_i (i.e. the minimum difference between the predicted values and the measured values of the dependent variable across the range of the independent variable). Multiple linear regression is effectively OLS estimation with multiple independent variables, which is accomplished through matrix algebra (Jenkins-Smith et al. 2017). The result of this estimation process is a predictive surface that accounts for changes across multiple independent variables, versus the two-dimensional predictive line that predicts the dependent variable using only one independent variable that is produced by OLS estimation. Note that multiple linear regression differs from multivariate regression, which is a statistical method that compares multiple dependent variables using one independent variable (Jenkins-Smith et al. 2017).

Though useful for identifying relationships between many independent variables and a dependent variable, multiple linear regression is not infallible. There are a series of assumptions that must be met for multiple linear regression to be feasible. First, the slope estimates B_{ij} must be linear, or constant, across the dataset. Non-linearity is tested for using residual plots, as the standard errors of the residuals should be evenly distributed around the predictive surface that is developed by the regression. Further, the value of any given independent variable should be fixed, or not systematically related to the value of the other independent variables (Jenkins-Smith et al. 2017). This can be tested for using correlation tables, which should show that the correlation coefficients between the independent variables are all less than .80, as well as Variance Inflation Factor (VIF) estimation, which should show values less than 10. The three major assumptions of OLS also apply to multiple linear regression, which are that the residuals must have identical distributions, be independent of the independent variables and other residuals, and that errors should be normally distributed. These assumptions are all tested using residual plots (Jenkins-Smith et al. 2017).

When the core assumptions of multiple linear regression are satisfied, this analysis can be used to compare many independent variables that may have competing effects on a dependent variable. Multiple regression accounts for partial effects of independent variables on one another, which are a result of the shared variance between the dependent and independent variables. An example of this exists between the relationship between the number of fire trucks that respond to a fire and the number of deaths a fire causes. Though an analysis of only these two variables might find a positive relationship between the number of fire trucks that respond to a fire and the number of fire deaths, a third variable, the size of the fire, drives the variance in both the number of trucks that respond to a fire as well as the number of deaths (Jenkins-Smith et al. 2017). Multiple linear regression can remove the effect of other independent variables on each individual independent variable and the dependent variable, thus removing these partial effects. This process only works to an extent, as independent variables that perfectly predict each other result in a condition of multicollinearity that can violate the assumption of fixed independent variables (Jenkins-Smith et al. 2017). These variables can be removed if multicollinearity is discovered through correlation tables or VIF estimation.

The final step of multiple linear regression analysis is to apply a *t*-test to the slope coefficients, to identify their statistical significance. For this study, the null hypothesis is that the independent variables tested in my model do not have any effect on the SPC word or color scores, or in other words that the value of each slope coefficient in the model will be equal to zero. The *t*-test was performed by finding the ratio of the slope coefficient to the standard error of the slope coefficient, which produces the *t* value for the coefficient (Jenkins-Smith et al. 2017). A statistically significant result would suggest that the null hypothesis should be rejected in favor of the alternative hypothesis, or that some relationship exists between the independent variable and the SPC word or color scores. The estimate of each slope coefficient, their p-values based on the t-test, and the standard error of each slope coefficient was then reported in the results.

Chapter 3 – Results and Findings

CTA Thematic Analysis

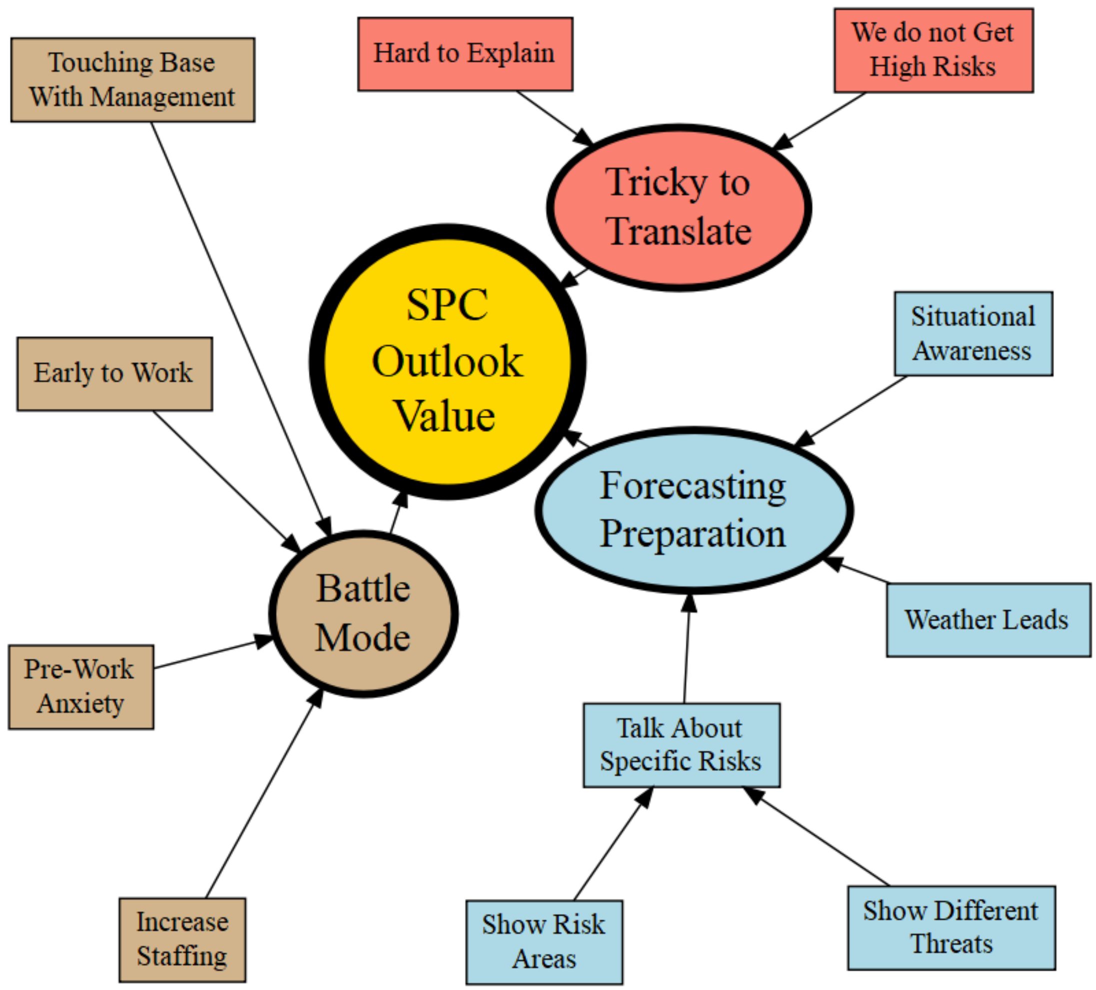


Figure 7: Thematic map compiled from the CTA interview data, with themes and subgroups grouped by color shading. Arrows indicate how subthemes build into overarching major themes.

Though my thematic analysis of the CTA interviews, three major themes emerged that described the value that broadcast meteorologists glean from the SPC convective outlook (Fig. 7). First, "Forecasting Preparation" relates to how broadcasters used the outlook to communicate risk areas and threats, dictate to their producers when weather coverage would lead the newscast, and how the outlook helped them maintain situational awareness on severe weather days. The second theme, “Battle Mode”, describes forecaster’s changes in behavior with increasing outlook levels, as they mentioned touching base with management more often, coming in earlier to work, increasing their staffing, and feeling anxious and concerned about upcoming severe weather. Finally, “Tricky to Translate” captures the forecaster’s issues with explaining the meaning of the different risk categories, as well as their frustration that the scale does not accurately capture the range of severe weather events in less severe-prone regions of the country. These themes are supported by excerpts from participant testimony, and overall paint a picture of a widely used, but imperfect, forecast product.

The first major subtheme that defines Forecasting Preparation is “Talk about Specific Risks”, which is further broken down into two main ideas. Seven broadcasters described using the SPC outlook to “Show Different Threats”, using a variety of methods to display the risk posed by tornado, hail, and wind threats. Some of those participants described repackaging the outlook into a bar graph that displayed the risk of each severe weather hazard from low to high, which they would then show their viewers. MET6 used one such graph to show that “the threat of hail is this, [the] threat of tornado is this,” while MET11 used their bar graph to show the “specific threats for viewers to worry about.” Others, such as MET5 and MET7, would show the risk map on air, and like MET11 would “show the hatched areas for different threats” to explain to their viewers which weather hazards were most threatening. Overall, by showing the hazard breakdown or by reinterpreting the outlook into a bar graph, these broadcasters were able to use the convective outlook to help them describe to their viewers exactly what kinds of threats to expect on a severe weather day.

A subset of these broadcasters highlighted the importance of not just using the outlook to show which weather threats to expect, but also to “Show Risk Areas”. The three broadcasters that mentioned displaying the original outlook instead of bar graphs specifically noted that the outlook allows them to “show risk areas… [and] general timing and likelihood” (MET12). MET7 also preferred to “show the Slight Risk map and where the area is, talk about specific risks,” but lamented that “risk areas do not show timing.” Though only reported by three of the 15 participants, the importance of communicating the where of a threat helps explain why these broadcasters prefer to show the original outlook instead of focusing more on repackaging it into bar graphs.

Beyond using the SPC outlook to communicate severe weather hazards to the public, 13 of the 15 participants reported that their stations use the product to help determine when "Weather Leads" their news broadcasts. Eight of the 13 directly reference weather coverage leading the newscast when specific outlook tiers are issued for their area, such as MET4 leading on Slight Risk days or MET6 on Moderate and High Risk days. Other broadcasters linked the outlook to their station's weather alert advertising, such as for MET7, where they "have a 'storm watch' stinger on hit intros and can turn banners red for an 'Alert Day'." Seven participants specifically mentioned either "Weather Alert" or "Code Red" theming being added to broadcasts when SPC outlooks are issued above a specific tier. These decisions are generally made in coordination with station producers, as MET8 describes "[emailing an] alert to producers and other news ops to identify [the] day as a weather day." The SPC outlook helps these broadcasters define the expected weather risk better for their producers, allowing them to change the theming of the news show to better focus on the severe weather risk on a storm day.

One last direct use of the SPC outlook for the broadcasters in this study can be found in their wake-up routines, as 11 of the 15 broadcasters reported using the outlook as a "Situational Awareness" tool. Generally, participants reported using the outlook to orient themselves to the potential weather impacts of a day before diving into a comprehensive forecasting process, to "make sure nothing crazy changed overnight" (MET12). Of the 11 participants that reported performing this check in on the outlook, five reported doing so on No Risk days, while the other six reported starting to do so for Slight or High and Moderate Risk level days. Checking the SPC outlook as part of waking up and getting ready for work was a popular use of the product for these broadcasters and helped them catch any sudden forecast changes that may have occurred while they were sleeping.

In addition to these examples of broadcasters directly using the SPC outlook, there were also more subtle changes in behavior that the participants reported, as they entered a sort of "Battle Mode" on higher severe weather risk days. The first major change comes in the form of broadcasters increasingly "Touching Base with Management" on higher SPC risk level days. Fourteen of the 15 broadcasters interviewed reported this increase in contact with their producers and weather teams, with MET15 only mentioning contact with their station management for Moderate and High Risk days. Generally, broadcasters reported having little contact on a No

Risk day, such as how MET18 "[makes] an appearance, [and will] usually send out an email if I think there's going to be a need to change the regular schedule," when they arrive at work. However, MET18 also has "pretty liberal control of the weather center" on Slight Risk days, and management will allow them to "go over commercials or work around them," while on Moderate or High Risk days they will "call our assignment manager… and make suggestions for pre-deployment of other news crews." MET9 follows the same pattern, where on arrival to work on a No Risk day they will only briefly update their producer, whereas on Slight risk days they add a check in with their storm chasers, and on High Risk days they "brief producers [and] staff, and check in with chasers [to suggest target areas and test their video equipment]… [talk] to everyone." Overall, when more significant weather is forecast by the SPC, these broadcasters would take more and more control of the newsroom activity, to better cover any potential thunderstorm threats that would occur.

The second behavior change that broadcasters reported was one they indirectly shared, as they reported in their CTA interviews that they would arrive "Early to Work" on days with higher SPC outlook levels. For this analysis, I compared participants reported time of arrival to work on Slight Risk and Moderate and High Risk days to their reported time of arrival on No Risk days, compiling the results in Table 3. Some participants, such as MET7, did not report any change in their arrival time to work, but as the averages show, participants arrived roughly 45 minutes early on Slight Risk days, and two hours early on Moderate or High Risk days.

Table 3: The time of arrival to work each participant reported, as compared to their arrival time on a "No Risk" day.

Arrival Time at Work by Risk Level compared to a No Risk Day		
MET ID	**Work Arrival Slight Risk**	**Work Arrival Moderate/High Risk**
MET4	2 hours early	4 hours early
MET5	Same time as no risk	0.5 hours early
MET6	1 hour early	2 hours early
MET7	Same time as no risk	Same time as no risk
MET8	0.5 hours early	1 hour early
MET9	2 hours early	3.5 hours early
MET10	0.5 hours early	2 hours early
MET11	Same time as no risk	0.5 hours early
MET12	Same time as no risk	3.5 hours early
MET13	Same time as no risk	5 hours early
MET14	1 hour early	1 hour early
MET15	1.5 hours early	2 hours early
MET16	1 hour early	2 hours early
MET17	1.25 hours early	2.25 hours early
MET18	1 hour early	1 hour early
Average	**0.78 hours early**	**2.02 hours early**

In addition to arriving at work early, these broadcasters also reported that they would "Increase Staffing" on days with higher risks issued by the SPC convective outlook, as comparted to normal operations. Though eight of the 15 broadcasters suggested that they generally did not have additional meteorologists on duty during Slight risk days (see Table 4), only one morning meteorologist, MET17, reported no change to their staffing on High and Moderate risk days. The average number of additional broadcasters brought into the station for Slight risk days (where 0-1 is counted as 0.5 for averaging purposes, etc.) was 0.79, while on average the broadcasters interviewed brought in 1.79 additional broadcasters for Moderate or High Risk days, with many reporting that such days were all hands on deck.

Table 4: The number of additional broadcast meteorologists at work each participant reported, as compared to their staffing on a "No Risk" day.

Additional Broadcast Meteorologists on Duty by Risk Level compared to a No Risk Day		
MET ID	**Slight Risk**	**Moderate/High Risk**
MET4	0	1-2
MET5	1	2-3
MET6	0	1-2
MET7	2	2-3
MET8	0	2
MET9	0-1	1
MET10	2	3
MET11	0	0
MET12	1-2	2-3
MET13	0	1
MET14	1	1-2
MET15	0-1	1
MET16	1-2	1-2
MET17	1	2-3
MET18	0	1
Average	0.79	1.79

Finally, as part of the "Battle Mode" effects that broadcasters reported feeling on higher SPC outlook risk days, seven of the broadcasters reported feeling "Pre-Work Anxiety," in particular for high and moderate risk days. This generally revealed itself in admissions that High Risk events were anxiety- and nerve-inducing, or in complaints of not being able to sleep well before these severe weather days. MET4 reported having trouble with anxiety before work, as well as difficulty "winding down" after significant days. MET8, from a less storm-prone region, also reported feeling this way on Slight Risk days, and that they would be "nervous, [it's] so nerve-wracking with severe weather potential." Both MET8 and MET18 discussed that they would often suffer from a lack of sleep the night before a High or Moderate Risk event, with MET12 going as far as "to take a sleeping pill when I get home to ensure that I get a full eight hours sleep." As MET6 mused, broadcasters "feel like [they're] juggling a lot, trying to communicate with [the] newsroom, viewers, and stay on top of what's changing." Though they did not directly attribute this stress to being from SPC outlook, broadcasters were asked to describe how they feel on days with High and Moderate Risks, suggesting that the forecast from the SPC is at least indirectly tied to their stress.

Finally, broadcasters mentioned that though they generally appreciate the SPC convective outlook, there are times where it can be "Tricky to Translate." The first of the two translation troubles broadcasters reported is related to the outlook being "Hard to Explain." Nine participants reported issues with translating the outlook's meaning, both with words and percentages. MET18 finds that the words of the outlook do not match up with the threat they are trying to communicate:

"From a public understanding standpoint people don't know the difference between the SPC categories. [The KISS (Keep It Simple Stupid) principle is important when you only have 2 min [to present the weather]. Moderate Risk is hard to explain, and slight tells the wrong message to the public, it's actually an elevated risk versus a normal day."

Others, like MET17, agree that the words for the categories don't always match up, as their "local definition of Moderate/High would be a day where actual watches are issues and storms are lined up across the whole DMA (Designated Market Area)." MET16 also reported that they "struggle with categories," while MET8, a bilingual meteorologist, mentioned that "there's no way of translating [into Spanish] because it's so technical." Though the convective outlook's probability graphics could also be used to explain a severe weather threat, "News Directors have a hard time understanding probabilities, they would blow off a 10% tornado probability due to the hurricane probabilities [that go as high as 100%]" (MET16). MET11 also mentions that they "will show the hatched areas for the different threats but not the percentages because I don't want to explain it." Overall, these broadcasters were frustrated with the way that the SPC convective outlook words and probabilities lined up with the impacts they were trying to communicate to their viewers, frustrating their efforts to explain the threat of severe weather.

The other major issue multiple broadcasters discussed was the infrequency of the higher tiers of the SPC outlook, some even saying that "We don't get High Risks." The four broadcasters that reported this issue were from more northern and eastern regions, including places like New England where a High Risk outlook has never been issued. MET17's comment about the categories not lining up with the impacts in their region is a good example of this, as they add that they "don't get many of these Moderate/High Risk days… Unheard of." This sentiment is mirrored by MET13 and MET15, who both reported that the "baseline for Slight is so high" that the actions they take for High and Moderate Risk days are effectively the same as those that they would take on Slight days. This rarity also can lead to suspicion, as MET16 suggested that after their "first High Risk ever was in January, [and after] 4 Moderates in 20 years, [there is a] high bust potential." For these broadcasters, the SPC convective outlook fails to accurately represent the variety of severe weather events they face, as their worst days do not compare to the extreme severe weather outbreaks that can occur in different parts of the country.

Focus Group Thematic Analysis

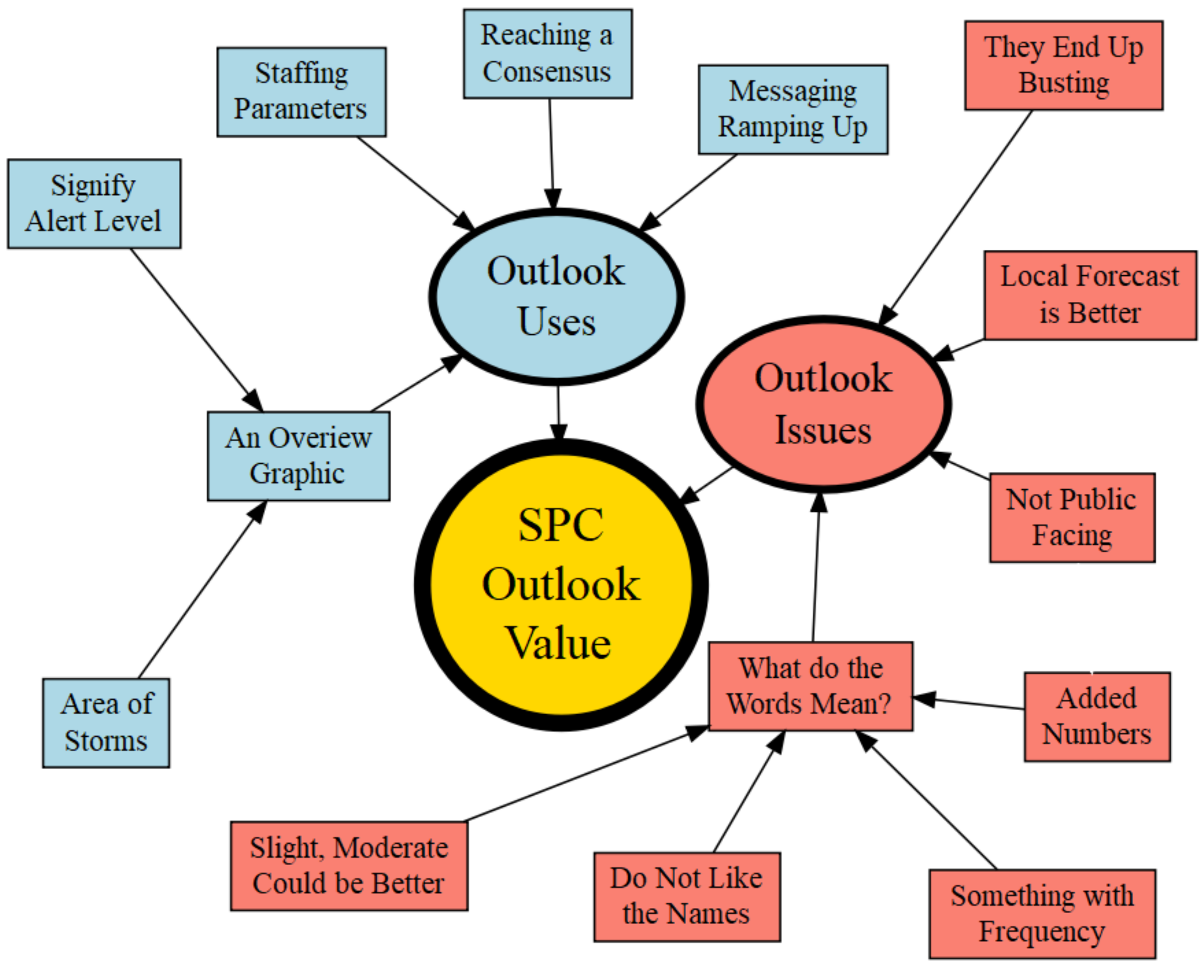


Figure 8: Thematic map compiled from the focus group interview data, with themes and subgroups grouped by color shading. Arrows indicate how subthemes build into overarching major themes.

Much like the CTA analysis, the thematic analysis of the focus group interviews revealed themes linked to the SPC convective outlooks' value to broadcasters as well as the issues they have with the product as it is now (Fig. 8). First, the broadcasters reported several key "Outlook Uses" that allowed them to better communicate severe weather threats, noting that the outlook was useful as "An Overview Graphic" that could show the "Area of Storms" and "Signify Alert Level." It was also a key part of "Reaching a Consensus" on their final severe weather forecasts, as well as determining when to start "Messaging Ramping Up" in advance of storms and for setting “Staffing Parameters.” Broadcasters did also report some “Outlook Issues,” however, suggesting that “They End Up Busting” on higher end outlooks, and that sometimes their “Local Forecast is Better.” Further, the broadcasters also suggested that the outlook is “Not Public Facing,” which is related to their confusion over “What do the Words Mean?” Issues with the outlook words came in several different groups of concerns, ranging from generally reporting that they “Do Not Like the Names” to pointing out that “Slight, Moderate Could be Better.” Others reported how they “Added Numbers” to supplement the words, while some used “Something with Frequency” to explain the message of the outlook to their viewers. Much like the CTA interviews, the data overall from the focus groups suggest a flawed but valued product that more frequently shapes behind the scenes action versus on the air communication.

Broadcasters generally described their use of the SPC convective outlook as “An Overview Graphic” for a severe weather day in one of two ways. The first group of five participants used the product to show the “Area of Storms” graphically, for example “a specific part of the viewing area that we’re concerned about convection” (MET7). Others used the outlook areas to show that “the chance for severe weather is in the western regions, eastern regions maybe not so much” (MET1), or to help steer them “towards the right places to be talking about” (MET2). MET7, in particular, described using the outlook as an image to tie together their weather story, following up a surface map with fronts and then bringing up the outlook overlay with those fronts to explain why their area was expected to have severe weather.

This group of broadcasters that used the outlook to show threat area overlapped some with the second group of six broadcasters that described using the outlook to “Signify Alert Level.” MET5 explained that “when you go past slight and you start getting into enhanced and moderate, you can reinforce to the audience that this might be a little busy in the afternoon.” Using the outlook to suggest that a severe weather event might be more significant than what viewers are used to was a strategy also employed by MET1, who would “point it out if we’ve got a Moderate Risk… [to say that] this is actually really unusual.” Further, the probabilistic breakdowns by hazard type in the Day 1 Outlook allowed these broadcasters to say to their viewers that “the tornado risk is higher with this event, the wind risk is whatever with this event” (MET6) or that “even though its mainly a wind and hail threat, there does exist a tornado threat”

(MET5). As MET4 summarized, the outlook is “a really good broad brush tool to use, to kind of signify alert level” to these broadcasters’ audiences during severe weather events.

Beyond presenting the outlook directly to their publics, eight of the broadcasters interviewed in the focus groups also mentioned that the outlook was crucial for “Reaching a Consensus” in their own severe weather forecasts. As MET1 points out, the outlook is “like a confidence booster… they are seeing what I’m seeing, we’re on the same page.” These broadcasters like to see “the same thing [the SPC is] seeing” (MET6), and will even be “texting friends, ‘did you see this look at this! SPC says this!’” (MET4). They trust the veracity and experience that the outlook channels, and thus use it “to get general information and to see what’s going on” (MET8). MET2 summarizes this idea, as they explain that:

“I almost use it more like I’m trying to reach a consensus forecast - you’ve got your own thoughts and you’ve got people whose job it is to look at severe weather, and you’re hearing their thoughts and you’re seeing where you agree and disagree.”

This consensus forming process allows broadcasters to gauge the certainty of a severe weather forecast outcome by using the SPC outlook and their own as a human ensemble forecast. As their confidence in an impactful severe weather event grows, the SPC outlook also acts to help these participants know when to begin “Messaging Ramping Up” for their audiences. Eight broadcasters reported that they and their stations would take certain actions related to SPC outlooks, as for some “once we see Moderate we’re issuing our first alert” (MET8), while for others “any time part of our viewing area in a Marginal, that’s our Code Red trigger” (MET3). Several broadcasters mentioned even longer–range actions, as MET9 described “we have a bar behind [our 7 day forecast]… I would pop it up Yellow in advance… As we got closer, I probably would have bumped that up to an Orange, and then by day 3, I’d do Orange or Red.” The outlooks help these broadcasters calibrate their messaging as well, as for an Enhanced Risk day, “we would not be ramping up as fast [as a High Risk day]. The ramp up would be closer to the event as confidence increased” (MET5). For these broadcasters, “it’s all about when that next big impact is” (MET3) and the SPC outlook helps them better adjust their response to the severity of that impact.

Indeed, all nine broadcasters mentioned that the SPC outlook also helps them adjust their “Staffing Parameters” to ensure that enough broadcasters are ready to assist to cover of a severe weather event, though the level of planning varies in detail and lead time. MET1 described that "in terms of planning if we're outlooked to day 3 or day 4, I'll look at the schedule and figure out who we can move around." Others, like MET8, make staffing decisions "the day prior, or even two days out… some of us may have to extend our hours to make sure we're covered." These actions are often preceded by discussion among the weather team, as MET6 mentioned that "Day 6, nothing changes. But [I'm] starting to talk more among the meteorologists." Further, planning also varies by the SPC Risk issued for the day. MET2 explained that High Risk days are "an all hands on deck kind of thing," while for Enhanced Risk days "we would probably send some crews [to the storms] and at least stage one or two reporters," and for Marginal Risk days "we'll have two people in the office." Across the board, these broadcasters relate the SPC Risk level they face to the level of preparedness they need to reach for effective coverage when those storms finally arrive.

Though the SPC outlooks do help broadcasters show the coverage of a potential threat and help them confidently plan their messaging strategy ahead of storms, the broadcasters I interviewed also shared some "Outlook Issues" that they felt the SPC could address. First, six of the broadcasters felt that the outlooks could build up to outcomes that would not come to pass, or in other words, that "They End Up Busting." This perception stems from some lower outlook levels feeling more impactful overall than higher ones for these broadcasters. For some, "it's usually the Enhanced that are worst for us. When they don't end up upgrading to a Moderate or High" (MET3). Others expressed frustration that "we've had Slight Risk [days] and we'll have 4 or 5 tornadoes, we've had a Moderate Risk [days] and we've had nothing" (MET4). Two separate broadcasters even jokingly said "That's the kiss of death, when its High like that" (MET7). This frustration stems from a central issue, where MET1 describes that, "if I went on TV and told someone there's a Slight Risk for severe weather today, and a 39–mile long F3 went through [my city], I wouldn't show my face the next day!" Broadcasters seek to maintain credibility in the eyes of their audience, and if the outcome of a weather event does not match the perception the audience holds based on the weather forecast, this credibility can be lost. Because broadcasters feel that higher outlook levels can underperform in terms of impacts, while some middle- or lower-level outlooks can overperform, they seek to use language they feel better represents the storm threat they expect instead of what the outlook states. In these cases, broadcasters may "handle [an Enhanced Risk] the same way we handled the High Risk day" (MET8).

Potentially related to this concern for convective outlooks busting low or high on storm impacts, some broadcasters also mentioned feeling that their "Local Forecast is Better." Five broadcasters mentioned feeling that their local insight often overwrote the threat forecast by the SPC, including all three of the meteorologists from the Northeast that were interviewed. MET9 explained that they felt "like a local forecaster is always going to have kind of a better idea, than a national forecaster," to which MET8 added "we get all these EF0s or very weak tornadoes, and the risk for tornadoes is pretty much never there." Part of the issue may stem from local preconceptions of what is a "High Risk" day, such as when MET2 suggested that "a Moderate Risk doesn't happen very often [for us], and we have big events that are outside of Moderate Risk days and High Risk days." Similar to the concern about forecast busts, these broadcast meteorologists feel that the SPC outlook does not properly capture their "big" days, and they end up relying more on their local forecast knowledge to anticipate when those days will happen.

Further, six of the broadcasters had concerns about showing the unedited convective outlook to their viewers because the product is "Not Public Facing" in nature. Similar to the concerns about forecast busts, MET7 explained that viewers "don't get that it has to do with the likelihood that severe weather at a specific location – if you ask anybody if you just had a Slight Risk for severe weather, could we see tornadoes? Most people on the street are gonna say no." This concern for viewers placing greater emphasis on impacts was shared by MET8, who stated, "I feel like 5% [tornado risk] for [my area is] high but for a typical person its low." MET1 mentioned that SPC outlook information "is important for me, but it's not terribly important for the viewer." Indeed, many of these broadcasters felt that their job was to translate the outlook into a more digestible format, and view it as "a tool for us, and then it's our job to break it down into simpler terms, or more "street slang" if you will, for people who are watching" (MET2).

All three of these concerns with the status of the outlook can be traced to one central issue that all nine broadcasters described, which is that they have trouble explaining "What do the Words Mean?" These concerns grouped into four subthemes, the first of which was voiced by four participants that simply "Do Not Like the Names." MET7 explained that while "generally I like the outlooks, I don't like the names of the threat levels." MET4 voiced a similar opinion, suggesting that "I think the way [the outlook is] named is not a good choice." However, seven of the broadcasters were more precise with their criticisms, narrowing their suggestions to "Slight, Moderate Could Be Better." Overall, the broadcasters reported that though they like the 5-level outlook more than the older 3-level outlook, "Enhanced sounds much worse than Moderate" (MET7). Others felt that Slight did not accurately capture the risk that storm days could pose, as "slight is really indicating that there's a good chance of severe weather today, particularly where we are" (MET2). MET2 went on to say further that Slight is "the biggest thing that keeps it from being public facing, to me, is that it's just not a word that people would associate it with being an impactful day." Similarly, MET6 felt that Moderate could also be misconstrued, as "[viewers say] hey it's just a Moderate Risk," which actually represents the penultimate threat level issued in the outlook. Overall, these broadcasters were hesitant to endorse the words currently used in the SPC outlook, for fear that their meaning would confuse viewers.

Participants unhappy with the meanings of the SPC outlook words also used to other communication techniques to translate the risk messages of the outlook to their viewers. These methods fell into two camps, as three broadcasters mentioned that they "Added Numbers" when they shared the outlook, while four others described using "Something with Frequency" to portray the severe weather threat. Those that preferred numbers used a 1 to 5 scale. MET6 described how "we'll show the word but we will verbally say the number," while others like MET7 explained that "at my station we ignore the words, we just use 1 2 3 4 5." Even so, numbers did not provide a perfect solution to these broadcasters, because "we can still get tornadoes with a 1!" (MET7). On the other hand, the broadcasters that preferred frequency-based terminology felt that "you can kind of point to marginal and slight as isolated, enhanced or moderate is widespread or scattered" (MET3). The coverage- or frequency–based words helped broadcasters "convey that this has to do with likelihood of severe weather at a geographical point, versus overall threat" (MET7). Again, frequency language was not without flaws, as words like "numerous" and "widespread" could be confusingly similar for viewers (MET4). Overall, broadcasters' reported that their issues with the current SPC outlook words led them to experiment with new ways of discriminating between the outlook levels, though they also feared that the lack of consistency across these methods "just leads to more confusion" (MET7).

Survey Data Statistical Analysis

In addition to these thematic analyses, I also completed a detailed statistical analysis of survey data collected from several thousand members of the public on their interpretation of the SPC outlook. To perform this analysis, I first plotted the distributions for participant's preferred words and colors for the five ranking levels and their ranking scores (Figs. 9-10). These distributions revealed some potential points of confusion for participants, as well as a few parts of the SPC scale that work more effectively.

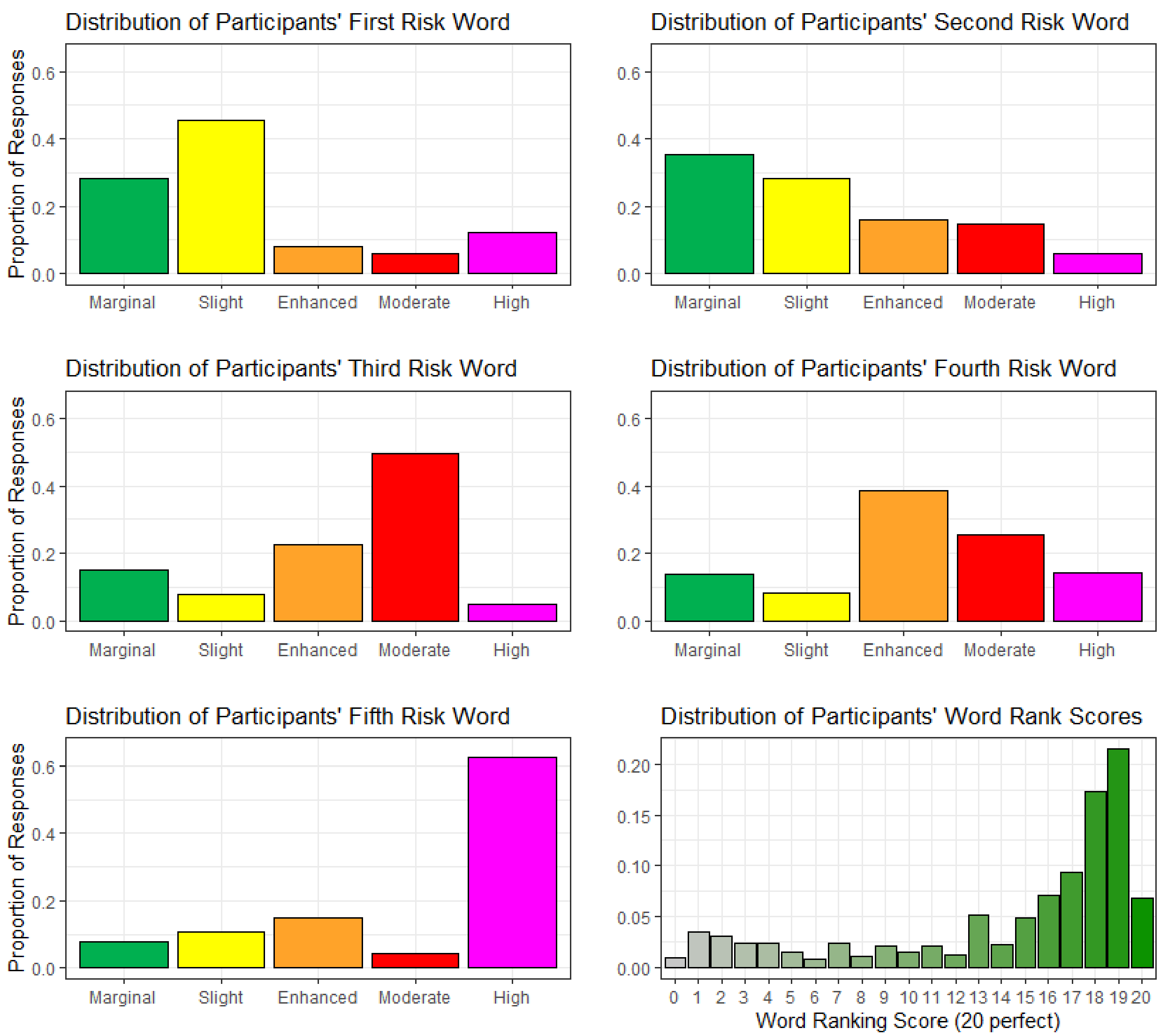


Figure 9: Graphs of the frequencies with which participants ranked each of the 5 available SPC outlook words for each of the 5 open positions for words, as well as the distribution of participants' SPC word ranking scores.

Overall, participants interchanged the positions of Marginal and Slight, as well as Moderate and Enhanced. The first four risk words as favored by the participants were thus Slight, Marginal, Moderate, and Enhanced, in that order. The distribution of participants for the second and fourth risk words were also even across the word options relative to the first and third words, which suggests the participants were confused about what word fit best for those levels. The fifth risk word was overwhelmingly chosen to be High in this sample, with 63% of participants selecting the word for that level. The distribution of participant word ranking scores was found to be somewhat skew left, with most participants making minor mistakes, like interchanging adjacent words, and very few making significant mistakes, like switching the placement of Marginal and High.

Table 5: The number of participants that ranked the order of the SPC words, for the seven groups with *n* greater than 90. The words participants selected for each position are labeled, as well as the number *n* of participants for each group, and their group's SPC word rank score. The correct rank order is highlighted by dashed border.

1st Word	2nd Word	3rd Word	4th Word	5th Word	n	Rank Score
Slight	Marginal	Moderate	Enhanced	High	505	18
Marginal	Slight	Moderate	Enhanced	High	369	19
Slight	Marginal	Enhanced	Moderate	High	224	19
Marginal	Slight	Enhanced	Moderate	High	205	20
Slight	Marginal	Moderate	High	Enhanced	158	16
Slight	Moderate	Marginal	Enhanced	High	107	15
Slight	Enhanced	Marginal	Moderate	High	105	17
Marginal	Slight	Moderate	High	Enhanced	99	17

Table 5 presents the largest groups of participants' ranking order of the SPC words, with the number of participants that ranked the words in that order and their rank score also listed. The most common ranking order participants reported swaps the rank order of Marginal and Slight, as well as Enhanced and Moderate (as seen in Fig. 9). The correct order of the SPC words was the fourth most popular ranking. Rankings that interchanged only enhanced and moderate and only marginal and slight were chosen more frequently by participants. High is consistently perceived to be the highest risk word except in the fifth most popular ranking group.

Figure 10: Graphs of the frequencies with which participants ranked each of the 5 available SPC outlook colors for each of the 5 open positions for colors, as well as the distribution of participants' SPC color ranking scores.

Participant responses by rank for the SPC colors were closer to the official rank order used in the convective outlook (Fig. 10). Green, yellow, and orange were the most popular colors for the first three ranks, though 22% and 17% of participants chose magenta for the second and third ranks, respectively. For the fourth rank, red was the third most popular choice at 19% of participants, behind magenta and orange, which were nearly evenly split at 32% and 28%. Red was chosen by 62% of participants as the fifth risk color, which may be related to the relatively low number of participants who chose red for the fourth rank level. Participants' color ranking scores were more spread out than the word ranking scores, likely due to the errors in placing magenta, but similarly were heavily skewed to the left. Notably, while only 6.8% of participants correctly ranked the outlook words, 11.7% of participants were able to correctly rank the colors.

Table 6: The number of participants that answered by ranking order of the SPC colors, for the seven groups with *n* greater than 90. The words participants selected for each position are labeled, as well as the number *n* of participants for each group, and their group's SPC color rank score. The correct rank order is highlighted by a dashed border.

1st Color	2nd Color	3rd Color	4th Color	5th Color	n	Rank Score
Green	Yellow	Orange	Magenta	Red	477	19
Green	Yellow	Orange	Red	Magenta	352	20
Green	Magenta	Yellow	Orange	Red	277	14
Green	Yellow	Magenta	Orange	Red	202	17
Green	Orange	Yellow	Magenta	Red	154	18
Green	Magenta	Orange	Yellow	Red	97	13

The color rank order results show that a plurality of participants switched the positions of magenta and red in their preferred color order (see table 6). The official color order used by the SPC was the second most popular order for these participants. Interestingly, red and green were almost universally chosen as the fifth and first ranked colors by participants, respectively, with the position of magenta varying most from the second to third rank positions. The large number of respondents that placed magenta in the second rank level likely contributed to the small peak in ranking order scores around 13 and 14 (see Fig. 10).

Next, the independent variables were individually compared to participants' scores for the word and color ordering task using a series of violin plots. As compared to participants' word scores, positive relationships appear to exist across numeracy, education, and tornado responsiveness (Figs. 11-12). A negative relationship appears to exist across age for word scores, though no clear relationships exist across weather salience, disaster preparedness, or subjective understanding of the difference between tornado watches and warnings. White participants appear to score higher on the word ordering task than Black or African American or Other racial group participants, while individuals that can correctly identify between tornado watches and warnings score higher than those who cannot. Non-Hispanic participants also score higher on the word ordering task than Hispanic participants. No clear differences in SPC word scores exists between males and females, or between the different NWS regions.

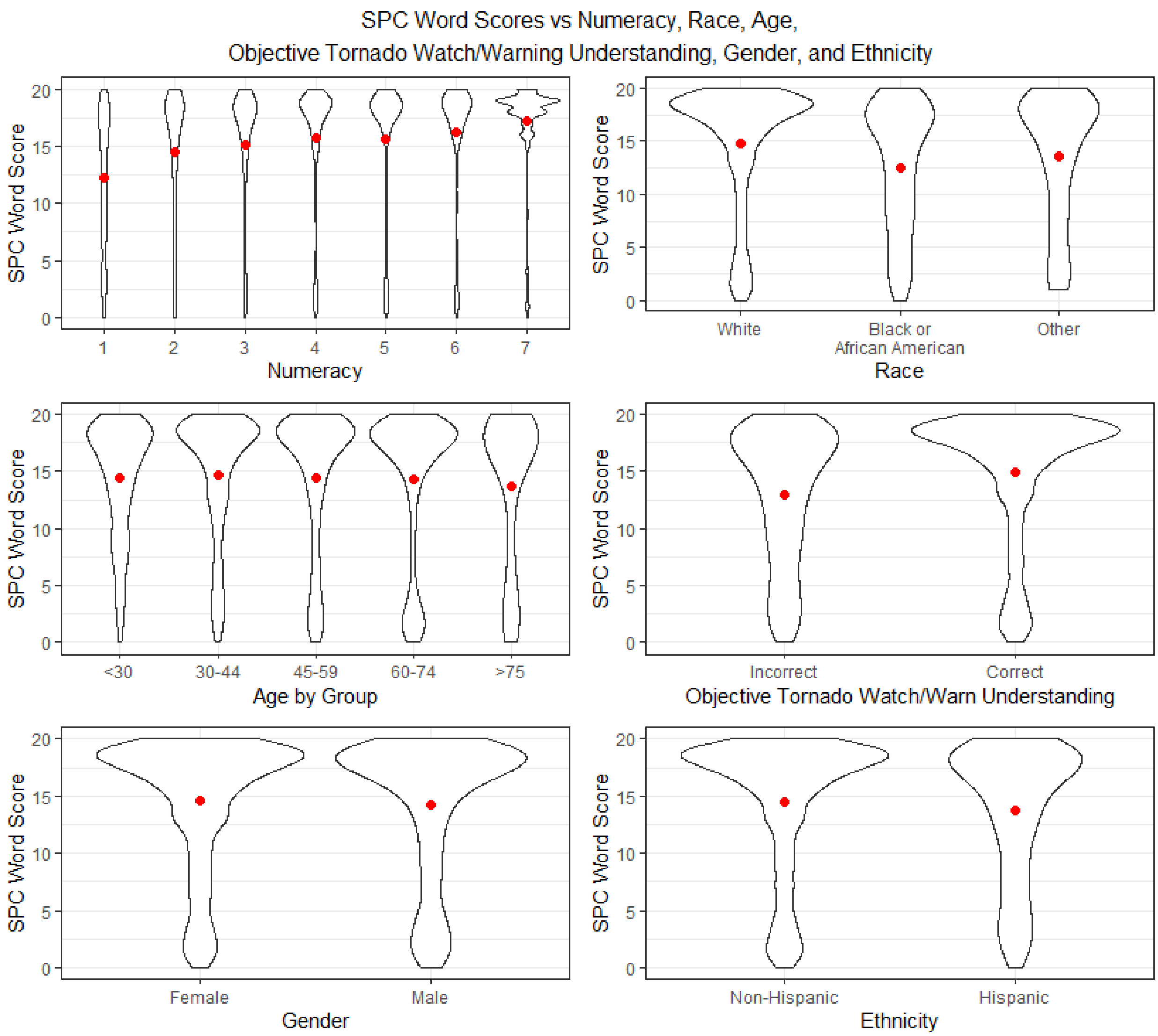


Figure 11: **Graphs of the distributions of participant word ranking scores across numeracy, race, age, objective understanding of the difference between tornado watches and warnings, gender, and ethnicity. Red dots represent the mean values for each distribution, while the width of each violin plot relates to the number of participants with scores in that part of the distribution.**

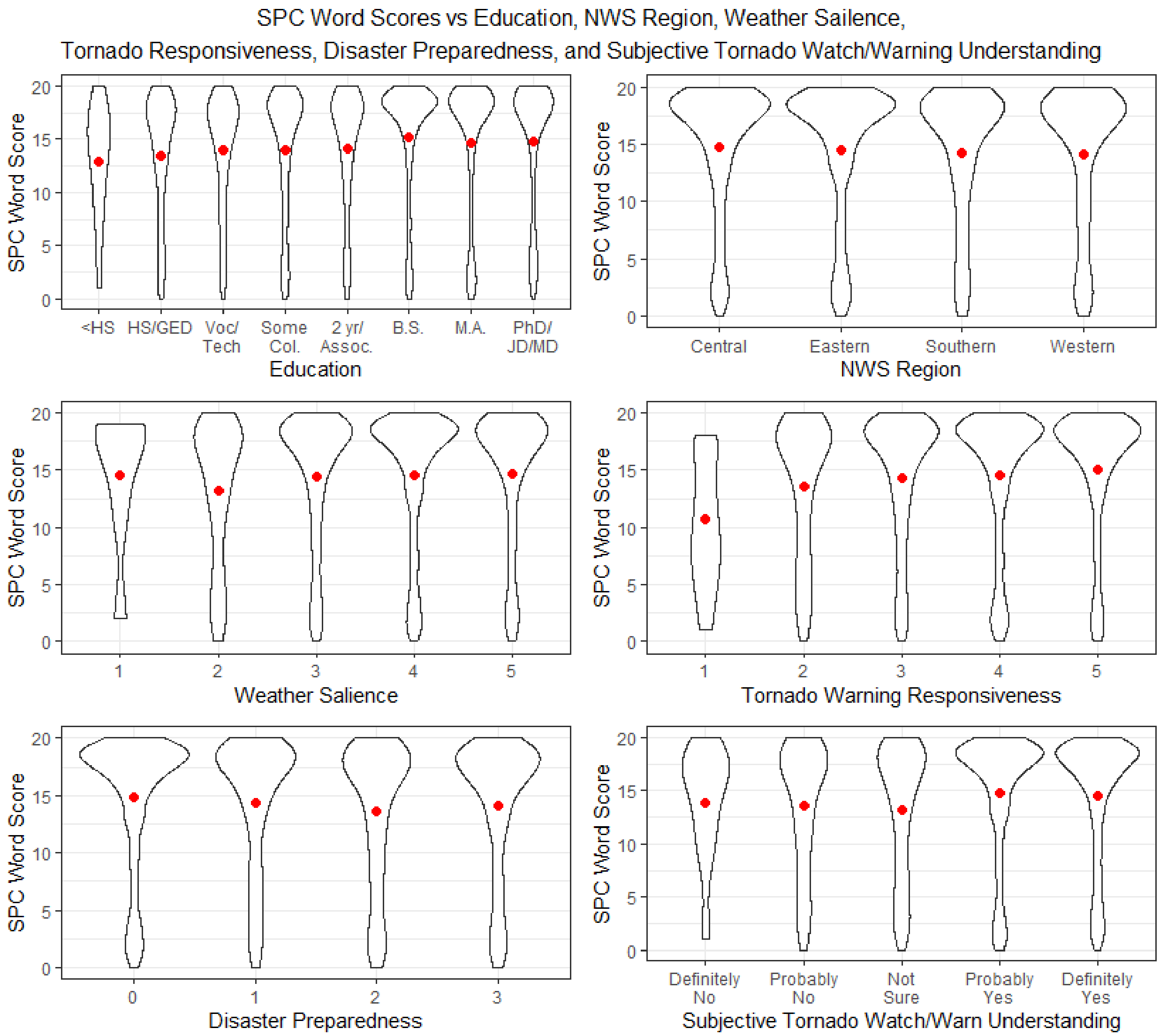


Figure 12: **Graphs of the distributions of participant word ranking scores across education, NWS region, weather salience, tornado warning responsiveness, disaster preparedness, and subjective understanding of the difference between a tornado watch and warning. Red dots represent the mean values for each distribution, while the width of each violin plot relates to the number of participants with scores in that part of the distribution. Higher values for weather salience, tornado warning responsiveness, and disaster preparedness suggest greater individual weather salience, warning responsiveness, and disaster preparedness.**

As for color scores compared to the independent variables, scores appeared to have a positive relationship with numeracy, education, and weather salience (Figs. 13-14). A weak negative relationship appeared to exist between color score and age, though no clear relationship existed between color score and tornado warning responsiveness, disaster preparedness, and subjective understanding of the difference between tornado watches and warnings. Similar to the word scores, White participants appear to have overall higher color scores than Black or African American and Other race participants, as do participants that can correctly identify between tornado watches and warnings score higher versus those who cannot. Non-Hispanic participants appear to score higher on the color ordering task than Hispanic ones, but no clear relationships exist across gender or NWS region in these plots.

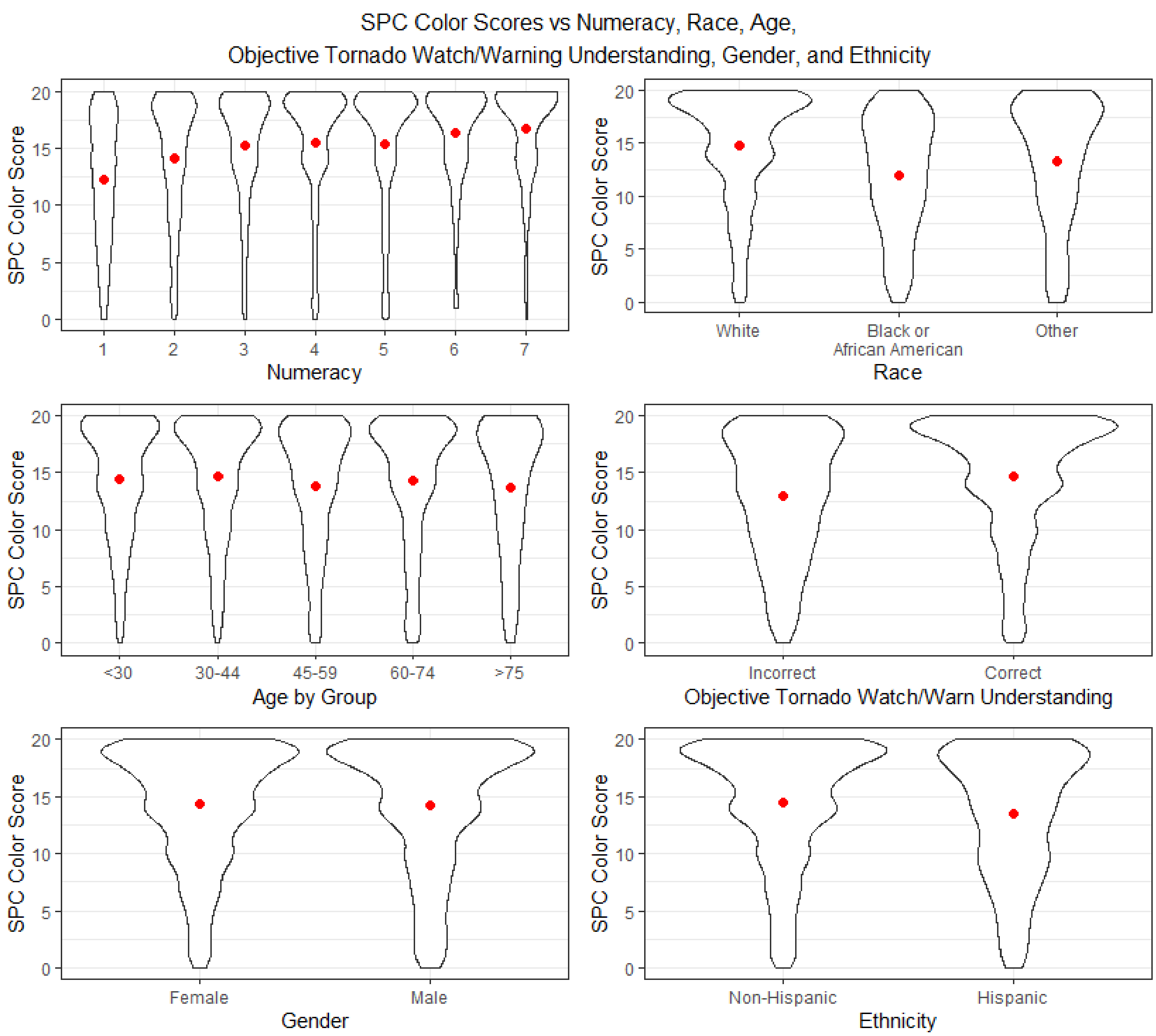


Figure 13: Graphs of the distributions of participant color ranking scores across numeracy, race, age, objective understanding of the difference between tornado watches and warnings, gender, and ethnicity. Red dots represent the mean values for each distribution,

while the width of each violin plot relates to the number of participants with scores in that part of the distribution.

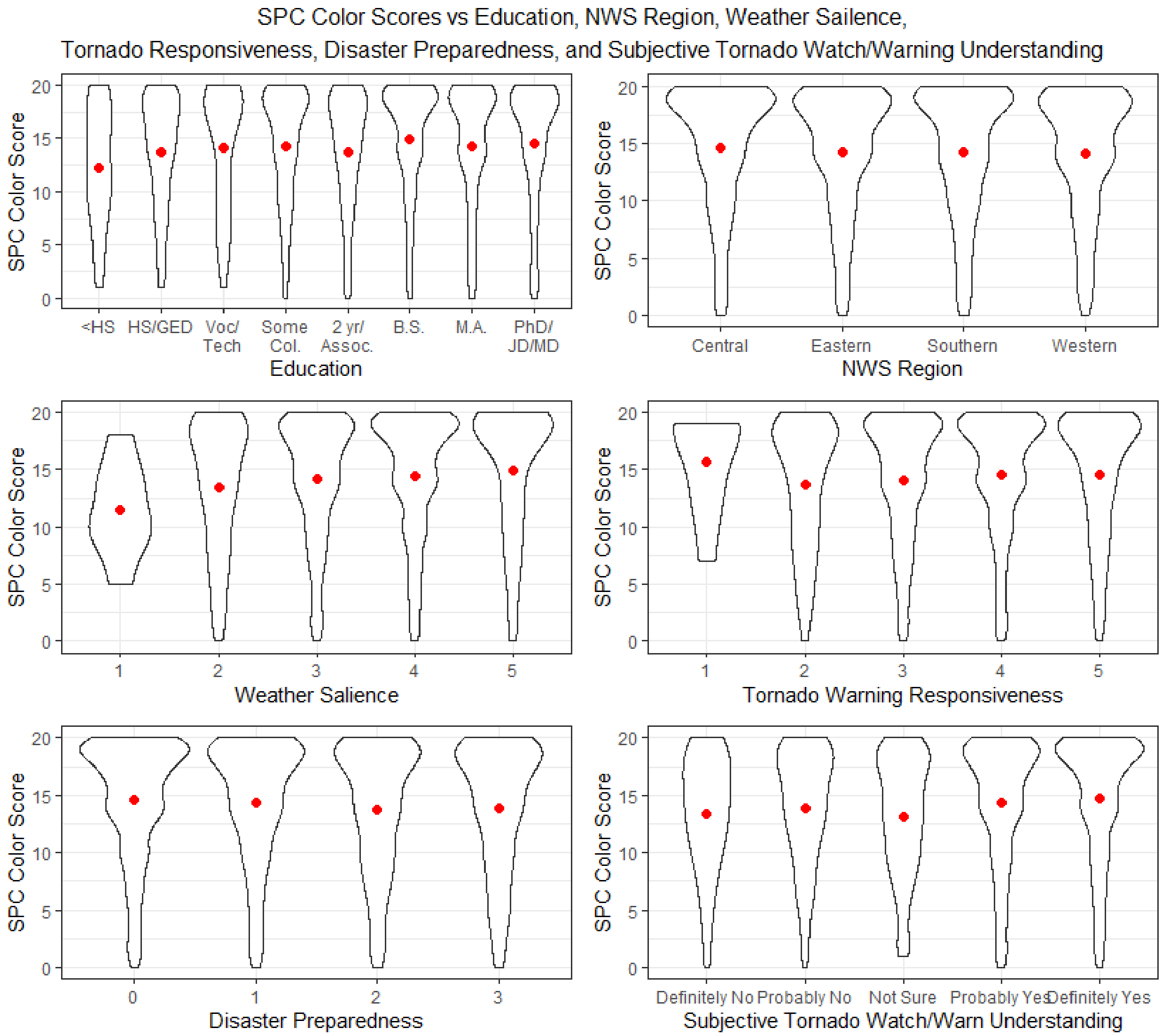


Figure 14: Graphs of the distributions of participant color ranking scores across education, NWS region, weather salience, tornado warning responsiveness, disaster preparedness, and subjective understanding of the difference between a tornado watch and warning. Red dots represent the mean values for each distribution, while the width of each violin plot relates to the number of participants with scores in that part of the distribution. Higher values for weather salience, tornado warning responsiveness, and disaster preparedness suggest greater individual weather salience, warning responsiveness, and disaster preparedness.

Before being entered into the multiple linear regression models, the independent variables were compared using a correlation table to identify any potential multicollinearity issues. Though some variables, such as numeracy and education, exhibit weak correlation, all inter-variable correlations fall below the 0.80 mark that is suggestive of strong correlation and thus possible multicollinearity (Table 7).

Table 7: The statistical output from a correlation test with all 12 independent variables. Values over 0.80 are suggestive of a strong correlation, which may suggest multicollinearity. Values indicate the correlation between the variables that match the column and row of each value.

	AGE	GENDER	ETHNICITY	RACE	EDUCATION	NUMERACY	NWS REGION	WEATHER SALIENCE	TORNADO RESPONSIVENESS	DISASTER PREPAREDNESS	SUBJECTIVE TOR. WATCH/WARN	OBJECTIVE TOR. WATCH/WARN
AGE	1.00	0.16	-0.08	0.00	0.13	0.10	0.04	0.03	0.18	-0.03	0.11	0.12
GENDER	0.16	1.00	-0.05	0.00	0.15	0.23	0.03	0.02	0.01	0.05	0.07	0.01
ETHNICITY	-0.08	-0.05	1.00	0.03	-0.06	-0.15	0.14	-0.01	-0.05	0.07	-0.07	-0.05
RACE	0.00	0.00	0.03	1.00	0.01	-0.13	0.12	-0.06	-0.05	0.03	-0.16	-0.13
EDUCATION	0.13	0.15	-0.06	0.01	1.00	0.32	0.03	0.07	0.00	0.02	0.03	0.09
NUMERACY	0.10	0.23	-0.15	-0.13	0.32	1.00	-0.01	0.02	0.04	-0.11	0.04	0.19
NWS REGION	0.04	0.03	0.14	0.12	0.03	-0.01	1.00	-0.08	-0.07	0.06	-0.10	-0.05
WEATHER SALIENCE	0.03	0.02	-0.01	-0.06	0.07	0.02	-0.08	1.00	0.25	0.16	0.26	0.09
TORNADO RESPONSIVENESS	0.18	0.01	-0.05	-0.05	0.00	0.04	-0.07	0.25	1.00	0.14	0.28	0.14
DISASTER PREPAREDNESS	-0.03	0.05	0.07	0.03	0.02	-0.11	0.06	0.16	0.14	1.00	0.09	-0.04
SUBJECTIVE TOR. WATCH/ WARN	0.11	0.07	-0.07	-0.16	0.03	0.04	-0.10	0.26	0.28	0.09	1.00	0.21
OBJECTIVE TOR. WATCH/ WARN	0.12	0.01	-0.05	-0.13	0.09	0.19	-0.05	0.09	0.14	-0.04	0.21	1.00

Table 8: The statistical output from the multiple linear regression models that compared demographics to SPC word and color rank scores. Discrete variables with multiple values, such as NWS Region, compare a reference value (in this case Central Region SPC word and color scores) to the other values in that variable (in this case Eastern, Southern, and Western Regions). Values displayed in the model columns are slope estimates, or the increase in Word or Color Score per unit of the independent variable, while the number of stars defines the statistical significance of each estimate. Values in parenthesis are the standard error of each slope estimate. All estimates are unitless.

INDEPENDENT VARIABLE	WORD SCORE MODEL	COLOR SCORE MODEL
Age	-0.028*** (0.006)	-0.028*** (0.006)
Female vs Male	-0.676** (0.223)	-0.356 (0.203)
Not Hispanic vs Hispanic	-0.193 (0.298)	-0.472 (0.272)
White vs Black	-1.210*** (0.341)	-1.892*** (0.311)
White vs Other Race	-0.490 (0.376)	-0.736* (0.343)
Education	0.130* (0.062)	0.014 (0.057)
Numeracy	0.649*** (0.072)	0.628*** (0.065)
Central Region vs Eastern Region	0.156 (0.299)	-0.058 (0.273)
Central Region vs Southern Region	-0.100 (0.314)	0.055 (0.286)
Central Region vs Western Region	-0.182 (0.338)	-0.123 (0.308)
Weather Salience	0.091 (0.167)	0.214 (0.152)
Tornado Responsiveness	0.350* (0.142)	0.156 (0.130)
Preparedness	-0.275** (0.096)	-0.178* (0.088)
Tornado Watch/Warning Comprehension	0.159 (0.131)	0.260* (0.119)
Incorrect vs Correct Tornado Watch/Warning Comprehension	1.193*** (0.267)	1.035*** (0.243)
Model Adjusted R^2 Score	0.076	0.089

Reported numbers are the slope estimate and standard error (in parenthesis)
*r<0.05, **r<0.01, ***r<0.001

These independent variables were then compared to the word and color scores in a series of Multiple Linear Regression models (Table 8), to understand the statistical significance of their apparent relationships. The two models had R^2 scores of 0.076 for the word model and 0.089 for the color model, suggesting there are other unknown independent variables that may also influence SPC outlook ranking interpretation. This model revealed no significant linear relationships between participant's ranking scores and their ethnicity, NWS region of residence, or weather salience, contrary to the individual comparisons (Figs. 11-14). Age, race, numeracy, disaster preparedness, and objective tornado watch/warning comprehension were found to have significant relationships at the r<0.05 level with both the word and color ranking scores. Gender, education, and tornado responsiveness had significant relationships with only the word ranking scores, while subjective tornado watch/warning comprehension was significantly related to color ranking scores. Additionally, VIF for the independent variables in both models did not exceed 10 for any variable, suggesting that multicollinearity is not an issue in these models and that the assumption of fixed independent variables is met (Tables 9-10). Residual plots for the continuous variables also suggest that linear relationships are present for all tested independent variables, and that residuals are normally, independently, and identically distributed (Fig. 15).

Table 9: The unitless output of the VIF calculation for each independent variable in the word score model. Values greater than 10 are indicative of potential multicollinearity between variables.

SPC WORD SCORE VIF	GVIF
AGE	1.120134
GENDER	1.109444
ETHNICITY	1.070773
RACE	1.151457
EDUCATION	1.158592
NUMERACY	1.278716
NWS REGION	1.140256
WEATHER SALIENCE	1.145558
TOR. RESPONSIVENESS	1.188962
PREPAREDNESS	1.083101
SUBJECTIVE TOR. WATCH/WARN	1.220169
OBJECTIVE TOR. WATCH/WARN	1.115259

Table 10: The unitless output of the VIF calculation for each independent variable in the color score model. Values greater than 10 are indicative of potential multicollinearity between variables.

SPC COLOR SCORE VIF	GVIF
AGE	1.120134
GENDER	1.109444
ETHNICITY	1.070773
RACE	1.151457
EDUCATION	1.158592
NUMERACY	1.278716
NWS REGION	1.140256
WEATHER SALIENCE	1.145558
TOR. RESPONSIVENESS	1.188962
PREPAREDNESS	1.083101
SUBJECTIVE TOR. WATCH/WARN	1.220169
OBJECTIVE TOR. WATCH/WARN	1.115259

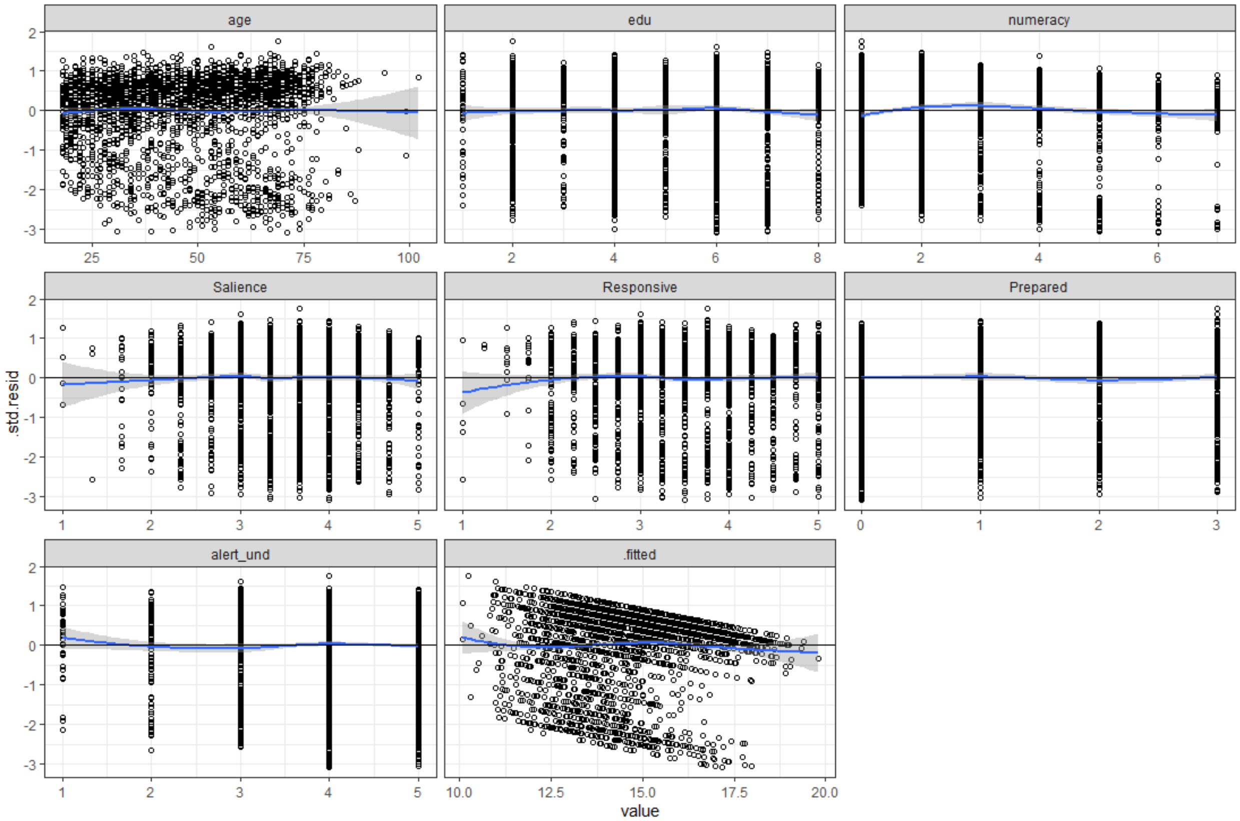


Figure 15: **Graphs of the distributions of the residuals for each of the continuous independent variables tested in this study. Standard error is displayed on the y-axis of each plot, with number values of each variable on the x-axis. Displayed residuals are for age, education, numeracy, weather salience, tornado warning responsiveness, disaster preparedness, subjective tornado watch vs warning understanding, and model fitted values, from left to right and top to bottom row.**

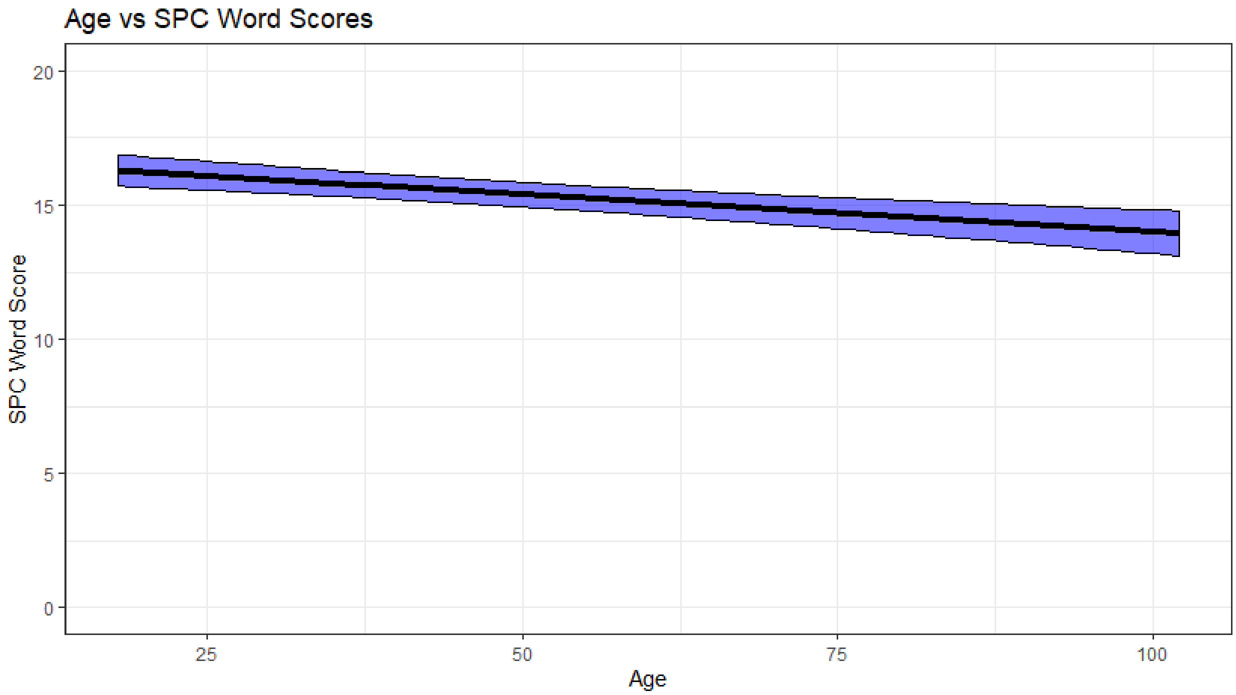


Figure 16: Graph of the slope of the relationship between Age and SPC word score, with all other independent variables held equal.

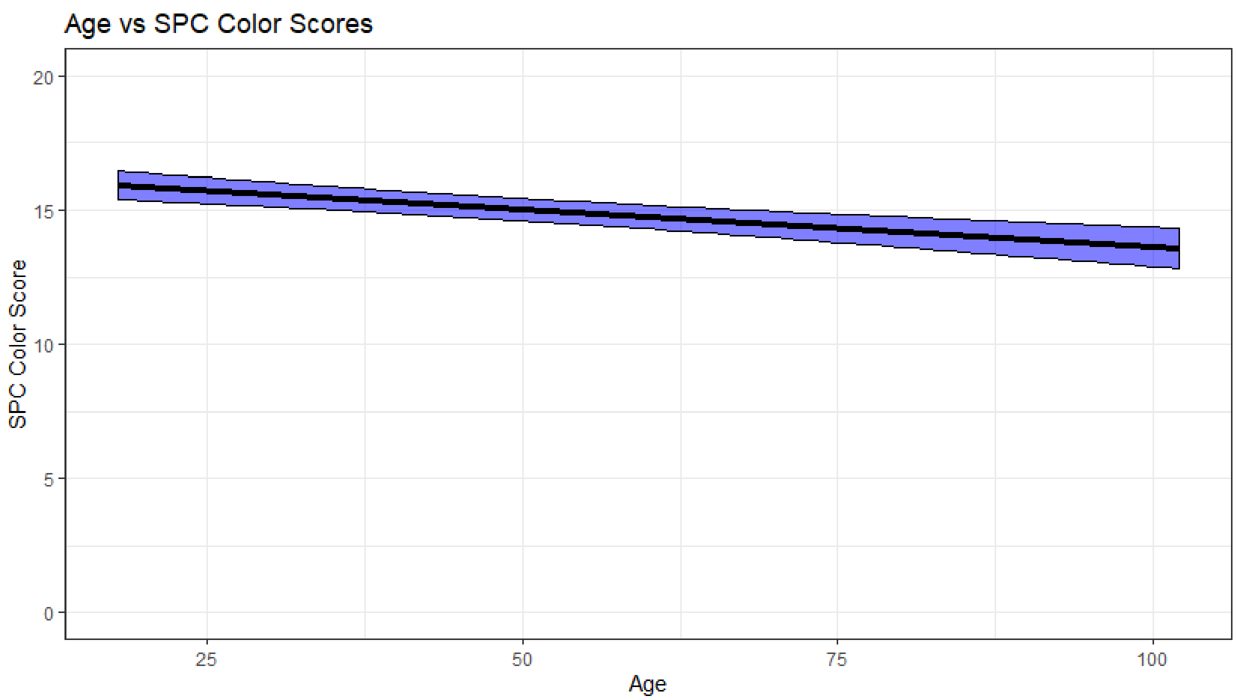


Figure 17: Graph of the slope of the relationship between Age and SPC color score, with all other independent variables held equal.

When age was compared to SPC word ranking scores, a moderate negative slope was found. Older participants reported lower SPC word rank scores (age = 103, SPC word score = 13.9) than younger participants (age = 18, SPC word score = 16.3; Fig. 16). This relationship may be weakly influenced by the low number of participants over age 75, though analysis of the violin plots (Fig. 11) suggests that a negative relationship overall should be expected. The second model compared SPC color ranking scores to the dependent variables and found a statistically significant negative relationship between the SPC score and age. Older individuals scored lower on the ranking task (Age = 100, SPC color score = 13.6) than younger participants (Age = 18, SPC color score = 15.9; Fig. 17). As with the word scores, there were few participants over age 75 which could potentially skew the slope of the relationship and make the magnitude of this relationship somewhat uncertain. The difference across the ages is significant enough to suggest that older individuals have more difficulty ranking the SPC convective outlook words and colors than younger ones.

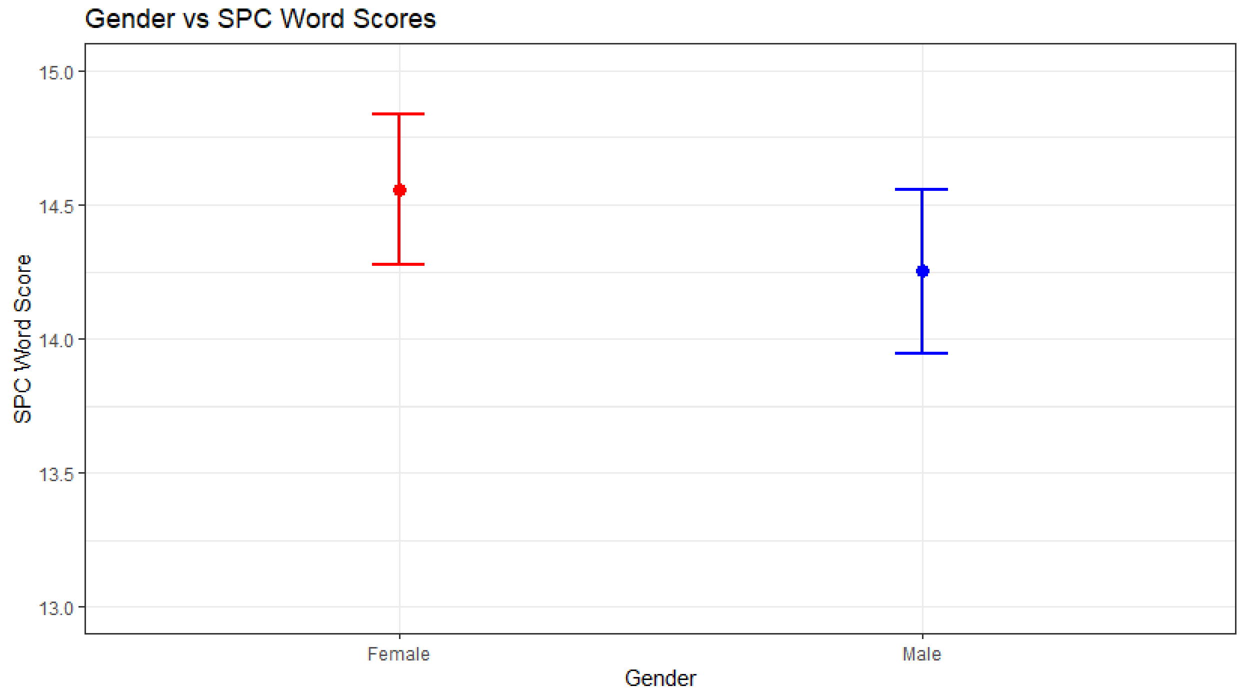


Figure 18: **Graph of the difference in average SPC word score between Female and Male participants. Points display the mean value of SPC word score for females and males, while the error bars display the 95% confidence interval around that estimate based on the standard error.**

Gender was also significantly related to SPC word score. Female participants (SPC word score = 14.6) scored just over 0.25 points higher on the 20 point scale than male participants (SPC word score = 14.3; Fig. 18). This difference suggests minimal real-world differences may exist between the two genders in SPC word interpretation. No statistically significant relationship was found between gender and SPC color scores, suggesting only minor influences of gender on outlook interpretation overall.

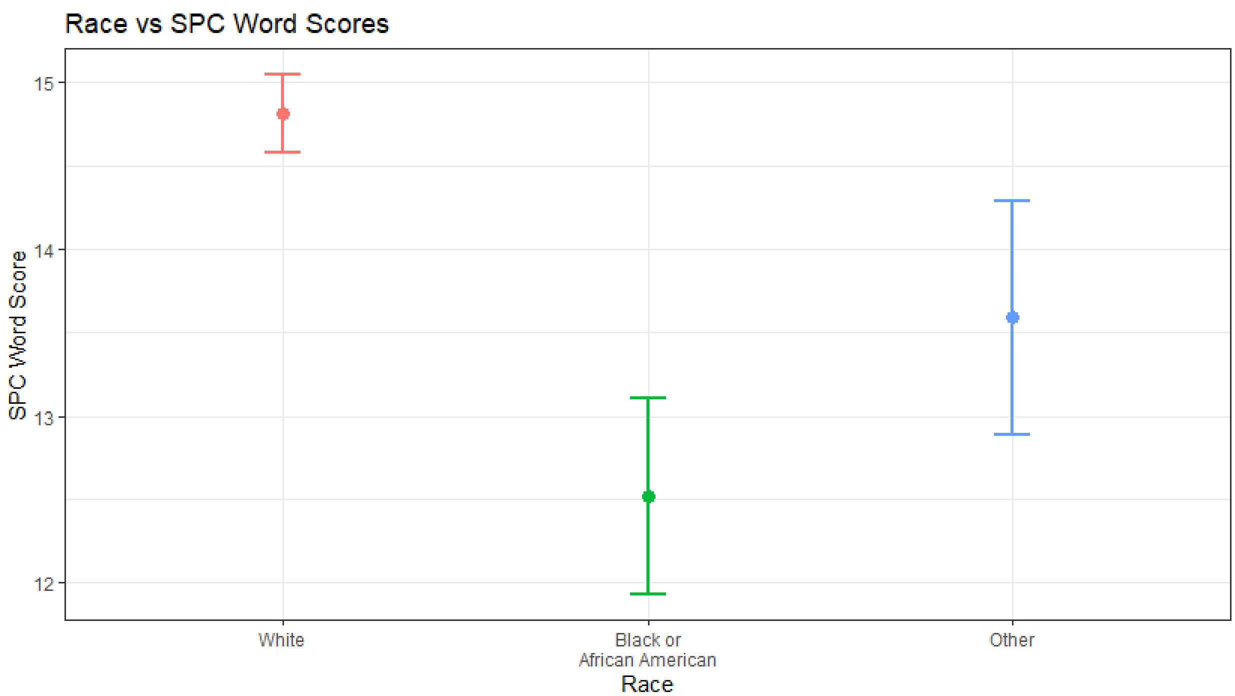


Figure 19: **Graph of the difference in average SPC word score between White, Black or African American, and participants of Other races. Points display the mean value of SPC word score for Whites, Blacks, and Other racial groups, while the error bars display the 95% confidence interval around that estimate based on the standard error.**

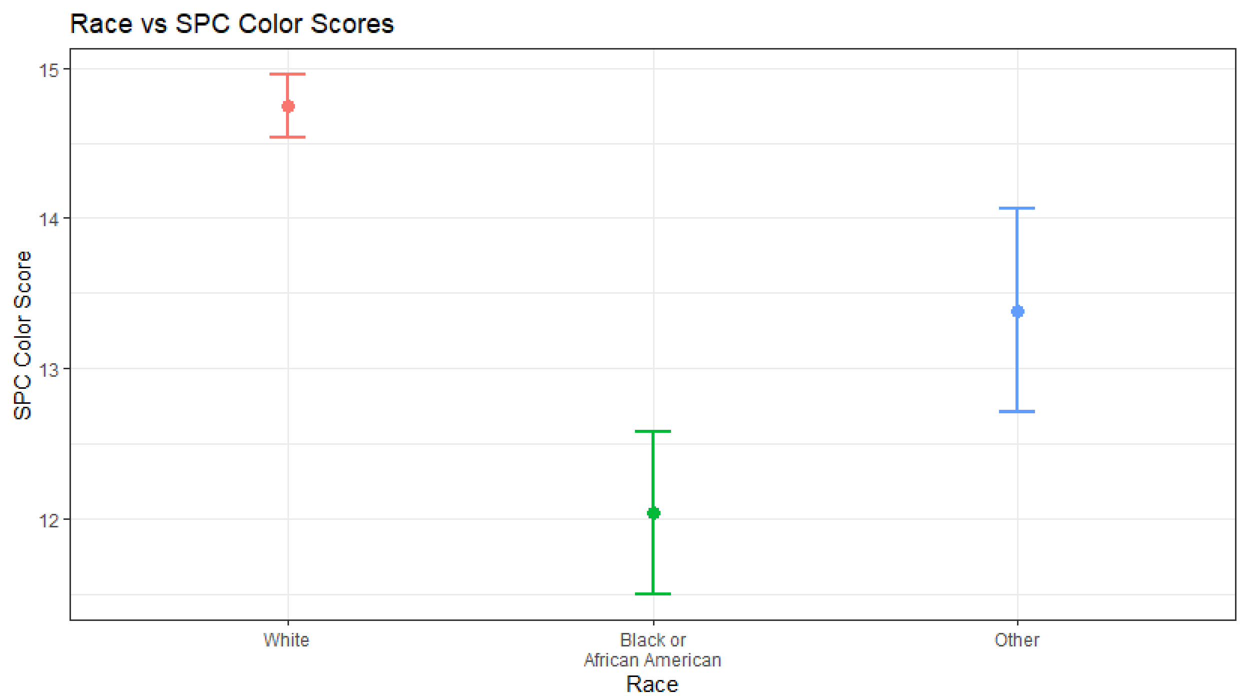


Figure 20: **Graph of the difference in average SPC color score between White, Black or African American, and participants of Other races. Points display the mean value of SPC color score for Whites, Blacks, and Other racial groups, while the error bars display the 95% confidence interval around that estimate based on the standard error.**

White respondents (SPC word score = 14.8) scored significantly higher than Black or African American respondents (SPC word score = 12.5) and those who identified as other races (SPC word score = 13.6; Fig. 19) on average. Black or African American participants on average performed 2.3 points worse on the 20-point score scale. Though other races scored slightly higher on average, they still trailed White respondents by 1.2 points. For the SPC color score model, White participants scored much higher (SPC color score = 14.7) than Black or African American participants (SPC color score = 12.0) and participants that identified as any other racial group (SPC color score = 13.4; Fig. 20). The difference between SPC color ranking scores was a 2.7-point difference, larger than the difference in the word scores. Overall, race appears to be strongly related to SPC convective outlook interpretation, which may prevent vulnerable populations from effectively comprehending advance warning of severe weather days.

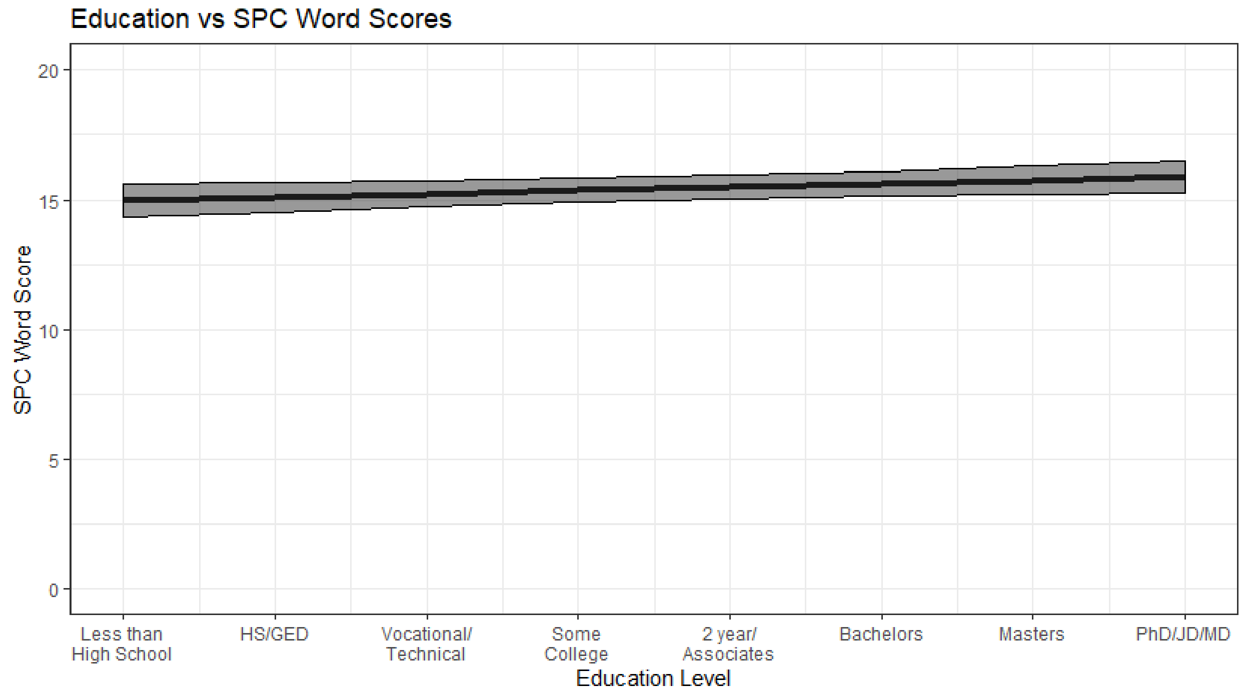


Figure 21: Graph of the slope of the relationship between Education and SPC word score, with all other independent variables held equal.

A statistically significant increase in SPC word scores was also noted across increasing levels of education. Participants with less than a High School education scored lower (SPC word score = 15.0) than those with their PhD or other advanced degree (SPC word score = 15.9; Fig. 21). With only a 0.9-point change in scores across the range of education, however, this appears to be a relatively minor effect. No statistically significant relationship was found between education and color scores, which suggests that education level has only a weak influence on outlook understanding.

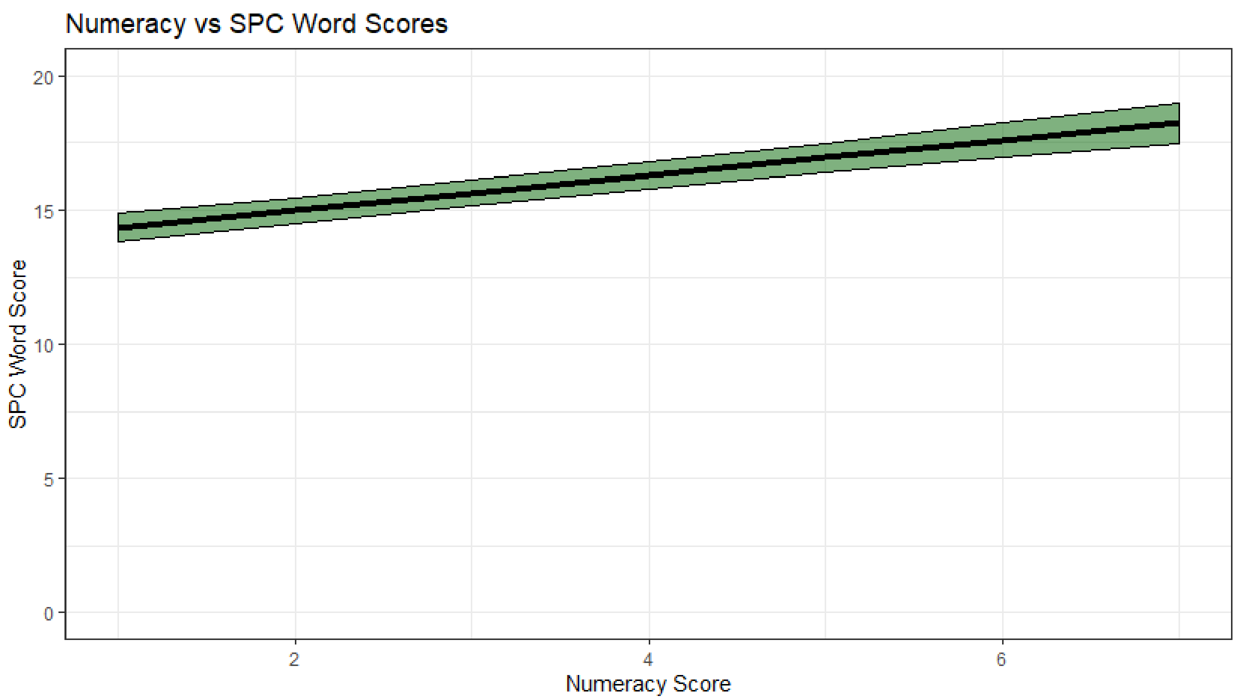


Figure 22: Graph of the slope of the relationship between Numeracy and SPC word score, with all other independent variables held equal.

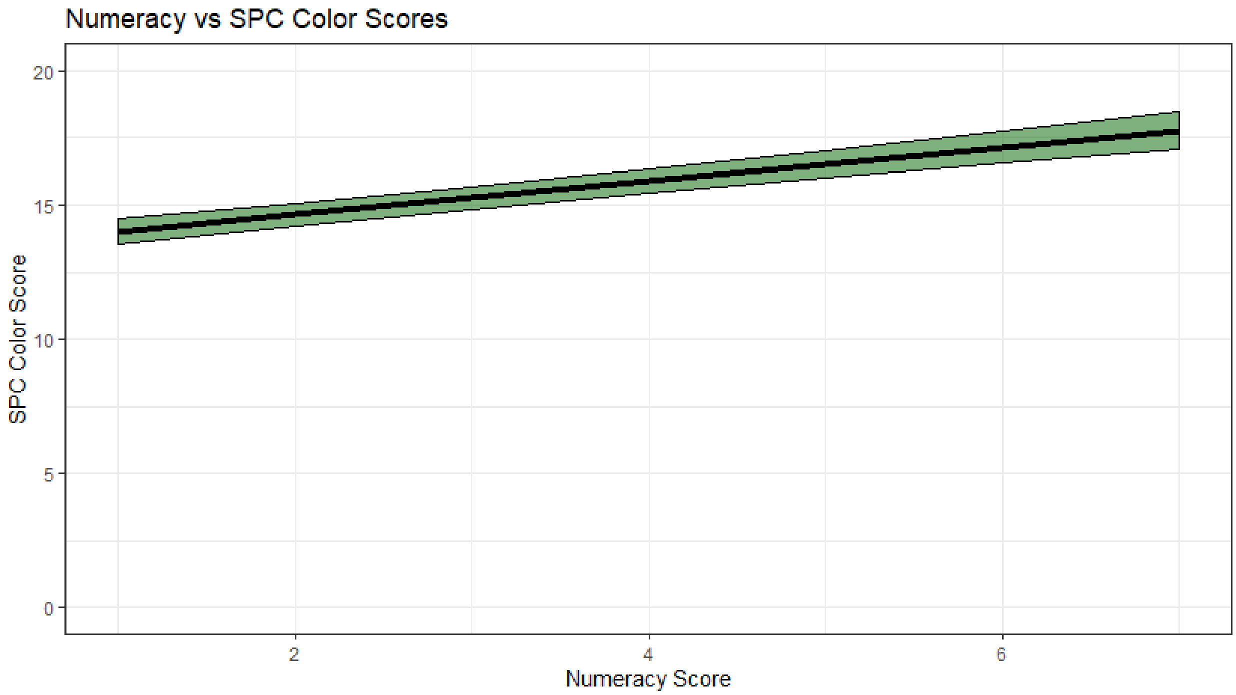


Figure 23: Graph of the slope of the relationship between Numeracy and SPC color score, with all other independent variables held equal.

A much larger effect was found across participants with different numeracy scores. Participants with lower numeracy scores (Numeracy Score = 1) recorded much lower SPC word scores on average (SPC word score = 14.3) compared to participants with higher numeracy scores (Numeracy Score = 7, SPC word score = 18.2; Fig. 22). This increase of almost four points on average is both statistically significant and operationally meaningful. Participants with lower numeracy scores are more confused about the correct ordering of the convective outlook's words from lowest to highest risk. A similar relationship was also found with SPC color scores. Participants with lower numeracy scores also scored lower (Numeracy Score = 1, SPC color score = 14.0) than those with higher numeracy scores (Numeracy Score = 7, SPC color score = 17.8; Fig. 23). The results from these two models suggest numeracy is strongly related to SPC convective outlook interpretation, more so than any of the other variables tested.

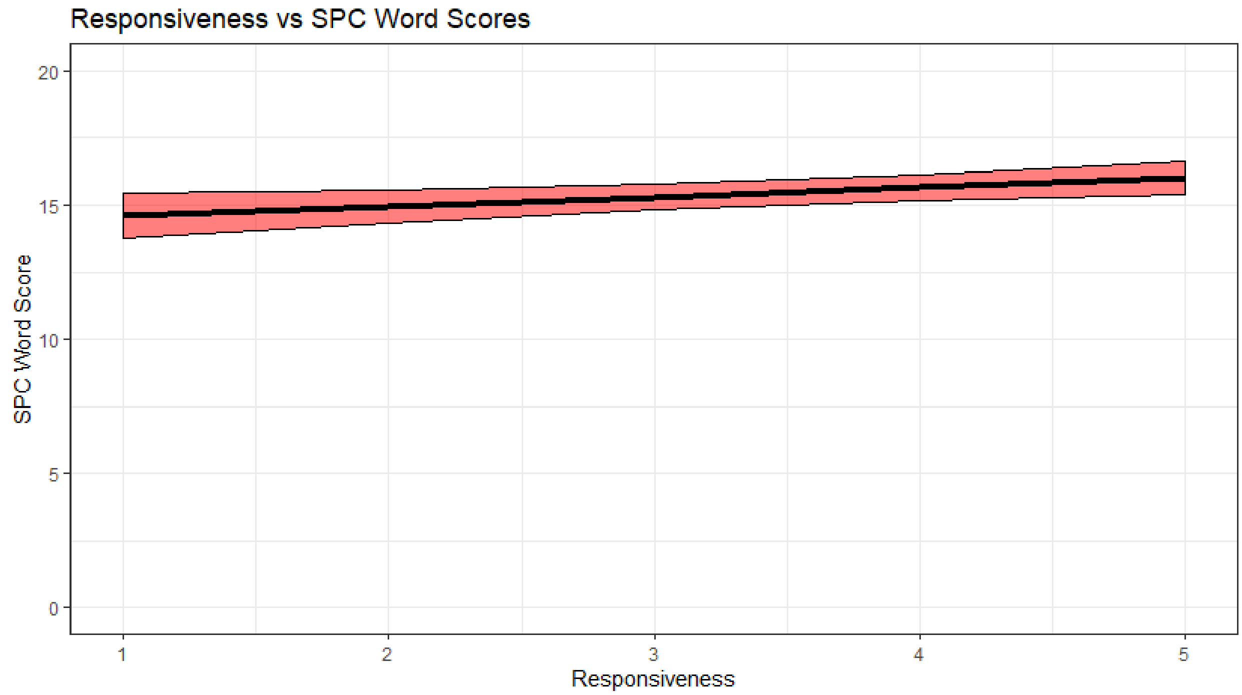


Figure 24: Graph of the slope of the relationship between Tornado Warning Responsiveness and SPC word score, with all other independent variables held equal.

Participant's reported tornado warning responsiveness was also related to their SPC word scores, though with some caveats. Participants that reported being very responsive to tornado warnings (Responsivenss = 5) had higher SPC word scores (SPC word score = 16.0) than those who reported being less responsive (Responsiveness = 1, SPC word score = 14.6; Fig. 24). Similar to age however, very few individuals recorded responsiveness scores less than two. This introduces some uncertainty in the strength of the relationship with so few available data points from which to build a linear model. Additionally, the identified slope is not very steep, which suggests that tornado warning responsiveness is not strongly related to SPC word score like numeracy is. No significant relationship existed between tornado responsiveness and SPC color ranking scores, which suggests tornado responsiveness is not strongly related to outlook understanding.

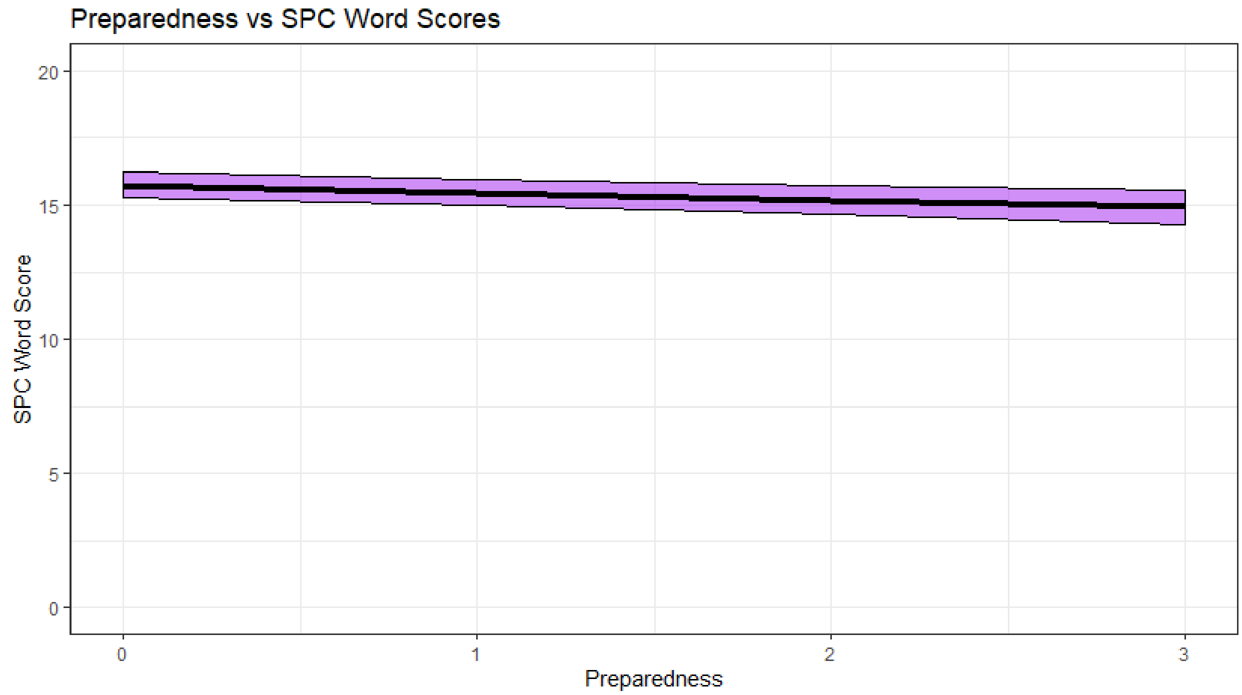


Figure 25: Graph of the slope of the relationship between Disaster Preparedness and SPC word score, with all other independent variables held equal.

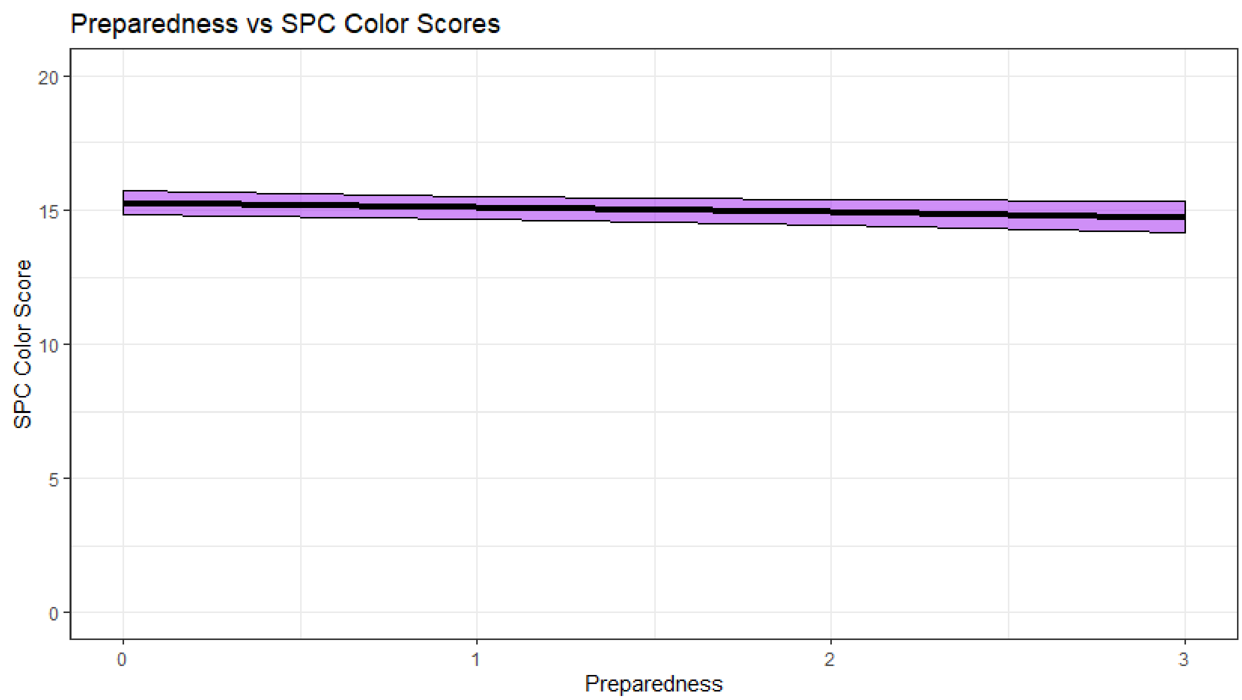


Figure 26: Graph of the slope of the relationship between Disaster Preparedness and SPC color score, with all other independent variables held equal.

Disaster preparedness of participants was negatively related to their SPC word scores, unlike responsiveness. Participants that reported having a disaster kit, a small emergency kit, and a disaster communication kit (Preparedness = 3) scored lower (SPC word score = 14.9) than those without (Preparedness = 0, SPC word score = 15.7; Fig. 25). Though statistically significant, this is a very shallow slope with a 0.8 point difference across the four levels of preparedness. Similar findings emerged in the SPC color model, as more prepared participants also scored slightly lower (Preparedness = 3, SPC color score = 14.7) than those who were less prepared (Preparedness = 0, SPC color score = 15.3; Fig. 26). Overall, disaster preparedness does not appear to have a strong relationship with SPC outlook comprehension, with a nearly flat slope present in both models.

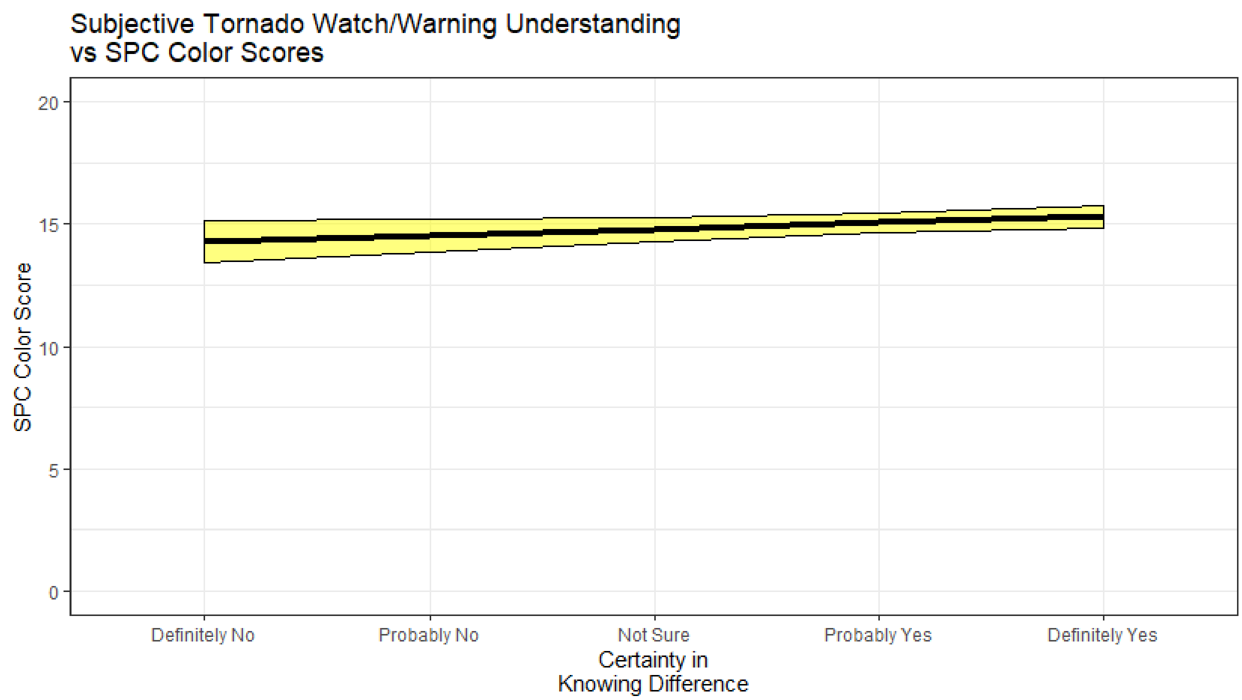


Figure 27: Graph of the slope of the relationship between Subjective Tornado Watch/Warning Understanding and SPC color score, with all other independent variables held equal.

Though not significantly related to SPC word scores, participant's subjective tornado watch/warning interpretation ability was significantly related to their SPC color ranking scores. Participants more confident in their ability to tell tornado watches from tornado warnings scored higher (Certainty in Knowing Difference = Definitely Yes, SPC color score = 15.3) than those who were less confident (Certainty in Knowing Difference = Definitely No, SPC color score = 14; Fig. 27). Similar to education, this change of about one point in the 20-point color ranking score is not very large. Few participants believed they would be unable to properly discern tornado watches from warnings, which makes the slope of this relationship more uncertain. Thus, subjective tornado watch/warning interpretation ability does not appear to have a strong relationship with SPC outlook interpretation.

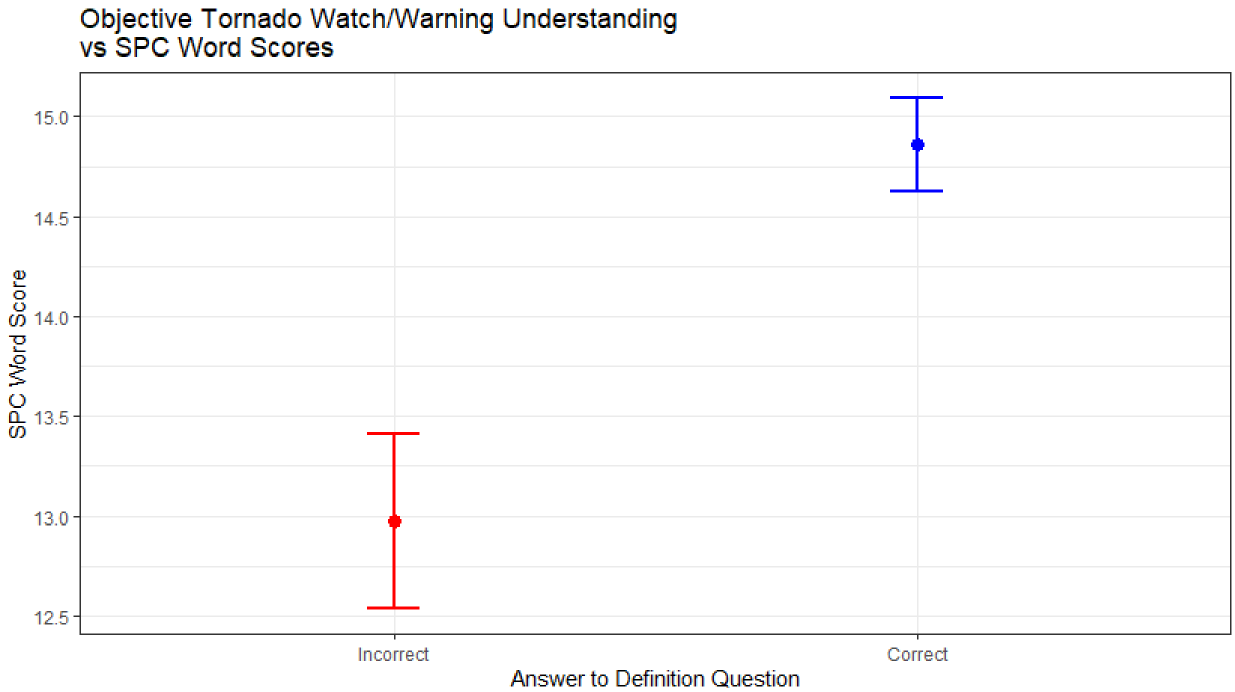


Figure 28: **Graph of the difference in average SPC word score between individuals who incorrectly and correctly define the difference between tornado watches and tornado warnings. Points display the mean value of SPC word score for individuals who incorrectly and correctly answered the tornado watch or warning prompt question, while the error bars display the 95% confidence interval around that estimate based on the standard error.**

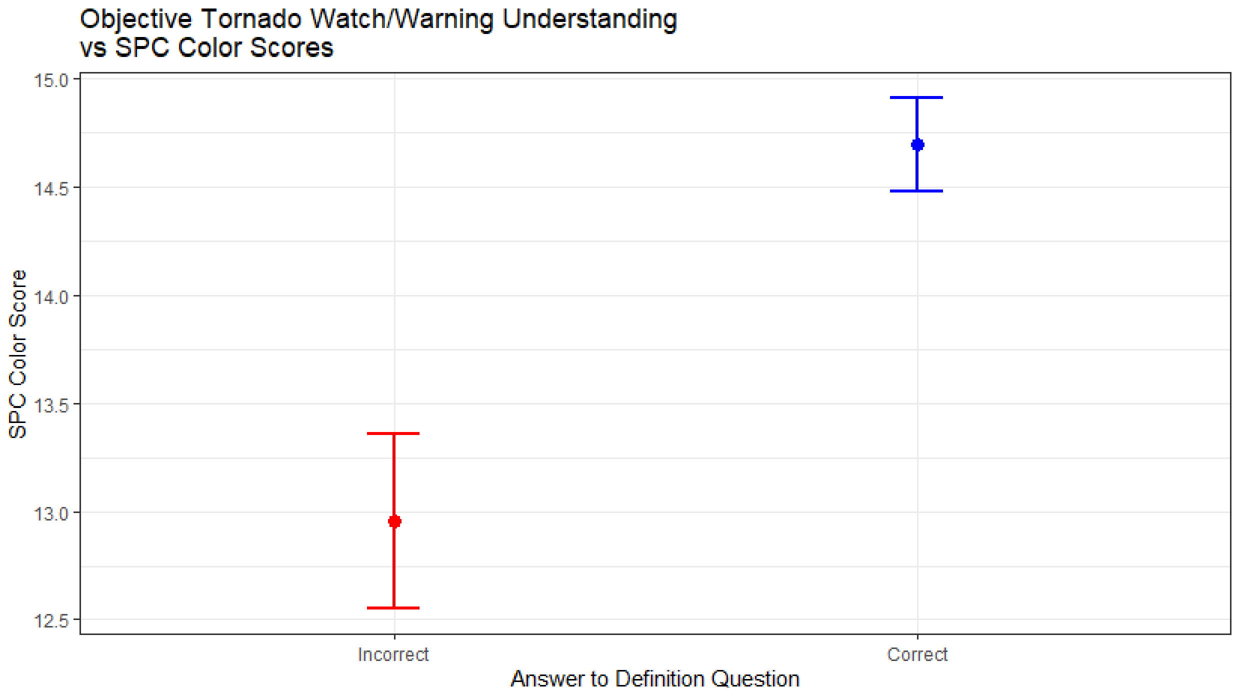


Figure 29: Graph of the difference in average SPC color score between individuals who incorrectly and correctly define the difference between tornado watches and tornado warnings. Points display the mean value of SPC word score for individuals who incorrectly and correctly answered the tornado watch or warning prompt question, while the error bars display the 95% confidence interval around that estimate based on the standard error.

Finally, for the model of SPC word scores, a statistically significant difference in scores existed between participants that could correctly identify tornado watches from tornado warnings and those who could not. Individuals that correctly answered their definition question scored much higher (SPC word score = 14.9) than those who did not (SPC word score = 13.0; Fig. 28). This nearly two-point gap is comparable in size to the other large effects found in this study, which suggests that participant's knowledge of the tornado watch and warning system is strongly related to their ability to interpret the SPC convective outlook words. Similar results were found for SPC color interpretation. Participants able to discern tornado watches from warnings correctly scored higher (SPC color score = 14.7) than those who could not (SPC color score = 13.0; Fig. 29). Combined with the SPC word model results, this suggests familiarity with the tornado communication system is related to overall SPC outlook comprehension ability.

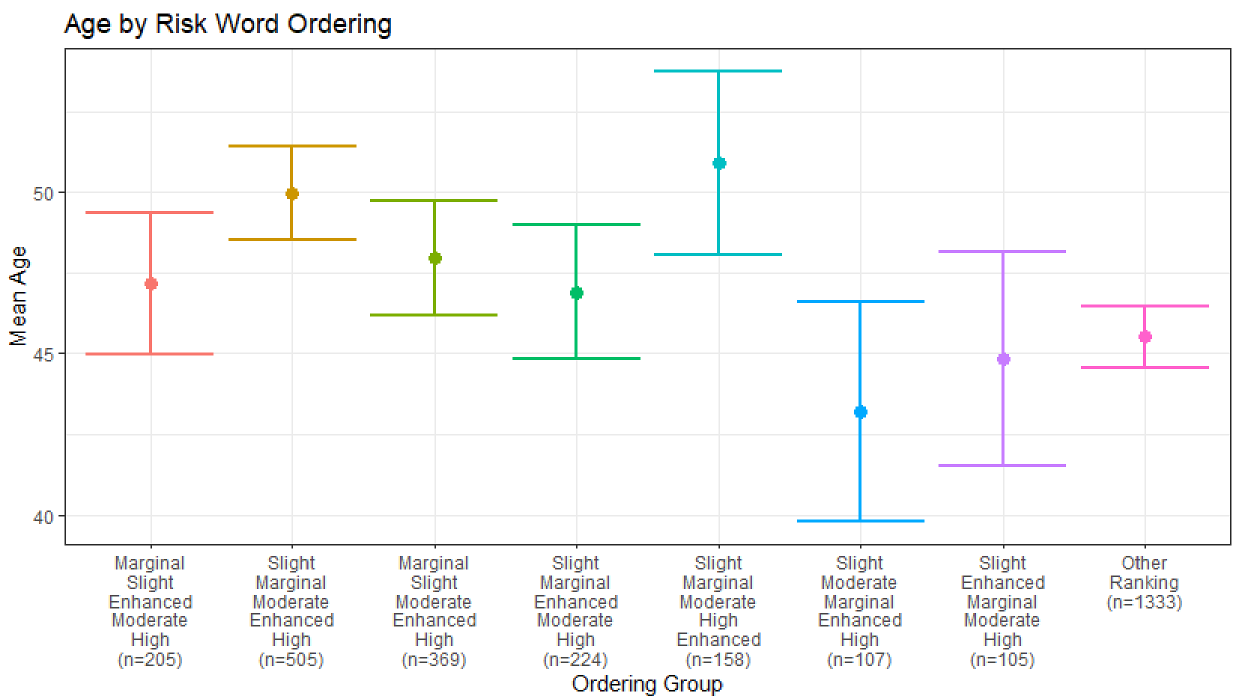


Figure 30: **Graph of the difference in average age across the different SPC word ordering groups as defined in table 5. The correct ordering group is the furthest to the left. Points display the mean age for individuals in each word order group, while the error bars display the 95% confidence interval around that estimate based on the standard error.**

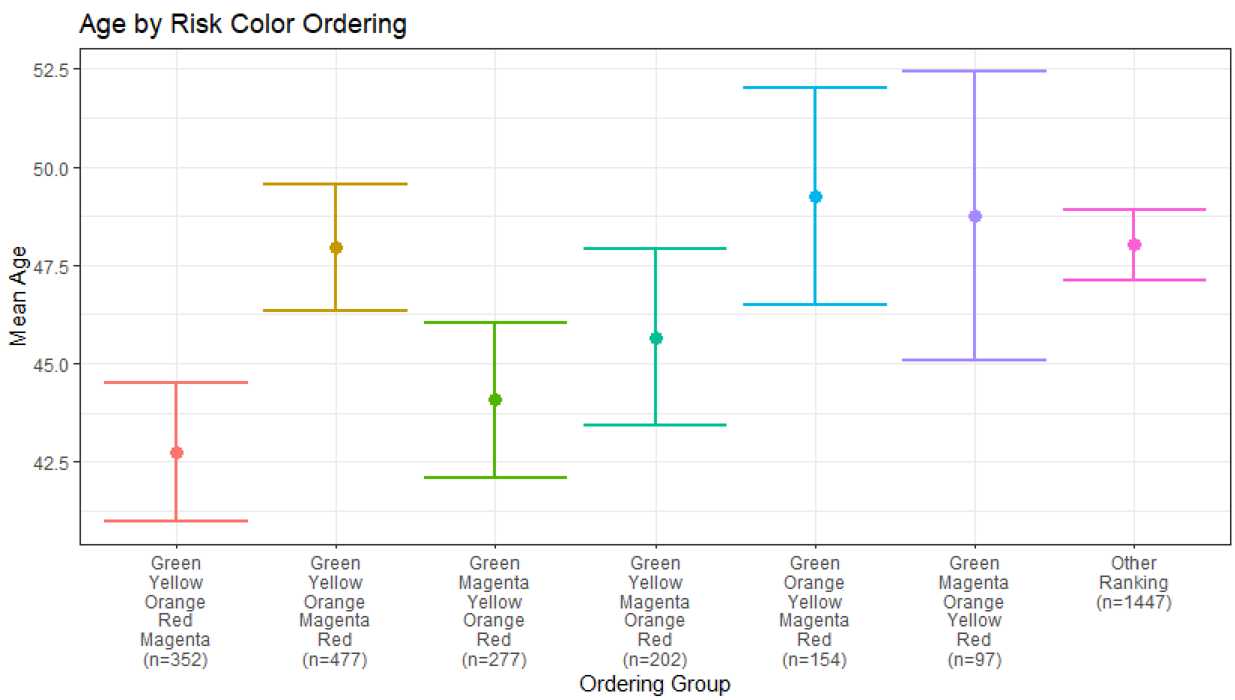


Figure 31: **Graph of the difference in average age across the different SPC color ordering groups as defined in table 6. The correct ordering group is the furthest to the left. Points display the mean age for individuals in each color order group, while the error bars display the 95% confidence interval around that estimate based on the standard error.**

In addition to the multiple regression models, I compared the average values of age and numeracy across the ordering groups (Tables 5-6) to identify differences that may exist across those groups. Little relationship was found across word and color ordering for age (Figs. 30-31), with variation in average age appearing relatively random and generally within the error bars of each distribution, but more significant differences appeared for the distributions of numeracy.

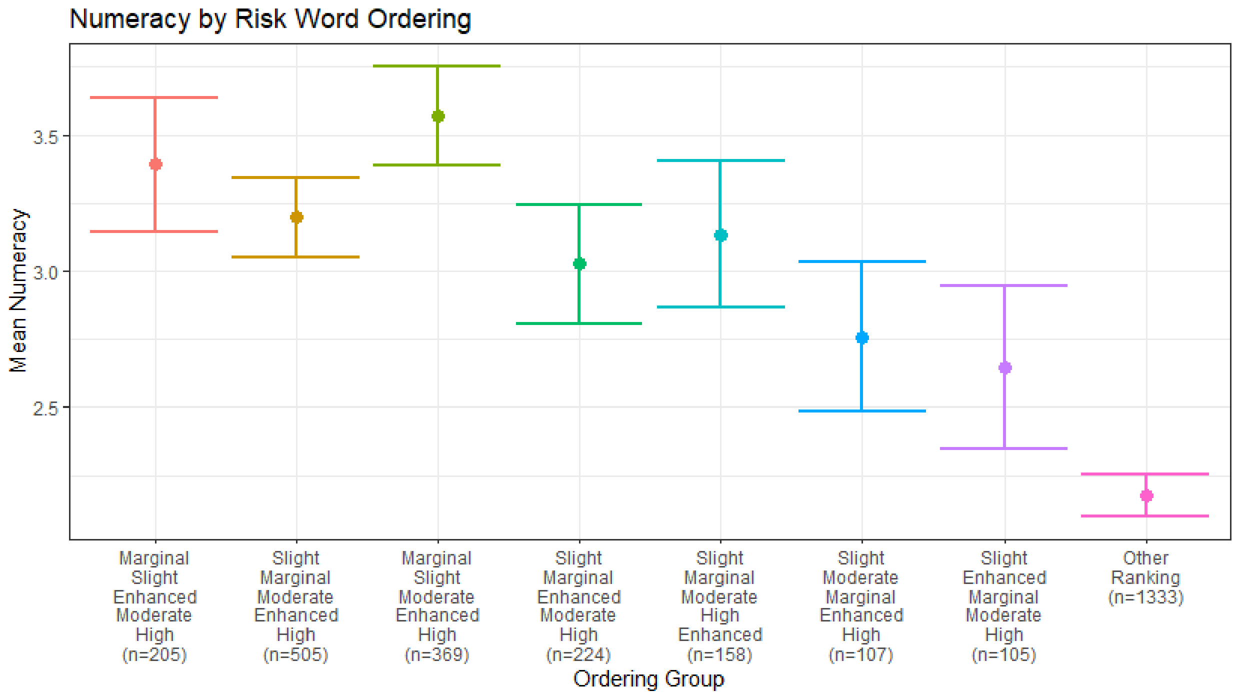


Figure 32: **Graph of the difference in average Numeracy across the different SPC word ordering groups as defined in table 5. The correct ordering group is the furthest to the left. Points display the mean numeracy for individuals in each word order group, while the error bars display the 95% confidence interval around that estimate based on the standard error.**

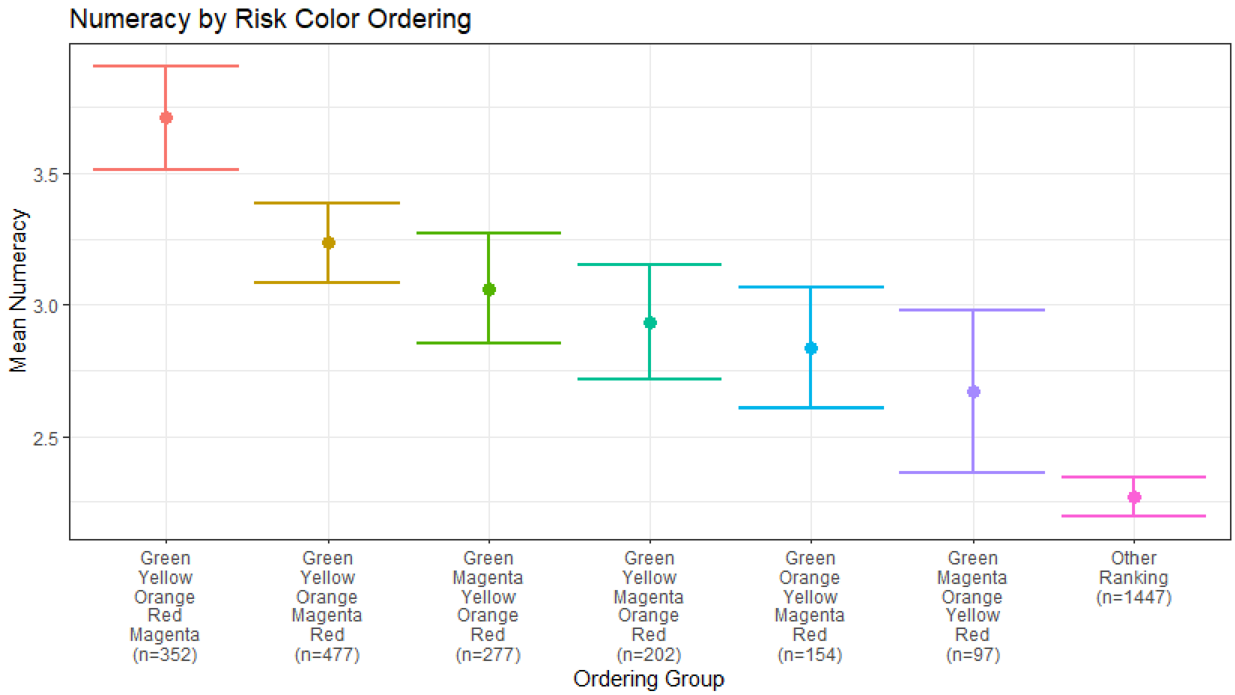


Figure 33: **Graph of the difference in average Numeracy across the different SPC color ordering groups as defined in table 6. The correct ordering group is the furthest to the left. Points display the mean numeracy for individuals in each color order group, while the error bars display the 95% confidence interval around that estimate based on the standard error.**

In general, individuals that correctly ordered the SPC words and colors, or were in the largest order groups, achieved higher numeracy scores through the BNT-S scoring measure (Fig. 32-33). This was particularly noticeable across the different color ordering groups. Participants who correctly ordered the colors also reported much higher numeracy scores than the other groups. This evidence further suggests that numeracy is significantly related to participant accuracy in the SPC word and color ordering tasks and is strongly related to SPC convective outlook interpretation.

Chapter 4 – Discussion

Review of Major Findings

Though the SPC convective outlook has existed for over 50 years, a knowledge gap exists around how users outside of governmental actors generate value from the information contained within these forecasts. Though previous work into tornado warnings and hurricane information could only suggest how broadcast meteorologists and the public to whom they present understood the product, this study shows that overall broadcasters view the outlook as a valuable tool for spotting and acting in advance of severe weather events. However, the way the outlook defines the forecast of storm coverage is flawed. As a result, broadcasters worry that the outlook will confuse their viewers, which is further reflected in the survey data. Further, some demographics have more difficulty interpreting the product when compared to White, numerate, younger individuals that know the difference between tornado watches and warnings.

Though the CTA and focus group interview formats were very different, parallel themes emerged from both. Broadcasters appreciated the SPC outlook as a general guide to where severe weather was expected, how serious it could be, and that could be quickly adapted and shared with their viewers. For both interview groups, participants reported that they used the outlooks to help guide management decisions, declare "Red Alert" days when outlooks were issued for their area, or manage their staffing to ensure more of their weather team were in the office on days when severe weather was expected. The outlook also helped both groups as a situational awareness tool to increase their confidence in their own interpretation by sharing with them the forecast thoughts of expert severe weather forecasters. These explanations point towards value generation by broadcasters, as the outlook allows them to make decisions and take actions, such as alert the public more effectively to upcoming severe, schedule enough staff to cover severe weather without overworking their teams, and improve forecast confidence.

Similar issues were revealed by the CTA and focus group interviews. Both groups struggled to explain the meaning of the SPC outlook words to their viewers and management. In the focus group interviews, broadcasters suggested that the words Slight and Moderate undersell their respective risks. These broadcasters also mentioned that they have issues with the highest tiers of the outlooks. Specifically, broadcasters suggested that Moderate and High Risks more frequently "bust" with less impactful outcomes than expected. Others explained that High Risk events were so rare for them that Enhanced or Moderate days were their top-tier severe event days. The focus groups were able to investigate these issues in a bit more detail and revealed that broadcasters felt that the current version of the outlook feels like it is not a public facing product, and that their local knowledge of meteorology would sometimes perform better than the outlook.

Though the survey data did not include questions that could investigate the accuracy of the outlook levels, the issues broadcasters identified with the outlook words certainly emerged in statistical analysis. Participants struggled to correctly order the outlook words, with the largest groups of participants switching the placement of Slight and Marginal, Moderate and Enhanced, or even both pairs of words. The participants were much more successful at ordering the outlook colors. The only group larger than the correct order group interchanging the placement of red and magenta. This lined up well with broadcaster concerns about the SPC outlook words, suggesting that confusion in their interpretation may reduce the value members of the public can generate from the outlook in its current state. Finally, the correlation table across the independent variables showed low correlation values (below 0.3) between variables like region, race, age, gender, ethnicity, and education versus objective tornado watch/warning comprehension, contrary to prior studies of warning comprehension (Powell and O'Hair 2008; Jauernic and Van den Broeke 2017; Allan et al. 2019; Ripberger et al. 2019). Future work may seek to more deeply investigate these relationships within this survey data.

Overall, my hypotheses into how broadcasters use and perceive the usefulness of the convective outlook, and how the public interprets the outlook, were supported by the study findings. There were six major hypotheses for this study:

- **H1:** Broadcasters use the SPC Outlook as a visual aid to describe severe weather threats in their shows, as well as to drive conversations with station management to prepare for severe weather television coverage.
- **H2**: Broadcasters will voice frustrations with the lack of forecast detail contained within the convective outlook, with additional information on timing and intensity most needed to improve the outlook.
- **H3:** Broadcasters will mention adding information or reinterpreting the convective outlook before presenting it to their viewers based on their concerns about what information the outlook lacks.

- **H4:** White, educated, middle-aged men living in tornado-prone regions will best comprehend the SPC outlook when asked to rank the outlook colors and words from least to greatest risk.
- **H5:** More numerate and weather salient individuals will best comprehend the SPC outlook when asked to perform the outlook ranking task.
- **H6:** Broadcaster concerns about public interpretation of the SPC outlook words and colors will be reflected in the analysis of the public survey data.

First, as suggested by research into how broadcasters are the primary conduit of weather information for the public (Pietrycha and Fox 2004, Schumacher et al. 2010, Drobot et al. 2014), broadcasters reported using the outlook to communicate severe threats and to define their coverage strategy to best inform their viewers. Consistent with Reed and Senkbeil (2019) and theorized in the second hypothesis, the lack of impact-oriented information in the outlook may lead broadcasters to report frustrations with more frequent "bust" events for the higher risk outlooks when expected severe impacts fail to materialize. This could be due to the inherent volatility of high-end severe weather events, where one or two missing atmospheric conditions can change an outcome from a damaging tornado outbreak to harmless clouds and rain. For the third hypothesis, similar to what Broad et al. (2007), Morrow et al. (2008), and Demuth et al. (2009) found with the cone of uncertainty product from the NHC, broadcasters in this study reported that they experiment with the format of the outlook to present its information to their viewers. Though their methods were varied, some common methods included converting the outlook into bar graphs by hazard, replacing the outlook words with numbers or other words, and developing their own risk graphics that covered similar areas to the outlook though with different colors and levels.

The fourth hypothesis was found to be partly true through statistical analysis, as White participants did score higher on the ranking tasks, but younger individuals had higher scores and education was not strongly correlated with SPC word and color ranking scores. Regionality but NWS Region surprisingly had no effect on participants' SPC ranking scores, which was also the opposite of what I had hypothesized. Numeracy, as postulated by the fifth hypothesis, did have a strong relationship with the ranking scores, whose correlation with education may explain why the latter variable was related to tornado warning comprehension in previous studies (Jauernic and Van den Broeke 2017; Allan et al. 2019; Ripberger et al. 2019). The other half of the fifth hypothesis, that suggested weather salience would be positively correlated with SPC ranking scores, was not supported by my analysis. Finally, broadcasters' concerns about the SPC outlook words confusing the public more than the outlook colors were validated in the public survey, supporting the sixth hypothesis of this study.

Implications/Future Research

It is important to note that broadcasters' concerns about how well viewers could interpret the standard convective outlook did match the data I collected from the public, as was expected to be the case due to their relatively close relationship. This finding further supports the idea that future research into the development of products like the SPC convective outlook should involve broadcaster input from the earliest stages of their design. Broadcasters may be able to help avoid potential communication pitfalls during the early development of prototype level identifiers or words, streamlining the public testing process to a smaller number of vetted prototypes. Future work should also determine how attributable the SPC outlook is to some of the behaviors identified by broadcasters in this study, such as increasing the amount of discussion with or changing the decisions made by station management, or the anxiety some broadcasters reported feeling before Moderate and High Risk days. It is important to identify whether these findings are a result of the outlook itself or are more related to the broadcaster's feelings based on their own forecast. Finally, broadcaster interpretations of the SPC outlook should continue be studied with public surveys and compared to public interpretation of NWS products. Broadcast meteorology provides a unique testing ground for live experimentation with the visual communication of threat information with the general public.

The results of the public survey also imply the existence of potential areas of improvement in the SPC outlook. Future research should investigate the significant gaps in comprehension of the outlook across individuals with different levels of numeracy and of different racial groups. These demographics displayed large swings in ranking accuracy that could belie social vulnerabilities in the nation's severe weather messaging paradigm. If the SPC outlook words are to be changed, public data on the interpretation of any prototype words must be collected, to ensure that these demographic vulnerabilities are addressed and that the product is usable by as much of the community as possible. Any change to the outlook should also be studied as a potential source of confusion for users, and measures that could reduce confusion during the implementation of any change should also be investigated. Furthermore, this study did not attempt to determine individuals' ability to interpret the complete graphical product produced by the SPC that has words and colors combined. Investigations in the future should study the public interpretation of the full product as well as the individual risk words and colors, to better simulate the kinds of encounters individuals can have with the outlook product.

Finally, the findings reported in this study have implications for the development process of upcoming products currently being developed at the SPC and through the FACETs initiative. First, the development of any new words or level indicators to be used in the outlook should be carefully prototyped with broadcaster input, followed by rigorous testing through surveys and focus groups with representative samples of the US public. Such a development process should ensure greater understanding of the outlook across its many user groups, from emergency managers and public sector professionals to private businesses, members of the public, and of course broadcasters. A series of prototype words or level indicators should show a sequentially increasing level of perceived risk and expected likelihood of action across the five tiers of the outlook, which would suggest accurate comprehension of the information in the forecast. Additionally, broadcasters' laments over outlook forecasts busting due to impacts not matching the level of seriousness of the issued risk level could be addressed somewhat by the upcoming forecast severity product currently being tested in the HWT (P. Marsh 2020, personal communication). These findings suggest that expanding the scope of the testing of this new product to broadcasters, as well as to the public through rigorous surveys, may improve the effectiveness of the final version of the product with users from all sectors. Overall, the SPC outlook appears to have clear value to broadcasters in the form of a decision aid but could be improved through user input and further increase its value.

Study Limitations

Though the design of the interviews and surveys performed in this study was intended to maximize the value and credibility of its findings, there are still limitations that should be addressed when interpreting these results. First, focus group interviews, CTAs, and public surveys are observational in nature, and thus cannot establish causation, only correlation between variable changes. I cannot say that the SPC outlook directly causes broadcasters to experience anxiety about higher risk days, as there are a multitude of other uncontrolled variables, such as prior severe weather events, high end event frequency, and the broadcaster's own forecast for the event that could also contribute to that anxiety and are not experimentally controlled. Similarly, it is possible that unmeasured confounding variables exist that could result in the increase in SPC outlook interpretation scores with increasing numeracy, such that numeracy is not related to the scores at all. As such, these results should be taken as suggestive of potential relationships, and do not experimentally confirm the existence of direct relationships.

Other limitations exist within the sampling method, particularly for the broadcasters interviewed here. These participants were sampled using a convenience method, where they were contacted by study organizers or heard of the research through twitter or other social media apps where study openings were announced. This means that participants may already have a more favorable view of the SPC and NWS than the population of broadcasters overall, which could sway results in a more positive direction. There are also some limitations to the thematic analysis of broadcaster's interviews, as recommended practice has multiple coders compare their coding of data to ensure impartiality of the coding scheme. Due to the combined factors of limited time and the COVID-19 crisis creating a barrier to such collaborative work, coding comparisons across multiple coders were not performed with this data. In an attempt to mitigate this issue the coded data was reviewed multiple times by the main coder, with time gaps between each session. The coding process was also journaled and recorded, to aid in potential future efforts to replicate this work.

Beyond the interviews and thematic analyses, some limitations were also present in the design of the public survey. Though the sampling strategy was stratified and administered by an acclaimed third-party survey administrator, the nature of survey data can lead to errors in responses. For example, the ordering task did not include a "do not know" option and randomly assigned each word and color to a random position in the administered survey, meaning some erroneous responses created by random generation and not participant input cannot be identified in the data. I also did not test interpretation of the full convective outlook and instead asked participants to rank the words and colors alone. Participant interpretation of the full outlook graphic may differ due to these two elements being combined and displayed in a way that may reveal their true ranking more effectively. Further, the home setting for most survey takers may reduce their focus on the survey and lead them to incorrectly read or answer survey questions. Finally, the analysis of the survey data has some flaws as well, particularly with the design of the "ranking score" used in this study. The score penalizes misplacements of the words and colors more distant from their original position more than closer misplacements (e.g. the greatest error points are earned by swapping the position of High and Marginal, while the lowest is earned by swapping any two adjacent words, like High and Moderate). It could be argued that mistaking the position of the Slight and Enhanced Risk levels could be more significant of an error than mistaking Marginal and High, as Slight and Enhanced days are more common and still have large differences in impact severity and coverage. Different scores that prioritize different errors could paint a completely different picture of public SPC outlook interpretation, so it is important to keep the limitations of this score design in mind when interpreting this study's findings.

Conclusions

In summary, the SPC convective outlook is a widely used and recognized NWS product for broadcast meteorologists, who appreciate the product and incorporate it into their forecast process. The outlook helps broadcasters plan their staffing, work with station management before severe weather events, and present what areas are under what risk of severe weather. It is not a perfect product. However, broadcasters feel that the outlook does not adequately capture the range of severe weather events that can occur in some regions, and that higher end forecasts can fail to verify in terms of notable impacts. Broadcasters also suggest that the outlook words, particularly Slight and Moderate, do not adequately convey the seriousness of the severe weather coverage that both levels are representing, and are confusing to viewers. This concern is validated in surveys of the public, who have difficulty ranking the first four words used in the outlook correctly from highest to lowest. Demographic factors also appear to impact outlook comprehension ability, as white, numerate, and younger participants who could correctly determine the difference between a tornado watch and a tornado warning had higher SPC word and color ranking scores than other groups.

Though it is tempting to assume that the solution to these issues would be to increase severe weather education for the general public, these results suggest that working with broadcast meteorologists, as well as with members of the public through surveys and interviews, would be a more cost and time-effective method of increasing SPC outlook value. This product could help to prepare individuals for severe weather well in advance of the issuance of a tornado warning by helping them spread sheltering or evacuation efforts over a greater period of time, as long as the message the product contains is well received by the public and other users at risk. Recent efforts by the SPC to improve their messaging suggest that the research needed to bring these improvements about is possible, and that the SPC can meet populations vulnerable to severe weather halfway by designing weather messaging around their needs.

References

Allan, J. N., 2018: Numeracy vs. intelligence: A model of the relationship between cognitive abilities and decision making. M.S. Thesis, Dept. of Psychology, The University of Oklahoma.

Allan, J. N., J. T. Ripberger, W. W. Wehde, M. J. Krocak, C. Silva, and H. Jenkins-Smith, 2019: Geographic distributions of extreme weather risk perceptions in the United States. University of Oklahoma Center for Risk and Crisis Management. Link: https://github.com/oucrcm/wxsurvey/blob/master/geographic%20distributions%20of%20extreme%20weather%20risk%20perceptions/manuscript.pdf.

Allan, J. N., J. T. Ripberger, V. T. Ybarra, and E. T. Cokely, 2017: The Oklahoma Warning Awareness Scale: A psychometric analysis of a brief self-report survey instrument. *Proceedings of the Human Factors and Ergonomics Society 2017 Annual Meeting*, Austin, TX, Hum. Factors, 1203-1207, DOI:10.1177/1541931213601783

Anderson, A. A., T. A. Myers and E. W. Maibach, 2013: If they like you, they learn from you: How a brief weathercaster-delivered climate education segment is moderated by viewer evaluations of the weathercaster. *Wea. Clim. Soc.*, **5**, 367-377, DOI: 10.1175/WCAS-D-12-00051.1

Ash, K. D., R. L. Schumann III, and G. C. Bowser, 2014: Tornado warning trade-offs: Evaluating choices for visually communicating risk. *Wea. Clim. Soc.*, **6**, 104-118, DOI: 10.1175/WCAS-D-13-00021.1

Bostrom, A., R. Morss, J. Demuth, H. Lazrus and J. K. Lazo, 2018: Eyeing the storm: How residents of coastal Florida see hurricane forecasts and warnings. *Int. J. Risk Reduct.*, **30(A)**, 105-119. DOI: 10.1016/j.ijdrr.2018.02.027

Bostrom, A., R. Morss, J. K. Lazo, J. L. Demuth, H. Lazrus and R. Hudson, 2016: A mental models study of hurricane forecast and warning production, communication, and decision-making. *Wea. Clim. Soc.,* **8**, 111-129. DOI: 10.1175/WCAS-D-15-0033.1

Boyatzis, R. E. 1998: *Transforming Qualitative Information: Thematic analysis and code development.* Sage Publications, Inc., 184 pp.

Braun, V. and V. Clarke, 2006: Using thematic analysis in psychology. *Qual. Res. Psychol.*, **3(2)**, 77-101. https://doi.org/10.1191/1478088706qp063oa

Breen, R. L., 2006: A practical guide to focus-group research. *J. Geog. Higher Edu.*, **30(3)**, 463-475, DOI: 10.1080/03098260600927575

Broad, K., A. Leiserowitz, J. Weinkle, and M. Steketee, 2007: Misinterpretations of the "Cone of Uncertainty" in Florida during the 2004 hurricane season. *Bull. Amer. Meteor. Soc.*, **88(5)**, 651-668. DOI: 10.1175/BAMS-88-5-65I

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. [Accessed online at: https://www.nssl.noaa.gov/users/brooks/public_html/papers/prague2k1.pdf]

Bryant, B., M. Holiner, R. Kroot, K. Sherman-Morris, W. B. Smylie, L. Stryjewski, M. Thomas, and C. I. Williams, 2014: Usage of color scales on radar maps. *J. Operational Meteor.,* **2**, 169-179, DOI: http://dx.doi.org/10.15191/nwajom.2014.0214

Cappucci, Matthew, 2020: The National Weather Service issues highly accurate thunderstorm forecasts. The public doesn't understand them. *The Washington Post*, Accessed 28 July 2020, https://www.washingtonpost.com/weather/2020/06/10/storm-prediction-center-risk-categories/

Center for the Promotion of Health in the New England Workplace (CPH-NEW), 2020: Tips for facilitating focus groups. Accessed 4 April 2020, https://www.uml.edu/docs/FG%20Tips%20sheet_RK_tcm18-167588.pdf

Cokely, E. T., A. Feltz, S. Ghazal, J. N. Allan, D. Petrova, and R. Garcia-Retamero, 2018: Skilled Decision Theory: From intelligence to numeracy and expertise. *The Cambridge Handbook of Expertise and Expert Performance.* K. A. Ericsson, R. R. Hoffman, A. Kozbelt, and A. M. Williams, Ed., Cambridge University Press, 476-505.

Cokely, E. T., M. Galesic, E. Schulz, S. Ghazal, and R. Garcia-Retamero, 2012: Measuring risk literacy: The Berlin Numeracy Test. *Judgem. Decis. Mak.,* **7(1),** 25-47.

Corfidi, S. F., 1999: The birth and early years of the Storm Prediction Center. *Wea. Forecasting*, **14**, 507-525, https://doi.org/10.1175/1520-0434(1999)014%3C0507:TBAEYO%3E2.0.CO;2

Daniels, G. L. and G. M. Loggins, 2007: Conceptualizing continuous coverage: A strategic model for wall-to-wall local television weather broadcasts. *J. Appl. Comm. Res.*, **35(1)**, 48-66, DOI: 10.1080/00909880601065680

Demuth, J. L., B. H. Morrow, and J. K. Lazo, 2009: Weather forecast uncertainty information: An exploratory study with broadcast meteorologists. *Bull. Amer. Meteor. Soc.*, **90(11)**, 1133-1146. DOI: 10.1175/2009BAMS2787.1

Demuth, J. L., R. E. Morss, B. H. Morrow and J. K. Lazo, 2012: Creation and communication of hurricane risk information. *Bull. Amer. Meteor. Soc.*, **93(8)**, 1614-1618. DOI: 10.1175/BAMS-D-11-00150.1

Drobot, S., A. R. S. Anderson, C. Burghardt, and P. Pisano, 2014: U.S. public preferences for weather and road condition information. *Bull. Amer. Meteor. Soc.,* **95(6)**, 849-860, DOI: 10.1175/BAMS-D-12-00112.1

Drost, R., M. Casteel, J. Libarkin, S. Thomas, and M. Meister, 2016: Severe weather warning communication: Factors impacting audience attention and retention of information during tornado warnings. *Wea. Clim. Soc.*, **8**. DOI: 10.1175/WCAS-D-15-0035.1

Edwards, R. and F. Ostby, 2015: Time line of SELS and SPC. Storm Prediction Center, accessed Feb 28 2020, https://www.spc.noaa.gov/history/timeline.html

Ernst, S., D. LaDue, and A. Gerard, 2018: Understanding emergency manager forecast use in severe weather events. *J. Operational Meteor.,* **6(9)**, 95-105, DOI: https://doi.org/10.15191/nwajom.2018.0609

Fereday, J. and E. Muir-Cochrane, 2006: Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *Int. J. Qual. Methods*, **5(1)**, 80-92.

Finley, J. P., 1884: Tornado Predictions. *Amer. Meteor. J.,* **1**, 84-88.

Grams, J., B. Bunting, and S. Weiss, 2014: SPC Convective Outlooks. Storm Prediction Center, April 9th, 2019, DOI: https://www.spc.noaa.gov/misc/SPC_probotlk_info.html

Hahn, B. B., E. Rall, and D. W. Klinger, 2003: Cognitive task analysis of the warning forecaster task. Order No. RA1330-02-SE-0280, 28pp. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.536.9648&rep=rep1&type=pdf

Hammer, B. and T. W. Schmidlin, 2002: Response to warnings during the 3 May 1999 Oklahoma City tornado: Reasons and relative injury rates. *Wea. Forecasting*, **17**, 577-581, https://doi.org/10.1175/1520-0434(2002)017%3C0577:RTWDTM%3E2.0.CO;2

Henson, R., 2010: *Weather on the Air: A History of Broadcast Meteorology.* American Meteorological Society, 241pp.

Herman, G. R., E. R. Nielsen, and R. S. Schumacher, 2018: Probabilistic verification of Storm Prediction Center convective outlooks. *Wea. Forecasting*, **33**, 161-184, DOI: 10.1175/WAF-D-17-0104.1

Hitchens, N. M. and H. E. Brooks, 2012: Evaluation of the Storm Prediction Center's day 1 convective outlooks. *Wea. Forecasting*, **27**, 1580-1585, DOI: 10.1175/WAF-D-12-00061.1

Hitchens, N. M., H. E. Brooks, and M. P. Kay, 2013: Objective Limits on Forecasting Skill of Rare Events. *Wea. Forecasting*, **28**, 525-534, DOI: 10.1175/WAF-D-12-00113.1

Hitchens, N. M. and H. E. Brooks, 2014: Evaluation of the Storm Prediction Center's convective outlooks from day 3 through day 1. *Wea. Forecasting*, **29**, 1580-1585, DOI: 10.1175/WAF-D-13-00132.1

Hitchens, N. M. and H. E. Brooks, 2017: Determining criteria for missed events to evaluate significant severe convective outlooks. *Wea. Forecasting*, **32**, 1321-1328, DOI: 10.1175/WAF-D-16-0170.1

Hoffman, R. R., 2005: Protocols for Cognitive Task Analysis. Florida Institute for Human and Machine Cognition, 109pp. [accessible online at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1021.7915&rep=rep1&type=pdf]

Imy, D. and R. Edwards, 2013: SPC severe weather outlooks. Storm Prediction Center, accessed Feb 28 2020, https://www.spc.noaa.gov/misc/SPC_3-tier_probotlk_info.html

Jauernic, S.T. and M. S. Van Den Broeke, 2017: Perceptions of tornadoes, tornado risk, and tornado safety actions and their effects on warning response among Nebraska undergraduates. *Nat. Hazards*, **80**, 329-350, DOI: https://doi.org/10.1175/WCAS-D-16-0031.1

Jenkins-Smith, H., J. Ripberger, G. Copeland, M. Nowlin, T. Hughes, A. Fister, W. Wehde, and J. Davis, 2017: *Quantitative research methods for political science, public policy, and public administration: 4th edition with applications in R*. 177pp. [accessible online at https://bookdown.org/ripberjt/qrmbook/]

Keul, A. G. and A. M. Holzer, 2013: The relevance and legibility of radio/TV weather reports to the Austrian public. *Atmos. Res.*, **122**, 32-42. http://dx.doi.org/10.1016/j.atmosres.2012.10.023

Kirschenbaum, S. S., 2003: Comparative cognitive task analysis: The cognition of weather forecasting. Proceedings, *Human Factors and Ergonomics Society 47th Annual Meeting,* **47**, 473-477, https://doi.org/10.1177%2F154193120304700347

Krueger, R. A., 2002: Designing and conducting focus group interviews. Accessed 9 April 2020, https://www.eiu.edu/ihec/Krueger-FocusGroupInterviews.pdf

Lazo, J. K. and L. G. Chestnut, 2002: Economic value of current and improved weather forecasts in the U.S. household sector. Report to the NOAA office of Policy and Planning, U.S. National Oceanic and Atmospheric Administration.

Lazo, J. K., R. E. Morss, and J. L. Demuth, 2009: 300 billion served: Sources, perceptions, uses, and values of weather forecasts. *Bull. Amer. Meteor. Soc.,* **90(6)**, 785-798 DOI: 10.1175/2008BAMS2604.1

Leiserowitz, A. A., E. W. Maibach, C. Roser-Renouf, G. Feinberg, and P. Howe, 2012: Climate change in the American mind: American's global warming beliefs and attitudes in September 2012. Yale Project on Climate Change Communication Rep.m 31 pp.

[Available online at https://environment.yale.edu/climate-communication-OFF/files/Climate-Beliefs-September-2012.pdf]

Lindell, M. K., and R. W. Perry, 2012: The protective action decision model: Theoretical modifications and additional evidence. *Risk Anal.*, **32**, 616–632, https://doi.org/10.1111/j.1539-6924.2011.01647.x.

Lipkus, I. M. and J. G. Hollands, 1999: The visual communication of risk. *J. Natl. Cancer Inst.,* **25**, 149-163

Mason, J. B. and J. C. Senkbeil, 2015: A tornado watch scale to improve public response. Wea. Climate Soc., 7, 146–158, https://doi.org/10.1175/WCAS-D-14-00035.1.

Maddox, R. A. and C. A. Crisp, 1999: The Tinker AFB Tornadoes of March 1948. *Wea. Forecasting*, **14**, 492-499, https://doi.org/10.1175/1520-0434(1999)014<0492:TTATOM>2.0.CO;2

McCarthy, D. and K. P. Tarp, 2020: Covering the Storm: Broadcasting from the NWS Storm Prediction Center and other National centers. Accessed at https://www.spc.noaa.gov/publications/mccarthy/nwa-spc.pdf

Meiners, T. and C. M. Lusuriello, 2015: Millennials in Broadcast Meteorology. *43rd conf. on Broadcast Meteorology*, Raleigh, NC, Amer. Meteor. Soc., J3.5, https://ams.confex.com/ams/43BC3WxWarn/webprogram/Paper272685.html

Militello, L., W. Wong, S. Kirschenbaum, and E. Patterson, 2011: Systematizing discovery in Cognitive Task Analysis. *Informed by knowledge: Expert performance in complex situations*. K. L. Mosier and U. M. Fischer, Ed. Psychology Press, Taylor and Francis Group, 287-303.

Miran, S. M., C. Ling, L. Rothfusz, 2018: Factors influencing people's decision-making during three consecutive tornado events. *Int. J. Disaster Risk Reduct.*, **28**, 150-157, DOI: https://doi.org/10.1016/j.ijdrr.2018.02.034

Mohammed, I. A., 2012: Thematic analysis: A critical review of its process and evaluation. *W. East J. Soc. Sci.,* **1(1)**, 39-47.

Morrow, B. H., J. K. Lazo, J. Rohme, and J. Feyen, 2015: Improving storm surge risk communication: Stakeholder perspectives. *Bull. Amer. Meteor. Soc.*, **96(1)**, 35-48, DOI:10.1175/BAMS-D-13-00197.1

Morrow, B. H., J. L. Demuth, and J. K. Lazo, 2008: Communicating weather forecast uncertainty: An exploratory study with broadcast meteorologists. *36th conf. on Broadcast Meteorology*, Denver, CO, Amer. Meteor. Soc., https://www.researchgate.net/publication/280611657_Communicating_weather_forecast_uncertainty_An_exploratory_study_with_broadcast_meteorologists

Morss, R. E., J. L. Demuth, A. Bostrom, J. K. Lazo, and H. Lazrus, 2015: Flash flood risks and warning decisions: A mental models study of forecasters, public officials, and media broadcasters in Boulder, Colorado. *Risk Anal.,* **35(11)**, 2009-2028, DOI: 10.1111/risa.12403

Murphy, A. H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281-293, DOI: https://doi.org/10.1175/1520-0434(1993)008%3C0281:WIAGFA%3E2.0.CO;2

National Centers for Environmental Information (NCDC), 2020: Billion-dollar weather and climate disasters: Time series. NOAA, Accessed 28 July 2020, https://www.ncdc.noaa.gov/billions/time-series/US

National Oceanographic and Atmospheric Administration (NOAA), 2011: NWS Central Region service assessment: Joplin, Missouri, tornado. National Weather Service, 35 pp. [available online at https://www.weather.gov/media/publications/assessments/Joplin_tornado.pdf]

National Oceanographic and Atmospheric Administration (NOAA), 2014: NWS Central Region service assessment: May 2013 Oklahoma tornadoes and flash flooding. National Weather Service, 35 pp. [available online at https://www.weather.gov/media/publications/assessments/13oklahoma_tornadoes.pdf]

Nowell, L. S., J. M. Norris, D. E. White, and N. J. Moules, 2017: Thematic analysis: Striving to meet the trustworthiness criteria. *Int. J. Qual. Methods*, **16**, 1-13. DOI: 10.1177/1609406917733847

NWS, 2020a: The May 20, 2013 Newcastle – South Oklahoma City – Moore EF-5 tornado. Accessed 6 March 2020, https://www.weather.gov/oun/events-20130520

NWS, 2020b: The May 31-June 1, 2013 tornado and flash flooding event. Accessed 6 March 2020, https://www.weather.gov/oun/events-20130531

Pennesi, K., 2007: Improving forecast communication: Linguistic and cultural concerns. *Bull. Amer. Meteor. Soc.*, **88(7)**, 1033-1044, DOI:10.1175/BAMS-88-7-I033

Pietrycha, A. E. and M. A. Fox, 2004: Effective use of various communication methods during a severe convective outbreak. *NWA Digest,* **28**, 59-64. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.176.7267&rep=rep1&type=pdf

Powell, S. W., and H. D. O'Hair, 2008: Communicating weather information to the public: People's reactions and understandings of weather information and terminology. Third Symp. on Policy and Socio-Economic Research, New Orleans, LA, Amer. Meteor. Soc., https://ams.confex.com/ams/88Annual/webprogram/Paper132939.html.

Reed, J. R. and J. C. Senkbeil, 2019: Perception and comprehension of the extended forecast graphic: A survey of broadcast meteorologists and the general public. *Bull. Amer. Meteor. Soc.* DOI: 10.1175/BAMS-D-19-0078.1

Ripberger, J. T., C. L. Silva, H. C. Jenkins-Smith, J. Allan, M. Krocak, W. Wehde, and S. Ernst, 2020: Exploring community differences in tornado warning reception, comprehension, and response across the United States. *Bull. Amer. Meteor. Soc.*, DOI: 10.1175/BAMS-D-19-0064.1

Ripberger, J. T., C. L. Silva, H. C. Jenkins-Smith, and M. James, 2015: The influence of consequence-based messages on public responses to tornado warnings. *Bull. Amer. Meteor. Soc.*, **96(4)**, 577-590, DOI:10.1175/BAMS-D-13-00213.1

Ripberger, J. T., M. J. Krocak, W. W. Wehde, J. N. Allan, C. Silva, and H. Jenkins-Smth, 2019: Measuring tornado warning reception, comprehension, and response in the United States. *Wea. Climate Soc.,* **11**, 863 – 880, DOI: 10.1175/WCAS-D-19-0015.1

Rothfusz, L. P., R. Schneider, D. Novak, K. Klockow-McClain, A. E. Gerard, C. Karstens, G. J. Stumpf, and T. M. Smith, 2018: FACETs: A proposed next generation paradigm for high-impact weather forecasting. *Bull. Amer. Meteor. Soc.*, **99(10)**, 2025-2043, https://doi.org/10.1175/BAMS-D-16-0100.1

Sandlin, L., 2013: The unfriendly sky. *Storm kings: The untold history of America's first tornado chasers.* Pantheon Books, 217–238.

Schultz, D. M., E. C. Gruntfest, M. H. Hayden, C. C. Benight, S. Drobot, and L. R. Barnes, 2010: Decision making by Austin, Texas, residents in hypothetical tornado scenarios. Wea. Climate Soc., 2, 249–254, https://doi.org/10.1175/2010WCAS1067.1.

Schumacher, A. B., J. Braun, S. D. Miller and J. L. Demuth, 2010: Multidisciplinary analysis of an unusual tornado: Meteorology, climatology, and the communication and interpretation of warnings. *Wea. Forecasting*, **25**, 1412-1429. DOI: 10.1175/2010WAF2222396.1

Schutz, W. C., 1958: FIRO: A three-dimensional theory of interpersonal behavior. Rinehart.

Schwartz, L. M., S. Woloshin, W. C. Black, and H. G. Welch, 1997: The role of numeracy in understanding the benefit of screening mammography. *Ann. Intern. Med.*, *127(11)*, 966-972.

Sherman-Morris, K., 2005: Tornadoes, television, and trust – A closer look at the influence of the local weathercaster during severe weather. *Environmental Hazards*, **6(4)**, 201-210, DOI: 10.1016/j.hazards.2006.10.002

Silva, C., , J. T. Ripberger, H. Jenkins-Smith, M. Krocak, S. Ernst, and A. Bell, 2019: Establishing a baseline: Public reception, understanding, and responses to severe weather forecasts and warnings in the contiguous United States. Reference report, 33pp.

Stelter, B., 2019: Weatherman who defied 'Code Red' alerts is out of a job. CNN, Accessed 24 July 2020, https://www.cnn.com/2019/06/13/media/joe-crain-sinclair-code-red/index.html

Suffern, P., K. Harding, V. Brown, J. Keeney, K. Stammer, J. Stefkovich, 2013: May 2013 Oklahoma tornadoes and flash flooding. Service Assessment, 63pp. https://www.weather.gov/media/publications/assessments/13oklahoma_tornadoes.pdf

Storm Prediction Center (SPC), 2020: Feb-15-2003 1300 UTC Day 1 convective outlook. Accessed Feb 28 2020, https://www.spc.noaa.gov/exper/archive/event.php?date=20030215

Stewart, A. E., 2009: Minding the weather: The measurement of weather salience. *Bull. Amer. Meteor. Soc.*, **90(12)**, 1833-1842. DOI: 10.1175/2009BAMS2794.1

Stewart, A. E., J. K. Lazo, R. E. Morss, and J. L. Demuth, 2012: The relationship of weather salience with the perceptions and uses of weather information in a nationwide sample of the United States. *Wea. Clim. Soc.*, **4,** 172 – 189, DOI: 10.1175/WCAS-D-11-00033.1

Stough, S., E. M. Leitman, J. L. Peters, and J. Correia Jr., 2012: The role of the Storm Prediction Center products in decision making leading up to severe weather events. 18pp. https://www.spc.noaa.gov/publications/leitman/stough.pdf

Trujillo, J. E., O. Bermudez, P. T. Marsh, and E. M. Leitman, 2020: The Storm Prediction Center Spanish Language Initiative. *8th Symposium on the Weather, Water, and Climate Enterprise*, Boston, MA, Amer. Meteor. Soc., J5.6, https://ams.confex.com/ams/2020Annual/meetingapp.cgi/Paper/364505

Weather Prediction Center (WPC), 2020: Excessive rainfall forecasts. Accessed Feb 28 2020, https://www.wpc.ncep.noaa.gov/qpf/excess_rain.shtml

Williams, C. A., A. J. Grundstein, and J. So, 2020: Should severe weather graphics wear uniforms? Understanding the effects of inconsistent convective outlook graphics on members of the public. *15th Symposium on Societal Applications: Policy, Research, and Practice.*, Boston, MA, Amer. Meteor. Soc., 11A.1, https://ams.confex.com/ams/2020Annual/meetingapp.cgi/Paper/365011

Williams, C. A., P. W. Miller, A. W. Black, and J. A. Knox, 2017: Throwing caution to the wind: National weather service wind products as perceived by a weather-salient sample. *J. Operational Meteor.*, **5(9)**, 103-120, DOI: https://doi.org/10.15191/nwajom.2017.0509

Wilson, K., 2008: Television weathercasters as science communicators. *Public Understand. Sci.,* **17**, 73-87. DOI: 10.1177/0963662506065557

Zhao, Xiaoquan, E. Maibach, J. Gandy, J. Witte, H. Cullen, B. A. Klinger, K. E. Rowan, J. Witte, and A. Pyle, 2014: Climate change education through TV weathercasts. *Bull. Amer. Meteor. Soc.*, **95(1)**. DOI:10.1175/BAMS-D-12-00144.1

Appendix A: Blank CTA interview guide

Cognitive Task Analysis & Severe Weather Coverage Questionnaire

2019 HWT PHI Project with TV

Instructions
Describe your staffing situation and normal procedures for days with a high/moderate risk of severe convective weather, slight risk of severe convective weather, and no risk of severe convective weather. Please include any prep work done before arriving at the station, any hair/makeup routines, taping daily promotional material, calling in extra staff, etc.

Examples:

8am	*Pre-work prep*	*Wake up, check models, send out Facebook post*	*5m*
10am	*Appearance*	*Wake up, shower, blow-dry hair, eat, etc.*	*3h 30m*
1:30pm		*Arrive at station, post on Facebook, start forecasting*	*1h 30m*
3pm		*Input forecast numbers into automated graphics*	*15m*
3:45pm		*Make show*	*30m*

No Risk Day

Extra personnel you would call in or request:

Use the tab key to advance to the next cell; the table will automatically build new rows.

Time	Activity/Task	Description of Activity/Task	Length of Time

Slight Risk Day (e.g., few storms, but those that form could be bad)

Extra personnel you would call in or request:

Use the tab key to advance to the next cell; the table will automatically build new rows.

Time	Activity/Task	Description of Activity/Task	Length of Time

Moderate/High Risk Day (e.g., many storms, many could be bad)

Extra personnel you would call in or request:

Use the tab key to advance to the next cell; the table will automatically build new rows.

Time	Activity/Task	Description of Activity/Task	Length of Time

Appendix B: Focus group interview guide

- Welcome

Welcome to the focus group, and thank you so much for attending! We are really excited to have this discussion on your use of SPC convective outlooks, and how these outlooks let you generate value for your viewers. For many years, research into the goodness of SPC outlooks has focused on the quantitative accuracy and quality of these forecasts. This study is one of the first attempts to understand the value (the positive outcomes that forecast users are able to generate from a forecast) of SPC outlooks. We want to hear about your experiences and attitudes about these outlooks, and how they change the way you present your forecasts to the public. As we realize that this information can be sensitive, your identity will not be disclosed to anyone outside the research team, to protect your ability to fully express your views. I will be asking questions that focus on how SPC outlooks impact your workflow, how you take the information in these outlooks and repackage it for your viewers, and what you are satisfied or dissatisfied about with respect to these outlooks. The goal of these questions will be to stimulate discussion between you and the other members of the focus group, while I will be tracking the time spent discussing each question and making sure we are able to discuss all of these topics. I will be able to repeat questions, and clarify parts of the question, but I will aim to contribute to the discussion as little as possible. I will also be recording this discussion, so please speak clearly and remember that body language such as nodding will not be recorded, so vocalization of your opinions is critical to reviewing the discussion later. We also ask that the discussion prioritize avoiding interrupting other group members, as it can be difficult to interpret those parts of the discussion later if multiple voices are overlaid in the recording. We will now test the recording equipment by introducing ourselves, and begin an overview of the main topic.

- Topic Overview

The first convective outlooks were issued in 1955 by the Severe Local Storm Warning Service (SELS), evolving over the decades into the modern 1-8 day Convective Outlook product issued by the Storm Prediction Center (SPC, Corfidi 1999, Grams et al. 2014). These outlooks, issued by 0400 CST for the 4-8 day period, 1200 UTC for the day 3 period, 0600 and 1730 UTC for the day 2 period, and 0600, 1300, 1630, 2000, and 0100 UTC for the day 1 period, are designed to forecast the expected coverage of severe weather for the forecast period (Hitchens and Brooks 2014, Grams et al. 2014). The information contained in these outlooks is designed to increase as a severe weather event approaches, starting with the day 4-8 outlooks, which contain only threshold areas for 15 and 30% chances of severe weather occurring within 25 miles of a point. The day 3 outlooks add thresholds of 5%, 15%, 30%, and 45%, as well as the 10% or greater chance of significant severe within 25 miles of a point threshold (defined as tornadoes of EF2 strength or greater, hail larger than 2 inches in diameter, or winds in excess of 74 mph). Day 2 outlooks add a 60% threshold, and the day 1 outlooks break down the thresholds for the different types of severe weather, keeping the day 2 thresholds for hail and wind while adding 2% and 10% thresholds for tornadoes. These probabilities are broken down into categorical risk areas, titled Marginal, Slight, Enhanced, Moderate, and High, which convey increasing threat of severe weather occurring within 25 miles of a point contained within these areas (fig.1, Grams et al. 2014).

Since these outlooks can convey information that Emergency Managers, Broadcast Meteorologists, and the general public can use to protect their lives and property, multiple studies have made efforts to define and improve the goodness of these outlook products. Hitchens and Brooks (2012) initially found that day 1 categorical outlook areas have decreased over time, and have become better placed, thus reducing the number of false alarms while maintaining consistent probability of detection. Hitchens et al. (2013) further identified that SPC forecaster skill had grown over the years, which explains the increases in outlook precision and accuracy after the mid-1990s; while Hitchens and Brooks (2014) identified similar improvements in the day 2 and 3 outlooks to those that occurred in the day 1 outlooks. These studies combined suggest that SPC forecast quality has made great strides in the past three decades, and that convective outlooks, from both the categorical and probabilistic standpoint, have become skillful forecasts of severe weather events.

However, forecast quality is only one type of forecast goodness. Murphy (1993) also discusses forecast value, or the economic and social benefit of decisions made using forecast information, as another valuable dimension of goodness. There are not currently any published studies of the forecast value of SPC convective outlooks, though these outlooks have begun to be shared widely by TV news organizations like The Weather Channel and the ABC News network. Broadcasters have also been identified as the number one source of weather information, and are often the trusted face that members of the public turn to when weather threats occur (Lazo et al. 2009, Suffern et al. 2013, Morss et al. 2015). Though this information is being shared with the general public's preferred information source, recent work has found that increased collaboration between broadcasters and the NWS is needed to improve the communication of weather threats (Morss et al. 2015). This focus group discussion seeks to help us better understand the value that broadcast meteorologists working for news organizations create through the use of SPC convective outlooks, by investigating the ways that broadcasters receive and use the outlook products. This discussion will thus act as a means of evaluating the performance of outlooks from a different perspective of forecast goodness.

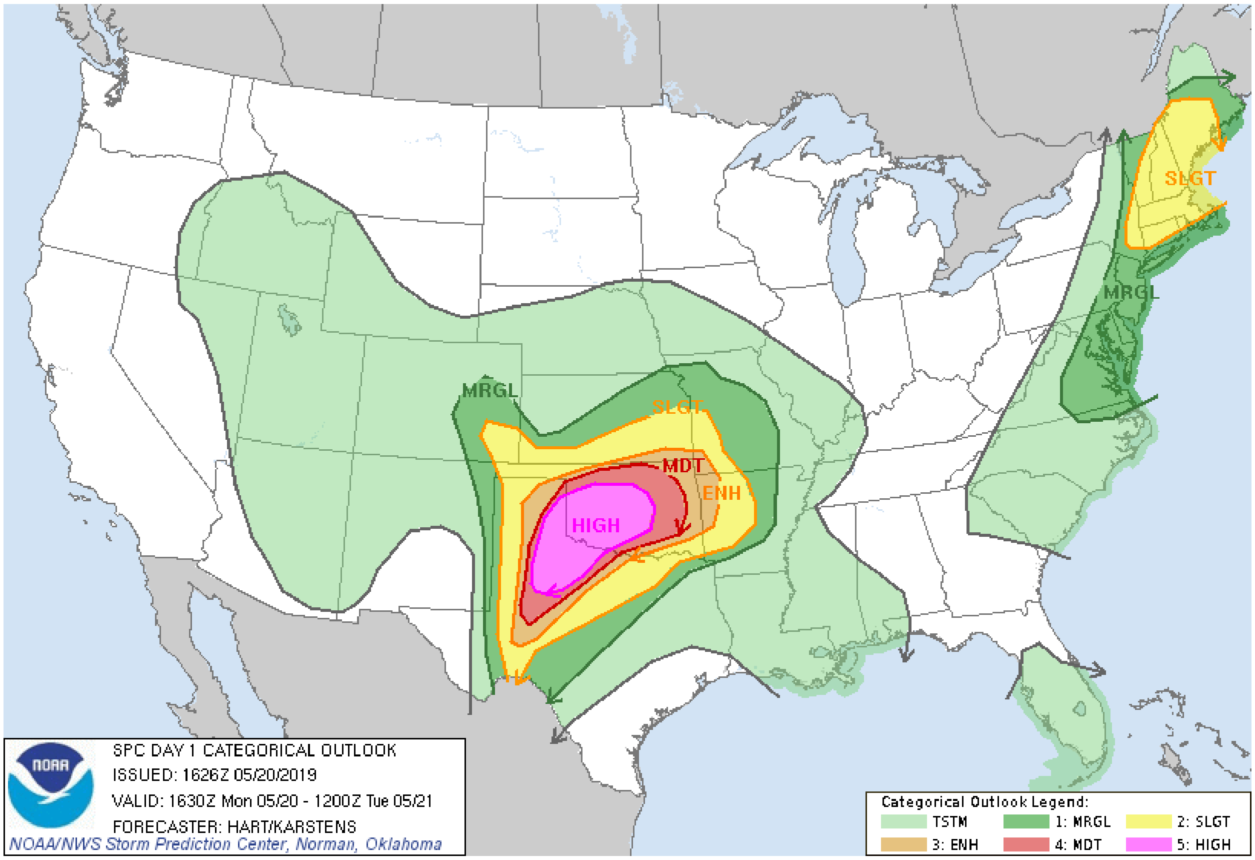


Figure 1: Example of a Day 1 SPC Convective Outlook, issued May 20th, 2019.

- **Statement of Ground Rules**

We would like to take this moment to remind everyone of the ground rules of this discussion, so that we can foster a constructive environment for discussion.

- Participation in the focus group is voluntary.
  - If any topic makes you uncomfortable, you can abstain from the discussion.
- There are no right or wrong answers in this discussion - only differing points of view.
- Please be respectful of other's positions, even if you do not agree with them.
- Listen carefully when others are speaking, and speak one at a time so that the voice recording is as clear as possible.
- We will use a first name basis in this conversation.
- Please try to stay on topic during the conversation, I will act as a moderator to ensure that we do not stray from the questions though the rest of the discussion will be between participants.
- Please try to silence communications devices during the conversation (e.g. cellphones, laptops, etc)
- Please do not discuss the details of the focus group conversation with others outside this group, to protect the privacy of your fellow participants.

- **Assurance of Confidentiality**

I would like to take a moment to reiterate the importance of confidentiality in this research, and how your personal information will be protected throughout the research process. As part of the process for applying to join this conversation, you were asked to complete a consent form. This form explained that there are not risks or benefits to you from his research, and that your voluntary participation is required. Your responses will also be kept confidential, though you can choose to withdraw your data at any time by contacting us.

- ## Questions for Discussion (have reflection, use of examples, drawings, ratings scales)

Opening Q (a round robin question)

- Can you introduce yourself and tell us where you are from, how long you have worked as a broadcast meteorologist, and a brief story of how you got into weather? (round robin)

Introductory questions

- What SPC products do you currently use? How do you access SPC products?
- Are SPC convective outlooks a useful product to you? What do these outlooks tell you about an upcoming severe threat?
- Do you like the way that the SPC outlooks present hazard information?

Transition questions

- What is the most significant severe weather event you have had to be on air for? What was the SPC outlook category for that day? What did you learn from forecasting and broadcasting during that event? (round robin)

Key questions

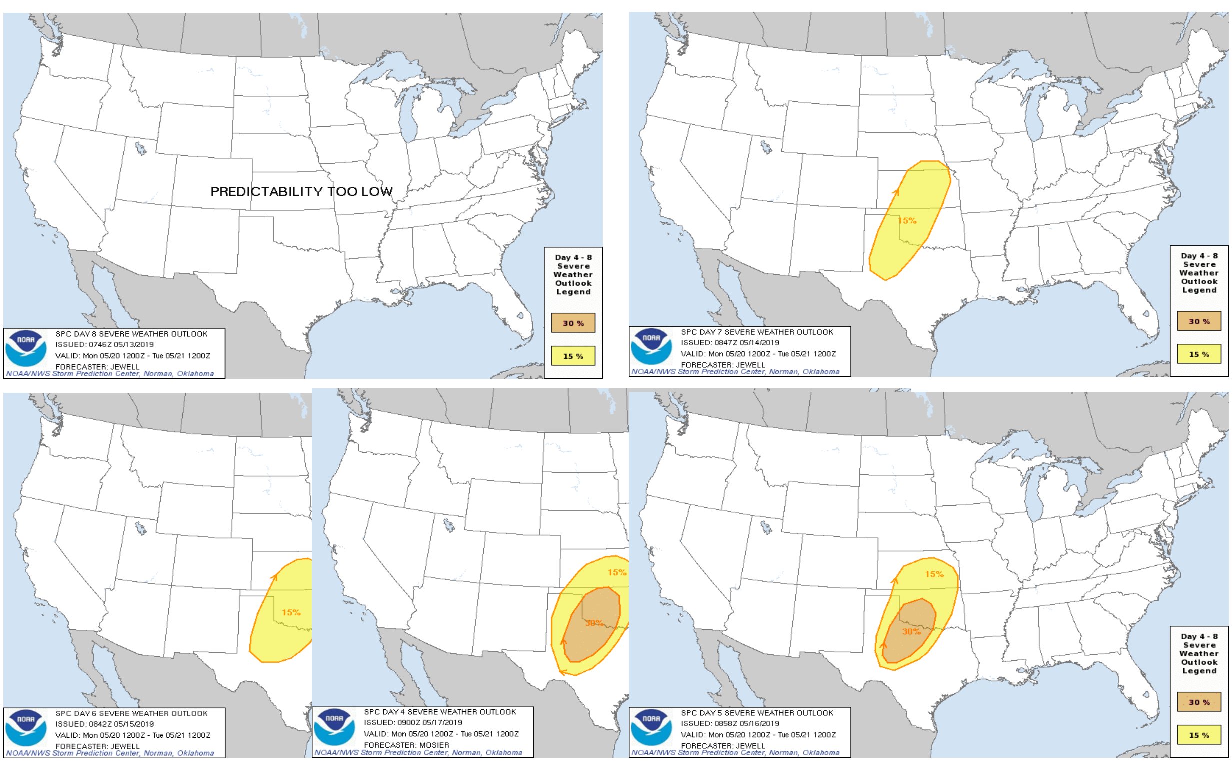
- (Show example of gen thunder, marginal/slight, and moderate/high days for OKC area with days leading up included)
- If the outlook for OKC in this image was issued for your home market, how would you change your messaging for these three different events?
- How would your daily routine differ between these three events?
- For each day leading up to these events, what would be the most important message for you to get to your viewers in ten seconds or less?

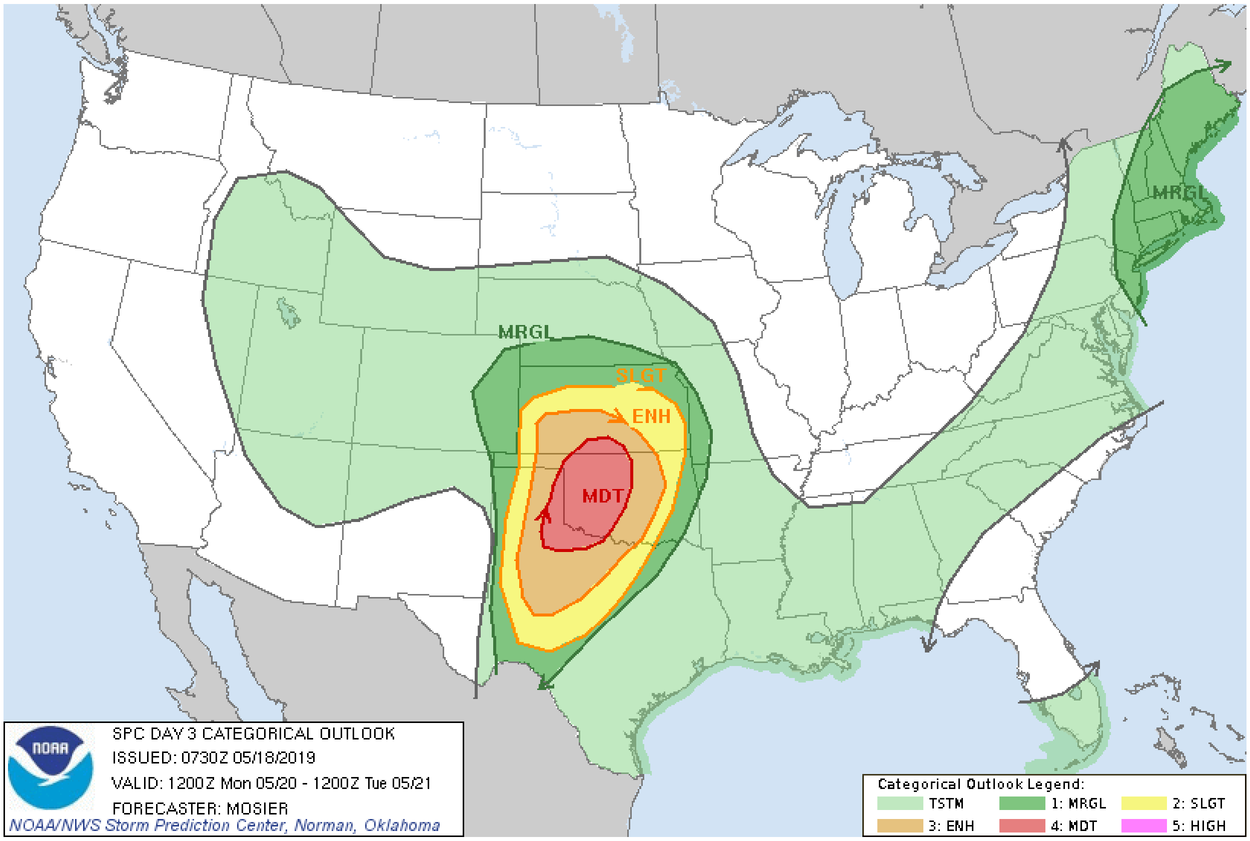
End question - moderator gives summary of thoughts, have we missed anything

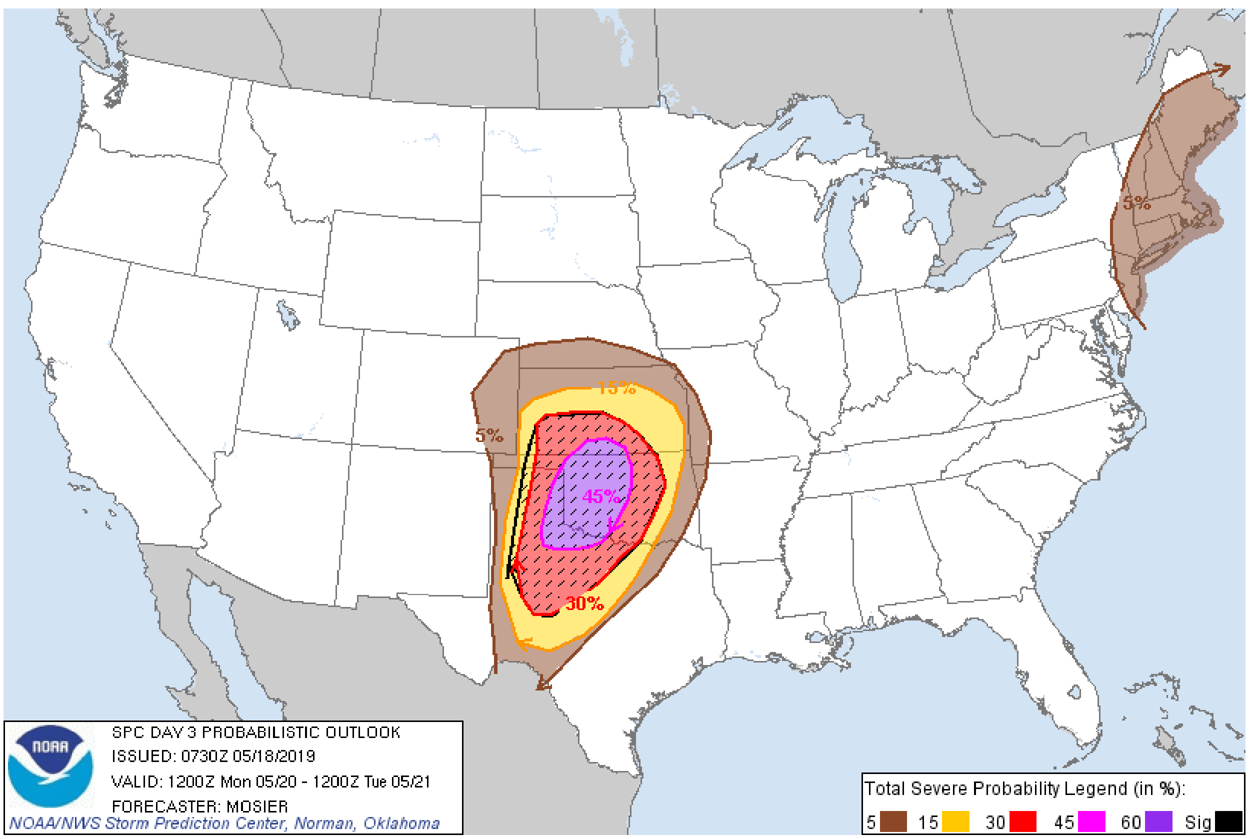
- How would you change the information shared in SPC convective outlooks, if you had total control over the forecast?
- Did we miss any important ideas?

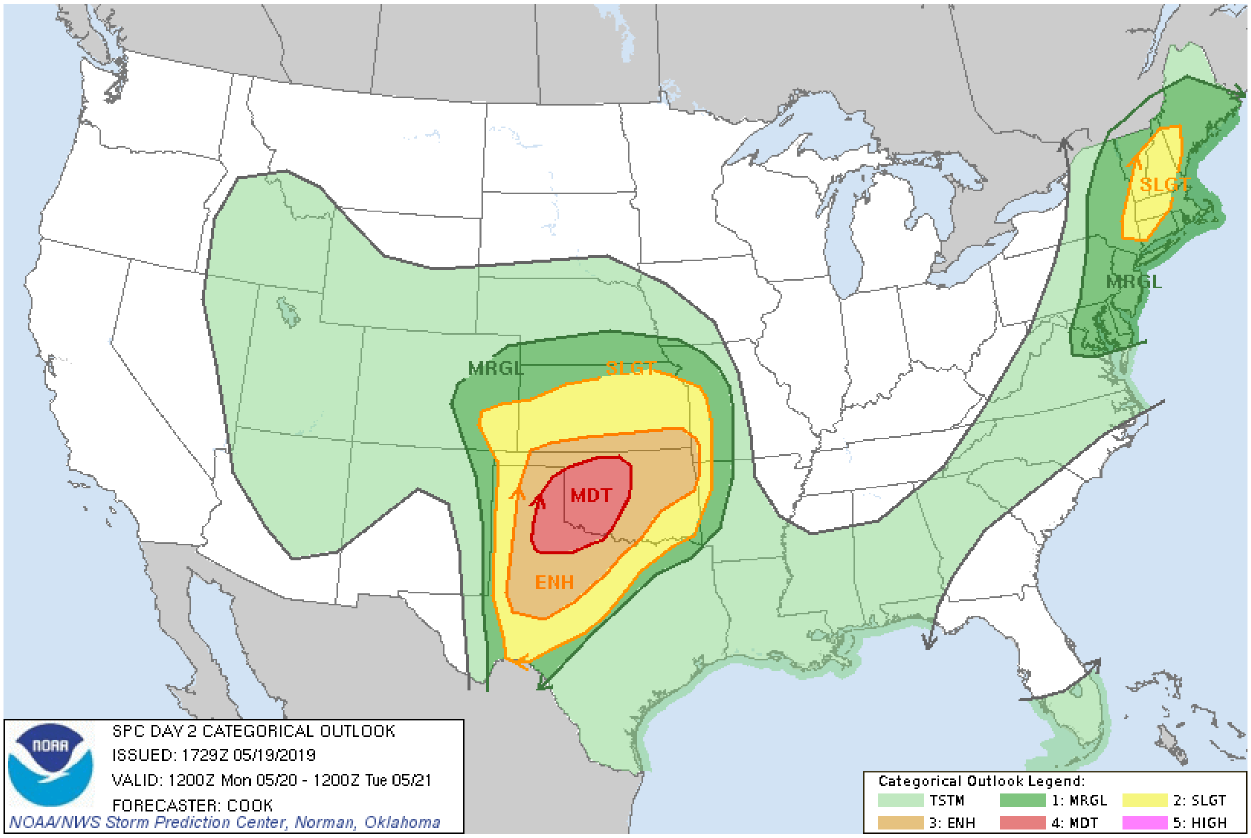
- ## Obtain Background Info with Example Outlooks (following pages)

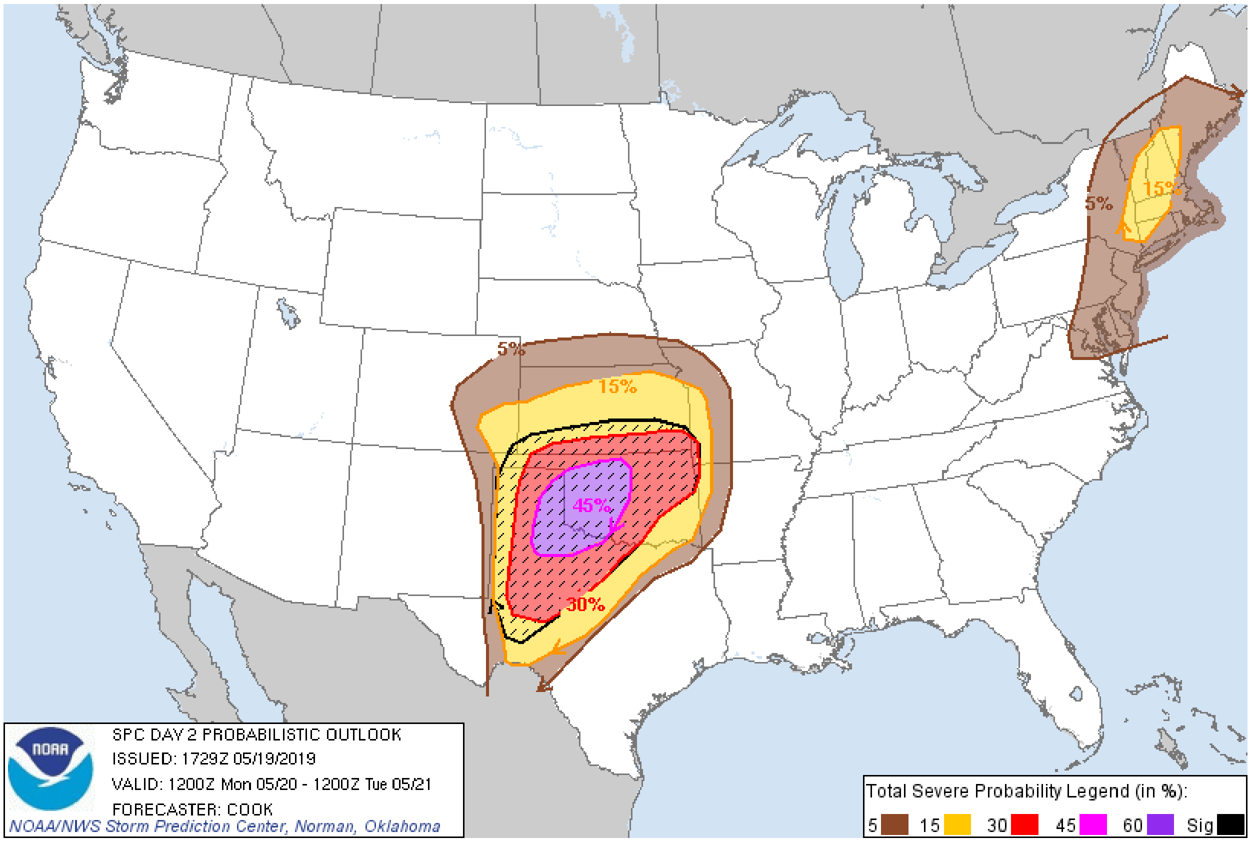
High risk day:

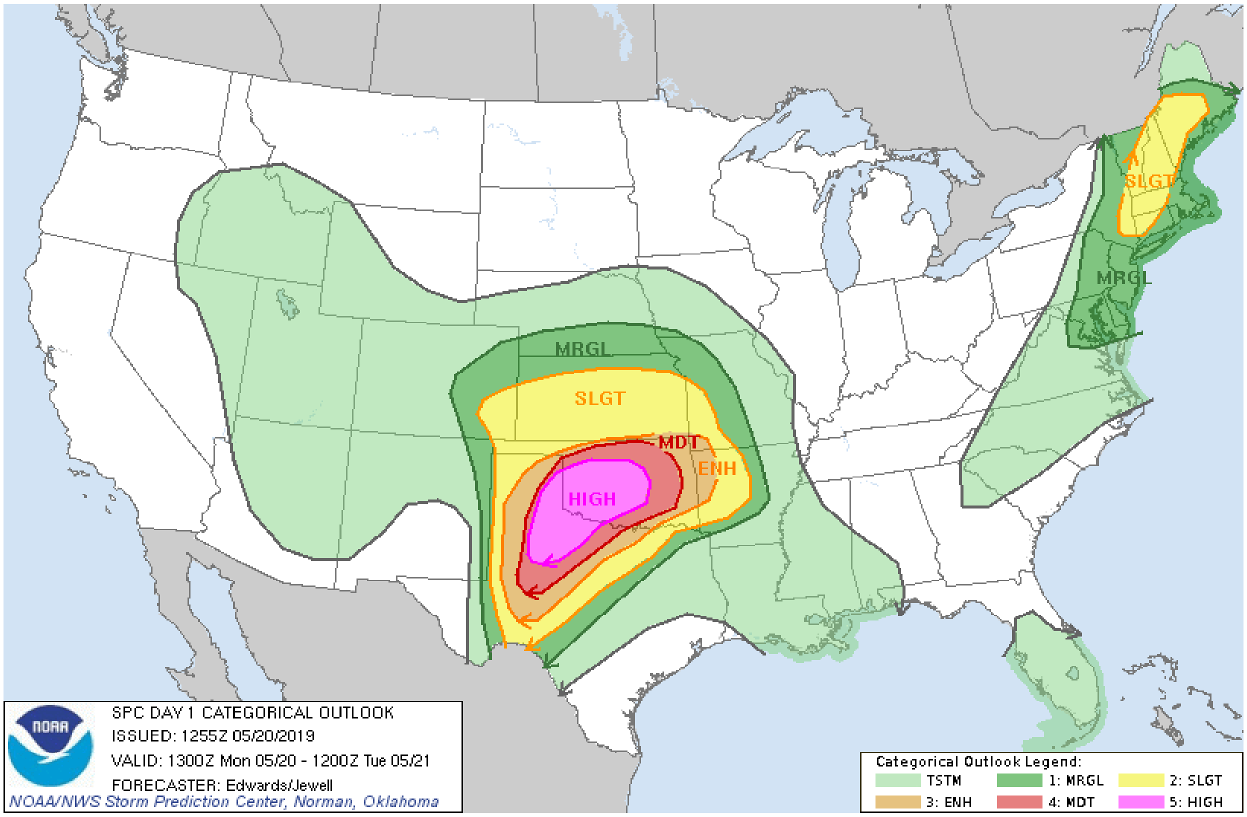


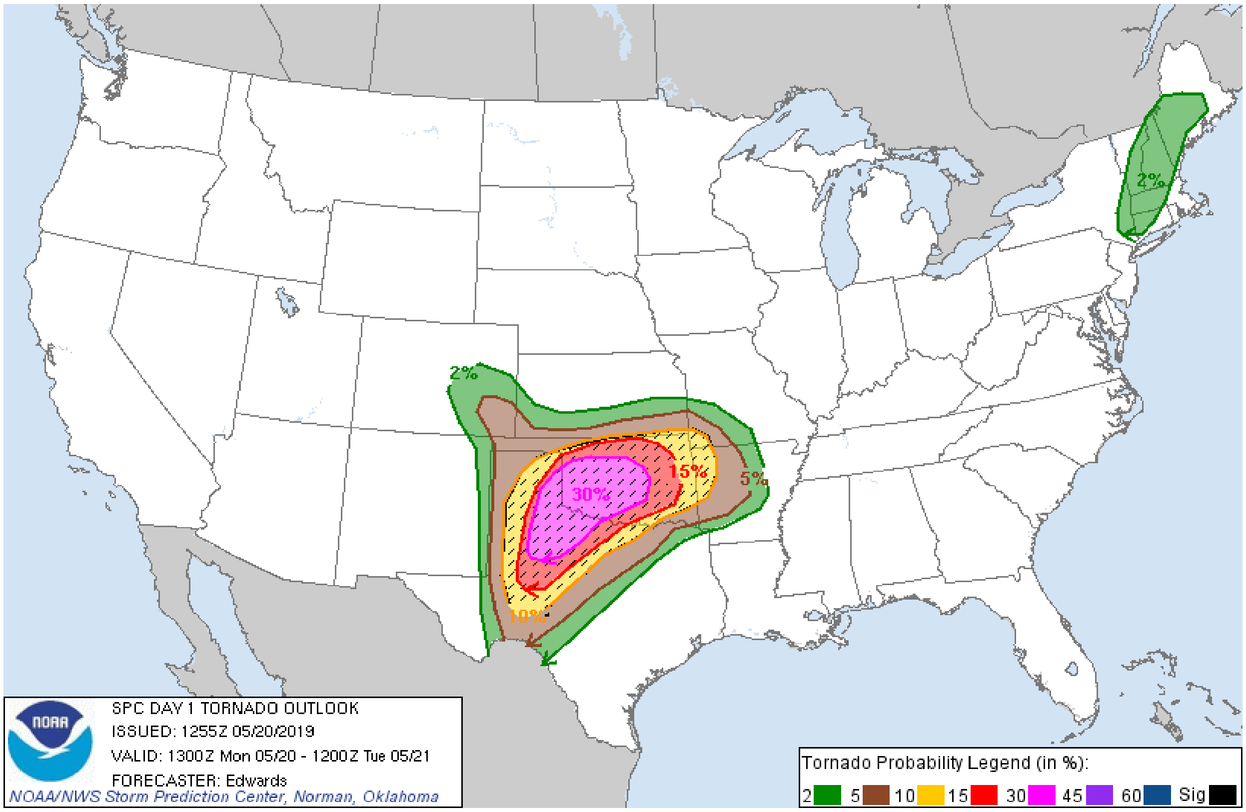

MRGL
MRGL
SLGT
ENH
MDT
SPC DAY 3 CATEGORICAL OUTLOOK
ISSUED: 0730Z 05/18/2019
VALID: 1200Z Mon 05/20 - 1200Z Tue 05/21
FORECASTER: MOSIER
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Categorical Outlook Legend:
TSTM
1: MRGL
2: SLGT
3: ENH
4: MDT
5: HIGH

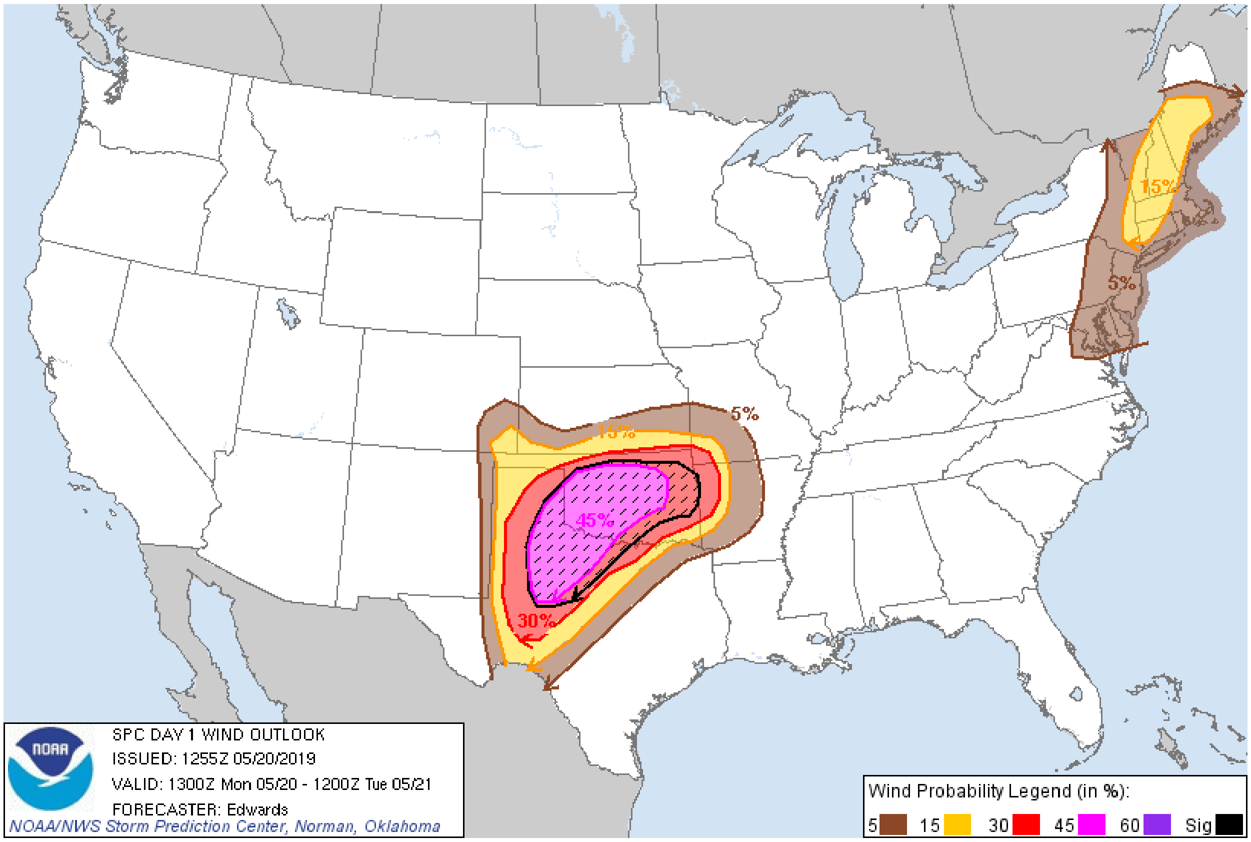

5%
15%
5%
45%
30%
5%
SPC DAY 3 PROBABILISTIC OUTLOOK
ISSUED: 0730Z 05/18/2019
VALID: 1200Z Mon 05/20 - 1200Z Tue 05/21
FORECASTER: MOSIER
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Total Severe Probability Legend (in %):
5
15
30
45
60
Sig

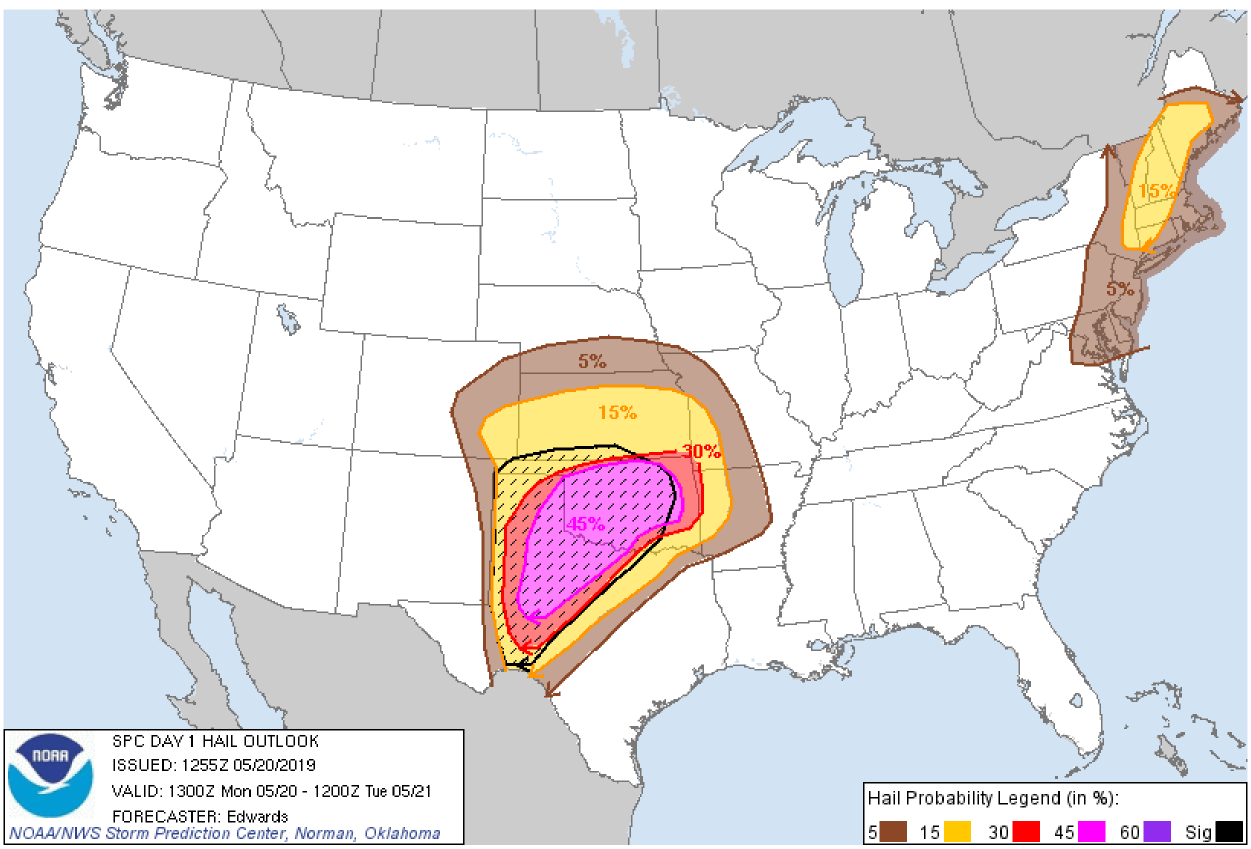

MRGL
SLGT
MDT
ENH
SLGT
MRGL
SPC DAY 2 CATEGORICAL OUTLOOK
ISSUED: 1729Z 05/19/2019
VALID: 1200Z Mon 05/20 - 1200Z Tue 05/21
FORECASTER: COOK
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Categorical Outlook Legend:
TSTM
1: MRGL
2: SLGT
3: ENH
4: MDT
5: HIGH


5%
15%
45%
30%
5%
15%
SPC DAY 2 PROBABILISTIC OUTLOOK
ISSUED: 1729Z 05/19/2019
VALID: 1200Z Mon 05/20 - 1200Z Tue 05/21
FORECASTER: COOK
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Total Severe Probability Legend (in %):
5 15 30 45 60 Sig

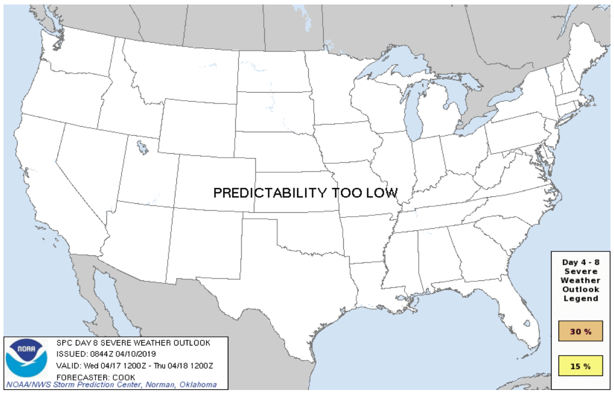

MRGL
SLGT
MDT
ENH
HIGH
SLGT
MRGL
SPC DAY 1 CATEGORICAL OUTLOOK
ISSUED: 1255Z 05/20/2019
VALID: 1300Z Mon 05/20 - 1200Z Tue 05/21
FORECASTER: Edwards/Jewell
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Categorical Outlook Legend:
TSTM
1: MRGL
2: SLGT
3: ENH
4: MDT
5: HIGH

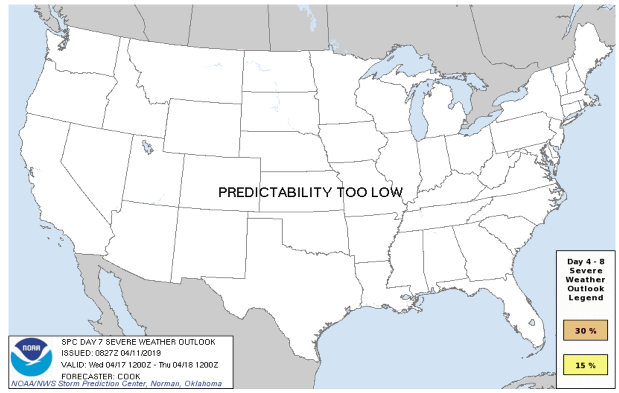

2%
2%
15%
5%
30%
10%
SPC DAY 1 TORNADO OUTLOOK
ISSUED: 1255Z 05/20/2019
VALID: 1300Z Mon 05/20 - 1200Z Tue 05/21
FORECASTER: Edwards
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Tornado Probability Legend (in %):
2 5 10 15 30 45 60 Sig

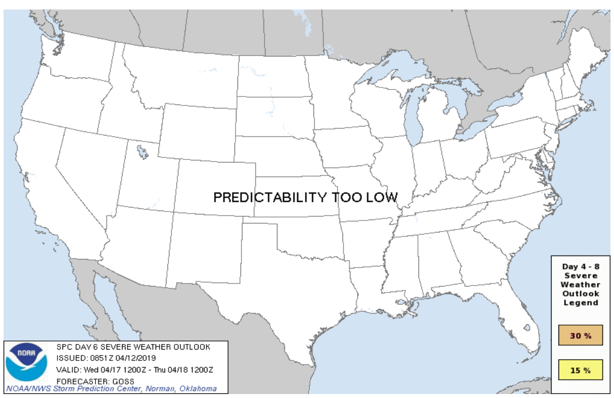

5%
15%
45%
30%
15%
5%
SPC DAY 1 WIND OUTLOOK
ISSUED: 1255Z 05/20/2019
VALID: 1300Z Mon 05/20 - 1200Z Tue 05/21
FORECASTER: Edwards
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Wind Probability Legend (in %):
5 15 30 45 60 Sig

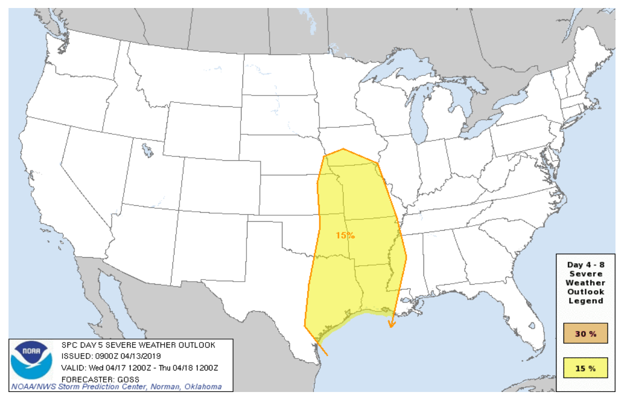

5%
15%
30%
45%
15%
5%
SPC DAY 1 HAIL OUTLOOK
ISSUED: 1255Z 05/20/2019
VALID: 1300Z Mon 05/20 - 1200Z Tue 05/21
FORECASTER: Edwards
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Hail Probability Legend (in %):
5 15 30 45 60 Sig

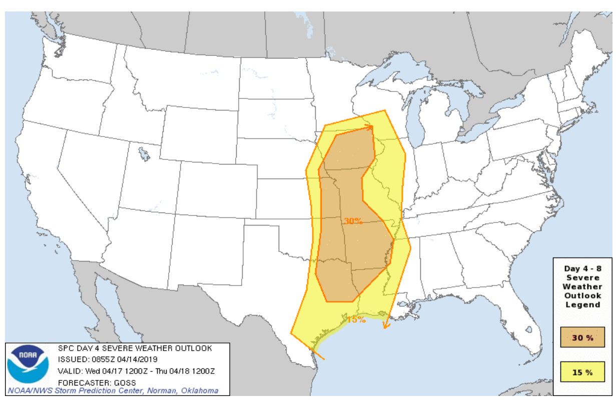
Slight/Enhanced Day:

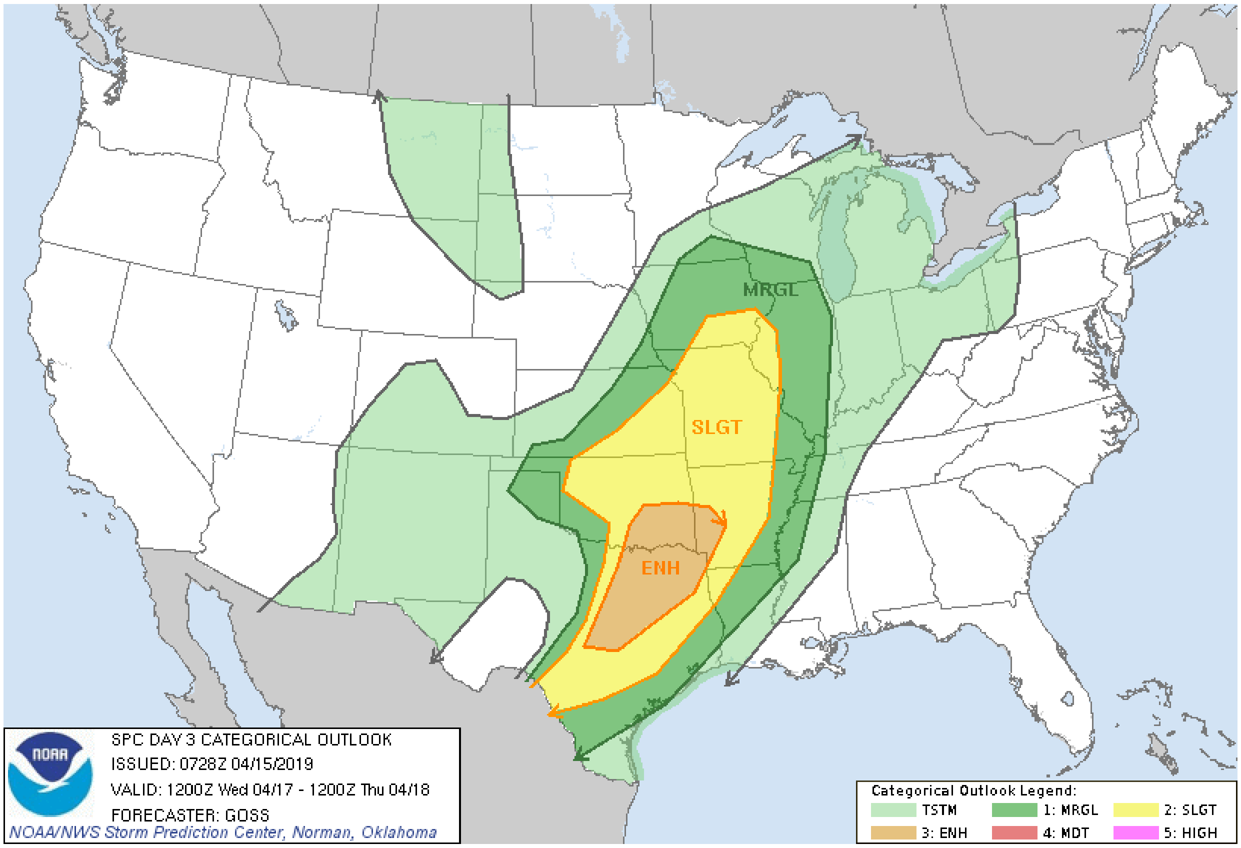


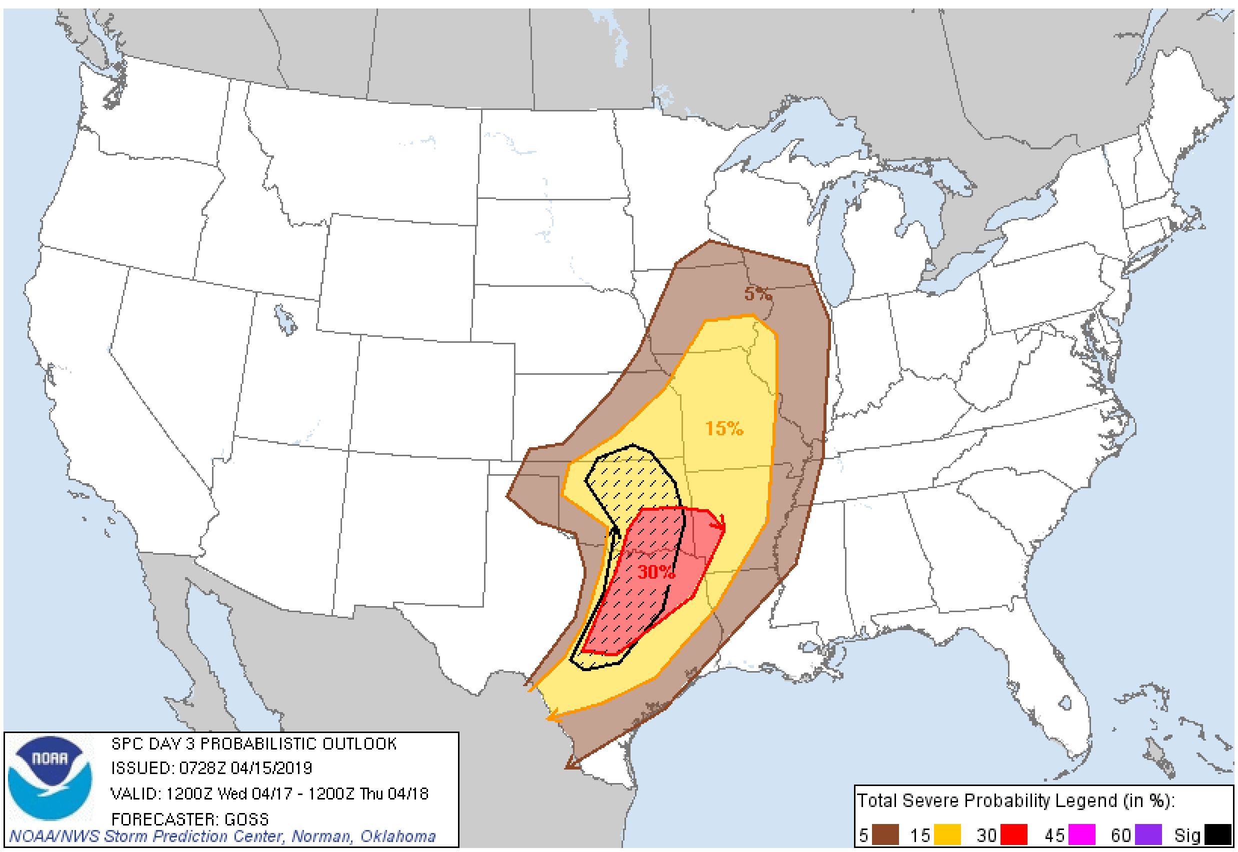


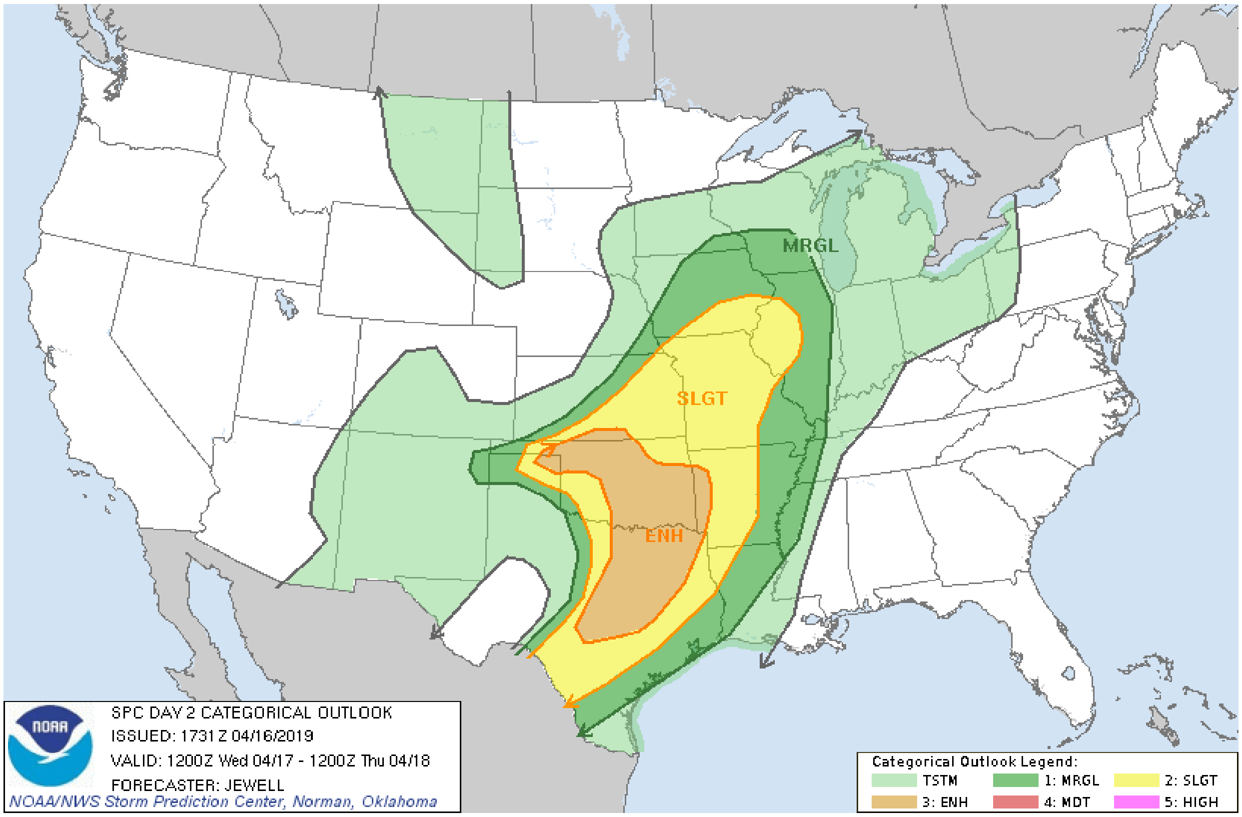


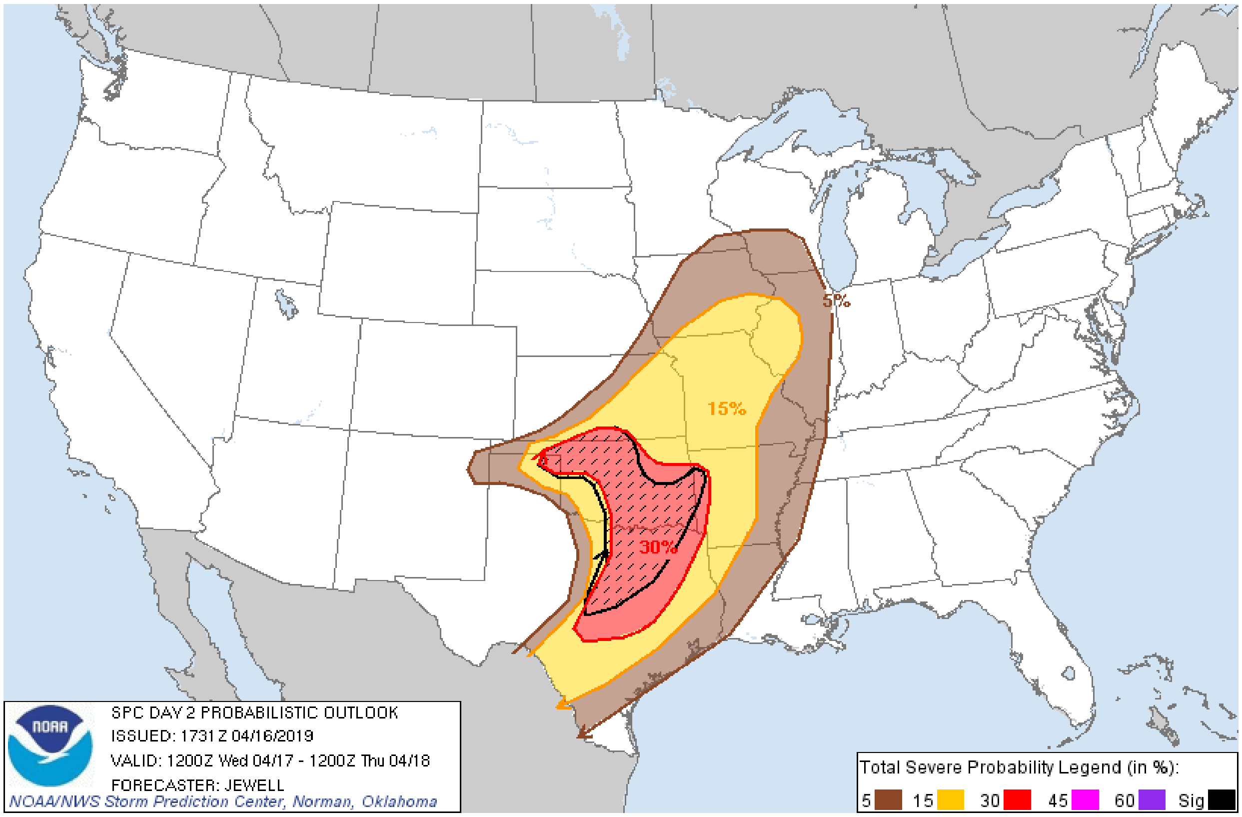


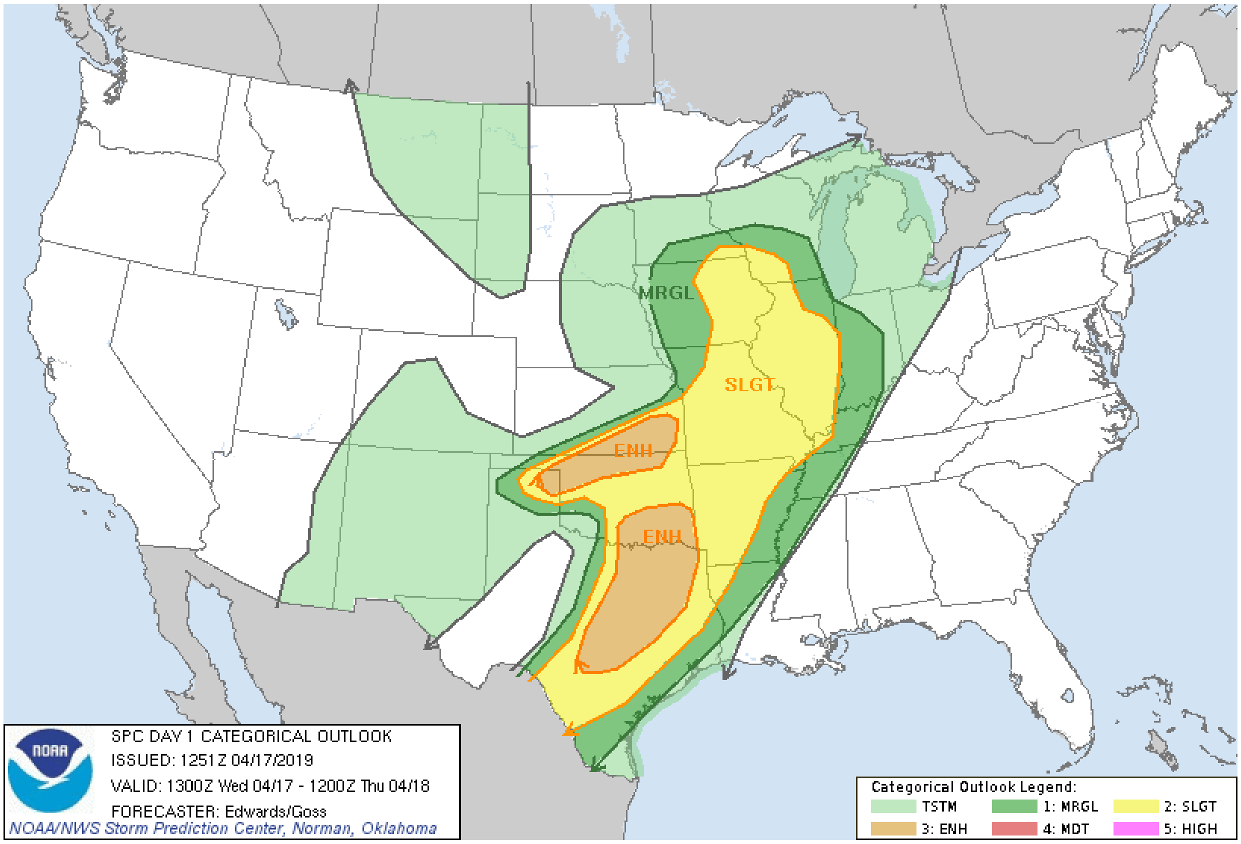


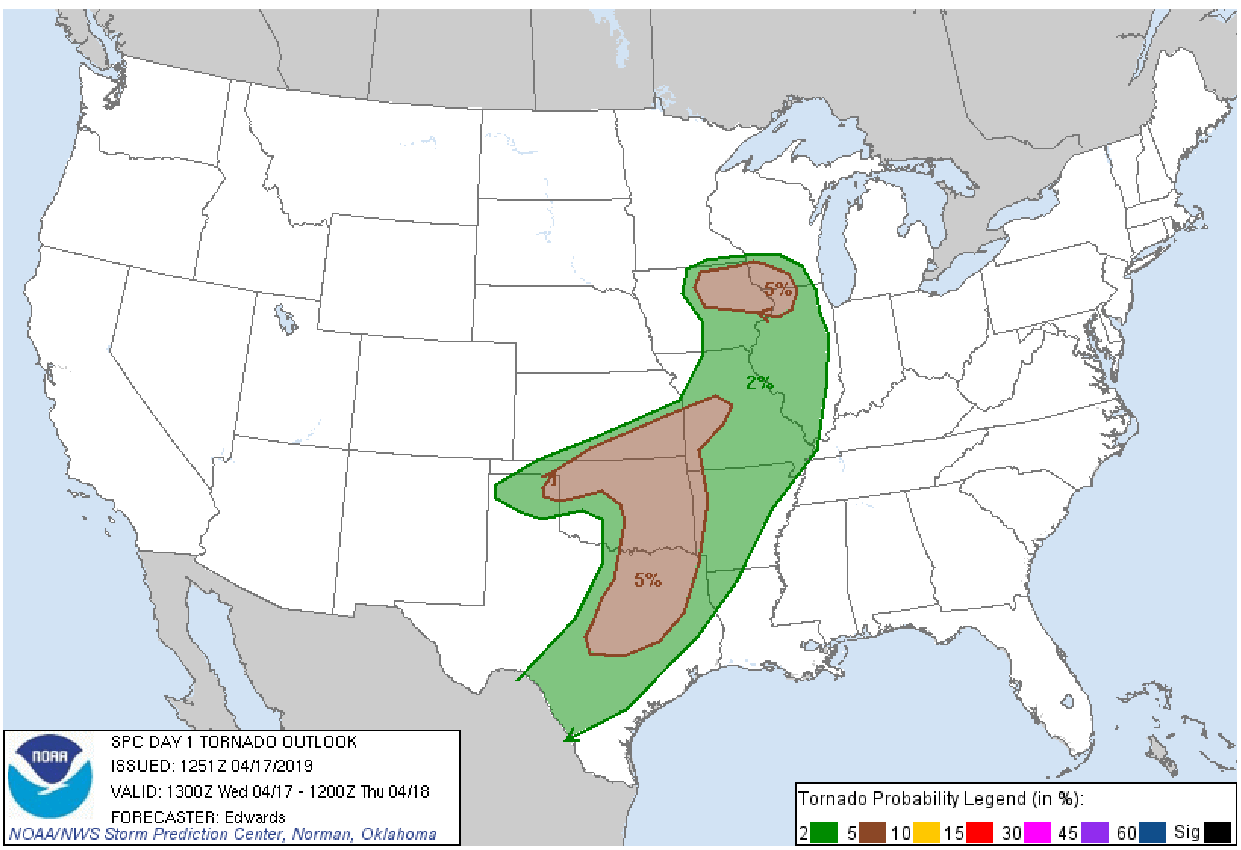

MRGL
SLGT
ENH
SPC DAY 3 CATEGORICAL OUTLOOK
ISSUED: 0728Z 04/15/2019
VALID: 1200Z Wed 04/17 - 1200Z Thu 04/18
FORECASTER: GOSS
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Categorical Outlook Legend:
TSTM
1: MRGL
2: SLGT
3: ENH
4: MDT
5: HIGH

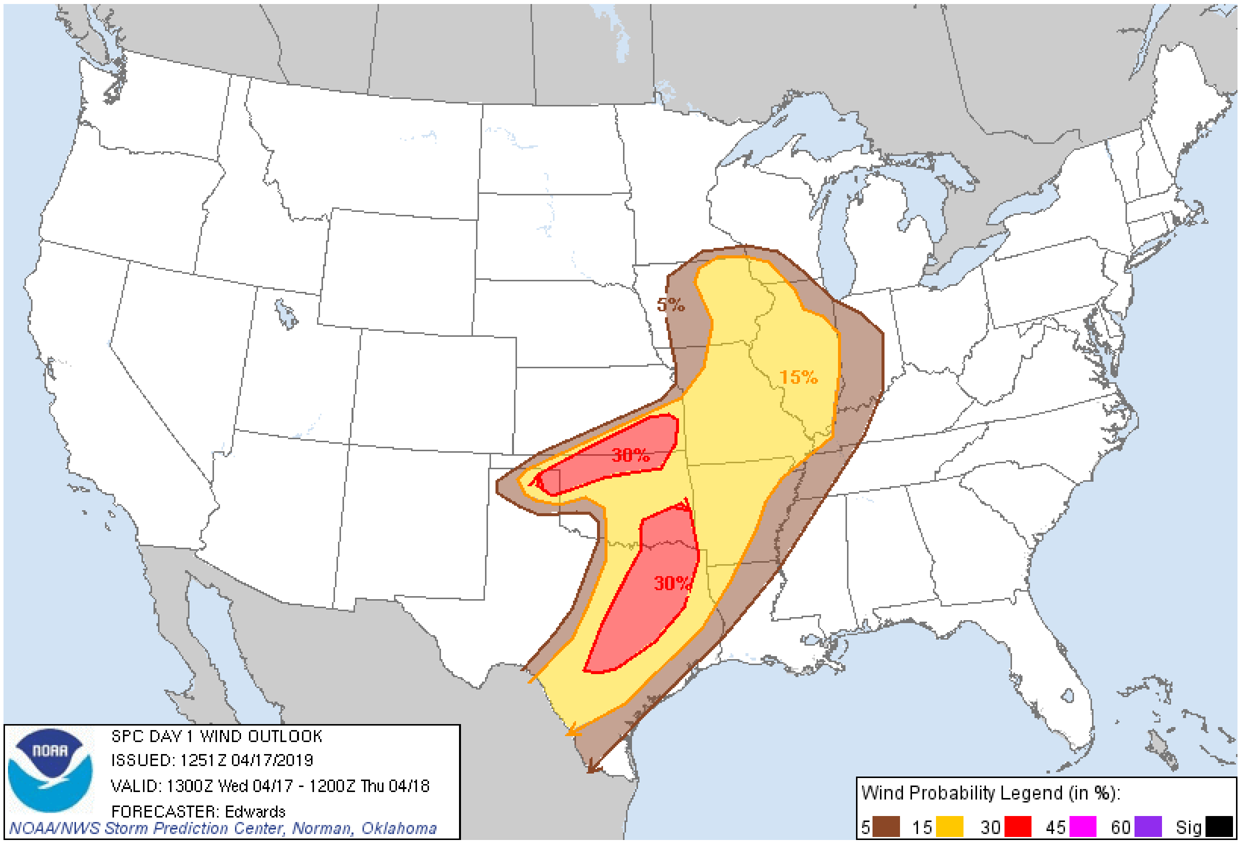

5%
15%
30%
SPC DAY 3 PROBABILISTIC OUTLOOK
ISSUED: 0728Z 04/15/2019
VALID: 1200Z Wed 04/17 - 1200Z Thu 04/18
FORECASTER: GOSS
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Total Severe Probability Legend (in %):
5
15
30
45
60
Sig


MRGL
SLGT
ENH
SPC DAY 2 CATEGORICAL OUTLOOK
ISSUED: 1731 Z 04/16/2019
VALID: 1200Z Wed 04/17 - 1200Z Thu 04/18
FORECASTER: JEWELL
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Categorical Outlook Legend:
TSTM
1: MRGL
2: SLGT
3: ENH
4: MDT
5: HIGH

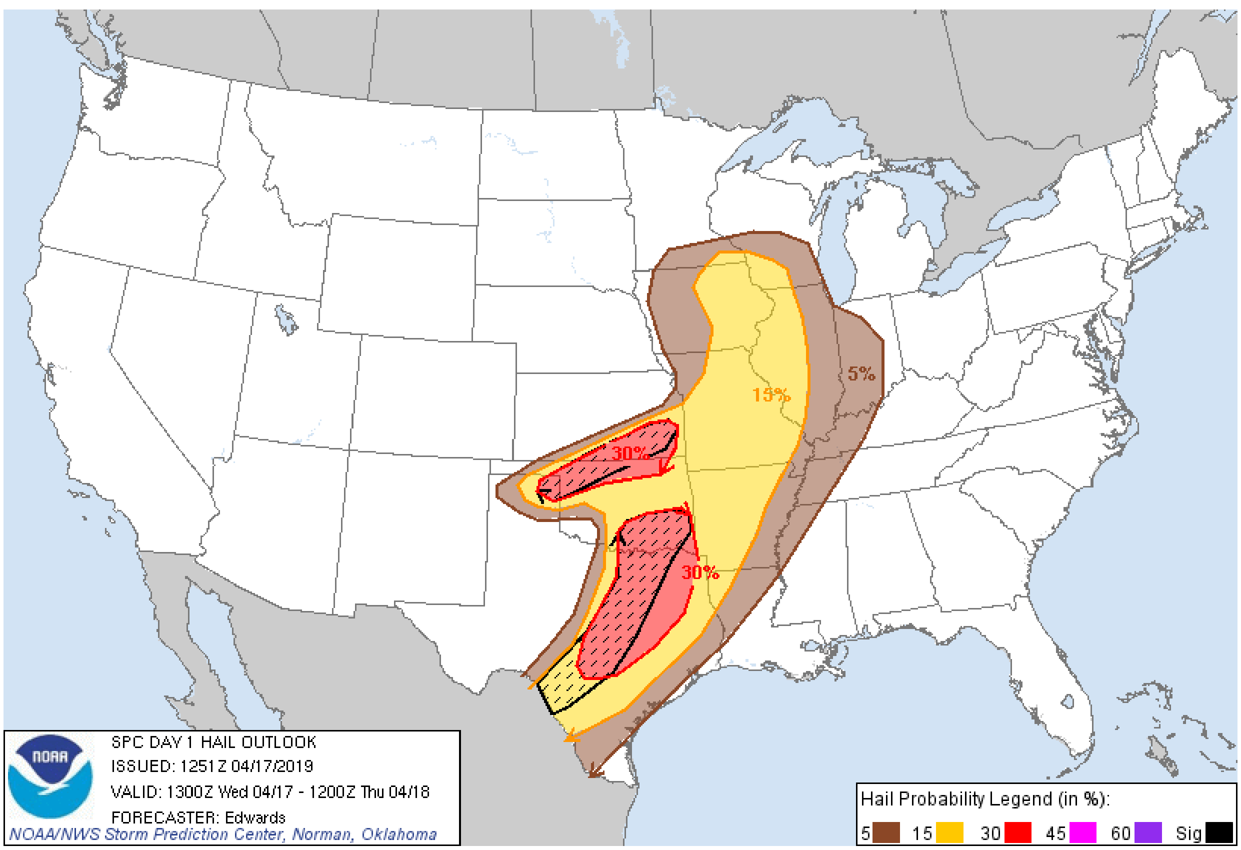

5%
15%
30%
SPC DAY 2 PROBABILISTIC OUTLOOK
ISSUED: 1731 Z 04/16/2019
VALID: 1200Z Wed 04/17 - 1200Z Thu 04/18
FORECASTER: JEWELL
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Total Severe Probability Legend (in %):
5 15 30 45 60 Sig


MRGL
SLGT
ENH
ENH
SPC DAY 1 CATEGORICAL OUTLOOK
ISSUED: 1251Z 04/17/2019
VALID: 1300Z Wed 04/17 - 1200Z Thu 04/18
FORECASTER: Edwards/Goss
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Categorical Outlook Legend:
TSTM
1: MRGL
2: SLGT
3: ENH
4: MDT
5: HIGH


5%
2%
5%
SPC DAY 1 TORNADO OUTLOOK
ISSUED: 1251Z 04/17/2019
VALID: 1300Z Wed 04/17 - 1200Z Thu 04/18
FORECASTER: Edwards
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Tornado Probability Legend (in %):
2 5 10 15 30 45 60 Sig

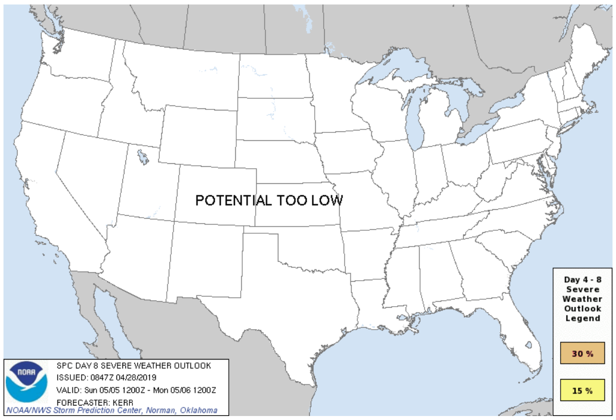


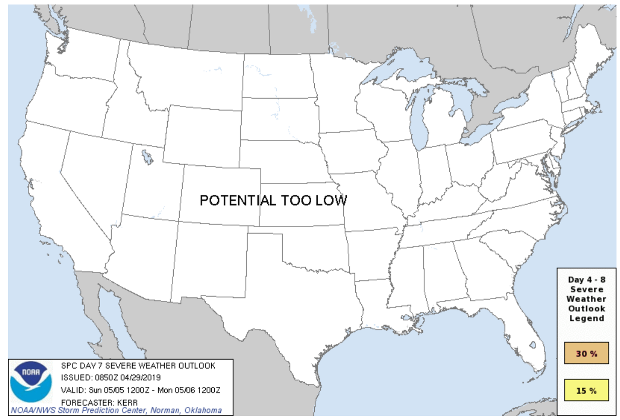
5%
15%
30%
30%
SPC DAY 1 WIND OUTLOOK
ISSUED: 1251Z 04/17/2019
VALID: 1300Z Wed 04/17 - 1200Z Thu 04/18
FORECASTER: Edwards
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Wind Probability Legend (in %):
5 15 30 45 60 Sig

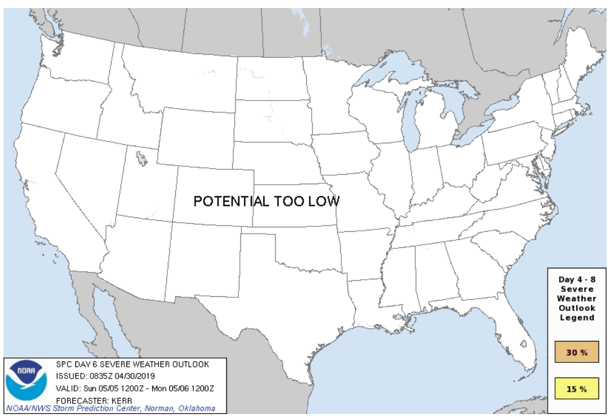


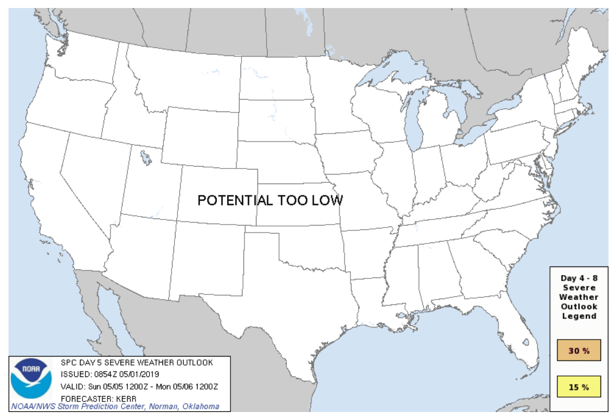
5%
15%
30%
30%
SPC DAY 1 HAIL OUTLOOK
ISSUED: 1251Z 04/17/2019
VALID: 1300Z Wed 04/17 - 1200Z Thu 04/18
FORECASTER: Edwards
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Hail Probability Legend (in %):
5 15 30 45 60 Sig

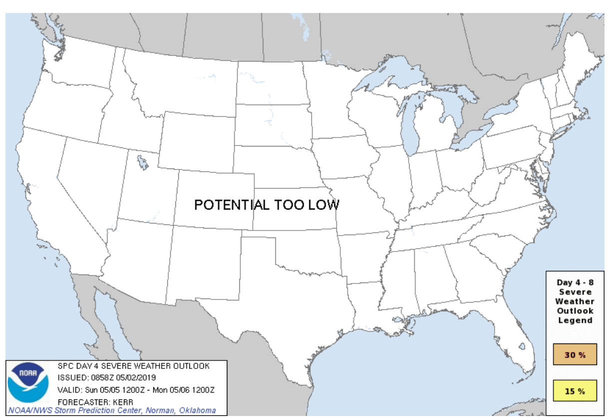
Marginal Event

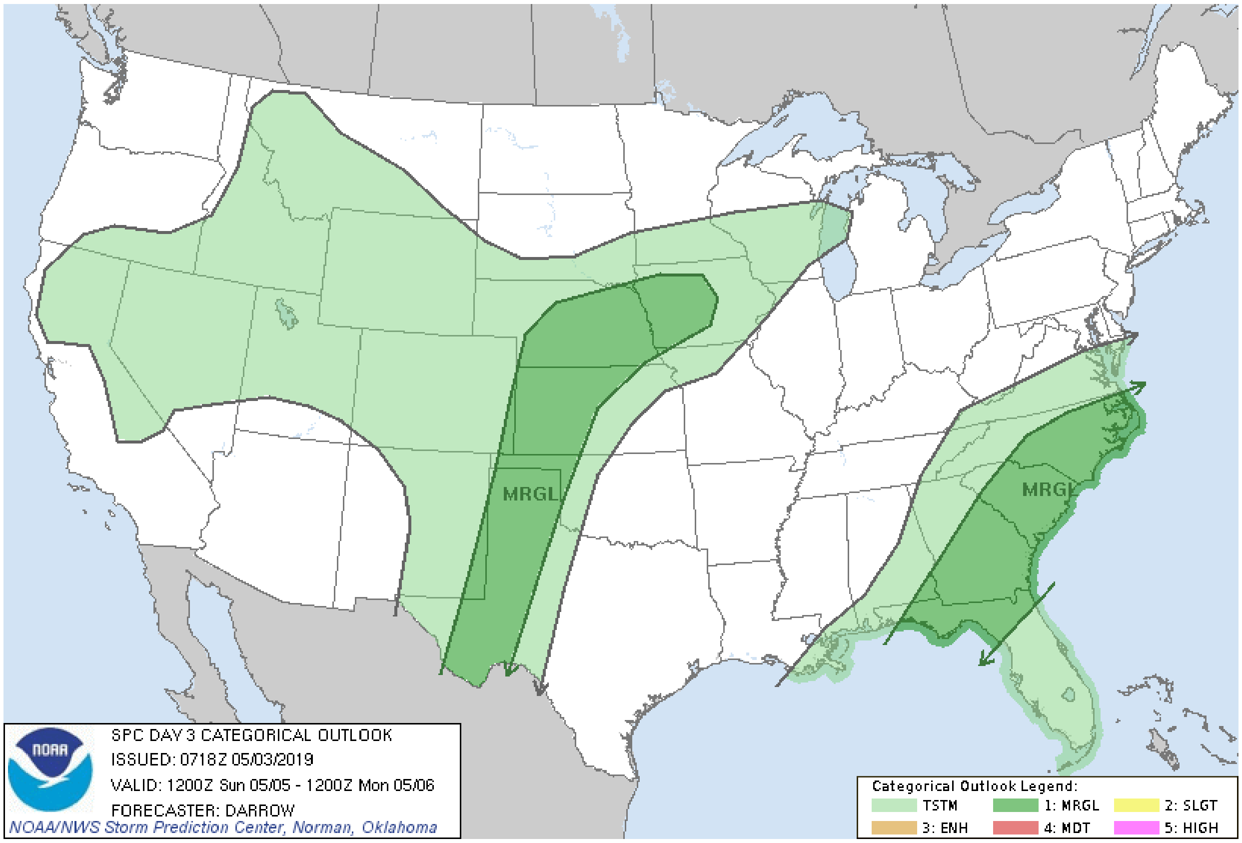


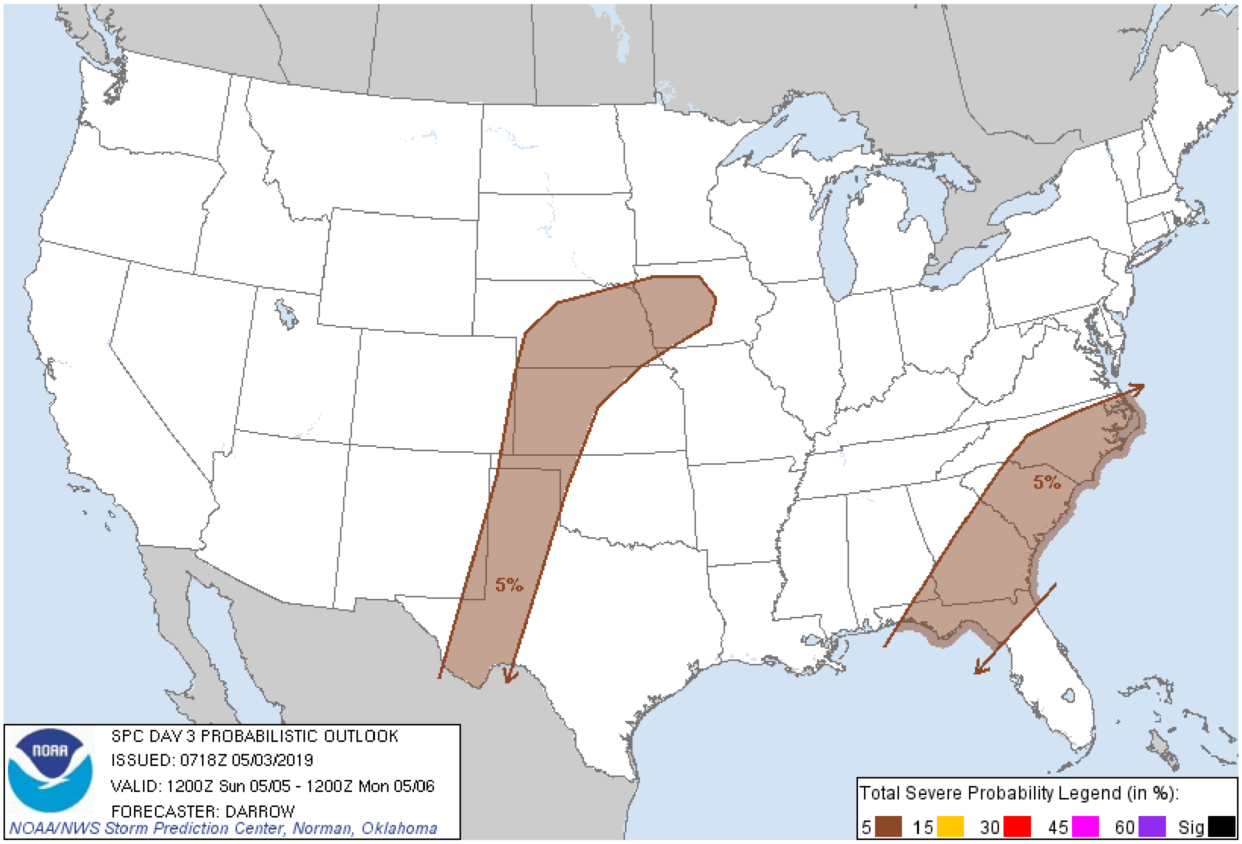


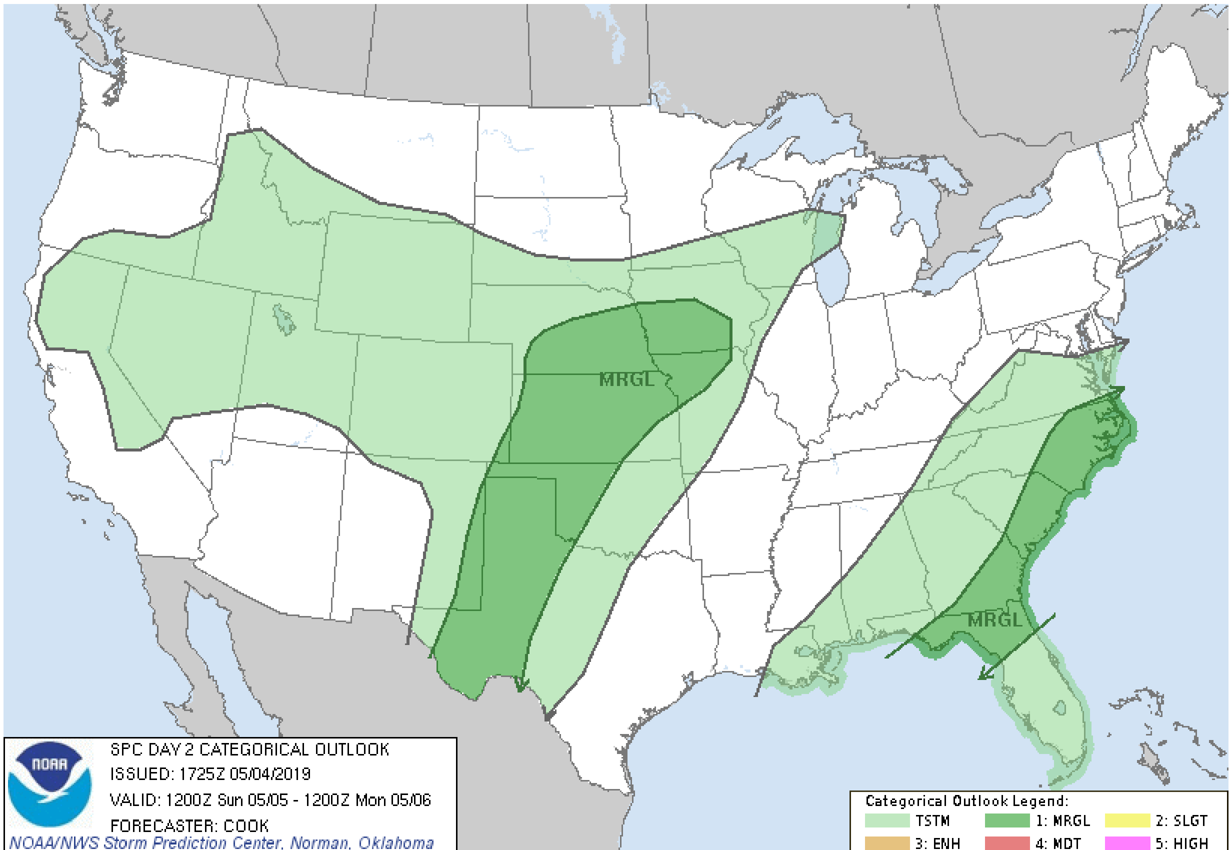


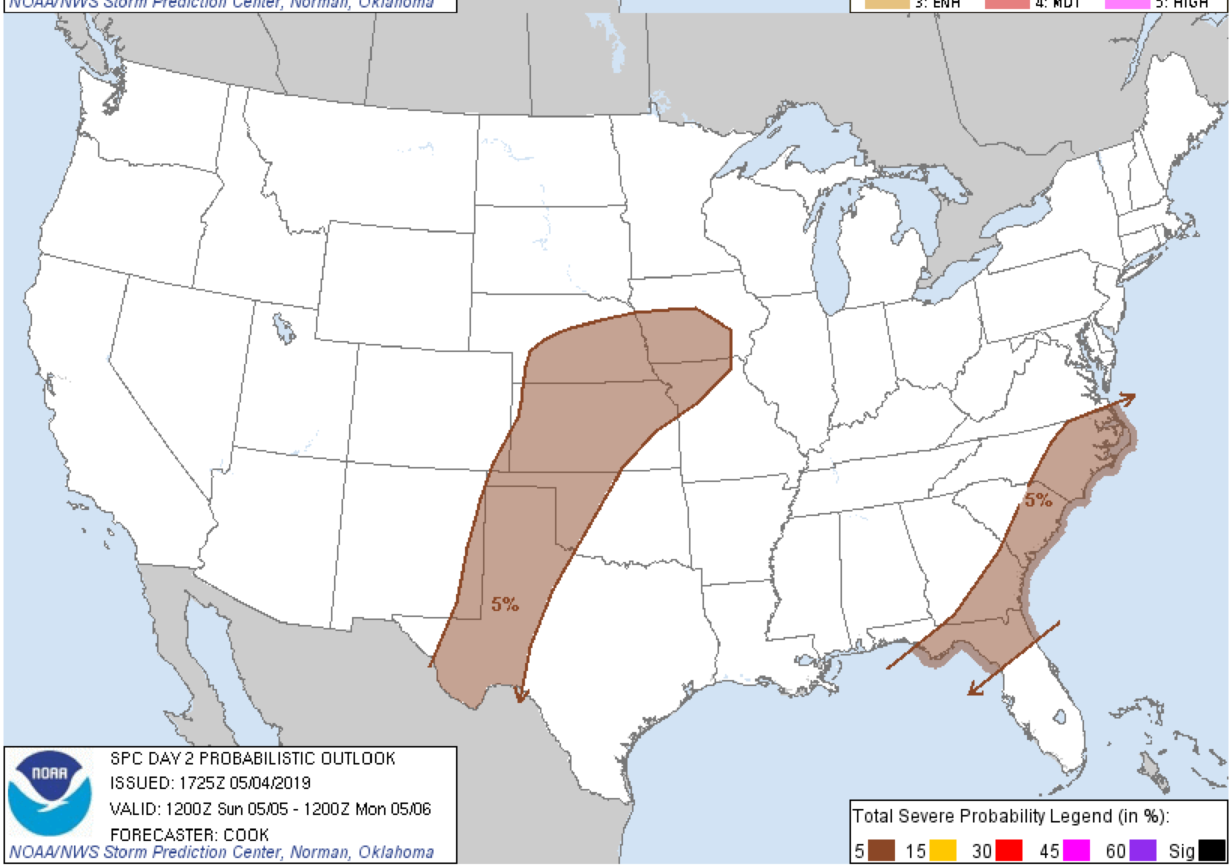


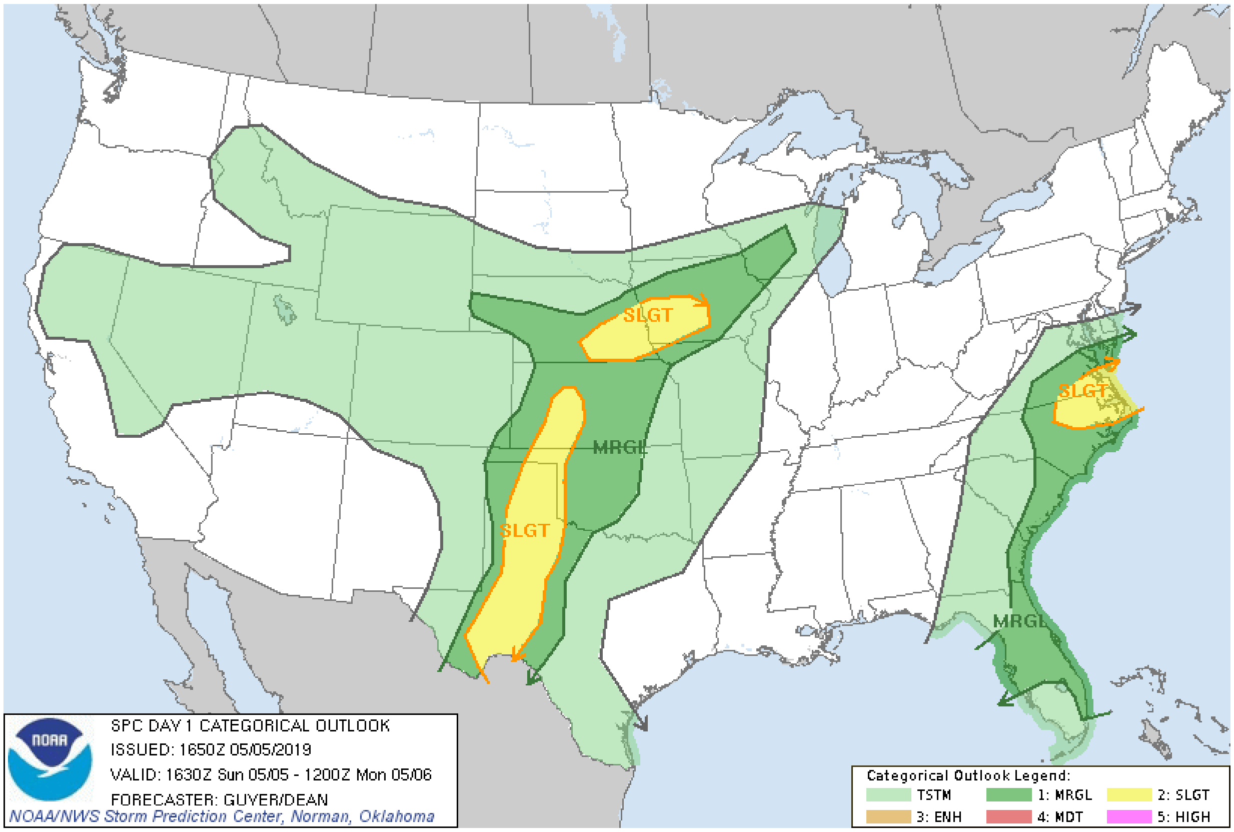


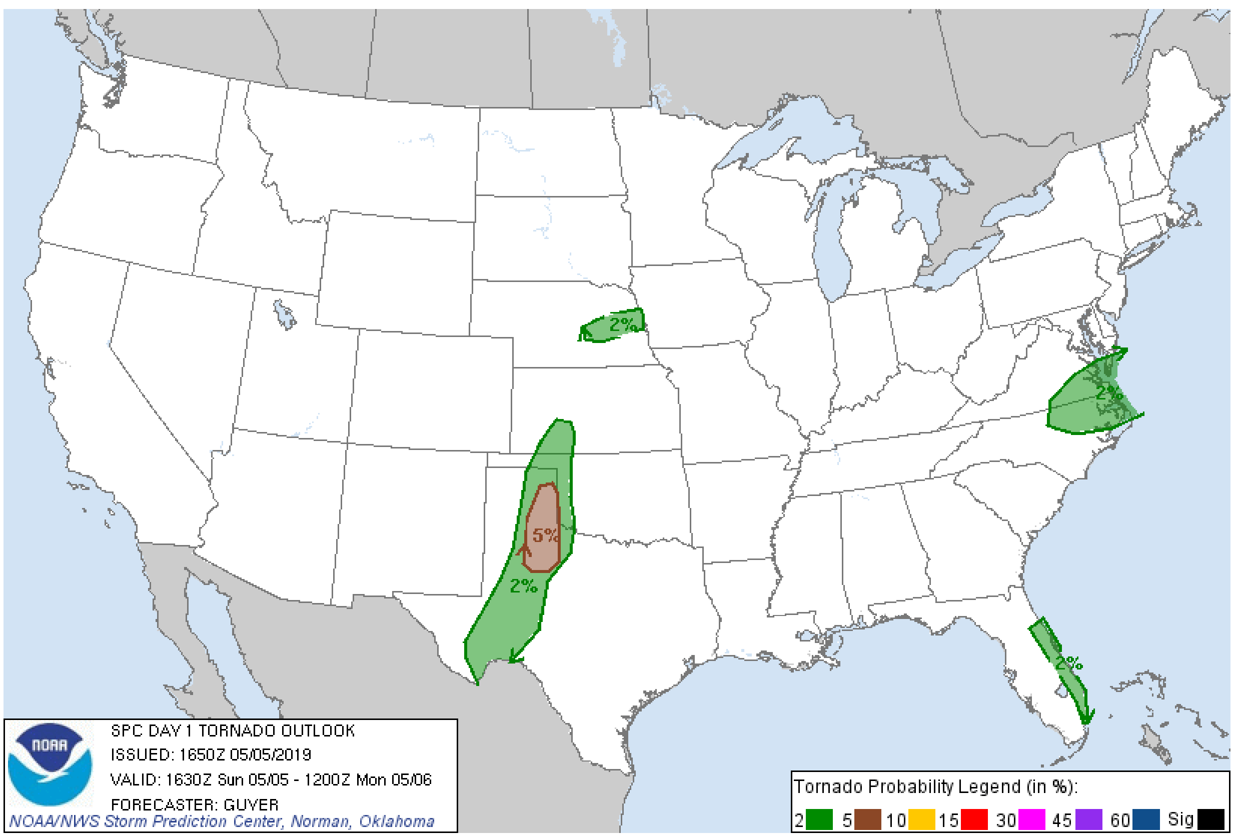

SPC DAY 3 CATEGORICAL OUTLOOK
ISSUED: 0718Z 05/03/2019
VALID: 1200Z Sun 05/05 - 1200Z Mon 05/06
FORECASTER: DARROW
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
MRGL
MRGL
Categorical Outlook Legend:
TSTM
1: MRGL
2: SLGT
3: ENH
4: MDT
5: HIGH

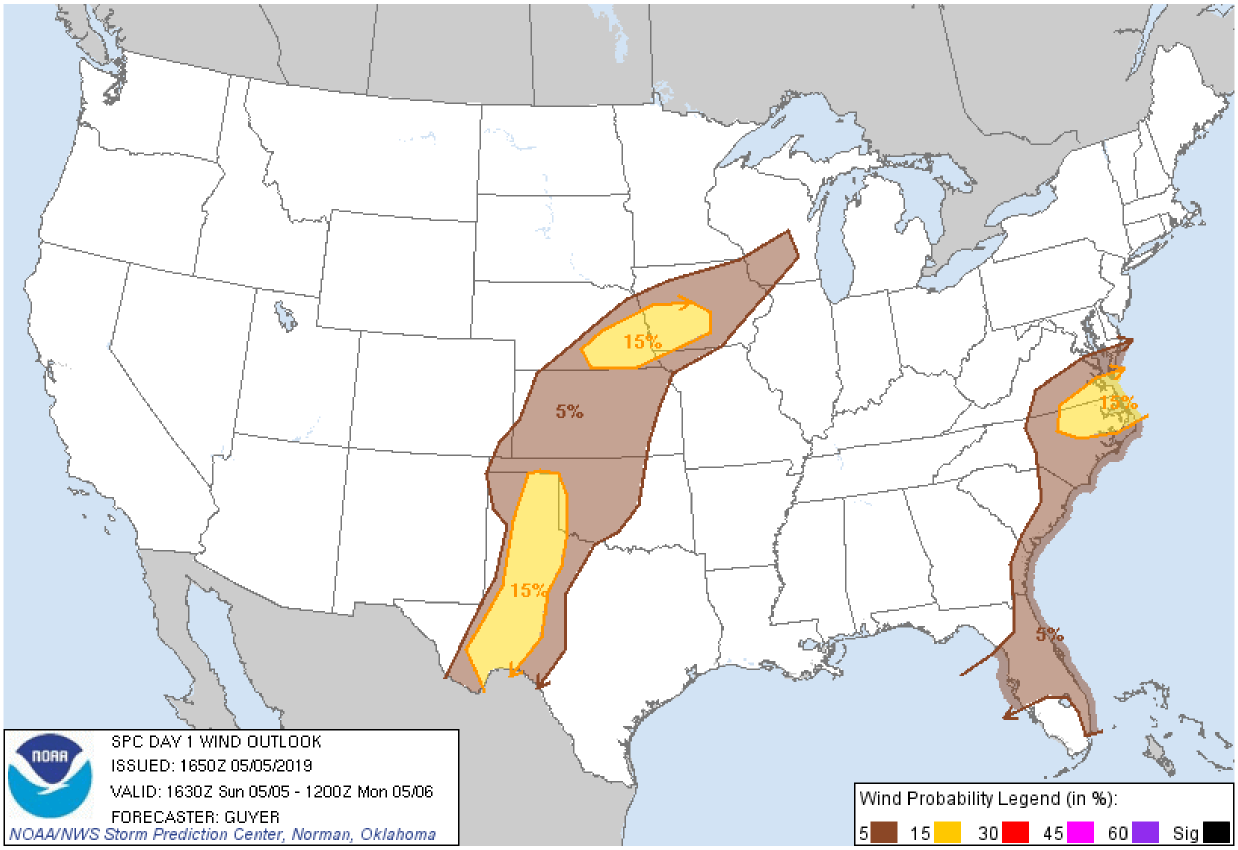

SPC DAY 3 PROBABILISTIC OUTLOOK
ISSUED: 0718Z 05/03/2019
VALID: 1200Z Sun 05/05 - 1200Z Mon 05/06
FORECASTER: DARROW
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
5%
5%
Total Severe Probability Legend (in %):
5
15
30
45
60
Sig

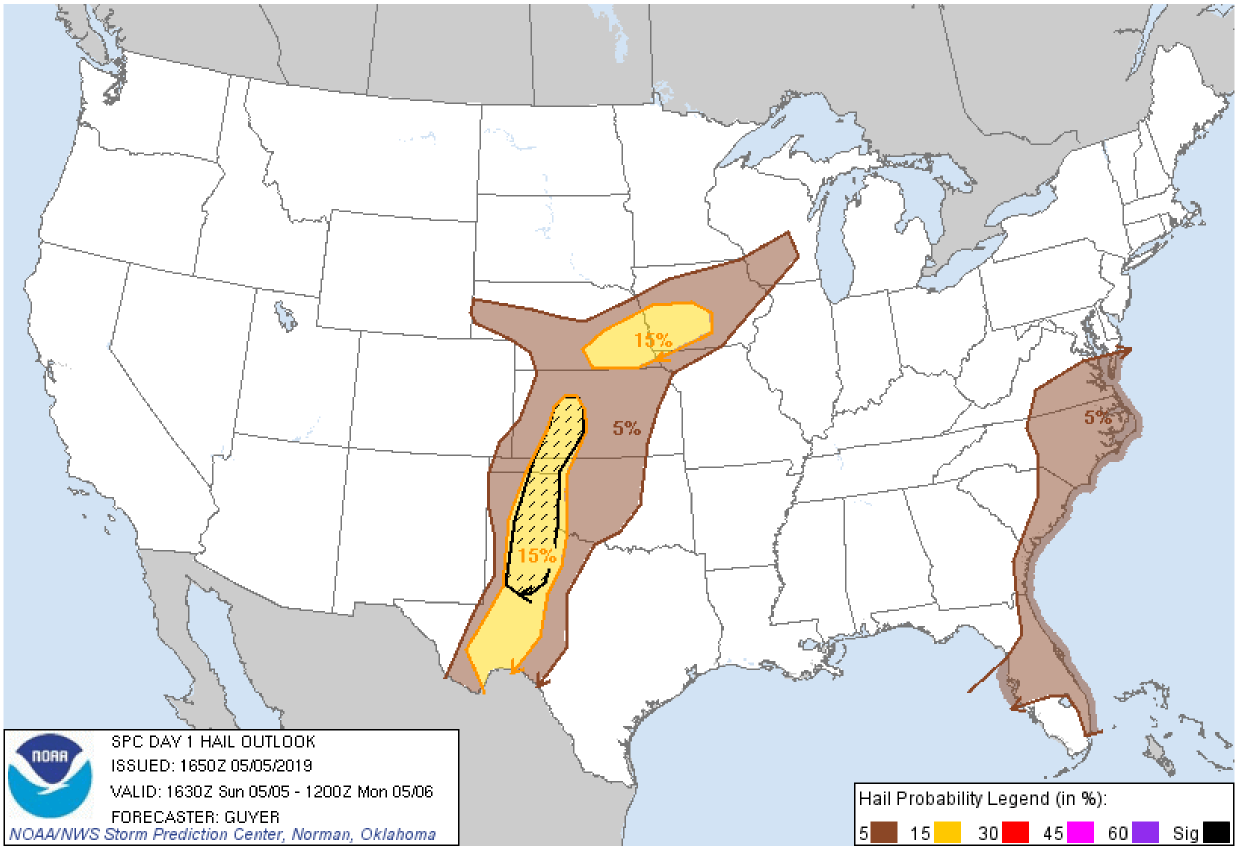

MRGL
MRGL
SPC DAY 2 CATEGORICAL OUTLOOK
ISSUED: 1725Z 05/04/2019
VALID: 1200Z Sun 05/05 - 1200Z Mon 05/06
FORECASTER: COOK
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Categorical Outlook Legend:
TSTM
1: MRGL
2: SLGT
3: ENH
4: MDT
5: HIGH


5%
5%
SPC DAY 2 PROBABILISTIC OUTLOOK
ISSUED: 1725Z 05/04/2019
VALID: 1200Z Sun 05/05 - 1200Z Mon 05/06
FORECASTER: COOK
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Total Severe Probability Legend (in %):
5
15
30
45
60
Sig


SLGT
SLGT
MRGL
SLGT
MRGL
SPC DAY 1 CATEGORICAL OUTLOOK
ISSUED: 1650Z 05/05/2019
VALID: 1630Z Sun 05/05 - 1200Z Mon 05/06
FORECASTER: GUYER/DEAN
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Categorical Outlook Legend:
TSTM
1: MRGL
2: SLGT
3: ENH
4: MDT
5: HIGH


2%
2%
5%
2%
2%
SPC DAY 1 TORNADO OUTLOOK
ISSUED: 1650Z 05/05/2019
VALID: 1630Z Sun 05/05 - 1200Z Mon 05/06
FORECASTER: GUYER
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Tornado Probability Legend (in %):
2 5 10 15 30 45 60 Sig


15%
5%
15%
15%
5%
SPC DAY 1 WIND OUTLOOK
ISSUED: 1650Z 05/05/2019
VALID: 1630Z Sun 05/05 - 1200Z Mon 05/06
FORECASTER: GUYER
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Wind Probability Legend (in %):
5 15 30 45 60 Sig


15%
5%
15%
5%
SPC DAY 1 HAIL OUTLOOK
ISSUED: 1650Z 05/05/2019
VALID: 1630Z Sun 05/05 - 1200Z Mon 05/06
FORECASTER: GUYER
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Hail Probability Legend (in %):
5 15 30 45 60 Sig