Possible Influences on National Weather Service Tornado Warnings

TREY J. HOLIDAY*

National Weather Center Research Experiences for Undergraduates Program
Norman, Oklahoma,
University of Louisiana Monroe
Monroe, LA

MICHAEL E. BALDWIN

Cooperative Institute for Severe and High-Impact Weather Research and Operations, NOAA/NWS/Storm Prediction Center
Norman, Oklahoma

HAROLD E. BROOKS

NOAA/OAR/National Severe Storms Laboratory
Norman, Oklahoma

KRISTIN M. CALHOUN

NOAA/OAR/National Severe Storms Laboratory
Norman, Oklahoma

THEA N. SANDMÆL

Cooperative Institute for Severe and High-Impact Weather Research and Operations, NOAA/OAR National Severe Storms Laboratory
Norman, Oklahoma

ABSTRACT

Tornado warnings issued by meteorologists rely on radar and environmental data to provide life-saving information, but uncertainties can result in shorter lead times and occasional false alarms. Hence, this study evaluates the performance of tornado warnings issued by the National Weather Service in 2018 using probability of detection, false alarm ratio, success ratio, bias, and critical success index. Both radar data and values from a machine learning-based tornado probability algorithm (TORP) are analyzed for each tornado warning before or near the warning to identify factors contributing to accurate warnings or false alarms. TORP detections are based on a 0.006 s^1 azimuthal shear threshold, which uses 0.5° tilt radar data to display a tornado probability for forecasters to use. Thresholds for TORP, rotational velocity, and azimuthal shear were created to provide forecasters with recommendations to aid in the decision-making process. Additionally, population density was examined as a potential factor affecting warning performance, where the success ratio was generally low for sparsely populated areas and increased for densely populated areas. Together, this work contributes to enhancing the accuracy and effectiveness of future warning systems and operational forecasting.

1. Introduction

Tornado warnings issued by the National Weather Service (NWS) are well known to the public, providing life-saving information if rotation is detected by Weather Surveillance Radar-1988 Doppler radars (WSR-88D). NWS meteorologists consider a variety of radar and environmental variables to aid in their decision-making process for issuing warnings. These decisions can come with uncertainty, leading to shorter lead times and false alarms due to various external factors.

NWS forecasters are concerned for the well-being of the public (Bostrom et al. 2016); (Brooks and Doswell 2002) and also carry the burden of not wanting to issue a warn-

*Corresponding author address: FirstName LastName, University of Louisiana Monroe, 700 University Ave, Monroe, LA 71209 (can be CAPS or your university address)
E-mail: holidaytj@warhawks.ulm.edu

ing to avoid false alarms (Doswell 2004). However, being the backbone of the operational field of meteorology, there have been improvements over time. With the implementation of WSR-88D radars in 1986, the probability of detection (POD) increased from 0.25 to nearly 0.70 in the early 2000s. POD has started to remain steady, but then started to decrease in the mid to late 2010s (Brooks and Correia 2018). Various scientists in the field have researched this decrease in POD in the mid to late 2010s, and there are multiple conclusions about this decrease. One is that large reductions in the false alarm ratio (FAR) correlate with large reductions in POD (Brooks 2004). Another possible cause of this decrease is the changes in the warning system, one with different thresholds to issue tornado warnings, and the duration of the warning was reduced from 45 to 30 minutes (Brooks and Correia 2018).

These changes to the warning protocol highlight the critical decisions that NWS forecasters must make, often under time pressure, as they weigh radar data, external cues, and societal impacts to issue lifesaving warnings. Forecasters are often forced to make split-second decisions with compromised information, whether that is distance from radar or combination of sidelobes. However, the influence of population density and tornado probabilities on the warning decision-making process remains underexplored.

The primary goal of this study is to explore the relationship between various factors and tornado false alarms, with a focus on radar characteristics and tornado probabilities associated with tornado-warned storms from 2018. This involves investigating population density, radar range, and meteorological quantities using radar data and the tornado probability algorithm (TORP). In addition, the study also assesses whether a radar data threshold should be established on the basis of this relationship.

2. Literature Review

Radar Data

NWS meteorologists primarily use reliable radar data and storm reports from real-time observations to issue tornado warnings (Durage et al. 2013). The Warning Decision Training Division (WDTD), tasked with tornado warning training for NWS forecasters, has compiled its findings to create warning recommendations to get quantitative estimates on tornado probabilities, and most of these recommendations were based on work done by (Gibbs 2016) and (Thompson and Coauthors 2017). Forecasters can have different thresholds and warning philosophies (Karstens and Coauthors 2018), so having research provided with data for radar thresholds can help decrease FAR and increase POD. (Bentley et al. 2021) showed this by creating thresholds for 0.5 elevation scan peak rotational velocity (referred to as V_{rot}) and significant tornado parameter (STP). They found that creating $V_{\text{rot}} \geq 30$ kts and $STP \geq 0$ as a threshold for a tornado warning would have led to more warnings issued and fewer missed events associated with $POD = 0.68$ and $FAR = 0.63$, which are improvements from the 2016–18 database.

Beyond thresholds, all tornado events are different and will continue to pose challenges to NWS meteorologists when issuing warnings. Since the first radar observation from the Weather Surveillance Radar network (WSR-57) in the 1950s (Stout and Huff 1953), radar knowledge of severe weather has evolved, and how this knowledge can be useful for decision-making in issuing warnings. Early signs of this knowledge came from distinguishing features of a tornadic storm, including reflectivity (Z_H), hook echo signature, radial velocity (V_r), tornadic vortex signature (TVS; (Chisholm 1973; Fujita 1973; Burgess et al. 1975; Brown et al. 1978; Markowski 2002; Brown and Wood 2012)). With the new addition of WSR-88D radars, upgrades have been made to these radars, including the implementation of enhanced resolution for data ((Brown et al. 2002, 2005; Torres and Curtis 2007)) and dual-polarization ((Istok and Coauthors 2009; Saxion and Ice 2012)). With these upgrades and advancements in Doppler radar, (Ryzhkov et al. 2005) developed the tornadic debris signature (TDS), characterizing an area of low correlation coefficient (ρ_{HV}) and differential reflectivity (Z_{DR}) affiliated with a TVS.

TORP Algorithm

As the previous discussion shows, today's NWS forecasters synthesize an abundance of radar data in real-time, and investigating these products can become a distraction to the forecaster trying to issue a warning (Boustead and Mayes 2014). To address this issue, TORP was developed to synthesize various radar data variables to enhance confidence in warning decisions (Sandmæl and Coauthors 2023). TORP identifies objects using radar data from the 0.5° tilt, applying a 0.006 s^{-1} azimuthal shear (AzShear) threshold. AzShear, a derivative of velocity along the radar beam, is derived using linear least squares and quantifies rotation. It then extracts radar data from the center of the object's AzShear maximum to calculate and display the tornado probability, aiding the tornado warning decision-making process (Sandmæl and Coauthors 2023).

External Factors

Even with algorithmic support for tornado warning decision-making, external factors still play a critical role in shaping how storms are assessed and whether warnings are issued. Warning severe weather is a mix of art and science, where forecasters assemble data provided to them via technology (Daipha 2015). But, even with the advanced technology we have today, it still cannot predict the weather perfectly (Hoffman et al. 2017). (Kim et al. 2022)

showed that meteorologists make decisions from meteorological factors, but also from personal factors. Forecasters' personal factors can include a variety of thresholds that they perceive as good for tornado warning of a storm, which vary from forecaster to forecaster. Forecasters had a significant dependency on the radar velocity couplet in decision-making. In addition, storm reports from storm chasers can also influence decision-making, depending on whether the forecaster believes it is considerable (Kim et al. 2022). Forecasters can also make decisions on a storm-by-storm basis. Forecasters are less likely to warn the first tornado of each storm, leading to a shorter lead time than the subsequent tornadoes for the same storm. In addition, storms that produce only one tornado have a POD = 0.568 with a lead time of 15 min, making them poorly warned (Chamberlain et al. 2023). Watches produced by the Storm Prediction Center (SPC) are in coordination with local NWS Weather Forecast Offices (WFOs) that can be Severe, Tornado, or Particularly Dangerous Situation (PDS). Operational warning decisions are sometimes guided by upstream products from the SPC. The separation between watch types is clear across all types, PDS having the highest POD and Success Ratio (SR), while no watch has the lowest POD and SR (Krocak and Brooks 2021).

The distance from a WSR-88D radar significantly impacts the likelihood of a tornado warning being issued. Tornadoes occurring more than 100 km from the radar are less likely to be warned. Even when tornadoes move into densely populated areas but remain far from radar, the warning failure rate remains high, about 35%. This suggests that tornadoes in sparsely populated regions, especially those far from radar coverage, are even less likely to be detected and warned (Brotzge and Erickson 2010).

3. Data & Methods

To examine the relationship between meteorological and external factors contributing to tornado false alarms, this study utilizes multiple databases for analysis. Storm Data from the National Centers for Environmental Information (NCEI) was used for tornado warning verification. Tornado warning data, including polygon shapes, issuance and expiration times, and warning text, were obtained from the NWS. This paper uses a database comprising all 2,010 tornado warnings issued by the NWS in 2018. Additionally, population resolution data were sourced from the National Aeronautics and Space Administration (NASA) Socioeconomic Data and Applications Center (SEDAC). The following results were obtained using this methodological framework and provide insight into what influences forecasters' decision-making for warning tornadoes.

a. Population

The SEDAC database includes the area covered by the warning in kilometers and the population inside the warning. The population count has three resolutions: 30-second, 2.5-minute, and 15-minute, with the 30-second being the highest resolution and the 15-minute being the lowest resolution. Population Density is calculated as:

$$\frac{\text{30-second resolution}}{\text{Area Covered by the Warning}}$$

. The 30-second resolution will be used as the population within the warning polygon, since it is the finest resolution and gives the most accurate estimation of the population count. Population Density will be separated into three percentiles: Bottom 33% (< 7 people/km^2), Middle 33% (7–26 people/km^2), and Top 33% (> 26 people/km^2). There are 688 warnings in the Bottom 33%, 665 in the Middle 33%, and 677 in the Top 33%.

b. Performance Metrics

Tornado warnings can either verify (tornado occurred within the warning) or not verify (tornado did not occur within the warning). To evaluate warning performance, we will use POD, FAR, SR, bias, and critical success index (CSI) for verification methods. POD quantifies the fraction of events that were successfully warned of all warnings issued and is defined as:

$$\text{POD} = \frac{H}{H+M}$$

where H is the number of hits (verified warnings) and M is the number of misses (tornadoes with no warning).

In addition, POD can be categorized into two subsets: POD_1, when the warning is issued before the tornado occurs, and POD_2, when a warning is issued after the tornado occurs but before it ends (Brooks and Correia 2018).

In 2018, there were a total of 428 misses accounted for, but 325 were used for the population density warning verification. The other 103 misses did not have population density recorded, so we were not able to correlate them to the population density thresholds.

FAR measures the ratio of warnings that were not verified and is defined as:

$$\text{FAR} = \frac{F}{H+F}$$

where F is the number of false alarms (warnings were issued, but nothing occurred).

SR measures the ratio of forecasted warnings that were successfully verified, defined as:

$$\text{SR} = 1 - \text{FAR}$$

Bias reflects whether tornado occurrences were over-predicted or under-predicted by the forecast. A bias value greater than 1.0 indicates over-prediction, while a value less than 1.0 indicates under-prediction. It is defined as:

$$\text{Bias} = \frac{H+F}{H+M}$$

CSI is often referred to as the accuracy score for the forecast, and measures the hits to the total number of predictions excluding true negatives (TN). It is defined as:

$$\text{CSI} = \frac{H}{H+F+M}$$

The base rate, often referred to as the "random forecast", measures the proportion of observed events relative to the total number of forecast opportunities. It represents the success rate a forecaster would achieve by issuing a "yes" warning for every event. The base rate is defined as:

$$\text{Base Rate} = \frac{H+M}{H+F+M+TN}$$

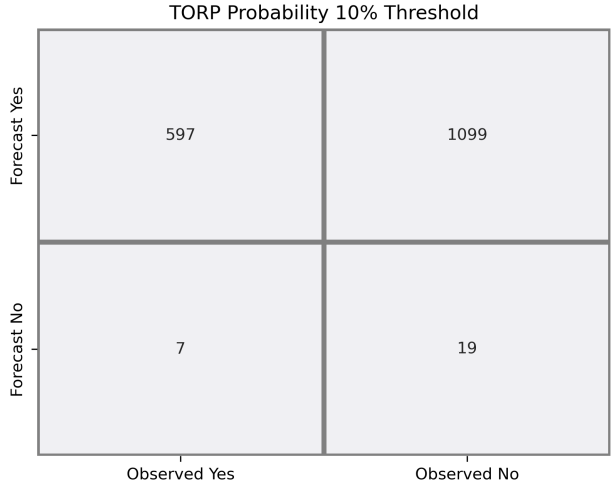

FIG. 1. Contingency table for the 10% TORP threshold, showing the number of hits, false alarms, misses, and true negatives based on this threshold.

All of our threshold findings used a 2×2 contingency table, where a threshold was set and any value at or above it was considered a warning issued, while anything below it was not. These contingency tables allowed us to calculate all of the performance metrics mentioned earlier. A sample contingency table for the 10% TORP threshold is shown in Fig. 1. A hit occurs when a forecaster issues a tornado warning (forecast yes) and a tornado occurs (observed yes). A false alarm occurs when a forecaster issues a tornado warning, but no tornado occurs (observed no). A miss occurs when a forecaster does not issue a tornado warning (forecast no) and a tornado does occur (observed yes). Lastly, a true negative occurs when a forecaster does not issue a tornado warning and no tornado occurs.

c. TORP

Although TORP was not operationally available to forecasters in 2018, it was used in this study to assess detection thresholds and match warning decisions. NWS warning polygons were matched with associated TORP detections (Sandmæl and Coauthors 2023). TORP uses a 0.006 s^{-1} AzShear threshold to create a detection, and not every warning had a TORP detection. There is also the possibility of multiple TORP detections within a warning polygon. Within this study period of 2018, 1,722 warnings were associated with TORP detections near or before the warning. TORP detections immediately before warning issuance were given preference to provide a better comparison to what the forecaster had access to before issuing the warning. In addition, 288 warnings did not have TORP detections due to the rotation being under 0.006 s^{-1} AzShear threshold (Sandmæl and Coauthors 2023) or radar data errors. To connect the closest TORP detection before or near when the warning was issued, a code was developed to match warnings and detections using their shared identifiers. Each warning and TORP detection includes a unique ID consisting of the issuing WFO and the warning order number. Once matches are made, the closest TORP detection will be based on the closest time difference between the timestamp of the TORP detection and the warning. If multiple detections were found in the same location, the detection with the highest AzShear value was selected.

4. Results

Notice, the threshold results only use warnings that had TORP detections since radar data is available for before or near when the warning was issued. In addition, these threshold results use the contingency table to create thresholds from hits, false alarms, misses, and true negatives.

These thresholds are plotted on a performance diagram where SR is on the x-axis, and as you increase SR, warnings become more precise. POD is on the y-axis, and as you increase in POD, warnings catch more of the tornado events that occurred. CSI is a gradient on this diagram, where as you go to the top right corner, warnings are more accurate and bias lines are indicated by the dashed lines plotted diagonally, where above 1 is an indication of over-prediction, while less than 1 is under-prediction.

A TORP threshold of 10% aligns closely with the base rate, where the SR is approximately 0.35. As the TORP threshold increases, SR generally increases while the POD decreases (Fig. 4). At the 20% threshold, performance is slightly above the base rate, with a POD near 1.0 due to the low threshold allowing for nearly all events to be detected. However, as the threshold increases to 30% and 40%, performance begins to diverge more noticeably from the base rate line. Specifically, the 30% TORP threshold has a POD of approximately 0.8 and an SR near 0.4, while the 40%

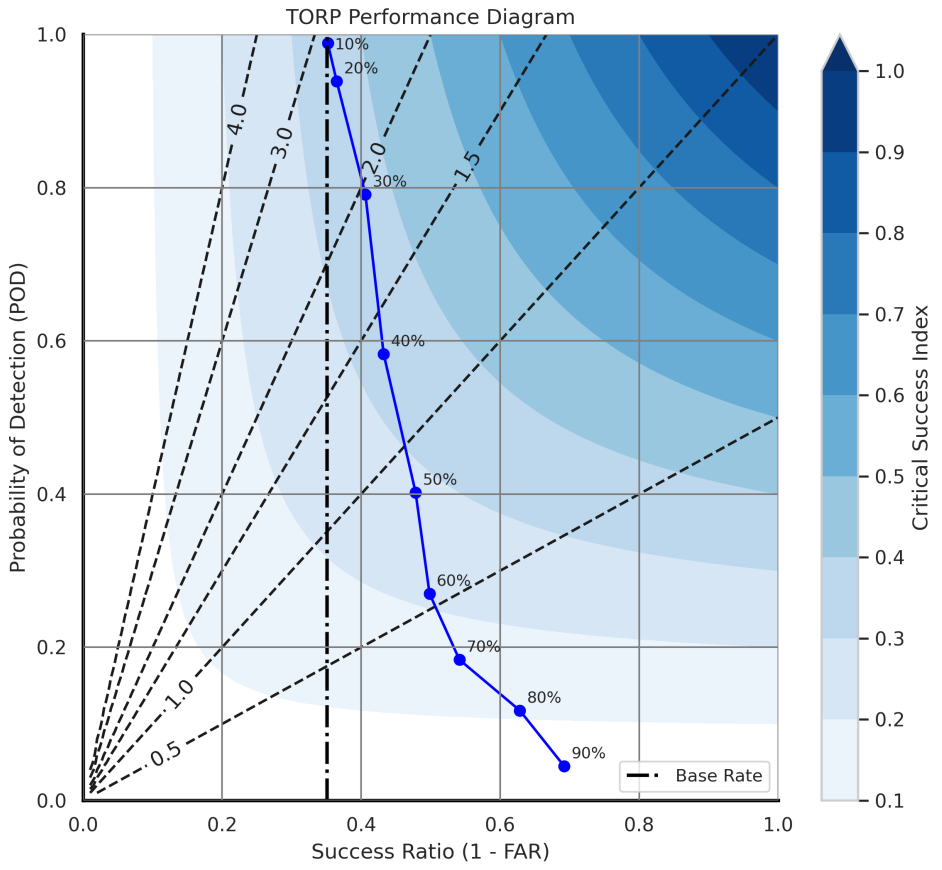

FIG. 2. Performance diagram for TORP detection thresholds in 2018. Each blue dot represents a different threshold for TORP, and the blue line represents TORP's overall performance in 2018. Base rate is the dashed dotted line to show the bare minimum success ratio for forecasters if every storm were warned.

threshold further increases SR beyond 0.4 but sees a decrease in POD to around 0.6. Both the 30% and 40% thresholds fall within the 0.3 Critical Success Index (CSI) contour. None of the TORP thresholds exceed the 0.3 CSI contour, but the 30% threshold is closest to the 0.4 contour. When compared to the 2018 overall performance using TORP detections (shown in Fig.. 4), the 30% and 40% thresholds offer improved performance. The overall TORP performance in 2018 had a POD near 0.5 and SR around 0.35, both of which are exceeded by the 30% and 40% threshold values (Fig. 4).

The 0.006 s^{-1} AzShear threshold was not plotted in Fig. 4. The 0.007 s^{-1} and 0.008 s^{-1} AzShear thresholds are barely ahead of the base rate; the SR of the base rate is 0.35. As we move towards the 0.009 s^{-1} threshold, the threshold starts to move away from the base rate. Using 0.010 s^{-1} AzShear threshold, POD is near 0.65 and SR is above 0.4. A 0.011 s^{-1} AzShear threshold decreases in POD significantly compared to the 0.010 s^{-1} threshold, but the SR increases. As the AzShear threshold increases, SR increases, and POD decreases (Fig. 4).

V_{rot} is the gate-to-gate shear between the maximum outbound and inbound velocity, where the thresholds are in knots. Shown in Fig. 4, each V_{rot} threshold is ahead of the base rate, with some thresholds approaching the base rate as we keep increasing the threshold. Previous research by Gibbs (2016) and Thompson (2017) uses V_{rot} as a threshold for 30 kts and $STP \geq 0$. Looking at warnings in 2018, when TORP had a detection before or near the warning, the 30kts threshold has a POD of 0.7, and SR is near 0.4, but not a huge difference between the base rate and the threshold (Fig. 4). As shown in Figs. 1 and 4, increasing the threshold for the radar variable results in a decrease in POD and an increase in SR. However, in Fig. 4, for V_{rot} thresholds, the values begin to curve back toward the base rate. The 40 kts and 45 kts V_{rot} thresholds are decreasing

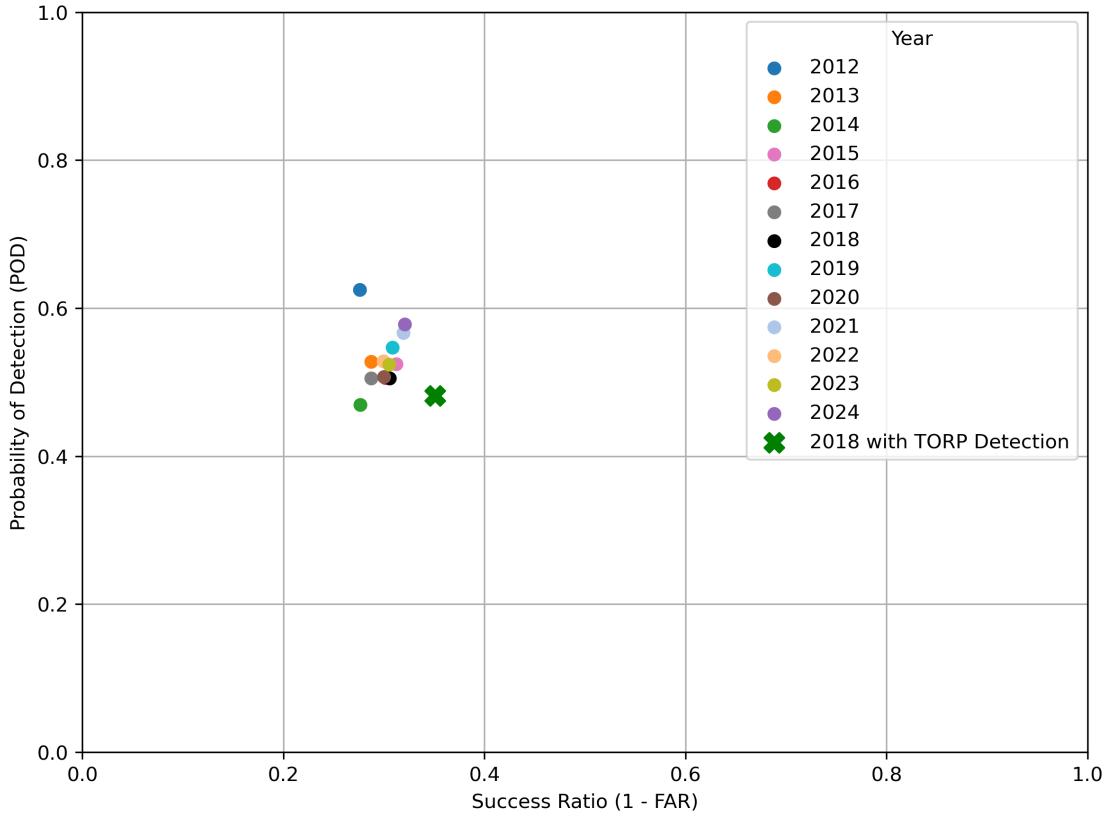

FIG. 3. Scatterplot for Overall NWS Performance from 2012 - 2024, including 2018 Overall Performance only using TORP detections, where SR is on the x-axis and POD is on the y-axis.

in POD, but also decreasing SR as well. POD for 40 kts is near 0.3, and SR is near 0.4. For 45 kts, POD is near 0.2 and SR is near 0.4 as well.

Each population density percentile subset includes values for both POD_1 and POD_2, as defined in the Methods section. The Bottom 33% has the lowest POD values among all subsets, with POD_1 near 0.6 and POD_2 near 0.8 (Fig. 4). The Middle 33% shows the highest POD values, with POD_1 approaching 0.8 and POD_2 near 0.9. In contrast, the Top 33% has the highest SR, with values approaching 0.4. In terms of CSI contours, the Top 33% falls within the 0.3 CSI contour, while both the Middle and Bottom 33% percentiles are within the 0.2 CSI contour. When compared to the bias reference line, all percentiles lie near or above a bias of 2.0.

5. Discussion

Warnings inherently involve societal factors that can determine people's decision-making and interpretation of information (Schumacher et al. 2010). These same factors may also influence forecaster behavior. Additionally, weak tornadoes in rural areas often go undetected or are much more difficult to confirm (Brotzge and Erickson 2010). Forecasters may be influenced by second opinions, such as a reliable report from a storm spotter or input from another forecaster, especially when the decision is difficult. These inputs can prompt the issuance of a warning (Kim et al. 2022).

Even though false alarms likely do not lead to mistrust in the NWS (Ripberger et al. 2015), warnings still impact many and have lasting effects on communities. Warnings are a complex system that involves prediction, detection, decision, dissemination, and response. The effectiveness of the warning depends on these factors, as it creates a standard of its capability to alert people affected. In an era of improved technology, investments should be strategically focused on creating highly efficient tornado warnings (Brotzge and Donner 2013). This research tends to look at other possible factors in the warning decision-making that are unexplored or what previous research was looking ahead to.

a. Main Findings

Using TORP between 30% and 40% seems to be a good threshold for forecasters' decision-making when compared to the 2018 overall performance with TORP in Fig. 4. The 30% and 40% TORP thresholds are approaching the 0.4 CSI contour, indicating improved performance

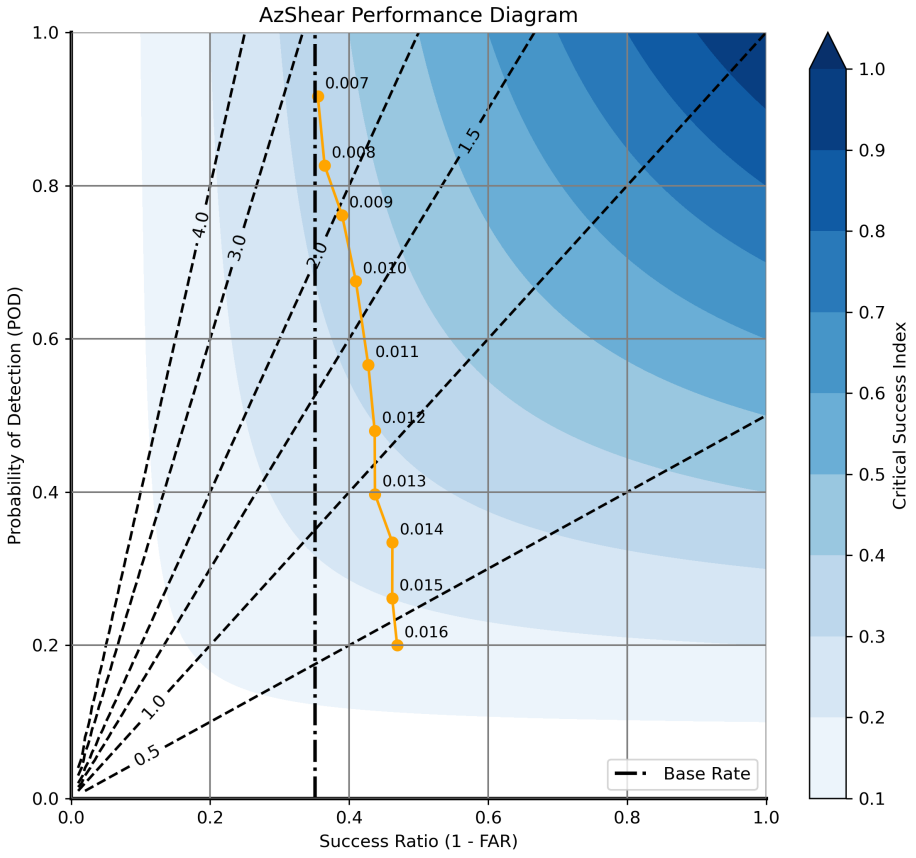

FIG. 4. Performance diagram for AzShear detection thresholds in 2018. Each orange dot represents a different threshold for AzShear, and the orange line represents AzShear threshold overall performance in 2018.

and greater forecast skill. In addition, while comparing to overall performance when only using TORP detections (Fig. 4), the TORP 30-40% performance is better. Even though the frequency bias will be twice as much as a tornado occurring, forecasters are still validating a miss as costly (Fig. 4). 10% shows no use since it is on the base rate, and 20% is extremely near it, but there is an exponential increase for SR as you increase the threshold for TORP, and whether to warn based on those thresholds.

TORP is based on a 0.006 s^{-1} AzShear threshold, where it can only create a detection whenever that threshold is reached, which is why 0.006 s^{-1} is not shown in Fig. 4. 0.007 s^{-1} is on the base rate, showing that a threshold of 0.007 s^{-1} will be worse than a random forecast. From a forecaster standpoint, a threshold can be made for a 0.010 s^{-1} AzShear where the POD is around 0.65 and SR is 0.4, giving forecasters a higher probability of forecasting an accurate tornado warning. When comparing it to the overall performance only using TORP detections (Fig. 4), a 0.010 s^{-1} AzShear threshold is more accurate. An Azshear threshold of 0.011 s^{-1} could also be used due to the frequency bias being lower and SR being higher, but POD below 0.6 will become a concern (Fig. 4). As you keep increasing the thresholds, SR increases, but POD decreases due to being very selective and warning less until there is overwhelming evidence.

V_{rot} has been used as a threshold for forecasters since its research by Gibbs (2016) and Thompson (2017), where 30 kts and an $STP \geq 0$ should be considered for a warning. Looking at Fig. 5, 30 kts is in the .3 CSI contour, where it has a high POD and near a 0.4 SR, showing this threshold should not be changed. Compared to the overall 2018 performance using only TORP detections (Fig. 4), the 30 kt threshold (Fig. 4) shows improved POD and SR. In addition, the V_{rot} curve goes back near the base rate (Fig. 4), due to receiving more misses and fewer hits, decreasing POD and SR.

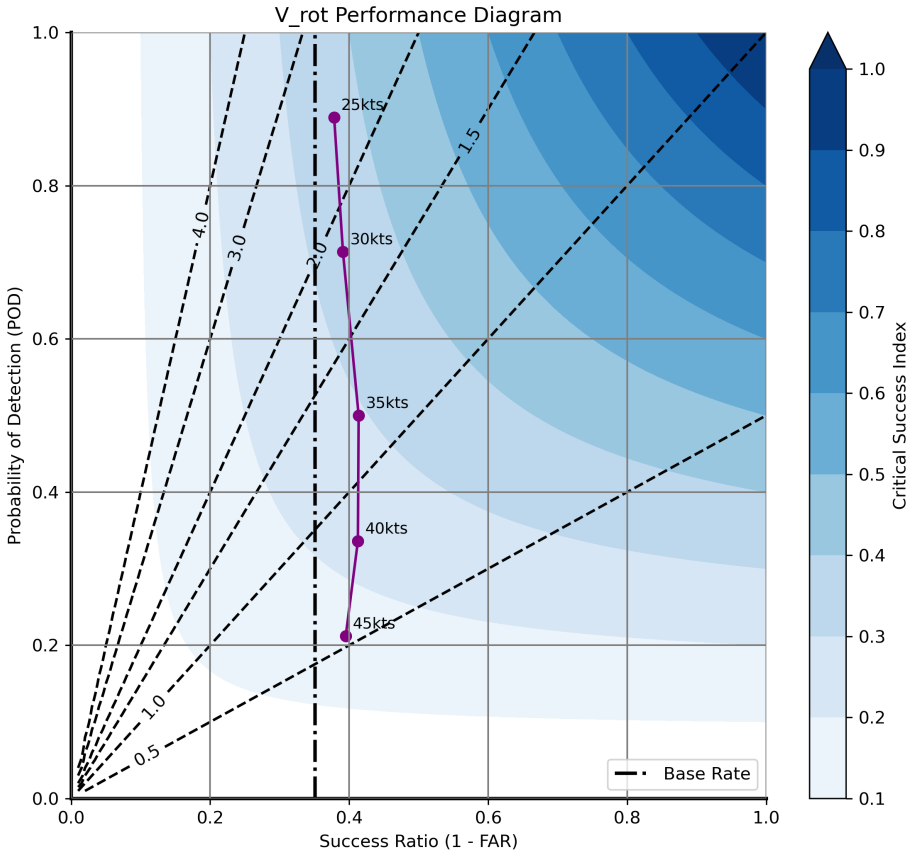

FIG. 5. Performance diagram for V_{rot} detection thresholds in 2018. Each purple dot represents a different threshold for V_{rot}, and the purple line represents V_{rot} threshold overall performance in 2018.

This research also aimed to explore whether population density had any effect on tornado warning decision-making, and as shown in Fig. 4, there is an apparent relationship. As we increase in population density, FAR decreases, making the SR increase. From the Bottom 33% POD_1, we see a substantial POD and SR difference compared to the Middle and Top 33% (Fig. 4). A possible conclusion to this substantial difference is that there are fewer reports for the forecaster. Without having that second opinion, fewer reports can lead to POD decreasing and an increase in the FAR. In addition, another possible conclusion is that these are mostly farmland areas that affect fewer people. With it affecting fewer people, forecasters might just warn of it and worry less about false alarms since fewer people are being affected by this storm. This is vice versa for the more densely populated areas, where forecasters have more reports and care about decreasing false alarms in more populated areas, which helps SR. The Middle 33% have a higher POD_1 and POD_2 than any of the percentiles (Fig. 4). This conclusion is still unknown as to why this is happening, but possibly due to the sample size. POD_1 has a lower POD for all of the percentiles since it is whenever a warning came before the tornado occurred. POD_2 has a higher POD due to its definition of warning of a storm whenever the tornado is in progress or before it ends, so it is more likely to detect a tornado. Overall, forecasters think misses are costly relative to false alarms due to the frequency bias being above or near 2.0 (Fig. 4). Lastly, the base rate could be changing with each population density threshold, but this limitation will be discussed later in the paper.

b. Limitations

While this study gives implications on the various factors discussed in the forecaster's decision-making, multiple limitations need to be addressed. For instance, TORP cannot be compared to NWS performance in 2018 due to the 0.006 s^{-1} AzShear threshold that TORP has to create

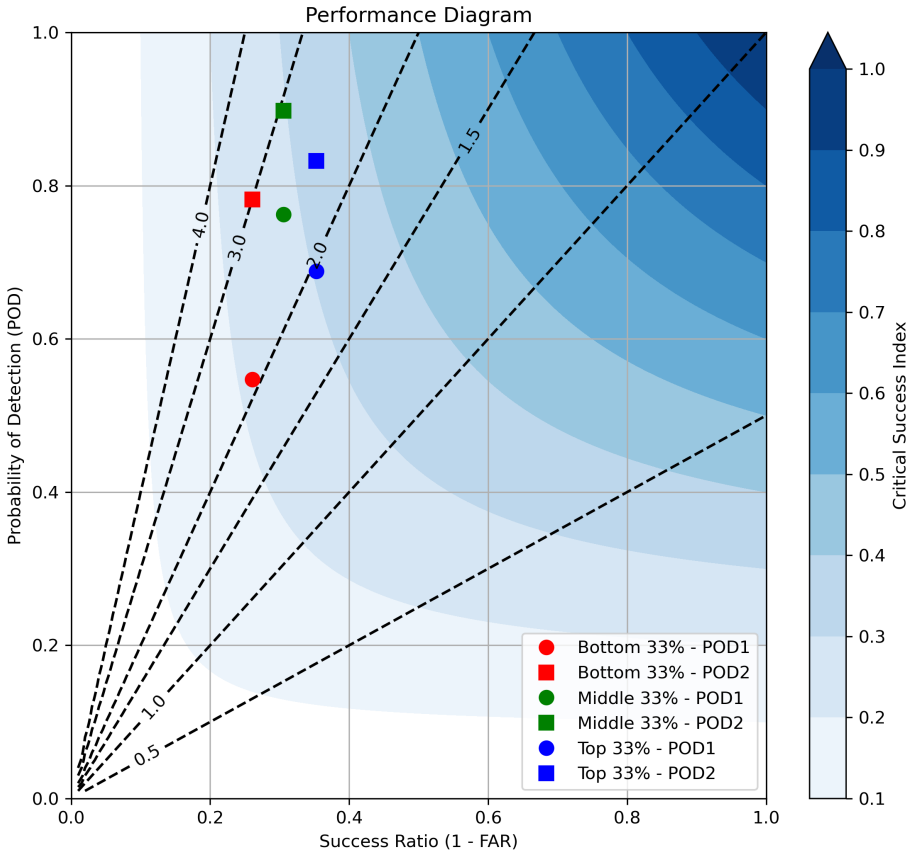

FIG. 6. Performance diagram for population density percentiles in 2018. Each color represents a different 33rd percentile for population density. The circles are for POD_1, and the squares are for POD_2.

any detection. Anything below that threshold will not have a TORP detection due to its weak rotation, where a bias could occur between the NWS and TORP warning performance. Comparing the two will not give any significant evidence of whether TORP is outperforming the NWS or vice versa.

Second, the 0.006 s^{-1} AzShear threshold used in TORP may exclude false alarm cases that warrant further study, such as those involving tropical cyclone tornadoes, where false alarms are common. Additionally, if TORP fails to run or radar data is missing, no radar input is available near the time of warning issuance.

Third, as discussed previously in the section above, the base rate for population density cannot be calculated due to the dataset that was being used did not truly have a true negative case. Options were available to decide what could be considered a true negative; for example, severe thunderstorm (SVR) warnings that did not produce a tornado could be considered a true negative. This would sway the base rate to an extremely low number due to the large amount of SVR warnings issued in 2018. But, it can be done with the NASA SEDAC population database and a warning polygon to be able to compute population density to see the count of true negatives.

6. Summary and conclusions

A total of 2,010 tornado warnings issued by the NWS in 2018 were examined, where 1,722 were used in performance metrics for thresholds due to having a TORP detection before or near when the warning was issued and having radar data associated with them. The rest of the warnings did not have a TORP detection associated with them due to not meeting the 0.006 s^{-1} AzShear threshold. Various factors were analyzed to see if there is a relationship to better understand false alarm cases in tornado warnings, for example, TORP, AzShear, Vrot, and population density. In summary, the most important takeaways from this study were the following:

- TORP around 30% - 40% is a good threshold to use for issuing warnings with POD ranging from 0.6 - 0.8 and SR being above 0.4. In addition, 0.010 s^{-1} AzShear threshold can be made to forecasters as a recommendation.

- Previous research by (Gibbs 2016) and (Thompson and Coauthors 2017) has made 30 kts V_{rot} another recommendation for forecasters, which has been proven successful within 2018 NWS warnings before it was operational to forecasters.

- Population density shows a clear relationship with warning performance, where higher density often corresponds to better apparent performance. In sparsely populated areas, fewer reports can lead to higher FAR, and forecasters may be less concerned about false alarms due to the smaller number of people affected, while vice versa for densely populated areas.

Future studies should consider other societal factors: social vulnerability and demographics, which could influence the decision-making regarding tornado warnings. In addition, focusing more on cases where false alarms are the greatest, for example, weaker (EF0-1) tornadoes, or when a TORP detection did not occur due to the AzShear threshold. Furthermore, another useful thing that can be studied is possibly if the population density had different base rates, and it is possible to calculate the base rate by finding the true negatives.

With the advancement of technologies like phased-array radar (PAR), which offers rapid updates, full-volume scans, forecasters gain increased confidence in making timely and accurate warning decisions (Heinselman et al. 2015). As such, continued integration of technology and high-resolution radar data stands to enhance the effectiveness of tornado warning operations. TORP aims to serve as a guidance tool for forecasters to enhance warning performance (Sandmæl and Coauthors 2023), and although it was not operational in 2018, its capability to filter and prioritize radar data shows potential in reducing information overload and supporting more effective decision-making by forecasters.

Acknowledgments. This work was generously supported by funding from NSF AGS-20250267, and by the NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA11OAR4320072, U.S. Department of Commerce. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, NOAA, or the U.S. Department of Commerce. We would like to thank Daphne LaDue (CAPS) and Alex Marmo (CAPS) for their guidance and support throughout this project and the NWC REU. Additional thanks to the NSSL, CIWRO, and NOAA for their institutional support. We also acknowledge the Tornado Probability Algorithm (TORP), developed by Thea Sandmæl, which played a significant role in guiding the direction of this research.

References

Bentley, E. S., R. L. Thompson, B. R. Bowers, J. G. Gibbs, and S. E. Nelson, 2021: An analysis of 2016–18 tornadoes and national weather service tornado warnings across the contiguous united states. *Weather and Forecasting*, **36**, 1909–1924, doi:10.1175/WAF-D-20-0241.1.

Bostrom, A., R. E. 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. *Weather, Climate, and Society*, doi:10.1175/WCAS-D-15-0033.1, in press.

Boustead, J. M., and B. Mayes, 2014: The role of the human in issuing severe weather warnings. *27th Conference on Severe Local Storms*, American Meteorological Society, Madison, WI, URL https://ams.confex.com/ams/27SLS/webprogram/Paper254547.html, 4B.2.

Brooks, H. E., 2004: Tornado-warning performance in the past and future: A perspective from signal detection theory. *Bulletin of the American Meteorological Society*, **85**, 837–843, doi:10.1175/BAMS-85-6-837.

Brooks, H. E., and J. J. Correia, 2018: Long-term performance metrics for national weather service tornado warnings. *Weather and Forecasting*, **33**, 1501–1511, doi:10.1175/WAF-D-18-0120.1.

Brooks, H. E., and C. A. I. Doswell, 2002: Deaths in the 3 may 1999 oklahoma city tornado from a historical perspective. *Weather and Forecasting*, **17**, 354–361, doi:10.1175/1520-0434(2002)017⟨0354:DITMOC⟩2.0.CO;2.

Brotzge, J., and W. Donner, 2013: The tornado warning process: A review of current research, challenges, and opportunities. *Bulletin of the American Meteorological Society*, **94**, 1715–1733, doi:10.1175/BAMS-D-12-00147.1.

Brotzge, J., and S. Erickson, 2010: Tornadoes without nws warning. *Weather and Forecasting*, **25**, 159–172, doi:10.1175/2009WAF2222270.1.

Brown, L. R., B. A. Flickinger, E. Forren, D. M. Schultz, D. Sirmans, P. L. Spencer, V. T. Wood, and C. L. Ziegler, 2005: Improved detection of severe storms using experimental fine-resolution wsr-88d measurements. *Weather and Forecasting*, **20**, 3–14, doi:10.1175/WAF-832.1.

Brown, L. R., L. R. Lemon, and D. W. Burgess, 1978: Tornado detection by pulsed doppler radar. *Monthly Weather Review*, **106**, 29–38, doi:10.1175/1520-0493(1978)106,0029:TDBPDR.2.0.CO;2.

Brown, L. R., V. T. Wood, and D. Sirmans, 2002: Improved tornado detection using simulated and actual wsr-88d data with enhanced resolution. *Journal of Atmospheric and Oceanic Technology*, **19**, 1759–1771, doi:10.1175/1520-0426(2002)019,1759:ITDUSA.2.0.CO;2.

Brown, R. A., and V. T. Wood, 2012: The tornadic vortex signature: An update. *Weather and Forecasting*, **27**, 525–530, doi:10.1175/WAF-D-11-00111.1.

Burgess, D. W., L. R. Lemon, and R. A. Brown, 1975: Tornado characteristics revealed by doppler radar. *Geophysical Research Letters*, **2**, 183–184, doi:10.1029/GL002i005p00183.

Center for International Earth Science Information Network - CIESIN - Columbia University, 2018: Population estimation service, version 3 (pes-v3). Palisades, NY, accessed 19 September 2024, https://doi.org/10.7927/H4DR2SK5.

Chamberlain, J. U., M. D. Flournoy, M. J. Krocak, H. E. Brooks, and A. K. Anderson-Frey, 2023: Analyzing tornado warning performance during individual storm life cycles. *Weather and Forecasting*, **38**, 773–785, doi:10.1175/WAF-D-22-0142.1.

Chisholm, A. J., 1973: *Alberta hailstorms Part I: Radar case studies and airflow models*. No. 14, Meteorological Monographs, American Meteorological Society, 1–36 pp.

Daipha, P., 2015: *Masters of Uncertainty: Weather Forecasters and the Quest for Ground Truth*. University of Chicago Press, 280 pp.

Doswell, C. A. I., 2004: Weather forecasting by humans—heuristics and decision making. *Weather and Forecasting*, **19**, 1115–1126, doi:10.1175/WAF-821.1.

Durage, S. W., S. C. Wirasinghe, and J. Ruwanpura, 2013: Comparison of the canadian and us tornado detection and warning systems. *Natural Hazards*, **66**, 117–137, doi:10.1007/s11069-012-0168-7.

Fujita, T. T., 1973: Proposed mechanism of tornado formation from rotating thunderstorms. *Eighth Conference on Severe Local Storms*, American Meteorological Society, Denver, CO, 191–196.

Gibbs, J. G., 2016: A skill assessment of techniques for real-time diagnosis and short-term prediction of tornado intensity using the wsr-88d. *Journal of Operational Meteorology*, **4**, 170–181, doi:10.15191/nwajom.2016.0413.

Heinselman, P. L., D. S. LaDue, D. M. Kingfield, and R. Hoffman, 2015: Tornado warning decisions using phased-array radar data. *Weather and Forecasting*, **30**, 57–78, doi:10.1175/WAF-D-14-00042.1.

Hoffman, R. R., D. S. LaDue, M. Mogil, P. J. Roebber, and J. G. Trafton, 2017: *Minding the Weather: How Expert Forecasters Think*. MIT Press, 470 pp.

Istok, M. J., and Coauthors, 2009: Wsr-88d dual polarization initial operational capabilities. *25th Conference on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, American Meteorological Society, Phoenix, AZ, URL https://ams.confex.com/ams/pdfpapers/148927.pdf, 15.5.

Karstens, C. D., and Coauthors, 2018: Development of a human machine mix for forecasting severe convective events. *Weather and Forecasting*, **33**, 715–737, doi:10.1175/WAF-D-17-0188.1.

Kim, J., A. A. Seate, B. F. Liu, D. Hawblitzel, and T. Funk, 2022: To warn or not to warn: Factors influencing national weather service warning meteorologists' tornado warning decisions. *Weather, Climate, and Society*, **14**, 697–708, doi:10.1175/WCAS-D-20-0115.1.

Krocak, M. J., and H. E. Brooks, 2021: The influence of weather watch type on the quality of tornado warnings and its implications for future forecasting systems. *Weather and Forecasting*, **36**, 1675–1680, doi:10.1175/WAF-D-21-0052.1.

Markowski, P. M., 2002: Hook echoes and rear-flank downdrafts: A review. *Monthly Weather Review*, **130**, 852–876, doi:10.1175/1520-0493(2002)130⟨0852:HEARFD⟩2.0.CO;2.

NOAA/National Weather Service, 1950: Storm events database. Accessed May 2025, https://www.ncdc.noaa.gov/stormevents.

Ripberger, J. T., C. L. Silva, H. C. Jenkins-Smith, D. E. Carlson, M. James, and K. G. Herron, 2015: False alarms and missed events: The impact and origins of perceived inaccuracy in tornado warning systems. *Risk Analysis*, **35**, 44–56, doi:10.1111/risa.12262.

Ryzhkov, A. V., T. J. Schuur, D. W. Burgess, and D. S. Zrnic, 2005: Polarimetric tornado detection. *Journal of Applied Meteorology and Climatology*, **44**, 557–570, doi:10.1175/JAM2235.1.

Sandmæl, T. N., and Coauthors, 2023: The tornado probability algorithm: A probabilistic machine learning tornadic circulation detection algorithm. *Weather and Forecasting*, **38**, 445–466, doi:10.1175/WAF-D-22-0123.1.

Saxion, D. S., and R. L. Ice, 2012: New science for the wsr-88d: Status of the dual-polarization upgrade. *28th Conference on Interactive Information Processing Systems*, American Meteorological Society, New Orleans, LA, Vol. 5, URL https://ams.confex.com/ams/92Annual/webprogram/Manuscript/Paper197645/NEXRAD_DP_Status_28th_IIPS_Jan2012.pdf.

Schumacher, R. S., D. T. Lindsey, A. B. Schumacher, 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. *Weather and Forecasting*, **25**, 1412–1429, doi:10.1175/2010WAF2222396.1.

Stout, G. E., and F. A. Huff, 1953: Radar records illinois tornadogenesis. *Bulletin of the American Meteorological Society*, **34**, 281–284.

Thompson, R. L., and Coauthors, 2017: Tornado damage rating probabilities derived from wsr-88d data. *Weather and Forecasting*, **32**, 1509–1528, doi:10.1175/WAF-D-17-00.

Torres, S. M., and C. D. Curtis, 2007: Initial implementation of super-resolution data on the nexrad network. *21st International Conference on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology*, American Meteorological Society, San Antonio, TX, URL https://ams.confex.com/ams/pdfpapers/116240.pdf, 5B.10.