

## FORECASTERS' FORUM

### Simulating Tornado Probability and Tornado Wind Speed Based on Statistical Models

ARIEL E. COHEN<sup>a</sup>

*NOAA/NWS/NCEP/Storm Prediction Center, Norman, Oklahoma*

JOEL B. COHEN

*NiSource, Inc., Columbus, Ohio*

RICHARD L. THOMPSON AND BRYAN T. SMITH

*NOAA/NWS/NCEP/Storm Prediction Center, Norman, Oklahoma*

(Manuscript received 20 November 2017, in final form 30 March 2018)

#### ABSTRACT

This study presents the development and testing of two statistical models that simulate tornado potential and wind speed. This study reports on the first-ever development of two multiple regression-based models to assist warning forecasters in statistically simulating tornado probability and tornado wind speed in a diagnostic manner based on radar-observed tornado signature attributes and one environmental parameter. Based on a robust database, the radar-based storm-scale circulation attributes (strength, height above ground, clarity) combine with the effective-layer significant tornado parameter to establish a tornado probability. The second model adds the categorical presence (absence) of a tornadic debris signature to derive the tornado wind speed. While the fits of these models are considered somewhat modest, their regression coefficients generally offer physical consistency, based on findings from previous research. Furthermore, simulating these models on an independent dataset and other past cases featured in previous research reveals encouraging signals for accurately identifying higher potential for tornadoes. This statistical application using large-sample-size datasets can serve as a first step to streamlining the process of reproducibly quantifying tornado threats by service-providing organizations in a diagnostic manner, encouraging consistency in messaging scientifically sound information for the protection of life and property.

### 1. Introduction

Recent significant advances have occurred in the meteorological community's ability to reproducibly evaluate and quantify the threat for tornadoes. In an age of increasing need for accurate, high-precision forecasts, this work is becoming increasingly integrated into daily National Weather Service operations. [Smith et al. \(2012\)](#) and [Thompson et al. \(2012\)](#) laid the initial groundwork for relating storm-scale characteristics and near-storm environments to tornado damage ratings with a large sample size, while also providing detailed literature

reviews including references documenting past research that has quantified tornado threat. Within an expansive dataset containing 22 901 tornado and significant severe thunderstorm events, of which 10 753 events were tornadoes, they manually assigned a convective mode to each of the parent storms based on WSR-88D data. Additionally, attributes of the near-storm environment were considered, including the significant tornado parameter (STP; [Thompson et al. 2012](#)), through the use of the Storm Prediction Center (SPC) mesoanalysis system ([Bothwell et al. 2002](#)). The net result of this work has been a better understanding of the relationship between the convective mode and the environment in which the tornadic storms occur. Moreover, geographic variability of these tornado-associated attributes was a major focus of their work.

[Smith et al. \(2015\)](#) extended the aforementioned work forward to deriving conditional probabilities of

<sup>a</sup> Current affiliation: National Weather Service, Topeka, Kansas.

Corresponding author: Ariel E. Cohen, ariel.cohen@noaa.gov

tornado damage ratings—conditional upon the known occurrence of a tornado. This diagnostic approach to quantifying tornado damage rating probabilities takes full advantage of the attributes previously discussed, with the addition of radar-based low-level rotational velocity, and is one of the first known attempts to provide explicit probabilistic information about tornado damage ratings based on the presence of a tornadic storm and its interaction with the background environment. This work provides an important level of quantification for increasing or decreasing confidence in potential tornado impacts from convective elements known to be producing tornadoes. Accordingly, this yields reproducible guidance for forecasters to express varying levels of potential tornado impact in critical warnings and statements based on individual storm-scale and environmental characteristics.

Even more recently, [Thompson et al. \(2017\)](#) demonstrated that tornado damage rating probabilities can be determined by considering many of the storm-scale and environmental attributes that were also documented by [Smith et al. \(2015\)](#). However, using a large dataset of severe weather reports from 2014 to 2015, [Thompson et al. \(2017\)](#) compared attributes associated with tornadic storms to nontornadic severe thunderstorms, yielding the first attempt at quantifying tornado damage ratings, conditional only on the presence of a supercell or quasi-linear convective system (QLCS) with low-level cyclonic rotation. The [Thompson et al. \(2017\)](#) sample allows quantification of tornado damage rating probabilities through varying the peak low-level rotational velocity, height above radar level, circulation diameter, a “clear and/or tight” categorical and binary assessment that is subjectively determined and characterizes the visual clarity and spatial coherency of the circulation, and the presence of a dual-polarization (dual-pol) tornadic debris signature ([WDTD 2016](#)), along with other attributes (e.g., mesoanalysis STP and convective mode). This approach has permitted reproducible tornado threat assessment for severe storms based on storm attributes and environmental conditions. Such findings continue to solidify our understanding of the propensity for ongoing convection to produce tornadoes.

The series of aforementioned studies (i.e., [Smith et al. 2012, 2015](#); [Thompson et al. 2012, 2017](#), and references therein) outline the present state of research on tornado quantification leading up to the present. Meanwhile, the present study continues to extend the foundation that the previously mentioned studies have laid, specifically in terms of providing a multivariable analysis of tornado probabilities and tornado wind speeds based upon damage ratings. In particular, this present work will employ multivariable regression analysis to develop linear statistical models ([Pindyck and Rubinfeld 1981](#)) that

simultaneously combine many of the variables that [Smith et al. \(2015\)](#) and [Thompson et al. \(2017\)](#) found to explain conditional and unconditional tornado probabilities. One statistical model is constructed to simulate tornado probabilities in a diagnostic manner, and another statistical model is constructed to simulate tornado wind speeds, based upon tornado damage ratings. The purpose of these models is to apply the results of the previous studies that concisely combine several factors known to influence tornado probability and rating assessment, in order to provide additional operational insight into assessing tornado potential and impact from severe thunderstorms. By simply providing readily ascertainable information from WSR-88D and the SPC mesoanalysis system as input to these models, statistical simulations of the tornado threat can be reproducibly integrated into critical messaging.

## 2. Methodology

The full dataset of tornado, severe hail, and severe wind events documented by [Thompson et al. \(2017\)](#) was first separated into two datasets: tornado and nontornado severe events. Subsequently, each of the events was assigned a random number, and then random-numbered events were sorted by ascending random number within separate tornado and nontornado datasets. Within the tornado dataset, the first 500 tornado events, sorted by ascending random number, were assigned to an independent tornado dataset. Likewise, within the nontornado dataset, the first 500 nontornado events, sorted by ascending random number, were assigned to an independent nontornado dataset. Cases with missing significant tornado parameter information were removed, leaving a total of 498 tornado events and 499 nontornado events, for a total independent database size of 997 events. The remaining tornado and nontornado events were combined into a large training dataset comprising a total of 5206 events (tornado and severe hail/wind), serving as the basis for the development of the tornado probability model. The 5206 events consist of a total of 1025 tornado events, which are considered separately in the development of a statistical model that simulates tornado wind speeds. As a result, the size of the independent dataset is about 19% that of the training dataset.

The statistical model for simulating tornado wind speed given a tornado is a basic multivariable, linear regression form:

$$y = \left( \sum_{i=1}^m a_i x_i \right) + b, \quad (1)$$

where  $y$  corresponds to the tornado wind speed simulated, and  $a_i$  represents each of the regression coefficients

corresponding to the predictors  $x_i$ , used as inputs for the model. The number of input predictors is referred to as  $m$ , and  $b$  represents the intercept value of the model. An important consideration here is that the regression is performed on the average wind speed for each tornado damage rating. This treats the tornado wind speed as the predictand of the regression, with the regression formulated by averaging the minimum and maximum wind speeds corresponding to the damage rating for each event. As such, enhanced Fujita (EF) tornado damage ratings of 0, 1, 2, 3, 4, and 5 were assigned as tornadoes with wind speeds of  $75 \text{ mi h}^{-1}$  (averaging  $65$  and  $85 \text{ mi h}^{-1}$ ),  $98 \text{ mi h}^{-1}$  (averaging  $86$  and  $110 \text{ mi h}^{-1}$ ),  $123 \text{ mi h}^{-1}$  (averaging  $111$  and  $135 \text{ mi h}^{-1}$ ),  $151 \text{ mi h}^{-1}$  (averaging  $136$  and  $165 \text{ mi h}^{-1}$ ),  $183 \text{ mi h}^{-1}$  (averaging  $166$  and  $200 \text{ mi h}^{-1}$ ), and  $218 \text{ mi h}^{-1}$  (averaging  $201$  and  $234 \text{ mi h}^{-1}$ ), respectively. This is done to ensure that the real-number predictand is indeed consistent with a continuous variable (i.e., wind speed), as opposed to the discrete EF-scale ratings. The simulated EF-scale ratings can then be subsequently determined from the simulated tornado wind speeds, if desired. The aforementioned wind speeds are based on the operational EF scale, with the derived EF scale used only for the upper limit of the EF5 damage rating. Wind speeds are provided in miles per hour ( $\text{mi h}^{-1}$ ) herein owing to the National Weather Service's use of this wind speed scale for operational forecasting means. Note the conversion between miles per hour, knots (kt), and meters per second is as follows:  $1 \text{ mi h}^{-1} = 0.447 \text{ m s}^{-1} = 0.869 \text{ kt}$ .

For simulating tornado probabilities, a binary logistic regression equation is used:

$$y = \frac{1}{1 + \exp\left\{-\left[\left(\sum_{i=1}^m a_i x_i\right) + b\right]\right\}}, \quad (2)$$

where each of the variables used in this regression model has the same meaning as in the linear Eq. (1). The logistic regression analysis was chosen for simulating tornado probabilities because it binds the simulated probabilities between 0 and 1 (multiplied by 100% to yield units of percent) and, more importantly, because this methodology enables the conversion of binary values (0 and 1) to a continuous range of probabilities.

For both tornado probabilities and tornado wind speed simulations, the regressions are performed on the training dataset and then tested on the independent dataset, as a simulation on a mutually exclusive dataset from that which was used to create the model. Both Eqs. (1) and (2) are addressed in numerous statistics textbooks (e.g., [Pindyck and Rubinfeld 1981](#)).

For the case of simulating tornado probabilities, candidate predictors are those individual variables that [Smith et al. \(2015\)](#) and [Thompson et al. \(2017\)](#) found to offer utility in explaining tornado probabilities: height of sampled circulation above radar level (arl, measured in ft), peak average rotational velocity ( $v_{\text{rot}}$ , measured in kt), circulation diameter (dist, measured in n mi;  $1 \text{ n mi} = 1.852 \text{ km}$ ), clear/tight designation (ct), and effective-layer significant tornado parameter (stp). Only one of these predictors, ct, is binary (treated as 0 for a diffuse circulation and 1 for a clear and tight circulation). The probability simulation model is crafted by performing multivariable regression of tornado occurrence on the various predictor inputs from the training dataset. The formulation of this model treats tornado occurrence as binary (0 for a nontornadic event and 1 for a tornadic event). The regression process effectively determines an expected value for the predictand, which corresponds to a probability of tornado occurrence, based on the inputs used in creating the model. Note that the probabilities derived from this model need to be multiplied by 100% to attain units of percent.

For the determination of simulated tornado wind speed, the candidate inputs are the same as those used to simulate tornado probabilities, except the dual-pol tornadic debris signature (tds) variable is added as a predictor for this model. We treat tds as a binary predictor, with possible values of 0 (no tds present) or 1 (tds present). The tornado wind speed model performs multivariable linear regression of actual tornado wind speeds in the training dataset on the various predictor inputs.

The choice to incorporate tds into the statistical model for tornado wind speed, but not for tornado probability, reflects the difference in a priori knowledge of tornado occurrence between the two models. For the tornado probability model, it is assumed that the simulation of tornado probability makes no a priori assumption of tornado occurrence. This effectively generalizes the utility of this model to cases where no prior knowledge of a tornado is available. Furthermore, if a tds<sup>1</sup> were to become evident, that would set the tornado probability to 100% in practice. In the case of the statistical model simulating tornado wind speeds, the regression is performed on known tornado cases and thus this model is applicable to the condition of tornado occurrence being met. The tds can provide real-time evidence of meeting this condition, thus its incorporation in this model.

<sup>1</sup> A tornadic debris signature can persist just beyond the life span of a tornado.

TABLE 1. Regression coefficients and  $P$  values corresponding to multiple regression analysis for deriving the tornado probability statistical model.

	Regression coefficient	$P$ value
$v_{rot}$	0.0552	<0.001
dist	-0.684	<0.001
ct	0.835	<0.001
stp	0.0473	0.058
intercept	-2.51	<0.001

For both the tornado probability and tornado wind speed statistical models, variables that offer little explanatory power are removed from the respective models, before creating final versions of the regression models. This is accomplished by setting a  $P$ -value threshold of 0.10, with predictors offering  $P$  values at or above 0.10 being removed prior to the formulation of the final regression model. Restricting variable incorporation to a  $P$ -value threshold at or below 0.10, while arbitrarily selected, represents a level of at least marginal significance that the variable explains the regression model for which it is being fit. After each of the statistical models is created and finalized with the associated attributes presented, the inputs from the independent dataset are used to statistically simulate the tornado probability and tornado wind speed for each constituent event. Differences between the simulation and what actually occurred (i.e., the value of performing the model simulation on a dataset for which actual outcomes are known) are also subsequently presented. Finally, these statistical models are applied to specific cases addressed by [Thompson et al. \(2017\)](#), and the results are subsequently displayed.

### 3. Results and discussion

The tornado probability model yields the regression coefficients and corresponding  $P$  values listed in [Table 1](#). The  $P$  values are determined using the standard normal distribution ([Pindyck and Rubinfeld 1981](#)), and predictors that are retained following the  $P$ -value-based filtration process specified earlier are  $v_{rot}$ , dist, ct, and stp. The Somers'  $D$  statistic ([Somers 1962](#)) is used as a measure of the goodness of fit owing to its assessment of a probabilistic model, and this statistic can range from -1 to 1. Somers'  $D$  statistic values closer to 1 and -1 imply a better-fitting model, while values closer to 0 imply a worse-fitting model. In this case, the Somers'  $D$  statistic is found to be 0.58. While this value, in this case, may not be considered high or strong, it may still suggest that this regression approach could provide some *practical* value, even if the exact model does not

TABLE 2. Regression coefficients and  $P$  values corresponding to multiple regression analysis for deriving the tornado wind speed statistical model.

	Regression coefficient	$P$ value
arl	$5.96 \times 10^{-4}$	0.009
$v_{rot}$	0.588	<0.001
stp	1.50	<0.001
tds	12.2	<0.001
intercept	60.7	<0.001

offer a perfect statistical fit. The approximations and inherent error to inputs of the forecast process are associated with some degree of error in any model.

Encouragingly, the physical implications of these regression coefficients are all reasonable and consistent with the behaviors that [Smith et al. \(2015\)](#) and [Thompson et al. \(2017\)](#) identified in their work regarding tornado probabilities. Specifically, an increase in tornado probabilities can be explained by increasing  $v_{rot}$ , decreasing dist, a clear/tight designation, and/or increasing stp. Based on the  $P$  values provided,  $v_{rot}$ , dist, and ct suggest the strongest relationship to tornado probability assessment, followed by stp.

Similar to the analysis performed for the tornado probability model, [Table 2](#) presents the attributes of the tornado wind speed model. In this case,  $P$  values are determined using the standard normal distribution ([Pindyck and Rubinfeld 1981](#)). Predictors that are retained following the  $P$ -value-based filtering process specified earlier are arl,  $v_{rot}$ , tds, and stp. The  $R$ -squared statistic for this linear regression is 0.38. In a similar regard with the tornado probability model, the  $R$ -squared value for the tornado wind speed model still may not be considered high or strong, though *practical* value may still be extracted from this model, even if it does not provide exact tornado wind speeds perfectly. Once again, the physical implications of these regression coefficients are generally reasonable. For instance, an increase in tornado wind speed can be explained by increasing  $v_{rot}$ , increasing stp, the presence of a tds, and/or lowering arl. Based on the  $P$  values provided,  $v_{rot}$ , stp, and tds suggest the strongest relationship to simulated tornado wind speeds, followed by arl. Also, similar to the tornado probability model, simulated tornado wind speeds can be appropriately scaled within the context of traditional tornado damage ratings (i.e., EF0–5).

Applying the model developed for tornado probabilities with corresponding regression coefficients in [Table 1](#) to the independent dataset yields appreciable separation between probabilities corresponding to tornado events (total of 498 tornado events) and

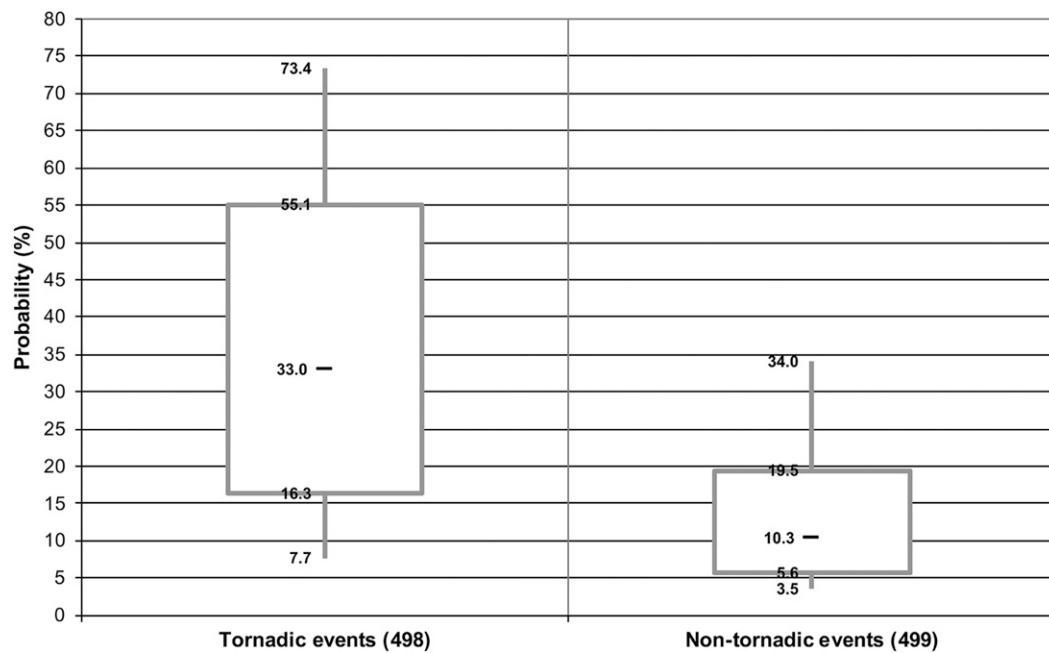


FIG. 1. Box-and-whisker plots corresponding to the simulated tornado probabilities from the independent dataset for (left) tornadic and (right) nontornadic events. Boxes correspond to interquartile ranges (25th–75th percentiles) with the embedded marker indicating the median (50th percentile) of the distributions, and whiskers extending to the 10th and 90th percentiles of the distributions. The plotted values are in percent, following multiplication of the tornado probabilities by 100%. Sample sizes for each of the distributions are provided below the horizontal axis in parentheses.

probabilities corresponding to nontornado events (total of 499 nontornado events). This is illustrated by the box-and-whisker plots shown in Fig. 1. In fact, the 25th percentile of the tornado probability for the tornado events is similar to the 75th percentile of the tornado probability for the nontornado events. This implies meaningful separation in practice between the tornado and nontornado events for the simulated probabilities, as well as the utility offered by this statistical model. While there is notable separation between the tornadic and nontornadic distributions, especially between the interquartile ranges, one is cautioned that there is still some overlap between the distributions that is exacerbated in practice owing to the proportionally greater number of severe thunderstorms with circulations that do not produce tornadoes compared to those that do produce tornadoes.

Applying the model developed to compute simulated tornado wind speeds using the regression coefficients in Table 2 to the independent dataset suggests that this model is generally reasonable in simulating wind speeds, as shown in the box-and-whisker plot presented in Fig. 2. There is a general tendency for increasing the actual estimated tornado peak wind speeds based on damage to be associated with increasing computed tornado wind

speed based on the statistical simulations. For significant tornadoes (EF2+), this model offers a general underestimate of tornado wind speeds, and this could be explained by the small sample size of higher-end tornadoes in the training dataset, yielding comparatively minimal influence in the corresponding model. This could be motivation for developing a model with more similar sample sizes representing each of the tornado wind speeds. Regardless, the application of the tornado wind speed model on the independent dataset definitively suggests a signal for trends in model output simulations to correspond to similar trends in reality.

Additionally, both the tornado probability and tornado wind speed models are applied for three of the cases illustrated by Thompson et al. (2017). Two of these cases, shown in their Figs. 5 and 12, provided and annotated further in Figs. 3 and 4, respectively, correspond to significant tornado occurrence. In both cases, the tornado probability model is simulated to produce probabilities of nearly 100%, and tornado peak wind speed estimates based on damage that represent significant tornadoes (EF2+). In both of these cases, these models perform very well in accurately highlighting high confidence of significant tornado occurrence. The third case, shown in Thompson et al.'s (2017) Fig. 13,

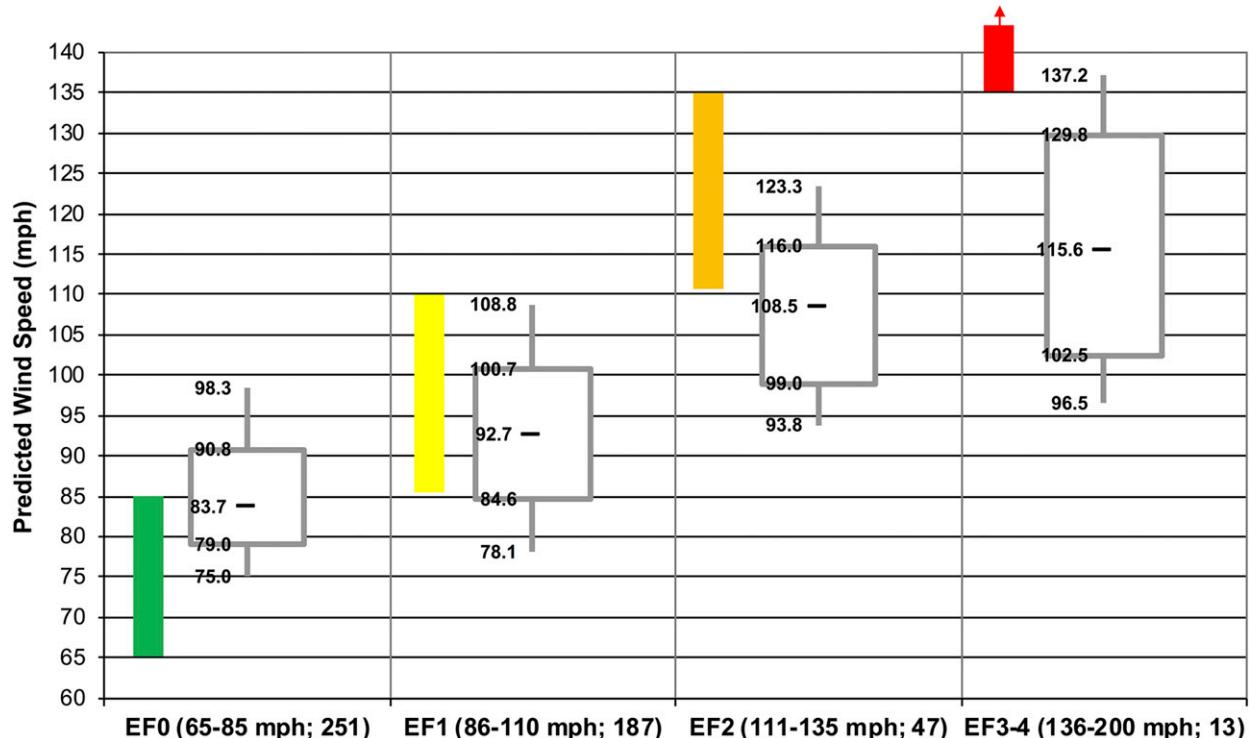


FIG. 2. Box-and-whisker plots corresponding to the simulated tornado wind speeds from the independent dataset for tornadic events. These tornadic events are binned along the horizontal axis by the EF scale (EF3 and EF4 grouped together), with corresponding wind speeds and sample sizes provided in parentheses. Boxes correspond to interquartile ranges (25th–75th percentiles) with the embedded marker indicating the median (50th percentile) of the distributions, and whiskers extend to the 10th and 90th percentiles of the distributions. Colored bars represent the range of binned tornado wind speeds that each EF rating or rating group represents: green for EF0, yellow for EF1, orange for EF2, and red for EF3 and EF4 (extending above the plotted vertical axis range).

provided and annotated further in Fig. 5, corresponds to a nontornado event. In this case, the simulated tornado probability is minimal (i.e., 7%), suggesting another accurate assessment of the likelihood of this circulation to produce a tornado. Output from a tornado wind speed simulation is not provided for this case, as the tornado wind speed statistical model was developed and conditioned on the existence of a tornado.

As a consideration regarding the quality of observations serving as the foundation for both the training and independent datasets, there are ultimately flaws in the tornado database that translate to a limitation of this analysis. Such flaws include an underrating bias, which has direct implications on the opportunity for this regression to represent “truth.” Kingfield and LaDue (2015), Strader et al. (2015), and Alexander and Wurman (2008) all collectively highlight some of the flaws of the tornado database. While this database is the official documented source of tornado information, corresponding results from the statistical models need to be considered within the context of the flaws of the tornado database. Additionally, the tornado wind

speed model incorporates the maximum EF rating along the entire path of each tornado event and the maximum average rotational velocity, which are not necessarily collocated. Some of the variability and possible error in the simulated tornado wind speeds could be explained by this lack of collocation. Moreover, the modest magnitude of the Somers’  $D$  statistic and  $R$ -squared statistic might serve as motivation to use a more precise matching of  $v_{rot}$  with damage indicators for well-surveyed tornadoes in areas with sufficient damage-indicator density to reveal actual tornado intensity.

#### 4. Conclusions

This study is the first to develop an applications-based, multivariate statistical model based on the databases created by Smith et al. (2012), Thompson et al. (2012), Smith et al. (2015), and Thompson et al. (2017). In particular, using the database from Thompson et al. (2017), multiple regression analysis was used to develop statistical models that simulate tornado probabilities and tornado wind speed based on the peak average

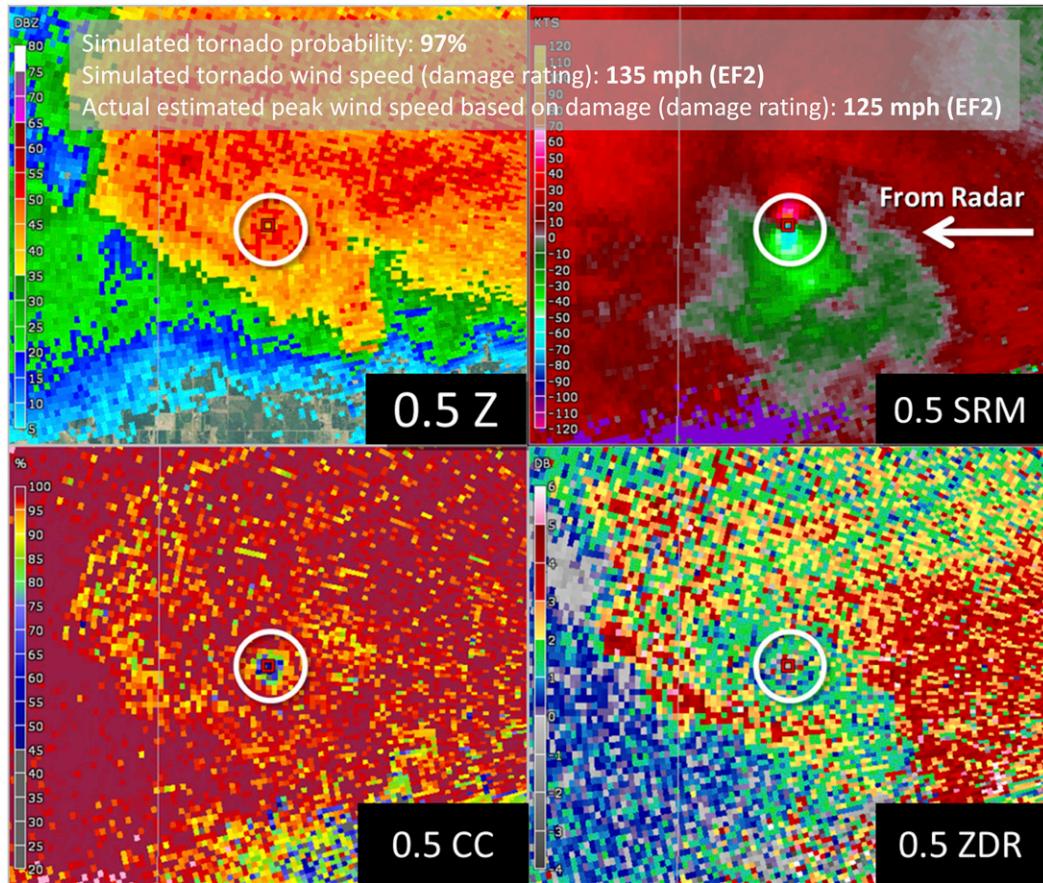


FIG. 3. Figure 5 from [Thompson et al. \(2017\)](#), with an annotation provided at the top indicating the output from the tornado probability model and tornado wind speed model for the circulation illustrated here. The actual estimated peak wind speed based on damage is also annotated. For this example, the following inputs characterize the regression equations [Eqs. (1) and (2)]:  $arl = 900$  ft,  $v_{rot} = 92.8$  kt,  $dist = 0.1$  n mi,  $ct = 1$ ,  $tds = 1$ , and  $stp = 4.72$ .

rotational velocity, height of circulation above radar level, circulation diameter, whether the circulation was characterized as clear/tight, and the effective-layer significant tornado parameter, with the presence or absence of a dual-pol tornadic debris signature considered for the tornado wind speed model. These statistical models yield relatively modest  $R$ -squared statistics though the signs and  $P$  values corresponding to the regression coefficients are generally consistent with physical implications addressed by [Thompson et al. \(2017\)](#) and [Smith et al. \(2015\)](#).

Furthermore, within the general context of the storm-scale circulation's and environment's propensity to support tornadoes and certain simulated tornado peak wind speeds, these models offer practical insights that can be communicated as various levels of confidence for tornado impacts in real-time severe weather events. This is affirmed through simulating these models on an independent dataset—separate from the training dataset

upon which the models were constructed. The simulations yield relatively strong separation between computed tornado probability simulations corresponding to tornado events and nontornado events. Additionally, it is apparent that, with increasing actual tornado wind speeds based on documented EF-scale rating, the statistical model also simulates increasing tornado wind speeds for the independent dataset, implying at least some degree of consistency between simulation and reality. Applications of the models to specific cases addressed by [Thompson et al. \(2017\)](#) reveal convincing support for the validity of this model for the opposite ends of circulation strength and impact spectra represented by these cases.

This work is intended to be a step in applying the dataset presented by [Thompson et al. \(2017\)](#). The models created from this dataset could be integrated into an operational workflow, permitting quantifications of tornado threat. This work has also highlighted avenues for additional research. For example,

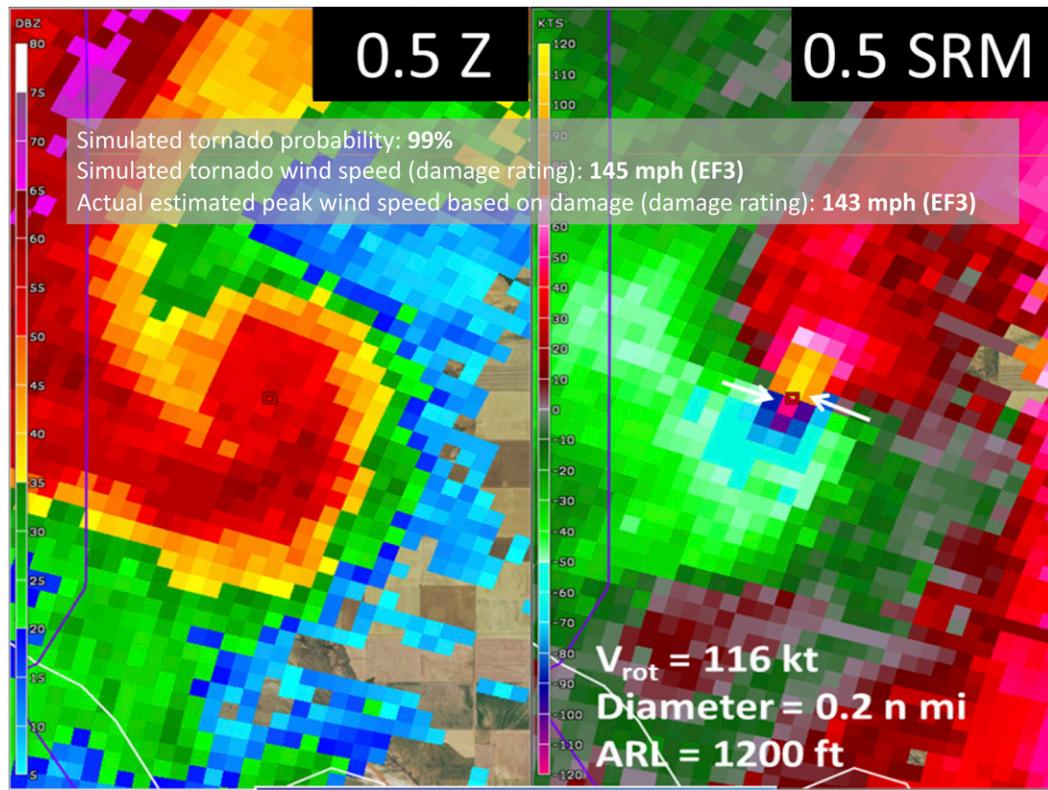


FIG. 4. As in Fig. 3, but corresponding to Fig. 12 in Thompson et al. (2017). For this example, the following inputs characterize the regression equations [Eqs. (1) and (2)]:  $arl = 1200$  ft,  $v_{rot} = 116$  kt,  $dist = 0.2$  n mi,  $ct = 1$ ,  $tds = 1$ , and  $stp = 1.94$ .

the development of more sophisticated models could be a focus of subsequent research. Models founded on larger sample sizes within individual wind speed bins may also provide an opportunity to better represent the higher-end tornado threats. Varying sample-size constituents of the training and independent datasets could assist in building more sophisticated models via testing, as well. Incorporation of convective mode could also be a focus of additional model development. The notion of developing separate models for different geographic divisions of the country could also be considered, owing to the spatially varying nature of the damage indications that influence both tornado occurrence documentation and tornado damage rating documentation. Also, since the tornado wind speed statistical simulation was built solely on the tornado dataset, its use is conditional upon knowledge of a tornado's occurrence (e.g., that associated with a tornadic debris signature or a report). A more sophisticated model could mitigate the conditionality of a known tornado occurrence to simulate a wind speed in cases where it is unknown whether a tornado is occurring.

Ultimately, the work that Thompson, Smith, and their collaborators have done, as cited throughout this work,

has opened the doors for a plethora of research stemming from the severe thunderstorm and tornado database. This represents a first attempt at directly applying their work into a form that can be instituted into operational practice for simulating tornado potential and wind speed in a diagnostic manner. It is cautioned that no attempt has been made to specify the prognostic—or forecast—utility of these models. Rather, given the present state of the storm circulation and near-storm environment, these regression models provide simulation utility, based on present storm-scale and environmental conditions contextualized in terms of event-based documentation from the past. Ultimately, such work serves as the foundation for evolving warning services, including such initiatives as “warn on forecast” (Stensrud et al. 2009) and Forecasting a Continuum of Environmental Threats (FACETs; Rothfusz et al. 2013). These initiatives represent clear paths to build from diagnostic information regarding tornado potential, perhaps following the methodology outlined in this study, and then project forward in time and space. This would require ensuring more stringent, reproducible bounds on what is defined as a clear and tight circulation, for which

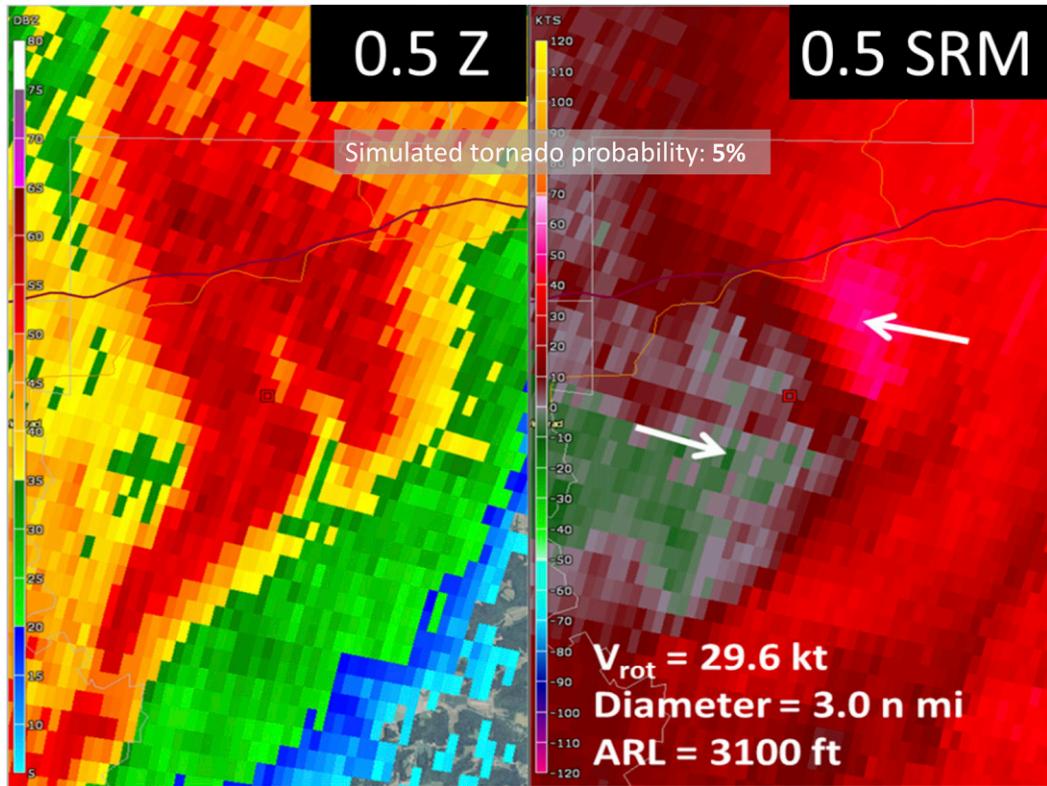


FIG. 5. As in Fig. 3, but corresponding to Fig. 13 in Thompson et al. (2017) without the simulated tornado wind speed and without the actual estimated peak wind speed based on damage, as no tornado occurred. For this example, the following inputs characterize the regression equations [Eqs. (1) and (2)]:  $arl = 3100$  ft,  $v_{rot} = 29.6$  kt,  $dist = 3.0$  n mi,  $ct = 0$ ,  $tds = 0$ , and  $stp = 0$ .

the subjectivity in the present work offers a limitation. Automated systems could potentially permit such reproducibility by leveraging automated rotational velocity computations. The infusion of the methodologies presented in this study with convection- and environment-diagnosing systems may yield numerous opportunities to enhance the spatiotemporal precision of accuracy of high-impact weather information. It is our hope that these statistical models applied to physically based results can become infused into infrastructure that supports accurate and consistent messaging of critical severe weather threats.

**Acknowledgments.** The authors greatly appreciate Israel Jirak's (Storm Prediction Center) careful oversight of this work and feedback that contributed to its improvement. The lead author expresses deep gratitude to Alan Gerard (National Severe Storms Laboratory), Chauncy Schultz (National Weather Service, Bismarck, ND), and Ryan Bunker (National Weather Service, Topeka, KS, and University of Oklahoma) for their encouragement of this work. Three anonymous reviewers provided feedback that also greatly improved

this manuscript. The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the authors and do not necessarily reflect the views of NOAA or the Department of Commerce.

## REFERENCES

Alexander, C. R., and J. Wurman, 2008: Updated mobile radar climatology of supercell tornado structures and dynamics. *24th Conf. on Severe Local Storms*, Savannah, GA, Amer. Meteor. Soc., 19.4, [https://ams.confex.com/ams/24SLS/techprogram/paper\\_141821.htm](https://ams.confex.com/ams/24SLS/techprogram/paper_141821.htm).

Bothwell, P. D., J. A. Hart, and R. L. Thompson, 2002: An integrated three-dimensional objective analysis scheme in use at the Storm Prediction Center. *21st Conf. on Severe Local Storms/19th Conf. on Weather Analysis and Forecasting/15th Conf. on Numerical Weather Prediction*, San Antonio, TX, Amer. Meteor. Soc., JP3.1, <https://ams.confex.com/ams/pdfpapers/47482.pdf>.

Kingfield, D. M., and J. G. LaDue, 2015: The relationship between automated low-level velocity calculations from the WSR-88D and maximum tornado intensity determined from damage surveys. *Wea. Forecasting*, **30**, 1125–1139, <https://doi.org/10.1175/WAF-D-14-00096.1>.

Pindyck, R. S., and D. L. Rubinfeld, 1981: *Econometric Models and Economic Forecasts*. 2nd ed. McGraw-Hill, 630 pp.

Rothfusz, L. P., E. Jacks, J. T. Ferree, G. J. Stumpf, and T. M. Smith, 2013: Next-generation warning concept: Forecasting a Continuum of Environmental Threats (FACETs). *Second Conf. on Weather Warnings and Communication*, Nashville, TN, Amer. Meteor. Soc., 3.4, <https://ams.confex.com/ams/41BC2WxWarn/webprogram/Paper225900.html>.

Smith, B. T., R. L. Thompson, J. S. Grams, C. Broyles, and H. E. Brooks, 2012: Convective modes for significant severe thunderstorms in the contiguous United States. Part I: Storm classification and climatology. *Wea. Forecasting*, **27**, 1114–1135, <https://doi.org/10.1175/WAF-D-11-00115.1>.

—, —, A. R. Dean, and P. T. Marsh, 2015: Diagnosing the conditional probability of tornado damage rating using environmental and radar attributes. *Wea. Forecasting*, **30**, 914–932, <https://doi.org/10.1175/WAF-D-14-00122.1>.

Somers, R. H., 1962: A new asymmetric measure of association for ordinal variables. *Amer. Sociol. Rev.*, **27**, 799–811, <https://doi.org/10.2307/2090408>.

Stensrud, D. J., and Coauthors, 2009: Convective-scale warn-on-forecast system: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1499, <https://doi.org/10.1175/2009BAMS2795.1>.

Strader, S. M., W. Ashley, A. Irizarry, and S. Hall, 2015: A climatology of tornado intensity assessments. *Meteor. Appl.*, **22**, 513–524, <https://doi.org/10.1002/met.1482>.

Thompson, R. L., B. T. Smith, J. S. Grams, A. R. Dean, and C. Broyles, 2012: Convective modes for significant severe thunderstorms in the contiguous United States. Part II: Supercell and QLCS tornado environments. *Wea. Forecasting*, **27**, 1136–1154, <https://doi.org/10.1175/WAF-D-11-00116.1>.

—, and Coauthors, 2017: Tornado damage rating probabilities derived from WSR-88D data. *Wea. Forecasting*, **32**, 1509–1528, <https://doi.org/10.1175/WAF-D-17-0004.1>.

WDTD, 2016: Convective storm structure and evolution: Analyzing tornado scale signatures. Radar and Applications Course (RAC), Warning Decision Training Division, [wdtd.noaa.gov/courses/rac/severe/tornadic-signatures/presentation.html](http://wdtd.noaa.gov/courses/rac/severe/tornadic-signatures/presentation.html).