The Tornado Probability Algorithm: A Probabilistic Machine Learning Tornadic Circulation Detection Algorithm

Thea N. Sandmæl,[a,b] Brandon R. Smith,[a,b] Anthony E. Reinhart,[b] Isaiah M. Schick,[a,b,c] Marcus C. Ake,[a,b,c] Jonathan G. Madden,[a,b] Rebecca B. Steeves,[a,b] Skylar S. Williams,[a,b] Kimberly L. Elmore,[a,b] and Tiffany C. Meyer [a,b]

[a] *Cooperative Institute for High-Impact and Severe Weather Research and Operations, the University of Oklahoma, Norman, Oklahoma*

[b] *NOAA/OAR National Severe Storms Laboratory, Norman, Oklahoma*

[c] *The University of Oklahoma School of Meteorology, Norman, Oklahoma*

Corresponding author: Thea N. Sandmæl, thea.sandmael@noaa.gov

Isaiah M. Schick's current affiliation: AccuWeather, Wichita, KS

Skylar S. Williams's current affiliation: AvMet Applications Inc., Reston, VA

Tiffany C. Meyer's current affiliation: Unidata, University Corporation for Atmospheric Research, Boulder, CO

Early Online Release: This preliminary version has been accepted for publication in *Weather and Forecasting*, may be fully cited, and has been assigned DOI 10.1175/WAF-D-22-0123.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

ABSTRACT: A new probabilistic tornado detection algorithm was developed to potentially replace the operational tornado detection algorithm (TDA) for the WSR-88D radar network. The Tornado Probability algorithm (TORP) uses a random forest machine learning technique to estimate a probability of tornado occurrence based on single-radar data, and is trained on 166,145 data points derived from 0.5°-tilt radar data and storm reports from 2011-2016, of which 10.4% are tornadic. A variety of performance evaluation metrics show a generally good model performance for discriminating between tornadic and non-tornadic points. When using a 50% probability threshold to decide whether the model is predicting a tornado or not, the probability of detection and false alarm ratio are 57% and 50%, respectively, showing high skill by several metrics and vastly outperforming the TDA. The model weaknesses include false alarms associated with poor-quality radial velocity data and greatly reduced performance when used in the western United States. Overall, TORP can provide real-time guidance for tornado warning decisions, which can increase forecaster confidence and encourage swift decision making. It has the ability to condense a multitude of radar data into a concise object-based information read-out that can be displayed in visualization software used by the National Weather Service, core partners, and researchers.

SIGNIFICANCE STATEMENT: This study describes the Tornado Probability algorithm (TORP) and its performance. Operational forecasters can use TORP as real-time guidance when issuing tornado warnings, causing increased confidence in warning decisions, which in turn can extend tornado warning lead times.

1. Introduction

Data from weather radars have been used to observe and forecast severe storms in the United States for many decades (Whiton et al. 1998a,b). Since the first radar observation of a tornadic storm and the subsequent installation of the Weather Surveillance Radar network (WSR-57) in the 1950s (Stout and Huff 1953), the U.S. radar network has undergone several expansions and upgrades. As the radar network evolved, so has our understanding of the radar representation of severe storms and how this knowledge can be used to support the decision to issue tornado warnings (Smith and Holmes 1961; Donaldson 1970; Lemon et al. 1977). Some early discoveries that were – and continue to be - significant for issuing tornado warnings include the reflectivity (Z_H) hook echo signature and the weak echo region, as well as the 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).

The establishment of the Next Generation Weather Radar (NEXRAD) network of Weather Surveillance Radar-1988 Doppler (WSR-88D; Crum and Alberty 1993) radars in the 1990s led to a large increase in tornado-warning skill as a result of the new information provided to National Weather Service (NWS) forecasters, most evident by a considerable improvement in the probability of detection of tornadoes (Simmons and Sutter 2005). Since the initial deployment, several significant upgrades to the WSR-88D network have occurred, including the implementation of super-resolution data (Brown et al. 2002, 2005; Torres and Curtis 2007) and the dual-polarization upgrade (Istok et al. 2009; Saxion and Ice 2012).

These advances in both radar and data-processing technology have aided in research that furthers the understanding of Doppler radar signatures associated with storm-scale processes that involve tornadoes. Ryzhkov et al. (2005) introduced the tornadic debris signature (TDS), which is characterized by an area of depressed correlation coefficient (ρ_{HV}) and differential reflectivity (Z_{DR}) associated with a TVS. The TDS signature is a robust tornado detection tool that is used in NWS tornado-warning operations (Van Den Broeke 2017; Warning Decision Training Division 2022a). The vertical extent of the TDS is also used to infer potential tornado intensity, which is incorporated into decisions concerning impact-based tornado warning tags (Bodine et al. 2013; Gibbs 2016; Warning Decision Training Division 2022b).

More recent work has shown how radar signatures change with the evolution of tornadic storms, including both the pre-tornadic and post-tornadic periods, as well as how they differ from non-tornadic storms. Sandmæl et al. (2019) evaluated storm-based extrema of a multitude of variables, including azimuthal and range derivatives of V_R, and showed statistical differences between tornadic and non-tornadic storms, as well as between the pre-tornadic, currently tornadic, and post-tornadic periods. Similar results were found by Lyza et al. (2022), who examined the azimuthal shear evolution of tornadic supercells during a historical tornado outbreak. The vertical alignment of the azimuthal shear has also been found to separate tornadic and non-tornadic storms through analysis of a large sample of supercells, as well as separating the tornadic and post-tornadic periods of tornado-producing supercells (French and Kingfield 2019; Homeyer et al. 2020). Moreover, Homeyer et al. (2020) showed that the storm-motion relative orientation of mid-to-low-level Z_{DR} and the relative orientation of regions of low-level enhanced Z_{DR} and specific differential phase (K_{DP}) could indicate whether a supercell has tornadic potential, the latter of which was also observed by Loeffler et al. (2020). The size of the Z_{DR} column can also separate pre-tornadic supercells from non-tornadic supercells (Van Den Broeke 2020). Polarimetric radar variables have also been linked with tornado demise. Segall et al. (2022) found that decreases in maxima in Z_{DR} arcs (Kumjian and Ryzhkov 2008) and the separation between the orientation angles in Z_{DR} and K_{DP} were associated with tornado dissipation.

With the discovery of radar signatures associated with tornadoes, there have been efforts to create automated algorithms to identify them to aid operational forecasters. The TVS algorithm was the first tornado detection single-radar algorithm to be a part of the WSR-88D operational system products, which used V_R to identify areas of tight rotation (NEXRAD Joint System Program Office 1985; Crum and Alberty 1993). To utilize the improved capabilities of the WSR-88D radars since the TVS algorithm was developed, the National Oceanic and Atmospheric Administration (NOAA) National Severe Storms Laboratory (NSSL) introduced the Tornado Detection Algorithm (TDA; Mitchell et al. 1998) in the late 1990s, which provided a marked improvement over the TVS algorithm and has been a part of the operational NEXRAD Level III products since. Radar dual-polarization moments were not available and therefore not implemented during the development of the original TDA algorithm, but are frequently used in NWS tornado warning operations today in their base form (Bentley et al. 2021) or through ancillary products that they are ingested into (Smith et al. 2016; Cintineo et al. 2020). Warning forecasters now face an increasing amount of weather information, and investigating all of the products and tools available can be time consuming and overwhelming (Karstens et al. 2015; Boustead and Mayes 2014). Advancement in automated techniques such as machine learning can be used to condense available model and observational data concerning high-impact weather by highlighting features that might be of interest to the warning forecaster (McGovern et al. 2017; Lagerquist et al. 2017). As machine learning is increasingly applied to the severe storms nowcasting problem (Lagerquist et al. 2020; Mecikalski et al. 2021; Gensini et al. 2021), there have been successful transitions of machine learning products from research to operations (e.g., Cintineo et al. 2020).

This paper describes a new probabilistic Tornado Probability algorithm (TORP) developed for the WSR-88D network to potentially replace the TDA as a NEXRAD Level III product, which utilizes a random forest (RF; Ho 1998; Breiman 2001) machine learning technique and several base, dual-polarization, and derived Doppler radar products to provide fast, real-time probabilistic detection of tornadoes. TORP uses an object-based framework, employing V_R-derived linear least-square derivative (LLSD; Mahalik et al. 2019) azimuthal shear (AzShear) fields to determine areas with higher magnitudes of rotation to generate storm objects. TORP refines the abundance of radar information available in real time into an easy-to-interpret product that provides both a situational-awareness tool for forecasters to use in tornado-warning operations, and a device to accelerate decision making by enhancing forecaster confidence during the tornado warning issuance process. This publication gives an overview of the TORP algorithm, which includes the construction of the RF and details on how the algorithm operates. A thorough performance evaluation of TORP and how it compares to the performance of the operational TDA is also presented. Additionally, a discussion of how TORP can be used in operations and its potential benefits to the tornado warning operations process is provided.

2. Tornado Radar Algorithms

To put TORP's functionality in context, this section will provide an overview of the TDA (Mitchell et al. 1998), the TDS algorithm (Snyder and Ryzhkov 2015), and ProbSevere's ProbTor (Cintineo et al. 2020).

a. TDA

The TDA is an operational NEXRAD Level III product that is calculated and output through the Radar Product Generator (RPG) maintained by the NOAA NWS Radar Operations Center (NWS Radar Operations Center Applications Branch 2021). The TDA is a single-radar algorithm that provides the location of a TVS or an elevated TVS within a default range of 100 km from a radar (Mitchell et al. 1998). Tornado objects are defined using multiple thresholds of $\mathrm{V_R}$ difference between adjacent radar gates at a constant range thresholded by $\mathrm{Z_H}$. The object creation method first defines single-tilt 2D detections by identifying areas of shear characterized by large values of $\mathrm{V_R}$ difference, which are later combined into 3D detections with data from different tilts below a specified altitude. TDA detections are tracked in time through other algorithms run at the RPG (the Mesocyclone Detection Algorithm (MDA; Stumpf et al. 1998) or Storm Cell Identification and Tracking Algorithm (SCIT; Johnson et al. 1998) depending on what data are available).

b. TDS

TDS detection was developed as a proposed classification in the Hydrometeor Classification Algorithm (HCA; Park et al. 2009; Snyder and Ryzhkov 2015). The TDS classification is decided by applying thresholds and weighting by fuzzy logic to five radar-derived variables: AzShear, ρ_{HV}, $\mathrm{Z_{DR}}$, $\mathrm{Z_H}$, and a differential phase shift (Φ_{DP}) texture parameter. The HCA is an areal product that assigns the TDS classification to individual radar gates. These gates can be isolated to create TDS tracks, which are accumulated in a similar way to the operational Multi-Radar Multi-Sensor (MRMS; Smith et al. 2016) rotation tracks. While the TDS classification category is not operational, the principles and TDS signatures presented in Snyder and Ryzhkov (2015) are used by operational meteorologists as one of the tools to identify tornadoes. The Radar and Applications course offered to NWS forecasters by the Warning Decision Training Division has four guidelines in their TDS detection method, which mimics the HCA TDS classification thresholding technique

(Warning Decision Training Division 2022a); 1) identify a $\mathrm{V_R}$ couplet, 2) check for a nearby area of low ρ_{HV} values, which 3) overlaps with $\mathrm{Z_H}$ values above 35 dBZ with 4) $\mathrm{Z_{DR}}$ values near 0.

c. ProbTor

ProbTor is a part of the ProbSevere algorithm, which provides storm-based probabilities for severe wind, hail, and tornado hazards (Cintineo et al. 2014, 2020). With a 2-min temporal resolution, ProbTor provides probabilities of a tornado occurring within the next 60 min. The probabilities are calculated using machine learning, specifically a naïve Bayesian classifier (in version 2), with information from MRMS AzShear, Earth Networks Total Lightning Network (ENTLN) flash density, and environmental variables from Rapid Refresh (RAP; Benjamin et al. 2016) model data. ProbTor objects are tied to ProbSevere objects that are created using a watershed technique with MRMS composite $\mathrm{Z_H}$, which do not require rotation and can range in size from a relatively small cell to a large-scale linear system.

3. TORP Overview

a. TORP Data and Methods

1) Algorithm Inputs

TORP has a range of required and optional inputs, as well as adjustable parameters, which are outlined in Table 1.

(i) Single-Radar Products TORP operates by calculating tornado probabilities from Level II single-radar data, which can be obtained from the NOAA Big Data Program Amazon Web Services (AWS; Ansari et al. 2018) NEXRAD storage (NOAA National Weather Service Radar Operations Center 1991). All base radar fields, which include $\mathrm{Z_H}$, $\mathrm{V_R}$, velocity spectrum width (SW), as well as the dual-polarization products Φ_{DP}, ρ_{HV}, and $\mathrm{Z_{DR}}$, are processed using a 3×3 median filter before the algorithm extracts the object-based values of these fields. The $\mathrm{V_R}$ data are dealiased using the dealiasing methods from WSR-88D RPG Build 19 (Jing and Wiener 1993; Zittel 2019; Losey-Bailor et al. 2019).

(ii) Single-Radar LLSD Products LLSD azimuthal, range, and total gradients (AzGradients, RanGrandients, and Gradients) were calculated for all single-radar moments following the methods

Table 1. List of required and optional inputs for TORP, including inputs in the text file formats comma-separated values (CSV), and extensible markup language (XML), as well as adjustable numerical and string inputs.

	Input	Description
Required	Random forest CSV file	This file will be provided by the developer and will not change unless the RF is retrained
	4-letter ICAO radar code	Text string specifying which radar to use
	Single-radar 0.5°-tilt data	Dealiased V_R, Φ_{DP}, Z_H, ρ_{HV}, velocity spectrum width, Z_{DR}
		Base product, and LLSD azimuthal, range, and total gradients, indexed by an XML file
Optional	RAP-derived sounding table	XML file with 0-6-km wind information
	Imputation data	CSV file with imputation values in case of missing data
	Tilt	Default: 0.5°
		Specifying which tilt the algorithm will run on
	Minimum number of object gates	Default: 4
		Requiring objects to have at least this many gates meeting the AzShear threshold
	Range limit threshold	Default: 160 km
		The radar range threshold used to remove or flag far-range objects
	Range limit	Default: On
		Removing objects outside of the range threshold
	Elevation threshold	Default: None
		A height threshold in km, which will remove objects above the specified threshold
	Z_H filters	Default: 20-dBZ thresholding, double despeckling, median filtering, and dilation filtering
		Object data will be thresholded by Z_H data with this filter
	Maximum merging radius	Default: 9,000 m
		Objects with centers within this distance of one another will be merged
	Radius for variable extraction	Default: 2,500 m
		The current RF is trained on variables extracted in the default radius
	Probability threshold	Default: 0
		Excluding objects below this probability
	Maximum tracking distance	Default: 9,000 m
		Objects will only be linked in time if they are within this distance of the forecasted position
	Maximum storm speed	Default: 15.5 m s^{-1}
		The maximum assumed speed used when tracking objects by using the track history
	Anti-cyclonic detections	Default: Off
		Anti-cyclonic objects can be evaluated if given an RF that is trained with anti-cyclonic objects
	Latitude, longitude, and time	Creating an object from manual time and location input to be evaluated by the RF
	Manual location distance	Default: 2,500 m
		Adjusts manual location to the nearest AzShear maximum within this distance

described in Mahalik et al. (2019). In addition to being used to calculate tornado probabilities with the other radar variables, LLSD AzShear is used to define high-rotation storm objects. Similar to how AzShear represents the azimuthal shear, the range gradient in V_R represents divergent shear (DivShear). Recent studies have shown the utility of DivShear signatures associated with tornadic storms, with case studies showing signatures during pre-tornadic periods for lower-rated quasi-linear convective system or tropical cyclone tornadoes (Mahalik et al. 2019; Sandmæl et al. 2019; Sandmæl and Reinhart 2022). These signatures include quadrupoles of extreme high and low values (Fig. 1F) and highly negative (convergent) low-level DivShear. Several LLSD gradient products show distinct signatures associated with TDSs, some of which occur without a TDS present as well, which is utilized when TORP is calculating its tornado probabilities. As an example, a selection of LLSD products are presented in Fig. 1 as a supercell is producing a tornado that caused EF5-rated damage (circled location). While not all of the LLSD gradients for each product are shown, signatures of note are dipoles of extreme low and high values of Φ_{DP}, ρ_{HV}, and SW AzGradients and RanGrandients, rings in Z_H and ρ_{HV} Gradients, and circles of enhanced Φ_{DP} and SW Gradients (not shown). These LLSD signatures coincide with known TVS and TDS signatures in Fig. 1A-D, which are the debris ball, V_R couplet, and areas of low ρ_{HV} and enhanced SW, respectively (e.g., Burgess et al. 1975; Ryzhkov et al. 2005; Bodine et al. 2013; Sandmæl et al. 2019). While most of the LLSD gradients are unique products, the Φ_{DP} RanGradient is calculated similarly to the operational K_{DP}. However, the Φ_{DP} RanGradient requires fewer processing steps and retains values higher than 12° km^{-1}, revealing the dipole of extreme values in the TDS location in Fig. 1J.

(iii) RAP Data Model analysis data from RAP (Benjamin et al. 2016) can be used to calculate 0-6-km mean storm motion to use as a first guess when using storm motion to track TORP objects in time. Hourly RAP data can be retrieved from NOAA National Centers for Environmental Information (NCEI).

2) Object Creation and Tracking

After initiation, TORP will create storm objects based on a 0.006-s^{-1} AzShear threshold. This threshold for creating automated objects is adjustable, though this default number is based on a statistical analysis of all non-quality-controlled storm reports from 2011-2018 from NCEI (NOAA National Weather Service 1950). The analysis showed that 92% of storm samples with an ongoing tornado and 68% of storm samples with any ongoing severe weather exceed this AzShear threshold.

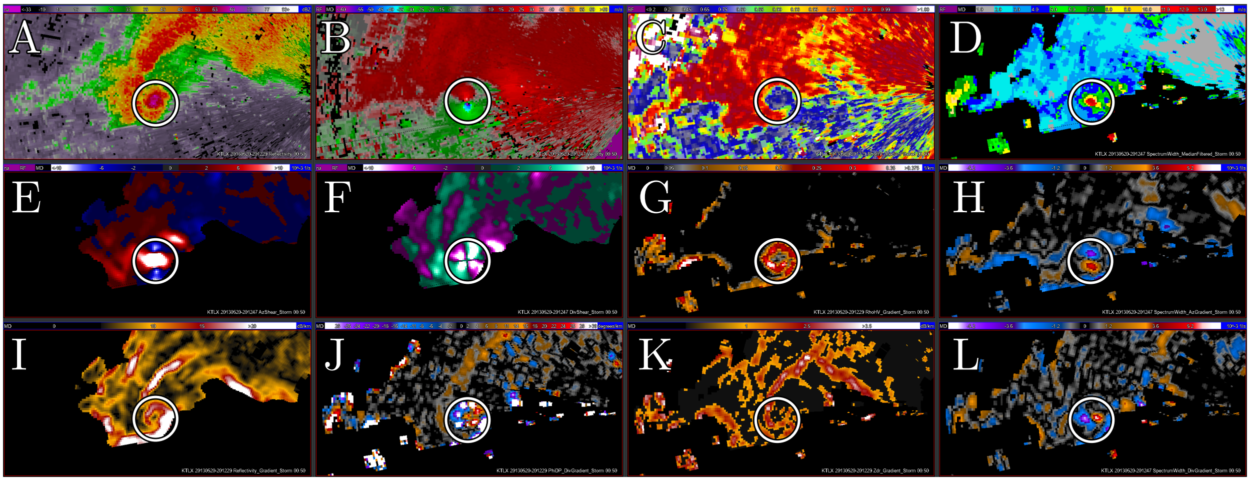


FIG. 1. Example of 0.5°-tilt single-radar products associated with the 20 May 2013 Moore, OK tornadic supercell at 2012 UTC with the EF5 tornado location circled. A) Raw Z_H, B) Dealiased V_R, C) Median-filtered ρ_{HV}, D) Median-filtered SW, E) AzShear, F) DivShear, G) ρ_{HV} Gradient, H) SW AzGradient, I) Z_H Gradient, J) Φ_{DP} RanGrandient, K) Z_{DR} Gradient, L) SW RanGrandient.

Initial objects are found by grouping the radar gates that indicated high rotation from AzShear by using a depth-first search recursive algorithm, which searches for neighboring pixels (Tarjan 1972). Any objects consisting of fewer than 4 gates are discarded to reduce the number of objects associated with noise in the V_R data. The objects are also limited by the object distance from the radar, which is 160 km by default. Outside of this range, the TORP RF model had limited samples to train on primarily because data from the nearest radar was always used when extracting the training data, which rarely exceeded 160 km (Fig. 2). The total number of data points exceeding 160 km in range is an order of magnitude lower than the count in each 20-km bin within 160 km of a radar. However, the range threshold is adjustable and can be turned off, which will trigger the algorithm to flag objects outside of the threshold rather than removing them.

The initial objects are masked with filtered Z_H data to further reduce noise. In addition to the 3×3 median filter that is applied to all base variables, the Z_H field that is used to mask the objects is filtered further by applying a 20-dBZ threshold, double despeckling, and a dilation filter. This combination of image-processing techniques and thresholding allows the removal of a large amount of V_R noise, especially near-radar ground clutter, while retaining objects in weak-echo regions by dilation of the higher Z_H values.

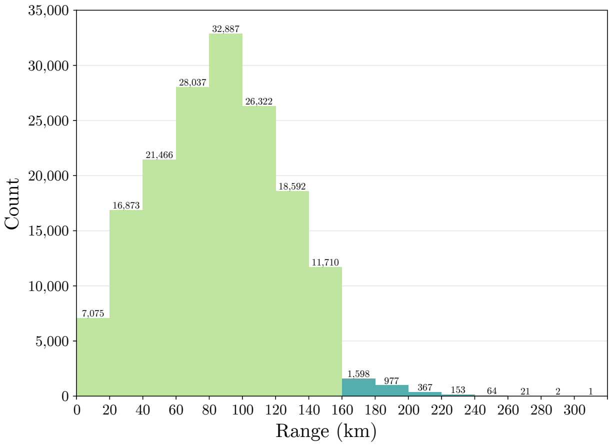


Fig. 2. Radar range counts for the 2011-2016 training dataset in bins of 20 km. The color shift signifies which bins are inside or outside of the default range limit.

Multiple detections of one threat are minimized by combining objects that are within either 9 km or two times the azimuthal spacing of the radar data, whichever is smaller. If two objects fall within this distance of one another, the two objects are merged by assigning the ID of the object with the strongest AzShear maximum to both objects. The 9-km threshold distance was chosen as an optimal distance to create storm-based objects after subjectively testing varying thresholds between 5 and 15 km, but it is an adjustable input that can be provided when initializing the algorithm. This could be a relevant adjustment to obtain individual tornado probabilities for investigating cycling supercell storms with multiple areas of rotation that are producing concurrent tornadoes.

The object merging concludes the object creation process. Radar data are extracted within a 2.5-km radius of each object center, which is defined as the location of the maximum AzShear within the object. The extraction radius was determined by estimating the areal extent of the single-radar signatures associated with large tornadoes to use as an upper limit. For reference, the circle encompassing the TVS in Fig. 1 is approximately 5 km in diameter. This circle captures all of the extreme values associated with the TVS/TDS in the example, with the exception of the Z_{DR} arc displayed in the Z_{DR} Gradient. The minimum, 25th percentile, median, 75th percentile, and maximum values of each radar product are calculated from the extracted values to be evaluated by TORP's RF model that provides the tornado probability. The objects can be filtered by the probability provided when the algorithm is started. However, the default is to keep all of the objects and let the user filter them by using a probability slider in a visualization tool.

Objects go through a tracking process to link objects in time to develop trends and show continuity. The tracking is performed so that objects with the highest rotation are evaluated first, which will lead to retaining information from the strongest circulation if storms are merging. To create forecasted positions for linking objects in time, the algorithm will use the mean storm motion calculated from the last four object locations along an object track or the 0-6-km mean storm motion if the object does not have a track history.

The algorithm can also be run by providing a user-input object location and time. This allows for manual objects to be created using a mouse click over an area of interest to obtain a tornado probability for any storm with available radar data, which could be used for real-time or post-event investigation of storms that do not meet the default AzShear threshold criteria.

3) Random Forest

TORP uses a machine learning RF model, which allows for rapid tornado probability calculations (¡0.1 s) when working with real-time radar data. Other advantages of using an RF include the lack of a need to normalize variables, which could add processing time, as well as having solutions for handling missing data, something that can occur occasionally with real-time data. It is also trivial to retrain region-specific RF models given ample data.

An RF is a supervised learning method for classification, meaning that it requires labeled objects to learn from, which in this case will be labels of “tornadic” and “non-tornadic”. Each object will have a set of radar predictors associated with it, which are defined as the variables that are presented to the machine learning technique in order to learn how to classify the object. The same predictors that were used during training have to be accessible to the algorithm in real time to make an informed decision with the highest accuracy. TORP is still able to run with lower accuracy when some radar data are missing as long as at least 75% of the expected predictors are available. In the event of missing predictors, statistical imputation is utilized. This method replaces the missing values with the mean value of the predictors from the training set.

The RF consists of a number of decision trees created in the model training process, and will use the predictors to determine its prediction for each tree. The model has assigned a threshold value for the predictor in each node in a tree (Fig. 3). The predictors extracted from the object to be evaluated by the RF will determine which branch the algorithm will follow. When the end of a branch is reached, also known as a "leaf", the RF will provide a fraction of training objects that were labeled as "tornadic" that reached this specific leaf. Each tree is based on a different subsample of the training dataset using bootstrap aggregating, or bagging, which is the process of extracting random subsets of the training data to avoid overfitting the model. The final tornado probability is computed by calculating the mean of the fractions from each decision tree, which is 500 trees for TORP's RF.

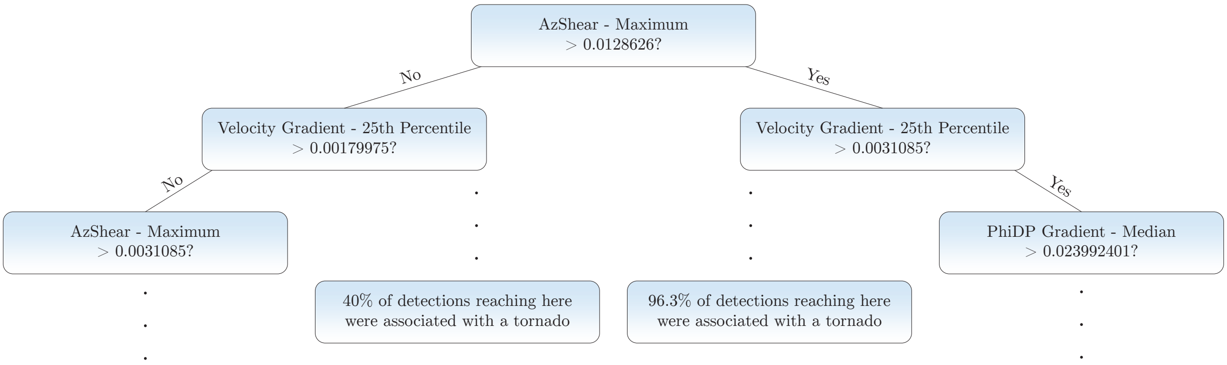

AzShear - Maximum
> 0.0128626?
No
Yes
Velocity Gradient - 25th Percentile
> 0.00179975?
Velocity Gradient - 25th Percentile
> 0.0031085?
No
Yes
AzShear - Maximum
> 0.0031085?
PhiDP Gradient - Median
> 0.023992401?
40% of detections reaching here
were associated with a tornado
96.3% of detections reaching here
were associated with a tornado

FIG. 3. Example of part of a decision tree in the TORP RF.

All available 0.5°-tilt single-radar moments and the LLSD gradients for each product were used to define predictors for the TORP RF model. The range from the radar to the object center in 20-km bins was also included in the list of predictors, as the values of the single-radar variables associated with tornadic events will likely vary based on distance from the radar due to resolution changes with range (Wood et al. 2001; Wood and Brown 1997).

In order to limit the number of predictors without affecting performance and to conduct a meaningful variable importance analysis, a correlation analysis was performed using the Pearson correlation coefficient (PCC), which was calculated by using Eq. 1,

$$PCC = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x}) \sum (y-\bar{y})}} \tag{1}$$

where x and y represent the two variables to be evaluated for correlation, and $\bar{x}$ and $\bar{y}$ are the mean of each variable data in the training dataset (Bravais 1844). One predictor for each correlated predictor pair in the training dataset correlated with a PCC above 0.7 was removed. The correlated predictor that eliminated the largest number of calculations was chosen for removal. The final set of 62 predictors are listed in Table 2, which includes the predictor importance rank and score. An exploratory analysis was performed to assess the performance impact of the predictors ranking from 40 to 62 to determine whether the predictor list should be reduced further. Removing these predictors showed a very slight decline in RF performance when training a new RF for every removed predictor. Since calculating the additional predictors is computationally inexpensive, all of the 62 predictors remain in the model.

TABLE 2. The 62 single-radar predictors used in TORP with importance ranking and the impurity-based predictor importance score. The score measures the importance based on a weighted measure of a predictor's contributions to the thresholding within each decision tree in the RF (Pedregosa et al. 2011).

Predictor	Importance Rank	Importance Score
AzShear 25th Percentile	34	0.005637
AzShear Maximum	1	0.293067
AzShear Median	5	0.031931
AzShear Minimum	8	0.023323
DivShear 75th Percentile	28	0.007405
DivShear Maximum	16	0.010551
DivShear Median	26	0.008254
DivShear Minimum	25	0.008544
Φ_{DP} AzGradient Median	39	0.004942
Φ_{DP} RadGradient 25th Percentile	12	0.015336
Φ_{DP} RadGradient 75th Percentile	3	0.044553
Φ_{DP} RadGradient Median	11	0.016041
Φ_{DP} RadGradient Minimum	42	0.004606
Φ_{DP} Gradient Maximum	30	0.006596
Φ_{DP} Gradient Median	13	0.014336
Φ_{DP} Gradient Minimum	50	0.003948
Φ_{DP} Median-Filtered Maximum	22	0.009087
Φ_{DP} Median-Filtered Minimum	27	0.008047
Radar Range Interval	17	0.010151
Z_H AzGradient Maximum	33	0.005642
Z_H AzGradient Median	19	0.009879
Z_H AzGradient Minimum	37	0.005061
Z_H RanGradient Median	23	0.009045
Z_H RanGradient Minimum	18	0.010106
Z_H Gradient Maximum	10	0.016867
Z_H Gradient Minimum	41	0.004718
Z_H Median-Filtered Maximum	4	0.043441
Z_H Median-Filtered Minimum	7	0.026069
ρ_{HV} AzGradient 25th Percentile	56	0.003166
ρ_{HV} AzGradient 75th Percentile	61	0.002342
ρ_{HV} AzGradient Median	62	0.001719
ρ_{HV} RanGradient Median	57	0.003160
ρ_{HV} Gradient Maximum	15	0.010873
ρ_{HV} Gradient Minimum	59	0.002890
ρ_{HV} Median-Filtered Maximum	14	0.012677
ρ_{HV} Median-Filtered Median	31	0.005964
ρ_{HV} Median-Filtered Minimum	9	0.019017
SW AzGradient 25th Percentile	47	0.004208
SW AzGradient 75th Percentile	48	0.004169
SW AzGradient Median	60	0.002754
SW AzGradient Minimum	44	0.004427
SW RanGradient 25th Percentile	40	0.004740
SW RanGradient 75th Percentile	29	0.006888
SW RanGradient Median	55	0.003266
SW RanGradient Minimum	21	0.009126
SW Gradient Minimum	51	0.003697
SW Median-Filtered Maximum	35	0.005352
SW Median-Filtered Minimum	20	0.009281
V_R Gradient 25th Percentile	2	0.136583
V_R Gradient Minimum	49	0.003996
V_R Median-Filtered Absolute Maximum	6	0.031014
V_R Median-Filtered Absolute Median	54	0.003290
V_R Median-Filtered Absolute Minimum	38	0.004978
Z_{DR} AzGradient Median	43	0.004538
Z_{DR} RanGradient 25th Percentile	52	0.003636
Z_{DR} RanGradient 75th Percentile	36	0.005239
Z_{DR} RanGradient Median	53	0.003406
Z_{DR} RanGradient Minimum	46	0.004408
Z_{DR} Gradient Minimum	58	0.003064
Z_{DR} Median-Filtered Maximum	24	0.008877
Z_{DR} Median-Filtered Median	32	0.005653
Z_{DR} Median-Filtered Minimum	45	0.004418

The RF training and testing process was performed using the Python machine learning library Scikit-learn and its Random Forest Classifier program (Pedregosa et al. 2011). The training and testing data were split by time period to avoid data contamination across the two datasets. The training data consisted of data from 2011-2016 and the testing data contained data from 2017-2018. A list of optimal hyperparameters was found by tuning the model by iterating through hyperparameter value ranges using the "GridSearchCV" method from Scikit-learn, which uses 5-fold cross validation to reduce the risk of overfitting the model (Table 3). The model accuracy remained the same for both the training (seen) and testing (unseen) data, indicating that the model is unlikely to suffer from overfitting. Once the training and testing of the RF was complete, it was converted into a CSV document that is read in by the algorithm once during its initialization, which takes approximately 0.5-0.6 s.

TABLE 3. List of the hyperparameters used in the Scikit-learn RandomForestClassifier for the final RF model. Hyperparameter descriptions can be found in the Scikit-learn RandomForestClassifier documentation (Pedregosa et al. 2011).

Hyperparameter	Value
bootstrap	True
class_weight	balanced_subsample
criterion	entropy
max_depth	10
max_features	None
max_leaf_nodes	None
min_impurity_decrease	0
min_impurity_split	None
min_samples_leaf	3
min_samples_split	5
min_weight_fraction_leaf	0
n_estimators	500
oob_score	False
random_state	None
warm_start	False

To create the training and testing data for the RF, all severe and sub-severe storm reports (tornado, hail, and thunderstorm wind) where dual-polarization data were available from 2011 to 2018 in the Storm Events Database were retrieved from NCEI (NOAA National Weather Service 1950). The storm reports were used to determine which time periods and WSR-88D radars to download single-radar data for. This was achieved by linearly interpolating each storm report location at every 1-min interval, and acquiring volume data from the nearest radar within 5 min of each interpolated point location, yielding 86,124 0.5°-tilt radar scans. Because the training dataset domain is defined by storm report locations, it will resemble the climatological distribution of storms in the Continental United States (CONUS). Consequently, the central and eastern portions of the CONUS contain the largest concentration of training data point locations (Fig. 4).

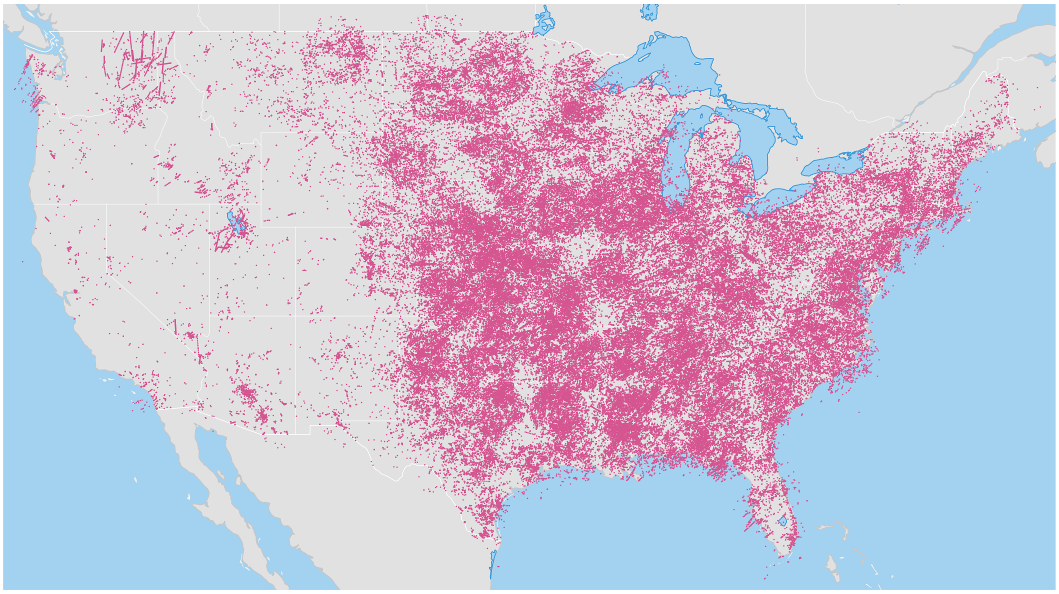


Fig. 4. Data point locations for the training dataset from 2011-2016. Each pink circle corresponds to a data point in or near CONUS. Not shown are 64 data points in Alaska, Hawaii, and Puerto Rico that were included in the training dataset, where the majority were non-tornadic.

Two methods were used to create objects to label for the training dataset; one using storm report locations and one using high-rotation objects following the methods of TORP's AzShear object creation. For the method using storm report locations, tornado reports were limited to those that are tagged as surveyed by the NWS to apply a level of quality control. Additionally, any storm reports that were used to define non-tornadic objects that occurred within an hour of *any* tornado report were excluded. This was done in order to avoid labeling tornadic storms as non-tornadic by using non-tornadic reports in close proximity to a tornado or a pre-tornadic storm.

Each interpolated report location was linked to the area of highest rotation, as defined by AzShear, from a 0.5°-tilt radar scan within 90 sec. The location of the AzShear maximum was also required to be within a variable distance from the data point dependent on how far removed the interpolated point was from the start or end location of the report. This linking method was used because the linearly interpolated path can significantly deviate from the actual tornado path (Fig. 5). Similar to the methods used by Kingfield and LaDue (2015), the default search radius of 10 km was increased by 25 m for each minute removed the interpolated location was from the report start or end location, which was determined after testing various radii on a limited selection of storms. This study used an increased minimum radius (10 km) compared with the 5-7.5-km radius used by Kingfield and LaDue (2015) partially due to including reports from sources other than the NWS. Roughly 2,000 out of 26,000 tornado reports were linked with AzShear maxima more than 7.5 km away. The center of each report object was defined by the location of the AzShear maxima within this search radius, and any duplicate identifications of the same AzShear maximum were removed. A preliminary RF model was trained using only these report objects, but was deemed to produce too many false alarms (non-tornadic objects with high tornado probabilities) in testing. The false alarms from 2011-2016 were introduced to the training dataset to train a new RF model to help the model identify these types of false alarms. These false alarm objects were defined by the AzShear object creation method that is described in the next paragraph, which is identical to how TORP creates automated objects.

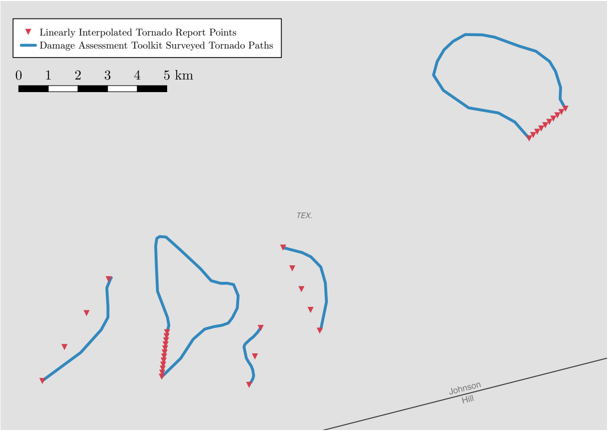

Linearly Interpolated Tornado Report Points
Damage Assessment Toolkit Surveyed Tornado Paths
0 1 2 3 4 5 km
TEX.
Johnson Hill

FIG. 5. Linearly interpolated storm reports (red triangles) compared to the surveyed tornado damage from the Damage Assessment Toolkit tornado paths for five tornadoes in Texas from 27 April 2015.

The second method of creating training objects involved performing an objective search for areas of high rotation over the entirety of every single-radar scan in the dataset. Objects of high rotation were defined as having AzShear values at or above a threshold of 0.006 s^{-1}. Over 98% of radar objects associated with NWS-surveyed quality-controlled tornado reports (excluding cases that were identified as land/waterspouts) in the training dataset were captured by this threshold. The high-rotation objects were labeled as "tornadic" if they were located within 5 min and 20 km of an interpolated tornado report point. While only tornado reports with the "NWS Storm Survey" source tag were used when defining objects following the storm-report identification method, tornado reports from all sources were used to label the high-rotation objects derived from the second object-definition method. This increases confidence in the validity of the tornado reports used in the overall training dataset, because reports that are not associated with rotation are excluded. However, this will still likely exclude some legitimate, but probably weak, tornadoes in the training set.

When only using the locations of non-tornadic storm reports, the preliminary RF assigned high probabilities to objects generated from noisy $\mathrm{V_R}$ data. Consequently, a new and final RF was trained with the original quality-controlled storm report objects, as well as objects created by the second method, which included tornadic objects and false alarm non-tornadic high-rotation objects. The false alarms were defined by being assigned a tornado probability above 50% by the preliminary RF trained on the storm-report objects only. After processing, the training dataset from 2011-2016 totaled 166,145 data points, where 17,336 points (10.4%) were tornadic events.

The 2017-2018 testing dataset included objects defined by both the storm report and high-rotation methods (99,581 data points), as well as additional data that were included to establish a thorough and representative evaluation of the model. These data sources were included to better reflect TORP's real-time performance, and to address some limitations of the training data used to create and tune the RF model. Data from three additional data sources were added to the testing dataset to ensure correct labeling of object types that were never presented to the model during training. These sources included the high-rotation objects with preliminary RF probabilities below 50%, storm objects based on storm reports from sources other than the NWS, and manually identified non-tornadic storm objects.

The first data source appended 137,593 high-rotation objects that were correctly labeled by the preliminary RF as non-tornadic. These objects were reintroduced to confirm that the final RF model also handled these objects appropriately. These non-tornadic objects were defined by having an

AzShear maximum meeting the 0.006-s^{-1} AzShear threshold, and by having no tornado reports within 20 km and 5 min of the object center.

The second data source added the tornado report types that were previously excluded from the testing dataset, which comprises all tornado reports not associated with a high-rotation object with a report source other than “NWS Storm Survey”. While the tornadic high-rotation objects in the training data were potentially associated with tornado reports from other sources, all of the previously excluded tornado report types were added to the complete testing dataset. Any duplicates across the objects defined from the tornado reports and the high-rotation objects were removed, and in order to apply some quality control, any of the tornado report objects with an AzShear maximum below 0.0009 s^{-1} (below 1st percentile of all values associated with severe storm reports from 2011-2018) were removed. This yielded an additional 2,096 tornadic data points.

The third and final additional data source included manually identified non-tornadic data points in or near tornadic environments, which were defined by storms present in the same radar scan as a storm with an ongoing tornado. Any manually identified object within 90 s and 15 km of a tornado was removed to avoid accidental extraction of tornadic data near a non-tornadic storm since the automated data extraction will center on any AzShear rotational maximum within the search radius. The number of manually identified non-tornadic objects totaled 18,067 after applying this filter. A caveat with this data is that it may introduce pre-tornadic objects that are within minutes of producing a tornado, which could for example lead to counting a 60% tornado probability attached to a storm that produces a tornado in the next radar scan as a false alarm. To accurately assess TORP’s pre-tornadic performance, the generation of a dataset containing 0-60 min pre-tornadic manually identified storm locations is currently in progress, which will be part of future work.

The testing dataset totaled 257,097 data points from all of the described storm-report objects, high-rotation objects, and manual objects, and included 17,632 (6.9%) tornadic data points.

4) Algorithm Outputs

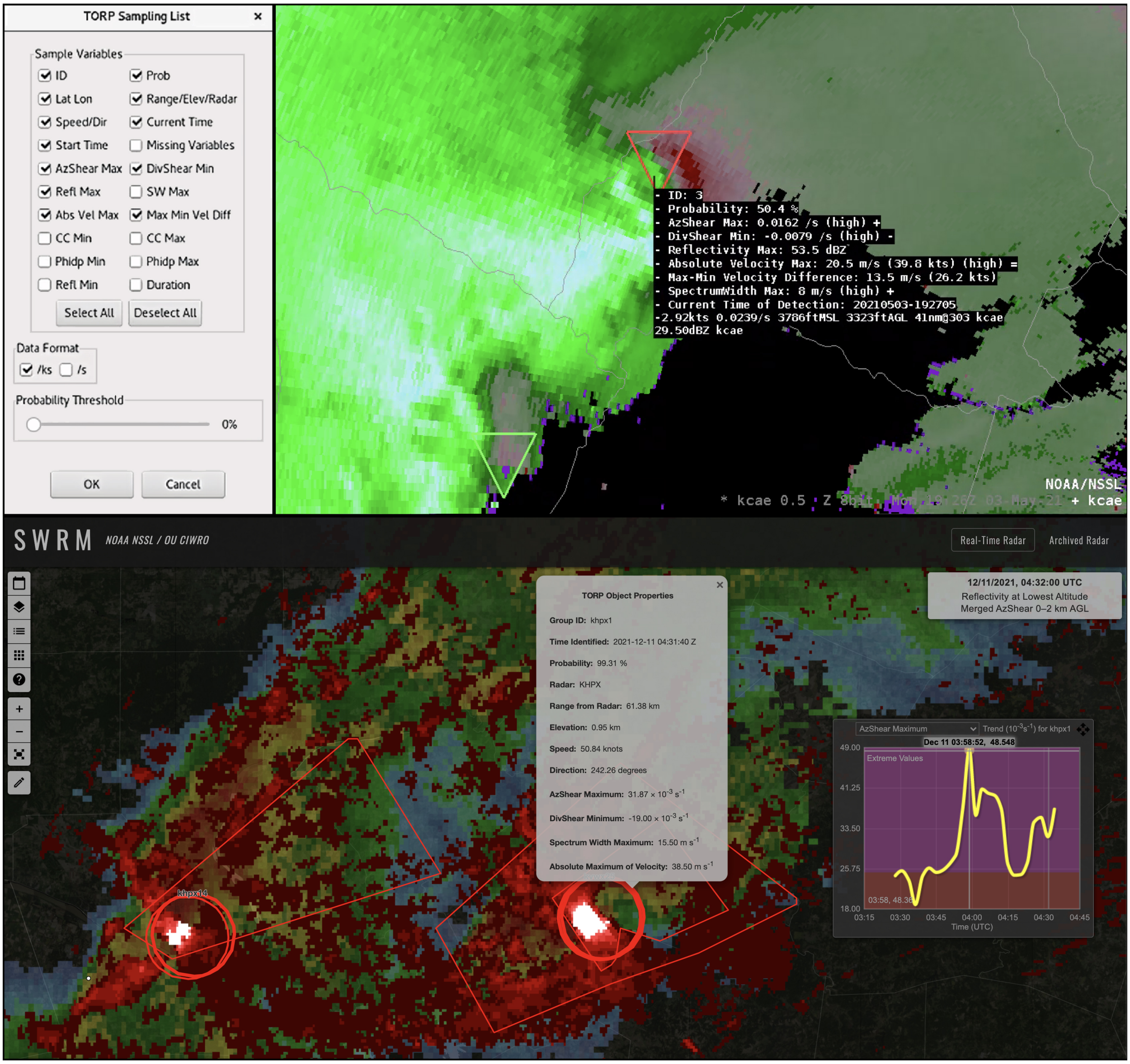
After the object tracking is complete, TORP will write out a text file that includes object location and other metadata, predictor information, trends, and climatological levels for certain predictors. TORP objects can be displayed in the Advanced Weather Interactive Processing System (AWIPS-

II), which is used by NWS forecasters operationally, or other visualization software. In AWIPS-II, a read-out with different information about the object, such as predictor values, becomes available when the user hovers their cursor over the object icon (top panel of Fig. 6). Other visualization features include probability filtering, predictor output customization, and trend graphs (Fig. 6). The climatological levels for select predictors are based on statistics from the storm reports included in TORP's RF training and testing dataset, and labeled with a categorical tag. These tags indicate whether a predictor value is "low", "medium", "high", or "extreme" to contextualize the predictors for users that may not be familiar with what the raw values signify. In the AWIPS-II readout, these are simply displayed as text next to predictor values, however they can also be used in a more visual way as in the Cooperative Institute for Severe and High-Impact Weather Research and Operations (CIWRO)/NSSL Severe Weather Research Map (SWRM), which uses them to shade predictor trend graphs according to the categorical level of the predictor (Fig. 6).

b. Differences Between TORP and other Tornado Algorithms

The goal of TORP is to replace the TDA, and while the guiding principle of tornado detection for the algorithms is the same, TORP and TDA are two vastly different algorithms. The TDA is a binary tornado detection algorithm with objects based on gate-to-gate AzShear from V_R. It uses different V_R-derived products to determine whether these objects are tornadic. The TDA also requires more than one radar tilt scan to build 3D detections. This is not required by TORP, which reduces the time required to create an output product. TORP instead allows the user to browse through detections by tilt to confirm vertical continuity. This method may be more sensitive to noise, however, TORP does have other methods for reducing noise through Z_H masking and by requiring at least a 30% tornado probability for an object within 30 km of a radar to be displayed.

TORP is probabilistic and evaluates all potentially tornadic high-rotation objects, while all output TDA detections are considered a tornado. Although the TDA can use radar Z_H to threshold objects, no information other than V_R-derived products is used to determine whether a circulation is tornadic, whereas TORP uses all available radar information, including dual-polarization products that were not available when the TDA was created. TORP also increases the default algorithm range from 100 km to 160 km away from a radar. The tracking processes for TORP and the TDA are similar, however the main difference is the amount of data that is saved for each object/detection for



TORP Sampling List

Sample Variables

☑ ID ☑ Prob
☑ Lat Lon ☑ Range/Elev/Radar
☑ Speed/Dir ☑ Current Time
☑ Start Time □ Missing Variables
☑ AzShear Max ☑ DivShear Min
☑ Refl Max □ SW Max
☑ Abs Vel Max ☑ Max Min Vel Diff
□ CC Min □ CC Max
□ Phidp Min □ Phidp Max
□ Refl Min □ Duration

Data Format

☑ /ks □ /s

Probability Threshold 0%

```
- ID: 3
- Probability: 50.4 %
- AzShear Max: 0.0162 /s (high) +
- DivShear Min: -0.0079 /s (high) -
- Reflectivity Max: 53.5 dBZ
- Absolute Velocity Max: 20.5 m/s (39.8 kts) (high) =
- Max-Min Velocity Difference: 13.5 m/s (26.2 kts)
- SpectrumWidth Max: 8 m/s (high) +
- Current Time of Detection: 20210503-192705
-2.92kts 0.0239/s 3786ftMSL 3323ftAGL 41nm@303 kcae
29.50dBZ kcae
```

```
* kcae 0.5 Z 8bit Mon 19:26Z 03-May-21 + kcae
```

TORP Object Properties

Group ID: khpx1

Time Identified: 2021-12-11 04:31:40 Z

Probability: 99.31 %

Radar: KHPX

Range from Radar: 61.38 km

Elevation: 0.95 km

Speed: 50.84 knots

Direction: 242.26 degrees

AzShear Maximum: 31.87 × 10^{-3} s^{-1}

DivShear Minimum: -19.00 × 10^{-3} s^{-1}

Spectrum Width Maximum: 15.50 m s^{-1}

Absolute Maximum of Velocity: 38.50 m s^{-1}

12/11/2021, 04:32:00 UTC
Reflectivity at Lowest Altitude
Merged AzShear 0–2 km AGL

FIG. 6. TORP visualization examples.

Top panel: AWIPS-II visualization for a storm system in South Carolina on 3 May 2021 at 1926 UTC.

Bottom panel: SWRM visualization with an AzShear maximum trend graph for an 11 Dec 2021 tornadic supercell in Kentucky at 0432 UTC.

each algorithm, with TORP storing data for all of the additional radar products it uses to calculate tornado probabilities.

An overview of the key differences between TORP and the TDA, HCA TDS, and ProbTor can be found in Table 4. While the differences between TORP and the TDA are the main focus due to the planned direct replacement, it is worth noting some of the similarities and contrasts between TORP and other tornado algorithms. TORP uses many of the same polarimetric products as the HCA TDS. The most important distinction is that the HCA TDS has fixed thresholds, while TORP's RF will adjust the thresholds of the predictors associated with TDSs depending on other predictors, and will weight the products differently. TORP and ProbTor are the most similar in function, since they are both machine learning-based algorithms that provide tornado probabilities. However, there are important distinctions that make them applicable in different situations, such as for linear storms where ProbTor often covers a greater area with a single probability since the objects are not based on rotation like TORP. TORP also takes advantage of polarimetric products, while ProbTor uses several environmental variables along with lightning and MRMS AzShear to derive predictive tornado probabilities.

Table 4. Overview of key differences between TORP and TDA, HCA TDS, and ProbTor.

	TORP	TDA	HCA TDS	ProbTor
Data	Single radar	Single radar	Single radar	Multi radar, lightning, model environment
Data type	Single-radar base and gradient products	Gate-to-gate shear, Z_H	AzShear and Z_H, polarimetric products	MRMS AzShear, ENTLN, RAP
Object Type	AzShear	Gate-to-gate shear	Gates with HCA debris classification	Watershed composite reflectivity
Probabilistic	Yes	No	No	Yes
Method	Machine learning: RF classifier	Thresholding	Thresholding with fuzzy logic	Machine learning: Naïve Bayesian classifier
Output	Icons with customizable hover-over table	Icons with permanent table	Tracks	Areal objects with hover-over table
Prediction	No, but planned	No	No	Yes, 0-60 minutes

c. Performance Evaluation Methods

1) Skill Scores

The performance of TORP was evaluated two-fold: 1) the individual performance of the algorithm and 2) performance when compared against the operational TDA. Due to TORP's automated objects being based on a 0.006-s^{-1} AzShear maximum threshold, any objects not meeting this threshold are assigned an automatic 0% tornado probability to reflect the performance of the automated algorithm. TORP has the capability to evaluate user-defined objects below this threshold, but to accurately demonstrate TORP's real-time performance, any tornado objects missed due to an absent rotation signature in the radar data will be counted as a false negative.

The top section of Table 5 shows the 2 × 2 contingency table that was used to calculate performance statistics when evaluating TORP deterministically at the 50% probability threshold in both an overall evaluation and for the dataset that is used when TORP is compared with the TDA performance. A true positive (tp; hit) is defined as when the algorithm predicts a tornado for an object associated with a tornado report. Similarly, a false positive (fp; false alarm) is when a tornado is predicted for an object, but not observed in the tornado report dataset. False negatives (fn; miss) are tornado report locations where the algorithm did not predict a tornado, and true negatives (tn; correct null) are objects that were correctly predicted to be non-tornadic.

Table 5. Contingency tables for TORP and TDA for testing data from 2017 and 2018. Each table section shows the contingency table for different data evaluations, where different conditions are used to determine what counts as a "predicted tornado" versus a "predicted no tornado".

	Tornado Report	No Tornado Report
Predicted Tornado	True Positive (tp)	False Positive (fp)
Predicted No Tornado	False Negative (fn)	True Negative (tn)
TORP ≥ 50%	9999	10156
TORP ¡ 50%	7633	229309
TORP 1-1 ≥ 50%	3811	780
TORP 1-1 ¡ 50%	2417	11105
TDA	731	1864
No TDA	3683	6645
TDA 1-1	731	30
No TDA 1-1	3683	6645

It is worth noting that storm reports are susceptible to human error, which can lead to errors in location or time (Trapp et al. 2006). Additionally, the absence of a report does not necessarily mean that the weather event did not happen. Potvin et al. (2019) estimated that 45% of tornadoes were not reported in their study domain due to population density bias. As a result, some objects that are identified as false positives are in reality true positives, and some true negatives are actually false negatives. In addition to these limitations, not all tornado reports have available radar data for portions of or over the entirety of the tornado path. Roughly 1% of the original tornado data points were missed due to data limitations.

Nine different performance metrics are calculated for the 2017-2018 datasets: Probability of detection (POD), false alarm ratio (FAR), probability of false detection (POFD), critical success index (CSI), bias, accuracy, Gilbert skill score (GSS), Heidke skill score (HSS), and Peirce skill score (PSS). The equation for each metric is listed in Table 6. POD is the percentage of tornadoes that are detected. FAR is the fraction of false positives to the total number of objects that are predicted to be tornadic. POFD is the number of false alarms to the total number of non-tornadic objects. CSI is often referred to as the skill score and measures the fraction of true positives to the total number of predictions excluding the true negatives. The bias shows whether a forecast of tornadoes is over-predicted or under-predicted. A bias above 1 is indicative of over-prediction, while under 1 is under-prediction. The accuracy measures the fraction of correct predictions to the total number of both tornadic and non-tornadic objects. GSS is similar to CSI, but GSS accounts for hits due to chance, whereas CSI does not (Gilbert 1884; Schaefer 1990). HSS compares the skill of the algorithm with the skill of a random forecast, and PSS estimates how well the algorithm separates tornadic and non-tornadic objects (Peirce 1884; Heidke 1926). For all three of the named skill scores, 0 indicates no skill.

2) Data Stratification

In addition to the overall performance evaluation, TORP's performance for different data categories is explored. The categories include storm type, tornado rating, and NWS Regions.

To allow data stratification for analyzing TORP's performance for different types of storm convective modes, a majority of the 2017-2018 radar scans that included tornadic data were manually analyzed to determine each individual storm's convective mode. The primary convective

TABLE 6. Performance metric formulas and performance values for TORP using 0.5°-tilt radar data at 50% and the TDA.

Performance Metric	Formula	Range	Perfect Score	TORP	TORP 1-1	TDA	TDA 1-1
Probability of Detection	$\mathrm{POD} = \frac{\mathrm{tp}}{\mathrm{tp} + \mathrm{fn}}$	0 - 1	1	0.5647	0.6119	0.1656	0.1656
False Alarm Ratio	$\mathrm{FAR} = \frac{\mathrm{fp}}{\mathrm{tp} + \mathrm{fp}}$	0 - 1	0	0.5040	0.1699	0.7183	0.0394
Probability of False Detection	$\mathrm{POFD} = \frac{\mathrm{fp}}{\mathrm{fp} + \mathrm{tn}}$	0 - 1	0	0.0473	0.0656	0.2191	0.0045
Critical Success Index	$\mathrm{CSI} = \frac{\mathrm{tp}}{\mathrm{tp} + \mathrm{fn} + \mathrm{fp}}$	0 - 1	1	0.3588	0.5438	0.1164	0.1645
Bias	$\mathrm{Bias} = \frac{\mathrm{tp} + \mathrm{fp}}{\mathrm{tp} + \mathrm{fn}}$	0 - ∞	1	1.1385	0.7372	0.5879	0.1724
Accuracy	$\mathrm{Accuracy} = \frac{\mathrm{tp} + \mathrm{tn}}{\mathrm{tp} + \mathrm{fn} + \mathrm{fp} + \mathrm{tn}}$	0 - 1	1	0.9232	0.8235	0.5708	0.6652
Gilbert Skill Score	$\mathrm{GSS} = \frac{(\mathrm{tp}\cdot\mathrm{tn}) - (\mathrm{fp}\cdot\mathrm{fn})}{\big((\mathrm{fp}+\mathrm{fn})\cdot(\mathrm{tp}+\mathrm{fp}+\mathrm{fn}+\mathrm{tn})\big)+\big((\mathrm{tp}\cdot\mathrm{tn})-(\mathrm{fp}\cdot\mathrm{fn})\big)}$	$-\frac{1}{3}$ - 1	1	0.3214	0.4112	-0.0288	0.1034
Heidke Skill Score	$\mathrm{HSS} = \frac{2\cdot\big((\mathrm{tp}\cdot\mathrm{tn})+(\mathrm{fp}\cdot\mathrm{fn})\big)}{(\mathrm{tp}+\mathrm{fn})\cdot(\mathrm{fn}+\mathrm{tn})+(\mathrm{tp}+\mathrm{fp})\cdot(\mathrm{fp}+\mathrm{tn})}$	-∞ - 1	1	0.4865	0.5827	-0.0593	0.1874
Peirce Skill Score	$\mathrm{PSS} = \frac{(\mathrm{tp}\cdot\mathrm{tn})+(\mathrm{fp}\cdot\mathrm{fn})}{(\mathrm{tp}+\mathrm{fn})\cdot(\mathrm{fp}+\mathrm{tn})}$	-1 - 1	1	0.5174	0.5463	-0.0535	0.1611

mode was determined by visual inspection of the Z_H representation of the storm structure. The storms were labeled as either discrete, linear, or tropical cyclone. This labeling was also completed for the non-tornadic objects in tornadic environments.

The tornado rating was extracted for each storm object associated with a tornado report. The data is separated by EF rating, including EFU (unknown rating). To test the statistical significance of the separation of the probability distributions between significant severe tornadoes (EF2+) and the other tornadoes (EF0-1 and EFU), statistical tests were conducted. These included one- and two-sided permutation, one- and two-sided Kolmogorov-Smirnov, and unequal variances t-test, which were performed using SciPy (Virtanen et al. 2020) and mlxtend (Raschka 2018),

The final categorical performance evaluation was performed for different NWS regions to investigate possible consequences of the training dataset climatology and what recommendations would apply to each region. One focus of this analysis was how the NWS Western Region differed from other regions due to the limited number of samples in this region, which includes all of the U.S. West Coast states, Idaho, Montana, Nevada, Arizona, most of Utah, and some counties in Wyoming.

4. Results

1) TORP Overall Performance

TORP's performance varies based on the probability threshold the user picks to distinguish between tornadic and non-tornadic objects, as shown by the teal line in the performance diagram (Fig. 7; Roebber 2009). Choosing a threshold between 40 and 70% will yield a CSI above 0.3. When evaluating tornado probabilities at or above 50% as a predicted tornado, the TORP POD is about 56.5%, while the FAR is 50.4% (Table 6). It shows high accuracy and has a slight over-forecasting bias, with skill scores indicating fair skill from different measures.

TORP's tornado probabilities only show accurate reliability over about 55% (Fig. 8A), and become more reliable with higher probabilities. For example, when TORP assigns a 95% probability, a tornado will occur close to 95% of the time that TORP assigns this probability. Fig. 8A also implies that any probability below 20% should be treated as non-tornadic most of the time, and therefore represents a satisfactory default threshold for when to hide/show TORP objects when displaying TORP using visualization software. Though TORP has an over-forecasting bias, Fig. 8B reveals that TORP assigns tornado probabilities below 20% an overwhelming majority of the time.

TORP is highly reliant on AzShear when assigning tornado probabilities, meaning the tornadoes with lower values for the AzShear maximum are much less likely to have a high tornado probability, which is reflected in the predictor importance (Table 2), where the AzShear maximum (0.29) is more than twice the weight of the second-ranked predictor (0.14) and at least an order of magnitude greater than the rest of the predictors (0.04 and lower).

Evaluating TORP with and without the manually created non-tornadic points in tornadic environments (20,273 points) revealed an FAR difference of only 2-3% (higher FAR with), implying that the algorithm handles non-tornadic points well that are in relatively close proximity to a tornado. However, as mentioned in Section 3a(3), one caveat by introducing this data is that the non-tornadic points may include pre-tornadic storms that are within 1 hour of producing a tornado. When examining the five worst false alarm detections within the manually identified non-tornadic data, as defined by the highest tornado probabilities, two of the false alarms were within 3-10 min of producing a tornado, which arguably is a desirable feature of a tornado probability algorithm. While V_R dealiasing was a potential issue with the other three high-probability false alarms, these

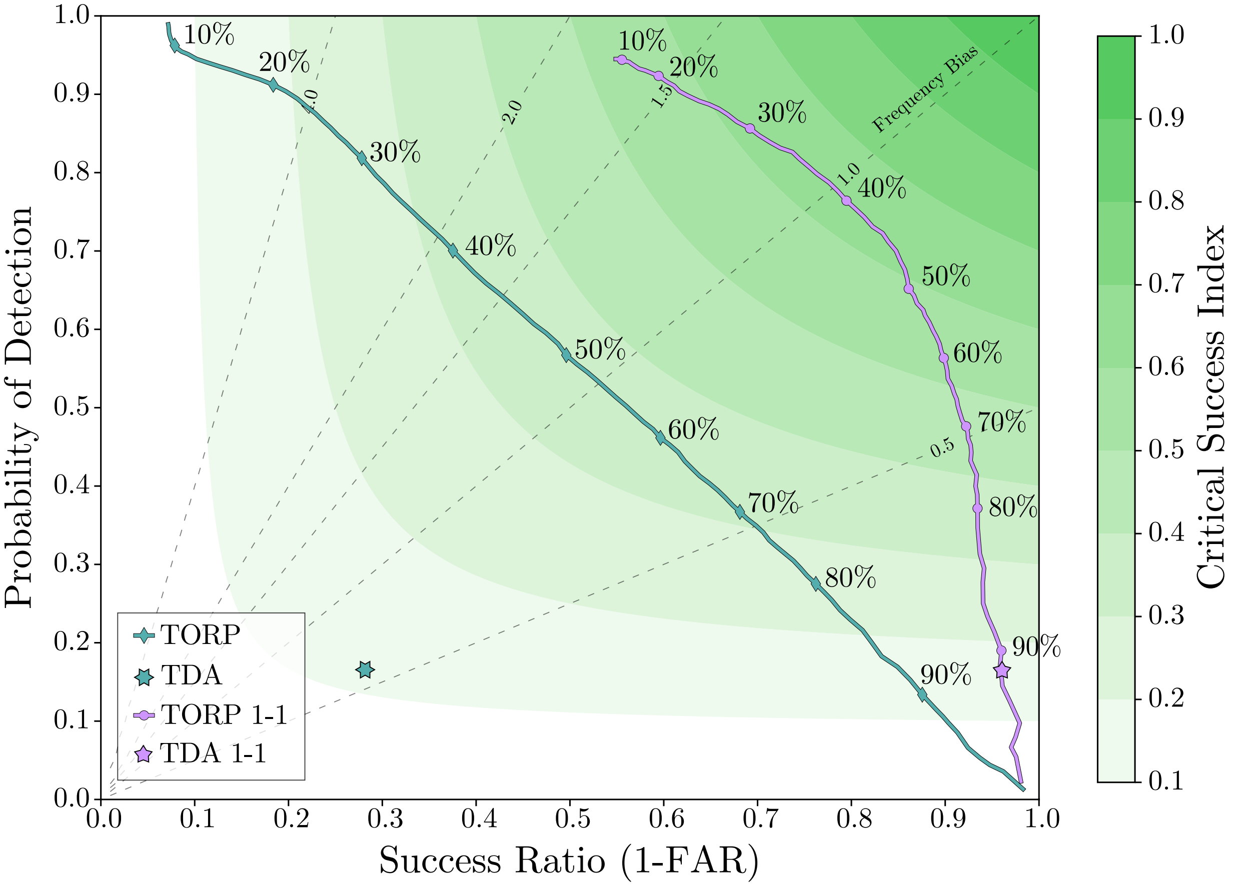


Fig. 7. Performance diagram for the TORP 2017-2018 test dataset based on 0.5°-tilt single-radar data. The lines depict TORP's performance, as defined by the POD as a function of success ratio (1 – FAR), at 1% intervals with every 10% signified by the markers. The teal line with diamond markers represents TORP's overall performance, and the lavender line with circle markers represents TORP's performance when compared 1-to-1 with the TDA. The teal six-pointed star shows the TDA's overall performance, and the lavender five-pointed star shows the TDA's performance for the 1-to-1 comparison with TORP. Performance diagram plotting code adapted from David John Gagne (https://github.com/djgagne).

objects were associated with strong supercells at the time of the highest tornado probability and it is possible that an objective analysis of the radar data predictors is just unable to discern differences between a tornadic and non-tornadic supercell with extreme AzShear rotation values that are above the 95th (84th) percentile when compared with climatological values based on all severe (tornado) storm reports from 2011-2018.

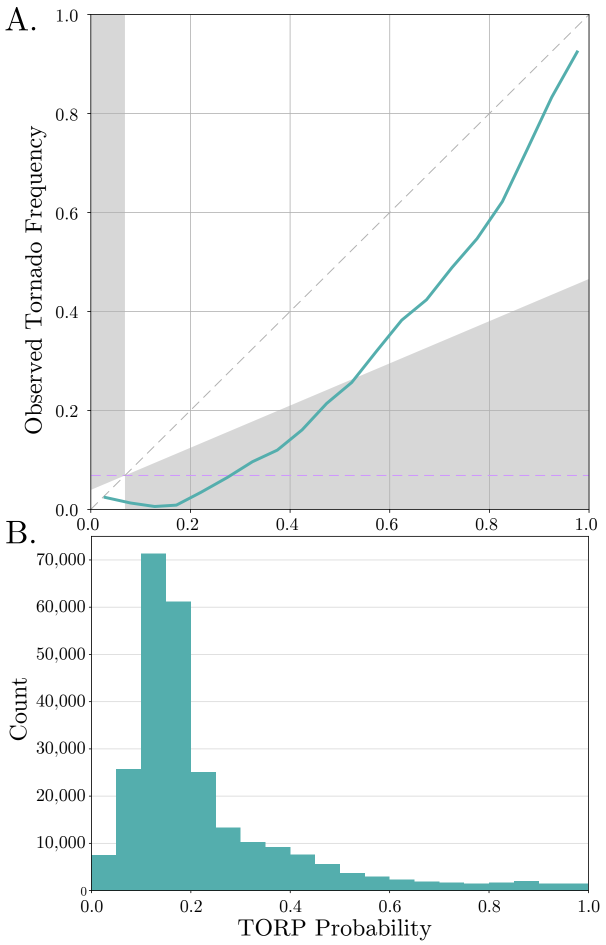


Fig. 8. Panel A is a reliability diagram for the TORP 2017-2018 test dataset. The teal line shows how TORP probabilities relate to the frequency of observed tornadoes in each 5% probability bin. The gray dashed line represents perfect reliability. The lavender dashed line depicts the overall fraction of tornadoes in the test dataset at 6.9%. The gray shaded areas represent poor reliability as defined by the Brier Skill Score (Brier 1950). The number of times TORP assigned probabilities in each bin is depicted in Panel B.

To justify TORP's 160-km default range limit, the performance at the 50% threshold was calculated at each 20-km bin (Fig. 9). The range limit can be turned off, but objects outside of the range limit show a drop in accuracy, and caution should be used beyond 140 km as evident by the spike in the FAR. TORP also has a slight reduction in POD with increasing range from the radar outside of 60 km, with a sharper decline after 160 km. This is likely due to the courser resolution due to changes in the radar beam width and the higher elevation at which the radar measurements are made.

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0123.1.

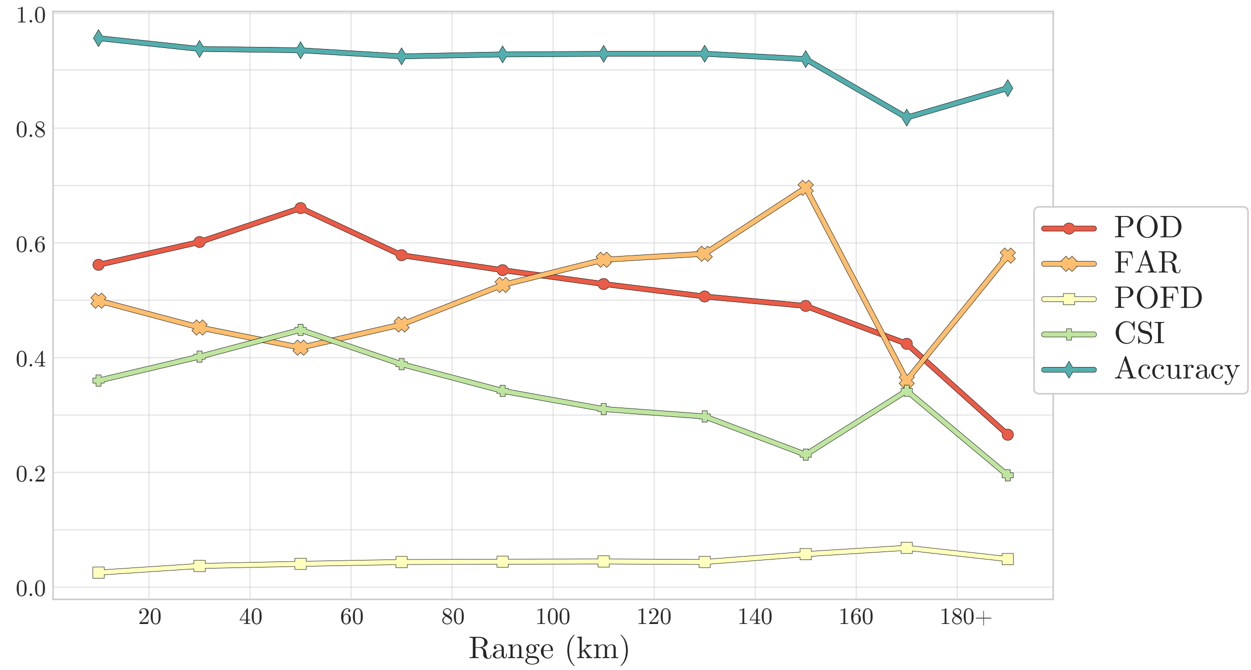


Fig. 9. Select performance metrics at 50% tornado probability as a function of range. The last bin includes every object beyond 180 km.

Overall, TORP shows consistently good performance in distinguishing between tornadic and non-tornadic objects for a very large multi-year dataset with 0.5°-tilt radar data. Fig. 10 shows the receiver operator characteristic (ROC) curve for the TORP test data, indicating skillful results with an area under the curve (AUC) of 0.89 (Mason 1982). The AUC indicates a perfect model at a value of 1, and one of no skill at 0.5 that is illustrated by the dashed line in Fig. 10. This favorable skill assessment is supported by the Gilbert, Heidke, and Peirce skill scores. TORP performs better than random chance and climatology by exceeding 0 for each score. Very limited testing on single-radar data with tilts between 0.3° and 1.9° shows skill in line with this performance evaluation that uses 0.5°-tilt data. More work is needed to evaluate the robustness of this result.

2) TORP Performance by Storm Type

TORP performance did not vary much between storm convective modes, though there are some differences worth noting. The raw numbers for the performance metrics related to non-tornadic objects do not reflect reality as only manually labeled storms were included, leaving no noise objects in the dataset. The datasets included 19,484 data points for discrete storms, 15,205 for

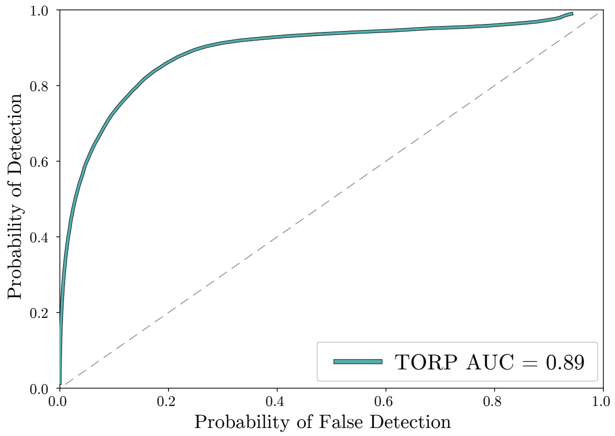


Fig. 10. Receiver operator characteristic curve for TORP in teal, i.e., POD as a function of POFD. The curve for no skill is shown as a dashed line.

linear storms, and 3,004 for tropical cyclones, with 49%, 43%, and 41% tornadic data, respectively. No restrictions or thresholds were applied to the tornadic data. At the 50%-probability threshold, POD and FAR were 57% and 10% for the discrete storms, 56% and 15% for linear storms, and 50% and 13% for tropical cyclones. Quasi-linear convective system (QLCS) events can produce false alarms caused by enhanced AzShear due to the angle between the storm feature and the radar (Mahalik et al. 2019). Fig. 11 shows several areas with probabilities near 50% without tornado reports, though it is possible that weak tornadoes may have been under-reported due to the diurnal variability in reporting and reporting's dependence on population density (Kelly et al. 1978). However, the probability of the objects near tornado reports were generally relatively high when compared with the other areas within the linear system.

While TORP performance is marginally better for discrete storms, further stratification of the discrete category into ordinary cells and supercells will likely increase the difference between the performance for supercells and other types of storms. However, it is encouraging to observe TORP detect non-warned QLCS tornadoes, as well as TC tornadoes, both of which pose forecasting challenges in tornado-warning operations.

3) TORP Performance by Damage Rating

An analysis of how tornado probabilities relate to the tornado damage rating was conducted (Fig. 12). Performing several statistical tests yielded confidence that the mean, median, and

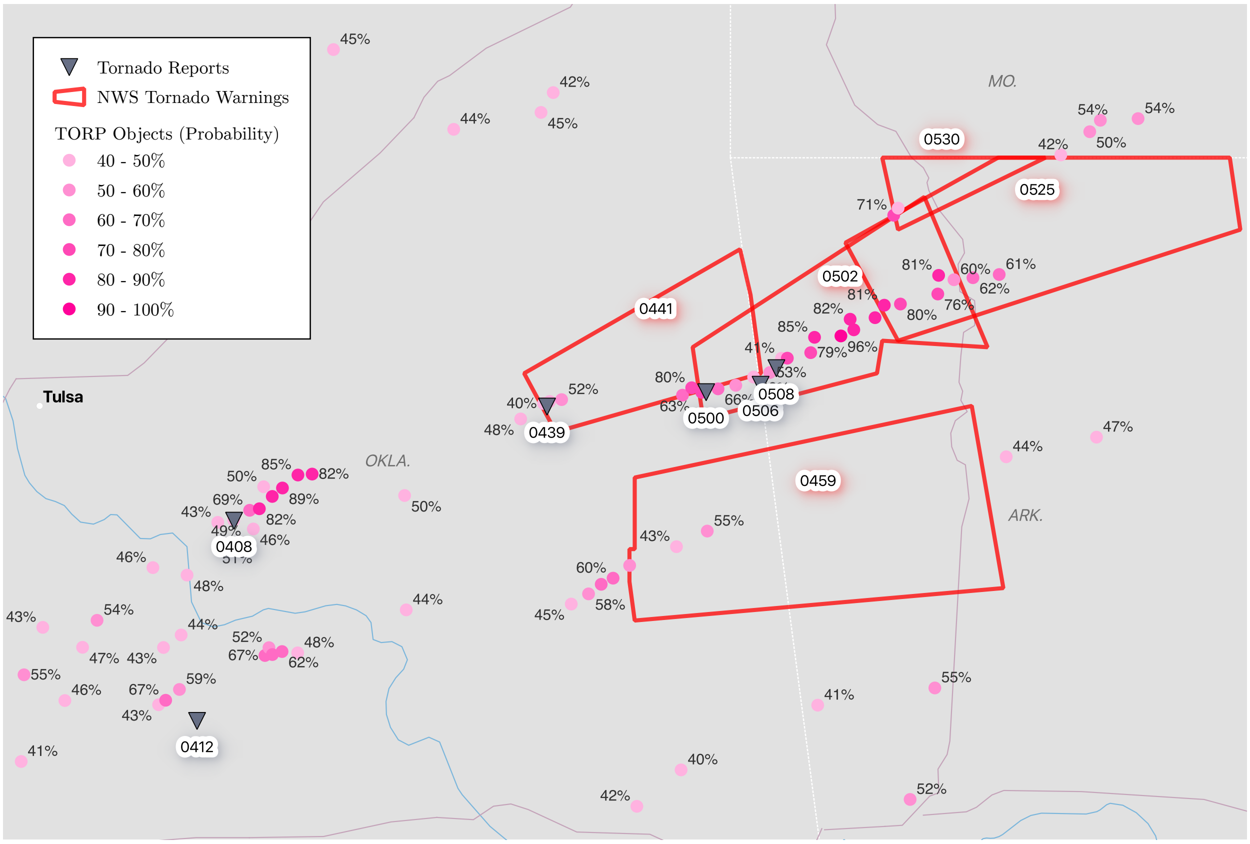


FIG. 11. Accumulation of TORP objects with tornado probabilities ≥ 40% for 2019-10-21 0330-0600 UTC. An overnight linear system moved through OK and AR, producing several tornadoes. Start time stamps are provided for both tornado warnings and reports in UTC.

distribution differences between significant severe (EF2+) tornadoes and other tornadoes (EF0-1 and EFU) are significantly different with no p-value above $8 \cdot 10^{-153}$. The one-sided permutation test and Kolmogorov-Smirnov test confirmed that the median and mean probability, as well as the probability distribution of the EF2+ population, are greater than those of the EF0-1 and EFU population with statistical significance. This indicates that TORP rarely misses more destructive tornadoes, and that most of the false negatives stems from tornadoes of lower or unknown ratings. It also implies that tornadoes capable of producing EF2+-rated damage are more likely to produce higher probabilities. This is likely due to the presence of TDS signatures measured by TORP's predictors since higher-rated tornadoes produce these signatures more frequently (Van Den Broeke and Jauernic 2014).

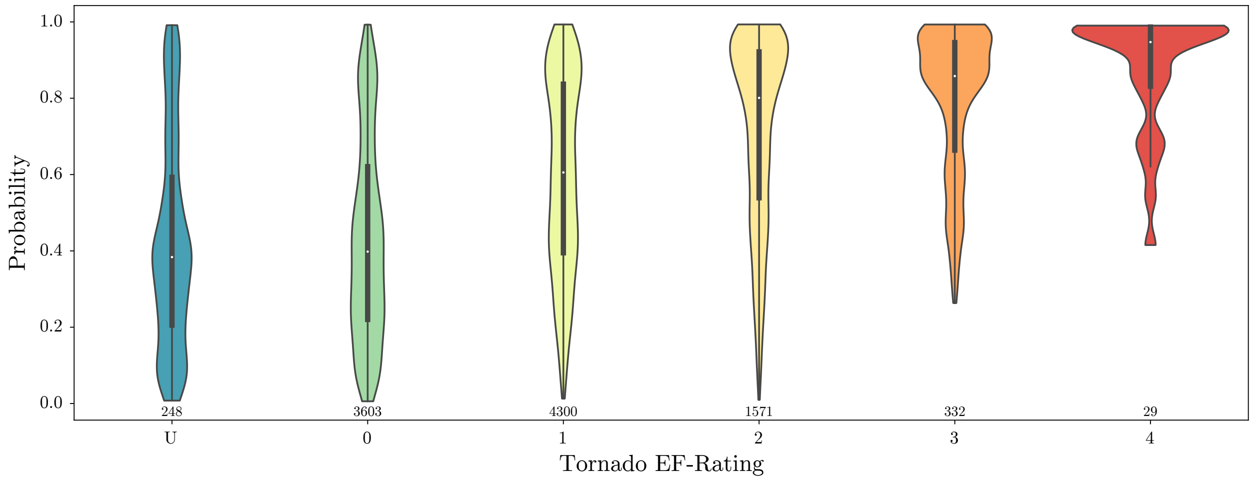


FIG. 12. TORP probability distributions for tornadoes of different EF-ratings illustrated by violin plots showing the 25th, 50th, and 75th percentiles for each population. The number of samples per category is provided below each violin.

While EF ratings are included in tornado reports in the Storm Events Database, the number only represents the maximum rating, while damage can vary greatly within a single tornado track (Burgess et al. 2014). The storm report data are not detailed enough to conclude whether the higher probabilities occur during the time period when the tornado actually produces the highest rated damage.

4) TORP Performance by Region

An analysis using NWS Regions was performed, which revealed that TORP's performance for the NWS Western Region is much lower than the other combined regions. Caution should be used if using the algorithm for this region, as a large reduction in skill was shown at the 50%-probability threshold with a POD of only 17% and a high FAR at 85% which is likely due to the limited samples from Western Region in the training data (10,777 points where 186 were tornadic) and the high fraction of tornadoes in this region being weak and/or landspouts from ordinary cells (Blier and Batten 1994). The POD increased to 27% at the trade-off of a 3% increase in FAR when evaluating the algorithm using a 40% probability, hence a probability threshold reduction when using TORP for Western Region storms (and storms in other regions with similar characteristics) is advisable if using the default RF. While the tornado probabilities are less useful in the NWS Western Region, the algorithm can still be used to condense relevant radar information into a short read-out or to extract trend information for vortices of interest (Fig. 6). Optimally, an NWS Western Region specific RF would be trained and evaluated, however, the training data sample size will be limited by the climatological occurrence of tornadoes in this region, and additional training data collection is not planned at this time.

a. TORP vs. TDA Performance

To be able to compare TORP with the TDA, the "Tornadic Vortex Signature" Level III product was downloaded from the NCEI (NOAA National Weather Service Radar Operations Center 1991; Mitchell et al. 1998). The TDA is a binary algorithm, where a detection is always considered to be a tornado. Therefore, non-tornadic reports in this dataset that did not have a TDA detection associated with them are considered non-tornadic objects for the purpose of this evaluation. Due to the TDA's binary nature, a threshold of 50% is used to indicate a predicted tornado by TORP in order to compare algorithm performance.

Both the operational TDA and TORP were tested on a subset of the 2017-2018 severe storm report dataset, consisting of surveyed tornadoes only and no non-tornadic reports within 1 hour of tornadogenesis. Additionally, only reports within 100 km of a radar were considered to achieve a fair comparison due to the TDA's range limitation. This evaluation is referred to as "TORP 1-1" and "TDA 1-1" in Tables 5 and 6 and Fig. 7. The algorithms are evaluated with the exact same dataset with a total of 11,089 data points, which is limited to when the TDA was available. For this dataset, TORP and the TDA are used to assign a tornado prediction to each severe storm report following the methods for tornado labeling of training dataset objects (objects are tornadic if within 5 min and 20 km of an interpolated report location), meaning non-tornadic objects from either algorithm that are not associated with a severe storm report are not considered in this evaluation to achieve a 1-to-1 comparison.

TORP performs very well within these constraints, and is skillful when distinguishing between tornadic and non-tornadic severe reports in separate convective environments (Table 6 and Fig. 7). The POD is slightly higher than TORP's actual performance, but imposing these limitations on the non-tornadic points sharply reduces the false alarms for TORP. The TDA comparatively has a much lower POD and FAR, indicating a high under-forecasting bias comparable to using TORP with a probability threshold above 90%, resulting in very few detected tornadoes.

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0123.1.

To reflect a more accurate representation of the TDA's performance, it was also evaluated with all false alarms included. The added false alarms are all of the TDA detections that were not associated with any storm report, tornado or otherwise. Shown as "TDA" in Table 6 and Fig. 7, this more realistic evaluation reveals a drastically increased FAR at over 71%. This evaluation is still excluding tornado reports not tagged as "NWS surveyed" from the missed tornadoes and tornadoes outside of a 100-km distance of a radar, so the actual POD is also likely lower. TORP offers a large skill improvement over the TDA by raising the POD and lowering the FAR substantially.

Overall, TORP addresses some of the limitations of the TDA discussed in Mitchell et al. (1998), such as increasing performance outside of 100 km, and achieves a much higher skill than the operational TDA by utilizing new technology. The use of V_R-derived AzShear instead of gate-to-gate shear resolves some data noise issues as well as the radar sampling issue with gate-to-gate shear at very short ranges. Mitchell et al. (1998) also brings up the need to adjust parameter thresholds depending on the environment and region, while the predictor thresholds in TORP are determined by the RF model. Regional and environmental differences can be adjusted for by lowering or increasing probability expectations. However, the interpretation of TORP probabilities could benefit from quantifying the potential relationship between its performance and environmental conditions, which would require additional work. The algorithm mainly running with its default settings may explain the TDA's poor performance on a large dataset compared with the performance for the case studies in Mitchell et al. (1998).

5. Summary and Conclusions

In summary, TORP:

1. Reads in single-radar data for a specified tilt, which by default is 0.5°, as well as the RF model configuration file, and an optional RAP sounding table to calculate 0-6-km storm motion as a first guess for the object tracking process.

2. Creates objects based on a 0.006-s^{-1} AzShear threshold, filters them using filtered Z_H, and merges objects close to one another.

3. Extracts radar data predictors within 2.5 km of the center of each object, which is defined as the location of the object's maximum AzShear.

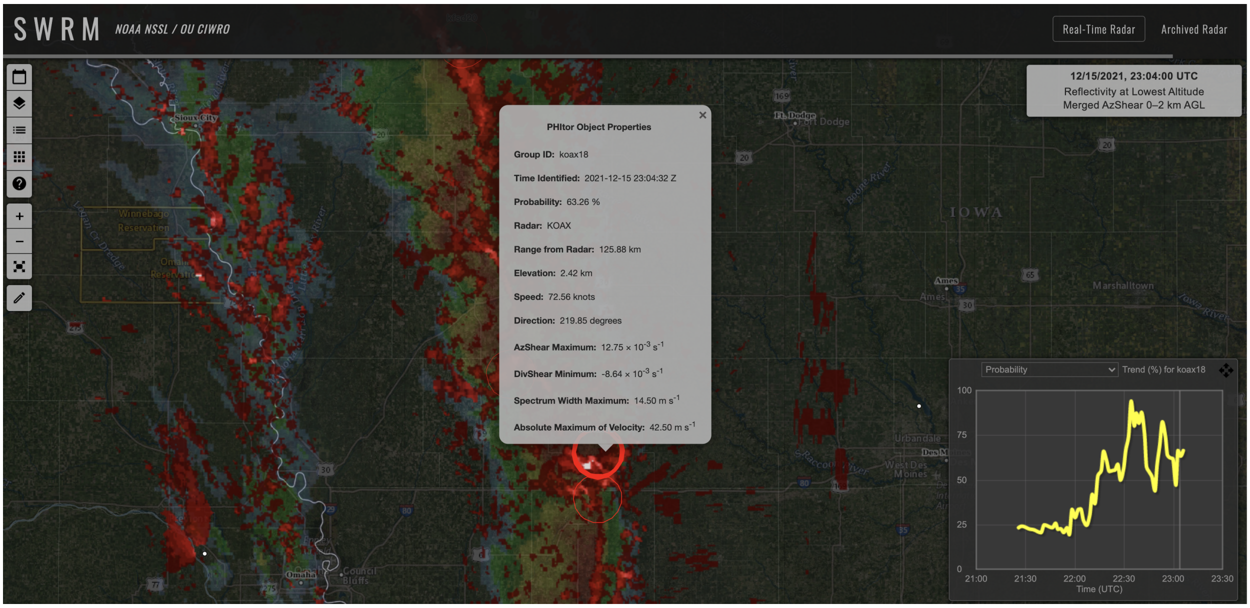
4. Supplies the predictor information to the RF model to calculate a tornado probability.

5. Tracks objects in time using a forecasted position based on the 0-6-km mean storm motion or the mean motion of up to the four most recent storm objects in an object's track history.

6. Outputs a text file that can be used to display the tornado probability objects in a visualization tool.

TORP is intended as a tool to summarize a multitude of radar data variables to encourage confidence in tornado warning decisions for operational forecasters. This could potentially lead to a lower forecaster workload or allow for a quick way to prioritize storms with the highest tornado probabilities. Overall, it performs well as shown by different performance metrics and, although not a 1-to-1 comparison, is comparable with 2016-2018 NWS tornado warning performance with a POD range of 50-62% and FAR of about 70% in the 2016-2018 period as analyzed by Brooks and Correia (2018, more recent years' performance obtained by personal correspondence with Harold Brooks) and Bentley et al. (2021). Notably, TORP is evaluated for continued tornado detection at instantaneous points in time with no temporal context, meaning that it is possible that it detects more than 50% of tornadoes, but does not label them as tornadoes for the whole duration of the storm report time window. In fact, Fig. 11 shows two objects below 50% during the 0500-0508 UTC tornado reports, and Fig. 13 shows the variability in the probability trend, which can change dramatically from one scan to the next. An algorithm-forecaster combination will likely produce a higher POD as the tornado warning duration will exceed the time scale of radar scan updates.

Though the main model evaluation is performed at a 50% threshold, a forecaster could potentially adjust their analysis of TORP objects based on the probability at which the POD and FAR occur that they deem optimal, and can use probability trends to assess potential increasing threats. The evaluation can also be adjusted by calculating performance for specific regions. While difference in storm modes overall did not affect the performance drastically at 50%, it is possible for forecasters to adjust their mental probability thresholds on a case-by-case basis. For example, if a tornado is reported with an object at 40%, a forecaster may pay closer attention to other storms in the same environment that produce similar probabilities. It is also possible to retrain an RF to a certain environment or region, as it is trivial to switch out the RF file. The only limitation to this approach would be the availability of data, as machine learning can be limited by insufficient training data.



12/15/2021, 23:04:00 UTC
Reflectivity at Lowest Altitude
Merged AzShear 0–2 km AGL

PHItor Object Properties

Group ID: koax18

Time Identified: 2021-12-15 23:04:32 Z

Probability: 63.26 %

Radar: KOAX

Range from Radar: 125.88 km

Elevation: 2.42 km

Speed: 72.56 knots

Direction: 219.85 degrees

AzShear Maximum: 12.75×10^{-3} s^{-1}

DivShear Minimum: -8.64×10^{-3} s^{-1}

Spectrum Width Maximum: 14.50 m s^{-1}

Absolute Maximum of Velocity: 42.50 m s^{-1}

FIG. 13. Probability trend example with a tornadic QLCS in NE/IA at 2021-12-15 from ~2130-2304 UTC. This object was associated with the 2250-2308 UTC Atlantic EF-2 tornado and 2234-2242 UTC 3 E Macedonia EF-1 tornado.

A potential avenue for future work could be to extract environmental information to determine possible correlation with TORP's tornado probabilities in order to provide quantified guidance for when to adjust probability expectations.

Currently, one of the biggest failure points of the algorithm is false alarms due to bad V_R data. Because TORP is heavily relying on data derived from V_R data, the algorithm predictions become less reliable when the V_R data is inaccurate. This is often due to dealiasing failure or side-lobe contamination from severe hail, which can sometimes be challenging to discern by forecasters with less experience in radar data examination (Boettcher and Bentley 2022). V_R spikes due to ground clutter and other non-meteorological features usually fall below 10-20% tornado probability when evaluated by the RF, but can occasionally produce a random tornado probability exceeding 50%, however, these objects can easily be disregarded by a forecaster. The largest concern with the obvious false alarms would be potential distrust in the algorithm, and sufficient training to make users aware of the algorithm's weaknesses is essential to operational success.

TORP has received favorable reviews by NWS forecasters in the Hazardous Weather Testbed Experimental Warning Program virtual 2021 experiment after several weeks of testing the algorithm in real-time and archived severe weather case studies (Sandmæl et al. 2022). Many forecaster suggestions for product improvement have already been incorporated in TORP, and inquiries for opinions on ongoing efforts for future features were well received. These features include incorporating pre-tornadic data and including probabilities of tornado impact by including population data to aid confidence in impact-based tornado warning damage tags, such as "considerable" and "catastrophic". The incorporation of pre-tornadic data includes the generation of a 0-60-min manually identified pre-tornadic dataset that is nearing completion, and will allow an analysis of lead time performance of TORP and the potential for training a pre-tornadic RF to make a distinction between the pre-tornadic and currently tornadic stages of a storm. Another planned effort is to expand the evaluation of TORP for single-radar tilts other than the 0.5° tilt, and train more RF models to improve performance if needed. This can be especially important for radars relying on different tilts to adequately sample tornadic storms.

The goal of TORP, similar to that of the original TDA (Mitchell et al. 1998), is to be a guidance tool for operational forecasters, and should contribute to enhanced performance using the combination of human knowledge and experience and the analytical advantages of machine learning. TORP will provide a tool for forecasters to assess and track potentially tornadic circulations, giving a quick overview of the current radar attributes of a circulation and how it has evolved with time. As forecasters see increasing probabilities and strengthening predictors, TORP can provide confidence in making swift tornado warning decisions. TORP can also provide guidance in situations where real-time ground reports are lacking, such as with overnight QLCS tornadoes, to reduce the number of non-warned tornadoes such as the two tornadoes south of Tulsa in Fig. 11.

Acknowledgments. The authors would like to thank the three anonymous reviewers and editor John T. Allen, as well as Eric D. Loken, for their invaluable feedback during the review process of this manuscript. Funding was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA21OAR4320204, U.S. Department of Commerce.

Data availability statement. The data used in this study are openly available from NOAA National Centers for Environmental Information at doi.org/10.7289/V5W9574V and www.ncdc.noaa.gov/stormevents.

References

Ansari, S., and Coauthors, 2018: Unlocking the potential of NEXRAD data through NOAA's Big Data Partnership. *Bull. Amer. Meteor. Soc.*, **99 (1)**, 189 – 204, https://doi.org/10.1175/BAMS-D-16-0021.1, URL https://journals.ametsoc.org/view/journals/bams/99/1/bams-d-16-0021.1.xml.

Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144 (4)**, 1669 – 1694, https://doi.org/10.1175/MWR-D-15-0242.1, URL https://journals.ametsoc.org/view/journals/mwre/144/4/mwr-d-15-0242.1.xml.

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. *Wea. Forecasting*, **36 (6)**, 1909 – 1924, https://doi.org/10.1175/WAF-D-20-0241.1, URL https://journals.ametsoc.org/view/journals/wefo/36/6/WAF-D-20-0241.1.xml.

Blier, W., and K. A. Batten, 1994: On the incidence of tornadoes in California. *Wea. Forecasting*, **9 (3)**, 301 – 315, https://doi.org/10.1175/1520-0434(1994)009⟨0301:OTIOTI⟩2.0.CO;2, URL https://journals.ametsoc.org/view/journals/wefo/9/3/1520-0434_1994_009_0301_otioti_2_0_co_2.xml.

Bodine, D. J., M. R. Kumjian, R. D. Palmer, P. L. Heinselman, and A. V. Ryzhkov, 2013: Tornado damage estimation using polarimetric radar. *Wea. Forecasting*, **28 (1)**, 139 – 158, https://doi.org/10.1175/WAF-D-11-00158.1, URL https://journals.ametsoc.org/view/journals/wefo/28/1/waf-d-11-00158_1.xml.

Boettcher, J. B., and E. S. Bentley, 2022: WSR-88D sidelobe contamination: From a conceptual model to diagnostic strategies for improving NWS warning performance. *Wea. Forecasting*, **37 (6)**, 853 – 869, https://doi.org/10.1175/WAF-D-21-0155.1, URL https://journals.ametsoc.org/view/journals/wefo/37/6/WAF-D-21-0155.1.xml.

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

Bravais, A., 1844: *Analyse mathématique sur les probabilités des erreurs de situation d'un point*. Impr. Royale.

Breiman, L., 2001: Random forests. *Machine learning*, **45 (1)**, 5–32.

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78 (1)**, 1 – 3, https://doi.org/10.1175/1520-0493(1950)078⟨0001:VOFEIT⟩2.0.CO;2, URL https://journals.ametsoc.org/view/journals/mwre/78/1/1520-0493_1950_078_0001_vofeit_2_0_co_2.xml.

Brooks, H. E., and J. Correia, 2018: Long-term performance metrics for National Weather Service tornado warnings. *Wea. Forecasting*, **33 (6)**, 1501 – 1511, https://doi.org/10.1175/WAF-D-18-0120.1, URL https://journals.ametsoc.org/view/journals/wefo/33/6/waf-d-18-0120_1.xml.

Brown, R. A., 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. *Wea. Forecasting*, **20 (1)**, 3 – 14, https://doi.org/10.1175/WAF-832.1, URL https://journals.ametsoc.org/view/journals/wefo/20/1/waf-832_1.xml.

Brown, R. A., L. R. Lemon, and D. W. Burgess, 1978: Tornado detection by pulsed Doppler radar. *Mon. Wea. Rev.*, **106 (1)**, 29 – 38, https://doi.org/10.1175/1520-0493(1978)106⟨0029:TDBPDR⟩2.0.CO;2, URL https://journals.ametsoc.org/view/journals/mwre/106/1/1520-0493_1978_106_0029_tdbpdr_2_0_co_2.xml.

Brown, R. A., and V. T. Wood, 2012: The tornadic vortex signature: An update. *Wea. Forecasting*, **27 (2)**, 525 – 530, https://doi.org/10.1175/WAF-D-11-00111.1, URL https://journals.ametsoc.org/view/journals/wefo/27/2/waf-d-11-00111_1.xml.

Brown, R. A., V. T. Wood, and D. Sirmans, 2002: Improved tornado detection using simulated and actual WSR-88D data with enhanced resolution. *J. Atmos. Oceanic Technol.*, **19 (11)**, 1759 – 1771, https://doi.org/10.1175/1520-0426(2002)019⟨1759:ITDUSA⟩2.0.CO;2, URL https://journals.ametsoc.org/view/journals/atot/19/11/1520-0426_2002_019_1759_itdusa_2_0_co_2.xml.

Burgess, D., and Coauthors, 2014: 20 May 2013 Moore, Oklahoma, tornado: Damage survey and analysis. *Wea. Forecasting*, **29 (5)**, 1229–1237.

Burgess, D. W., L. R. Lemon, and R. A. Brown, 1975: Tornado characteristics revealed by Doppler radar. *Geophys. Res. Lett.*, **2**, 183–184.

Chisholm, A. J., 1973: Alberta hailstorms part I: Radar case studies and airflow models. *Alberta Hailstorms, Meteor. Monogr.*, Amer. Meteor. Soc., 1–36.

Cintineo, J. L., M. J. Pavolonis, J. M. Sieglaff, L. Cronce, and J. Brunner, 2020: NOAA ProbSevere v2.0—ProbHail, ProbWind, and ProbTor. *Wea. Forecasting*, **35 (4)**, 1523 – 1543, https://doi.org/10.1175/WAF-D-19-0242.1, URL https://journals.ametsoc.org/view/journals/wefo/35/4/wafD190242.xml.

Cintineo, J. L., M. J. Pavolonis, J. M. Sieglaff, and D. T. Lindsey, 2014: An empirical model for assessing the severe weather potential of developing convection. *Wea. Forecasting*, **29 (3)**, 639 – 653, https://doi.org/10.1175/WAF-D-13-00113.1, URL https://journals.ametsoc.org/view/journals/wefo/29/3/waf-d-13-00113_1.xml.

Crum, T. D., and R. L. Alberty, 1993: The WSR-88D and the WSR-88D operational support facility. *Bull. Amer. Meteor. Soc.*, **74 (9)**, 1669 – 1688, https://doi.org/10.1175/1520-0477(1993)074⟨1669:TWATWO⟩2.0.CO;2, URL https://journals.ametsoc.org/view/journals/bams/74/9/1520-0477_1993_074_1669_twatwo_2_0_co_2.xml.

Donaldson, R. J., 1970: Vortex signature recognition by a Doppler radar. *J. Appl. Meteor. Climatol.*, **9 (4)**, 661 – 670, https://doi.org/10.1175/1520-0450(1970)009⟨0661:VSRBAD⟩2.0.CO;2, URL https://journals.ametsoc.org/view/journals/apme/9/4/1520-0450_1970_009_0661_vsrbad_2_0_co_2.xml.

French, M. M., and D. M. Kingfield, 2019: Dissipation characteristics of tornadic vortex signatures associated with long-duration tornadoes. *J. Appl. Meteor. Climatol.*, **58 (2)**, 317 – 339, https://doi.org/10.1175/JAMC-D-18-0187.1, URL https://journals.ametsoc.org/view/journals/apme/58/2/jamc-d-18-0187.1.xml.

Fujita, T. T., 1973: Proposed mechanism of tornado formation from rotating thunderstorms. American Meteorological Society, Denver, CO, 191–196.

Gensini, V. A., C. Converse, W. S. Ashley, and M. Taszarek, 2021: Machine learning classification of significant tornadoes and hail in the United States using ERA5 proximity soundings. *Wea. Forecasting*, **36 (6)**, 2143 – 2160, https://doi.org/10.1175/WAF-D-21-0056.1, URL https://journals.ametsoc.org/view/journals/wefo/36/6/WAF-D-21-0056.1.xml.

Gibbs, J. G., 2016: A skill assessment of techniques for real-time diagnosis and short-term prediction of tornado intensity using the WSR-88D. *J. Operational Meteor.*, **4 (13)**, 170–181, https://doi.org/http://dx.doi.org/10.15191/nwajom.2016.0413.

Gilbert, G. K., 1884: Finley's tornado predictions. *American Meteorological Journal. A Monthly Review of Meteorology and Allied Branches of Study (1884-1896)*, **1 (5)**, 166.

Heidke, P., 1926: Berechnung des erfolges und der güte der windstärkevorhersagen im sturmwarnungsdienst. *Geografiska Annaler*, **8 (4)**, 301–349, https://doi.org/10.1080/20014422.1926.11881138, URL https://doi.org/10.1080/20014422.1926.11881138.

Ho, T. K., 1998: The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20 (8)**, 832–844, https://doi.org/10.1109/34.709601.

Homeyer, C. R., T. N. Sandmæl, C. K. Potvin, and A. M. Murphy, 2020: Distinguishing characteristics of tornadic and nontornadic supercell storms from composite mean analyses of radar observations. *Mon. Wea. Rev.*, **148 (12)**, 5015 – 5040, https://doi.org/10.1175/MWR-D-20-0136.1, URL https://journals.ametsoc.org/view/journals/mwre/148/12/mwr-d-20-0136.1.xml.

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

Jing, Z., and G. Wiener, 1993: Two-dimensional dealiasing of Doppler velocities. *J. Atmos. Oceanic Technol.*, **10 (6)**, 798 – 808, https://doi.org/10.1175/1520-0426(1993)010⟨0798:TDDODV⟩2.0.CO;2, URL https://journals.ametsoc.org/view/journals/atot/10/6/1520-0426_1993_010_0798_tddodv_2_0_co_2.xml.

Johnson, J. T., P. L. MacKeen, A. Witt, E. D. W. Mitchell, G. J. Stumpf, M. D. Eilts, and K. W. Thomas, 1998: The storm cell identification and tracking algorithm: An enhanced WSR-88D algorithm. *Wea. Forecasting*, **13 (2)**, 263 – 276, https://doi.org/10.1175/1520-0434(1998)013⟨0263:TSCIAT⟩2.0.CO;2, URL https://journals.ametsoc.org/view/journals/wefo/13/2/1520-0434_1998_013_0263_tsciat_2_0_co_2.xml.

Karstens, C. D., and Coauthors, 2015: Evaluation of a probabilistic forecasting methodology for severe convective weather in the 2014 Hazardous Weather Testbed. *Wea. Forecasting*, **30 (6)**, 1551 – 1570, https://doi.org/10.1175/WAF-D-14-00163.1, URL https://journals.ametsoc.org/view/journals/wefo/30/6/waf-d-14-00163_1.xml.

Kelly, D. L., J. T. Schaefer, R. P. McNulty, C. A. Doswell, and R. F. Abbey, 1978: An augmented tornado climatology. *Mon. Wea. Rev.*, **106 (8)**, 1172 – 1183, https://doi.org/10.1175/1520-0493(1978)106⟨1172:AATC⟩2.0.CO;2, URL https://journals.ametsoc.org/view/journals/mwre/106/8/1520-0493_1978_106_1172_aatc_2_0_co_2.xml.

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 (5)**, 1125–1139, https://doi.org/10.1175/WAF-D-14-00096.1, URL https://doi.org/10.1175/WAF-D-14-00096.1.

Kumjian, M. R., and A. V. Ryzhkov, 2008: Polarimetric signatures in supercell thunderstorms. *J. Appl. Meteor. Climatol.*, **47 (7)**, 1940 – 1961, https://doi.org/10.1175/2007JAMC1874.1, URL https://journals.ametsoc.org/view/journals/apme/47/7/2007jamc1874.1.xml.

Lagerquist, R., A. McGovern, C. R. Homeyer, D. J. G. II, and T. Smith, 2020: Deep learning on three-dimensional multiscale data for next-hour tornado prediction. *Mon. Wea. Rev.*, **148 (7)**, 2837 – 2861, https://doi.org/10.1175/MWR-D-19-0372.1, URL https://journals.ametsoc.org/view/journals/mwre/148/7/mwrD190372.xml.

Lagerquist, R., A. McGovern, and T. Smith, 2017: Machine learning for real-time prediction of damaging straight-line convective wind. *Wea. Forecasting*, **32 (6)**, 2175 – 2193, https://doi.org/10.1175/WAF-D-17-0038.1, URL https://journals.ametsoc.org/view/journals/wefo/32/6/waf-d-17-0038_1.xml.

Lemon, L. R., R. J. Donaldson, D. W. Burgess, and R. A. Brown, 1977: Doppler radar application to severe thunderstorm study and potential real-time warning. *Bull. Amer. Meteor. Soc.*, **58 (11)**, 1187 – 1193, https://doi.org/10.1175/1520-0477(1977)058⟨1187:DRATST⟩2.0.CO;2, URL https://journals.ametsoc.org/view/journals/bams/58/11/1520-0477_1977_058_1187_dratst_2_0_co_2.xml.

Loeffler, S. D., M. R. Kumjian, M. Jurewicz, and M. M. French, 2020: Differentiating between tornadic and nontornadic supercells using polarimetric radar signatures of hydrometeor size sorting. *Geophys. Res. Lett.*, **47 (12)**, e2020GL088 242, https://doi.org/https://doi.org/10.1029/2020GL088242.

Losey-Bailor, A., W. D. Zittel, and Z. Jing, 2019: Improving Doppler velocity coverage on the WSR-88D by using low PRFs with 2DVDA. Tech. rep., NWS Radar Operations Center, 16 pp. URL https://roc.noaa.gov/wsr88d/PublicDocs/Publications/Losey-BailorEtAl2019_ImprovingDoppVelCoverageLowPRFs_39thICRM.pdf.

Lyza, A. W., M. D. Flournoy, and E. N. Rasmussen, 2022: Observed characteristics of the tornadic supercells of 27-28 April 2011 in the Southeast United States. *Mon. Wea. Rev.*, https://doi.org/10.1175/MWR-D-21-0274.1, URL https://journals.ametsoc.org/view/journals/mwre/aop/MWR-D-21-0274.1/MWR-D-21-0274.1.xml.

Mahalik, M. C., B. R. Smith, K. L. Elmore, D. M. Kingfield, K. L. Ortega, and T. M. Smith, 2019: Estimates of gradients in radar moments using a linear least squares derivative technique. *Wea. Forecasting*, **34 (2)**, 415–434, https://doi.org/10.1175/WAF-D-18-0095.1, URL https://doi.org/10.1175/WAF-D-18-0095.1, https://journals.ametsoc.org/waf/article-pdf/34/2/415/4866756/waf-d-18-0095_1.pdf.

Markowski, P. M., 2002: Hook echoes and rear-flank downdrafts: A review. *Mon. Wea. Rev.*, **130 (4)**, 852 – 876, https://doi.org/10.1175/1520-0493(2002)130⟨0852:HEARFD⟩2.0.CO;2, URL https://journals.ametsoc.org/view/journals/mwre/130/4/1520-0493_2002_130_0852_hearfd_2.0.co_2.xml.

Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30 (4)**, 291–303.

McGovern, A., K. L. Elmore, D. J. Gagne, S. E. Haupt, C. D. Karstens, R. Lagerquist, T. Smith, and J. K. Williams, 2017: Using artificial intelligence to improve real-time decision-making for high-impact weather. *Bull. Amer. Meteor. Soc.*, **98 (10)**, 2073 – 2090, https://doi.org/10.1175/BAMS-D-16-0123.1, URL https://journals.ametsoc.org/view/journals/bams/98/10/bams-d-16-0123.1.xml.

Mecikalski, J. R., T. N. Sandmæl, E. M. Murillo, C. R. Homeyer, K. M. Bedka, J. M. Apke, and C. P. Jewett, 2021: A random-forest model to assess predictor importance and nowcast severe storms using high-resolution radar–goes satellite–lightning observations. *Mon. Wea. Rev.*, **149 (6)**, 1725 – 1746, https://doi.org/10.1175/MWR-D-19-0274.1, URL https://journals.ametsoc.org/view/journals/mwre/149/6/MWR-D-19-0274.1.xml.

Mitchell, E. D. W., S. V. Vasiloff, G. J. Stumpf, A. Witt, M. D. Eilts, J. T. Johnson, and K. W. Thomas, 1998: The National Severe Storms Laboratory tornado detection algorithm. *Wea. Forecasting*, **13 (2)**, 352 – 366, https://doi.org/10.1175/1520-0434(1998)013⟨0352:TNSSLT⟩2.0.CO;2, URL https://journals.ametsoc.org/view/journals/wefo/13/2/1520-0434_1998_013_0352_tnsslt_2_0_co_2.xml.

NEXRAD Joint System Program Office, 1985: Next Generation Weather Radar (NEXRAD) algorithm report. Tech. rep., Silver Spring, MD, 738 pp. URL https://ia800309.us.archive.org/33/items/nextgenerationwe00nexr/nextgenerationwe00nexr.pdf.

NOAA National Weather Service, 1950: Storm Events Database. NOAA National Centers for Environmental Information, URL https://www.ncdc.noaa.gov/stormevents, accessed April 2019 - April 2020.

NOAA National Weather Service Radar Operations Center, 1991: NOAA Next Generation Radar (NEXRAD) Level 2 base data. NOAA National Centers for Environmental Information, accessed April 2019 - April 2020, https://doi.org/10.7289/V5W9574V.

NWS Radar Operations Center Applications Branch, 2021: Interface control document for the RDA/RPG. NWS Doc. 2620003AA, RDA build 20.0. Tech. rep., 193 pp. URL https://www.roc.noaa.gov/WSR88D/PublicDocs/ICDs/2620003AA.pdf.

Park, H. S., A. V. Ryzhkov, D. S. Zrnić, and K.-E. Kim, 2009: The hydrometeor classification algorithm for the polarimetric wsr-88d: Description and application to an mcs. *Wea. Forecasting*, **24 (3)**, 730 – 748, https://doi.org/10.1175/2008WAF2222205.1, URL https://journals.ametsoc.org/view/journals/wefo/24/3/2008waf2222205_1.xml.

Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, **12 (85)**, 2825–2830, URL http://jmlr.org/papers/v12/pedregosa11a.html.

Peirce, C. S., 1884: The numerical measure of the success of predictions. *Science*, **NS-4 (93)**, 453–454.

Potvin, C. K., C. Broyles, P. S. Skinner, H. E. Brooks, and E. Rasmussen, 2019: A Bayesian hierarchical modeling framework for correcting reporting bias in the U.S. tornado database. *Wea. Forecasting*, **34 (1)**, 15 – 30, https://doi.org/10.1175/WAF-D-18-0137.1, URL https://journals.ametsoc.org/view/journals/wefo/34/1/waf-d-18-0137_1.xml.

Raschka, S., 2018: Mlxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *The Journal of Open Source Software*, **3 (24)**, https://doi.org/10.21105/joss.00638, URL http://joss.theoj.org/papers/10.21105/joss.00638.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24 (2)**, 601 – 608, https://doi.org/10.1175/2008WAF2222159.1, URL https://journals.ametsoc.org/view/journals/wefo/24/2/2008waf2222159_1.xml.

Ryzhkov, A. V., T. J. Schuur, D. W. Burgess, and D. S. Zrnic, 2005: Polarimetric tornado detection. *J. Appl. Meteor.*, **44 (5)**, 557 – 570, https://doi.org/10.1175/JAM2235.1, URL https://journals.ametsoc.org/view/journals/apme/44/5/jam2235.1.xml.

Sandmæl, T. N., C. R. Homeyer, K. M. Bedka, J. M. Apke, J. R. Mecikalski, and K. Khlopenkov, 2019: Evaluating the ability of remote sensing observations to identify significantly severe and potentially tornadic storms. *J. Appl. Meteor. Climatol.*, **58 (12)**, 2569 – 2590, https://doi.org/10.1175/JAMC-D-18-0241.1, URL https://journals.ametsoc.org/view/journals/apme/58/12/jamc-d-18-0241.1.xml.

Sandmæl, T. N., and A. E. Reinhart, 2022: Using linear least square shear product signatures from single radars to evaluate tornado potential for quasi-linear convective system circulations. *Symposium on Radar Science in the Service of Earth System Predictability*, American Meteorological Society, Virtual, 14.3, URL https://ams.confex.com/ams/102ANNUAL/meetingapp.cgi/Paper/393253.

Sandmæl, T. N., B. R. Smith, J. W. Monroe, J. G. Madden, P. T. Hyland, and B. A. Schenkel, 2022: The 2021 Hazardous Weather Testbed Experimental Warning Program Radar Convective Applications Experiment: Evaluating the Tornado Potential Algorithm and the AzShear Rotation Detection Algorithm. *31st Conference on Weather Analysis and Forecasting (WAF)/27th Conference on Numerical Weather Prediction (NWP)*, American Meteorological Society, Virtual, J15B.4, URL https://ams.confex.com/ams/102ANNUAL/meetingapp.cgi/Paper/393261.

Saxion, D. S., and R. L. Ice, 2012: New science for the WSR-88D: Status of the dual-polarization upgrade. *28th Conf. on Interactive Information Processing Systems*, American Meteorological Society, New Orleans, LA, 5, URL https://roc.noaa.gov/wsr88d/PublicDocs/Publications/DP_Status_28th_IIPS_Jan2012.pdf.

Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting*, **5 (4)**, 570 – 575, https://doi.org/10.1175/1520-0434(1990)005⟨0570:TCSIAA⟩2.0.CO;2, URL https://journals.ametsoc.org/view/journals/wefo/5/4/1520-0434_1990_005_0570_tcsiaa_2_0_co_2.xml.

Segall, J. H., M. M. French, D. M. Kingfield, S. D. Loeffler, and M. R. Kumjian, 2022: Storm-scale polarimetric radar signatures associated with tornado dissipation in supercells. *Wea. Forecasting*, **37 (1)**, 3 – 21, https://doi.org/10.1175/WAF-D-21-0067.1, URL https://journals.ametsoc.org/view/journals/wefo/37/1/WAF-D-21-0067.1.xml.

Simmons, K. M., and D. Sutter, 2005: WSR-88D radar, tornado warnings, and tornado casualties. *Wea. Forecasting*, **20 (3)**, 301 – 310, https://doi.org/10.1175/WAF857.1, URL https://journals.ametsoc.org/view/journals/wefo/20/3/waf857_1.xml.

Smith, R. L., and D. W. Holmes, 1961: Use of Doppler radar in meteorological observations. *Mon. Wea. Rev.*, **89 (1)**, 1 – 7, https://doi.org/10.1175/1520-0493(1961)089⟨0001:UODRIM⟩2.0.CO;2, URL https://journals.ametsoc.org/view/journals/mwre/89/1/1520-0493_1961_089_0001_uodrim_2_0_co_2.xml.

Smith, T. M., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) severe weather and aviation products: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97 (9)**, 1617 – 1630, https://doi.org/10.1175/BAMS-D-14-00173.1, URL https://journals.ametsoc.org/view/journals/bams/97/9/bams-d-14-00173.1.xml.

Snyder, J. C., and A. V. Ryzhkov, 2015: Automated detection of polarimetric tornadic debris signatures using a hydrometeor classification algorithm. *J. Appl. Meteor. Climatol.*, **54 (9)**, 1861 – 1870, https://doi.org/10.1175/JAMC-D-15-0138.1, URL https://journals.ametsoc.org/view/journals/apme/54/9/jamc-d-15-0138.1.xml.

Stout, G. E., and F. A. Huff, 1953: Radar records Illinois tornadogenesis. *Bull. Amer. Meteor. Soc.*, **34 (6)**, 281–284.

Stumpf, G. J., A. Witt, E. D. Mitchell, P. L. Spencer, J. T. Johnson, M. D. Eilts, K. W. Thomas, and D. W. Burgess, 1998: The National Severe Storms Laboratory mesocyclone detection algorithm for the WSR-88D. *Wea. Forecasting*, **13 (2)**, 304 – 326, https://doi.org/10.1175/1520-0434(1998)013⟨0304:TNSSLM⟩2.0.CO;2, URL https://journals.ametsoc.org/view/journals/wefo/13/2/1520-0434_1998_013_0304_tnsslm_2_0_co_2.xml.

Tarjan, R., 1972: Depth-first search and linear graph algorithms. *SIAM journal on computing*, **1 (2)**, 146–160.

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, Vol. 5B.10, URL https://ams.confex.com/ams/pdfpapers/116240.pdf.

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0123.1.

Trapp, R. J., D. M. Wheatley, N. T. Atkins, R. W. Przybylinski, and R. Wolf, 2006: Buyer beware: Some words of caution on the use of severe wind reports in postevent assessment and research. *Wea. Forecasting*, **21 (3)**, 408–415.

Van Den Broeke, M. S., 2017: Potential tornado warning improvement resulting from utilization of the TDS in the warning decision process. *Journal of Operational Meteorology*, **5 (10)**.

Van Den Broeke, M. S., 2020: A preliminary polarimetric radar comparison of pre-tornadic and nontornadic supercell storms. *Mon. Wea. Rev.*, **148 (4)**, 1567 – 1584, https://doi.org/10.1175/MWR-D-19-0296.1, URL https://journals.ametsoc.org/view/journals/mwre/148/4/mwr-d-19-0296.1.xml.

Van Den Broeke, M. S., and S. T. Jauernic, 2014: Spatial and temporal characteristics of polarimetric tornadic debris signatures. *J. Appl. Meteor. Climatol.*, **53 (10)**, 2217 – 2231, https://doi.org/10.1175/JAMC-D-14-0094.1, URL https://journals.ametsoc.org/view/journals/apme/53/10/jamc-d-14-0094.1.xml.

Virtanen, P., and Coauthors, 2020: SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, **17**, 261–272, https://doi.org/10.1038/s41592-019-0686-2.

Warning Decision Training Division, 2022a: Radar & Applications Course - Convective storm structure and evolution - Analyzing tornadic scale signatures. URL https://training.weather.gov/wdtd/courses/rac/severe/tornadic-signatures/presentation_html5.html, accessed 10 Oct 2022.

Warning Decision Training Division, 2022b: Radar & Applications Course - Warning fundamentals - Warning content: Impact-based warnings. URL https://training.weather.gov/wdtd/courses/rac/warnings/warn-content/presentation_html5.html, accessed 10 Oct 2022.

Whiton, R. C., P. L. Smith, S. G. Bigler, K. E. Wilk, and A. C. Harbuck, 1998a: History of operational use of weather radar by U.S. Weather Services. part I: The pre-NEXRAD era. *Wea. Forecasting*, **13 (2)**, 219 – 243, https://doi.org/10.1175/1520-0434(1998)013⟨0219:HOOUOW⟩2.0.CO;2, URL https://journals.ametsoc.org/view/journals/wefo/13/2/1520-0434_1998_013_0219_hoouow_2_0_co_2.xml.

Whiton, R. C., P. L. Smith, S. G. Bigler, K. E. Wilk, and A. C. Harbuck, 1998b: History of operational use of weather radar by U.S. Weather Services. part II: Development of operational Doppler weather radars. *Wea. Forecasting*, **13 (2)**, 244 – 252, https://doi.org/10.1175/1520-0434(1998)013⟨0244:HOOUOW⟩2.0.CO;2, URL https://journals.ametsoc.org/view/journals/wefo/13/2/1520-0434_1998_013_0244_hoouow_2_0_co_2.xml.

Wood, V. T., and R. A. Brown, 1997: Effects of radar sampling on single-Doppler velocity signatures of mesocyclones and tornadoes. *Wea. Forecasting*, **12 (4)**, 928 – 938, https://doi.org/10.1175/1520-0434(1997)012⟨0928:EORSOS⟩2.0.CO;2, URL https://journals.ametsoc.org/view/journals/wefo/12/4/1520-0434_1997_012_0928_eorsos_2_0_co_2.xml.

Wood, V. T., R. A. Brown, and D. Sirmans, 2001: Technique for improving detection of WSR-88D mesocyclone signatures by increasing angular sampling. *Wea. Forecasting*, **16 (1)**, 177 – 184, https://doi.org/10.1175/1520-0434(2001)016⟨0177:TFIDOW⟩2.0.CO;2, URL https://journals.ametsoc.org/view/journals/wefo/16/1/1520-0434_2001_016_0177_tfidow_2_0_co_2.xml.

Zittel, W. D., 2019: Theory and concept of operations for multi-PRF dealiasing algorithm's VCP 112. Tech. rep., NWS Radar Operations Center Applications Branch, 13 pp. URL https://www.roc.noaa.gov/WSR88D/PublicDocs/NewTechnology/Theory_ConOps_VCP112.pdf.

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0123.1.