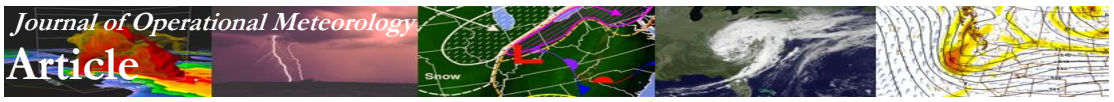
Gibbs, J. G., 2016: A skill assessment of techniques for real-time diagnosis and short-term prediction of tornado intensity using the WSR-88D. *J. Operational Meteor.*, **4** (13), 170–181, doi: http://dx.doi.org/10.15191/nwajom.2016.0413.



A Skill Assessment of Techniques for Real-Time Diagnosis and Short-Term Prediction of Tornado Intensity Using the WSR-88D

JUSTIN G. GIBBS
NOAA/National Weather Service, Warning Decision Training Division, Norman, Oklahoma

(Manuscript received 20 July 2016; review completed 1 September 2016)

ABSTRACT

Recent advancements in the science of tornado detection have allowed the National Weather Service (NWS) Warning Decision Training Division to incorporate real-time tornado intensity estimation into guidance available to NWS forecasters. This guidance focuses specifically on differentiating between strong/violent (EF2+) and weak (EF0–1) tornadoes. This study evaluates the skill of a portion of that guidance, specifically the quantification of the relationship between rotational velocity signatures and the height of tornadic debris signatures (TDSs) with observed EF-scale tornado damage.

The guidance is found to be sufficiently skillful at diagnosing tornado intensity. Perhaps most usefully, when attempting to differentiate between weak and strong/violent tornadoes in real-time, skill scores peak at the threshold of 20.57 m s^{-1} (40 kt) of rotational velocity when the velocity couplet is combined with a TDS. Skill sufficient for operational decision making also is evaluated and found in other permutations of rotational velocity, with and without a TDS, and the guidance regarding the height of the TDS. Beyond real-time diagnosis, several subjectively analyzed radar parameters show skill within the dataset at differentiating between strong/violent and weak tornadoes with lead times of 2–3 volume scans.

1. Introduction

The United States averages approximately 1200 tornadoes per year[1]. Strong to violent tornadoes (EF2+ on the Enhanced Fujita scale) make up a small portion of the overall number of tornadoes, but produce a majority of tornado injuries and fatalities (Simmons and Sutter 2011). Tornado detection and warning by the National Weather Service (NWS) is an integral and high-profile part of its overall mission (NWS 2014).

The NWS service assessment of the deadly Joplin, Missouri, tornado of 22 May 2011 (NWS 2011) identified the need for the organization to focus on discriminating between weak and strong tornado events, and to use more vivid, specific, and impact-based wording in its tornado warning products. The impact-based warning (IBW) program emerged as a result (www.weather.gov/impacts/). IBWs attempt to differentiate between relatively weak (EF0–1) and strong (EF2+) tornadoes.

Impact-based tornado warnings are separated into three tiers. The "base" tornado warning describes the detected threat and general protective action recommendations for tornadoes. Base tornado warnings convey a serious danger to life and property, and are designed for use in the current system for all tornado events. The "considerable" tornado tag invokes much stronger language, with more certain, specific, and vivid impact statements. They are designed to be used in EF2+ intensity tornadoes when such intensity can be diagnosed in real-time. The "catastrophic" tag is used when forecast offices make the decision to declare a "tornado emergency." The wording of the warning is very similar to the considerable tag but comes with specific mention of population footprints in the immediate path of the tornado as part of the emergency declaration statement. Warnings with the

[1] Refer to the National Centers for Environmental Information (NCEI) United States tornado climatology at www.ncdc.noaa.gov/climate-information/extreme-events/us-tornado-climatology.

Corresponding author address: Justin Gibbs, National Weather Service, 120 David. L Boren Blvd., Suite 2640, Norman, OK 73072
E-mail: justin.gibbs@noaa.gov

catastrophic tag are designed to be used when a tornado is confirmed, believed to be EF2+ intensity, and headed directly for a population footprint. Warning tags can be modified following the initial issuance to upgrade or downgrade the language used [Warning Decision Training Division (WDTD) 2016a].

Ripberger et al. (2015) examined the results of the consequence-based messaging used in IBW, and found that it would tend to make individuals in the path of a tornado more likely to take the proper protective action. The likelihood of proper protective action reached a point of diminishing returns when the strongest language was used, leading to a small decrease in protective action rate. The NWS adjusted its IBW messaging based on this study to use similar phrasing for considerable and catastrophic tag events (WDTD 2016a). Casteel (2016) found similar evidence supporting the use of enhanced language for significant tornadoes. Increased specificity also was shown to be linked to increased rates of protective action by Naegle and Trainor (2012).

A long history of physical science advancements made the task of real-time tornado intensity estimation possible. Brown et al. (1978) discovered the ability to use Doppler radar to identify tornadic vortex signatures (TVSs) in velocity data. Lemon and Doswell (1979) further refined our understanding of severe thunderstorm evolution and the structure of a mesocyclone as it relates to supercell tornadogenesis. Techniques also were developed and refined following the development of the Weather Surveillance Radar-1988 Doppler (WSR-88D) for using the radar to issue tornado warnings (Falk 1997).

Burgess et al. (2002) examined use of the strength of Doppler rotational velocity to estimate tornado intensity. Brown et al. (2002) showed that reducing the effective beamwidth of the WSR-88D greatly improved signature resolution of the radar. The so-called super-resolution velocity data (0.5° azimuth × 250 m range gates) produced by this change resulted in stronger velocity signatures of tornadoes when the tornado's core diameter was larger than the radar's effective beamwidth. This allowed for much improved resolution of the tornadic circulation. This altered the definition of rotational signatures detected by the WSR-88D. The TVS is currently defined as a signature where the radar effective beamwidth is less than the diameter of the tornado, whereas the tornado signature (TS) is greater than that of the effective beamwidth and the radar samples the tornado in multiple range gates (WDTD 2016b).

In an attempt specifically to address needs related to IBW implementation, Kingfield and LaDue (2015) quantified the skill that automated low-level radar rotational velocity [$V_{rot} = (V_{max} - V_{min})/2$] calculations have in discriminating between weak and strong tornadoes. The authors found that skill scores for V_{rot} peaked between 18 and 25.72 m s^{-1} (35 and 50 kt) in quality controlled data, but noted that considerable overlap still existed between strong and weak tornado events. However, overall skill scores were as high as the skill in differentiating between tornadic and nontornadic signatures.

Smith et al. (2015) produced probabilities of the report of a given EF-scale damage based on (i) observed V_{rot} values and (ii) environmental conditions evaluated with the significant tornado parameter (Thompson et al. 2012). The study showed skill at discriminating tornado intensity. Stronger velocity signatures were associated with stronger tornadoes, within a range of approximately 120–129.6 km (65–70 nm) from a WSR-88D radar site, which agrees with principles discussed by Dowell et al. (2005).

Polarimetric radars were found by Ryzhkov et al. (2005) to be capable of detecting debris lofted by tornadoes in near real-time. These tornadic debris signatures (TDS) were characterized by a relative drop in correlation coefficient (ρ_{hv}) and differential reflectivity (Z_{DR}) within a Doppler velocity couplet. Bodine et al. (2013) first explored TDS characteristics as a tool for real-time damage estimation. They found that a correlation did exist between TDS characteristics—specifically height and volume—and EF-scale damage. Entremont and Lamb (2015) and Van Den Broeke and Jauernic (2014) quantified this relationship, specifically the height of the TDS versus EF-scale. Stronger tornadoes showed a tendency to produce TDSs that extended higher into the storm. This proved simpler to classify in real-time than an estimate of the total volume of the TDS.

In 2012, WDTD updated its tornado warning guidance in response to the IBW program and the associated emerging physical science. This guidance has been updated frequently in response to emerging science, with the most recent version published in 2016 (WDTD 2016e). WDTD collaborated with NWS meteorologists to create a one-page summary of this guidance in 2014. This document was revised and included in WDTDs IBW guidance in 2016 (Fig. 1). This decision aid presented values of low-level V_{rot} and TDS height coupled with potential tornado intensity. The chart is intended to be used when sufficient

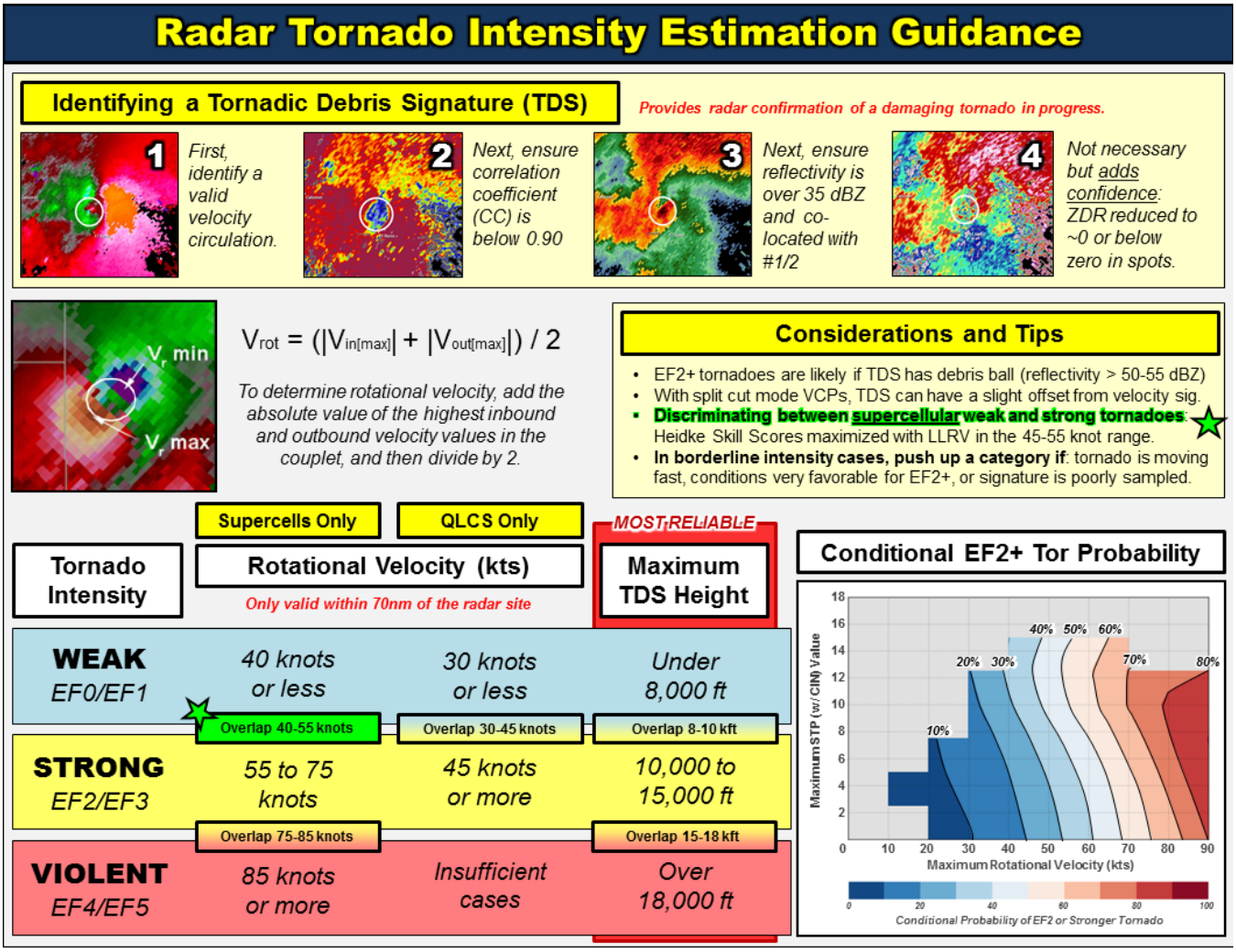


Figure 1. Radar tornado intensity estimation guidance currently in use by the NWS WDTD. The lower left quarter of the graphic contains rotational velocity and TDS height guidance for tornado intensity estimation. The top of the graphic contains the process for identifying a TDS. Below that, additional tornado intensity estimation tips, conditional environmental probabilities, and the process for calculating V_{rot} are included. *Click image for an external version.*

quality velocity data are present, preferably from the lowest available elevation angle. For example, V_{rot} values >28.29 m s^{-1} (55 kt) are likely to be associated with a strong tornado. The chart also reflects the frequent overlapping of values between ranges (Kingfield and LaDue 2015; Smith et al. 2015). For instance, V_{rot} values of 20.57–28.29 m s^{-1} (40–55 kt) lie within the overlap between weak and strong tornadoes. TDSs higher than 3048 m (10 000 ft) AGL were identified as likely to be associated with a strong tornado, while values of 2438–3048 m (8000–10 000 ft) AGL were considered to lie within the overlap region between strong and weak tornadoes (Van Den Broeke and Jauernic 2014; Entremont and Lamb 2015).

This study will evaluate the skill of this guidance. The focus will be on operational diagnosis of when a tornado is producing—or imminently will produce—EF2+ damage. Techniques also will be identified that show skill discriminating between weak and strong/violent tornado events 2–3 volumes scans prior to the onset of damage.

2. Data and methods

Tornado events were identified using the NCEI Storm Events Database (www.ncdc.noaa.gov/stormevents/) and tornado damage surveys. Events were matched to damage points listed in the NWS Damage Assessment Toolkit (DAT, apps.dat.noaa.gov/StormDamage/DamageViewer/). If no DAT points were

available, specific geographic references in the damage survey or NCEI's *Storm Data* entry were accepted. WSR-88D data then were examined to identify the storm responsible for producing the surveyed tornado. Once the storm was identified, the time was adjusted to match the rotational velocity couplet to a listed damage point. Couplets were identified using similar methods as in Smith et al. (2015) with combinations of velocity maxima exhibiting cyclonic or anticyclonic azimuthal shear within 9.26 km (5 n mi) and ≤45° angle from one another. The rotational velocity was calculated as $V_{rot} = (V_{max} - V_{min})/2$. The Gibson Ridge software (GR2 Analyst version 2.13) was used for interrogating WSR-88D data. Data were evaluated on the 0.5° elevation angle, except for three points where velocity aliasing—at close radar range—contaminated the 0.5° data, in which case 0.9° velocities were used. Velocity data then were subjectively evaluated for validity. Velocity data subjected to aliasing errors, range folding, apparent side-lobe contamination, or areas of reflectivity <20 dBZ were rejected (Piltz and Burgess 2009; WDTD 2016d).

For the purpose of this study, which is to evaluate the skill in diagnosing tornado intensity in real-time, individual velocity couplets were matched to individual damage points in DAT or listed within the storm survey. The elevation scan in which the couplet was closest, but prior to the damage listed in DAT, was used, and the V_{rot} and EF-scale damage at each point documented. This was done to attempt to control for tornadoes that moved over areas that were rated EF0 owing to a lack of damage indicators. This method ensured that the tornado associated with the velocity couplet struck an EF-scale damage indicator (Texas Tech University 2006) and that the NWS evaluated the damage. Couplets were not used that (i) had no corresponding damage points, (ii) were damage points that occurred where the 0.5° beam was >2438.4 km (8000 ft) above radar height, or (iii) were approximately 138.9 km (75 n mi) from a radar site.

Couplets then were evaluated for the presence of a TDS. If observed, the height that the TDS signature extended vertically into the TS couplet was documented and matched to the closest DAT point. Signatures were identified using WDTDs TDS criteria of a collocation of ρ_{hv} <0.90 within an area of cyclonic and convergent velocity data, and reflectivity values >35 dBZ. Low Z_{DR} values (close to zero) were used to add confidence to potentially ambiguous signatures (WDTD 2016e). Non-uniform beam filling, areas of large hail, and three-body scatter spikes occasionally made identification difficult. If there was doubt of whether the circulation was still debris, vertical continuity of the signature was evaluated by examining the elevation angle above the data in question. If a quasi-symmetric TDS signature persisted with height, it was assumed that a TDS was present at the lower elevation angle. If there was no vertical continuity, the event was considered to either not have a TDS (if it was on the 0.5° elevation angle) or the TDS was considered to have reached maximum height at the elevation angle below the last point in which it was observed.

The coupling of surveyed damage and a Doppler velocity couplet were documented at 432 points. For 84 (75) different tornado events, a TDS was not (was) observed for 218 (214) points. Strong/violent damage produced 212 points, and weak damage produced 220 points. Table 1 shows the distribution by EF-scale. Even though both weak and strong/violent tornadoes and TDS/no TDS pairs were sampled evenly, the TDS cases clearly represented a larger percentage of EF2+ cases. Similar results were found by Van Den Broeke (2015). Sixteen of the damage points were associated with convection produced by a squall-line/quasi-linear convective system (QLCS) per the definition by Trapp et al. (2005), while the rest were various forms of supercell or hybrid convection containing a persistent independent mesocyclone (Doswell and Burgess 1993; Smith et al. 2012).

Table 1. Number of cases included in the study by EF-scale rating sorted by whether they were or were not associated with a TDS.

EF-scale Rating	No TDS Signatures	TDS Signatures
4+	21	10
3	10	61
2	49	61
1	127	58
0	32	3
All	239	193

From there, data were evaluated in two ways. First, the V_{rot} and TDS heights were binned based on where they fit in the WDTD radar tornado intensity estimation guidance (Fig. 1). Table 2 describes the scoring bins for assigned V_{rot} and TDS values.

The observed damage points were assigned 1 for a weak tornado (EF0–1), 3 for a strong tornado (EF2–3), and 5 for a violent tornado (EF4–5). This scoring paradigm shows larger error when the observed damage was off by an entire category (e.g., strong versus weak), while partially crediting for when the signal falls in the adjacent "overlap" bin. QLCS tornadoes were scored using a slightly lower V_{rot} scale (Fig. 1). It

Table 2. Values of V_{rot} and TDS height and the associated scoring bins used for supercells within the study to be matched with DAT damage pairs. EF0–1 tornadoes were scored 1, EF2–3 tornadoes were scored 3, and EF4–5 tornadoes were scored 5. QLCS tornadoes were scored similarly but using lower V_{rot} values (Fig. 1).

V_{rot} in m s^{-1} (kt)	TDS Height in m (ft)	Scoring Bin
<20.57 (<40)	<2438 (<8000)	1
20.57–28.29 (40–55)	2438–3048 (8000–10 000)	2
28.30–38.58 (55–75)	3048–4572 (10 000–15 000)	3
38.59–43.72 (75–85)	4572–5486 (15 000–18 000)	4
≥43.72 (≥85)	≥5486 (≥18 000)	5

then was documented how well the observed damage matched predicted intensity. This was done by noting what values, for example, fell exactly within the correct bin, or one bin too strong, two too weak, etc. Situations where the observed bin matched the assigned bin were marked as 0. But, for example, an EF2 damage point with velocities that fell in the 20.57–28.29 m s^{-1} (40–55 kt) "overlap" bin would have been scored one "too weak," or –1.

Next, the skill scores were evaluated from the perspective of using the data to make operational IBW decisions in real-time. Probability of detection (POD), false alarm ratio (FAR), critical success index (CSI, Schaefer 1990), and Heidke skill score (HSS, Heidke 1926) were calculated for diagnosing an EF2+ tornado event based on a number of thresholds of rotational velocity strength and TDS height.

Going beyond real-time diagnosis, an attempt also was made to ascertain the predictability of EF2+ tornado events based on WSR-88D data. To accomplish this, 52 EF2+ tornado events were selected from events that occurred from 2013 to 2016. For volume scans collocated with the first EF2+ damage the following were documented:

- near gate-to-gate velocity, classified in this case as separated by no more than one range gate,
- the strongest V_{rot} below 1524 m (5000 ft), when available, and the distance between V_{max} and V_{min},
- the strongest V_{rot} anywhere in the storm, the distance between V_{max} and V_{min}, and the height of the center of that rotation, and
- the depth of a detectable mesocyclone with V_{rot} ≥15.43 m s^{-1} (30 kt).

The same data then were gathered for the five full volume scans preceding the observed damage. This provided data going back approximately 25 min before the onset of EF2 damage in volume coverage pattern (VCP) 212 or 12 (WDTD 2016c). If data were unavailable [such as velocity information below 1524 m (5000 ft) above radar height], or data quality issues made the data unreliable, they were excluded from the dataset. Because the desire was to obtain full volumetric information on a storm, 0.5° supplemental adaptive intra-volume low-level scans were not included in this portion of the study. The same methods were applied to 52 tornadoes that were surveyed to a peak intensity of EF1. Fifty of these cases exhibited V_{rot} ≥15.43 m s^{-1} (35 kt), which is approaching the lower end of the "overlap" between weak and strong tornadoes in the evaluated guidance. Approximately 65 of the 104 tornadoes used in this dataset were included in the previous dataset, but to keep a balance between the number of weak and strong/violent tornadoes, and TDS/no-TDS damage points, some of these events were not included in the previous dataset. The starting point for the EF1 tornado data was the volume scan matched to the onset of EF1 damage.

Among the EF2+ dataset, 46 of the tornadoes were produced by supercells while six were produced by QLCS mesovortices. Thirty-four of the weak tornado events were produced by supercells, six were somewhat ambiguous between embedded supercell and QLCS mesovortex status, and ten were more clearly QLCS in nature. All but two of the tornadoes that were produced by QLCS-like storms had V_{rot} >18 m s^{-1} (35 kt)—one from each dataset—which fit well with the existing supercell dataset.

Average values were calculated for the weak and strong/violent tornado datasets, and null hypothesis testing using Pearson's χ^2 test for independence (Rao 2002) was conducted for all variables at a number of different thresholds. Evaluation of the data and independence testing also were conducted for rate of change, trends, the overall value of the variables, and the persistence of values.

3. Analysis and discussion

a. Nowcasting tornado intensity skill

For the first evaluation, radar data were sorted into four categories: (i) velocity values that were not associated with a TDS, (ii) velocity values that were associated with a TDS, (iii) all velocity values, and (iv) the height of the TDS. Table 3 shows the results. The data were sorted by instances that matched the guidance based on their velocity/TDS height, or were weaker or stronger than these instances. For velocity

Table 3. The number of cases included in this study where the observed damage matched, or was too strong or too weak, compared to the listed parameters in WDTD's radar estimation of tornado intensity guidance.

	V_{rot} (No TDS)	TDS Height	V_{rot} (With TDS)	All V_{rot}
4 Too Strong	0	0	0	0
3 Too Strong	0	0	0	0
2 Too Strong	14	28	15	29
1 Too Strong	45	20	39	84
Guidance Matched Observations	111	82	96	207
1 Too Weak	26	28	42	68
2 Too Weak	19	54	16	35
3 Too Weak	3	2	4	7
4 Too Weak	0	0	2	0
Total Cases	218	214	214	432

signatures without a TDS, velocity values fell in the correct bin 111 of 218 times (50.9%) and came within one category 182 of 218 times (83.4%). For signatures associated with a TDS, velocity values were in the correct bin 96 of 214 times (44.8%) and within one bin 177 of 214 times (82.7%). Using the TDS height produced a velocity estimate that was in the correct bin 82 of 214 times (38.3%) and within one bin 130 of 214 times (60.7%). The total velocity dataset—with both TDS and non-TDS signatures—produced a correct estimate 207 of 432 times (47.9%) and came within one bin 359 of 432 times (83.1%).

To test whether a combination of the V_{rot} and TDS height provided increased skill over the parameters by themselves, the maximum and minimum of the two scores were collected for each point associated with a TDS. The maximum of the two values produced a correct result 93 of 214 times (43.4%) and was within one bin 162 of 214 times (75.7%). The minimum of the two values produced a correct result 84 of 214 times (39.2%) and was within one bin 145 of 214 times (67.7%).

In an attempt to capture real-time predictability, the TDS was measured at the time of damage. This is likely responsible for the lower number of matches associated with the TDS. This presumably is associated with the known lag time from the time a tornado begins producing damage—even at high intensity—and the time that the TDS reaches its full height, thus causing underestimates. Additionally, the vortex and associated ground damage may have weakened while debris had yet to fall from the storm (Van Den Broeke 2015), causing overestimates. These factors should be taken into consideration when using the results of this study operationally.

The data then were evaluated for skill in use as a decision tool for when to invoke considerable or catastrophic wording for use in NWS impact-based tornado warnings. Table 4 shows POD, FAR, CSI, and HSS for multiple iterations of V_{rot} sorted by all cases as well as by cases that did, and did not, have a collocated TDS. A "hit" was considered as EF2+ damage occurring at the time the data were gathered or before the next 0.5° radar slice occurred. The most skillful iteration appears to be $V_{rot} \geq 20.57$ m s^{-1} (40 kt) combined with a TDS. This iteration had a POD of 0.915 for EF2+ damage, a FAR of 0.190, a CSI of 0.752, and an HSS of 0.414. Higher V_{rot} values produced slightly lower FARs, but also resulted in an appreciable drop in POD for the dataset sampled.

With all events combined, the strongest HSS scores were produced by a V_{rot} of ≥ 20.57 m s^{-1} (40 kt), with 25.72 m s^{-1} (50 kt) very close behind. However, all three iterations of V_{rot} had lower skill scores and a relatively high FAR for EF2+ damage when no TDS was present. A significant contributing factor in this study may have been the short amount of time/distance any tornado/couplet would have had to hit a detectable EF2+ damage indicator before the next volume scan. This notion is supported by the fact that when debris is confirmed with a tornado via a TDS, lower rotational velocities appear to be associated consistently with EF2 damage within this dataset.

The height of the TDS alone also was evaluated (Table 5) and showed generally poorer scores, but a relatively low FAR. Heights other than those shown in Table 5 were evaluated, but did not produce significantly better results. Other more robust TDS studies still support it as a useful parameter for operational tornado intensity estimation, particularly in generally

Table 4. POD, FAR, CSI, and HSS metrics in discriminating between weak and strong/violent tornadoes for varying permutations of V_{rot} in m s^{-1} (kts) with and without a TDS present and with a combination of both TDS and non-TDS cases.

V_{rot} in m s^{-1} (kt)	20.57 (40)	25.72 (50)	28.29 (55)	20.57 (40)	25.72 (50)	28.29 (55)	20.57 (40)	25.72 (50)	28.29 (55)
TDS?	N	N	N	Y	Y	Y	All Cases	All Cases	All Cases
POD	0.711	0.470	0.322	0.915	0.750	0.633	0.858	0.679	0.550
FAR	0.607	0.508	0.475	0.190	0.147	0.118	0.462	0.253	0.210
CSI	0.338	0.307	0.243	0.752	0.670	0.584	0.440	0.556	0.495
HSS	0.241	0.295	0.228	0.414	0.397	0.347	0.468	0.461	0.428

Table 5. POD, FAR, CSI, and HSS metrics for using the listed height of TDSs above radar level for discriminating between weak and strong/violent tornadoes.

TDS Height	2438 m (8000 ft)	3048 m (10000 ft)
POD	0.673	0.50
FAR	0.195	0.12
CSI	0.287	0.475
HSS	0.234	0.250

being associated with stronger tornadoes (Van Den Broeke 2015).

b. Predictability of EF2+ damage at various lead times

The second evaluation involved assessing 104 tornado events for predictability and differentiation between weak and strong/violent tornado events with lead time. Fifty-two EF2+ tornado events and 52 EF1 tornado events were evaluated through subjective analysis of WSR-88D volumetric base data moments. These moments, their average values, and the number of cases that were accepted after quality control are presented in the appendix.

Pearson's χ^2 test for independence then was conducted on each dataset at each temporal iteration. Results allowing us to reject the null hypothesis of strong/violent and weak tornado dataset independence at $p \leq 0.01$ then were evaluated for operational utility.

Peak $V_{rot} \geq 25.72$ m s^{-1} (50 kt) in any of the three prior volume scans, hereafter 50VROT15, was found to have statistically significant differentiation between weak and strong/violent tornado events. Table 6 shows a contingency table of cases, which produced a p value <0.0001, with 50VROT15 associated with a higher number of strong/violent tornado events. Strengthening rotational velocity values (Table 7) also were found not to be independent of tornado intensity within the dataset with p values <0.005.

$V_{rot} \geq 20.57$ m s^{-1} (40 kt) that persisted during the prior three volume scans, hereafter P40VROT, also was found to have statistically significant differentiation between weak and strong/violent tornado events. The contingency table for this parameter (Table 8) also shows a much higher frequency of strong/violent tornado events when P40VROT was met. There is somewhat more overlap in these events, but a p value of 0.00029 suggests the values are not independent.

Table 6. Contingency table with the number of cases that did or did not have $V_{rot} \geq 25.72$ m s^{-1} (50 kt) in the three volume scans (50VROT15) prior to the onset of damage for strong and weak tornado events. Results of Pearson's χ^2 test are $\chi^2 = 35.1194$, degrees of freedom (df) = 1, and p <0.0001

50VROT15	EF2+ Tornado Events	EF1 Tornado Events
Yes	45	17
No	5	35

Table 7. Number of strong/violent and weak tornado cases that contained strengthening V_{rot} by the time prior to the onset of the event, using 5 min as an approximation for volume scan time. The p value of Pearson's χ^2 test between the two datasets is listed in the right column.

Strengthening V_{rot}	EF2+ Tornado Events (Cases)	EF1 Tornado Events (Cases)	p value
0–5 min	37 (50)	16 (48)	<0.0001
5–10 min	35 (50)	14 (48)	<0.0001
10–15 min	31 (49)	16 (48)	<0.0044
15–20 min	34 (49)	25 (48)	0.0982
20–25 min	28 (49)	19 (48)	0.1050

Table 8. Same as Table 6 except for $V_{rot} \geq 20.57$ m s^{-1} (40 kt) persisting for the three volume scans (P40VROT) prior to the onset of damage. The results of Pearson's χ^2 test are $\chi^2 = 17.4582$, df = 1, and p <0.0001

P40VROT	EF2+ Tornado Events	EF1 Tornado Events
Yes	37	17
No	13	35

A 2438-m (8000 ft) or deeper mesocylcone with V_{rot} in each elevation angle ≥ 15.43 m s^{-1} (30 kt) persisting between the last and second-to-last volume scan before EF2+ damage, hereafter PMESO30, also showed statistically significant differentiation. The time between the last and second-to-last volume scan is approximately 4–9 min in VCP 212 (WDTD 2016c). Table 9 shows this parameter's contingency table with a p value <0.0001.

Finally, near gate-to-gate $V_{rot} \geq 36.01$ m s^{-1} (70 kt), with bins separated by no more than one range gate

Table 9. Same as Table 6 except for a mesocyclone with V_{rot} ≥15.4 m s^{-1} (30 kt) and depth ≥2438 m (≥8000 ft) persisting for 2–3 volume scans (PMESO30). The results of Pearson's χ^2 test are χ^2 = 34.09, df = 1, and p <0.0001.

PMESO30	EF2+ Tornado Events	EF1 Tornado Events
Yes	34	6
No	16	46

Table 10. Same as Table 6 except for a mesocyclone with near gate-to-gate V_{rot} ≥36 m s^{-1} (70 kt) in any of the three preceding volume scans (70LLDV) prior to the onset of strong and weak tornado events. The results of Pearson's χ^2 test are χ^2 = 21.33, df = 1, and p <0.0001.

70LLDV	EF2+ Tornado Events	EF1 Tornado Events
Yes	29	7
No	22	45

and occurring in any of the preceding volume scans (hereafter 70LLDV), also showed statistical significance. Table 10 shows this parameter's contingency table with a p value <0.0001.

Combining any three of the previous four parameters, hereafter 3OF4, additionally showed statistically significant differentiation. Table 11 shows the contingency table for the combination of parameters with a p value of <0.0001.

Azimuthal shear—measured in this case as the velocity difference (ΔV) divided by the distance between the two pixels in which the shear was observed—did not show statistically significant differentiation at any time, value, or rate of change. The same holds true for the height that the peak V_{rot} was observed in any given volume scan, or its rate of change from one volume scan to the next. The distance between V_{max} and V_{min} was not found to have statistical significance at any time or value, but the rate of change of the width did appear to have some significance between the first and second volume scans preceding EF2+ damage (Table 12). Widths that decreased tended to be more frequently associated with strong/violent tornado damage within the next 5 min.

Skill scores were calculated for 50VROT15, P40VROT, PMESO30, 70LLDV, and 3OF4 (Table 13). The 50VROT15 parameter showed the strongest skill scores with a POD of 0.900, FAR of 0.274, CSI of 0.671, and HSS of 0.570. PMESO30 also showed strong HSS and CSI metrics, with a lower POD and lower FAR. P40VROT showed reasonable skill with a relatively high POD and modestly low FAR, as did 3OF4. Given the skill scores from this dataset, there appears to be operational utility in using these values for making considerable tag decisions in the 5–15 min prior to damage onset range.

Table 11. Same as Table 6 except for cases that contained at least 3 of the 4 conditions outlined in Tables 6, 8, 9, and 10 (3OF4) for strong and weak tornado events. The results of Pearson's χ^2 test are χ^2 = 17.403, df = 1, and p <0.0001.

3OF4	EF2+ Tornado Events	EF1 Tornado Events
Yes	32	12
No	18	40

Table 12. The average change in width (km with n mi in parentheses) between maximum V_{rot} pixels and the number of cases where the width of the mesocyclone increased for strong/violent and weak tornado events. Columns indicate the approximate time prior to the onset of EF2+ damage for EF2+ cases and the onset of EF1 damage for EF1 cases based on an approximation of the full volume scan time in VCP 12 or 212. The asterisk (*) denotes that Pearson's χ^2 test produced a p value of <0.01 between the two datasets.

Time Prior to the Onset of Strong/Violent or Weak Damage	Average Change of Mesocyclone Width in km (n mi)		# of EF2+ Cases Where Mesocyclone Width Increased	# of EF1 Cases Where Mesocyclone Width Increased
	EF2+ Cases	EF1 Cases		
0–5 min	–1.03 (–0.56)	0.62 (0.34)	18 (52)*	35 (52)*
5–10 min	–0.85 (–0.46)	0.38 (0.21)	22 (50)*	32 (52)*
10–15 min	–0.29 (–0.16)	0.05 (0.03)	21 (49)	24 (50)
15–20 min	–0.07 (–0.04)	0.20 (0.11)	24 (50)	26 (47)
20–25 min	0.61 (0.33)	–0.35 (–0.19)	29 (49)	18 (45)

Table 13. POD, FAR, CSI, and HSS metrics for the listed parameters out of the 104 cases in the study.

	50VROT15	P40VROT	PMESO30	70LLDV	3OF4
POD	0.900	0.740	0.680	0.568	0.640
FAR	0.274	0.314	0.150	0.194	0.272
CSI	0.671	0.552	0.607	0.500	0.516
HSS	0.570	0.412	0.566	0.435	0.410

4. Conclusions

This study is intended to evaluate WDTD guidance for making impact-based tornado warning decisions. The IBW system differentiates tornado events by weak and strong/violent events. This guidance is built off of decades of science, but focuses on recent work quantifying the relationship between tornado damage intensity and subjectively analyzed WSR-88D volumetric products—specifically V_{rot} (Kingfield and LaDue 2015; Smith et al 2015) and TDS height (Entremont and Lamb 2015; Van Den Broeke and

Jauernic 2014). This study, evaluating 432 surveyed damage points from 84 different tornadoes, finds that the guidance is skillful. Within 70 nm of the radar site, real-time weak and strong/violent tornado differentiation skill peaks at the threshold of $V_{rot} \geq 20.57$ m s^{-1} (40 kt) when the couplet is associated with a TDS. This V_{rot}–TDS coupling produces a POD of 0.915 and FAR of 0.190 for strong/violent tornado events. The height of the TDS was found to be somewhat less skillful by itself, but known limitations of how a TDS develops suggest that the very short temporal resolution produced by the methods of this study may have reduced its overall skill as compared with previous studies.

To examine for potential decision making skill with lead-time, concepts employed in real-time tornado intensity estimation also were evaluated in datasets prior to the onset of tornado damage. Volumetric WSR-88D data were evaluated in the five preceding volume scans for 52 strong/violent and 52 weak tornado events. This identified three other subjectively analyzed WSR-88D parameters that showed skill at differentiating between weak and strong/violent tornadoes before the onset of EF2+ damage. 50VROT15, P40VROT, PMESO30, 70LLDV, and 3OF4 were shown to have skill at differentiating between weak and strong/violent tornadoes within the dataset. This provides a glimpse at techniques that may allow forecasters to issue considerable tag warnings and statements prior to the onset of lowest-elevation angle signatures associated with EF2+ damage, especially in concert with existing warning decision methods. Future studies may be able to further sharpen these relationships.

The lead-time-related techniques differed from real-time estimation in that the full volumetric dataset was used in each volume scan rather than the lowest available elevation angle. V_{rot} detected in the low levels, below 1524 m (5000 ft), which was shown to be quite skillful at diagnosing current intensity, did not appear to have skill at differentiating future tornado damage within the dataset collected. The same was true for distance between maximum low-level velocity values below 1824 m (5000 ft) and overall azimuthal shear. In other words, whereas the 0.5° elevation angle did handle real-time tornado intensity estimation well, skillful lead-time required examining velocity data at higher levels of the storm.

Beyond the core findings, V_{rot} tended to increase in the prior three volume scans in strong/violent tornadoes as compared to weak tornadoes, and the distance between V_{max} and V_{min} tended to decrease in the preceding two volume scans. These values, however, clearly would need to be used in concert with other parameters to effectively contribute to an operational warning decision.

This study only looked at situations where a tornado was known to have occurred and was surveyed in some form by the NWS. This was done primarily as an effort to control for "open field" tornadoes that may go unreported or did not strike EF-scale damage indicators. Nevertheless, this study did not factor in potential false alarm couplets that have been factored into other studies and guidance such as by Smith et al. (2015) and WDTDs tornado warning guidance (2016e). This prior guidance remains valid and should continue to be factored into operational warning decision making. Presumably, skill scores would be at least somewhat lower when cases that do not produce a tornado are included.

The results of the study show that by using a combination of WSR-88D volumetric data—especially if combined with a thorough understanding of the current mesoscale environment (e.g., Grams et al. 2012; Thompson et al. 2012)—forecasters should be able to consistently identify significant tornadoes in progress within 120–130 km (65–70 nm) of a radar site. There also appears to be skill available at differentiating between strong/violent and weak tornado-producing storms with some lead time prior to the development of significant tornado damage. Primary limitations of these techniques include radar range, data sampling, and limitations of the EF-scale, such as available damage indicators negatively impacting the reliability of this and other studies.

Acknowledgments. NWS meteorologist Alex Lamers is acknowledged for the original development, design, and collaboration of revisions of the radar tornado intensity estimation guidance used in this paper (Fig. 1). The author also thanks Jim LaDue and Bobby Prentice of WDTD, as well as JOM editor Dr. Andrew Mercer and two anonymous reviewers, for their observations and suggestions that greatly improved this manuscript.

APPENDIX

Data for the Predictive Part of the Study

Tables A1–A8 present the average values of the eight subjectively analyzed WSR-88D parameters that formed the initial dataset for evaluating predictive skill in discriminating between weak and strong/violent tornadoes (section 3b). These data are included here to provide a clear vis-

ualization of the information that was gathered, the number of cases that were included, and some sense of what the data reflected prior to analysis. This database then was examined for operationally and statistically significant trends/values that are discussed in the main body of the paper.

Table A1. Average values of V_{rot} (m s^{-1} with kt in parentheses) for weak and strong/violent tornado events from the volume scan that significant damage (EF2+) began in the EF2+ cases or EF1 damage began in the EF1 cases, going back five volume scans. The number of cases with useable data is given in the column to the left of the values.

	V_{rot} in m s^{-1} (kt)			
	# Cases	EF2+ Events	# Cases	EF1 Events
Time of Damage	52	30.5 (59.2)	52	23.5 (45.8)
1 Volume Scan Prior	52	27.6 (53.6)	52	22.0 (42.8)
2 Volume Scans Prior	50	25.7 (49.9)	52	19.7 (38.4)
3 Volume Scans Prior	48	25.1 (48.7)	51	19.3 (37.6)
4 Volume Scans Prior	48	22.8 (44.3)	50	19.0 (36.9)
5 Volume Scans Prior	46	21.5 (41.8)	48	18.4 (35.7)

Table A2. Similar to Table A1 except for average values of azimuthal shear (m s^{-1}/km). The number of cases with useable data is listed in parenthesis.

	Azimuthal Shear (m s^{-1}/km)	
	EF2+ Events	EF1 Events
Time of Damage	52.26 (52)	25.43 (52)
1 Volume Scan Prior	39.28 (52)	23.76 (52)
2 Volume Scans Prior	19.82 (50)	21.28 (52)
3 Volume Scans Prior	21.04 (48)	20.86 (51)
4 Volume Scans Prior	17.05 (48)	20.45 (50)
5 Volume Scans Prior	22.95 (46)	19.81 (48)

Table A3. Same as Table A1 except for average values of the distance (km with n mi in parentheses) between V_{max} and V_{min}.

	Peak V_{rot} Width in km (n mi)			
	# Cases	EF2+ Events	# Cases	EF1 Events
Time of Damage	52	2.2 (1.2)	52	3.5 (1.9)
1 Volume Scan Prior	52	3.2 (1.8)	52	4.1 (2.2)
2 Volume Scans Prior	50	4.1 (2.2)	52	4.6 (2.5)
3 Volume Scans Prior	48	4.3 (2.3)	51	4.4 (2.4)
4 Volume Scans Prior	48	4.3 (2.3)	50	4.6 (2.5)
5 Volume Scans Prior	46	3.8 (2.1)	48	4.3 (2.3)

Table A4. Same as Table A1 except for average values of the beam height (m with ft in parentheses) at the location centered between the two highest observed values of V_{rot}.

	Peak V_{rot} Height in m (ft)			
	# Cases	EF2+ Events	# Cases	EF1 Events
Time of Damage	52	1733 (5687)	52	1737 (5700)
1 Volume Scan Prior	52	2177 (7144)	52	1923 (6310)
2 Volume Scans Prior	50	2516 (8255)	52	2133 (6999)
3 Volume Scans Prior	48	2602 (8539)	51	2210 (7252)
4 Volume Scans Prior	48	2594 (8511)	50	2275 (7466)
5 Volume Scans Prior	46	2678 (8789)	48	2873 (9429)

Table A5. Same as Table A1 except for average values of the near gate-to-gate V_{rot} (m s^{-1} with kt in parentheses).

	Near Gate-to-Gate V_{rot} in m s^{-1} (kt)			
	# Cases	EF2+ Events	# Cases	EF1 Events
Time of Damage	51	46.8 (91.5)	52	31.3 (60.9)
1 Volume Scan Prior	51	39.6 (76.9)	52	23.0 (44.7)
2 Volume Scans Prior	48	29.7 (57.8)	52	18.5 (36.0)
3 Volume Scans Prior	49	26.4 (51.5)	51	15.6 (30.4)
4 Volume Scans Prior	50	23.8 (46.3)	50	15.6 (30.4)
5 Volume Scans Prior	48	21.1 (41.2)	48	17.4 (33.9)

Table A6. Same as Table A1 except for average values of the depth (m with ft in parentheses) of a detectable mesocyclone with $V_{rot} \geq 15.4$ m s^{-1} (30 kt).

	Depth in m (ft) of Detectable Mesocyclone [$V_{rot} \geq 15.4$ m s^{-1} (30 kt)]			
	# Cases	EF2+ Events	# Cases	EF1 Events
Time of Damage	51	4353 (14 282)	52	3043 (9982)
1 Volume Scan Prior	51	4308 (14 133)	52	2943 (9965)
2 Volume Scans Prior	48	4163 (13 663)	51	2651 (8698)
3 Volume Scans Prior	48	3767 (12 358)	51	2410 (7907)
4 Volume Scans Prior	49	3463 (11 363)	50	2314 (7592)
5 Volume Scans Prior	46	3615 (11 861)	48	2342 (7687)

Table A7. Same as Table A2 except for average values of azimuthal shear (m s^{-1}/km) measured below 1524 m (5000 ft) above radar height.

	Azimuthal Shear (m s^{-1}/km) Measured Below 1524 m (5000 ft) Above Radar Height	
	EF2+ Events	**EF1 Events**
Time of Damage	49.21 (40)	26.03 (45)
1 Volume Scan Prior	27.24 (40)	14.49 (45)
2 Volume Scans Prior	25.80 (39)	14.26 (42)
3 Volume Scans Prior	21.69 (39)	23.75 (41)
4 Volume Scans Prior	16.95 (35)	11.88 (41)
5 Volume Scans Prior	17.35 (30)	12.05 (38)

Table A8. Same as Table A1 except for average values of V_{rot} (m s^{-1} with kt in parentheses) measured below 1524 m (5000 ft) above radar height.

	Below 1524 m (5000 ft) Above Radar Height V_{rot} m s^{-1} (kt)			
	# Cases	EF2+ Events	# Cases	EF1 Events
Time of Damage	40	29.8 (57.9)	45	22.3 (43.3)
1 Volume Scan Prior	40	25.9 (50.3)	45	19.7 (38.3)
2 Volume Scans Prior	37	23.0 (44.7)	42	18.4 (35.8)
3 Volume Scans Prior	40	21.6 (42.0)	41	17.3 (33.6)
4 Volume Scans Prior	37	19.5 (38)	41	16.6 (32.1)
5 Volume Scans Prior	31	16.9 (32.9)	38	15.6 (30.9)

REFERENCES

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**, 139–158, Crossref.

Brown, R. A., L. R. Lemon, and D. W. Burgess, 1978: Tornado detection by pulsed Doppler radar. *Mon. Wea. Rev.*, **106**, 29–38, Crossref.

____, 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**, 1759–1771, Crossref.

Burgess, D. W., M. A. Magsig, J. Wurman, D. C. Dowell, and Y. Richardson, 2002: Radar observations of the 3 May 1999 Oklahoma City tornado. *Wea. Forecasting*, **17**, 456–471, Crossref.

Casteel, M. A., 2016: Communicating increased risk: An empirical investigation of the National Weather Service's impact-based warnings. *Wea. Climate Soc.*, **8**, 219–232, Crossref.

Doswell, C. A., III, and D. W. Burgess, 1993: Tornadoes and tornadic storms: A review of conceptual models. *The Tornado: Its Structure, Dynamics, Prediction, and Hazards*, *Geophys. Monogr.*, No. 79, Amer. Geophys. Union, 161–172.

Dowell, D. C., C. R. Alexander, J. M. Wurman, and L. J. Wicker, 2005: Centrifuging of hydrometeors and debris in tornadoes: Radar-reflectivity patterns and wind-measurement errors. *Mon. Wea. Rev.*, **133**, 1501–1524, Crossref.

Entremont, C., and D. Lamb, cited 2015: Warning decision storm of the month: The relationship between tornadic debris signature height and tornado intensity: Operations case - April 28, 2014. [Available online at www.wdtb.noaa.gov/courses/SOTM/006-Apr15/presentation.html.]

Falk, K., 1997: Techniques for issuing severe thunderstorm and tornado warnings with the WSR-88D Doppler radar. NOAA Tech. Memo. SR-185. [Available online at www.srh.weather.gov/srh/ssd/techmemo/sr185.htm.]

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

Heidke, P., 1926: Calculation of the success and goodness of strong wind forecasts in the storm warning service. *Geogr. Ann.*, **8**, 301–349, Crossref.

Kingfield, D. M., and J. G. LaDue, 2015: The relationship between automated low-level velocity calculations from the WSR-88D and maximum tornado intensity determined from damage surveys. *Wea. Forecasting*, **30**, 1125–1139, Crossref.

Lemon, L. R., and C. A., Doswell III, 1979: Severe thunderstorm evolution and mesocyclone structure as related to tornadogenesis. *Mon. Wea. Rev.*, **107**, 1184–1197, Crossref.

Naegle, D. E., and J. E. Trainor, 2012: Geographic specificity, tornadoes, and protective action. *Wea. Climate Soc.*, **4**, 145–155, Crossref.

NWS, 2011: Joplin, Missouri, Tornado – May 22, 2011. NWS Central Region service assessment. [Available online at www.nws.noaa.gov/os/assessments/pdfs/Joplin_tornado.pdf.]

____, 2014: National Weather Service instruction 10-511. WFO severe weather products specification. [Available online at www.nws.noaa.gov/directives/sym/pd01005011curr.pdf.]

Piltz, S. F., and D. W. Burgess, 2009: The impacts of thunderstorm geometry and WSR-88D beam characteristics on diagnosing supercell tornadoes. Preprints, *34th Conf. on Radar Meteorology*,

Williamsburg, VA, Amer. Meteor. Soc., P6.18. [Available online at ams.confex.com/ams/pdfpapers/155944.pdf.]

Rao, C. R., 2002: Karl Pearson chi-square test: The dawn of statistical inference. *Goodness-of-Fit Tests and Model Validity*, C Huber-Carol, N. Balakrishnan, N. Nikulin, and M. Mesbah, Eds., Birkhäuser, 9–24.

Ripberger, J. T., C. L. Silva, H. C. Jenkins-Smith, and M. James, 2015: The influence of consequence-based messages on public responses to tornado warnings. *Bull. Amer. Meteor. Soc.*, **96**, 577–590, Crossref.

Ryzhkov, A. V., T. J. Schuur, D. W. Burgess, and D. S. Zrnic, 2005: Polarimetric tornado detection. *J. Appl. Meteor.*, **44**, 557–570, Crossref.

Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting*, **5**, 570–575, Crossref.

Simmons, K. M., and D. Sutter, 2011: *Economic and Societal Impacts of Tornadoes*. Amer. Meteor. Soc., 296 pp, Crossref.

Smith, B. T., R. L. Thompson, J. S. Grams, C. Broyles, and H. E. Brooks, 2012: Convective modes for significant severe thunderstorms in the contiguous United States. Part I: Classification and climatology. *Wea. Forecasting*, **27**, 1114–1135, Crossref.

____, ____, A. R. Dean, and P. T. Marsh, 2015: Diagnosing the conditional probability of tornado damage rating using environmental and radar attributes. *Wea. Forecasting*, **30**, 914–932, Crossref.

Texas Tech University, 2006: A recommendation for an Enhanced Fujita scale (EF-Scale). Wind Science and Engineering Center Report, Revision 2, 111 pp. [Available online at www.depts.ttu.edu/nwi/Pubs/FScale/EFScale.pdf.]

Thompson, R. L., B. T. Smith, J. S. Grams, A. R. Dean, and C. Broyles, 2012: Convective modes for significant severe thunderstorms in the contiguous United States. Part II: Supercell and QLCS tornado environments. *Wea. Forecasting*, **27**, 1136–1154, Crossref.

Trapp, R. J., S. A. Tessendorf, E. S. Godfrey, and H. E. Brooks, 2005: Tornadoes from squall lines and bow echoes. Part I: Climatological distribution. *Wea. Forecasting*, **20**, 23–34, Crossref.

Van Den Broeke, M. S., 2015: Polarimetric tornadic debris signature variability and debris fallout signatures. *J. Appl. Meteor. Climatol.*, **54**, 2389–2405, Crossref.

_____, S. T. Jauernic, 2014: Spatial and temporal characteristics of polarimetric tornadic debris signatures. *J. Appl. Meteor. Climatol.*, **53**, 2217–2231, Crossref.

WDTD, cited 2016a: Impact based warnings. Part 1: Rationale and motivation. [Available online at wdtd.noaa.gov/courses/ibw/lessons/ibw-lesson1/presentation.html.]

____, cited 2016b: Radar and applications course (RAC): Convective storm structure and evolution: Analyzing tornado scale signatures. [Available online at wdtd.noaa.gov/courses/rac/severe/tornadic-signatures/presentation.html.]

____, cited 2016c: Radar and applications course (RAC): Principles of meteorological Doppler radar: VCP selection. [Available online at www.wdtb.noaa.gov/courses/rac/principles/vcp-selection/presentation.html.]

____, cited 2016d: Radar and applications course (RAC): Velocity interpretation: Lesson: Velocity Dealiasing. [Available online at wdtb.noaa.gov/courses/rac/principles/dealiasing/presentation.html.]

____, cited 2016e: Tornado warning guidance 2016: Quick reference guide. 12 pp. [Available online at www.wdtb.noaa.gov/courses/woc/documentation/severe/twg16-reference-sheets.pdf.]