8.3 SIMULATED NWS TORNADO WARNING DECISIONS USING RAPID-SCAN RADAR DATA

*Pamela Heinselman, Daphne LaDue, Darrel M. Kingfield, Robert Hoffman, and Brent W. MacAloney II

1. INTRODUCTION

In Heinselman et al. (2012), the impact of radar update time on tornado warning decisions was examined. During the experiment forecasters worked as pairs on the same tropical supercell event in simulated real time with two different volumetric updates: 43 s and 4.5 min. The results indicated the potential for rapid-update phased array radar data to result in longer tornado warning lead times: 11–18 min vs 0–6 min. The results warrant further examination using a larger sample size and diversity of cases. The objective of the 2012 Phased Array Radar Innovative Sensing Experiment (PARISE) is to further investigate the extent to which rapid, adaptively scanned radar data aids forecaster ability to make warning decisions across a more diverse set of tough, potentially tornadic events.

As in Heinselman et al. (2012), this study focuses on weak tornadic events. Though EF0- and EF1-rated tornadoes by definition are less destructive to life and property than higher-rated tornadoes, they are also the most under warned (e.g., Brotzge and Erickson 2010). According to the National Weather Service (NWS) Performance Management System, during the last 5 years or so (1 January 2008 – 31 Oct 2012), EF0 and EF1 tornadoes that occurred across the Nation comprised 93.4% (2004 of 2146) of events with 0-min lead time. Furthermore, during this period the national mean tornado lead time of EF0 and EF1 events was 12.5 min, almost five minutes less than the lead time for EF2 and higher rated tornadoes (18.07 min). As a result, people in the path of EF0 and EF1 tornadoes were less likely to receive warning prior to these events.

In this paper we report results on the lead times for EF0 and EF1 tornado events resulting from forecaster use of rapid-scan phased array radar data during PARISE 2012. The paper also provides an overview of the experiment and a description of National Weather Service verification statistics computed for storm-based warnings.

2. RADAR DATA AND VISUALIZATION

2.1 Cases

The PARISE 2012 data set includes four cases sampled by the S-band phased array radar (PAR) at the National Weather Radar Testbed (Zrnić et al. 2007) (Table 1). Case longevity ranges from 18 to 52 min; tornadoes occurred during two of the four events (Table 1). The cases also met the many practical criteria for inclusion in this study, such as sufficient longevity and continuity prior to tornadogenesis, minimal velocity aliasing, and location within 120 km range of the PAR.

The scan strategy characteristics vary somewhat among events. On 11 May 2010, data collection focused on lower elevations where tornado cyclones and larger circulations associated with tornadoes are best observed. The scan strategy revisited the lowest four elevations twice between volumetric (22-elevation) scans, resulting in the following update times: volumetric 59 s and interlaced about 22 s (Table 1). Hence, the volumetric scan revisit time was about 1.8 min. On the three 2011 dates, storms were sampled with noninterlaced scan strategies with update times near 60 s (Table 1). Based on storm coverage and range from the radar, these volumetric update times were reduced during operations by running the adaptive digital signal processing algorithm for PAR timely scans (Heinselman and Torres 2011).

2.2 Radar Pre-Processing and Display

Base data display was handled using the Advanced Weather Interactive Processing System-2 (AWIPS-2). AWIPS-2 is currently replacing the AWIPS-1 architecture as the baseline forecasting platform at NWS Weather Forecast Offices (WFOs) across the country. Utilizing the AWIPS framework provided forecasters with access to PAR data within a familiar display and warning environment. This allowed for maximum focus to remain on product evaluation instead of software retraining.

For ease in data management and display, the four PAR cases were pre-processed using the Common Operations and Development Environment (CODE) Radar Product Generator software (Johnson et al. 1999). Utilizing CODE, we were able to generate AWIPS-readable reflectivity, velocity, and spectrum width products without data quality degradation.

3. EXPERIMENT DESIGN

3.1 Recruitment

Like all workforces, the NWS WFO staff includes individuals whose proficiency ranges from journeymen to expert levels. To explore how forecasters having differing levels of expertise may respond to and use rapid-scan data in their warning

* *Corresponding author address*: Pam Heinselman, NOAA NSSL, 120 David L. Boren Blvd., Norman, OK, 73072; e-mail: pam.heinselman@noaa.gov

Date	Duration (UTC)	Scan Strategy Update Time (s)	EF Rating and Duration (UTC)
11 May 2010	0035–0111	59 4 lowest elevations: 22	EF0: 0105–0109
14 April 2011	2055–2120	70	None
22 April 2011	2339–2358	54	None
22 May 2011	0050–0142	56	EF0: 0118–0120 EF0: 0129–0131 EF1: 0141–0147

Table 1. Case dates and times, radar update time(s), and tornado occurrence during the event.

process, we sought a distribution of participants that would enable us to contrast journeymen and experts. An initial proxy for identifying journeymen was to seek forecasters who had completed, within the last year, the Distance Learning Operations Course (DLOC). DLOC is the inaugural training course on the utilization of radar for forecasting and warning decision making. Expert forecasters were sought based on reputation among their superiors.

Participant recruitment focused on Central and Southern Regions of the NWS where tornadoes are climatologically most prevalent (e.g., Brooks et al. 2003). The two Science Support Division Chiefs were given an overview of the experiment and recruitment goals. They then worked with local offices to identify potential participants. The participant list they provided was used to select 12 forecasters based upon office location, availability, and NWS-assessed expertise category. Each forecaster was individually contacted and provided the opportunity to consent to participate. During the consent process, one declined. A replacement individual of the same expertise category was chosen from the list and consented to participate. In all reporting from this experiment, pseudonyms are used.

The participants ranged in experience from 1.3–19 years as qualified forecasters. The clear contrast in experience that we sought was not quite achieved: Six participants had at least 9-yrs experience, but the remaining six participants had as many as five-yrs experience and only one had taken DLOC within the last year. A slight majority were from NWS offices in the Southern Region. Participants had worked during 5–25 severe events in the previous year, but had not necessarily issued the warnings. One young forecaster (1.3-yrs experience) had only issued one tornado warning prior to this experiment.

3.2 Data Collection

Two participants travelled to Norman, Oklahoma for each of the six weeks of PARISE 2012. On the first morning of each week, they were provided an overview of the characteristics, capabilities and data collection strategies of the PAR, and an overview of the approach and findings of PARISE 2010 (Heinselman et al. 2012). The strategy for PARISE 2012 was then explained. Following the presentation, participants were given an overview of AWIPS-2, and provided about an hour on a workstation to familiarize themselves with the software. During that time, forecasters practiced loading PAR data from an archived case (02 May 2008) and drawing polygons via the AWIPS-2 Warning Generation (WarnGen) tool.

That afternoon and over the next 1.5 days, each participant individually worked four cases in simulated real time (Table 1), as if they were responsible for real-time warnings. Prior to each case, participants viewed a prerecorded weather briefing provided by Jim LaDue of the Warning Decision Training Branch. A few phone calls were carefully timed to provide spotter reports. Other aspects of operations were not simulated. While working each event, recordMyDesktop software recorded participant interactions with the AWIPS-2 software.

After each case, one researcher worked with each forecaster on the following. Each forecaster completed a Confidence Continuum (Heinselman et al. 2012). This instrument solicits scalar judgments of how the case compared to "usual" operations, with regard to both their understanding of the meteorology of the event, and their confidence in the depictions of storm evolution provided by the rapid-scan PAR data.

Next, each forecaster/researcher pair conducted the Recent Case Walkthrough, a method of cognitive task analysis (Crandall et al. 2006). The forecasters were first asked what goals they had achieved in that recent case. They then reviewed the video replay of their desktop activity and retrospected aloud about their reasoning and observations. The researcher prompted the forecaster to describe his or her actions and thought processes, and typed out a timeline of these.

After the Walkthrough, forecasters completed a National Aeronautics and Space Administration (NASA) Task Load Index (Hart 2006; NASA-TLX is available from: http://humansystems.arc.nasa.gov/groups/TLX/). This psychometric instrument solicits scalar judgments of six workload demands: (1) mental, (2) physical, (3) temporal, (4) performance, (5) effort, and (6) frustration.

Thereafter, the video was reviewed a second time. The descriptions developed in the first Walkthrough were refined, and details added as they were recalled. Reference was made to the draft timeline to support the forecaster in this process.

After completing that second Walkthrough, forecasters were asked (1) to identify key judgments during the case and the information used to make them, (2) whether information other than radar had or could have played a role in their decisions, (3) whether they had used any conceptual models and if so, to draw them, (4) where the case fell in the spectrum of their experience, and (5) what they did that was typical and atypical of their normal work processes.

Forecasters were not told what actually transpired in the cases until after all of these post-forecast tasks had been completed.

Some factors that could be anticipated to affect forecaster performance were addressed. To avoid a systematic unintended bias in performance from the order of cases, case order differed each week. Additionally, researchers switched places after completing two cases, so that each researcher and forecaster worked together.

4. STORM-BASED TORNADO VERIFICATION

4.1 Statistics

The implementation of storm-based warnings in October 2007 (Sutter and Erickson 2010) instigated reconsideration and reconstruction of tornado and warning statistics used by the NWS (B. MacAloney II, personal communication). County-based verification statistics used from 1986–2007 were based on the traditional 2x2 contingency table (Table 2). For warning verification purposes, this Table is still used to compute the false alarm ratio, defined as

$$FARatio = \frac{Z}{X + Z} \qquad (1)$$

(e.g., Wilks 2006). As noted by Barnes et al. (2009), an equivalent and perhaps clearer term is the probability of false alarm, POFA, which is used hereafter.

Traditional Contingency Table		Event Observed	
		Yes	No
Event Warned	Yes	X	Z
	No	Y	W

Table 2. Traditional 2x2 contingency table, where X is the number of verified warnings of events (hits), Y is the number of unwarned events (misses), Z is the number of unverified warnings (false alarms), and W is the number of verified null warnings.

Since the implementation of storm-based warnings, a new term, polygon probability of detection, PPOD, has been introduced, and the computation of tornado lead time, TLT, has been revised. Most importantly, these statistics are now defined with respect to the tornado path, rather than to the initial tornado location and time. As a result, their computation requires the creation of a path-relative 2x1 contingency table (Table 3). The two terms computed in this 2x1 contingency table are XP and YP, the number of verified and unverified points along the tornado path, respectively.

Path-relative Contingency Table		Observed Point Along Tornado Path
Event Pt Warned	Yes	XP
	No	YP

Table 3. Path-relative 2x1 contingency table, where XP is the number of verified points and YP is the number of unverified points along the tornado path.

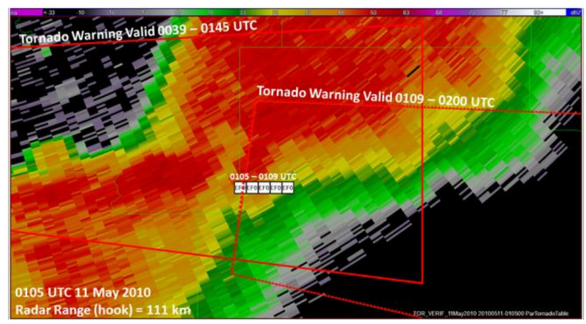

Figure 1. The 1-min-interval tornado tracks (white boxes) and warnings (red polygons) associated with an EF0-rated tornado near Millcreek, Oklahoma on 11 May 2010.

To assess XP and YP, point locations between the tornado beginning and ending points were determined at 1-min intervals by assuming that the tornado traveled in a straight line and at constant speed (Fig. 1). In this study, tornado reports were attained from the Storm Data webpage hosted by the Norman, Oklahoma Weather Forecast Office (http://www.srh.noaa.gov/oun/?n=stormdata). Although there are limitations to the use of Storm Data in tornado verification (Witt et al. 1998), the timing and location of tornadoes reported in Storm Data are reasonably consistent with circulation signatures seen in the PAR data.

Once the time-incremented tornado path is determined, every 1-min point was examined to assess whether a warning was valid at the time the event occurred (i.e., compute XP and YP). If more than one polygon was within the vicinity of the point, those polygons were combined spatially into one cohesive polygon. These points and the times associated with them were then used to compute the PPOD and TLT, respectively. In Fig. 1, all five points along the tornado path were verified by the first-issued tornado warning.

The PPOD is defined as the average percentage of event, i.e., tornado track, warned across all events:

$$PPOD = \frac{\sum_{n}^{1} \frac{XP(n)}{XP(n)+YP(n)}}{X+Y} \qquad (2)$$

The numerator of $PPOD$ is the percent of event warned, summed over the number (n) of tornado events and the denominator is the number of events observed.

The TLT is defined as the average lead time through the event's duration:

$$TLT = \frac{\sum_{p}^{1} LT(p)}{XP+YP} \qquad (3)$$

where LT(p) is the difference between the time at a given point and the time the warning was issued. If multiple warnings were valid when the event occurred, LT(p) is computed using the first warning issued (e.g., Fig. 1). At points without a valid warning, LT(p) is set to zero. For the case shown in Fig. 1, the POD is 1, FOFA is 0, and TLT is 28 min.

4.2 Tornado Lead Times

For each tornado, the TLTs resulting from participant use of rapid-scan PAR data were computed (Fig. 2). Additionally, the TLTs are compared to the two national mean lead times for tornadoes rated EF0 and EF1 (EF01LT; 12.55 min) and those rated EF2 and higher (EF2+LT;18.07 min) for the period 1 January 2008 to 31 October 2012. These lead times are appropriate for comparison as they are based on verification statistics since the implementation of storm-based warnings. Clearly, the mean lead time for the more destructive tornadoes is a more stringent measure for assessing impact of forecaster use of rapid-scan data on TLTs. A more ideographic comparison would be relative to each participant's verification statistics, or those of their office; this analysis is in progress.

For the 11 May 2010 EF0-rated tornado case (Table 1), the mean TLT is 24 min and 92% (11 of 12) of the TLTs exceed EF01LT and EF2+LT. These TLTs range from 25 to 29 min; the exception is an 8 min TLT (Fig. 2).

Three tornadoes occurred during the 22 May 2011 playback case (Table 1). Mean TLTs associated with each tornado exceed EF01LT. The mean TLT associated with the first tornado (EF0) is 14 min. The upper 50th percentile (7 of the 12) TLTs range from 18 to 24 min. The remaining 5 TLTs range from 0 to 10 min. In one case, an examination of the 0-min lead times revealed that the tornado occurred a few minutes prior to the warning. In the other case, the tornado occurred just outside of the west edge of the warning polygon.

The mean TLT for the second event (EF0) was 18 min, and individual TLTs ranged from 9 to 34 min (Fig. 2). Half of the TLTs were at least 18 min and 75% exceeded 12.5 min. This range of TLTs is associated, in part, with the issue time of the participant's first warning. The 34- and 9-min tornado lead times, for example, were both verified by 45-min tornado warnings, but the former warning (issued by participant Bob at 00:58 UTC) was issued 24 min prior to the latter warning (issued by participant Mike at 01:22 UTC). In other cases, both warning timing and longevity impact the tornado lead time. The 11.2-min warning lead time, for example, was verified by two 30-min warnings issued by Dirk: one expired and the other remained active during this second tornado event.

Like the 11 May 2010 case, the average TLT was 24 min, and 92% (11 of 12) of the third event's (EF1) TLTSs exceed 18 min, while 100% exceed 12.5 min. For this event, the use of rapid-scan data resulted in the highest TLTs, which range from 17.7–39 min.

4.3 PPOD and POFA

PPODs are computed across events, while the POFAs are computed across issued warnings (Fig. 3). In this study, the majority of the four tornado paths were verified by warnings, resulting in PPOD values 0.75 or higher (Fig. 3).

Given that half of the playback cases were null events, a POFA of 0.5 is an appropriate baseline value for assessing performance. POFAs lower than 0.5 indicate performance superior to random chance. As shown in Fig. 3, forecasters' POFA values ranged from 0 to 0.5, and 11 of 12 are below the 0.5 threshold. These findings suggest that the rapid-scan

PAR data helped some forecasters correctly discern the potential for tornadogenesis.

5. SUMMARY

Twelve NWS forecasters participated in the 2012 PARISE, which ran for six weeks during June – August 2012. The experiment's goal was to test whether rapid, adaptive sampling with the phased array radar at the National Weather Radar Testbed increases NWS forecasters' ability to effectively cope with tough tornado warning cases. Each forecaster worked four cases ranging from 18 – 52 min in length. Tornadoes were reported in 2 of the 4 cases, which allowed us to examine how rapid-scan data may help forecasters discern between tornadic and non-tornadic supercells, and the impact of the data on false alarms.

Verification statistics show that during the experiment 81% of tornado lead times exceeded the 12.5-min national mean lead time for EF0 and EF1 tornadoes that occurred during 1 January 2008 through 31 October 2012. 69% of lead times exceeded the 18-min national mean lead time for tornadoes rated EF2 and higher. The mean and median lead time across forecasters was 21 min. Polygon Probability of Detection values, defined as the average percent of tornado paths warned, were all 75% or higher. All False Alarm Ratios were 0.5 or lower. These quantitative results indicate that use of rapid-scan PAR data can improve tornado warning lead times for EF0- and EF1-rated events.

Acknowledgments. We thank the 12 NWS forecasters for their participation in this study, the Southern and Central Region SSD Chiefs, MICs, and SOOs for aiding recruitment, Les Lemon and Steve Martinaitis for their participation in our experiment test run, and Jimmy Correia for helping with data collection. We also thank the following colleagues for their contributions to this research. Experimental Warning Program leads Greg Stumpf, Travis Smith, and David Andra; A/V specialist James Murnan; GIS expert Ami Arthur; NWS verification expert Brent MacAloney II; WDSS-II expert Kiel Ortega; simulation expert Dale Morris; software experts Eddie Forren and Hoyt Burcham. Finally, thanks to Harold Brooks for his thought-provoking discussions on this research.

REFERENCES

Barnes, L. R., D. M. Schultz, E. C. Gruntfest, M. H. Hayden, C. C. Benight, 2009: False alarm rate or false alarm ratio? *Wea. Forecasting*, **24**, 1452–1454.

Brotzge, J., and S. Erickson, 2010: Tornadoes without NWS warning. *Wea. Forecasting*, **25**, 159–172.

Brooks, H. E., C. A. Doswell III, and M. P. Kay, 2003: Climatological estimates of local daily tornado probability for the United States. *Wea. Forecasting*, **18**, 626–640.

Crandall, B., G. Klein, and R. R. Hoffman, 2006: *Working Minds: A Practitioner's Guide to Cognitive Task Analysis*. Cambridge, MA: MIT Press, 332 pp.

Hart, S. G., 2006: NASA-Task Load Index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, Santa Monica, HFES, 904–908.

Heinselman, P. L., and S. M. Torres, 2011: High-temporal resolution capabilities of the National Weather Radar Testbed phased-array radar. *J. Appl. Meteor. Climatol.,* **50**, 579–593.

Heinselman, P. L., D. S. LaDue, and H. Lazrus, 2012: Exploring impacts of rapid-scan radar data on NWS warning decisions, *Wea. Forecasting*, **27**, 1031–1044.

Johnson, J.T., M. D. Eilts, A. White, W. Armstrong, T. J. Ganger, and M. Istock, 1999: The common operations and development environment: An environment for development and testing hydrometeorological applications. Preprints, *29th Conf. on Radar Meteorology*, Montreal, Quebec, AMS, pp. 65–68.

Sutter, D., and S. Erickson, 2010: The time cost of tornado warnings and savings with storm-based warnings. *Wea. Climate Soc.*, **2**, 103–112.

Witt, A., M. D. Eilts, G. J. Stumpf, E. D. Mitchell, J. T. Johnson, and K. W. Thomas, 1998: Evaluating the performance of WSR-88D severe storm detection algorithms. *Wea. Forecasting*, **13**, 513–518.

Zrnić, D. S., and Coauthors, 2007: Agile-beam phased array radar for weather observations. *Bull. Amer. Meteor. Soc.*, **88**, 1753–1766.

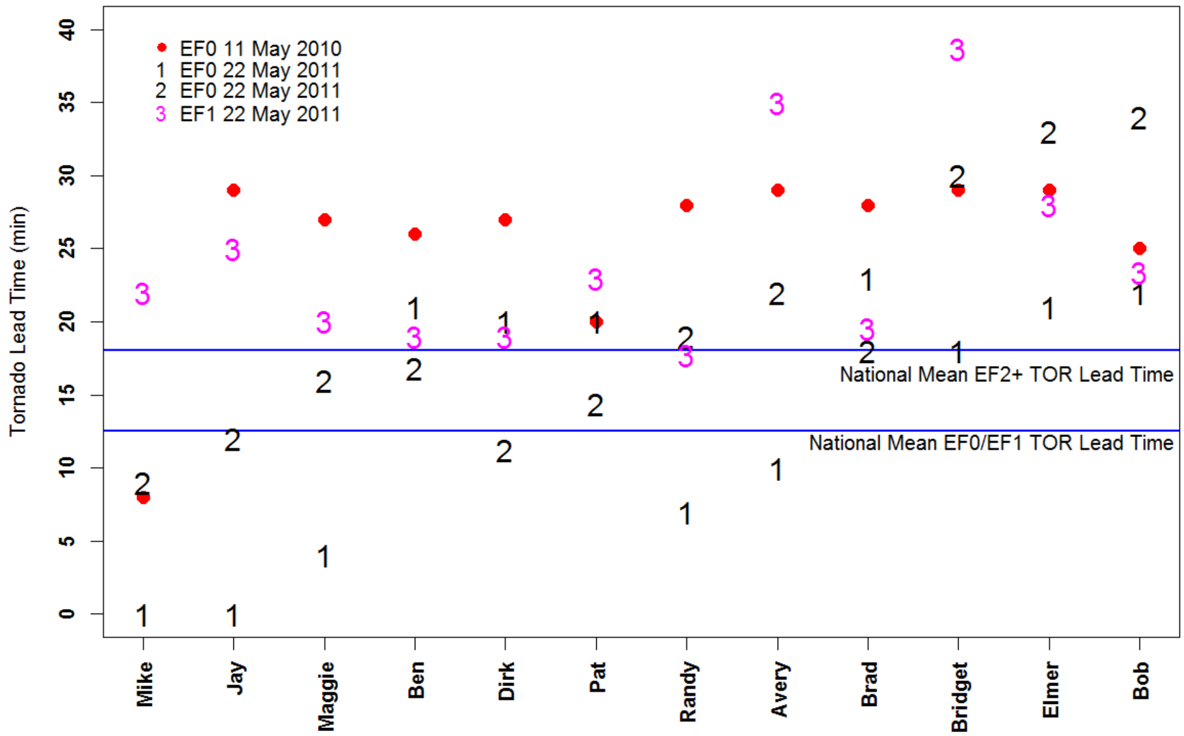

Figure 2. Distribution of tornado lead times (min) computed for 11 May 2010 and 22 May 2011 events: EF0-rated tornado on 11 May 2010 (red dot) and 3 tornadoes on 22 May 2011; EF-ratings listed in chronological order. Horizontal blue lines denote the mean national lead time computed from 1 January 2008 to 31 October 2013 for EF0 and EF1 tornadoes (12.55 min) and EF2 and higher tornadoes (18.07 min).

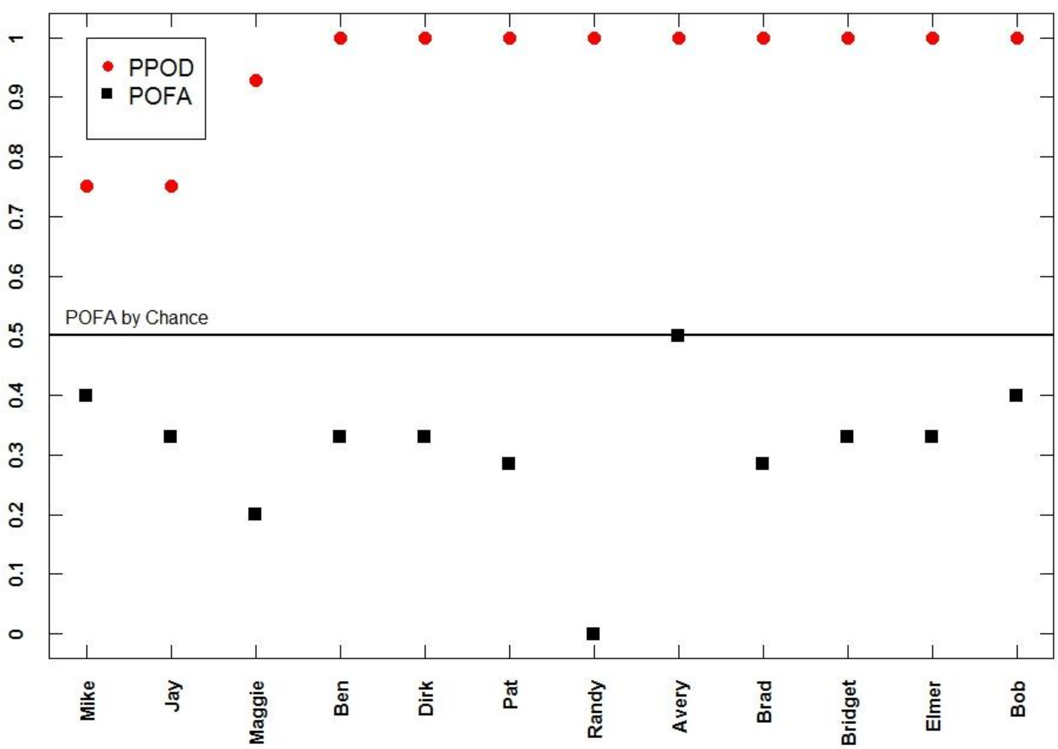

Figure 3. Distribution of polygon probability of detection (PPOD) and probability of false alarm (POFA) computed for 11 May 2010 and 22 May 2011 events. The horizontal line at 0.5 indicates the POFA attainable by chance.