Tornado Warning Decision Making with the Probabilistic Hazard Information Tool

Kerstin L. Gillespie[1], Daphne S. LaDue[2], and Christopher D. Karstens[3]

[1]National Weather Center Research Experiences for Undergraduates Program
Norman, Oklahoma
[2]Center for Analysis and Prediction of Storms
Norman, Oklahoma
[3]NOAA/NWS Storm Prediction Center
Norman, Oklahoma

ABSTRACT

The National Weather Service's (NWS) current warning system has been in place for several decades. Research has shown it has limitations with more precisely depicting the area of tornado threats, and the inability to update the warnings without reissuance, among others. This has motivated scientists to develop new ideas for warning systems, one of which is called the prototype Probabilistic Hazard Information project. Research done on this tool shows that it has limitations, but also the potential for adding to or replacing the current warning paradigm. Nine NWS forecasters were brought into the NOAA Hazardous Weather Testbed in 2017 to test the tool and warning strategy by working through several weather cases. For this study, three cases of talk-aloud data from the forecaster issuing tornado PHI objects were analyzed to determine how they made tornado warning and advisory decisions using the new tool. Findings show that forecasters were able to use the tool to accurately predict tornado formation. In contrast, tornado maintenance was not anticipated as accurately, with the tool showing their confidence decreasing over the duration of time that each PHI object would be in effect (usually 60 min). These findings indicate that PHI is useful for conveying the prediction of tornado formation and shows promise for future use in the NWS warning paradigm.

1. INTRODUCTION

The current National Weather Service (NWS) warning system has been in use for decades with minimal changes. Once issued for an entire county, warnings are now issued for particular storms (storm-based warnings, SBWs). This change to SBWs occurred about 13 years ago. A limitation that remains is that storm-based tornado warnings generally encompass an entire storm, with a buffer around it, thus many people included in the warning are not likely to be affected. This creates questions about whether new tools may allow forecasters to convey more precise geospatial warnings and if a new warning paradigm is warranted.

2. LITERATURE/BACKGROUND

Karstens et al. (2015) points out several limitations with storm-based warnings (SBWs). First, the warning implies that areas within the warning are equally under threat. Second, a warning is an instantaneous, static picture of the immediate and near-term threat from a dynamically evolving storm. Forecasters can trim their warnings but cannot extend them laterally or downstream. They must create a new warning for any area the original warning does not include.

A new forecasting strategy called FACETs (Forecasting A Continuum of Environmental Threats) is attempting to address the previously outlined issues (and many others; Rothfusz et al., 2018). The FACETs idea promotes the use of continually updated hazardous weather information that is based on probabilities. The goal is to keep warnings more up-to-date and accurate than they currently are.

People who have been under a tornado warning polygon without being directly affected by the threat can reasonably question the credibility of this system. They may form beliefs about warnings based upon those experiences and those beliefs may cause them to not perceive themselves as being under threat when they are (Klockow-McClain et al., 2014). Klockow-McClain studied the 27 April 2011 outbreak and the experiences with warnings when tornadoes occurred. Not all warnings resulted in a confirmed tornado (e.g., they are false alarms), though this

[1] *Corresponding author address:* Kerstin Gillespie, National Weather Center REU Program, 120 David L. Boren Blvd., Suite 2500, Norman, Oklahoma 73072, Klgille1@millersville.edu.

can be difficult for forecasters to correctly determine. Almost one third (.29) of the tornado warnings during the 03 May 1999 tornado outbreak were false alarms (Andra et al., 2002).

These shortcomings have prompted the development of the new, experimental Probabilistic Hazard Information warning paradigm that was tested at the NOAA Hazardous Weather Testbed in Norman, Oklahoma, from 2014–2017. This warning strategy and its associated tool were tested in experiments where forecasters were tasked with creating warning and advisory objects, which are more spatially refined than SBWs, on past and real-time events with PHI.

The Prototype PHI Tool is an experimental computer interface and warning strategy. It allows the user to create warning and advisory objects for three separate severe weather hazard categories: tornado, severe thunderstorm (hail and wind), and lightning. It requires its user to plot a curve of their confidence that a storm will produce that type of hazard for each 5-min increment of the PHI "warning." Together as one interface, there are two main aspects of the PHI Tool relevant to this study. The first is the decision to create a PHI object and the second is the curve of confidence that the forecaster creates and updates for that object.

Prior research by James et al. (2020) showed high mental demand and effort required for using the PHI tool. These challenges were quantified during the experiments with group members completing the NASA-TLX (Hart 2006) questionnaire throughout the experiment. The members of the testbed took this questionnaire to assess the mental workload of using the new tool. The members of the testbed took this questionnaire to assess the mental workload of using the new tool. Some of the reported complaints during use of the PHI tool include: hand pains from repetitive clicking and scrolling, confusion from having to track and update many objects at once, and issues with the speed and direction of the objects (James et al., 2020).

Creation and updating of tornado objects resulted in the highest mental workload of all three PHI object types (James et al. 2020). These objects were also completely manually produced, though guidance was available to help forecasters make their warning/advisory decisions. Of relevance here, the high workload was attributed, in part, because tornadoes can form and dissipate quickly. Thus, in a more continuous flow of information, there is an ongoing, constant monitoring required for tornado PHI "warnings."

Forecasters who worked with PHI reacted positively to the tool and strategy (Karstens et al., 2015). Some participants remarked that the new tool put less pressure on the forecasters to warn because PHI "warnings" could be issued at an advisory, or sub-warning level. Forecasters could begin to track the potential for a tornado from the first hints of threat and provide a more continuous flow of information throughout a potential tornado event.

In this project, data from 2017 were interpreted and analyzed from two of the three severe weather events. This research seeks to understand how far into the future a forecaster can accurately predict formation and maintenance of a tornado. These elements would help Emergency Managers, first responders, and the public to more effectively respond to severe weather.

3. DATA AND METHODS

In the NOAA Hazardous Weather Testbed project in 2017, forecasters created warnings and advisories in a lab-based environment. Each experiment lasted 3 weeks, with three forecasters working each 2-hr case each week. One forecaster focused on severe thunderstorm threats (severe wind gusts and hail), one focused on tornadoes, and the third on lightning.

Each case contained either real-time or displaced real-time weather data, which the experiment group used to create warnings and advisories on. These were communicated to Emergency Managers and Broadcast Meteorologists, who could also respond and ask questions, using tools such as NWSchat, as though it were a real-time situation. Forecasters were also provided with archived spotter reports, and archived tornado video.

This study focuses on two cases (described below). Each case presented different challenges to the experiment group. The data primarily being analyzed are the talk-aloud data written from the two cases described below. Only objects pertaining to tornadoes were studied

3.b. *Description of Cases*

- Topeka Case (NWS, 2016a)
- 25 May 2016: 2245-0045 UTC
- Brief tornado (2308–2309 UTC)

- Two long-track tornadoes with a nearly continuous path (0007–0020 UTC and 0020–0140 UTC)

This case contained one supercell thunderstorm that formed along a stationary front. Strong-to-violent tornadoes were not expected by the experiment groups that day. It turned out to become a very productive day when the storm suddenly intensified and produced a long track EF-3 tornado that started near Niles, KS. Then, the tornado dissipated and the storm produced another violent tornado (EF-4), this one lasting for 90 minutes.

- Norman Case (NWS, 2016b)
- 09 May 2016: 2030–2230 UTC
- Two tornadoes of interest, one tornado that tracked near Hennepin and Katie, OK (2106–2127 UTC); the other near Wynnewood, OK (2134–2213 UTC)

The primary focus during this case was the storm that produced the Katie and Wynnewood, Oklahoma, tornadoes. Two other tornadic storms served as additional workload for forecasters working the case.

3.c. *Description of Data Collection and Methods*

The three cases above were analyzed using the following methods. Talk-aloud data (Ericsson and Simon, 1993) from forecasters that had been transcribed from the live experiments and saved for research, were coded inductively (Boyatzis, 1998). Some codes captured the data, tools, and guidance forecasters used along with PHI to make their warning decisions. Codes also captured expressions of confidence (or lack thereof) and places where they struggled. The forecasters' work was compared and summarized across all 3 weeks using these codes. Raw PHI object data were then used to create plots for each forecaster's confidence curves, and their subsequent updates. Those curves were analyzed qualitatively to better understand what length of time the forecasters were comfortable with predicting the formation, and maintenance of a tornado.

3.d. *Limitations*

The analysis of talk-aloud data provides multiple clues about tornado decision-making while forecasters worked with PHI. A limitation found during the experiments was the participant's unfamiliarity with the new tool. Forecasters were introduced to both the warning philosophy and the computer interface at the beginning of their week of participation. They had no prior experience with either. Because we used talk-aloud data, we can account for these factors when evaluating forecaster's decision-making and warning performance. During the tests, there were several recorded instances of forecasters having trouble manipulating their PHI objects. In those cases, researchers assisted members of the group to the best of their ability to try and reduce these technical limitations.

4. FINDINGS

4.a Topeka Case 25 May 20016

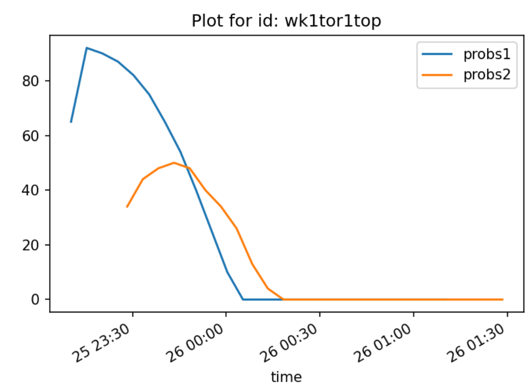

Figure 4.1 Week 1 tornado forecaster PHI object confidence curves for the object issued just after the report of a brief tornado (blue); the object was updated once (yellow).

In the first minutes of the case during week 1, the forecaster identified a boundary using spectrum width, and that there was some overhang (weak echo region). He asked in NWSChat if there were any spotters in Ottawa County, or any reports of funnel clouds. The response to this question, delayed 8 min by some experiment hiccups, was no. The first report came about 20 min into the case as a funnel cloud. Figure 4.1 combines the first confidence curve with the object update when the forecaster was responding to the brief, 90-sec tornado which occurred at 2308 (NOAA 2016). This tornado occurred shortly before the first warning (curve labeled probs1) was issued (2313 UTC). This was changed to an advisory (probs2) after the week 1 forecaster realized that the storm wasn't consistently strengthening yet. The forecaster responded to the report, "Oh, I guess so" before commenting that the storm-relative motion was "not very representative for a funnel cloud." He requested in NWSChat that spotters please let him know if the funnel cloud continues or gets closer to the ground. After this, the Warning Coordinator relayed a report from an NWS employee of a tornado on the ground. The researcher conducting the talk-aloud asks if that is surprising to the forecaster, to which they respond, "A little bit." The forecaster updated this object once, stating, "I'm not confident that that one is what I'm seeing in the radar. What was consistent is anything but consistent… I think what I want to do is take the probabilities on the first [object] and lower it."

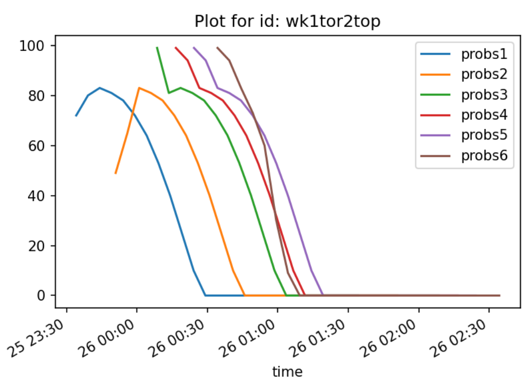

Figure 4.2 Week 1 Object 2 Topeka Case confidence curves. The first object was issued 3 minutes before the tornado reported at 2334 UTC (blue). This object had 5 updates(yellow, green, red, purple and brown).

Although the first object was associated with a circulation that no longer looked like it had tornadic potential, the forecaster said that the storm was reorganizing. He also noted that the WoF data supported the possibility that the storm would continue to pose a tornado threat. The forecaster began creating the first PHI object of the new tornado, remarking, "Well, I've got two different people seeing a funnel cloud. Last time it produced a brief tornado…" Stating, "we would be scared in my area," the forecaster created the new object with a curve that initially rose (Fig. 4.2: probs1), expressing an increase in confidence that a tornado might occur in the next 15–20 min. The first update (probs2), is issued the same way, with a rising confidence curve. The forecaster is attempting to provide lead time on tornado formation. By the third update, a tornado was confirmed, and the forecaster updated the object with an initial confidence near 100%. The confidence curve initially drops, but the forecaster then extends them out to remain at approximately 85% for about 20 min. This was done to indicate the idea that the tornado could remain in progress for that time, after which the forecaster's confidence dropped off to zero at 60 min.

In week two (not shown), a forecaster started producing the first tornado PHI object before the brief tornado at 2308 UTC, saying, "I think I am going to go ahead and just do a tornado advisory on it, just 'cause the environment it's in is ripe, and it's already got a good circulation aloft and some developing in the low levels, not really correlated in the right spot with the structure, but with the idea that it'll get that way. Warn on Forecast has some cells in the model that will maintain rotation with it down the road."

This forecaster was reasonably confident that the storm was not producing a tornado at that moment, but wanted to show potential that it could and was able to accomplish that goal using the PHI tool. When asked about how they drew the confidence curve they said, "I tried to show a maximum in confidence 30–45 minutes down the road with a somewhat decrease afterward." This resulted in the first curve's confidence rising before falling, similar to the first curve from week one. The update to this first curve was shaped in much the same way, with the forecaster remarking, "Still not really wrapping around much reflectivity into the [rear flank downdraft] or anything. I wouldn't worry about real fast tornadogenesis." They then decide to make the next object a warning after the Warning Coordinator reported a funnel cloud. Another member of the testbed remarked the tornado was reported as having dissipated after 1 min, before forecaster was able to issue the warning. It occurred 19 min after they issued the advisory.

In week three (Fig. 4.3), the forecaster fully utilized the frequent updating potential of PHI, rapidly updating the object many times in a short period of time and revealing the wavering organization of the storm before it produced the first of the two long-track tornadoes. During this period, a researcher asked them if they were happy with their PHI object. They said yes, and were "amazed" how well the PHI was tracking with the storm. They weren't sure, however, what to do for confidence until after 0000 UTC, because every time they increased confidence a new radar scan decreased it, and by the time the update with decreased confidence was issued, the storm would look more organized. They finished this first part of the case stating, "Well, now it's really unorganized," and so decided to decrease their confidence. They stated that they would keep it in the 40–60% range "because the environment supports it" (cyan curve). Figure 4.3 shows one object with 9 updates (first 10 curves) prior to the long-track tornado.

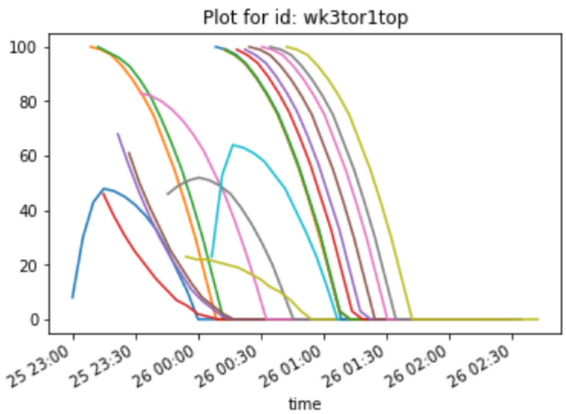

Figure 4.3: Week three Topeka Case Confidence curves. The first curve (blue) is from the brief tornado mentioned prior. This object had 16 updates (yellow, green, red, brown, etc.). All are shown in chronological order from left to right.

An EM commented in NWSChat that they liked the rapidly updating discussion and was told another update was on its way. By this time, a second tornado had touched down. It remained on the ground for 90 min (Figure 4.3: last seven curves). This second tornado reached EF-4 strength. The forecaster continued to rapidly update the object through the end of the case.

Figure 4.4 shows a summary of key codes indicating the aspects of the data or guidance that contributed to the forecaster's decisions and confidence regarding tornado formation. The top three rows are objects issued prior to the brief tornado, the next four show objects issued/updated prior to the EF-3, and the bottom three show objects/updates issued prior to the EF-4. Most of the objects issued in the Topeka case correctly anticipated a tornado and thus were marked as Yes. Accounting for experimental limitations described above, correct anticipation was indicated either verbally or in the discussion box of the PHI objects that a tornado would form with at least 5 min lead time. If the tornado warning was issued less than 5 min prior to tornado formation, or the forecaster anticipated formation that did not occur when they thought it would, it is marked as No. In all three weeks the forecasters made use of the Advisory capability of the PHI tool. They stated appreciation for the fact that the tool enables a forecaster to convey doubts about tornado formation, but also specify the location of an area of potential danger and how it is anticipated to move.

The green boxes show what data or guidance forecasters used to make each PHI object/update. Warn on Forecast (indicated as WOF) and the radar data (reflectivity and radial velocity) were verbalized most often. Reports of a tornado were verbalized just twice as important in PHI object updates, though this information often prompted an update or was included in the object discussion. ProbTor was used once in week 1 prior to the EF-3 tornado.

Did the forecaster accurately anticipate formation of a tornado?

Topeka	Brief tornado reported at 23:08 UTC				ProbTor	Rot	vr shear	WOF	Observation	radar/velocity	MRMS/satellite	SRM	Input from others
Week1	Nothing issued yet		no										
Week2	Tor1 22:50	Advisory	New object	no				x		x			
	Tor1U1 23:02	Advisory	Update	yes				x		x			
Week 3	Tor1 23:02	Advisory	New object	yes		x		x		x			x
	Niles tornado (EF3) forms at 00:07 UTC												
Week1	Tor2 23:36	warning	New object	yes				x	x				x
	Tor2U1 23:54	Warning	update	yes	x							x	
Week2	Tor1U6 00:05	warning	Update	yes						x	x		
Week 3	Tor1U8 23:56	Warning	Update	yes		x		x		x			
	Niles tor dissipates; Solomon tor (EF4) forms 00:20												
Week1	Tor2U3 00:19	Warning	update	yes~								x	
Week2	Tor1U7 00:15	warning	Update	yes		x			x	x			
Week 3	Tor1U13 00:19	Warning	Update	yes		x							x

Figure 4.4: Topeka case talk-aloud analysis prior to tornado formation. Products/information verbalized as important to the decision are marked with a green X. The column marked Yes/No indicates whether the forecaster correctly anticipated tornado formation.

Did the forecaster accurately anticipate maintenance of a tornado?

Topeka	Brief tornado reported at 23:08 UTC				ProbTor	Rot	vr shear	WOF	Observation	radar/velocity	MRMS/satellite	SRM	Input from others
Week1	Nothing issued yet			no*									
Week2	Tor1U1 23:02	Advisory	Update	no*				x		x			
Week3	Tor1 23:02	Advisory	New object	no*		x		x		x			x
	Niles tornado (EF3) forms at 00:07 UTC												
Week1	Tor2U1 23:54	Warning	update	yes	x							x	
Week2	Tor1U6 00:05	warning	Update	yes						x	x		
Week3	Tor1U8 23:56	Warning	Update	yes		x		x		x			
	Niles tor dissipates; Solomon tor (EF4) forms 00:20												
Week1	Tor2U3 00:19	Warning	update	no*								x	
Week2	Tor1U7 00:15	warning	Update	no*		x			x	x			
Week3	Tor1U13 00:19	Warning	Update	no*		x							x

Figure 4.5: As above, Topeka Case Talk-Aloud analysis for tornado maintenance. In this chart, the Yes/No column addresses whether the forecaster has accurately anticipated the maintenance of a tornado.

Figure 4.5 assesses forecasters' ability to project the maintenance of an existing tornado. The first tornado of this case was short-lived, lasting only 90 sec. In addition, initial tornadoes are notoriously difficult to warn for (Andra et al., 2002). The EF-3 tornado was strong and lasted approximately 13 min. The week 1 forecaster kept their warning because their confidence was increasing that the updraft intensity was increasing. They drew the curve to show high confidence of a tornado for the next 5–10 minutes, then after that left the curve as it was because the storm had a history of cycling. The forecaster declared, "This is probably not a great day for a 50-mile-long tornado. But you might get several shorter ones." For the EF-4 tornado, all of the forecasters had their confidence graphs at high levels for at least 15 min. Therefore, tornado maintenance was anticipated by all three forecasters.

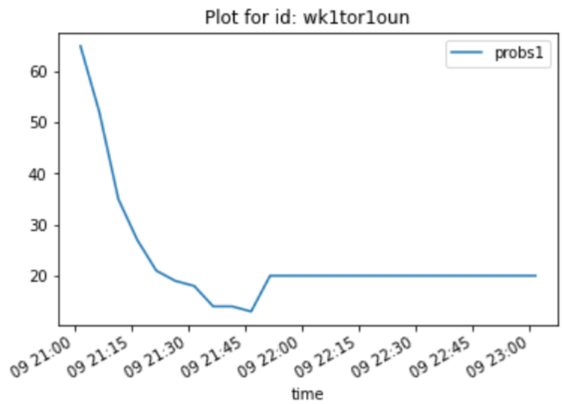

Figure 4.6: Week 1 OUN case: the first tornado object issued this day; this object was not updated. This was the only object issued before 2106 UTC when the Katie tornado officially touched down.

The final tornado of the case lasted 90 min. On a day calling for only a slight risk of severe weather (Storm Prediction Center, 2016), a long tracked and violent tornado was not anticipated. This was especially apparent when the week one forecaster remarked, "Might lower this down to 30 minutes. If you think of the model, that's what'll happen. But this [pointing at the screen] developing on the flank, I'd think this [circulation] is going to take over. But that could be totally wrong. I can't see where the boundary is." During week three, the forecaster had confidence that the final tornado could last for some time, but their estimate did not accurately predict the duration of the tornado. Researchers pointed out that new WOF data had come in. The forecaster said it was "suggesting in the next 10 minutes we're staying pretty strong... even the next 20... so this thing [tornado] could sustain itself for a while."

In the Norman Case, the Week 1 forecaster issued one PHI object for the first circulation on the storm in the OUN case (Fig 4.6). The forecaster remarked, "I'm not really confident, because it has been fluctuating a bit. So I'm going to keep it at over 50% to start and then bring it down a little bit." The forecaster then began to create and edit a second tornado object for the new circulation on the storm. By the time they finished issuing their second tornado object, the forecaster saw signs that a tornado was imminent. Confidence on the second tornado object was raised to ~87% (Figure 4.7, dark blue) after the forecaster stated, "It's gotten stronger, you can see kind of a hook here." Later in the case, the three final updates (red, purple, brown) became further apart when multiple storms drew forecasters' attention to other places.

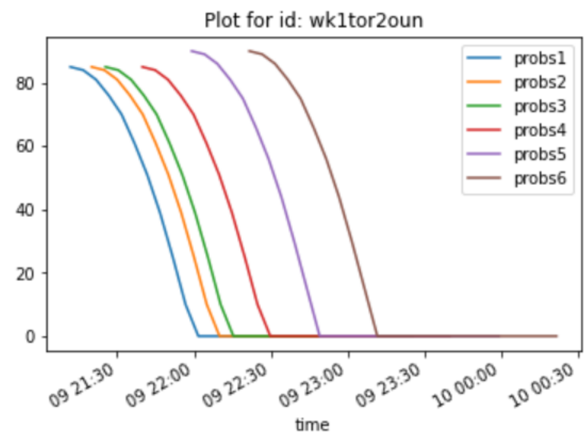

Figure 4.7 week one OUN graphs for the Katie storm: The creation of the object is shown in blue. Each update is issued in order of the legend displayed to the right of the curves.

In Week 2 of the Norman Case, (Fig. 4.8) the first object was created at 2056 UTC (probs1). The researcher asked how the forecaster felt about issuing that object. "Eh... Overall I'm feeling pretty good. Starting to wonder what's happening with this storm because it looks like we're getting some really rapid development on the southern flank. So that might actually overtake the northern part. So you might actually get development further south than the current tornado object shows."

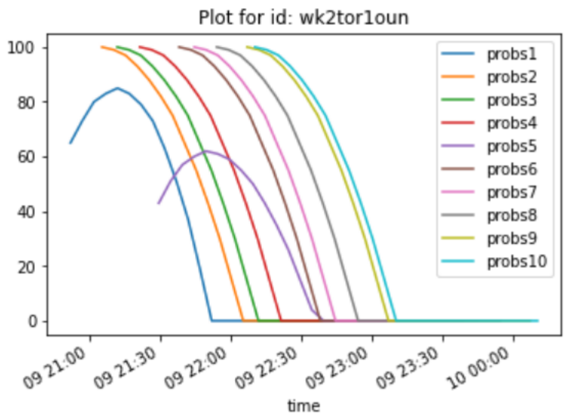

Figure 4.8 Week 2 OUN Graph for first tornado. First object seen in blue. All object updates are in the order seen in the legend to the right of the graph.

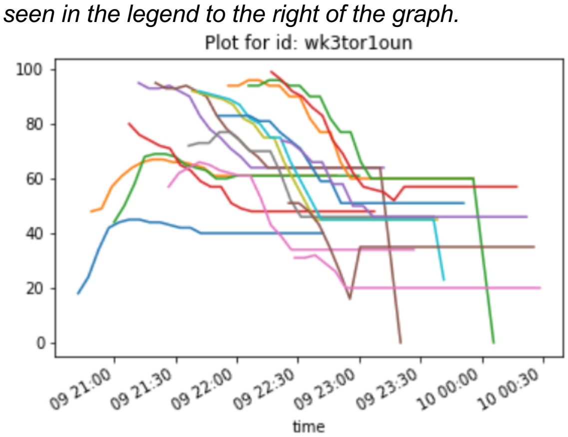

Figure 4.9 week 3 OUN Graph: The first object issued (in dark blue) was updated many times in this week of the testbed.

The researcher then proceeded to ask if they kept the warning length to 60 minutes. Their response was, “Yeah, I kept it at the 60 because I felt like… that would be one that would be the more dominant updraft. And depending on how this develops it still could be, it could all merge together.” The researcher asked the forecaster if they used an advisory or a warning to which the response is a warning. When asked why, the forecaster said, ”Well, at the time I thought I had rotation wrapping up, and mid-level rotation was strengthening and looked like that was coming down, and the low levels was starting to respond to that. But that was before I saw that there was new development further south, so that's gonna impact how that evolves." The Warning coordinator then reported a tornado on the ground 4 miles NNW of the town of Hennepin in Garvin County, OK. This is followed by another update (yellow). This tornado continued for the next 21 min.

Later, the Warning Coordinator reported the tornado had lifted just before it crossed the I-35. The forecaster updated the object (Figure 4.8, purple) saying, "Still think need a warning... something could form... so going with radar indicated ...so those are the probabilities I'm given but going to adjust up a little bit." A researcher then asks, “Did the WOF or the ProbTor play into the adjustments you just made?” to which the response is, "Certainly it allowed me to know that there's still a very good chance that something could happen in the future, provided that forecast information. Certainly no reason to drop the warning completely, though nothing's being observed, but that there's still the possibility of something occurring. People need to be aware of that." This tornado lasted another 39 min.

In week three (Fig. 4.9), ProbTor on the Wynnewood storm had risen to 9%, which the forecaster said was significant enough to warrant an advisory (blue). The forecaster drew their confidence curve by hand. This was the first time that technique had been seen in the experiment.

The forecaster then decided to update the PHI object for the main storm. When asked why, they responded that 10 min had elapsed since the last edit on that object. Because the ProbTor guidance was down 20% the forecaster adjusted their confidence curve upward (yellow) and they upgraded the object to a warning. The updated object was issued near 2106 UTC, when the tornado was later determined to have formed (NOAA 2016). Asked why they drew the graph the way they did, they answered they were not sure what the storm would do past 20 min, so the remainder of the curve was persistence-based.

The Warning Coordinator then reported a chaser observing a multi-vortex tornado. During the next few minutes, the forecaster updated the tornado warning object, which was then issued at 2114 UTC. Just as this was issued, the Warning Coordinator relayed another report about footage of a large tornado west of I-35. The tornado forecaster remarked that the storm wasn’t as strong aloft as expected, but the faint hook had gotten stronger. The forecaster then said that the gate-to-gate shear had weakened. After seven minutes, the forecaster asked if there were any new reports on the main storm, adding, "Wouldn't be surprised to hear that the tornado is occluding out and a new circulation is forming." With that, the Warning Coordinator relayed a chaser report that the tornado in Garvin County had dissipated.

Anticipating formation (Fig. 4.10) was more straightforward in this case with the storm forming in an environment supportive of tornadoes. Both tornadoes lasted a reasonable amount of time, 21 and 40 minutes respectively. Strong tornadoes were expected that day, with early development moving into an area of Enhanced severe risk (Storm Prediction Center, 2016). One of the forecasters acknowledged this in week two, stating, "[The storm is] in a good place for severe weather and tornadoes. It's right where a dryline and that outflow come together." Because of environmental factors like this, forecasters correctly anticipated formation all three weeks. However, the forecaster in week one was not able to issue their PHI object before the tornado touched down due to being unfamiliar with the PHI tool.

Did the forecaster accurately anticipate formation of a tornado?													
OUN	Katie tornado forms 21:06				ProbTor	Rot	vr shear	WOF	Observation	radar/velocity	MRMS/satellite	SRM	Input from others
Week1	Tor1 21:05	warning	New object	no*		x	x				x	x	
Week2	Tor1 20:56'	warning	Update	yes	x			x		x			x
Week3	Tor1U2 21:02:	warning	Update	yes		x		x	x	x			
	NEXT TORNADO FORMED AT 21:34 UTC												
Week1	Tor2U2 21:30	warning	Update	yes				x	x	x		x	
Week2	Tor1U4 21:31'	warning	Update	yes		x		x	x	x			
Week3	Tor1U6 21:29:	warning	Update	yes					x	x			

Figure 4.10 OUN Tornado Formation Talk -Aloud analysis. The table is divided in the order of each tornado. The type of product issued is listed toward the center of the graphic. All products/ information used to assist with the decision is marked by a green X.

Did the forecaster accurately anticipate maintenance of a tornado?													
OUN	Katie tornado forms 21:06				ProbTor	Rot	vr shear	WOF	Observation	radar/velocity	MRMS/satellite	SRM	Input from others
Week1	Tor1 21:05			yes		x	x				x	x	
Week2	Tor1 20:56'			Yes	x			x		x			x
Week3	Tor1U2 21:02			yes		x		x	x	x			
	NEXT TORNADO FORMED AT 21:34 UTC												
Week1	Tor2 21:30			yes				x	x	x		x	
Week2	Tor1U4 21:31'			yes		x		x	x	x			
Week3	Tor1U6 21:29:			yes					x	x			

Figure 4.11 OUN Tornado Maintenance Talk-Aloud analysis.

Next, we looked at whether or not forecasters were able to project the maintenance of the tornado (Fig. 4.11). Forecasters working with the Norman Case had an expectation of severe weather that day. During week three, the forecaster said, "Yeah, there are several [WOF boxes] that are moving that way, increasing potential for a long track storm, apparently." The forecasters had more signals pointing to a day that could produce strong tornadoes and thus had more accurate ideas about tornado maintenance. Regarding the tools used to assist them in their work, ProbTor was once again only mentioned one time in the assessment of tornado maintenance.

4	23	2	22	33	46	3	11	14
ProbTor	Rot	vr shear	WOF	Observation	radar/velocity	MRMS	SRM	Input from others

Figure 4.12: Sum of all instances of a coded segment data across both cases. Counts indicate how often each of these were voiced by forecasters in their warning/advisory decisions. The tools listed include (from left to right respectively) ProbTor, signs of rotation, VR shear (a calculation of rotational shear; NOAA 2021), Warn on Forecast guidance, Radar/velocity raw data, Multi-Radar Multi-Sensor, storm relative motion, and any input from other members of the testbed, including researchers.

Summing codes across all PHI objects, four aspects of data or guidance were used most often in making warning decisions (Fig. 4.12): Radar/Velocity (*N*=46), observations (e.g., storm reports; *N*=33), signs of rotation (*N*=23), and WOF guidance (*N*=22). Rotation and VR shear two were also noted as important base datasets of warning decision making by Andra et al. (2002).

5. DISCUSSION & CONCLUSIONS

The preceding analysis showed that PHI allowed forecasters to express their correct anticipation of both tornado formation and maintenance for two tornadic cases studied in the 2017 NOAA HWT Prototype PHI Project. Other cases could yield different results. The analyses reported here suggest that PHI may add value in addition to or as a replacement for the current warning system. While it may possibly not ever be deployed as a public system, the research above shows potential is there for it to amend the current one as well.

These types of studies are important because further reduction in the number and total area of false alarms could potentially increase trust in the NWS's current warning system. Many users of the current warning system do not act without additional, corroborating evidence (e.g., Kuligowski 2020) because they do not believe the system is accurate.

The hope for the future is that PHI can increase lead time and decrease false alarms. This translates to saving lives because it would give people more time to react and a sliding scale of severity to follow instead of simply warning or no warning.

6. ACKNOWLEDGMENTS

Thanks go to my mentor, Dr Daphne LaDue, and to Dr. Chris Karstens for being so supportive throughout this analysis. My mentor implemented the talk-aloud protocol in 2017, capturing forecasters' thoughts as they learned about new tools and issued "warnings" in the PHI paradigm. Dr. Karstens captured the PHI objects and made screen videos available for my analysis and provided helpful guidance on how to proceed. Thanks also go to Dr. Mark Laufersweiler for teaching Software Carpentry and helping me work out how to code to create the graphs in Python. Without his assistance I wouldn't have been able to create them. Thanks also go to my fellow REU students for their camaraderie and support.

This work supported by the National Science Foundation under Grant No. AGS-2050267 and used data collected under U.S. Department of Commerce, National Oceanic and Atmospheric Association grants NA15OAR4590187 and NAISNWS4680019. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of NSF, NOAA, or the U.S. Department of Commerce.

7. REFERENCES

Andra, D. L., E. M. Quoetone, and W. F. Bunting, 2002: Warning decision making: The relative roles of conceptual models, technology, strategy, and forecaster expertise on May 3, 1999. *Weather and Forecasting*, **17,** 559-566.

Boyatzis, R. E., 1998: *Transforming Qualitative Information: Thematic Analysis and Code Development*. SAGE Publications, 184 pp.

Ericsson, K. A., and H. A. Simon, 1993: *Protocol Analysis: Verbal Reports as Data.* Revised ed. The MIT Press, 443 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.

James, J. J. J., C. Ling, C. D. Karstens, J. Correia Jr., K. M. Calhoun, T. C. Meyer, and D. LaDue, 2020: Forecasters' cognitive task analysis and mental workload analysis of issuing probabilistic hazard information (PHI) during facets PHI prototype experiment. *Weather and Forecasting*, **35**, https://doi.org/10.1175/WAF-D-19-0194.1.

Karstens, C. D. and Coauthors, 2015: Evaluation of a probabilistic forecasting methodology for severe convective weather in the 2014 Hazardous Weather Testbed. *Weather and Forecasting*, **30**, 1551-1570, http://dx.doi.org/10.1175/WAF-D-14-00163.1.

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

Kuligowski, E. D., 2020: Field research to application: a study of human response to the 2011, Joplin tornado and its impact on alerts and warnings in the USA. *Natural Hazards*, **102,** 1057-1076, https://doi.org/10.1007/s11069-020-03945-6.

NOAA, 2021: AWIPS Fundamentals D2D Tools, Accessed August 3, 2021, https://vlab.noaa.gov/web/oclo/awipsfundamentals?page=radar.

NOAA NCEI, 2016a Storm Events Database. Accessed July 27, 2021, https://www.ncdc.noaa.gov/stormevents/listevents.jsp?eventType=ALL&beginDate_mm=05&beginDate_dd=25&beginDate_yyyy=2016&endDate_mm=05&endDate_dd=26&endDate_yyyy=2016&county=ALL&hailfilter=0.00&tornfilter=0&windfilter=000&sort=DT&submitbutton=Search&statefips=20%2CKANSAS

NWS, 2016: The Severe Weather and Tornado Outbreak of May 9, 2016. Accessed July 26, 2021, https://www.weather.gov/oun/events-20160509

Storm Prediction Center, 2016a: Accessed July 27 2021, https://www.spc.noaa.gov/products/outlook/archive/2016/day1otlk_20160525_1630.html

Storm Prediction Center, 2016b: Accessed July 27, 2021, https://www.spc.noaa.gov/products/outlook/archive/2016/day1otlk_20160509_2000.html