

# Forecasters' Cognitive Task Analysis and Mental Workload Analysis of Issuing Probabilistic Hazard Information (PHI) during FACETS PHI Prototype Experiment

JOSEPH J. J. JAMES AND CHEN LING

*University of Akron, Akron, Ohio*

CHRISTOPHER D. KARSTENS

*NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma*

JAMES CORREIA JR.

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

KRISTIN CALHOUN, TIFFANY MEYER, AND DAPHNE LADUE

*NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma*

(Manuscript received 7 October 2019, in final form 12 May 2020)

## ABSTRACT

During spring 2016 the Probabilistic Hazard Information (PHI) prototype experiment was run in the National Oceanic and Atmospheric Administration (NOAA) Hazardous Weather Testbed (HWT) as part of the Forecasting a Continuum of Environmental Threats (FACETS) program. Nine National Weather Service forecasters were trained to use the web-based PHI prototype tool to produce dynamic PHI for severe weather threats. Archived and real-time weather scenarios were used to test this new paradigm of issuing probabilistic information, rather than deterministic information. The forecasters' mental workload was evaluated after each scenario using the NASA-Task Load Index (TLX) questionnaire. This study summarizes the analysis results of mental workload experienced by forecasters while using the PHI prototype. Six subdimensions of mental workload: mental demand, physical demand, temporal demand, performance, effort, and frustration were analyzed to derive top contributing factors to workload. Average mental workload was 46.6 (out of 100, standard deviation: 19, range 70.8). Top contributing factors to workload included using automated guidance, PHI object quantity, multiple displays, and formulating probabilities in the new paradigm. Automated guidance provided support to forecasters in maintaining situational awareness and managing increased quantities of threats. The results of this study provided understanding of forecasters' mental workload and task strategies and developed insights to improve usability of the PHI prototype tool.

## 1. Introduction

The severe convective weather warning system in the United States has maintained the same weather warning paradigm since the first successful issuance of a tornado warning in 1948 (Meyer 2003), comprising watches, warnings, and advisories. The current warning system comprises 122 National Weather Service (NWS) offices, and their affiliated national centers, continuously analyzing radar, satellite, lightning, surface observations,

and model data to produce watches and warnings for severe thunderstorms, tornadoes, floods, hail, and other extreme weather phenomena (Friday 1994). The current warning paradigm is static and conveys the location of a threat via a graphical polygon and city–town location text. Polygons can be trimmed as threats evolve until expiration of duration, and current directives and practice indicate that at least a few updates should be issued (Stern 2020).

The advancement of the weather warning methods requires a new means to graphically represent and communicate threat information to users. The Probabilistic

---

*Corresponding author:* Joseph J. J. James, [jjj27@zips.uakron.edu](mailto:jjj27@zips.uakron.edu)

DOI: 10.1175/WAF-D-19-0194.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

Hazard Information (PHI) object is the current mechanism being evaluated to communicate accurate and timely information regarding specific meteorological threats (Kuhlman et al. 2008; Stumpf et al. 2008; Karstens et al. 2018).

Recent advances in algorithm development make it possible to automatically locate severe weather hazards spatially, and thus graphically. Using a new suite of fused radar, environment, and statistics, these hazards can be tracked and projected into the future, and while accounting for uncertainty in those projections, used to create a PHI object (polygon, freehand or ellipse). These added capabilities, including probabilistic information, were put to the test as an improvement to deterministic warning polygons for both forecasters and end users, such as emergency managers and TV broadcasters (Karstens et al. 2018).

#### *a. The PHI prototype tool*

Within the PHI prototype tool, there are both automated and manual PHI objects available to the forecasters for severe thunderstorm and lightning threats (tornado threats do not have automated guidance available). Forecasters have the option to either manually create new PHI objects, modify any of the attributes (hazard, severity, motion, duration, shape, location, time, and forecasted probability trends) of an existing object, or block an automated object (Karstens et al. 2015). Information communicated by the PHI object include current and future threat probability, time of arrival and departure, threat type, and severity of threat. Another attribute of PHI objects is the discussion box attached to each PHI object, where forecasters can comment or discuss their reasons for assigning certain PHI attributes or comment on storm development. PHI objects can be automatically created by software (severe thunderstorm and lightning) and can be manually created, modified, and blocked by forecasters (Karstens et al. 2015). Radar-derived PHI objects will automatically track a storm even if it begins at below current warning threshold levels. According to the NWS, warning thresholds for severe thunderstorms are  $25.7 \text{ m s}^{-1}$  for wind and 2.5 cm for hail; for tornado warnings, radar detected or a spotter report.

Severe thunderstorm PHI objects are based on the ProbSevere model (Cintineo et al. 2014, 2013, 2018), developed by the National Oceanic and Atmospheric Administration (NOAA) and the Cooperative Institute for Meteorological Satellite Studies (CIMSS). ProbSevere data are computed from a compilation of numerical weather prediction (NWP), geostationary satellites, ground-based radars, and cloud-to-ground lightning, which identifies areas of convection and calculates

the probability that a convective area will produce severe weather. The model forecasts a probability of severe weather for a 90-min duration.

Lightning PHI objects are developed using a random forest algorithm trained with Multi-Radar/Multi-Sensor (MRMS), near-storm environment, and both in-cloud and cloud-to-ground lightning data from multiple lightning detection networks (Meyer et al. 2016; Calhoun et al. 2018).

#### *b. Mental workload*

Mental workload is the cognitive demand a system or task imposes on the user (Wickens et al. 2004). Mental workload analysis is important in system or product design to ensure manageable workload for users. Mental workload is a relative measure used to determine differences between similar designs or changes in design. Properly assessing the mental workload and implementing design changes into a software system may increase the system usability. If the mental workload is too high, the user may feel fatigued and quality of work will decrease. If the workload is too low, the user will not be as engaged in the process and may lose situational awareness.

This study implements the NASA Task Load Index (NASA-TLX), one of the most widely used and accepted method to measure mental workload (Hart and Staveland 1988; Hart 2006). The NASA-TLX has been used across many industries and is accepted as an accurate and reliable tool to measure mental workload (Akyeampong et al. 2014; Finomore et al. 2013). The NASA-TLX has been used to evaluate driver workload in autonomous vehicle systems (Hooey et al. 2018) and to manage workload of robotic surgery operators and increase efficiency (Walters and Webb 2017). In addition, the NASA-TLX has been used to evaluate mental workload in implementing a conflict detection and resolution advisory system for air traffic controllers (Trapsilawati et al. 2016). Other applications include control interface designs in nuclear power plants (Yan et al. 2017) and evaluating mental workload in virtual environments in comparison to traditional environments (Burigat and Chittaro 2016).

The goal of this study is to understand forecasters' tasks in issuing PHI objects, analyzing mental workload experienced by forecasters and summarizing any task strategy that forecasters' develop in managing PHI objects.

## **2. Methodology**

The 2016 PHI prototype experiment implemented practitioners' cycles (Hoffman et al. 2010), which is an

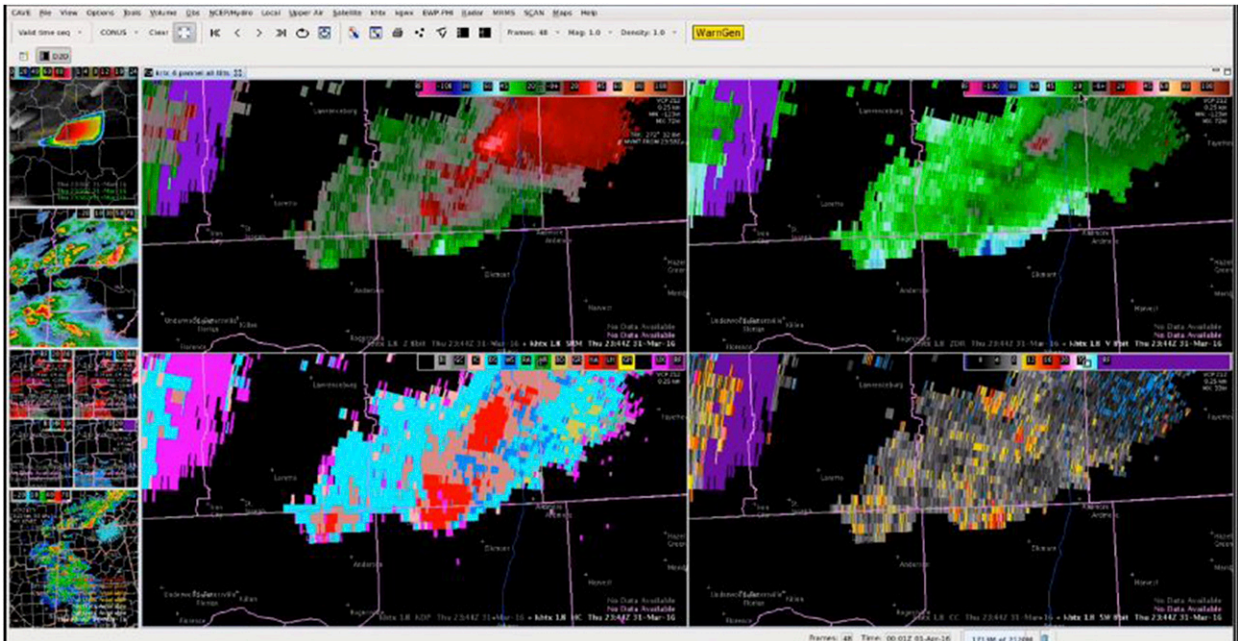


FIG. 1. 2016 PHI prototype hazardous weather experiment setup.

iterative design process that allows for rapid product improvement. After each week of testing, improvements were made to the system for the next week of testing, thus allowing researchers to analyze quickly how improvements affected aspects of the system and overall usability.

#### a. Hazardous Weather Testbed design

The 2016 Probabilistic Hazard Information (PHI) prototype experiment was conducted 9 May–10 June 2016 at the National Weather Center in Norman, Oklahoma. The experiment took place in the Hazardous Weather Testbed (HWT), a specifically designed test laboratory to complete experimental testing and development of new weather warning products and systems. The testing area consists of six dual 27-in. monitor Linux workstations.

During the experiment, Forecasters used the PHI prototype tool in conjunction with the Advanced Weather Interactive Processing System (AWIPS II) (Fig. 1). AWIPS II provides the standard radar displays and information that the forecasters use currently in operation at weather forecast offices (WFOs). Figure 1 shows the forecaster workstation with AWIPS II on the left screen and the PHI tool on the right screen (Fig. 3). Forecasters used the PHI tool with automated guidance to create and manage PHI objects to convey threat information on tornado, severe thunderstorm, and lightning threats. Threat objects produced by forecasters are populated in the Enhanced Data Display (EDD) and

displayed in another room to the emergency managers and media broadcasters for decision support (Wolfe 2014) (Fig. 2). The EDD tool is a web-based tool created to display and provide detailed PHI information to the end users. This study focuses on the forecaster side of the experiment, for understanding how forecasters use the PHI prototype tool to produce and manage PHI objects.

#### b. Procedure

On the first day of each week of the experiment participants were presented with an overview of the Forecasting A Continuum of Environmental Threats (FACETs) project (Rothfus et al. 2015) and tutorials for the PHI prototype tool. At the end of the first day, participants completed a hands-on training session with the PHI tool. Days 2, 3, and 4 were offset shifts from 1300 to 2200 local time (LT), to take advantage of the favored time for severe weather events. The first scenario each day was an archived case specifically chosen to test the capability of the PHI tool. The second session of each day utilized real-time severe weather in the CONUS. NASA-TLX surveys were completed after each archived or real-time weather event. Guided group debriefs took place after the event. The guided sessions covered topics of interest that resulted from forecaster decisions or situations that happened during the previous weather event. For each case, three forecasters were in charge of three types of threats. These responsibilities would rotate each day so each forecaster worked on all



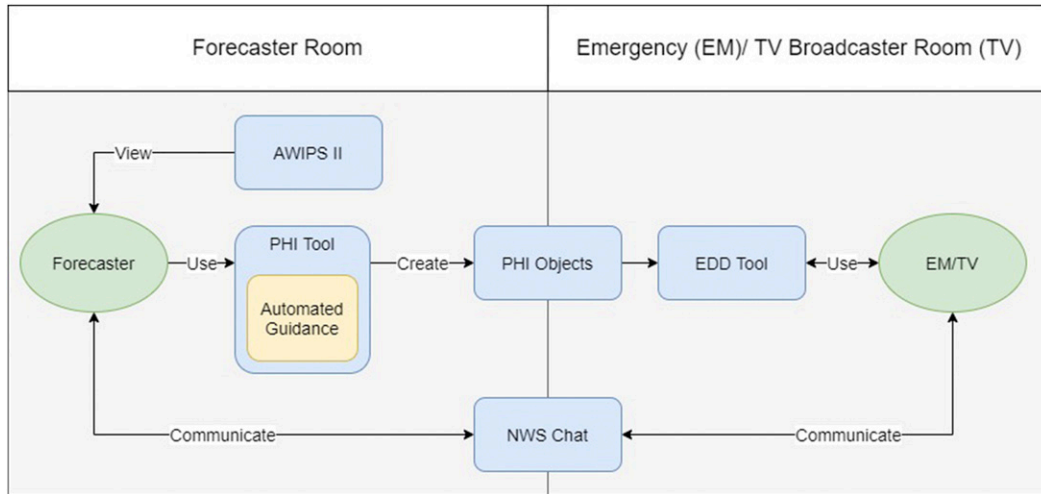


FIG. 2. AWIPS II four panel display of the HWT forecaster station. Forecasters had the opportunity to set up procedures as desired in AWIPS II before beginning a case. This panel varied for each forecaster and could be adjusted depending on the environmental conditions of the case.

three hazards over the course of the week. Following 3 days of PHI scenarios, researchers conducted a comprehensive guided discussion regarding the PHI tool interface, creation, and management of PHI objects, forecaster thoughts on the weather warning paradigm shift, and the logistics of the experiment. Throughout the week, each forecaster completed three archived

cases and four real-time severe weather events. The duration of each event varied from 1.5 to 2 h.

*c. Participants*

The PHI prototype experiment was conducted for three weeks. Three new forecasters participated each week. Participants were chosen from a pool of applicants

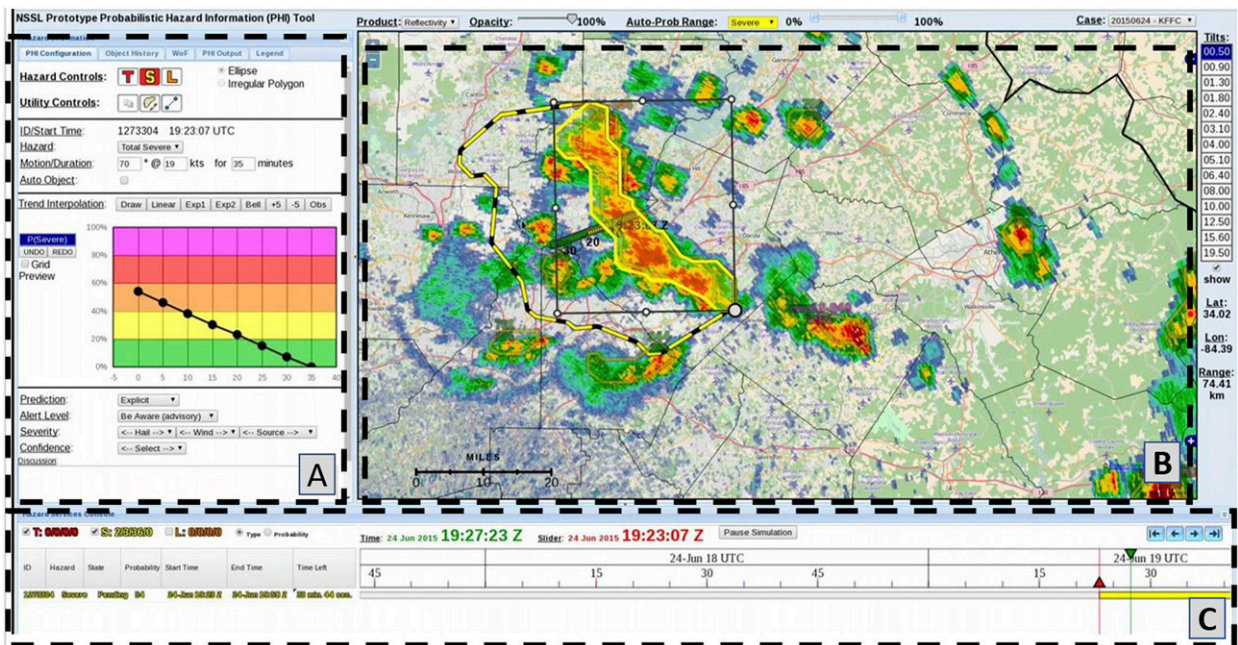


FIG. 3. PHI tool panel of the HWT forecaster station. (a) The Hazard Information Display (HID) allows forecasters to select PHI object attributes such as motion vector, probability trend, threat attributes, and add discussion. (b) The PHI tool spatial display allows forecasters to see PHI objects overlaid with reflectivity or velocity products on a map. Users can use the cursor to manipulate object vertices and motion vectors by dragging the PHI object. (c) The console allows forecasters to scroll through past time up to current and view all the start and finish times of PHI objects they have edited.

within the NWS. Considerations in forecaster selection included: region, position, warning experience, and motivation for participation. Nine forecasters participated: seven men and two women. The average participant age was 42.7 years. Participants had an average of 13.9 years of warning experience, ranging from 2 to 20+ years and represented 8 different weather forecast offices.

#### d. PHI prototype tool design

The PHI prototype was developed as a web browser-based tool (Fig. 3) (Rothfus et al. 2014; Karstens et al. 2015). This interactive tool allowed users to create and manage PHI objects for tornado, severe thunderstorm, and lightning threats. A PHI object is a geographically outlined area, within or around a storm, which represents a threat. The PHI tool control panel (area B in Fig. 3) gives the forecaster options to describe the threat motion and duration, probability trend, action level, severity, and confidence. There is also a discussion box for additional comments and information. The tool includes an interactive radar map (area B in Fig. 3) on the right side to help create, modify, and manage PHI objects. On the bottom of the PHI tool screen, the console (area C in Fig. 3) allows forecasters to track PHI objects they are managing and see when the objects will end.

#### e. Automated guidance

The PHI tool used automated guidance from ProbSevere and ProbLightning to create and track PHI objects, along with providing a probability level for each object (Karstens et al. 2015). The 2016 PHI prototype tool provided automatic object creation and tracking for severe thunderstorms and lightning threats.

The PHI tool allowed forecasters to take control of aspects of automated PHI objects as they deemed necessary. If the forecasters did not believe the PHI objects accurately represented the threat, forecasters could choose to take over all aspects of an automated object or just override aspects that were not accurately representing the hazard. As described by Karstens et al. (2018), there were four levels of automation that the forecaster could use:

- Level 1: Forecaster generates all probabilistic forecasts by manually creating a PHI object, with no involvement of automated guidance.
- Level 2: Forecaster optionally uses automated guidance to generate probabilistic forecasts. Automated guidance is running, but all aspects of a PHI object, including probability, size and shape of object, and motion vector can be overridden.

- Level 3: Forecaster partially overrides automation. Automated guidance is running, and all attributes except the mechanical attributes (size, shape, motion vector, and duration) of a PHI object can be overridden.
- Level 4: Forecaster observes automatic probabilistic forecast generation without any intervention. Automated guidance is running and is generating probabilistic forecasts.

#### f. Description of archived cases

Three archived cases were selected for use during the PHI experiment. The severe weather situations were chosen to present extreme and marginal weather conditions that help researchers to understand the capabilities and limits of the PHI prototype tool and provide challenges to the forecasters.

Case 1: 6 May 2015, in Oklahoma City, Oklahoma.

The situation included multiple large severe supercells with tornadoes (Fig. 4).

Case 2: 31 March 2016, in Huntsville, Alabama. The situation included merging supercells with multiple tornadoes (Fig. 5).

Case 3: 24 June 2015, in Atlanta, Georgia. The situation included many dispersed pulse storms (Fig. 6).

#### g. Instruments and analysis

##### 1) TASK ANALYSIS

Hierarchical task analysis (HTA) (Stanton et al. 2013) was used to develop a task model to evaluate the performance of PHI object creation and management. The PHI process was broken down to basic elemental tasks that forecasters needed to complete to create or update a PHI object. Analysis was performed to measure and analyze the time forecasters spent creating and updating objects as well as the length of time between updating objects.

Video recording of each session was completed using a built-in computer function, RecordMyDesktop, Ver 3.8 (Varouhakis 2007), to record each screen. Tripod mounted video cameras with microphones recorded an “over the shoulder” view and forecaster discussion during the experiment. Screen recordings were annotated with usability software (Morae by TechSmith, Ver 3.3.4), based on forecaster interaction with the PHI tool. Cases analyzed included nine sessions for three hazards, three different cases, and nine different forecasters. All actions and decisions by the forecaster during PHI creation and management were annotated. Each PHI object is designated a unique numerical identifier that remains the same as the object is tracked,



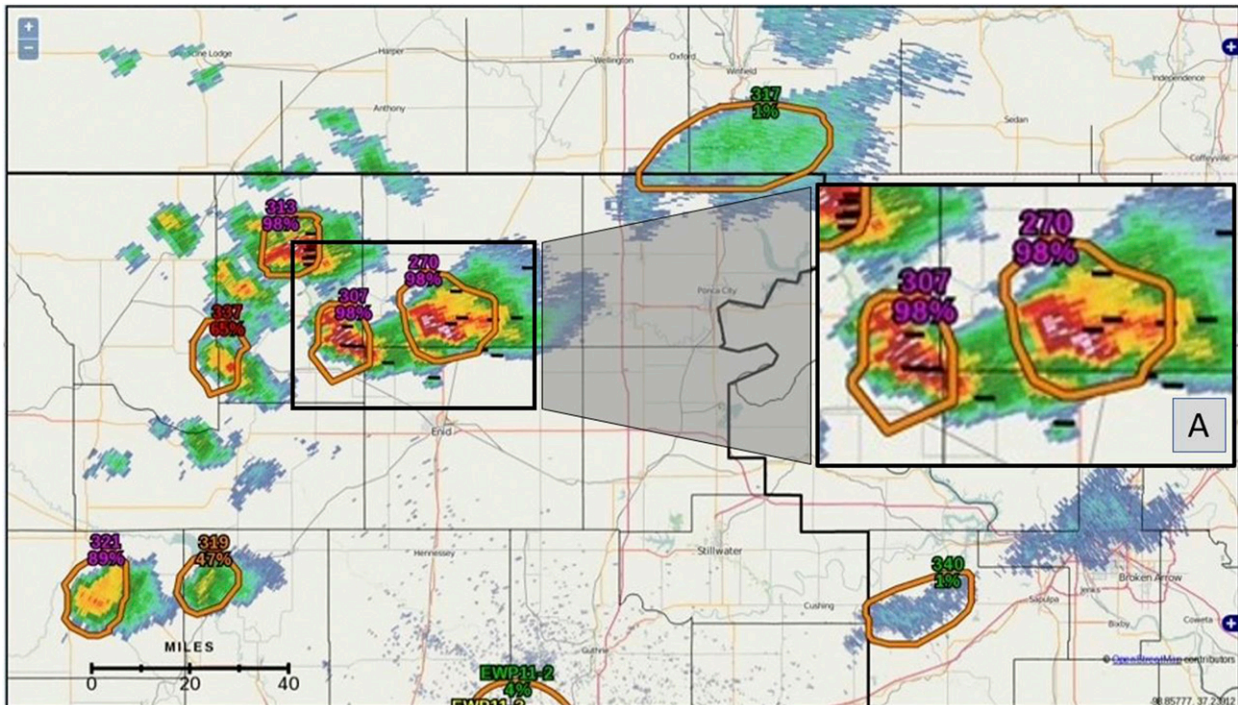


FIG. 4. Archived Case 1 was selected from 6 May 2015, in Oklahoma City, OK, featuring multiple large supercells with tornadoes. The PHI tool is displayed with a U.S. map background showing cities and roads. The radar reflectivity layer is shown with orange lightning PHI objects overlaid. Shown in the inset (denoted by A), each PHI object has an associated three-digit number and a probability percentage.

updated, and managed by the forecaster. A forecaster was able to issue updates to the same PHI object as often as necessary. The resulting analysis output was the number of unique PHI objects the forecaster interacted with, time spent interacting with each object per update/issue, and the number of total PHI objects or updates issued. A process flow was developed for automated objects and manual PHI objects based on the top-down design of the PHI tool (Fig. 7). The PHI tool interface was designed to provide a workflow similar to current warning software used by forecasters. During the experiment, researchers were on hand to answer questions and troubleshoot issues with PHI tool, as well as make suggestions for PHI tool usage.

## 2) NASA-TLX MENTAL WORKLOAD INSTRUMENT

The NASA-TLX workload index is a questionnaire-based workload rating tool (Hart and Staveland 1988). The tool measures six subdimensions of mental workload: mental demand, physical demand, temporal demand, performance, effort, and frustration. Mental demand is how much mental activity a user requires to complete tasks, including thinking, decision-making,

remembering data, or completing calculations. Physical demand defines how much the user has to move the mouse, click, and bring up different displays. Temporal demand is how much time pressure the user felt to get tasks completed. Performance defines how successful the user felt they were at accomplishing goals or tasks. Effort gauged how hard they had to work to accomplish a level of performance. Frustration includes how stressed, annoyed, or irritated the user felt while completing a task. The prompts for each subdimension were tailored for the experiment.

The analysis of workload includes a weighting dimension used to calculate an overall workload score. The questionnaire was modified slightly by adding a question: "What made it so?" after each subdimension. Forecasters wrote optional text responses to provide further explanation as to what events or situations contributed to their workload score. The raw scores of the mental workload ranged from 0 to 100, with 0 indicating extremely low workload and 100 denoting extremely high. The ratings were averaged from all the archived sessions for each of the six subdimensions of workload. The pairwise comparison among subdimensions produced the importance factors, which were averaged for each subdimension.

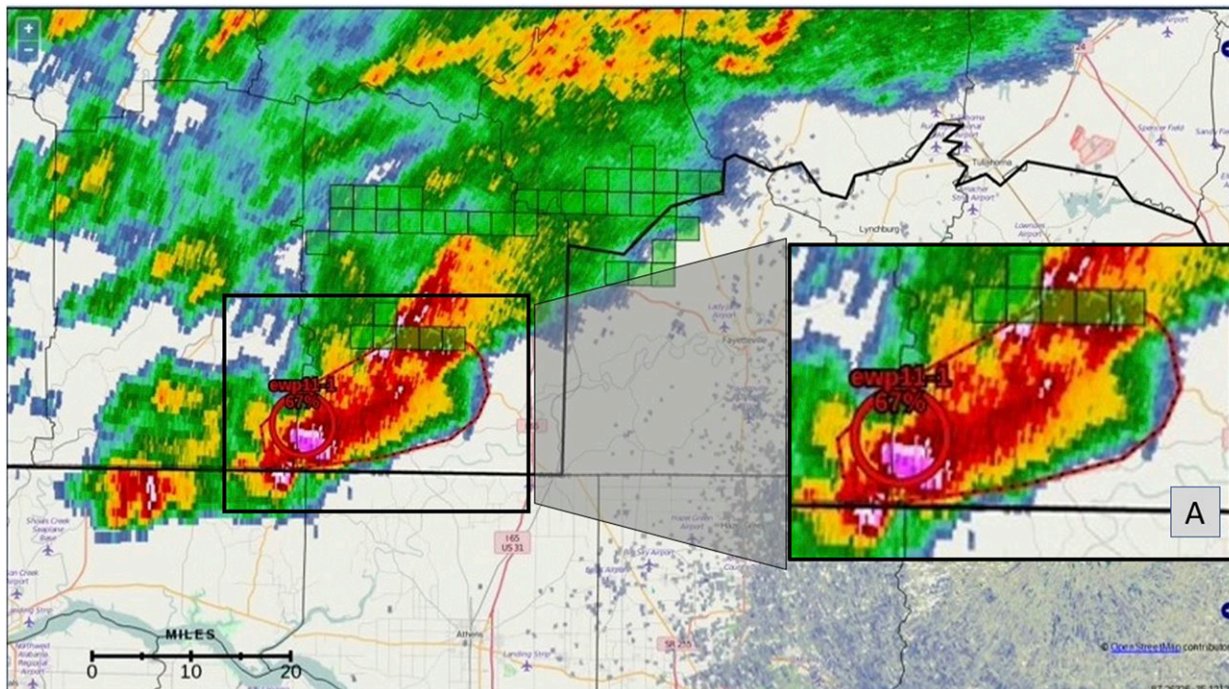


FIG. 5. Archived Case 2 was selected from 31 Mar 2016, in Huntsville, AL, featuring merging supercells with tornadoes. The PHI tool is displayed with a U.S. map background showing cities and roads. The radar reflectivity layer is shown with a red tornado PHI object overlaid. Shown in the inset (denoted by A), each PHI object has an associated identification number and a probability percentage.

### 3) THEMATIC ANALYSIS

Thematic analysis is an analysis method often used in psychology for reporting patterns and themes within sets of qualitative data (Braun and Clarke 2006; Guest et al. 2012; Boyatzis 1998). End of week discussion and NASA-TLX comment responses were analyzed for their qualitative content. Patterns and themes were then developed for topics of common concern among participants.

## 3. Results and analysis

### a. Task analysis

The task analysis results are summarized in Table 1. Task analysis was completed only on archived case. The “number of objects” is the average number of unique objects the forecasters “interacted with” during a single scenario. “Interacting with” means a forecaster made a decisive change in an object characteristic, probability trend graph, or text. The “number of updates” is the average total number of updates issued by the forecaster during each scenario. This includes all updates issued from the forecaster, from manually created or automatically created objects. The “average time per issue” is the average duration (in seconds) that the forecaster

interacted with an object prior to issuing an object or updating an object and clicking the “issue” button.

The results show lightning as having the greatest number of unique hazard objects per case, with an average of 14.5 objects and an average time to update of 189.2 s. Forecasters issued the least number of tornado objects with an average of 4.3 objects. However, tornado objects required an average of 302.5 s to update. The tornado objects were not created by automation and tornado objects were targeting a much smaller spatial area than severe thunderstorm or lightning objects. Instead of a polygon, forecasters were using circles and ellipses to target individual areas of rotation and tornadic development. Severe thunderstorm objects required the least amount of time to update, with approximately 163 s per update.

Data points that were repetitive or contained mistakes were removed. Sometimes a forecaster would select an option, and then almost immediately change the option to another value; for example when deciding wind speed or hail size. The time stamp on the latest occurring decision in a step is the time used to calculate a particular step duration. A detailed task analysis was carried out to account for all available decision steps (Fig. 7). The aggregated time steps were averaged for all PHI object creation and update instances with regard to hazard.



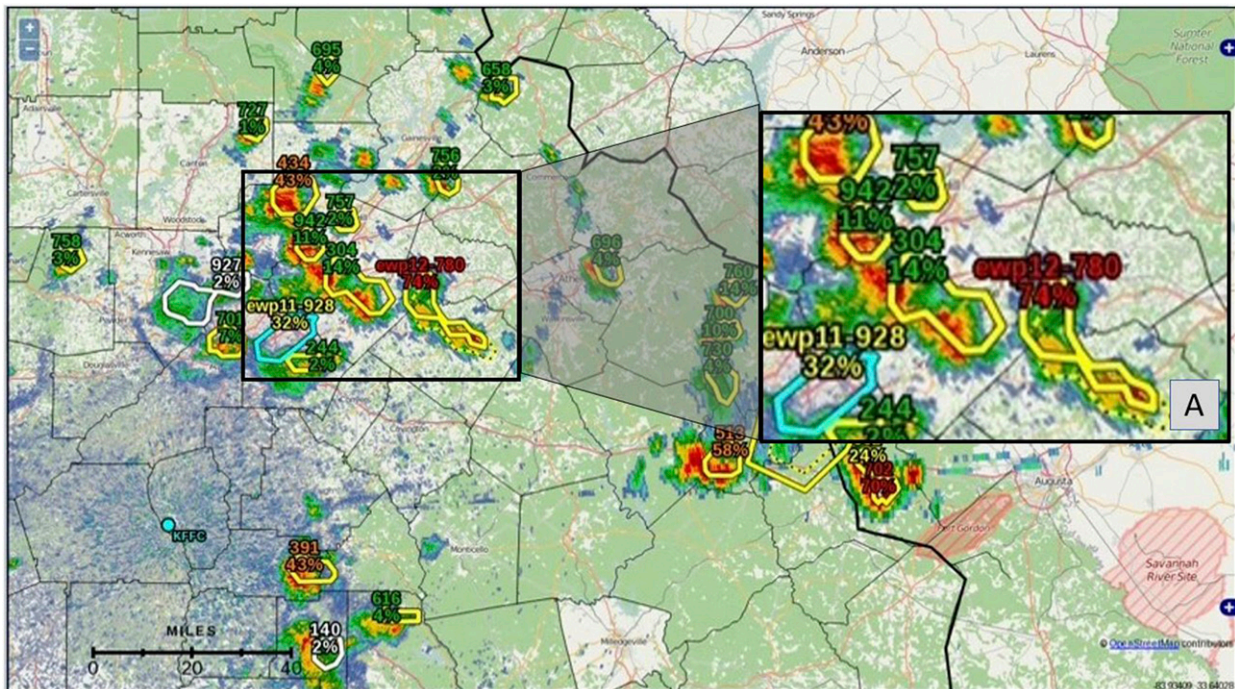


FIG. 6. Archived Case 3 was selected from 24 Jun 2015, in Atlanta, GA, featuring dispersed pulse storms. Radar reflectivity is shown displayed over a road and city map. Yellow severe thunderstorm PHI objects are shown along with a white PHI object (denoting a blocked object) and a blue PHI object, showing a suggested PHI object from the automated guidance. Shown in the inset (denoted by A), each PHI object has an associated identification number and a probability percentage.

Forecasters updated aspects and attributes of PHI objects as they determined to be necessary. If an attribute was not updated, the previous attribute parameter was carried over, so some instances of PHI object updates did not include or skipped some steps. These instances were not accounted in the average for that process step time. The average represents the average time of all instances of forecaster interaction with a particular action.

Severe thunderstorm and lightning follow a similar time step path (Fig. 8); however, tornado PHI objects took 100–150 s longer to produce. Drawing the probability trend and deciding action level for tornado objects each took longer than 70 s. The same two steps in producing lightning and severe thunderstorm information required less than 20 s for probability trend chart and less than 15 s for deciding action level.

In case debriefs, forecasters stated they could comfortably manage 4–5 PHI objects of tornado, lightning, or severe thunderstorm hazards at a time. In general, tornado and severe thunderstorm objects were within the range of manageable object numbers; however, there were significantly more objects capable of producing cloud-to-ground lightning than either severe or tornado. The average unique number of objects a forecaster interacted with during a case reflects the

difference, ranging from 4.3 objects with possible tornadoes to 7.3 severe thunderstorm objects and 14.5 objects with a lightning threat per case (Table 1). The automation created and managed more objects than the forecaster interacted with, forecasters prioritized interaction with objects with a higher probability of producing severe weather or cloud-to-ground lightning. The number of updates per PHI object varied, 2.9 for severe thunderstorm, 2.1 for tornado, and 1.5 for tornado. Forecasters would issue several updates on one PHI object over the course of the scenario or sometimes just one update depending on the weather development or forecaster workload, there was no specific guidance given to forecasters on how often or when to update objects.

#### b. Mental workload analysis by subdimensions

NASA-TLX data were analyzed using the average of the raw score and calculation of the importance of each of the six subdimensions of mental workload. The importance factor was calculated using 15 pairwise comparisons. Each subdimension was compared to each of the other subdimensions and overall importance was calculated for each subdimension. The mental workload level and relative importance factors are shown for each subdimension (Fig. 9).



| PHI process<br>PHI object selection options  | Severe<br>(sec)                                       | Tornado<br>(sec) | Lightning<br>(sec) |
|--|---|------------------|--------------------|
|  | <b>Motion/Duration</b><br>(Options: 3 boxes, <typed>) | 24.75            | 27.91              |
| <b>Trend Interpolation</b><br>(Options: Adjust Probability trend graph)  | 13.51   | 71.4             | 18.85              |
| <b>Prediction</b><br>(Options: Explicit, Persistence)  | 15.78   | n/a              | 15.72              |
| <b>Alert Level</b><br>(Options: Be aware, Take action)   | 12.6  | 70.75            | 13.68              |
| <b>Warning Threshold</b><br>(Options: 0-100 typed)   | 4.58  | 14.64            | 9.14               |
| <b>Severity</b><br>(Severe Options: Wind - 0 - 100 mph, Hail - 0 - 3 in)<br>(Tornado Options: none, weak, strong)<br>(Lightning Options: none, sparse, frequent) | Wind 8.77<br>Hail 11.05                               | 3.62             | 7.11               |
| <b>Source</b><br>(Options: none, radar indicated, observed, predicted)   | 9.67  | 9.71             | 6.00               |
| <b>Confidence</b><br>(Options: none, low, med-low, med, med-high, high)  | 12.73   | 16.33            | 12.14              |
| <b>Discussion</b><br>(Options: typed)  | 40.12   | 61.78            | 64.07              |

FIG. 7. 2016 PHI object creation/update matrix and average duration. The time for each decision step in seconds for each hazard (severe thunderstorm, tornado, and lightning) is shown. The workflow begins at the top of the chart with “Motion/Duration” and ends with “Discussion” at the bottom of the chart.

TABLE 1. Task analysis results of PHI objects.

| Hazard type                  | Severe<br>thunderstorm | Tornado | Lightning |
|------------------------------|------------------------|---------|-----------|
| No. of objects               | 7.33                   | 4.3     | 14.5      |
| No. of updates               | 21.5                   | 9       | 21.5      |
| Updates per object           | 2.9                    | 2.1     | 1.5       |
| Avg time per update (s)      | 163                    | 302.5   | 189.2     |
| Frequency of<br>update (min) | 5.5                    | 13.3    | 5.5       |

The mean mental workload was 49.9 [out of 100, standard deviation (std dev): 28.1], with a range of 70.8. The most mentioned contributing factors for increased workload in working with the PHI prototype tool included learning to use automated guidance, number of objects/storms to keep track of, multiple displays, and formulating probabilities. An increased standard deviation is reflective of a variety of severe weather event cases and variation in forecaster experience.

Summaries of each of the six subdimensions of mental workload are shown as follows:

- 1) MENTAL DEMAND: AVERAGE: 64.9, STD DEV: 25.6, RANGE: 90

Working with automation was cited as one of the major contributors to mental demand. Forecasters stated that the automation, for severe thunderstorm and

lightning products only, gave a good first guess or indication of an area to prioritize. Forecaster 6 stated, “The auto detected objects for severe potential helped in finding areas of concern and also were generally good with the motion and speed of the cells. This helped save time and thought.” The object tracking was generally good, which reduced workload. However, if the automation and tracking were not very accurate, adjusting tracking required higher mental demand because the forecaster would have to take over the object manually. This happened frequently during pulse storm situations. Forecasters also attributed increased mental demand to unfamiliarity with the new paradigm and PHI tool. Forecasters stated higher workload is normal during severe weather operations.

- 2) PHYSICAL DEMAND: AVERAGE: 58.6, STD DEV: 28, RANGE: 92

A significant amount of mouse clicking and moving between multiple monitors contributed to physical demand. The PHI tool required a lot of clicking and selecting many options to produce PHI. Forecasters also noted that workload was increased when there were larger numbers of objects and they had to click on each one to review the object details. Forecaster 6 explained his workload rating as, “Physical demand was not too bad. Tool was easy to use and manipulate for the most part. It wasn’t anymore (sic) difficult than conventional

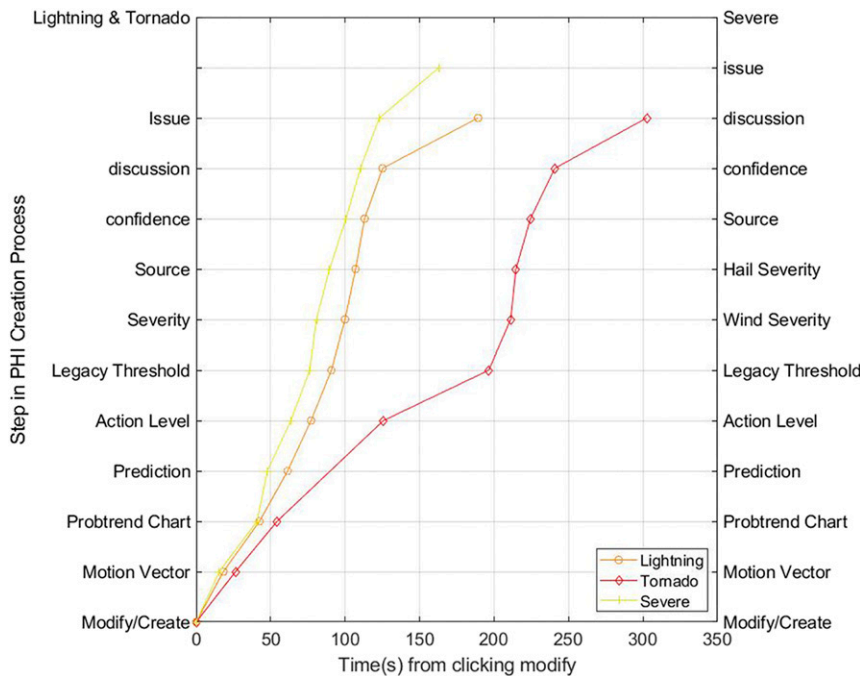


FIG. 8. Task-time analysis for respective hazard PHI object creation. Object is initiated at “modify,” with time equals zero and ends with the forecaster selecting “issue.”

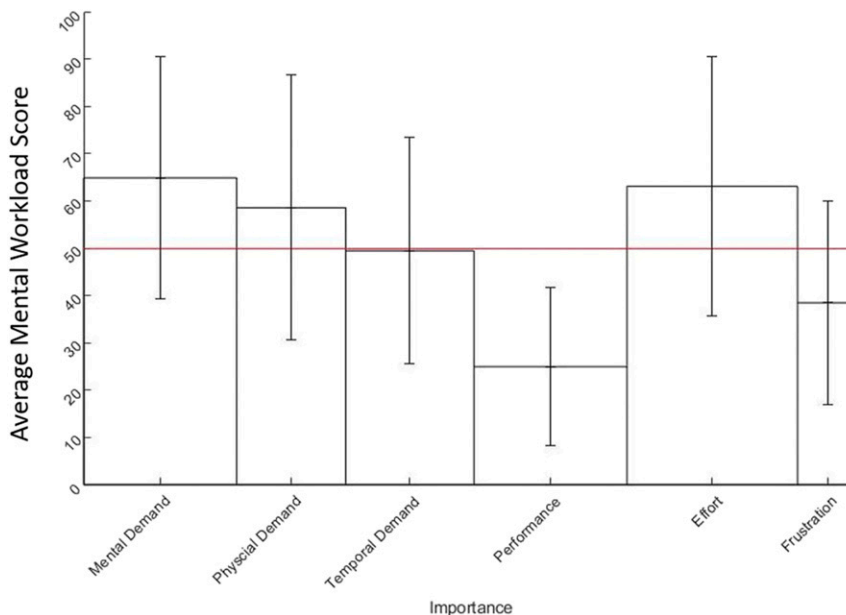


FIG. 9. 2016 PHI prototype experiment overall mental workload. The width of the bars indicates the importance of workload subdimension. The red line shows the overall average workload score for the experiment. Standard deviations are shown for each subdimension.

warning.” Another participant attributed the physical workload to adjusting the shape and motion of the PHI objects. Forecaster 7 commented that, “The prototype page requires a significant amount of mouse clicks. Trying to change a shape or adjust a motion can be a difficult task and occasionally requires one to start over.”

- 3) TEMPORAL DEMAND: AVERAGE: 49.5, STD DEV: 24, RANGE: 86

The number of objects monitored was the main contributing factor to temporal demand. Forecasters stated that the automation assisted in tracking and monitoring many different PHI objects. Forecaster 9 stated, “the pace was busy to brisk and did get a bit frantic at times as I tried to correlate incoming reports into the probs as multiple storms became tornadic around the same time.”

- 4) PERFORMANCE: AVERAGE: 25, STD DEV: 16.7, RANGE:70

The factors that contributed to performance were the realization of using the new PHI tool to forecast events accurately, address all important PHI objects, and provide increased lead time. (Note a lower rating in this subdimension corresponds to better performance.) Forecaster 6 commented, “I feel that, in the time allowed, that (sic) I was able to warn on most of the significant severe storms. The PHI automated shapes helped with this (showing areas of concern, severe

probabilities) and the relative ease and quickness of issuing the warnings. In a real event, if we had sectorized the CWA to allow me to focus on fewer storms, then I think I would have been able to put it as a 10 or 0.”

- 5) EFFORT: AVERAGE: 63.1, STD DEV: 27.3, RANGE: 91

The majority of effort was contributed to finding a location on a different display for the PHI tool and the radar display, due to a lack of geographic correspondence between displays. Other contributing factors were determining the probabilities and tracking of PHI objects. Forecaster 5 stated, “I felt my effort was high, primarily mental, to be situationally aware, interrogate storms, create warning in PHI, and to provide descriptive updates on why I was putting out the warnings.”

- 6) FRUSTRATION: AVERAGE: 38.5, STD DEV: 21.5, RANGE: 90

Frustration was caused by learning and using the new PHI tool. Forecaster 1 stated, “the frustration was not with accomplishing the tasks and making the decisions, it was with the whole shift in thinking and adjusting to a newer way of doing things in severe mode.” Forecasters felt especially frustrated when the automation did not produce the PHI object they wanted. When a forecaster manually took control of an object, they were frustrated that they had to maintain that object from that point forward.



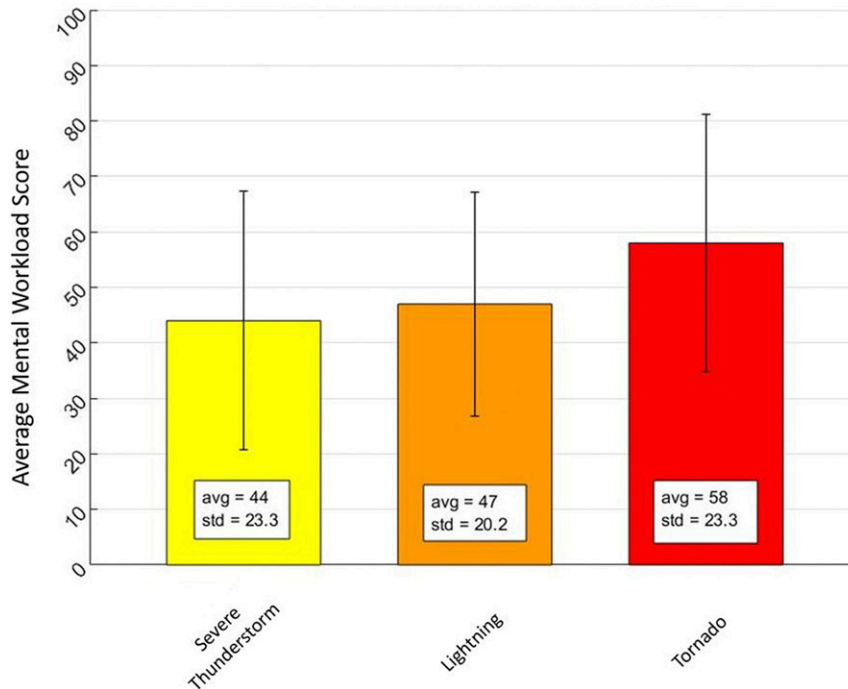


FIG. 10. Average mental workload by hazard type. The figure shows average workload for lightning, severe thunderstorm, and tornado. The whiskers denote standard deviation.

### c. Mental workload analysis by hazards

Figure 10 shows workloads for each hazard type. Severe thunderstorm and lightning hazards resulted in workload averages of 44 and 47, respectively. Average workload for tornado hazard was 58.

#### 1) TORNADO

The average mental workload for working on tornado PHI objects was 58 out of 100 (std dev: 23.3). No automation was implemented in the tornado PHI creation and management process. Forecasters attributed increased workload for tornado over other hazards to increased life-threatening impacts. Tornadoes could develop and touch down quickly and impact a smaller area, thereby requiring continuous monitoring. Therefore, forecasters had to be more diligent in interrogating the velocity products to produce short-term forecasts. The tornado PHI objects that forecasters were creating were often circles or ellipses, capturing the areas of rotation. The areas covered by the tornado PHI objects were much smaller than traditional tornado warnings areas, and often targeted individual velocity couplets.

#### 2) SEVERE THUNDERSTORM

The average mental workload for working on severe thunderstorm PHI objects was 44 out of 100 (std dev: 23.2). Severe thunderstorm PHI had automated objects

based on the ProbSevere algorithm. Severe storms did not have the same critical impacts as tornadoes. However, hail and high winds were still a concern. Severe storm development was more apparent and large cells were easier to interrogate for producing longer-term forecasts, in the 30–90-min range.

#### 3) LIGHTNING

The average mental workload for working on lightning PHI objects was 47 out of 100 (std dev: 20.2). Lightning PHI objects used trained algorithms to create automated objects. Lightning introduced a new threat for forecasters to produce PHI objects because forecasters do not issue warnings for lightning. Other challenges included situations where cloud-to-ground (CG) lightning was not always contained within automated objects, and it was unclear to the forecaster if the goal was to contain every single CG strike or contain a percentage of CG strikes within an object.

### d. Mental workload analysis by cases

#### 1) CASE 1: MULTIPLE LARGE SUPERCELLS WITH TORNADOES

The average mental workload for Case 1 (Fig. 11) was 39.2 (std dev: 27.5) for severe thunderstorms, 72.1 (std dev: 18.4) for tornado, and 57.5 (std dev: 26.7) for

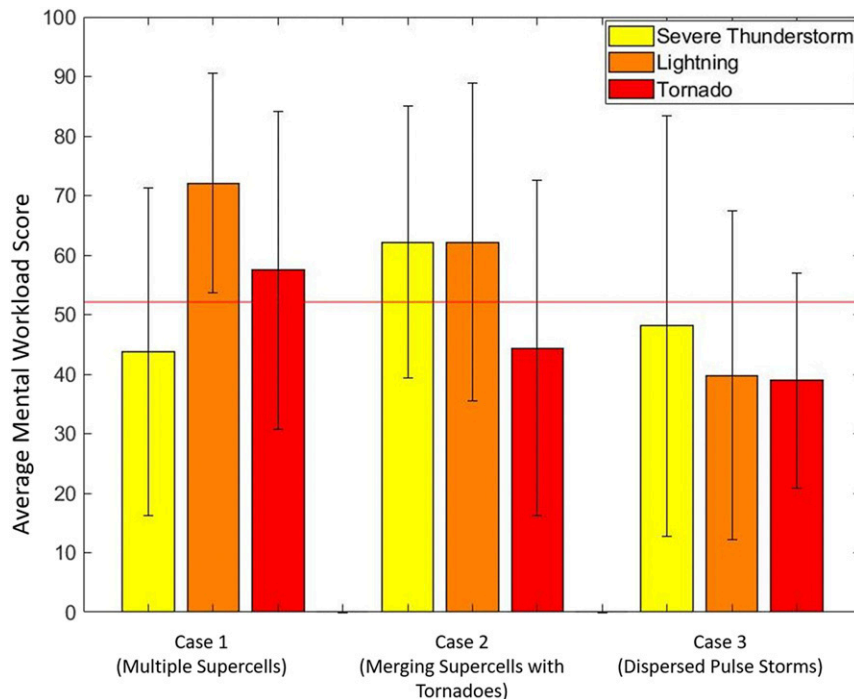


FIG. 11. Mental workload by hazard type and case. Whiskers show standard deviation.

lightning (Table 2). Forecasters were challenged to interrogate and assess probability and threat attributes for multiple large supercells with associated tornado threats. Forecasters were comfortable increasing initial probabilities higher than the automated guidance because the storms were well developed and likely to continue. Forecasters were unsure how to determine future development of probability over time. This led to frequent updates of the probability trend graph. Forecasters often used a bell curve probability trend prediction for future probabilities (the default probability trend is a linear decrease from the current ProbSevere probability).

## 2) CASE 2: MERGING SUPERCELLS WITH TORNADOES

The average mental workload for Case 2 (Fig. 11) was 44.8 (std dev: 22.9) for severe thunderstorms, 62.2 (std dev: 26.7) for tornado, and 57.5 (std dev: 28.1) for lightning (Table 2). Forecasters were challenged with a large developed supercell with embedded tornadoes and supercells merging with other storms. The merging of supercells required forecasters to maintain awareness of the automation. When the automation merged two storm objects of the same type, the automation would either keep only one object number or create a new object number. There was no notification when this would happen. A forecaster could be working in another

storm area and come back to check up on the objects they previously worked on and found that the object had been merged into a new object. This situation was very frustrating for forecasters, especially when they had adjusted or modified a specific object. As a result, a blue suggestion object, as seen in Fig. 12, was developed to inform forecasters that the automation wanted to merge or in some cases split an object and the blue object would represent the proposed action. Forecasters could then accept or block the suggested blue object. If the forecaster accepted the suggested object, they could link it to the previous object that they worked on to carry over the threat information and storm history.

## 3) CASE 3: DISPERSED PULSE STORMS

The average mental workload for Case 3 (Fig. 11) was 48.1 (std dev: 35.3) for severe thunderstorms, 39.8 (std dev: 27.6) for tornado, and 38.9 (std dev: 18.0) for lightning (Table 2). The case challenged forecasters with widespread, pulse thunderstorms where small storms rapidly developed and were capable of producing cloud-to-ground lightning and strong winds associated with microbursts before quickly dissipating. The tornado threat was not a critical feature of this weather scenario. After the initial triage, forecasters had to frequently retriage to pick up any rapidly developing storms. Forecasters used traditional radar interrogation methods, but also used the automated probabilities from ProbSevere

TABLE 2. Mental workload by hazard and case (out of 100).

|   | Severe thunderstorm | Tornado | Lightning |
|---|---------------------|---------|-----------|
| Case 1 (multiple large supercells with tornado) | 39.2                | 72.1    | 57.5      |
| Case 2 (merging supercells with tornado)        | 44.8                | 62.2    | 44.4      |
| Case 3 (dispersed pulse severe thunderstorms)   | 48.1                | 39.8    | 38.9      |

and ProbLightning to triage storms. Forecasters could quickly scan the map and see if any storms had increased in probability and needed attention. Due to the short-lived nature of storms, forecasters often had to adjust object duration and end objects early.

#### 4. Discussion

##### a. Mental workload

The overall mental workload from the NASA-TLX was an average of 49.9 (out of 100) with higher workload assessed for the subdimension of mental demand, 64.9, and effort, 63.1. The average workload was not excessively high to cause concern, as it is similar to the average mental workload based on a meta-analysis of mental workload ratings reported in 237 publications (Grier 2015). Forecasters had not used PHI prior to the testbed, thus some of the higher workload could be attributed to unfamiliarity with the process.

Mental demand was one of the workload subdimensions that resulted in higher overall mental workload. When the forecasters took over the objects completely, the number

of tasks they had to complete to issue an update to PHI increased. Increasing the reliability of automation could reduce the mental workload level during severe weather operations. Forecasters stated they normally experience higher mental workload during severe weather event.

Effort also contributed significantly to overall mental workload, mainly due to interacting with AWIPS II on one screen and the PHI tool on the other screen. This amount of workload was due to a limitation in the experimental setup. This result does lend some insight into tracking similar locations across multiple displays or software windows. A design recommendation would be cohesive location identification across displays or a software link to quickly correlate locations. This would help provide a truer testing environment in future software development.

Mental workload associated with tornado hazard was more than 10 points higher than those of lightning or severe thunderstorm hazards. The tornado PHI objects were much smaller geographically and there was no tornado automation available. Forecasters carefully examined areas of tornadic potential and used ellipse or

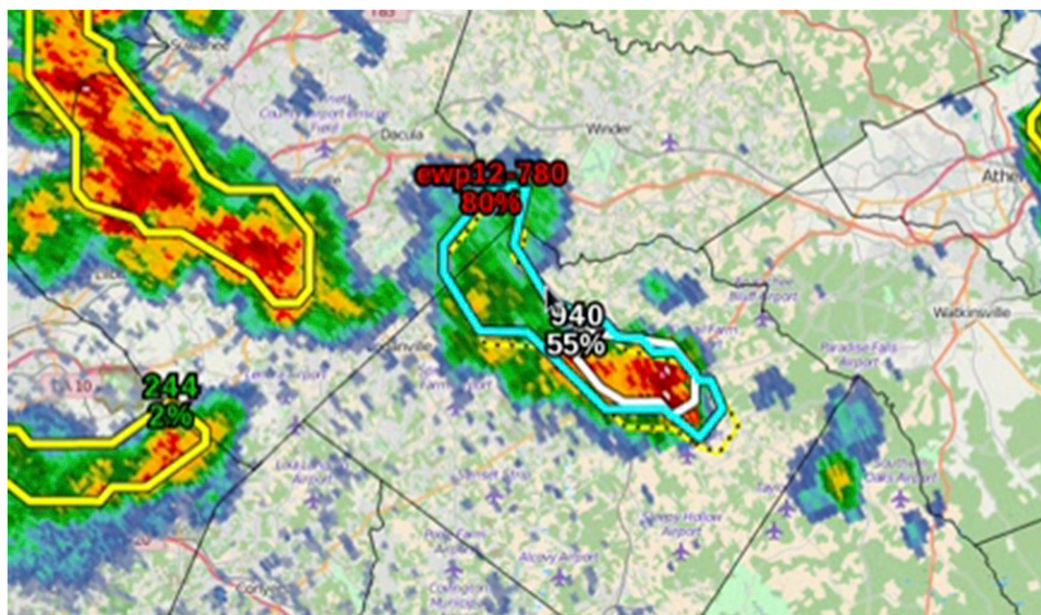


FIG. 12. Automated blue suggestion object for merging or splitting storm objects. This provides feedback to users when the automation suggests a merge or split of an object a forecaster has edited.



circle drawing tools to create tornado PHI objects. The sizes of all the PHI objects varied, but generally tornado objects were less than half the size of severe objects; depicting a more spatially specific objects may lead to higher workload.

#### *b. Formulating PHI in the new paradigm*

Forecasters discussed that formulating PHI in the new paradigm was a contributing factor to mental workload. According to forecasters' debriefs, the new PHI paradigm involved an increased level of interrogation of individual storms and increased situational awareness of all storms. Maintaining PHI objects required frequent updating and monitoring. Based on storm analysis, assigning a probability was a challenging aspect of the new paradigm. With the current static and deterministic weather warning system, forecasters have extensive experience making the binary decision to warn or not warn. The new PHI tool challenged forecasters to think of communicating threat risk in terms of probability. This challenge involves more than just assigning a single probability, but also shows how that probability will evolve over the duration of the PHI threat object. Additionally, this new paradigm allows forecasters to start communicating threat information to end users even when a hazard is subsevere and is not at a threat level for which they would normally produce a warning. This ability commanded more awareness of weather development before a storm reached severe criteria.

#### *c. Use of automated guidance*

Forecasters rely on first guess guidance from automated objects to initiate PHI objects, the majority of the time. Only 5% of the time forecasters created completely manual objects (this excludes tornado, which did not have automated guidance). Once a forecaster manually took over an aspect of automation, that aspect could not be returned to automation and the forecaster had to maintain it for the remainder of the scenario, or until the object expired or was ended by the forecaster. This was also true for manually created objects; once a forecaster created a manual object, they were responsible for all aspects of that object for the lifetime of the object. Forecasters also had the option to delete an object and allow the automation to replace it with an automatically generated object.

#### *d. Use of automation*

The automated objects for severe thunderstorm and lightning helped forecasters prioritize their interrogation. Forecasters stated in NASA-TLX discussions that they would focus on the automated objects with higher probabilities first and those objects were often the most

important storms. Their statements were supported by their frequent utilization of automated objects. Forecasters used automated objects the majority of the time to initiate objects and issue updates.

Forecasters found that, in some situations, the automated guidance had difficulty accurately tracking the motion and size of the storm. In these situations, forecasters were required to maintain a manual control of some objects. Manual override of automated objects provided users with greater control, but increased mental workload resulted because the user was required to maintain the object for the rest of its duration. A "return to automation" function was included in the following year's prototype design to relieve forecasters' hesitation to take over automated objects and reduce the mental workload of handling many low automation-level objects (Karstens et al. 2018).

Forecasters stated in debriefs that they could comfortably manage 4–5 objects at a time. Table 1 shows unique object numbers of up to 14 lightning hazard objects for a weather scenario. However, this does not mean forecasters were interacting with these objects continuously during the event. Forecasters updated some objects many times and some objects once. Working with multiple objects required efforts in maintaining situational awareness of these objects, which lead to increased mental workload. Meanwhile, frequent updates on objects also increased mental workload.

One of the issues that the forecasters identified was a lack of notification or feedback when the automation would split an object into two objects or merge two objects into one object, thus losing the history of forecaster discussions added to PHI objects. Forecaster 6 stated, "there were some bugs, such as issued objects being merged into auto objects, and thus disappearing." During week 2 of the experiment, researchers implemented a blue "suggested" object (Fig. 12). When the automation suggested a merge or split, forecasters could accept the suggestion and choose to link historical data or block the suggestion. This revision gave forecasters more control over the automated objects, and forecasters subsequently considered these changes favorably.

#### *e. Creating PHI objects*

Forecasters liked the ability to communicate information-rich warning products quickly via the PHI prototype software. The PHI paradigm allows accurate threat tracking and the ability to quickly provide updates. Forecasters would triage storms to identify the most significant storms and quickly issue a PHI object, then complete a more thorough analysis of storms with the possibility of these storms becoming severe in the near future.

Another change in the PHI paradigm is continuous update. There was no set update time; as soon as an

object was updated, a forecaster could update it again or move directly to another object. Traditionally, forecasters would issue a warning and provide any updates through issuing a “Special Weather Statement.” The PHI paradigm creates the opportunity for forecasters to update a PHI object as frequently as they determine necessary. The average frequency of update on any managed PHI objects was 5.5 min for severe thunderstorm, 13.5 min for tornado objects, and 5.5 min for lightning objects.

PHI attributes included a threat level as warning level or advisory and a probability threshold for a legacy warning to drop out. Forecasters spent considerable amount of time choosing between warning level, advisory level, and deciding legacy threshold, especially for tornado. These attributes required forecasters to make a “binary” warning decision along with providing the probabilistic forecast within the PHI paradigm. This challenge increased with the tornado threats, as tornado PHI objects were generally covered a much smaller geographic area than current tornado warning polygons. Additionally, the potential for loss of human life increased the importance of an accurate forecast. Forecasters determined a probability, as a percentage, for the PHI objects (current and future), and assigned a confidence level (low–high) for their forecast. The probability was a determination based on meteorological probability of storm impact, while confidence was a subjective evaluation of a forecasters’ confidence in the PHI objects. It was found that forecasters conflated the two factors, probability, and confidence, as representing the same or a similar measurement (Eastern Research Group 2016). If these two options conflicted (i.e., a low probability and a high confidence), decision-makers, such as emergency managers, were often confused regarding the risk of threat. In PHI experiments in later years, we decided to use forecaster confidence to represent the probability (Karstens et al. 2018).

The current deterministic warning paradigm lacks assigning probabilities of occurrence on the hazards. As such, training could be developed to calibrate forecasters to probabilities of hazards given certain radar or satellite-derived characteristics. Probability and probability trends could be mapped to certain kinds of storm characteristics to aid forecasters in deciding how to construct probability trends.

Forecasters often used the discussion box to describe reasons for increasing, decreasing, or changing threat probabilities. This often required reiterating threat attributes from previous object issuance and the new developments in selected attributes. This reiteration caused higher memory load and contributed to increases in mental workload. Later revisions of the tool in following years facilitated retrieval of previously issued PHI information.

## 5. Conclusions

The 2016 PHI prototype experiment was conducted as part of the FACETs project to advance NWS hazardous weather warning capabilities. Forecasters’ mental workload and task strategy were analyzed using the PHI prototype tool to communicate hazardous weather information in a time-sensitive warning environment with various hazardous weather situations. The PHI paradigm challenged forecasters in forecasting probabilities over duration of threat objects. Analysis showed that, in general, mental workload was manageable for forecasters with the help of automated guidance when working with multiple threats. Mental workload associated with tornado threat was found to be higher than workload for severe thunderstorm or lightning hazards. Mental demand and Effort were found to be the highest mental workload subdimensions. We propose that these workload issues can be alleviated through better design of human-automation interaction. System design limitations were found in lack of return to automation function and confusion between forecaster confidence and threat probability. Later revisions of the PHI prototype tool addressed several workload and usability issues to continuously improve design of the PHI tool within the FACETs framework (Karstens et al. 2018).

*Acknowledgments.* The authors thank many people in their support and development of the PHI prototype tool: Gabe Garfield, Darrel Kingfield, Amy McGovern, Kodi Nemunaitis-Berry, Holly Obermeier, Casandra Shivers, Shadya Sanders, Justin Sieglaff, Mike Pavolonis, James Hocker, Susan Jasko, Gina Eosco, Kim Klockow, Harold Brooks, Robert Hoffman, Israel Jirak, and Greg Stumpf. The authors also thank the participation of many NWS forecasters, emergency managers, and broadcast meteorologists for their valuable input, expertise, and feedback for further improvement. This study was supported by Grants NOAA-OAR-OWAQ-2015-2004230, OAR-USWRP-R2O FACETs PoW, and NAISNWS4680019.

## REFERENCES

- Akyeampong, J., S. Udoka, G. Caruso, and M. Bordegoni, 2014: Evaluation of hydraulic excavator human-machine interface concepts using NASA-TLX. *Int. J. Ind. Ergon.*, **44**, 374–382, <https://doi.org/10.1016/j.ergon.2013.12.002>.
- Boyatzis, R. E., 1998: *Transforming Qualitative Information: Thematic Analysis and Code Development*. Sage Publ., Inc., 184 pp.
- Braun, V., and V. Clarke, 2006: Using thematic analysis in psychology. *Qual. Res. Psychol.*, **3**, 77–101, <https://doi.org/10.1191/1478088706qp063oa>.
- Burigat, S., and L. Chittaro, 2016: Passive and active navigation of virtual environments vs. traditional printed evacuation maps: A comparative evaluation in the aviation domain.

- Int. J. Hum. Comput. Stud.*, **87**, 92–105, <https://doi.org/10.1016/j.ijhcs.2015.11.004>.
- Calhoun, K. M., and Coauthors, 2018: Cloud-to-ground lightning probabilities and warnings within an integrated warning team. *Special Symp. on Impact-Based Decision Support Services*, Austin, TX, Amer. Meteor. Soc., 4.4, <https://ams.confex.com/ams/98Annual/webprogram/Paper329888.html>.
- Cintineo, J. L., M. J. Pavolonis, J. M. Sieglaff, and A. K. Heidinger, 2013: Evolution of severe and nonsevere convection inferred from GOES-derived cloud properties. *J. Appl. Meteor. Climatol.*, **52**, 2009–2023, <https://doi.org/10.1175/JAMC-D-12-0330.1>.
- , —, —, and D. T. Lindsey, 2014: An empirical model for assessing the severe weather potential of developing convection. *Wea. Forecasting*, **29**, 639–653, <https://doi.org/10.1175/WAF-D-13-00113.1>.
- , and Coauthors, 2018: The NOAA/CIMSS ProbSevere model: Incorporation of total lightning and validation. *Wea. Forecasting*, **33**, 331–345, <https://doi.org/10.1175/WAF-D-17-0099.1>.
- Eastern Research Group, 2016: NWS hazard simplification project: Engagement at NOAA's 2016 Hazardous Weather Testbed to Collect Feedback on Prototypes Developed at the 2015 HazSimp Workshop. Eastern Research Group, Arlington, VA, Tech. Rep., 32 pp., [https://www.weather.gov/media/hazardsimplification/Final\\_HazSimp%20Testbed%20Report.pdf](https://www.weather.gov/media/hazardsimplification/Final_HazSimp%20Testbed%20Report.pdf).
- Finomore, V., T. Shaw, J. Warm, G. Matthews, and D. Boles, 2013: Viewing the workload of vigilance through the lenses of the NASA-TLX and the MRQ. *Hum. Factors*, **55**, 1044–1063, <https://doi.org/10.1177/0018720813484498>.
- Friday, E. W., 1994: The modernization and associated restructuring of the National Weather Service: An overview. *Bull. Amer. Meteor. Soc.*, **75**, 43–52, [https://doi.org/10.1175/1520-0477\(1994\)075<0043:TMAARO>2.0.CO;2](https://doi.org/10.1175/1520-0477(1994)075<0043:TMAARO>2.0.CO;2).
- Grier, R. A., 2015: How high is high? A meta-analysis of NASA-TLX global workload scores. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, **59**, 1727–1731, <https://doi.org/10.1177/1541931215591373>.
- Guest, G., K. M. MacQueen, and E. E. Namey, 2012: *Applied Thematic Analysis*. Sage Publishing, 320 pp.
- Hart, S., 2006: NASA-task load index (NASA-TLX); 20 years later. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, **50**, 904–908, <https://doi.org/10.1177/154193120605000909>.
- , and L. E. Staveland, 1988: Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Human Mental Workload*, P. A. Hancock and N. Meshkati, Eds., Advances in Psychology, Vol. 52, North-Holland, 139–183, [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9).
- Hoffman, R. R., S. V. Deal, S. Potter, and E. M. Roth, 2010: The practitioner's cycles, Part II: Solving envisioned world problems. *IEEE Intell. Syst.*, **25**, 6–11, <https://doi.org/10.1109/MIS.2010.89>.
- Hooley, B. L., D. B. Kaber, J. A. Adams, T. W. Fong, and B. F. Gore, 2018: The underpinnings of workload in unmanned vehicle systems. *IEEE Trans. Hum. Mach. Syst.*, **48**, 452–467, <https://doi.org/10.1109/THMS.2017.2759758>.
- Karstens, C. D., and Coauthors, 2015: Evaluation of a probabilistic forecasting methodology for severe convective weather in the 2014 Hazardous Weather Testbed. *Wea. Forecasting*, **30**, 1551–1570, <https://doi.org/10.1175/WAF-D-14-00163.1>.
- , and Coauthors, 2018: Development of a human-machine mix for forecasting severe convective events. *Wea. Forecasting*, **33**, 715–737, <https://doi.org/10.1175/WAF-D-17-0188.1>.
- Kuhlman, K. M., T. M. Smith, G. J. Stumpf, K. L. Ortega, and K. L. Manross, 2008: Experimental probabilistic hazard information in practice: Results from the 2008 EWP spring program. *24th Conf. on Severe Local Storms*, Savannah, GA, Amer. Meteor. Soc., 8A.2, <https://ams.confex.com/ams/pdfpapers/142027.pdf>.
- Meyer, T., K. M. Kuhlman, D. M. Kingfield, and D. J. Gagne II, 2016: Using random forest technique to create cloud-to-ground lightning probabilities. *28th Conf. on Severe Local Storms*, Portland, OR, Amer. Meteor. Soc., 146, <https://ams.confex.com/ams/28SLS/webprogram/Paper301841.html>.
- Meyer, W. B., 2003: Marlene Bradford: Scanning the skies: A history of tornado forecasting. *Isis*, **94**, 779–780, <https://doi.org/10.1086/386500>.
- Rothfus, L. P., C. Karstens, and D. Hilderband, 2014: Next-Generation Severe Weather Forecasting and Communication. Amer. Geophys. Union, accessed 23 February 2017, <https://eos.org/science-updates/next-generation-severe-weather-forecasting-communication>.
- , T. M. Smith, and C. D. Karstens, 2015: Forecasting a Continuum of Environmental Threats (FACETs): The science and strategic implementation plan for a watch/warning paradigm change. *Proc. Third Symp. on Building a Weather-Ready Nation: Enhancing Our Nation's Readiness, Responsiveness, and Resilience to High Impact Weather Events*, Phoenix, AZ, Amer. Meteor. Soc., 6.4, <https://ams.confex.com/ams/95Annual/webprogram/Paper266005.html>.
- Stanton, N., P. Salmon, and L. Rafferty, 2013: *Human Factors Methods: A Practical Guide for Engineering and Design*. Ashgate Publishing Ltd., 592 pp.
- Stern, A. D., 2020: National Weather Service Instruction 80-303.NOA/NWS/Department of Commerce, 59 pp., <http://www.nws.noaa.gov/directives/sym/pd0800303curr.pdf>.
- Stumpf, G. J., T. M. Smith, K. Manross, and D. L. Andra, 2008: The experimental warning program 2008 spring experiment at the NOAA Hazardous Weather Testbed. *24th Conf. on Severe Local Storms*, Savannah, GA, Amer. Meteor. Soc., 8A.1, [https://ams.confex.com/ams/24SLS/techprogram/paper\\_141712.htm](https://ams.confex.com/ams/24SLS/techprogram/paper_141712.htm).
- Trapsilawati, F., C. D. Wickens, X. Qu, and C.-H. Chen, 2016: Benefits of imperfect conflict resolution advisory aids for future air traffic control. *Hum. Factors*, **58**, 1007–1019, <https://doi.org/10.1177/0018720816665941>.
- Varouhakis, J., 2007: recordmydesktop. Accessed 3 March 2017, <http://recordmydesktop.sourceforge.net>.
- Walters, C., and P. J. Webb, 2017: Maximizing efficiency and reducing robotic surgery costs using the NASA task load index. *AORN J.*, **106**, 283–294, <https://doi.org/10.1016/j.aorn.2017.08.004>.
- Wickens, C. D., S. E. Gordon Becker, Y. Liu, and J. D. Lee, 2004: *An Introduction to Human Factors Engineering*. 2nd ed. Prentice Hall, 608 pp.
- Wolfe, J. P., 2014: An open source approach to communicating weather risks. *10th Free and Open Source (FOSS4G) Conf.*, Portland, OR, FOSS4G, Open Source Geospatial Foundation (OSGeo), <https://doi.org/10.5446/31617https://av.tib.eu/media/31617>.
- Yan, S., C. C. Tran, Y. Chen, K. Tan, and J. L. Habiyaemye, 2017: Effect of user interface layout on the operators' mental workload in emergency operating procedures in nuclear power plants. *Nucl. Eng. Des.*, **322**, 266–276, <https://doi.org/10.1016/j.nucengdes.2017.07.012>.