

# TOWARD A USER-CENTERED DESIGN OF A WEATHER FORECASTING DECISION-SUPPORT TOOL

ELIZABETH M. ARGYLE, JONATHAN J. GOURLEY, ZACHARY L. FLAMIG,  
TRACY HANSEN, AND KEVIN MANROSS

Incorporating usability evaluation into the iterative design of decision-support tools, such as Hazard Services, can improve system efficiency, effectiveness, and user experience.

In the weather enterprise, the forecaster's role is one that often requires managing large amounts of information under significant time pressures (Daipha 2015). Hazard Services is a software toolkit currently under development that is intended to streamline the forecasting process and assist forecasters in maintaining situation awareness. Currently in development, Hazard Services integrates existing forecasting functionality into a single user interface within the Advanced Weather Interactive Processing System II (AWIPS-II) display platform. Hazard Services is also intended to facilitate communication between decision-makers throughout the weather domain by allowing end users to share hazard-related information between forecast desks and offices.

While the initial interface design built upon other National Weather Service (NWS) tools in order to leverage user experience, Hazard Services introduces several new workflow processes. Forecasters may train to use new display systems, but the Hazard Services project affords unique opportunities to craft the interface to complement user needs. In software development, user-centered design (UCD) has great

potential to improve user acceptance and productivity (Buie and Murray 2012). UCD describes the general process of ensuring that a product matches end user needs (Abrams et al. 2004). In a UCD framework, iterative cycles of design and evaluation have the potential to resolve usability issues in product design.

Considering user requirements during a system's design stage not only promotes suitability for purpose, but it may also reduce project costs by minimizing the need for changes at the end of the project's time line (Nielsen 1992). One component of user-centered design is usability evaluation, which is most effective when conducted in iterative cycles of design and evaluation (Nielsen 1992). Usability is defined as "the extent to which a system, product, or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specific context of use" (ISO 2017). Although usability is often associated with the concept of intuitiveness, it also encompasses learnability, memorability, user experience, efficiency of use, and low error-rate measures (Holzinger 2005; Nielsen 1992). Frequently, a trade-off must occur

when selecting which aspects of usability are critical to the product's success.

Evaluations provide feedback on aspects including the degree to which a novice can learn to use the product, how quickly users can complete tasks, the number and types of errors that users make using the system, and the overall user experience (U.S. DHHS 2016). Usability evaluation often reveals issues throughout the design process, thereby improving the system's effectiveness and user acceptance (Holzinger 2005). Evaluation methods range from informal to formal and from automatic to empirical and may use one to several evaluators and novice to expert users (Nielsen 1994b). Opinions differ regarding timing of evaluations; Holzinger (2005) suggests that observational methods, like cognitive walk-throughs and activity analysis, have utility throughout the development process, but Norman (2006) argues that such techniques provide only value prior to the initial design stage. Heuristic evaluation, a technique in which usability is assessed against design recommendations, is known to find the greatest number of usability issues; this method requires that participants have user interface design expertise (Jeffries et al. 1991). After prototype development, evaluation and testing methods allow developers to assess components of the interface in order to update the original design.

In the public sector, usability evaluation has been applied to website design and information management software (Buie and Murray 2012; Mintmire et al. 2013). Several scholars have applied usability methods

to the user-centered design of weather, climate, and forecasting decision-support systems (Oakley and Daudert 2016; Timofeyeva-Livezey et al. 2015; Ling et al. 2015). Oakley and Daudert (2016) used a usability evaluation methodology to inform the design of a climate information website, resulting in a design more targeted to the needs of its end users. Similarly, Ling et al. (2015) used a task-and-questionnaire-based approach to compare mental workload and usability between Warning Generation software (WarnGen) and an alternative prototype. In the present work, we applied a user testing methodology to evaluate Hazard Services. Analysis of error rates, task response times, and questionnaire responses supported recommendations for user interface development. Through illustrating a research approach for usability assessment, we also present lessons learned that may inform the design of future meteorological decision-support systems.

**HAZARD SERVICES.** Currently, NWS forecasters use three different applications for warning on hazardous weather: WarnGen for short-fused hazards, such as tornadoes and thunderstorms; Graphical Hazard Generator (GHG) for longer-fused hazards, such as winter weather and hurricanes; and RiverPro for river flood events (Raytheon 2016). Each application has a unique interface, menus, and process that forecasters must learn and occasionally juggle during a shift. The Hazard Services system merges these disparate applications into one common interface and workflow.

The effort to unify the AWIPS hazard-generation applications began in 2004, involving a number of NWS discussions and workshops to refine the concept. In 2009, developers of the Earth System Research Laboratory's Global Systems Division (GSD) joined the effort and, in a kickoff workshop involving all stakeholders (forecasters, partners, social scientists, and software developers), designed a prototype user interface. The interface is implemented in a web-based format to allow for rapid cycles of user feedback and refinement. This iterative process continued until 2011, when work began to transition the application to AWIPS-II. In 2012, Raytheon contractors joined the effort and while stakeholder roles have fluctuated, iterative user feedback continues to be an integral part of the process.

To create the interface, developers analyzed the legacy applications and abstracted the common workflow for identifying a hazard area and time frame and specifying additional hazard characteristics. The system provides a user-extensible framework for ingesting models and guidance while applying algorithms to produce a first-guess indication of an impending hazard. Forecasters record relevant attributes and

**AFFILIATIONS:** ARGYLE\*—Cooperative Institute for Mesoscale Meteorological Studies, and NOAA/National Severe Storms Laboratory, and School of Industrial and Systems Engineering, University of Oklahoma, Norman, Oklahoma; GOURLEY—NOAA/National Severe Storms Laboratory, and School of Meteorology, University of Oklahoma, Norman, Oklahoma; FLAMIG—Cooperative Institute for Mesoscale Meteorological Studies, and NOAA/National Severe Storms Laboratory, and School of Meteorology, University of Oklahoma, Norman, Oklahoma; HANSEN—NOAA/Global Systems Division, Boulder, Colorado; MANROSS—CIARA, Colorado State University, Fort Collins, and NOAA/Global Systems Division, Boulder, Colorado

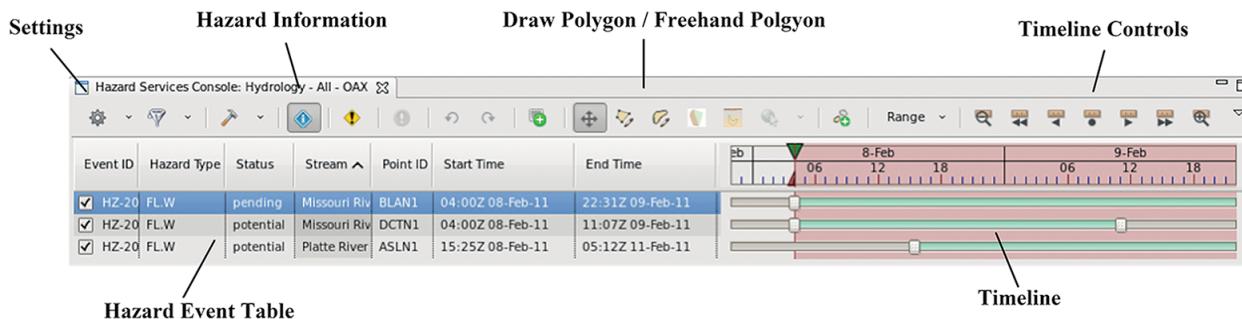
\* **CURRENT AFFILIATION:** Institute for Aerospace Technology, University of Nottingham, Nottingham, United Kingdom

**CORRESPONDING AUTHOR E-MAIL:** Jonathan Gourley, [jj.gourley@noaa.gov](mailto:jj.gourley@noaa.gov)

*The abstract for this article can be found in this issue, following the table of contents.*

DOI:10.1175/BAMS-D-16-0031.1

In final form 20 July 2016  
©2017 American Meteorological Society



**FIG. 1. The original Hazard Services console.**

then this information is transformed into actionable messages for partners, including emergency managers, broadcasters, and the public. An additional goal of Hazard Services is to allow forecasters to communicate threat information beyond what is currently supported in the legacy text products, moving toward using multiple forms of communication. Work continues toward an operational capability.

The user interface, shown in Fig. 1, is integrated directly into the AWIPS-II visualization platform. The primary interface, or “console,” displays information related to individual hazard alerts (hereafter, hazards), including start time, end time, hazard type, and a number of user-customizable metadata. In addition, the console allows users to view hazards (e.g., watches, warnings, advisories) in a timeline format, which allows users to maintain awareness of hazard status. Finally, a row of interactive icons, located at the top of the console, allows users to create hazards, alter them, and manipulate aspects of the display.

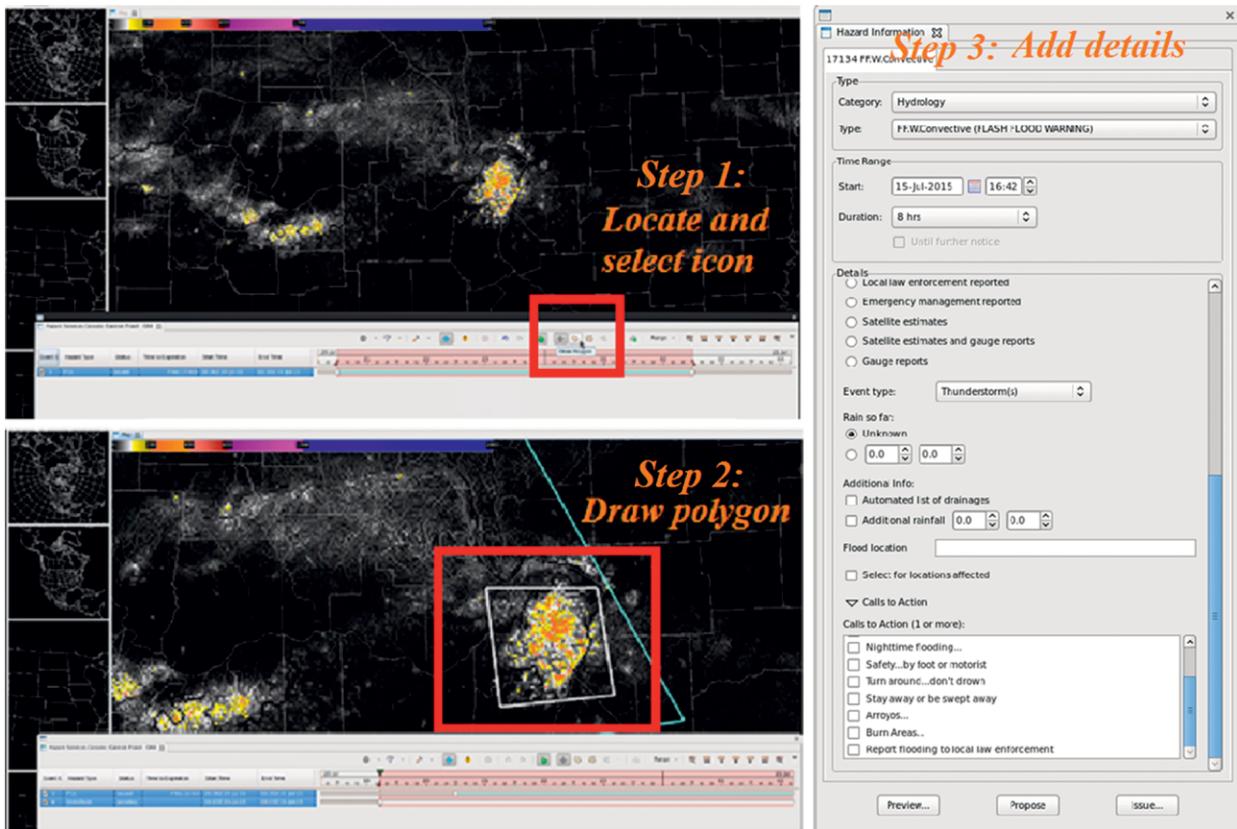
**METHOD. Participants.** Eighteen NWS forecasters participated in the Hazard Services evaluation. Some practitioners have found that five participants can

uncover 80% of usability errors (Virzi 1992). However, for quantitative usability studies, a larger sample size can increase statistical confidence in the findings (Nielsen 1994a; Faulkner 2003).

Participants were recruited from a population of forecasters participating in the Hydrometeorological Testbed–Hydrology in July 2015 (Martinaitis et al. 2017). Participants had between 2.5 and 25 years of professional forecasting experience (mean  $\mu = 11.75$ , standard deviation  $\sigma = 8.01$ ), and each had used the AWIPS-II system. At the time of the study, Hazard Services was in development. Thus, none of the participants had used it prior to the evaluation, and no training was provided prior to the study. We considered this to be a positive aspect, as it allowed us to assess usability as experienced by novice users. Yen and Bakken (2009) found that while UCD experts found errors related to interface design, subject matter experts were more likely to identify issues affecting task performance.

**Task selection.** Phrased as challenges, the tasks were presented individually to each participant via a web-based interface. The tasks, described in Table 1, included entering data, issuing flash flood watches

TABLE 1. Analysis of error rates and response time per task.				
Task No.	Description	No. of errors	Mean (std dev) of response time (min)	Sample size
1	Open hazard services	17	0.67 (0.38)	18
2	Add two columns to the workflow console	61	1.95 (1.13)	18
3	Draw and issue a hazard (watch 1)	56	6.37 (4.35)	18
4	Draw and issue a hazard (warning 1)	46	4.60 (2.35)	16
5	Draw and issue a hazard (warning 2)	9	5.50 (9.10)	15
6	Extend area of warning 1	14	1.32 (0.78)	11
7	Reduce area of warning 1	13	1.87 (2.57)	8
8	Update duration of watch 1	10	1.47 (1.62)	9
9	Update duration of warning 1	9	1.40 (0.83)	8
10	End warning 2	1	0.33 (0.18)	9



**FIG. 2.** The three steps involved in creating and issuing a hazard (tasks 3–5).

and warnings, editing existing watches and warnings, and canceling warnings. Tasks were selected through discussions with forecasters and Hazard Services developers. After reviewing the interface and discussing workflow processes with members of the development team, we selected 10 tasks representative of how a user might use the system in an operational setting.

**Study materials.** Participants received a static hydrologic visualization, displayed in AWIPS-II, to guide their work; an example is found in Fig. 2. Once each participant activated Hazard Services, the console appeared on screen below the unit streamflow visualization. Note that the unit streamflow product is a component of the Flooded Locations and Simulated Hydrographs (FLASH) project that is introducing new tools for flash flood forecasting across the United States (Gourley et al. 2017). A website interface displayed task instructions individually, and a desktop recording software captured videos of participant interactions with Hazard Services.

Screen-recording software captured user error rates, types, and task completion times. In addition to error-rate and task completion time measures, an adapted Questionnaire for User Interaction Satisfaction, version 7.0 (QUIS), assessed the system's

user interface (Chin et al. 1988). The QUIS was selected over standardized instruments, such as the System Usability Scale (SUS; Brooke 1996) and the Post-Study System Usability Questionnaire (PSSUQ; Lewis 2002) because of its ability to collect user opinions on highly specific design elements.

There were 58 questions across five categories: screen design, system transparency, learnability, multimedia, and system capabilities. Each question presented a phrase alongside a semantic differential, where 1 was associated with a negative response and 9 was associated with a positive response (Osgood and Luria 1954). In one question, participants declared their overall impression of the system and asked them to select a score where 1 was associated with "Frustrating" and 9 was associated with "Satisfying" (Chin et al. 1988). Additionally, the questionnaire requested subjective open-ended feedback for each design aspect.

**Procedure.** We evaluated the Hazard Services interface with a moderated multiple-user simultaneous testing (MUST) approach to user testing, which assesses system usability by measuring performance outcomes while participants use the interface by completing one or more tasks (Jokela et al. 2006; Nielsen 1994b, 2007).

User testing produces quantitative outcomes, such as error rates and task response times (Abrams et al. 2004). The MUST framework incorporates several collocated participants who work independently on user testing tasks and is recommended for quantitative evaluations constrained by time (Nielsen 2007). This approach was chosen because of the time constraints imposed by the co-participation in the simultaneous test bed study; the test bed method incorporated Hazard Services into the watch and warning process [see Martinaitis et al. (2017) for more information]. As the evaluation sought feedback from novice users, this necessitated the user testing to occur before any formal training or practice occurred.

All tests occurred within the Hazardous Weather Testbed at the National Weather Center. Participants worked at individual workstations spaced approximately 3–6 ft (1–2 m) apart, and each MUST session included six users. Researchers introduced the study and explained Hazard Services' purpose. Participants completed an informed consent form approved by the University of Oklahoma's Institutional Review Board. Researchers then instructed the participants to open the list of tasks and complete them. At any given point, one moderator and up to three additional observers took note of participant interactions with the software. Although some moderated usability evaluations ask the participants to speak about their thought processes during the study, this practice is not recommended when using the MUST approach, as it can bias the responses of nearby participants. However, if a participant became confounded by a particular task or encountered a catastrophic system failure, the moderator privately discussed it with the participant and assisted with resolving the issue so the participant could proceed to the next task. After finishing all the tasks, participants completed the QUIS.

**RESULTS.** *Error rates and task times.* Observational data revealed several common errors that participants encountered. Errors were defined as any action that deviated from a set of actions that would successfully accomplish the task. For example, the first task—launching Hazard Services within AWIPS-II—could be accomplished only by clicking the yellow button at the top-right corner of the interface. Here, errors included actions such as clicking the File menu in AWIPS-II or by launching WarnGen instead. Response times, also captured from the screen recordings, were measured from the time the task description was first displayed on screen to the point that the participant had accomplished the task. In forecasting, reducing response time for common tasks can lead to

reductions in lead time throughout the entire warning decision process. Error rates and mean response times for the tasks are presented in Table 1.

Task 2, customizing the console to include columns for valid time event codes (VTEC) and issue time, resulted in the greatest number of errors ( $n = 61$ ). Of these, 12 unique error types included clicking an incorrect icon, searching within the AWIPS-II menus for solutions instead of within the Hazard Services menus, and closing Hazard Services. To customize the console, users could right click the column header, triggering a drop-down menu from which users could then select new column types, or open a control window through the settings icon. Participants used both methods without guidance, though the majority used the former. Participant feedback revealed that this method was most frequently used because of its similarity to traditional spreadsheet-based interactions.

The three hazard issuance tasks (tasks 3–5) required similar interactions and assessed whether skill improved with practice. A comparison of tasks 3 (issuing a watch), 4 (issuing warning 1), and 5 (issuing warning 2) revealed a positive learning curve. The watch and warning issuance tasks involved three steps: 1) reading the instructions and clicking the Draw Polygon or Draw Freehand Polygon icon, 2) drawing the polygon, and 3) entering hazard information in the associated dialog window and clicking the Issue Hazard button; an example of this process is shown in Fig. 2. Task 3 was the first opportunity participants had to explore the hazard creation process, so not surprisingly, the task saw a high number of errors ( $n = 56$ ), primarily in the first step. Errors ranged from reversible, such as selecting “watch” instead of “warning” in the hazard information box or clicking an incorrect button, to irreversible, such as inadvertently issuing a colleague's warning or the catastrophic system failure that caused the entire AWIPS-II platform to crash.

Participants completed basic hazard issuance successfully with minimal practice, as reflected by the reduction in the mean response time and error count between tasks 3 and 5. Nevertheless, participants made several errors even with practice ( $n = 9$ ), but several of these may be attributable to individual circumstances. The two errors identified in the first step of task 5 appeared to be memory related; these errors involved looking for a function in the wrong place despite performing the identical previous task correctly. In the third step of task 5, participants made several data entry mistakes and accidentally issued more than one hazard due to a multiple selection in the console.

For the remainder of the tasks, performance generally improved. Error rates decreased as participants

**TABLE 2. The five highest- and lowest-scoring aspects of Hazard Services, assessed with the QUIS (lower scores indicate negative perceptions).**

	Category	Design aspect	Mean (std dev)
Highest scores	System capabilities	Information display rate	7.80 (1.10)
	System capabilities	System processing speed	7.67 (1.51)
	System capabilities	System response time	7.33 (1.51)
	Terms and system information	Terminology used for basic tasks (consistency)	6.67 (2.34)
	Multimedia	Color scheme adequacy	6.60 (1.95)
Lowest scores	Terms and system information	Ability to control feedback level	3.33 (1.03)
	System capabilities	System failure frequency	3.17 (1.72)
	General reactions	Overall reaction during use (ranging from frustrated to satisfied)	2.83 (1.47)
	Terms and system information	System transparency during operations	2.67 (1.37)
	Terms and system information	Clarity of error messages	2.33 (1.51)

became more familiar exploring the interface. Nevertheless, as shown in Table 1, sample size decreased throughout the study because of system performance and data quality factors. Several participants experienced catastrophic system failure, which required a workstation reset, resulting in an incomplete testing session. Others completed all 10 tasks, but they were removed in post hoc analysis because of issues with screen-recording legibility.

**Questionnaire results.** Following task completion, participants completed the QUIS to provide feedback on the design of the Hazard Services interface. Responses to the 58 questions were then analyzed by taking the mean and standard deviation of the scores. A collection of the highest- and lowest-scoring design aspects is presented in Table 2.

Three of the five highest-scored items emerged from the system capabilities section. Participants were generally pleased with the rapidity of the system’s response to their interactions (system response time). Participants felt similarly regarding the software’s processing speed, or system speed. In operational forecasting, display systems must be able to update at speeds matching the frequency of environmental and atmospheric model updates and observations. However, participants experienced system failures that catastrophically affected usability. Participants frequently had to restart the system in order to complete their tasks; in operations, this could significantly impact lead time during severe events.

Participants also indicated that elements related to terms and system information could benefit from further attention. The lowest score related to error message clarity. Participants found that instructions for correcting errors were more often confusing than

clear and were somewhat unhelpful. Furthermore, participants found that the internal system had poor transparency—the degree to which a user could tell what the system was doing internally. However, participants generally agreed that the length of delay during operational processing was acceptable ( $\mu = 6.67$ ,  $\sigma = 1.63$ ). While error messages require improvement, participants also responded that terminology used in dialog boxes and instructional labels were appropriately consistent throughout the different interface components ( $\mu = 6.67$ ,  $\sigma = 2.34$ ).

The QUIS revealed that learning to use Hazard Services by trial and error ranged in challenge level. The process of completing basic alerting tasks was often sequenced in a logical manner ( $\mu = 5.80$ ,  $\sigma = 0.84$ ). Based on several of the midrange scores, it was determined that the Hazard Services interface and workflow adequately corresponded to preexisting systems’ workflows. Participants reported that in the span of the 10 tasks, with no prior experience, the time it took to learn the system was neither too slow nor too fast ( $\mu = 4.50$ ,  $\sigma = 1.38$ ).

The screen design section addressed layout, legibility of characters and graphics, screen sequencing, and ease of maneuvering between windows. Participants felt that information was logically arranged within the AWIPS-II display and Hazard Services windows ( $\mu = 6.50$ ,  $\sigma = 1.22$ ) and users were not overloaded with information ( $\mu = 6.50$ ,  $\sigma = 1.52$ ). Still, participants assigned a lower score to the predictability of screen sequencing ( $\mu = 4.50$ ,  $\sigma = 1.52$ ), which indicated a departure from previous hazard management systems. One area that could lead to improvements in usability was the means to progress through work-related tasks, which participants responded was not clearly marked ( $\mu = 4.67$ ,  $\sigma = 2.34$ ). From a UCD perspective, using

design features to guide the user through work-related tasks can help to reduce limitations of workload and human memory (Krug 2000).

**Subjective feedback.** The QUIS also collected open-ended feedback related to each design section within the questionnaire, which produced actionable information for interface design. Participants largely felt that the layout of the interface and data entry windows was appropriate for achieving their goals. Participants commented that drawing polygons and issuing warning text was straightforward; this corroborates the response time analysis. Several participants stated that the polygon creation interactions were similar to those in previous systems. This similarity may reduce the need for extensive training on this process, as it leverages existing knowledge of users.

In line with this, one participant wrote, “without using it too much, [she] felt [she] picked up basic commands relatively easily,” while another stated that, “the process followed close enough to WarnGen not to be too confusing.” Nevertheless, while some participants could use features like the polygon creation with little practice, features like freehand drawing were more challenging. Others experienced a steeper learning curve, and several participants expressed frustration with the timeline feature. While one of the timeline’s purposes was to assist users in maintaining situation awareness, few participants understood how to manipulate it as intended, and when user errors occurred, recognizing one’s errors was not intuitive.

**DISCUSSION.** This study’s purpose was to identify usability issues in Hazard Services with a sample of experienced weather forecasters. As anticipated, the user testing evaluation and questionnaire revealed several areas for interface improvement.

**Recommendations for atmospheric science software development.** Assessing usability during the development phase of Hazard Services provided insight into design features that led to user errors. However, it also revealed several features that promoted task performance within the interface. For both positive and negative aspects, these lessons learned may have utility for other developers of atmospheric and hydrologic decision-support software.

First, we recommend that software designers seek to reduce the workload imposed by new systems. Minimizing human memory load during task performance can improve a system’s efficiency and may be accomplished by conveying information to the user that directs them to follow an appropriate workflow

(Nielsen 1992; Krug 2000). Efficient task flow was a core consideration during the design phase for Hazard Services, and the user testing results demonstrated several areas where further assessment would benefit usability. Based on the errors observed during user testing, we hypothesize that memory load could be reduced by using visual guidance through the hazard information and issuing dialog windows to direct users’ attentions through the critical steps. While few participants had difficulty entering data into the hazard information dialog window, several were challenged by the “product staging” window following the data entry step. To proceed, the participant needed to click the Continue button. While this may be necessary for organizational reasons, several participants experienced confusion and returned to the Hazard Information panel, failing to issue the hazard in a timely manner.

Based on the current findings and Nielsen’s (1992) usability heuristics, we also recommend differentiating the labels on navigational buttons within the hazard issuance workflow. This aligns with Oakley and Daudert’s (2016) best practice of using consistent and expressive labels to direct system users. However, as an interface for work-related activities, Hazard Services users would need training on system functions and integrating the new system with existing organizational best practices. Nevertheless, Shneiderman (2003, p. 15) states that developers can reduce the need for basic training by “[bridging] the gap between what [users] know and what they need to know.” In practice, this could translate into using on-screen instructions, graphical guidance, or reverse functions for error correction. Such design methods can complement tutorials, making training more effective and preempting errors made by experienced users.

Reducing workload may also be accomplished by improving the alignment between system features and user expectations; this has been referred to as mapping (Nielsen 1994a). In practice, natural mappings can reflect meaning through the use of expected locations, graphical elements (such as color or shape), or motion. In the present case, participants struggled with tasks requiring icon identification and indicated issues related to recognition. When developing interfaces, the most usable icons are able to convey meaning to the user about the type of function they represent; however, ensuring this is notoriously challenging (Moyes and Jordan 1993).

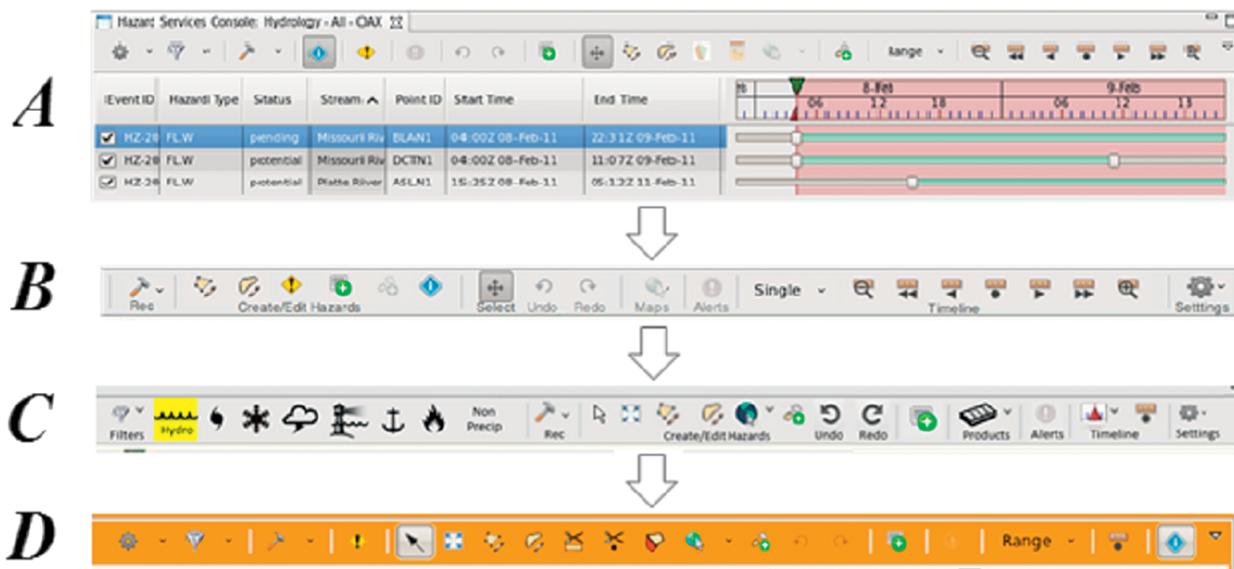
Ultimately, we encourage developers of atmospheric and hydrologic decision-support systems to put usability and design heuristics into practice when developing user interfaces. For example, color coding has been recognized as an effective means for reducing

user memory workload and enhancing task performance within visual displays (Braun et al. 1995; Yeh and Wickens 2001; Hegarty 2011). Likewise, interface layout itself can facilitate strong task performance. In this study, some participants made navigational errors by searching for the Draw Polygon function in the Settings drop-down menu. This may have been partly due to the Settings icon's prime location on the far left of the icon row, which may have drawn user attention away from the required icons. In light of this, we recommend that software developers consider frequency-of-use and functional grouping heuristics to improve the natural mappings between user expectations and tangible interactions. Indeed, in the design phase following this evaluation, these findings were considered alongside other user requirement assessments, resulting in a new icon row prototype shown in Fig. 3.

*Hazard Services and the iterative development process.* Based on evaluation findings and design heuristics, we developed a proposed redesign of the icon row in the console. In the original design, shown in Fig. 3a, icons were arranged with settings and filters on the left edge and timeline controls on the right edge. Between these, icons for hazard creation were interspersed among icons used for data management and display. For example, the icon for adding geometry to an existing hazard (ninth icon from the right in Fig. 3a) was separated from the polygon drawing icons by three data management icons. Searching for the correct icons contributed to increases in task performance

time during the user testing, so to address this, the proposed design shown in Fig. 3b relocated several of the icons to align with usage frequency. Frequently used icons, particularly the hazard creation icons, were brought to the far left of the list, thus reducing the time and precision needed to reach them. Likewise, icons were grouped by functional category. While timeline manipulation icons were already collectively located, hazard creation icons were originally interspersed between other types of icons. Last, functional groups and certain icons received additional labels. In the original design, users could hover over each icon to activate a label, but this increased task performance time as participants hunted for the correct icons. We attempted to reduce interaction time and human memory load by placing labels directly under the icons.

The suggested changes were submitted to the Hazard Services team. The team, composed of system designers, software developers, and end users, discussed the potential changes extensively. Since the experiment did not include prior training in Hazard Services, these discussions produced new insights related to usability concerns of expert users and technical constraints on new designs. Discussions within the Hazard Services Integrated Work Team produced a further-refined icon row layout (Fig. 3c). The third prototype incorporated the labeling conventions and functional groupings while adding several new icons to clarify functions that had previously been hidden in drop-down menus. However, software development is subject to a number of constraints. The software



**FIG. 3.** Design iterations based on user and developer feedback, including (a) the original prototype, (b) the proposed icon row layout based on user testing results, (c) an adapted proposal based on the development team's feedback on (b), and (d) the proposal intended for implementation at the time this paper was written.

engineers took into account the aforementioned recommendations as far as time and technical constraints would permit, and they developed a third implementation-ready layout. The resulting design is shown in Fig. 3d and was initially implemented in 2016. Further discussion and iterative refinement of the toolbar will continue throughout the development process.

**CONCLUSIONS.** In this work, we assessed Hazard Services' usability with novice users, which provided insight into initial reactions and the learning curve. Going forward, evaluating the next iteration of the interface with expert users would potentially reveal differences in behavior as experience with the system increases. While the current sample possessed expertise in forecasting procedures, it is likely that experienced Hazard Services users would identify more usability issues related to system capability. In addition, future work could extend this methodology toward understanding the influence of design on collaborative work. Hazard Services is unique in that it facilitates teamwork among forecasters. While team interactions were outside the scope of this work, future development would benefit from user feedback on the collaborative interface.

The findings reflect the importance of informative labeling, error prevention, and design heuristics on the usability of Hazard Services. Evaluation results not only inform the design for interface updates but also support training requirements. As it currently exists, the system leverages existing user knowledge gained through working with WarnGen. It is our belief that an effective training course would focus on the more novel elements, such as best practices for incorporating the timeline feature into the decision-making process or using some of the automated functions for hazard issuance.

Each of the legacy applications (WarnGen, GHG, and RiverPro) is complex in its own right, and the unification of them, while streamlining the process, requires deep knowledge of the complexities and underlying software architectures to retain functionality. While the study was application oriented, we believe that the methods and design recommendations can be generalized to the development of other weather forecast decision-support tools. When obtaining user feedback for the application, two things are advised. First, all stakeholders, including the designers and software developers, need to be involved to provide understanding of the unification and to guide the process to work around technical constraints. Second, a user-centered design should be used alongside training programs in order to promote effective and efficient

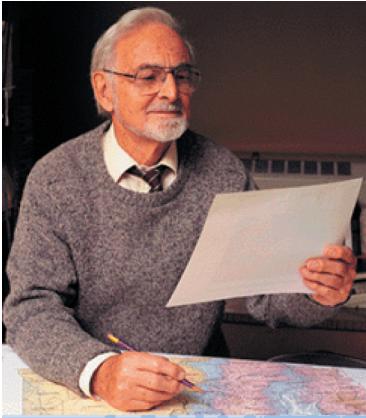
system use. Usability evaluation adds value throughout the software development process by identifying issues affecting performance and user acceptance. User testing provides empirical support for effective and efficient user interface designs. User feedback throughout the design and testing process not only ensures that a product is usable but adds value to the weather forecasting and alerting process by ensuring that systems promote task performance and error prevention.

**ACKNOWLEDGMENTS.** Funding for this research was provided by the Disaster Relief Appropriations Act of 2013 (P.L. 113-2), which provided support to the Cooperative Institute for Mesoscale Meteorological Studies at the University of Oklahoma under Grant NA14OAR4830100 and the Hydrometeorological Testbed Program under Grant NA15OAR4590158. The authors would like to acknowledge and thank Tiffany Meyer, Mike Magsig, Randa L. Shehab, Ziho Kang, Chris Golden, Steven Martinaitis, Race Clark, and Ami Arthur for their support and feedback during the usability evaluation.

## REFERENCES

- Abras, C., D. Maloney-Krichmar, and J. Preece, 2004: User-centered design. *Encyclopedia of Human-Computer Interaction*, W. Sims Bainbridge, Ed., Sage Publications, 445–456.
- Braun, C. C., P. B. Mine, and N. C. Silver, 1995: The influence of color on warning label perceptions. *Int. J. Ind. Ergon.*, **15**, 179–187, doi:10.1016/0169-8141(94)00036-3.
- Brooke, J., 1996: SUS: A quick and dirty usability scale. *Usability Evaluation in Industry*, P. W. Jordan et al., Eds., Taylor & Francis Ltd., 189–194.
- Buie, E., and D. Murray, 2012: *Usability in Government Systems: User Experience Design for Citizens and Public Servants*. Morgan Kaufmann, 440 pp.
- Chin, J. P., V. A. Diehl, and K. L. Norman, 1988: Development of an instrument measuring user satisfaction of the human-computer interface. *CHI '88: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, J. J. O'Hare, Ed., ACM, 213–218, doi:10.1145/57167.57203.
- Daipha, P., 2015: From bricolage to collage: The making of decisions at a weather forecast office. *Sociol. Forum*, **30**, 787–808, doi:10.1111/socf.12192.
- Faulkner, L., 2003: Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behav. Res. Methods Instrum. Comput.*, **35**, 379–383, doi:10.3758/BF03195514.
- Gourley, J., and Coauthors, 2017: The Flooded Locations and Simulated Hydrographs (FLASH) project: Improving the tools for flash flood monitoring and

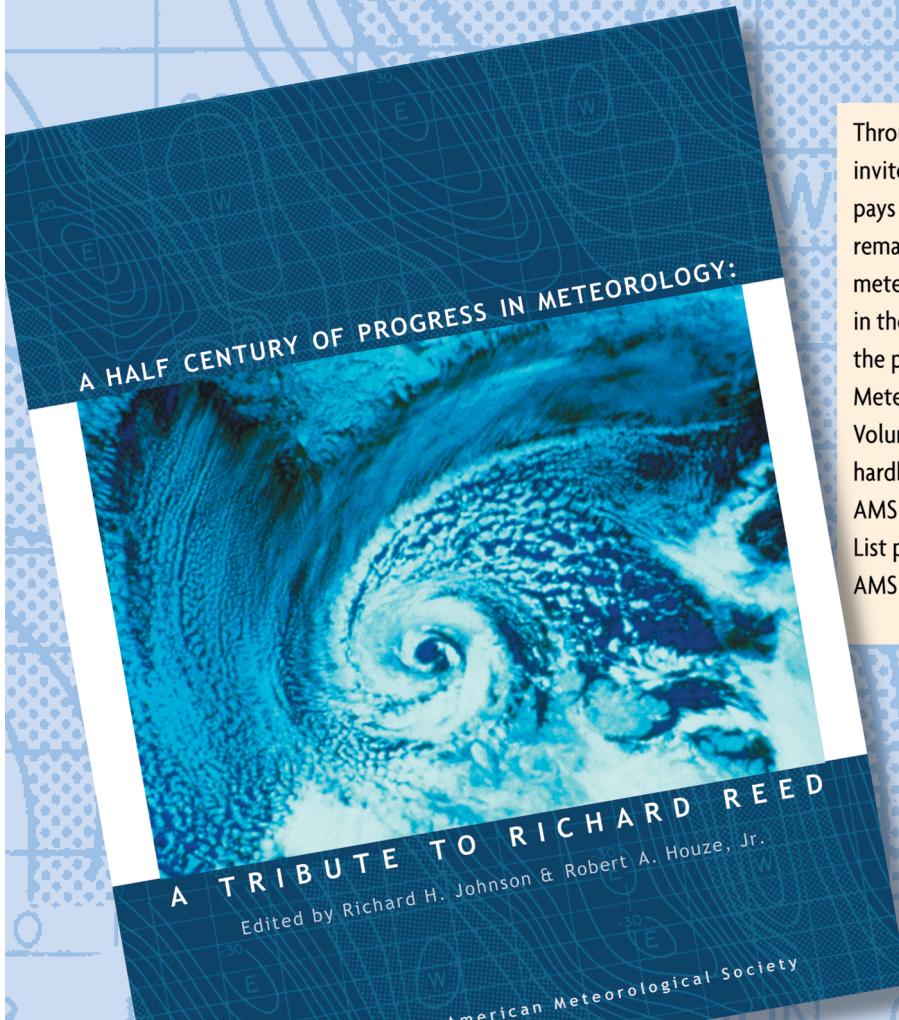
- prediction across the United States. *Bull. Amer. Meteor. Soc.*, **98**, 361–372, doi:10.1175/BAMS-D-15-00247.1.
- Hegarty, M., 2011: The cognitive science of visual-spatial displays: Implications for design. *Top. Cognit. Sci.*, **3**, 446–474, doi:10.1111/j.1756-8765.2011.01150.x.
- Holzinger, A., 2005: Usability engineering methods for software developers. *Commun. ACM*, **48**, 71–74, doi:10.1145/1039539.1039541.
- ISO, cited 2017: Ergonomics of human-system interaction—Part 210: Human-centred design for interactive systems. International Organization for Standards Doc. ISO 9241-210:2010, Accessed 10 Jan 2017. [Available online at [www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-1:vl:en](http://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-1:vl:en).]
- Jeffries, R., J. R. Miller, C. Wharton, and K. Uyeda, 1991: User interface evaluation in the real world: A comparison of four techniques. *CHI '91: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 119–124, doi:10.1145/108844.108862.
- Jokela, T., J. Koivumaa, J. Pirkola, P. Salminen, and N. Kantola, 2006: Methods for quantitative usability requirements: A case study on the development of the user interface of a mobile phone. *Pers. Ubiquitous Comput.*, **10**, 345–355, doi:10.1007/s00779-005-0050-7.
- Krug, S., 2000: *Don't Make Me Think: A Common Sense Approach to Web Usability*. Pearson Education, 216 pp.
- Lewis, J. R., 2002: Psychometric evaluation of the PSSUQ using data from five years of usability studies. *Int. J. Hum.-Comput. Interact.*, **14**, 463–488, doi:10.1080/10447318.2002.9669130.
- Ling, C., L. Hua, C. D. Karstens, G. J. Stumpf, T. M. Smith, K. M. Kuhlman, and L. Rothfusz, 2015: A comparison between WarnGen system and probabilistic hazard information system for severe weather forecasting. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, **59**, 1791–1795, doi:10.1177/1541931215591387.
- Martinaitis, S. M., and Coauthors, 2017: The HMT Multi-Radar Multi-Sensor Hydro Experiment. *Bull. Amer. Meteor. Soc.*, **98**, 347–359, doi:10.1175/BAMS-D-15-00283.1.
- Mintmire, E., C. Ling, and R. Shehab, 2013: Usability evaluation of a paratransit management system. *Proceedings of the 2013 Industrial and Systems Engineering Research Conference*, A. Krishnamurthy and W. K. V. Chan, Eds., IIE, 2000–2009.
- Moyes, J., and P. W. Jordan, 1993: Icon design and its effect on guessability, learnability, and experienced user performance. *People and Computers VIII: Proceedings of the HCI '93 Conference*, J. L. Alty, D. Diaper, and S. Guest, Eds, British Computer Society Conference Series, Vol. 7, Cambridge University Press, 49–60.
- Nielsen, J., 1992: The usability engineering life cycle. *Computer*, **25**, 12–22, doi:10.1109/2.121503.
- , 1994a: *Usability Engineering*. Elsevier, 362 pp.
- , 1994b: Usability inspection methods. *CHI '94: Conference Companion on Human Factors in Computing Systems*, ACM, 413–414, doi:10.1145/259963.260531.
- , 2007: Multiple-user simultaneous testing (MUST). Nielsen Norman Group, accessed 20 June 2016. [Available online at [www.nngroup.com/articles/multiple-user-simultaneous-testing/](http://www.nngroup.com/articles/multiple-user-simultaneous-testing/).]
- Norman, D., 2006: Why doing user observations first is wrong. *Interactions*, **13**, 50, doi:10.1145/1142169.1142199.
- Oakley, N. S., and B. Daudert, 2016: Establishing best practices to improve usefulness and usability of web interfaces providing atmospheric data. *Bull. Amer. Meteor. Soc.*, **97**, 263–274, doi:10.1175/BAMS-D-14-00121.1.
- Osgood, C. E., and Z. Luria, 1954: A blind analysis of a case of multiple personality using the semantic differential. *J. Abnorm. Soc. Psychol.*, **49**, 579–591, doi:10.1037/h0054362.
- Raytheon, 2016: AWIPS CAVE-D2D user's manual: AWIPS II operational build 13.4.1. Raytheon Doc. AWP.MAN.UM.A2-OB13.4.1, 609 pp. [Available online at [www.unidata.ucar.edu/software/awips2/doc/CAVE\\_D2D\\_Manual.pdf](http://www.unidata.ucar.edu/software/awips2/doc/CAVE_D2D_Manual.pdf).]
- Shneiderman, B., 2003: *Leonardo's Laptop: Human Needs and the New Computing Technologies*. MIT Press, 281 pp.
- Timofeyeva-Livezey, M., F. Horsfall, A. Hollingshead, J. Meyers, and L.-A. Dupigny-Giroux, 2015: NOAA Local Climate Analysis Tool (LCAT): Data, methods, and usability. *Bull. Amer. Meteor. Soc.*, **96**, 537–545, doi:10.1175/BAMS-D-13-00187.1.
- U.S. DHHS, 2016: Usability evaluation basics. U.S. Department of Health and Human Services. [Available online at [www.usability.gov/what-and-why/usability-evaluation.html](http://www.usability.gov/what-and-why/usability-evaluation.html).]
- Virzi, R. A., 1992: Refining the test phase of usability evaluation: How many subjects is enough? *Hum. Factors*, **34**, 457–468, doi:10.1177/001872089203400407.
- Yeh, M., and C. D. Wickens, 2001: Attentional filtering in the design of electronic map displays: A comparison of color coding, intensity coding, and decluttering techniques. *Hum. Factors*, **43**, 543–562, doi:10.1518/001872001775870359.
- Yen, P.-Y., and S. Bakken, 2009: A comparison of usability evaluation methods: Heuristic evaluation versus end-user think-aloud protocol—An example from a web-based communication tool for nurse scheduling. *AMIA Annu. Symp. Proc.*, **2009**, 714–718.



# A Half Century of Progress in Meteorology: A Tribute to Richard Reed

edited by **Richard H. Johnson and Robert A. Houze Jr.**

with selections by: **Lance F. Bosart Robert W. Burpee Anthony Hollingsworth  
James R. Holton Brian J. Hoskins Richard S. Lindzen John S. Perry Erik A. Rasmussen  
Adrian Simmons Pedro Viterbo**



Through a series of reviews by invited experts, this monograph pays tribute to Richard Reed's remarkable contributions to meteorology and his leadership in the science community over the past 50 years. 2003.

Meteorological Monograph Series, Volume 31, Number 53; 139 pages, hardbound; ISBN 1-878220-58-6; AMS Code MM53.

List price: \$80.00

AMS Member price: \$60.00

**ORDER ONLINE:** [bookstore.ametsoc.org](http://bookstore.ametsoc.org) or see the order form at the back of this issue

# AMS BOOKS

RESEARCH

APPLICATIONS

HISTORY

## AMS MEMBERS GET FREE

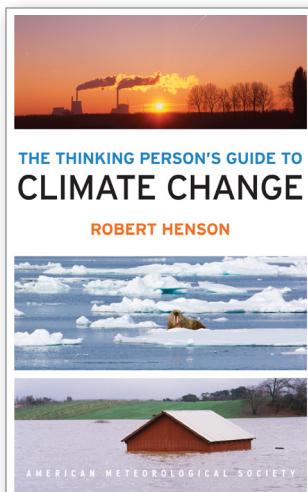
### CLIMATE

## The Thinking Person's Guide to Climate Change

ROBERT HENSON

This fully updated and expanded revision of *The Rough Guide to Climate Change* combines years of data with recent research. It is the most comprehensive overview of climate science, acknowledging controversies but standing strong in its stance that the climate is changing—and something needs to be done.

© 2014, PAPERBACK, 520 PAGES,  
ISBN: 978-1-935704-73-7  
LIST \$30 MEMBER \$20

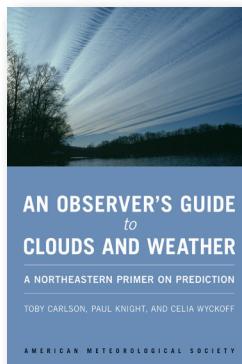


### GUIDES

## An Observer's Guide to Clouds and Weather:

A Northeastern Primer on Prediction

TOBY CARLSON, PAUL KNIGHT,  
AND CELIA WYCKOFF



With help from Penn State experts, start at the beginning and go deep. This primer, intended for both serious enthusiasts and new meteorology students, will leave you with both refined observation skills and an understanding of the complex science behind the weather: the ingredients for making reliable predictions of your own. It connects fundamental meteorological concepts with the processes that shape weather patterns, and will make an expert of any dedicated reader.

© 2014, PAPERBACK, 210 PAGES,  
ISBN: 978-1-935704-58-4 LIST \$30 MEMBER \$20

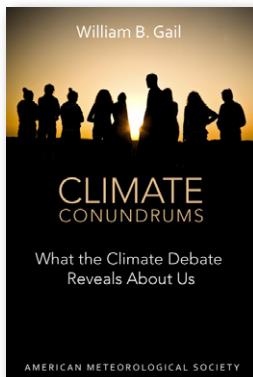
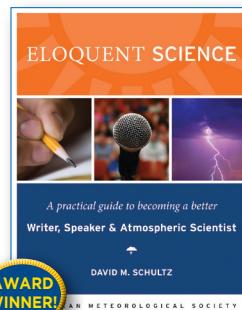
## Eloquent Science: A Practical Guide to Becoming a Better Writer, Speaker, and Atmospheric Scientist

DAVID M. SCHULTZ

The ultimate communications manual for undergraduate and graduate students as well as researchers in the atmospheric sciences and their intersecting disciplines.

© 2009, PAPERBACK, 440 PAGES,  
ISBN 978-1-878220-91-2

LIST \$45 MEMBER \$30

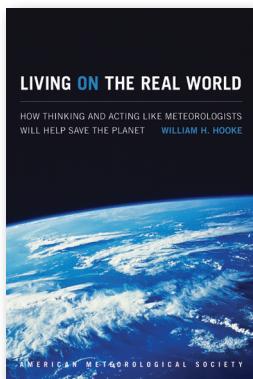


## Climate Conundrums: What the Climate Debate Reveals about Us

WILLIAM B. GAIL

This is a journey through how we think, individually and collectively, about humanity's relationship with nature, and more. Can we make nature better? Could science and religion reconcile? Gail's insights on such issues help us better understand who we are and find a way forward.

© 2014, PAPERBACK, 240 PAGES,  
ISBN: 978-1-935704-74-4 LIST \$30 MEMBER \$20



## Living on the Real World: How Thinking and Acting Like Meteorologists Will Help Save the Planet

WILLIAM H. HOOKE

Meteorologists focus on small bits of information while using frequent collaboration to make decisions. With climate change a reality, William H. Hooke suggests we look to the way meteorologists operate as a model for how we can solve the 21st century's most urgent environmental problems.

© 2014, PAPERBACK, 272 PAGES, ISBN 978-1-935704-56-0 LIST \$30 MEMBER \$22

### TEXTBOOK

## Midlatitude Synoptic Meteorology: Dynamics, Analysis, and Forecasting

GARY LACKMANN

This textbook links theoretical concepts to modern technology, facilitating meaningful application of concepts, theories, and techniques using real data.

©2011, PAPERBACK, 360 PAGES,  
ISBN 978-1-878220-10-3

LIST \$100 MEMBER \$75 STUDENT MEMB. \$65

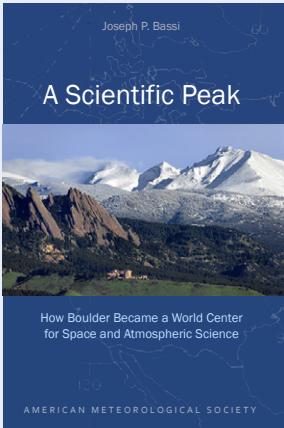


Midlatitude Synoptic Meteorology Teaching CD  
More than 1,000 PowerPoint Slides.

© 2013, CD, ISBN 978-1-878220-27-1 LIST \$100 MEMBER \$75



To order: [bookstore.ametsoc.org](http://bookstore.ametsoc.org), 617-226-3998, or use the order form in this magazine



## A Scientific Peak: How Boulder Became a World Center for Space and Atmospheric Science

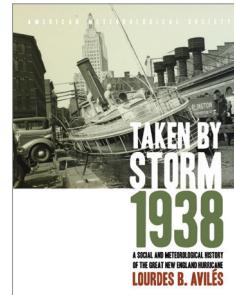
JOSEPH P. BASSI

How did big science come to Boulder, Colorado? Joe Bassi introduces us to the characters, including Harvard sun-Earth researcher Walter Orr Roberts, and the unexpected brew of politics, passion, and sheer luck that during the Cold War era transformed this "Scientific Siberia" to home of NCAR and NOAA.

© 2015, PAPERBACK, 264 PAGES, ISBN: 978-1-935704-85-0

LIST PRICE: \$35.00 MEMBER PRICE: \$25.00

## HISTORY



## Taken by Storm, 1938: A Social and Meteorological History of the Great New England Hurricane

LOURDES B. AVILÉS

The science behind the 1938 Hurricane, which hit New England unannounced, is

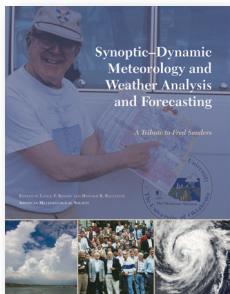
presented here for the first time along with new data that sheds light on the motivations of the Weather Bureau forecasters. This compelling history successfully weaves science, historical accounts, and social analyses to create a comprehensive picture of the most powerful and devastating hurricane to hit New England to date.

© 2013, HARDCOVER, 288 PAGES, ISBN: 978-1-878220-37-0

LIST \$40 MEMBER \$30

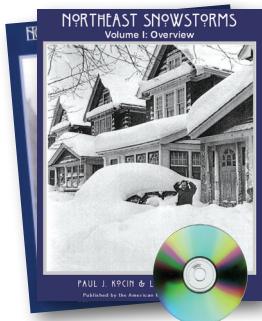
## METEOROLOGICAL MONOGRAPH SERIES

AMS Meteorological Monographs Vols. 1-55 (1947-2010) are available via Springer at [springer.com/ams](http://springer.com/ams), with select volumes available in print via AMS at [bookstore.ametsoc.org](http://bookstore.ametsoc.org). Starting in 2016 with Vol. 56, new monographs will be open access, available via the AMS Journals Online site.



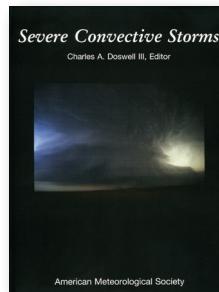
## Synoptic-Dynamic Meteorology and Weather Analysis and Forecasting: A Tribute to Fred Sanders

EDITED BY LANCE F. BOSART AND HOWARD B. BLUESTEIN



## Northeast Snowstorms (Volume I: Overview, Volume II: The Cases)

PAUL J. KOCIN AND LOUIS W. UCCELLINI



## Severe Convective Storms

EDITED BY CHARLES A. DOSWELL III

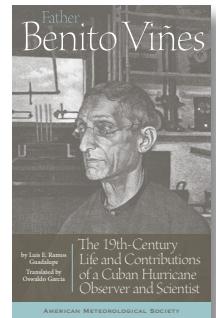
## Father Benito Viñes: The 19th-Century Life and Contributions of a Cuban Hurricane Observer and Scientist

LUIS E. RAMOS GUADALUPE  
TRANSLATED BY OSWALDO GARCIA

Before Doppler radar and weather broadcasts, Spanish Jesuit Benito Viñes (1837-1893) spent decades observing the skies at Belen Observatory in colonial Cuba. Nicknamed "the Hurricane Priest," Viñes taught the public about the weather and developed the first network of weather observation stations in the Caribbean, groundwork for the hurricane warning systems we use today.

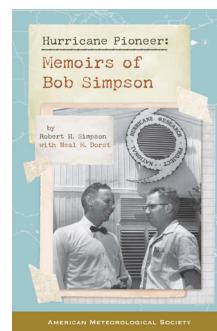
© 2014, PAPERBACK, 172 PAGES

ISBN: 978-1-935704-62-1 LIST \$20 MEMBER \$16



## Hurricane Pioneer: Memoirs of Bob Simpson

ROBERT H. SIMPSON AND NEAL DORST



In 1951, Bob Simpson rode a plane into a hurricane—just one of the many pioneering exploits you'll find in these memoirs. Bob and his wife Joanne are meteorological icons: Bob was the first director of the National Hurricane Research Project and a director of the National Hurricane Center. He helped to create the Saffir-Simpson Hurricane Scale; the public knows well his Categories 1-5. Proceeds from this book help support the AMS's K. Vic Ooyama Scholarship Fund.

© 2015, PAPERBACK, 156 PAGES

ISBN: 978-1-935704-75-1 LIST \$25 MEMBER \$20

## Booksellers, groups, or for examination copies:

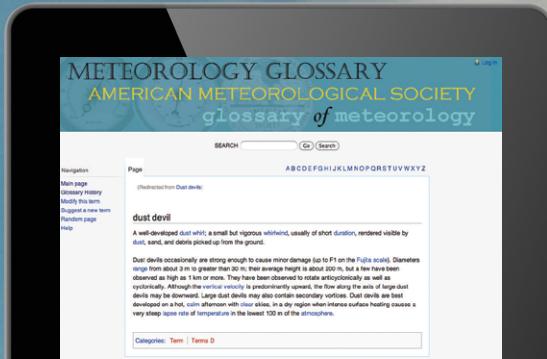
The University of Chicago Press:  
1-800-621-2736 (US & Canada)  
773-702-7000 (all others)  
[custserv@press.uchicago.edu](mailto:custserv@press.uchicago.edu)



Find out from the authoritative source

for definitions of meteorological terms.

[ What's a dust devil? ]



# THE AMERICAN METEOROLOGICAL SOCIETY Online Glossary of Meteorology

With over 12,000 meteorological terms,  
you'll be able to look up definitions  
online any time, any place, anywhere.

<http://glossary.ametsoc.org/wiki>

Also available in hardcover and  
CD formats at the AMS Bookstore,  
[www.ametsoc.org/amsbookstore](http://www.ametsoc.org/amsbookstore).

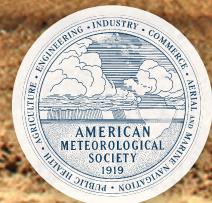


Photo: Sam Colantonio