Collaborative Exploration of Storm-Scale Probabilistic Guidance for NWS Forecast Operations

KATIE A. WILSON,[a,b,c] PATRICK C. BURKE,[a] BURKELY T. GALLO,[b,d] PATRICK S. SKINNER,[a,b] T. TODD LINDLEY,[e] CHAD GRAVELLE,[f] STEPHEN W. BIEDA III,[g] JONATHAN G. MADDEN,[a,b] JUSTIN W. MONROE,[a,b] JORGE E. GUERRA,[h] AND DALE A. MORRIS[b,i]

[a] *NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma*
[b] *Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma, Norman, Oklahoma*
[c] *RAND Corporation, Santa Monica, California*
[d] *NOAA/NWS Storm Prediction Center, Norman, Oklahoma*
[e] *NOAA/NWS Weather Forecast Office, Norman, Oklahoma*
[f] *NOAA/NWS Southern Region Headquarters, Fort Worth, Texas*
[g] *NOAA/NWS Headquarters, Silver Spring, Maryland*
[h] *Project Canary, Denver, Colorado*
[i] *NOAA/NWS Warning Decision Training Division, Norman, Oklahoma*

(Manuscript received 14 October 2023, in final form 3 January 2024, accepted 4 January 2024)

ABSTRACT: The operational utility of the NOAA National Severe Storm Laboratory's storm-scale probabilistic Warn-on-Forecast System (WoFS) was examined across the watch-to-warning time frame in a virtual NOAA Hazardous Weather Testbed (HWT) experiment. Over four weeks, 16 NWS forecasters from local Weather Forecast Offices, the Storm Prediction Center, and the Weather Prediction Center participated in simulated forecasting tasks and focus groups. Bringing together multiple NWS entities to explore new guidance impacts on the broader forecast process is atypical of prior NOAA HWT experiments. This study therefore provides a framework for designing such a testbed experiment, including methodological and logistical considerations necessary to meet the needs of both local office and national center NWS participants. Furthermore, this study investigated two research questions: 1) How do forecasters envision WoFS guidance fitting into their existing forecast process? and 2) How could WoFS guidance be used most effectively across the current watch-to-warning forecast process? Content and thematic analyses were completed on flowcharts of operational workflows, real-time simulation interactions, and focus group activities and discussions. Participants reported numerous potential applications of WoFS, including improved coordination and consistency between local offices and national centers, enhanced hazard messaging, and improved operations planning. Challenges were also reported, including the knowledge and training required to incorporate WoFS guidance effectively and forecasters' trust in new guidance and openness to change. The solutions identified to these challenges will take WoFS one step closer to transition, and in the meantime, improve the capabilities of WoFS for experimental use within the operational community.

SIGNIFICANCE STATEMENT: A first-of-its-kind experiment brought together forecasters from local weather forecast offices and national centers to examine the experimental Warn-on-Forecast System's (WoFS's) potential applications across watch-to-warning scales. This experiment demonstrated that WoFS can provide great benefit to forecasters, though a few challenges remain. Benefits provided by WoFS frequently overlap roles and responsibilities at local and national scales, suggesting the potential for enhanced cross-office collaboration. The challenges anticipated for WoFS operational use are far fewer than the benefits, and some solutions to these challenges are now being implemented. Finally, the mixed-methods experimental framework described herein also provides guidance for future collaborative experiments in testbed research that examine impacts of new technologies across NWS entities.

KEYWORDS: Social Science; Ensembles; Forecasting techniques; Numerical weather prediction/forecasting; Operational forecasting; Probability forecasts/models/distribution

1. Introduction

Forecasting natural hazards is an inherently uncertain and dynamic process (Lorenz 1969; Doyle et al. 2019). When severe weather hazards are possible, NWS operational meteorologists access an abundance of information in the form of knowledge and expertise (e.g., conceptual models and semantic experience), real-time observations (e.g., surface, radar, and satellite), and automated guidance (e.g., numerical and statistical weather prediction models) (Daipha 2015). This information is used to diagnose the current state of the atmosphere, consider uncertainties, and put forth a judgment of future weather conditions (Murphy 1993). The NOAA Forecasting a Continuum of Environmental Threats (FACETs; Rothfusz et al. 2018) framework aims to improve this process with an increased flow of probabilistic information across the timeline

Gallo's current affiliation: 16th Weather Squadron, Offutt Air Force Base, Bellevue, Nebraska.

Corresponding author: Katie Wilson, katiew@rand.org

DOI: 10.1175/WAF-D-23-0174.1

of a weather event and for all users in the value chain. For severe weather events, the FACETs framework could especially support the forecast process between the official NWS watch and warning products, which typically spans a timeframe ranging from several hours to minutes before potential severe weather impacts.

The NOAA National Severe Storms Laboratory has been leading this analysis and forecasting effort through their development of the Warn-on-Forecast System (WoFS; Stensrud et al. 2009; Heinselman et al. 2023). WoFS runs a 900-km^2 domain nested inside the High-Resolution Rapid Refresh model and is relocatable throughout the CONUS to target hazardous weather or other weather phenomena (e.g., fire weather). Ensemble analyses are cycled every 15 min and forecasts are launched every 30 min, projecting out three to six hours. The location of the domain is prioritized depending on research goals, and typically aligns to locations likely to experience hazardous weather. The probabilistic guidance is based on 18 ensemble members and includes products for a range of hazards including hail, wind, tornadoes, and rainfall. These products provide improved specificity for timing, location, and intensity relative to currently available operational products, with specificity increasing with decreasing lead time and after convective initiation (Guerra et al. 2022).

To examine the operational utility of WoFS guidance on different NWS forecast processes, the research team has collaborated with NWS local Weather Forecast Offices (WFOs) and national centers since 2017 in a variety of test and evaluation activities. These activities have included studies conducted in the NOAA Hazardous Weather Testbed (HWT), including a survey to investigate meteorologists' understanding and interpretation of WoFS guidance (Wilson et al. 2019), an experiment to capture access patterns of WoFS guidance during forecast tasks (Wilson et al. 2021; Gallo et al. 2022), and an experiment to evaluate use of WoFS guidance for short-term probabilistic flash flood prediction (Yussouf et al. 2020; Martinaitis et al. 2022). Alongside these testbed studies, test and evaluation activities have also been dedicated to real-time use of WoFS guidance at NWS WFOs and national centers. These activities have included a 2-yr study of WoFS use at the Weather Prediction Center's (WPC's) Metwatch desk (Wilson et al. 2023) and in-person training and liaising with the Storm Prediction Center (SPC). Additionally, the research team engaged in a real-time, virtual collaboration with nine NWS Southern Region WFOs initially, and all NWS regions beginning in 2022 during weather events when experimental WoFS guidance was available to support their operations. This collaboration resulted in timely interpretive support from the model developers, quick technical troubleshooting, and exchanges of ideas driven by the realities of real-world severe weather operations (Burke et al. 2022; Skinner et al. 2023a).

Through these numerous research activities described above, forecasters at WFOs, SPC, and WPC have each demonstrated the important utility of WoFS in their forecast processes. Additionally, forecasters have anecdotally shared their visions for how WoFS could impact future operations. These demonstrations and visions have highlighted potential applications of WoFS that could enhance forecast processes at WFOs and national centers in similar ways (e.g., providing increased information on the timing and location of possible severe weather in the several hours preceding a severe weather event), as well as applications that will serve the specific forecasting responsibilities for their offices (e.g., quantifying and sharing potential rainfall amounts in WPC Mesoscale Precipitation Discussions). To more fully understand how a transition of WoFS into NWS operations could impact the workflow of multiple NWS entities, a next step in our test and evaluation research was to bring these NWS meteorologists together in a collaborative experiment. We endeavored to answer two questions: 1) How do forecasters envision WoFS guidance fitting into their existing forecast process? and 2) How could WoFS guidance be used most effectively across the current watch-to-warning forecast process?

In addition to building on prior WoFS research, this study documents a unified approach to exploring new operational guidance with WFO and national center forecasters, and provides a methodological and logistical framework for executing these types of experiments. This approach is atypical of prior NOAA Hazardous Weather Testbed studies, which have previously fallen into either WFO or national center focus areas (Calhoun et al. 2021; Gallo et al. 2017). This study provides encouragement and guidance for more collaborative experimentation in future testbed research examining impacts of new technologies across NWS entities and across watch and warning scales.

2. Methods

A mixed-methods approach was used to explore operational applications of WoFS in an experiment during winter 2021. First, practical exposure to WoFS was gained through forecasters working hands-on and immersive simulated real-time weather events. Second, reflective discussions were held in a series of three focus groups, and each session captured individual and group sentiments relating to the study's research questions. A pre-experiment survey was also issued to 1) obtain information on participants' prior WoFS and forecasting experience and to 2) collect flow charts depicting each participant's forecasting workflow for later use in the focus groups. Sixteen NWS forecasters participated in this experiment over four weeks, and two additional NWS forecasters served in supporting roles. Each week, participant groups included two forecasters from nine of the Southern Region WFOs, a lead forecaster from SPC, and a Metwatch Desk[1] forecaster from WPC. All participants had at least some prior exposure to WoFS guidance prior to the experiment. The experiment was conducted as a 4-day virtual engagement and included overview presentations, simulations, focus groups, a debrief, and an opportunity to provide post-experiment anonymous feedback (Table 1).

[1] The WPC Metwatch desk is responsible for composing mesoscale precipitation discussions covering heavy rainfall that may lead to flash flooding within 6 h.

TABLE 1. Experiment schedule of virtual activities repeated over four weeks.

Day	Activity
Monday	• Facilitator and IT preparation day
Tuesday	• Welcome and overviews • Case descriptions and AWIPS-2 demonstration • Simulation, Case 1 (3.5 h): "Getting Situated"
Wednesday	• Simulation, Case 2 (8 h): "A Day in the NWS Office"
Thursday	• Focus group, session 1: The existing forecast process • Focus group, session 2: Comparing WoFS use at WFOs, SPC, and WPC
Friday	• Focus group, session 3: The visionary forecast process • Experiment debrief • Participant feedback survey

a. *Simulation design*

Participants completed two simulated real-time cases (described in section 2d). To offer briefings like those that occur at shift change, participants viewed online briefing forecast videos that an NWS collaborator prepared for the experiment. Participants were also given a written description of their designated role and responsibilities for the duration of each case, which were assigned based on real-world forecast duties. Roles included a WFO mesoanalyst, WFO decision support services (DSS) forecaster, an SPC lead forecaster, and a WPC Metwatch desk forecaster. For the duration of each case, participants and facilitators remained online together in a Google Meet room and in a Google Chat room. Participants were encouraged to engage with one another however they felt comfortable, and instant accessibility to facilitators allowed for quick troubleshooting of technical and nontechnical issues.

All participants had access to traditional observations (e.g., radar, satellite, surface and upper air) and model guidance e.g., from WoFS as well as the High-Resolution Rapid Refresh (Dowell et al. 2022) and High-Resolution Ensemble Forecast (Roberts et al. 2019) systems. Forecasters were encouraged to issue products typical of their role with decisions limited to a custom county warning area identified for each case. Issued products included watches, mesoscale discussions, DSS graphics, warnings, and NWS chats. Due to the high levels of workload required for warning issuance in these cases (as observed in a pilot experiment), warning issuance was assigned to an NWS forecaster who served in a supporting role. The issuance of these NWS products was important for allowing participants to engage in decision-making processes typical of normal operations.

b. *Virtual forecasting interface*

Given the virtual nature of this experiment, an operating platform enabling simulated real-time viewing of the cases while performing usual job functions was required. In past years, the NOAA HWT technological infrastructure supported these types of simulations using the Advanced Weather Interactive Processing System Two (AWIPS-2; Kingfield and Magsig 2009), which is an interactive forecast display system that WFOs use for most of their forecast functions and product creation. As of 2021, national centers were less reliant upon AWIPS-2 for product creation, though they had some familiarity with its interface and capabilities. Moreover, at the time of preparation for this experiment, the NOAA HWT was actively working to develop cloud-based AWIPS-2 to better support virtual experiments. The research team pursued this option and became the first NOAA HWT experiment to use AWIPS-2 in the cloud.

A single instance of cloud-based AWIPS-2 was assigned to each participant. Since these instances were not connected, participants were instructed when to begin the simulation together, and they were able to load, loop, and zoom data on multiple panes. WarnGen was available to issue and update warnings. For all other product issuances, forecasters prepared visuals in AWIPS-2. Participants then completed graphics and text in a shared Google Slides document, which allowed them to view a chronology of all products issued during the cases.

The simulations were built and run using the Weather Event Simulator version 2 (WES-2) Bridge software within AWIPS-2. As the simulations progressed through time, the Weather Event Simulator Scripting Language (WESSL) tool alerted participants of local storm reports, as documented in *Storm Data* (https://verification.nws.noaa.gov/). Reports included information about the time, location, reporter, and any other additional details (e.g., the type of damage observed). Additionally, the WESSL prompted participants each hour to open a package that included archived SPC mesoanalysis graphics for categories including instability, wind shear, composite indices, and observations. Also included in this package were skew *T*–log*p* diagrams from observed soundings and plan view upper air maps from observed data. Finally, participants were provided links each hour to "newly arriving" web-based model output from the High-Resolution Ensemble Forecast (Roberts et al. 2019) and WoFS (Skinner et al. 2023b) via the Google Chat Room.

c. *Familiarization activities*

Prior to beginning the first simulation case, time was spent ensuring that all participants were provided with a working knowledge of WoFS and AWIPS-2. During Tuesday morning, an overview presentation on WoFS was provided, which included a question-and-answer session. Next, participants completed an hour-long tutorial of AWIPS-2, learning how to overcome technical hurdles for producing and capturing forecast products alongside a written guide of best practices and frequently asked questions. The cloud-based AWIPS-2 learning curve was steepest for SPC participants, as SPC uses AWIPS-2 the least in operations and SPC lead forecasters had no recent experience working in WFOs. Facilitators included the Product Generation Palette as an add-on to the cloud AWIPS-2 instances. This application, obtained from WPC, migrates some of the drawing capabilities of the national center AWIPS (N-AWIPS) to AWIPS-2. Thus, WPC and SPC forecasters were able to use their native drawing tools for features (e.g., scalloped lines and weather map symbols). To further support ease of access, the most frequently used fields and products were sought from WPC and SPC participants prior to the experiment and AWIPS-2 procedures were created for specific participants. Finally, for the first time, a subset of WoFS products were displayable within the AWIPS-2 interface.

d. Case selection and descriptions

Cases were selected to 1) include severe weather and heavy rainfall forecast challenges that would give SPC and WPC forecasters nontrivial workloads, 2) provide a mixture of useful WoFS signals, misleading WoFS signals, and noteworthy trends characteristic of typical WoFS performance, and 3) limit the likelihood of participants having a priori knowledge. Finally, events were required to offer sufficient lead-in time for forecaster analysis (>30 min), and be of a duration that would fit the allotted time for each case.

Two cases were chosen from 2018 so they would not be fresh in memories, and both occurred outside of the NWS Southern Region from which all WFO participants were drawn. The WoFS configuration in 2018 was very comparable to warm season configurations in later years. The dates and domains of the selected cases were 1) 19 July 2018 in Iowa and 2) 28 June 2018 over the Dakotas and eastern Montana.

1) Case 1: Iowa, 19 July 2018

This case offered an opportunity to study how forecasters might use WoFS to catch up to an event whose severity exceeded expectations in real operations. The experiment simulation ran for 3.5 h from 1800 UTC. Severe weather occurred in a relatively narrow zone (~150 km) near and northward of a warm front (Figs. 1a–e). As initial updrafts formed in the vorticity-rich environment, numerous nonsupercell tornadoes were reported near Des Moines, Iowa, beginning at 1930 UTC. Experiment groups issued the first watch from 1915 to 1950 UTC, and the first warning from 1952 to 1958 UTC. A long-track supercell produced an EF3 tornado at Pella, Iowa, from 2101 to 2113 UTC. Other storms produced shorter-lived tornadoes, including an EF3 at Marietta, Iowa, from 2124 to 2127 UTC. The SPC convective outlooks for this event included slight risk but did not emphasize the tornado risk. The first experiment warnings were issued from 1952 to 1958 UTC, following initial reports of the nonsupercell tornadoes.

2) Case 2: The northern plains, 28 June 2018

This case was ideal for a "Day in the NWS Office," owing to its long duration, multiple hazards, and characteristic WoFS performance. The experiment simulation ran for 6.5 h from 2230 UTC, and severe storms were present inside the WoFS domain for more than 9 h (Figs. 1f–j). During this event, rich moisture advected into the western and central Dakotas and eastern Montana along a warm frontal zone. A broad area of 3000–5000 J kg^{-1} mixed-layer convective available potential energy developed. Flow aloft was moderately strong, and a longwave ridge axis developed eastward while an upstream deep-layer trough deepened during the day. By evening the western Dakotas were beneath southwesterly flow aloft and just downstream of a deepening 850-hPa low and 700-hPa trough.

Severe thunderstorms initiated between 2100 and 0100 UTC 1) along the warm front in North Dakota, 2) within upslope flow into northeast Wyoming and southeast Montana, and 3) within the area of greater height falls from central to northeast Montana. Participants were given no existing watches or warning at the start of the simulation. A real-world tornado watch was in effect and a severe thunderstorm warning was in effect for an isolated storm in North Dakota. The first experiment warning for this or any storm in the fictional county warning area was issued at 2329–2334 UTC; the first real-world local storm reports inside the county warning area occurred at 2347 UTC. A mixture of supercell and linear convective modes were observed until consolidation into a large, bowing mesoscale convective system by 0500 UTC 29 June 2018.

e. Focus groups

Focus groups were used for qualitative inquiry to capture participants' experiences, ideas, and concerns as they reflected on their individual and group use of WoFS in the two cases. This interactive method supported participants in expressing their shared and differing views without being pressured to make decisions or to form consensus (Liamputtong 2011). Additionally, focus groups have previously been demonstrated as a successful method for obtaining meaningful insight into the potential impacts of new information for operational meteorologists (e.g., Demeritt et al. 2007; Demuth et al. 2009; Wilson et al. 2017; Houston et al. 2020).

Three semistructured focus groups were conducted over Thursday and Friday of each week (Table 1). The focus groups each lasted up to two hours, and discussions were guided with an initial task and then a set of prewritten questions (the appendix). The initial tasks included modification of previously prepared flowcharts to indicate integration of WoFS guidance into their workflows, as well as virtual whiteboard sessions to generate ideas relating to the applications, challenges, and future visions for use of WoFS in operations. The semi-structured approach ensured that specific research questions were addressed while also allowing for interaction among participants and exploration of related topics. Two moderators took turns leading each discussion. These moderators have substantial experience using qualitative methods to collect NWS forecasters' insights during HWT experiments. Several support facilitators were also present to engage with participants, listen to discussions, and ask further questions. Moderators were also responsible for encouraging an inclusive discussion and ensuring all participants had opportunities to share perspectives. All focus groups were recorded and later transcribed for analysis.

f. Data collection and analysis

Data shared in this article include all participants' verbal and written communication from the two simulations (44 h of recording) and three focus groups (18 h of recording) conducted each week. Content analysis on the simulation transcripts was conducted by building familiarization with the text, developing a codebook, and identifying the location, frequency, and context of those codes. The precise and relatively

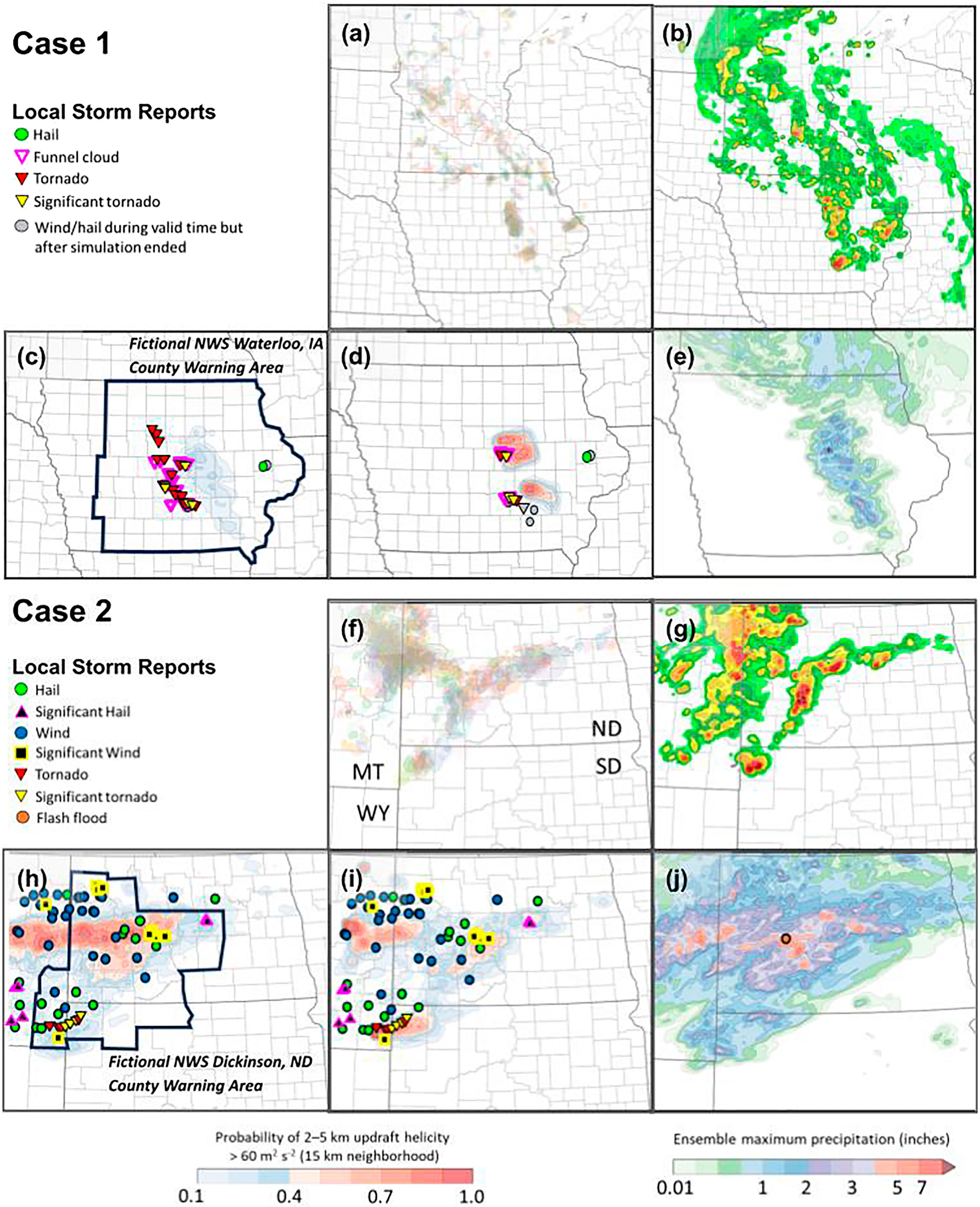

FIG. 1. Snapshots of WoFS guidance for (a)–(e) Case 1 (19 Jul 2018) and (f)–(j) Case 2 (28 Jun 2018), including 40-dB*Z* simulated composite reflectivity paintballs in (a) and (f), ensemble member simulated composite reflectivity in (b) and (g), probability of 2–5-km updraft helicity > 60 m^2 s^{-2} in (c) and (h) and the same probability for subsequent WoFS runs in (d) and (i), and ensemble hourly maximum rainfall in (e) and (j).

objective interpretation of text can provide a quantitative description of the ideas within (Neuendorf 2018).

To ensure consistent interpretation and applications of the codebook across the analysis team, four analysts worked in pairs to independently code random samples of the transcript. Units of data were first identified in these samples and typically represented the exchange of a single idea. Both analysts coded the text within those units and then calculated interrater reliability using the kappa statistic (McHugh 2012). The kappa statistic value can range from −1 to +1. A value ≤ 0 represents no agreement, 0.01–0.20 as none to slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement (McHugh 2012). In this study, acceptable levels of agreement were achieved when the kappa statistic indicated at least substantial agreement, though in most instances almost perfect agreement was achieved following one iteration of recoding (if necessary). Once interrater reliability had been established, one analyst then coded the entire transcript. The four analysts shared this responsibility across the eight transcripts.

For the focus group transcripts, a thematic analysis approach was applied. Unlike the content analysis approach, codes were not created or counted at the unit scale across transcripts. Instead, individual analysts identified big-picture and recurring ideas across the different focus groups, including both unique and overlapping perspectives (Neuendorf 2018). Themes emerged through an analysis and grouping of these ideas. Each of the focus group transcripts were reviewed for the presence of identified themes, and the prevalence of each theme across focus groups was noted (i.e., ranging from inclusion in 1–4 focus groups).

3. Results

The results draw on participants' experiences using WoFS during the two cases, as well as information shared in the focus group activities and discussions. Given the integration of these activities during the experiment (i.e., focus group discussions depended on the cases worked), the results are interwoven and based on data collected from both methods.

a. *WoFS in the existing forecast process*

To document participants' existing forecast processes prior to this experiment, all participants created a flowchart documenting their own workflow for a typical warm-season severe weather day. The first focus group session opened with an activity that gave participants an opportunity to review and modify these flowcharts to indicate if, when, and how their use of WoFS guidance would change their existing forecast process. Participants then shared their pre and post-simulation flowcharts with the group and explained their general workflow and the reasons for making any modifications.

1) Presimulation flowcharts

While unique to each participating meteorologist, the 16 flowcharts included common elements key to their forecast processes (Fig. 2). First, all participants discussed the importance of becoming situationally aware when initially coming on shift by receiving an office briefing and reviewing current observations. An important first question for most participants was whether convection was ongoing. If so, participants typically used current observations to form expectations and to check whether the current public-facing forecast products were sufficient for anticipated weather impacts.

Following this initial assessment and whether convection was already ongoing, participants described proceeding with their forecast funnel (assessing the synoptic scale down to the mesoscale) and identifying mesoscale features that could influence storm development. Participants described a workflow that then typically encompassed a three-phase cycle: analysis, forecast, and assessment (Fig. 2). The *analysis phase* included a review of observational data (e.g., satellite, radar, station observations, and antecedent conditions) and, if time permitted, the completion of a hand analysis. Next, the *forecast phase* included a review of the event's predicted evolution as indicated in large-scale models and convection allowing models (CAMs), as well as a comparison of model output to the analysis phase to determine which models were performing best. This comparison supported forecasters in making mental adjustments to their expectations for the range of potential ways the event could evolve, such as the event's potential timing, location, storm mode, and/or hazard(s). During the *assessment phase*, participants combined their takeaways from the analysis and forecast phases with their conceptual models to update their expectations of the weather event. Once expectations were formed, participants reviewed whether the status of the event had changed and whether any action was needed. If action was not needed, participants returned to the beginning of the three-phase cycle and repeated their analysis, forecast, and assessment. If action was needed, the steps that followed depended on whether the participant was from a local office or a national center (Fig. 2).

Actions for local WFO participants included conveying information and evolving needs to the warning forecaster, updating graphics and the hazardous weather outlook, communicating updates via NWS Chat and social media outlets, and communicating expectations internally with national centers. At the national centers, action included communicating expectations internally with the affected offices (e.g., local WFOs and River Forecast Centers) and determining whether a mesoscale discussion was required. For SPC, prewatch internal coordination with local WFOs also began as they created, refined, and prepared to disseminate the watch. Once the necessary actions were completed, participants returned to analysis and began the three-phase cycle again.

2) WoFS use during simulated real-time cases

The two cases provided participants with an opportunity to explore WoFS guidance use together and across different roles, responsibilities, and offices. Participants' verbalizations and written messages to one another offer a window into the collaborative forecast processes. The research team developed and coded for a set of nine topics that arose during participants' interactions, as well as the nature of those interactions (Table 2; Fig. 3). A "Single Speaker" code was included to capture occasions when participants voiced a thought without receiving a response. Each group of participants experienced different levels of interactions,

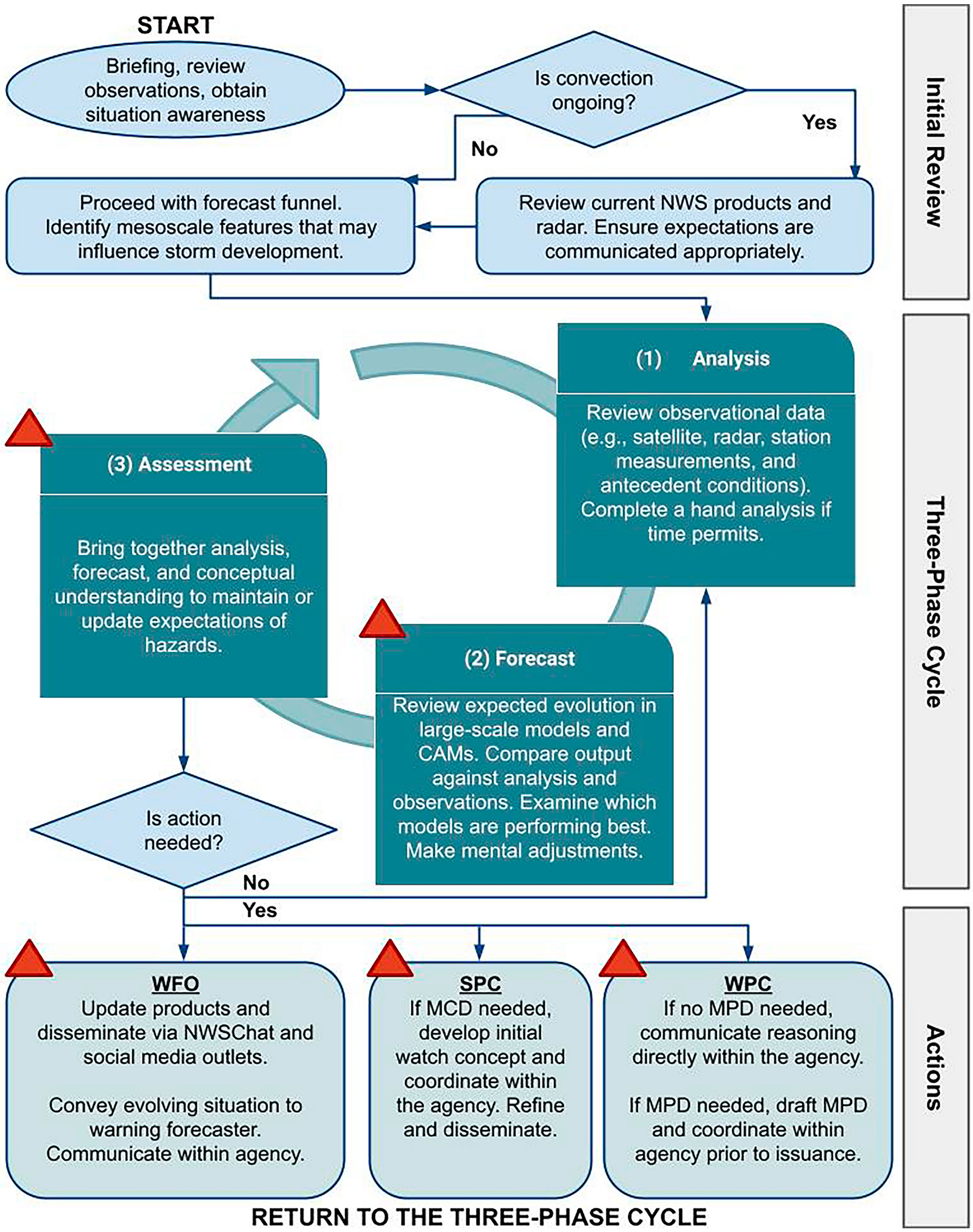

FIG. 2. A flowchart depicting common aspects of NWS forecasters' workflows for a typical warm-season, severe weather day. Red triangles indicate where NWS forecasters reported WoFS guidance most substantially modified their workflow.

and for both simulated cases the first week had the lowest level of interaction while the fourth week had the highest level of interaction (Fig. 4).

Across all transcripts, the most infrequent codes were Queries and Science Discussion. Queries related to technical and nontechnical issues during the simulations and general questions about the activity. Science Discussion was generated from general questions about WoFS that were not specific to the weather events being worked. The intraoffice code had the highest frequency count, meaning that the nature of interactions was more often between participants within the WFO (although all participants could hear the interactions). Interoffice collaboration included communication between participants from the WFO, SPC, and/or WPC and typically occurred less frequently (Fig. 5).

TABLE 2. Codebook for simulation recording content analysis.

Code	Description
Queries	Finding data, navigating the interface, asking about specifics of the experiment (e.g., issuing products), troubleshooting interface problems
Science Discussion	Asking general questions related to WoFS
WoFS	Discussing aspects of WoFS guidance specific to the event
Mesoscale Analysis	Discussion of mesoscale features and boundaries
Storm-Scale Analysis	Discussion of storm-scale characteristics unrelated to warning activity (e.g., radar signatures, satellite features, lightning activity)
Warning Operations	LSRs, tracking/triaging warning activity, discussing issuance decisions, and warning details
Collaboration (intraoffice)	Collaboration among WFO participants
Collaboration (interoffice)	Collaboration between WFO and national center (SPC and/or WPC) participants
Single Speaker	Only one person talking (i.e., thinking aloud)

The content of these interactions was coded most frequently as Storm-Scale Analysis and Warning Operations. The WoFS code occurred with lower frequency indicating that WoFS guidance was discussed more intermittently compared to other topics.

Examining the context in which WoFS discussions arose is important for understanding how WoFS guidance fits into the broader forecast process. During the first case, WoFS was coded in 63 units of all transcript data. Likewise, WoFS was coded in 122 units of transcript data in Case 2 (which was also twice as long as Case 1). The co-occurrence of the WoFS code with the other eight codes varied from week to week (Fig. 6). WoFS discussions occurred more often within the WFO rather than across offices. Still, all groups did discuss WoFS between the WFO and national centers, and the week 4 group especially demonstrated a noteworthy level of cross-office collaboration when sharing WoFS-related information. In fact, their WoFS-related collaborations

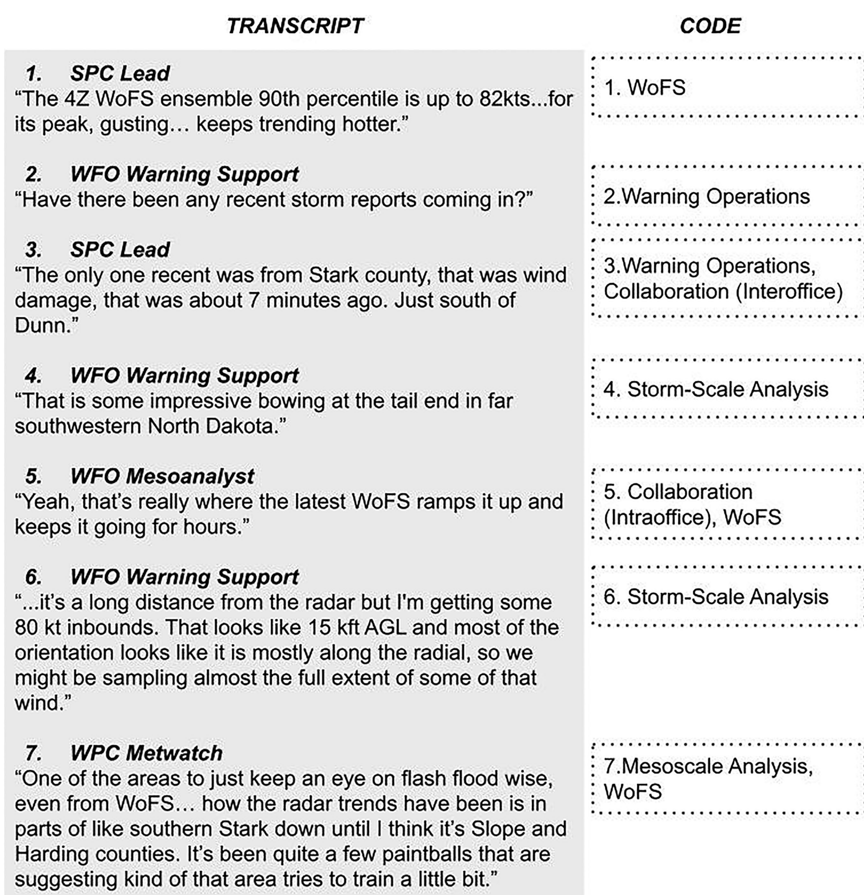

FIG. 3. An example of content analysis, with (left) units of data identified in the transcript and (right) codes identified for each of those units.

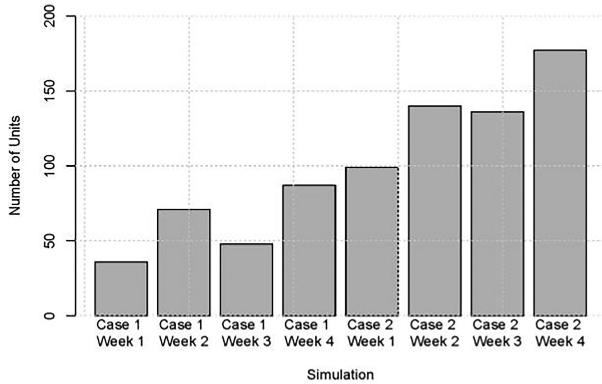

FIG. 4. Number of transcript units coded per case, per week.

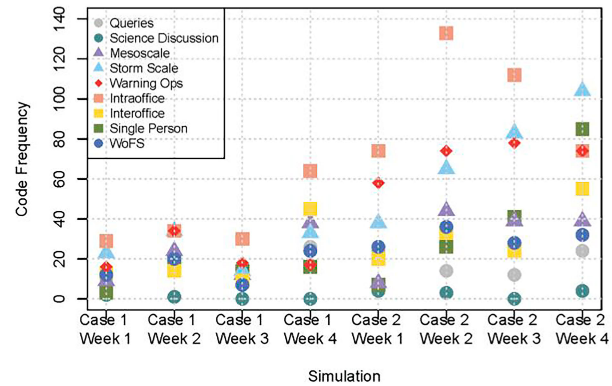

FIG. 5. Code frequency for each case across each week.

during Case 2 occurred more across the WFO and national centers rather than within the WFO (Fig. 6). WoFS-related interoffice collaboration occurred for multiple reasons. For example, participants chose to express concern about a particular storm or the likelihood of a hazard occurring (e.g., WPC informed the WFO of their concern for flash flooding). Across all weeks, interoffice collaboration also took place during coordination of national center products. WoFS guidance was used to guide the creation of these products, and in week 4, WFO participants noted high WoFS updraft helicity values predicted outside of the preliminary watch area. The SPC participant then used their feedback to adjust the watch boundaries farther east.

All groups discussed WoFS guidance in tandem with their mesoscale and storm-scale analysis (Fig. 6). For example, participants considered the positions of storms relative to fronts and other boundaries, how environmental forcing would affect storm mode and track, and the potential influence of antecedent conditions (e.g., ground wetness) in storm evolution. At the storm scale, participants compared WoFS guidance to real-time observations of storms to identify where storms were initiating and intensifying, how storm features in radar data compared to those in WoFS guidance, and any difference between observed storm evolution and WoFS predicted storm evolution. Examples of outcomes from this WoFS use include the following:

- Noticing trends indicative of increasing hazard potential [e.g., 1) tornado likelihood increasing with higher WoFS 2–5-km updraft helicity values from run to run combined with storms moving into an area of greater instability, and 2) flash flooding possible when considering a combination of the WoFS 90th percentile quantitative precipitation forecast with areas of wet soil/low flash flood guidance].
- Accounting for positional error of storms (e.g., storms were located one county too far north in the WoFS guidance).
- Anticipating motion of storms (e.g., WoFS highlighted deviant storm motion following an outflow boundary and the potential for splitting storms).

Participants' use of WoFS guidance was also discussed in a warning operations context, especially in Case 2 when warnings were issued more frequently (Fig. 6). Participants used WoFS guidance to relay hazard information either through additional text in warnings (e.g., adding wording to describe torrential rainfall) or through the warning type that was chosen (e.g., deciding to issue a tornado warning rather than a severe thunderstorm warning). Storm motion in WoFS was also used to guide the shape and track of the warning polygon, and participants used the WoFS guidance to coordinate consistency between warnings and DSS messaging. Finally, when LSRs were received, participants used them to verify both warning decisions and WoFS performance.

Figure 7 presents an example of WoFS usage during experiment week 2 for the cyclic, tornado-producing storm that began in southeast Montana in Case 2. A WFO forecaster summarized their initial impression of WoFS output aloud for the group at 2257 UTC, noting, "…that's gonna be moving east into Carter County…WoFS guidance is really kind of hitting on that…quite a few WoFS members are trying to generate some pretty big hail as it moves into Carter and maybe into Hardin County…" The SPC participant issued a tornado watch at 2314 UTC and discussed reasoning aloud, including, "…some of the WoFS stuff, it tries to develop a big outflow blob in southeast Montana…given the extreme buoyancy and even a little bit of an increase in shear in the evening…there could still be a fairly substantial tornado…" A WFO forecaster shared an AWIPS-based plot of WoFS 4-h updraft helicity swaths in the Google chat room for participants to see at 2327 UTC. The WFO participants determined a tornado warning was needed at 0034 UTC. They then spent 10 minutes communicating the growing threat in a chat message directed to partners such as emergency managers and in a forecast graphic (not shown) meant for public consumption; both the chat message and graphical downstream threat area were informed by WoFS without mentioning WoFS directly. A second graphic was issued at 0121 UTC following the second tornado warning at 0116 UTC (Fig. 7). The warning polygon is shown in yellow. The projected threat area in red extends 95 km and 60 min farther downstream than the warning.

3) POSTSIMULATION FLOWCHARTS

The post-simulation flowchart activity during the first focus group was designed to capture ways in which participants believe WoFS may or may not alter their existing forecast

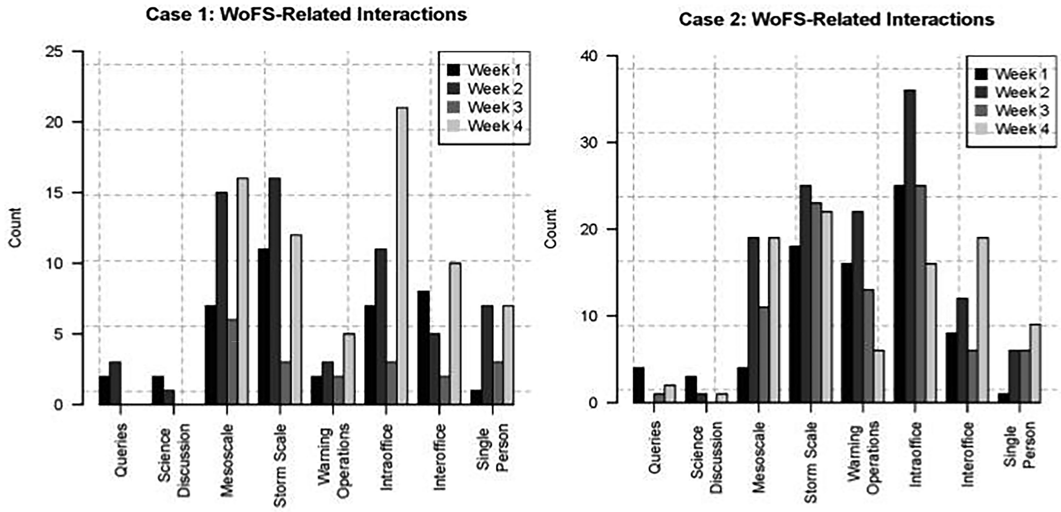

FIG. 6. Code counts for units of transcript data that were also coded for WoFS. Case 1 included a total of 63 WoFS-coded units, and Case 2 included a total of 122 WoFS-coded units.

process. All participants reflected on their experiences during the two cases, and some participants also drew on their prior experiences of using WoFS in real-world operations. In describing their post-simulation flowcharts, two thirds of participants agreed or strongly agreed that the editing task was helpful for capturing changes to their forecast and decision-making processes that resulted from their use of WoFS guidance. One participant had already included WoFS in their pre-simulation flowchart because they routinely used the guidance when it is experimentally available for real-time operations. Four participants who did not find this task helpful explained that they already use CAMs in their forecast process and that WoFS did not change the logic of their process (please note the related training discussion in section 4). This reasoning points to a need for WoFS-related training as there are key features that distinguish it from existing CAMs (see section 4).

Comparing participants' pre- and post-simulation flowcharts, most national center participants made only small modifications to their flowcharts, though one participant each from WPC and SPC and five participants from WFOs made notable modifications. Workflow areas participants identified as being most likely impacted by WoFS guidance are indicated in Fig. 2. One

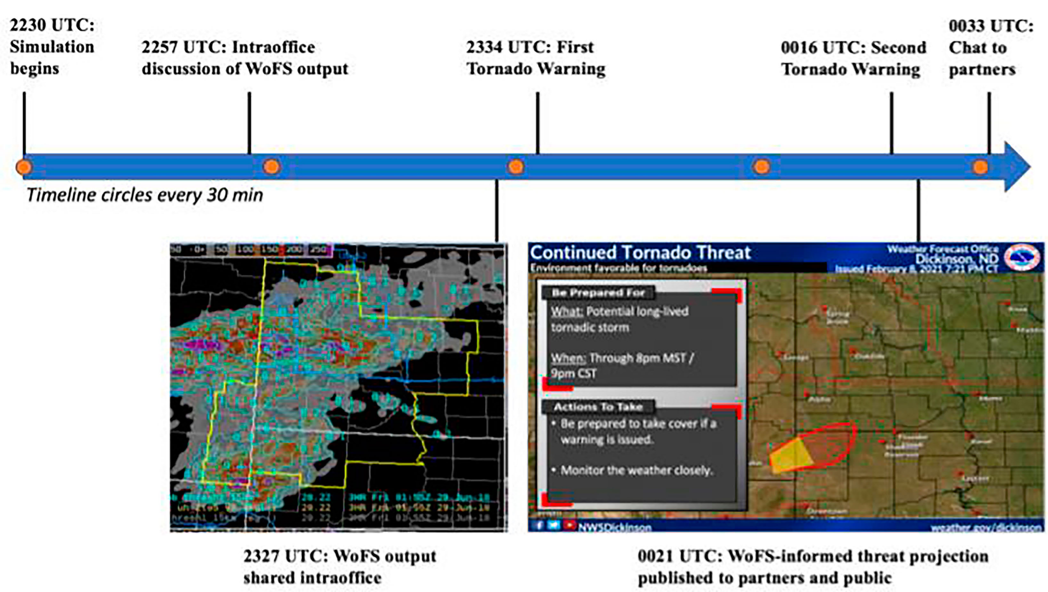

FIG. 7. A timeline depiction of week 2 simulated operations for the cyclic, tornado-producing storm in southeast Montana and northwest South Dakota. (bottom left) The WoFS 90th percentile of 2–5-km updraft helicity (s^{-1}). (bottom right) A graphical forecast denoting the area under increasing tornado threat over the next 90 min.

area of modification was in the *forecast phase*, where participants described being able to view and assess a range of possible scenarios in the ensemble output. Through an evaluation of the different WoFS ensemble members, participants could view the underlying uncertainty (including the best case, worst case, and most likely scenarios) and understand why different scenarios were predicted. The viewing of both probabilistic (i.e., exceedance threshold) and deterministic (i.e., member viewer) guidance is a preferred approach also documented in Demuth et al. (2020). Participants' understanding then supported enhancements to the *assessment phase*, where they could use WoFS to challenge, confirm, and/or adjust their mental models. WoFS was also used during the *assessment phase* to form expectations for how storms would trend over the next few hours and what hazards would likely occur, including their location, timing, type, and magnitude.

Participants from the WFOs, SPC, and WPC all described making modifications in their workflows specific to actions undertaken at their respective offices. For WFO participants, information from WoFS could be relayed to the radar meteorologist and DSS specialist to enhance internal warning operations. Enhancements related to operations planning included determining how to sectorize storms based on expected storm mode, anticipating the number of staff required, and determining what external communications were necessary given convective trends. WPC participants reported that the WoFS guidance would better enable collaboration within their office to ensure consistency across the Excessive Rainfall Outlook and Mesoscale Precipitation Discussion products. Finally, SPC participants shared that receiving WoFS updates every 30 min would help them assess whether updates to Mesoscale Convective Discussions, watches, and outlooks were needed as the event was unfolding. One SPC participant described how WoFS guidance could drive a new framework for providing routine updates to these products.

b. *Applications and challenges across the watch and warning scales*

In the second focus group participants shared their thoughts on applying WoFS across the NWS watch (national center) and warning (WFO) scales. The group was asked to place their ideas on a shared virtual whiteboard that outlined two Venn diagrams. These whiteboards held spaces dedicated for "just local offices" and for "just national centers," as well as a joint, overlapping space (Fig. 8). Overall, many more applications were identified compared to challenges.

1) APPLICATIONS

The first Venn diagram asked participants what the most useful applications of WoFS guidance could be (Fig. 8a). The most frequently shared ideas included better predicted event timing (including convective initiation and event duration), identifying the location of the greatest risk of high-impact weather, enhancing DSS of near-term hazards to the public and partners, and establishing a common reference point for WFO and national center collaboration. Participants explained that WoFS serving as a common reference point will improve consistency between WFOs' and national centers' expectations and messaging. Half of the focus groups also shared that WoFS could better enable forecasters to visualize a range of possible outcomes (including the worst-case scenario, especially for rainfall), anticipate the convective mode over the next several hours, and identify the type and magnitude of primary hazards. Finally, WoFS applications reported by just one of the four focus groups included using the 5-min product time steps to enhance situation awareness of the event's evolution, integrating guidance with observations to affirm or modify conceptual models, and conducting environment analysis with WoFS products, such as the point soundings.

Both WFO and national center participants also identified WoFS applications specific to their offices (Fig. 8a). WFO participants believe WoFS guidance will support more detailed communication graphics and improved DSS to core partners, and that associated messages could be conveyed earlier and with greater confidence. Additionally, increased specificity of storm location, path, hazard type, and severity could be provided using WoFS guidance. Furthermore, most focus groups explained that WoFS guidance could guide WFO planning, especially regarding their staffing strategy for warning operations over the next 1–3 h. For national centers, SPC and WPC participants reported that WoFS will increase the detail they can provide in their products on hazard type and timing. Additionally, as also described in the post-experiment flowchart activity, WPC participants reported that WoFS could improve internal consistency between products, and SPC suggested that routine WoFS updates could help refine the watch process.

2) CHALLENGES AND SOLUTION OPTIONS

Participants were asked to consider what aspects of WoFS use may be challenging should the system be transitioned into real-time operations. Here, we provide a review of the challenges identified (Fig. 8b), as well as solution options that participants, researchers, and operational partners either suggested or have previously explored. Many of these challenges and solution options were also discussed in the third focus group, which addressed forecasters' visionary forecast processes. The findings from both the second and third focus groups are therefore integrated in this discussion.

Challenge topics that all focus groups discussed include *trust, time,* and *data overload* (Fig. 8b). First, participants cautioned the importance of using WoFS to enhance their analysis process rather than replace it. The importance of being able to identify when WoFS is and is not performing well was stressed, such that participants could then assign an appropriate amount of weight to the guidance in their decision making. A connected challenge to accomplishing this real-time evaluation of WoFS is the required time and the potential for data overload given the many fields that WoFS produces. To mitigate these challenges, participants expect that they will need to prioritize a subset of WoFS products for different situations, which is a strategy that has successfully reduced WoFS workload impacts during a 2-yr evaluation of WoFS use during real-time WPC operations (Wilson et al. 2023). Additionally, synthesis of WoFS guidance will be possible with machine learning products that provide probabilistic

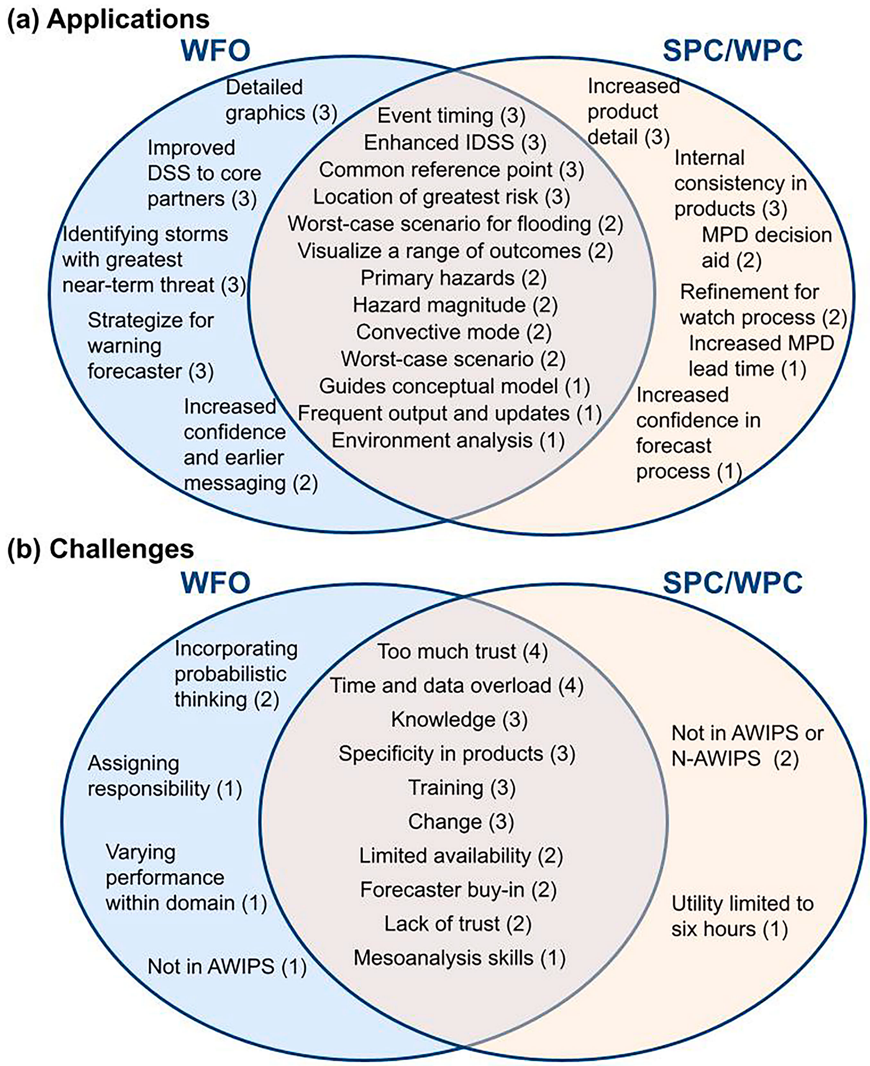

FIG. 8. Participants' shared ideas on (a) applications and (b) challenges of using WoFS guidance in NWS operations. The number of weeks reporting a specific idea is noted in parentheses.

guidance of specific hazards that are currently being developed (Flora et al. 2021; Loken et al. 2022).

Another suggested approach to reducing WoFS-related workload is to increase forecasters' abilities to understand, interpret, and apply the guidance in operations. Doing so will require overcoming an existing WoFS challenge related to *knowledge* and *training*. Overcoming this challenge is also important for establishing accurate and consistent interpretation across forecasters and offices. The release of the Foundations in Probabilistic Forecasting training curriculum through the Cooperative Program for Operational Meteorology MetEd program, which includes WoFS examples, is one step in this direction. Furthermore, the WoFS transition plan includes goals for extensive training through multiple avenues, one of which, a national webinar series aimed at forecasters, was initiated in 2023. Participants recommended training follow a format that includes best practices, a review of the characteristics and intricacies of WoFS, and an opportunity to build muscle memory of WoFS use in operations across a variety of weather scenarios through repeated exposure; the latter format is essential for establishing *trust* in the system, too.

The *limited spatial and temporal availability* of WoFS runs has presented an additional challenge for forecasters attempting to repeatedly use WoFS in operations. This limitation is especially challenging for WFOs whose area of responsibility may not lie within the WoFS domain-of-the-day, or who may experience high-end convective weather relatively infrequently. Since the HWT experiment, this challenge has been alleviated slightly with a cloud-based WoFS that can run more often and over dual domains (Martin et al. 2023). Additionally, the WoFS transition plan will outline a strategy for a concept of operations, which will include domain characteristics and placement decisions.

Another joint challenge identified in most focus groups was *change*, both in terms of changes to the forecast process and changes to the communication possibilities between offices and their communities (Fig. 8b). First, participants discussed the importance of obtaining *forecaster buy-in* on the validity of WoFS, which requires ensuring that users understand when WoFS does and does not perform well, as well as what WoFS guidance offers beyond operational models (e.g., the High-Resolution Rapid Refresh and High-Resolution Ensemble Forecast systems). The research team has been working closely with operational partners to build this understanding around real-time weather events (Bieda et al. 2022; Burke et al. 2022; Skinner et al. 2023a). Researchers and forecasters interact about WoFS guidance in real-time via a CONUS-wide online chat room and through informative case-studies reviewed in webinars. A second issue related to change is ensuring that the introduction of new guidance can still result in *consistent service* to partners across events. Varying levels of forecasters' willingness to integrate WoFS into event messaging could impact this consistency. Additionally, forecasters may have differing views on the *level of specificity* they should communicate in products. Best practices on the appropriate levels of spatial and temporal specificity to communicate from WoFS is needed, such that error is accounted for sufficiently while greater precision is offered.

Challenges unique to incorporating WoFS at WFOs were expressed in just one or two focus group sessions (Fig. 8b). First, two sessions discussed challenges around *incorporating probabilistic thinking* into their WFO workflow, such that forecasters may have tendencies to emphasize individual members and apply a deterministic mindset to their WoFS analysis. This challenge was documented in a separate examination of operational meteorologists' understanding of storm-scale ensemble forecast guidance (Wilson et al. 2019) and could be overcome through training and experience. Another WFO challenge discussed in one session was how WoFS *performance varies across the domain*, particularly prior to and near the time of convection initiation, therefore requiring an assessment of how well WoFS is handling each individual storm (Guerra et al. 2022). This unique characteristic of WoFS and the necessary analysis approach should be highlighted in forecasters' training and review material (Skinner et al. 2023b).

The WFO participants also shared two logistical challenges (Fig. 8b). First, forecast offices need to *identify which staff roles* have the ability to incorporate WoFS guidance into their current workload. Prior real-time experimentation of WoFS within WFOs has found that mesoanalysts are best suited to this role, and that further support could be provided remotely. For example, forecasters located in less-demanding locations can contribute to chat rooms, such as the NWS Central Region remote mesoanalysis room that has been active and growing over the past two years (Graham 2023; Ayd and Graham 2023). The second logistical challenge is WoFS not being available in *real-time operational platforms*. This limitation was also noted by national center participants. Since the timing of this HWT experiment, steps have been taken to implement and test a subset of WoFS products in AWIPS-2 at a WFO. The web-based WoFS also remains available to forecasters and has been heavily used in all real-time run seasons. Finally, national center participants noted that the 6-h WoFS forecasts have limited utility, since many of their forecast responsibilities and products are issued for *timeframes beyond the next six hours*. The WoF concept was intended to address the probabilistic model guidance void on the watch to warning timeline, and is based on the notion of developing an accurate analysis of individual thunderstorms so the ensemble can project them forward. Applying a WoF concept beyond six hours would likely require a new approach and is not currently a goal of the project.

4. Discussion and conclusions

This study brought together NWS forecasters from WFOs, SPC, and WPC in an experiment that explored use of WoFS guidance for forecast operations. The mixed-methods approach provided participants with an opportunity to learn about WoFS and experience using the guidance in a collaborative group setting. The simulated real-time cases offered hands-on WoFS experience that could then be reflected on during the focus group sessions. In a post-experiment survey, all 16 participants either agreed or strongly agreed that the simulations helped them to develop an understanding of how WoFS guidance can be applied in operations. They appreciated that the simulations mimicked operations yet were conducted in a risk-free environment with minimal distraction. Participants also enjoyed having time to explore the guidance and learn through trial and error about how to best interject WoFS guidance into their shared workflows on the watch to warning timeline.

Workflow was captured in the pre- and post-experiment flowchart tasks, which was then used to facilitate discussion in the first focus group. All but one participant reported that they either agreed or strongly agreed that the pre-experiment flowchart task was helpful for thinking about and sharing their forecast and decision-making process. Specifically, participants described that this task enabled them to self-reflect and become more self-aware of what they usually do on "autopilot." Analyzed together, forecast process steps representative of the participating group of forecasters in this experiment included the following:

- Initial assessment of the ongoing situation and implementation of the forecast funnel approach.
- A three-phase cycle of analysis, forecast, and assessment.
- Determination of necessary action steps, resulting in internal or external collaboration and communication of expectations.

For the post-experiment flowchart, four participants reported that they did not expect WoFS would impact their workflow because they treated it just like other CAMs. This feedback highlights the importance of educating future WoFS users on what makes this system unique relative to other convection allowing ensemble prediction systems; namely, the ability of rapidly cycled radar and satellite assimilation to produce accurate initial conditions of individual convective storms, which translates to more specific short-term probabilistic predictions of storm mode, track, and associated hazards (Guerra et al. 2022). In other words, WoFS has been designed to provide unique guidance, with greater spatial and temporal specificity, compared to other CAM forecast systems (Heinselman et al. 2023).

The numerous identified potential applications of WoFS for enabling the forecast process is encouraging. Many of the shared ideas overlap the responsibilities of WFOs and national centers, demonstrating the possibility for future enhancements to cross-office collaboration. Although use of WoFS is not necessarily resulting in a timeline shift for when products are issued, WoFS is allowing forecasters to be more specific and confident within their processes and is subsequently enhancing capabilities within the current operational framework. Some noteworthy examples of enhanced capabilities documented in this experiment include the following:

- Improved coordination by using WoFS as a common reference point within and across NWS offices and centers.
- Increased consistency in hazard expectations and messaging within and across NWS offices and centers.
- Enhanced messaging in products, including hazard magnitude and likelihood of a hazard occurring.
- Higher levels of situation awareness resulting in improved operations planning.

This experiment was also important for documenting the challenges participants believe already exist or anticipate existing for a successful implementation of WoFS into operations. As described in the results, discussions of many of these challenges have arisen from prior real time and testbed research, and so this experiment further underscored the importance of addressing these challenges. Potential issues focus on preparation for understanding WoFS (i.e., knowledge and training), real-time use of WoFS (i.e., trust, time, and data overload), and integration of WoFS into existing operations (i.e., assigning responsibility to WoFS and embracing the changes that WoFS can bring to the forecast process). Solutions have been proposed or are already underway, including a series of training webinars in 2023 that were attended by 99 different groups of NWS participants. Each solution will take WoFS one step closer to transition, and in the meantime, improve the capabilities of WoFS for experimental use within the operational community.

In a final reflection of this experiment, all participants agreed or strongly agreed that participating with forecasters from local offices and national centers gave them new insight into other's roles and responsibilities in the NWS. They shared that this experiment allowed them to observe and better understand other forecasters' operational processes and related challenges. This feedback highlights the importance of collaborative exploration of new forecasting capabilities, where forecasters can gain awareness of new technologies as well as the resulting impacts on the more expansive and human-led forecast process. For WoFS-related research, future studies should continue to implement a collaborative approach and expand these data collection efforts to a larger and more representative sample of NWS forecasters.

Acknowledgments. This experiment was possible due to the contributions of many individuals. We are grateful to the 16 NWS forecasters for their participation, to Laren Reynolds and Joseph Merchant for their supporting roles during the pilot and full experiment, and to Randy Bowers for preparing the forecast briefing videos. We are also thankful to the leadership at NOAA/NSSL, NWS Southern Region Headquarters, SPC, and WPC for their insightful discussions, and to Pam Heinselman, Adam Clark, and three AMS reviewers for their helpful feedback on this manuscript. Funding for this work was supported by the NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreements NA16OAR4320115, U.S. Department of Commerce.

Data availability statement. The simulation and focus group recordings, transcriptions, and content analyses are archived in password-protected cloud storage. Following the University of Oklahoma's human subject research guidelines and confidentiality agreements, these data nor data disclosing information about the experiment participants cannot be made available.

APPENDIX

Semi-Structured Focus Group Protocol

a. Focus group 1

Key questions: How do forecasters envision Warn-on-Forecast guidance fitting into their existing forecast process?

- Think about your existing forecast process, i.e., the initial flowchart you created.
- During these questions, compare your experiences in these simulations, and in particular the impact that WoFS guidance had, to your experiences working in normal operations.

1) FLOWCHART TASK

For your role, take your flowchart, copy and paste it into a new slide. Next, edit the aspects that you believe would be impacted by the availability of WoFS guidance. You may edit the flowchart, choose to focus on just a piece of the flowchart, or zoom in and focus on modifying a particular part of the process. Try to separate out the experimental aspects or constraints that were in place during the simulation and think about overall real-time workflow.

2) DISCUSSION ON WORKFLOW

[Show and explain flowcharts to illustrate ideas.]

1) In what ways did the availability of WoFS guidance alter your forecast processes as depicted in your flowcharts?
2) How would the flowchart alterations have impacted other forecast roles in your offices during the two events?

b. Focus group 2

Key questions: How can WoFS guidance be utilized most effectively across the current watch-to-warning forecast process?

- Think about your existing forecast process i.e., the initial flowchart you created.

- During these questions, compare your experiences of working together and the impact that WoFS guidance had on your experiences working in normal operations.

1) VENN DIAGRAM TASK

[Use the virtual whiteboard to answer the following questions.]

1) Based on your experiences in the two cases and our focus group this morning, can you describe what you believe to be some of the most useful applications of WoFS guidance for local offices and for national centers?
2) What are the challenges for utilizing WoFS guidance at local offices and national centers?

2) DISCUSSION ON APPLICATIONS AND CHALLENGES

1) Looking at the whiteboard, what are your takeaways regarding the separation versus overlap of WFOs versus National Centers?
2) How might WoFS guidance impact the existing flow of information from watches to warnings?
3) What will this impact of information flow mean for (i) NWS personnel and (ii) users outside of your NWS offices?
4) What personal or technological infrastructure exists or needs to be created to support suggestions made during this focus group?

c. Focus group 3

Key Questions: How could the availability of WoFS guidance fit into a visionary forecast process? What if the constraints of the existing forecast process were no more?

1) STICKY NOTE TASK

1) Using the sticky notes on the virtual whiteboard, note down what words come to mind when you think about how your role as a forecaster will evolve over the next decade.
2) What does this evolution mean for your use of WoFS guidance?

2) DISCUSSION ON THE VISIONARY FORECAST PROCESS

[Review sticky notes as a group.]

1) In what ways do you envision technology changing and what could these changes mean for use of WoFS guidance?
2) What are some of the experimental forecast guidance, products, or ideas that you hope will become operational, and what might they mean for use of WoFS guidance?
3) What are some of the limitations or constraints of your existing forecast process that might impact your desired use of WoFS? What would it take to remove these limitations or constraints?
4) What will it take for WoFS guidance to be implemented into your NWS office successfully in the next 5–10 years?

REFERENCES

Ayd, P., and R. A. Graham, 2023: WoFS in NWS Central Region remote mesoanalysis operations. *Special Symp. on Forecasting a Continuum of Environmental Threats (FACETs)*, Denver, CO, Amer. Meteor. Soc., 7.2, https://ams.confex.com/ams/103ANNUAL/meetingapp.cgi/Paper/410621.

Bieda, S. W., and Coauthors, 2022: A multioffice Warn-on-Forecast System R2O2R working group: A Science and Operations Officer's perspective. *12th Conf. on Transition of Research to Operations*, Houston, TX, Amer. Meteor. Soc., 12B.1, https://ams.confex.com/ams/102ANNUAL/meetingapp.cgi/Paper/391740.

Burke, P. C., and Coauthors, 2022: Collaborating to increase warning lead time using the Warn-on-Forecast System. *30th Conf. on Severe Local Storms*, Santa Fe, NM, Amer. Meteor. Soc., 44, https://ams.confex.com/ams/30SLS/meetingapp.cgi/Paper/407184.

Calhoun, K. M., K. L. Berry, D. M. Kingfield, T. Meyer, M. J. Krocak, T. M. Smith, G. Stumpf, and A. Gerard, 2021: The experimental warning program of NOAA's Hazardous Weather Testbed. *Bull. Amer. Meteor. Soc.*, **102**, E2229–E2246, https://doi.org/10.1175/BAMS-D-21-0017.1.

Daipha, P., 2015: *Masters of Uncertainty: Weather Forecasters and the Quest for Ground Truth*. The University of Chicago Press, 271 pp.

Demeritt, D. H., H. Cloke, F. Pappenberger, J. Thielen, J. Bartholmes, and M.-H. Ramos, 2007: Ensemble predictions and perceptions of risk, uncertainty, and error in flood forecasting. *Environ. Hazards*, **7**, 115–127, https://doi.org/10.1016/j.envhaz.2007.05.001.

Demuth, J. L., B. Morrow, and J. J. Lazo, 2009: Weather forecast uncertainty information. *Bull. Amer. Meteor. Soc.*, **90**, 1614–1618, https://doi.org/10.1175/2009BAMS2787.1.

——, and Coauthors, 2020: Recommendations for developing useful and usable convection-allowing model ensemble information for NWS forecasters. *Wea. Forecasting*, **35**, 1381–1406, https://doi.org/10.1175/WAF-D-19-0108.1.

Dowell, D. C., and Coauthors, 2022: The High-Resolution Rapid Refresh (HRRR): An hourly updating convection-allowing forecast model. Part I: Motivation and system description. *Wea. Forecasting*, **37**, 1371–1395, https://doi.org/10.1175/WAF-D-21-0151.1.

Doyle, E. E. H., D. M. Johnston, R. Smith, and D. Paton, 2019: Communicating model uncertainty for natural hazards: A qualitative systematic thematic review. *Int. J. Disaster Risk Reduct.*, **33**, 449–476, https://doi.org/10.1016/j.ijdrr.2018.10.023.

Flora, M. L., C. K. Potvin, P. S. Skinner, S. U. Handler, and A. McGovern, 2021: Using machine learning to generate storm-scale probabilistic guidance of severe weather hazards in the Warn-on-Forecast System. *Mon. Wea. Rev.*, **149**, 1535–1557, https://doi.org/10.1175/MWR-D-20-0194.1.

Gallo, B. T., and Coauthors, 2017: Breaking new ground in severe weather prediction: The 2015 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment. *Wea. Forecasting*, **32**, 1541–1568, https://doi.org/10.1175/WAF-D-16-0178.1.

——, and Coauthors, 2022: Exploring the watch-to-warning space: Experimental outlook performance during the 2019 Spring Forecasting Experiment in NOAA's Hazardous Weather Testbed. *Wea. Forecasting*, **37**, 617–637, https://doi.org/10.1175/WAF-D-21-0171.1.

Graham, R. A., 2023: Remote mesoanalyst support—A mutual and crowdsourcing success story. *13th Conf. on Transition of Research to Operations*, Denver, CO, Amer. Meteor. Soc., 11B.1, https://ams.confex.com/ams/103ANNUAL/meetingapp.cgi/Paper/413009.

Guerra, J. E., P. S. Skinner, B. Matilla, M. Flora, A. Clark, K. Knopfmeier, and A. Reinhart, 2022: Quantification of NSSL Warn-on-Forecast System skill by storm age using object-based verification. *Wea. Forecasting*, **37**, 1973–1983, https://doi.org/10.1175/WAF-D-22-0043.1.

Heinselman, P. L., and Coauthors, 2023: Warn-on-Forecast System: From vision to reality. *Wea. Forecasting*, **39**, 75–95, https://doi.org/10.1175/WAF-D-23-0147.1.

Houston, A. L., J. C. Walther, L. M. Pytlikzillig, and J. Kawamoto, 2020: Initial assessment of unmanned aircraft system characteristics required to fill data gaps for short-term forecasts: Results from focus groups and interviews. *J. Oper. Meteor.*, **8**, 111–120, https://doi.org/10.15191/nwajom.2020.0809.

Kingfield, D. M., and M. A. Magsig, 2009: Leveraging National Weather Service technology for collaboration and training. *25th Conf. on Int. Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, Phoenix, AZ, Amer. Meteor. Soc., P1.9, https://ams.confex.com/ams/pdfpapers/150678.pdf.

Liamputtong, P., 2011: *Focus Group Methodology: Principles and Practice*. SAGE Publications Ltd., 213 pp.

Loken, E. D., K. A. Wilson, T. Sandmael, A. J. Clark, K. M. Calhoun, A. E. Reinhart, P. Skinner, and P. C. Burke, 2022: Improving probabilistic watch-to-warning severe hazard guidance by merging the Warn-on-Forecast System with observations-based products using machine learning. *30th Conf. on Severe Local Storms*, Santa Fe, NM, Amer. Meteor. Soc., 42, https://ams.confex.com/ams/30SLS/meetingapp.cgi/Paper/407634.

Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307, https://doi.org/10.3402/tellusa.v21i3.10086.

Martin, J., and Coauthors, 2023: Migrating NSSL's Warn-on-Forecast System (WoFS) to a cloud-based architecture. *Ninth Symp. on High Performance Computing for Weather, Water, and Climate*, Denver, CO, Amer. Meteor. Soc., 4.1, https://ams.confex.com/ams/103ANNUAL/meetingapp.cgi/Paper/419831.

Martinaitis, S. A., and Coauthors, 2022: A path toward short-term probabilistic flash flood prediction. *Bull. Amer. Meteor.*, **104**, E585–E605, https://doi.org/10.1175/BAMS-D-22-0026.1.

McHugh, M. L., 2012: Interrater reliability: The kappa statistic. *Biochem. Med.*, **22**, 276–282, https://doi.org/10.11613/BM.2012.031.

Murphy, A. H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281–293, https://doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2.

Neuendorf, K. A., 2018: Content analysis and thematic analysis. *Advanced Research Methods for Applied Psychology: Design, Analysis, and Reporting*, P. Brough, Ed., Routledge, 221–223.

Roberts, B., I. L. Jirak, A. J. Clark, S. J. Weiss, and J. S. Kain, 2019: Post-processing and visualization techniques for convection-allowing ensembles. *Bull. Amer. Meteor. Soc.*, **100**, 1245–1258, https://doi.org/10.1175/BAMS-D-18-0041.1.

Rothfusz, L. P., R. Schneider, D. Novak, K. Klockow-McClain, A. E. Gerard, C. Karstens, G. J. Stumpf, and T. M. Martin, 2018: FACETs: A proposed next-generation paradigm for high-impact weather forecasting. *Bull. Amer. Meteor. Soc.*, **99**, 2025–2043, https://doi.org/10.1175/BAMS-D-16-0100.1.

Skinner, P. S., and Coauthors, 2023a: Efficient R2O2R for the Warn-on-Forecast System using a real time NWS chat room. *Special Symp. on Forecasting a Continuum of Environmental Threats (FACETS)*, Denver, CO, Amer. Meteor. Soc., 5.2, https://ams.confex.com/ams/103ANNUAL/meetingapp.cgi/Paper/421548.

——, and Coauthors, 2023b: Interpreting guidance from the Warn-on-Forecast System. Part I: Review of probabilistic guidance concepts, product design, and best practices. *J. Oper. Meteor.*, **11**, 110–131, https://doi.org/10.15191/nwajom.2023.1109.

Stensrud, D. J., and Coauthors, 2009: Convective-scale Warn-on-Forecast System: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1500, https://doi.org/10.1175/2009BAMS2795.1.

Wilson, K. A., P. L. Heinselman, and C. M. Kuster, 2017: Considerations for phased-array radar data use within the National Weather Service. *Wea. Forecasting*, **32**, 1959–1965, https://doi.org/10.1175/WAF-D-17-0084.1.

——, ——, P. S. Skinner, J. J. Choate, and K. E. Klockow-McClain, 2019: Meteorologists' interpretations of storm scale ensemble-based forecast guidance. *Wea. Climate Soc.*, **11**, 337–354, https://doi.org/10.1175/WCAS-D-18-0084.1.

——, B. T. Gallo, P. S. Skinner, A. Clark, P. Heinselman, and J. J. Choate, 2021: Analysis of end user access of Warn-on-Forecast guidance products during an experimental forecasting task. *Wea. Climate Soc.*, **13**, 859–874, https://doi.org/10.1175/WCAS-D-20-0175.1.

——, and Coauthors, 2023: The NOAA Weather Prediction Center's use and evaluation of experimental Warn-on-Forecast System guidance. *J. Oper. Meteor.*, **11**, 82–94, https://doi.org/10.15191/nwajom.2023.1107.

Yussouf, N., K. A. Wilson, S. M. Martinaitis, H. Vergera, P. L. Heinselman, and J. J. Gourley, 2020: The coupling of NSSL Warn-on-Forecast and FLASH systems for probabilistic flash flood prediction. *J. Hydrometeor.*, **21**, 123–141, https://doi.org/10.1175/JHM-D-19-0131.1.