Wilson, K. A., N. Yussouf, P. L. Heinselman, A. Orrison, R. Otto, P. S. Skinner, K. Knopfmeier, B. Matilla, and M. Erickson, 2023: The NOAA Weather Prediction Center's use and evaluation of experimental Warn-on-Forecast System guidance,. *J. Operational Meteor.,* **11** (7), 82–94, doi: https://doi.org/10.15191/nwajom.2023.1107.

Journal of Operational Meteorology
Article

The NOAA Weather Prediction Center's Use and Evaluation of Experimental Warn-on-Forecast System Guidance

KATIE A. WILSON
Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma
NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma
RAND Corporation, Santa Monica, CA

NUSRAT YUSSOUF, PATRICK S. SKINNER, KENT KNOPFMEIER, BRIAN C. MATILLA
Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma
NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

PAMELA. L. HEINSELMAN
NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

ANDREW ORRISON, RICHARD OTTO
NOAA/NCEP/Weather Prediction Center

MICHAEL ERICKSON
NOAA/NCEP/Weather Prediction Center
Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder

(Manuscript received 4 March 2023; review completed 17 April 2023)

ABSTRACT

This study examines use of experimental Warn-on-Forecast System (WoFS) guidance for short-term flash flood prediction at the NOAA Weather Prediction Center's Meteorological Watch (Metwatch) desk. The WoFS guidance provides storm-scale ensemble forecasts for individual thunderstorms out to six hours and has previously shown great promise in its predictive skill for heavy rainfall events. Its operational utility was examined during 2019 and 2020 in a formal collaboration between Warn-on-Forecast scientists and Metwatch meteorologists. During that time, Metwatch meteorologists integrated real-time WoFS guidance into their Mesoscale Precipitation Discussion forecast processes and provided evaluations via a post-event survey. The survey queried impacts of WoFS guidance on their situational awareness, workload, and confidence, and Metwatch meteorologists also reported subjective assessments of model performance. Survey results highlighted the importance of viewing consistency in WoFS guidance across runs and agreement between WoFS guidance with conceptual models, other numerical weather prediction guidance, and observations. The use of WoFS tended to either maintain or slightly increase Metwatch meteorologists' workload, while also increasing their confidence (notably for events perceived as better predicted). Of the different forecast attributes evaluated, Metwatch meteorologists reported convective mode as the attribute best predicted by WoFS. Use of WoFS guidance supported Mesoscale Precipitation Discussion decision making, including the placement and spatial extent of the product and the level of specificity provided about the related flash flood threat(s).

1. Introduction

Heavy rainfall events that produce flooding pose a significant threat to life and are responsible for the second highest number of weather-related fatalities each year, with heat-related fatalities being most common (Ashley and Ashley 2008; NOAA 2021). Flash floods are defined as events that occur within six hours of the causative event (e.g., intense rainfall), though in some circumstances can occur in minutes (NOAA 2022a). The rapid onset and highly localized nature of flash floods make this hazard especially challenging to forecast, and can lead to very dangerous situations with little time to respond effectively. Flash flood fatalities most often occur when individuals are in their vehicles or engaged in outdoor recreations (Terti et al. 2017). The provision of forecast information that better helps individuals anticipate a flash flood is

Corresponding author address: Katie Wilson, 120 David L Boren Blvd., Norman, OK 73072
E-mail: katiew@rand.org

therefore crucial for enabling improved preparedness and response to these events.

National Weather Service (NWS) meteorologists at local Weather Forecast Offices (WFOs), River Forecast Centers (RFCs), and the Weather Prediction Center (WPC) support the nation's operational prediction of flash flooding (Burke et al. 2022). Local WFO meteorologists focus on their designated County Warning Areas and issue flash flood watches, warnings, and advisory products. The RFCs provide hydrologic guidance for timescales ranging from hours to months for hazards including flooding, snowmelt, and water supply (NOAA 2022b). At the national center level, WPC forecasts precipitation across the United States and issues Quantitative Precipitation Forecasts, Excessive Rainfall Outlooks (Erickson et al. 2021; Burke et al. 2022), and Mesoscale Precipitation Discussions (MPDs). The WPC Meteorological Watch (Metwatch) Desk is responsible for issuing the MPDs to alert, inform, and collaborate with WFO meteorologists for regions that are likely to be impacted by flash flooding within the next six hours. During the MPD decision-making process, Metwatch meteorologists use a combination of observational data (e.g., radar, satellite, and surface and upper air observations) and numerical weather prediction guidance to determine the flash flood threat. Mesoscale models [e.g., the Rapid Refresh (Benjamin et al. 2016)] are used to assess atmospheric instability and moisture trends, and high-resolution convection-allowing models [e.g., High-Resolution Rapid Refresh and the High-Resolution Ensemble Forecast (Dowell et al. 2022 and Kalina et al. 2021)] predict convective systems and their storm attributes (e.g., composite reflectivity, updraft helicity, and accumulated rainfall).

Although the predictability of extreme precipitation in these models is limited by its small-spatiotemporal scales, research efforts to explore storm-scale ensemble systems hold promise for more accurate forecasts of rainfall location, timing, and intensity (Schumacher 2017). One experimental ensemble system under active development is the NOAA National Severe Storms Laboratory Warn-on-Forecast System (WoFS). WoFS is an 18-member convective-scale ensemble analysis and forecast system with 3-km horizontal grid spacing, and provides probabilistic forecasts of individual storm hazards over the next 0–6 hr (e.g., Jones et al. 2016; Skinner et al. 2018; Yussouf et al. 2020a, b). During the period of this study, WoFS used the Advanced Weather Research and Forecast (ARW-WRF core) dynamic solver. The frequent assimilation of observations (every 15 min) in WoFS results in more rapidly updating storm-scale probabilistic guidance than what is available with current operational convection-allowing models (Yussouf et al. 2016; Lawson et al. 2018). The predictive accuracy of the system improves with storm age, such that older storms have been included in several data assimilation cycles and are forecast with greater accuracy (Guerra et al. 2022). The WoFS guidance is expected to enhance situational awareness and support the forecast decision-making process. In particular, the spatiotemporal scale of WoFS (covering a 900-km square domain out to six hours) and its demonstrated ability to provide skillful forecasts for short-term heavy rainfall events (Yussouf et al. 2016; Yussouf and Knopfmeier 2019; Yussouf et al. 2020a,b; Martinaitis et al. 2022) suggest that WoFS could provide important guidance to the WPC Metwatch Desk.

To explore potential applications of WoFS guidance for the Metwatch Desk's flash flood forecasting, WoFS scientists and WPC meteorologists committed to a multiyear joint research and operations collaboration. This collaboration began with WoFS scientists visiting WPC in 2017 and 2018 to share information on the Warn-on-Forecast program, build relationships with Metwatch meteorologists, and learn about Metwatch Desk responsibilities and workflows. Real-time WoFS guidance was made available to the Metwatch Desk via a web viewer in 2018, which enabled an exploration of what a "concept of operations" with an NWS national center could look like, an assessment of the type of WoFS guidance that was most helpful to Metwatch meteorologists, and feedback on ways to improve the product suite and display. WPC feedback resulted in the development of new WoFS products, including probability of exceedance for additional accumulated rainfall thresholds, the median percentile value, and the addition of environmental parameters such as MUCAPE (Most Unstable Convective Available Potential Energy) and midlevel wind and dewpoint maps.

The success of the 2018 WoFS real-time demonstration in the WPC Flash Flood and Intense Rainfall experiment (Barthold et al. 2015) resulted in a funded collaboration between WoFS scientists and Metwatch meteorologists. This collaboration ensured that Metwatch meteorologists would have access to real-time WoFS guidance during the summer months of

2019 and 2020 as well as enabled a formalized evaluation of WoFS utility for the MPD decision-making process. Metwatch meteorologists’ reflection on their use of this guidance was captured in a post-case evaluation survey, and addressed the following research questions: (1) What WoFS guidance information do Metwatch meteorologists seek when forecasting heavy rainfall events, and how does this information impact their MPD forecast process?; (2) How does integration of WoFS guidance impact Metwatch meteorologists’ workload and confidence during the MPD forecast process?; and (3) How do Metwatch meteorologists rate the performance of different WoFS guidance attributes, and for what reasons are different ratings chosen? These research questions were established based on prior interactions with Metwatch meteorologists and were designed to address both meteorological and human factor aspects of the forecast process. The investigation of these questions within the operational forecast environment enabled an assessment of WoFS guidance usage during real-time operations, where the broader responsibilities and challenges of day-to-day forecast duties could also be considered. This paper discusses findings related to each of these research questions, with a focus on Metwatch meteorologists' situational awareness, workload, and confidence resulting from usage of WoFS guidance, and their subjective perceptions of specific WoFS guidance attributes.

2. Methods

a. *Real-time WoFS runs*

During the 2019 and 2020 data collection seasons, real-time WoFS runs were provided approximately from May through September, including daily runs Monday through Friday during May and early June as part of the five-week NOAA Hazardous Weather Testbed Spring Forecasting Experiment (SFE; Gallo et al. 2017; Clark et al. 2020). For each daily WoFS run, testbed leaders chose the domain center point to focus on severe hail, severe wind, and tornado hazards. The chosen domain was oftentimes collocated with a flash flood risk, and therefore also served Metwatch Desk forecast responsibilities.

Beyond the annual SFE periods, the WoFS-WPC collaboration drove the placement of real-time WoFS runs. The WPC Senior Branch Forecaster (SBF) was tasked with sending a WoFS run request the day before an event. Initial requests were anticipated based on the Day 2 Excessive Rainfall Outlook. The SBF then made a final request early on the day of the event after the issuance of the new Day 1 Excessive Rainfall Outlook. This request included the preferred domain center point and start time for that day’s WoFS run. Real-time runs were prioritized for locations forecasted to have at least a slight risk (defined as 10%–20% at the time of this evaluation) of rainfall exceeding flash flood guidance (FFG) within 25 miles of a point; however, because of the sparse occurrence of Southwest Monsoon events in 2019, WoFS also ran in this region in 2020 for events predicted to have a marginal risk (5%–10%). For the combined 2019 and 2020 May through September periods, WoFS ran in real time for 67 separate events. The majority of these events were located in the Great Plains region, with the remaining events located in the Carolinas, Southeastern United States, and Four Corners Region (Fig. 1).

On the day of each event, WoFS was initialized as early as 1500 UTC with High-Resolution Rapid Refresh Ensemble (Dowell et al. 2022) boundary conditions. The first forecast was launched after one (two) hours of cycling in 2019 (2020), and six-hour forecasts were provided every hour until 0600 UTC. The WoFS guidance was made available via a real-time web viewer (https://wof.nssl.noaa.gov/realtime/). Forecasts loaded gradually on the web viewer, with each full six-hour forecast available after an approximate 40-min latency. Metwatch meteorologists could view any of the WoFS guidance available on the web viewer, though a group of products focused on rainfall prediction were listed in a “QPF” (Quantitative Precipitation Forecast) drop-down menu for simplicity. These products included the exceedance probability and percentile information for accumulated rainfall for different neighborhood sizes, durations, and thresholds. All other products, including WoFS guidance for environmental and storm attributes, were available in additional drop-down menus.

b. *Survey design*

The 2019–2020 survey was designed in tandem with two visits to WPC in 2018 and 2019. During these visits, two WoFS research scientists shadowed multiple Metwatch meteorologists on the operations floor to learn about their workflow and most used forecast guidance. Initial impressions of real-time WoFS guidance were gathered in a preliminary survey during

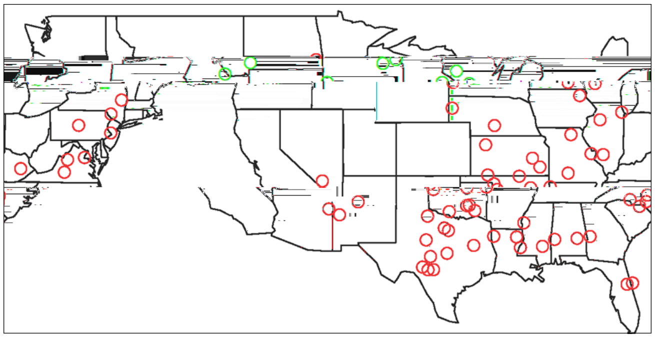

Figure 1. The center point locations for the WoFS 900-km-square domain during May–September 2019 and 2020 real-time run periods. *Click image for an external version; this applies to all figures and animations hereafter.*

the 2018 season. As part of a follow-on visit in 2019, WoFS scientists and lead Metwatch meteorologists worked together to modify survey questions for the official 2019–2020 data collection seasons. These modifications were made to do the following: (1) improve survey language to better reflect how Metwatch meteorologists describe aspects of their forecast process, (2) apply a specific framework to examine WoFS impacts on situational awareness, and (3) subjectively evaluate the separate attributes of WoFS performance through both quantitative ranking and qualitative reasoning.

The 2019–2020 survey questions were grouped into specific topics: forecast set-up, situational awareness, workload and confidence, and model performance ratings (see Appendix for all but forecast set-up). Forecasters were first asked to describe the forecast set-up to provide background and context for each event. The impact of WoFS guidance on their situational awareness was then queried using the Endsley (1995) model for dynamic decision making (Fig. 2). The suitability of the Endsley (1995) situational awareness model was determined based on its past successful application in operational meteorology research (Bowden and Heinselman 2016) and its theoretical focus on dynamically evolving events. This model describes three levels of situational awareness: perception, comprehension, and projection. An individual's projection informs subsequent decisions and actions. Specifically, Endsley (1995) describes situational awareness as "… *the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and a projection of their status in the near future.*" The

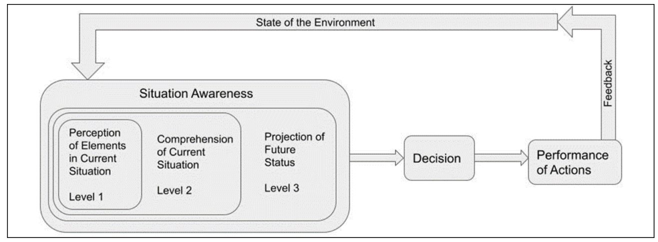

Figure 2. Simplified version of the Endsley (1995) model of situation awareness in dynamic decision making.

accumulating knowledge of the state of the environment is then fed back into the process of acquiring information, making sense of it, and predicting what will happen next. Metwatch meteorologists' acquired information from the WoFS guidance and its impact on forecast rationale and subsequent decisions was thus queried, providing insight on how WoFS guidance could impact situational awareness during the MPD decision-making process.

After responding to the qualitative situational awareness questions, Metwatch meteorologists used a Likert rating scale (Likert 1932) to rate their level of workload and confidence resulting from use of WoFS guidance during the event. Both ratings used a five-point scale ranging from greatly decreased to greatly increased (see Appendix). They also used a Likert rating scale to evaluate the performance of five forecast attributes: location, coverage, timing, intensity, and convective mode. Performance was rated on a five-point scale ranging from very poor to excellent (see Appendix), with Metwatch meteorologists guiding the language choice of these particular ratings. Metwatch meteorologists then provided important contextual insight into their subjective ratings of workload, confidence, and model performance through qualitative reasoning with each numeric rating.

c. *Data collection and analysis*

The 2019–2020 survey was created in Google forms and distributed to Metwatch meteorologists on the WoFS web viewer and via internal email at WPC. Metwatch meteorologists used the real-time WoFS guidance when it was available and completed the survey toward the conclusion of their forecast shift. All participation was voluntary, and the survey was approved under the University of Oklahoma's Institutional Review Board (#13320). In total, N=85 survey responses were collected for the 67 real-time

WoFS runs conducted in 2019 and 2020, with some events having multiple responses from consecutive shifts (e.g., an evening shift followed by a night shift). Twelve Metwatch meteorologists provided these responses, of whom five provided one response, five provided between two and eight responses, and two provided the remaining 59 responses. These two meteorologists provided the bulk (70%) of the responses due to their lead Metwatch responsibilities (meaning that they were most often assigned to this forecast desk) and their official collaborative roles in this project.

The use of both open-ended questions and Likert-scale ratings resulted in a database that was predominantly qualitative with some numeric assessments. The lead author collated and thematically analyzed the qualitative responses to identify the most prominent uses of WoFS guidance. Thematic analysis involved data familiarization followed by the identification of recurrent ideas (or "codes") within each question, and then grouping of these ideas into broader themes. Following the Braun and Clarke (2006) guide to using thematic analysis, themes were identified using a theoretical approach, such that patterns identified within the data were closely related to the specific survey questions, and therefore directly address the broader objectives of this study. Interpretations and overall findings related to these themes are described in the results below. For the rating questions, frequency count and rating distributions were analyzed, and these results were interpreted alongside forecasters' reasoning responses. Findings from the qualitative analysis were shared, discussed, and validated with the two lead Metwatch meteorologists to ensure accurate and fair representation of their WoFS guidance use during real-time operations.

3. Results

a. *Situational awareness*

1) Acquiring information from the WoFS guidance

The integration of WoFS guidance into Metwatch meteorologists' forecast process and the subsequent impacts on situational awareness began with their perception of the information available. The survey queried what specific information Metwatch meteorologists acquired for each event, and respondents typically provided the names of the products that were most frequently used. Together, the probability and percentile products provided Metwatch meteorologists with information about the likelihood, potential severity, and location of heavy rainfall. The most frequently reported product was the probability of rainfall accumulation for a variety of thresholds, with greater than 2" accessed most, followed by greater than 1", 3", and then 0.5". The next most reported product was the percentile values for rainfall accumulation, with the 90th percentile (i.e., reasonable maximum) accessed most, followed by the 50th percentile (i.e., median) and the 100th percentile (i.e., ensemble maximum). To determine whether heavy rainfall would result in flash flooding, Metwatch meteorologists often compared the amount of rainfall depicted in the percentile products to official FFG exceedance values. Use of additional WoFS products further supported assessments of the current situation, with Metwatch meteorologists reporting a secondary focus on products including the following: simulated radar reflectivity (for individual member solutions), reflectivity paintball plots (for values exceeding 40 dBZ), simulated satellite imagery (both infrared and water vapor), and mixed-layer convective available potential energy (MLCAPE).

In addition to sharing what products were accessed, Metwatch meteorologists described two important aspects of WoFS guidance that influenced confidence and willingness to integrate it into their forecast process: *consistency and agreement*. Consistency related to the trends viewed in a series of WoFS runs. The WoFS guidance was considered to be consistent if either a stationary trend in the products indicated the same outcome from run-to-run, or a strengthening or weakening trend indicated either an increasing or a decreasing threat of heavy rainfall from run-to-run. For example, in one event a Metwatch meteorologist noted that, "*The consistent run-to-run output of an organized (albeit progressive) convective line, and some subtle wetter trends was enough for me to have confidence in issuing a MPD*." By contrast, consistency in trends also led to decisions to not issue an MPD, with one meteorologist explaining that, "*Ultimately, the trends seen cycle to cycle tended to reflect a weakening trend or at least an inconsistent trend in the rainfall intensity… This allowed me to make a decision to not issue a MPD for the area despite the very close trigger for an issuance with the initial runs*."

The extent of agreement of WoFS guidance with observations, other high-resolution model guidance, and their own conceptual models directly influenced how much weight Metwatch meteorologists were willing to give to the WoFS guidance during the forecast process. In particular, greater weight was given to WoFS guidance when it matched radar observations, predicted aspects of an event not well captured in other model guidance, and confirmed their prior conceptual model for how the event would evolve. One Metwatch meteorologist explained that "*The WoFS output increased my confidence in my conceptual thinking of convective evolution given better alignment with my thoughts compared to some of the other hi-res models available (HRRR, ARW, NAM_NEST)*." Tendencies to discard WoFS guidance during an event occurred most often when ongoing storms were not well predicted in the WoFS guidance, with timing and location errors noted most frequently.

2) Building comprehension of the current event

Once Metwatch meteorologists had acquired information from the WoFS guidance and decided whether to incorporate it into their forecast process, they then moved forward with developing a more accurate and updated understanding of the current event. The WoFS guidance informed their conceptual models of the event, including aspects such as storm motion, convective mode, and the potential for repeated or training convection over a given location. Metwatch meteorologists reported important aspects for assessing the likelihood of flash flooding, including how organized and widespread convection was, its expected duration and intensity, and the progressive nature of it. One Metwatch meteorologist noted that "*The information helped to confirm for me the idea that there would be organized convection that would produce very heavy rainfall rates. There was plenty of instability exhibited, and the output suggested that convection would grow upscale through the afternoon hours*." Similarly, and by comparison, WoFS guidance was used to confirm the lack of a flash flooding threat, such that, "*The probabilities reinforced the idea of minimal risk of rainfall even approaching FFG... and the composite reflectivity showed the evolution/mode of convection and steady propagation of the cells*."

Even when WoFS guidance was evidently not predicting all aspects of an event accurately, Metwatch meteorologists extracted helpful information that was well predicted, and in some cases, made mental adjustments to account for errors in timing, location, or intensity. One Metwatch meteorologist decided to issue an MPD after noting that, "*Despite the timing and location being off compared to observed thunderstorm development around 2200 UTC in southwest OK, WoFS was consistent with the probability of a small cluster of thunderstorms developing with at least a moderate probability (50+%) of exceeding 3 inches, which would have been in excess of area FFG*." The willingness and ability of Metwatch meteorologists to draw value from the WoFS guidance and then make forecast decisions, despite observed errors in the forecast products, is encouraging. The authors have observed in other interactions with NWS forecasters that this approach tends to be employed once a comfortable level of operational familiarity with WoFS guidance use is established.

3) Forecast expectations and MPD decision making

Forecast expectations were focused on whether or not flash flooding was likely to occur, and whether or not an MPD should be issued. At least one MPD was issued in 45 of the responses, no MPD was issued in 37 of the responses, and the MPD decision was unclear in the final three responses. These survey results therefore provide balanced representation of WoFS guidance usage for events that were and were not projected to produce flash flooding.

In instances when an MPD was issued, the comprehension and resulting confidence from use of WoFS guidance directly impacted numerous aspects of the MPD. First, WoFS probability values were used to inform whether the chance of flash flooding in the MPD header would be labeled as "possible" or "likely," which is traditionally a decision made subjectively. Next, the WoFS guidance provided Metwatch meteorologists with additional information to form their forecast rationale in the MPD text, including an enhanced ability to offer greater specificity. The percentile values, for example, enabled Metwatch meteorologists to more readily and confidently quantify rainfall potential and highlight which locations were likely to receive the highest amount of rainfall. Metwatch meteorologists then reported that WoFS guidance helped to guide the initial placement, spatial extent, and timing of the MPD, and it guided downstream modifications of the MPD (e.g., an update to include additional areas in another MPD).

Although some decisions to issue an MPD were clear, other events gave Metwatch pause. For example, although a flash flood threat was evident in some events, it did not always fully meet the MPD criteria because the threat was either contained to just the next couple of hours or it was too localized. Still, Metwatch meteorologists wished to communicate the threat to local WFOs. During these instances, WoFS guidance proved to be a useful tool for facilitating this collaboration between WPC and local WFOs. One example of a Metwatch meteorologist using WoFS guidance in this way was summarized in the following statement: "*Later in the shift, as the flash flood threat remained somewhat isolated, I used WoFS output to support collaboration with the local offices via chat to mention only an isolated flash flood threat for the early overnight hours in AZ.*"

Metwatch meteorologists reported that an MPD was not issued in nearly half of the survey responses, and WoFS guidance provided added confidence in these situations. In the same way that WoFS guidance was used to determine that a flash flood threat existed, it was also used to confirm low threats for heavy rain fall. For example, comparisons of the WoFS percentile values of accumulated rainfall to non-WoFS operational thresholds such as FFG were particularly helpful.

4) Case examples

To demonstrate how WoFS was used to build situational awareness and make MPD decisions, two representative examples highlighting how Metwatch meteorologists acquired WoFS information, made sense of it, and then formed expectations for flash flooding potential are described below.

Decision to issue an MPD. A Metwatch meteorologist used WoFS to forecast activity during the Southwest Monsoon on 24 July 2020. The main forecast challenge was assessing rainfall rates and coverage of cell activity. Using the 2000–2200 UTC WoFS runs, the Metwatch meteorologist assessed the 50th, 90th, and maximum percentiles of accumulated rainfall, probability of exceedance values (>0.5", 1", and 2"), and environmental information including the ensemble mean values for MLCAPE and precipitable water. The MLCAPE and precipitable water values "reinforced" how moisture and instability were "entrenched" with one another and supportive of convection. Additionally, the probability of exceedance and percentile products both depicted potential for heavy rainfall (Fig. 3). This series of information led to the Metwatch meteorologists' conclusion that flash flooding was likely and that an MPD was required. The WoFS guidance was also used to quantify the rainfall potential in the MPD text (Fig. 4). After the event, this Metwatch meteorologist reflected that WoFS "*handled the monsoonal convection (basically in real time) so well compared to the other CAM [Convection Allowing Model] solutions…WoFS guidance greatly enhanced my confidence in the forecast process for my shift*."

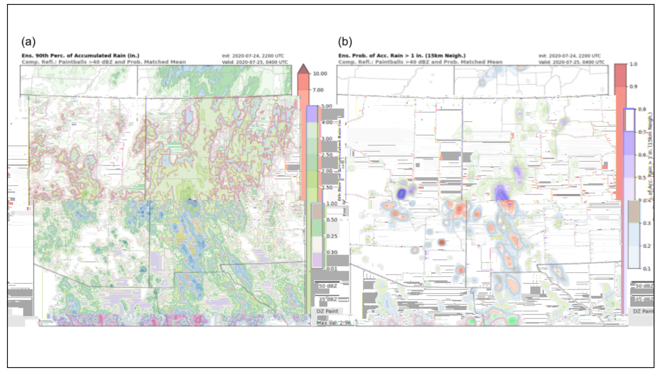

Figure 3. The ensemble 90th percentile of accumulated rainfall (a) and ensemble probability of accumulated rainfall >1" (b), both valid for the period 2200 UTC 24 July to 0400 UTC 25 July 2020.

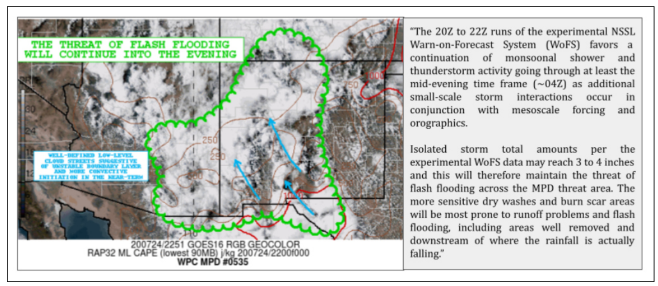

Figure 4. The MPD and associated WoFS-related text issued on 24 July 2020.

Decision to not issue an MPD. At the start of the Metwatch meteorologists' night shift on 26–27 July 2019, scattered convection was occurring in association with monsoonal activity. Multiple outflow boundaries were producing short-lived convection with heavy rain rates. The forecast challenge was determining how long convection would linger given decreasing instability and the limited training and longevity of storms. The Metwatch meteorologist viewed WoFS runs between 0000 and 0600 UTC and focused on a range of ensemble percentile values and probability of

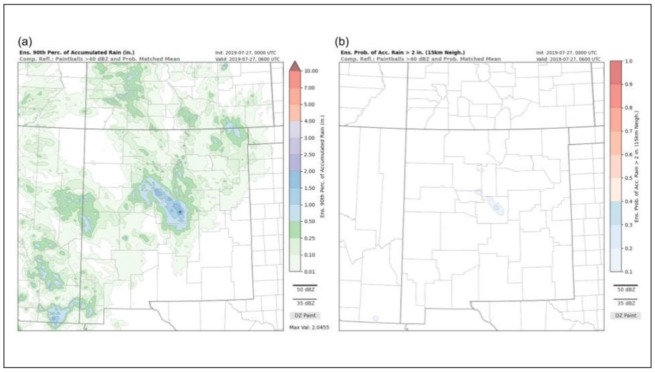

Figure 5. The ensemble 90th percentile of accumulated rainfall (a) and ensemble probability of accumulated rainfall >2” (b), both valid 0000–0600 UTC 27 July 2019.

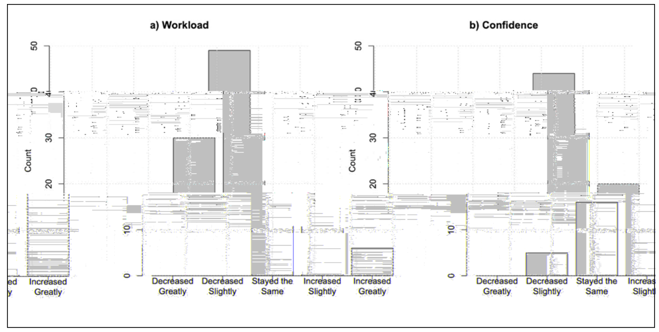

Figure 6. Summary of the surveyed forecaster (a) workload and (b) confidence ratings during 2019–2020 real-time WoFS guidance use.

exceedance values for accumulated rainfall (Fig. 5). The Metwatch meteorologist reported that although WoFS guidance showed 90th and 100th percentile values of 2”–3” of accumulated rainfall, the low probability associated with >2” of accumulated rainfall (up to 30%) were noteworthy.

Both the percentile and probability of exceedance information indicated a very limited and short-term flash flood threat, which corroborated the Metwatch meteorologist’s conceptual understanding of the event. The Metwatch meteorologist noted that “*My confidence was increased based on WoFS guidance and I was glad to see it matching my thinking, but I’m always cautious in the Southwest during the monsoon season*.” Although a decision to not issue an MPD was made, this Metwatch meteorologist used the WoFS guidance to communicate to the Albuquerque WFO that there was potential for flash flooding over the next few hours.

b. Workload and confidence

1) Workload

The incorporation of WoFS guidance into Metwatch meteorologists’ forecast processes had varying impacts on their workload (Fig. 6a). Most frequently, workload was reported to either increase slightly (*n*=49) or stay the same (*n*=30). In only a handful of cases (*n*=6) was workload reported to have increased greatly. The qualitative reasoning from the Metwatch meteorologists provides insight into why use of WoFS guidance resulted in differing impacts on workload.

When workload was reported to have stayed the same, three primary reasons were given. Metwatch meteorologists reported that they: (1) viewed WoFS as another piece of guidance and at times prioritized viewing it over other available high-resolution model guidance; (2) required only casual inspection of WoFS guidance to sufficiently gather the information they needed; and (3) adjusted their depth of WoFS guidance interrogation to a limited number of products during high-demand situations. Additional reasoning included having already built prior familiarity with the WoFS web viewer and products within, and having a slower-than-average (i.e., less workload) shift and thus time to view WoFS guidance.

For Metwatch meteorologists reporting an increase in workload, their reasoning centered on the extra time required to: (1) assess WoFS guidance, (2) compare it to observations and other high-resolution model guidance, and (3) mentally extract trends from multiple model runs. This increase in workload was noted especially when attempting to view every WoFS run, which requires a frequent investment in time and attention to keep up with the rapid-update design of WoFS. Additional reasons for increased levels of workload included: (4) a lack of familiarity with the WoFS guidance and web viewer; (5) the need to use multiple platforms to compare WoFS guidance to other data; and (6) inconvenient timing of WoFS guidance availability with forecast demands such as considering and/or constructing an MPD.

Despite experiencing an increase in workload, Metwatch meteorologists reported in a handful of events that this increase was minimized following their repeated use of it in operations. Specifically, they were better able to prioritize the most valuable WoFS guidance products and quickly synthesize the information gleaned into their conceptual models. This feedback aligns with the reasoning for events in which workload stayed the same, and suggests that strategies can be implemented to reduce workload impacts over time.

2) Confidence

The impact of WoFS guidance use on Metwatch meteorologists' confidence was recorded using the same Likert rating scale as for workload (Fig. 6b). For the majority of events captured in this survey (74%), confidence either slightly (*n*=44) or greatly (*n*=20) increased. For the remaining events, confidence either remained the same (*n*=16) or slightly decreased (*n*=5). In the five instances when Metwatch meteorologists' confidence decreased slightly, WoFS guidance was found to contradict conceptual models and/or other high-resolution model predictions. In two of these instances, these contradictions were also coupled with seeing that WoFS was not predicting ongoing convection well.

For the events in which use of WoFS guidance had no impact on Metwatch meteorologists' confidence, the reasoning was more varied. First, a lack of heavy rainfall threat meant that use of WoFS guidance in the forecast process did not impact confidence for some events. Second, confidence was not impacted if WoFS guidance provided the same information as other high-resolution model guidance. And finally, confidence was found to neither increase nor decrease if the WoFS guidance was perceived to have a mix of strengths and limitations for the same event (e.g., the location and timing were accurately predicted but the intensity was not perceived as accurate).

As discussed in Metwatch meteorologists' perceptions of WoFS guidance, agreement and consistency was highly valued. Slight increases in confidence were reported primarily due to agreement of WoFS guidance with conceptual models, other high-resolution model guidance, and real-time observations. Additionally, consistency in trends across a series of WoFS guidance runs increased Metwatch meteorologists' confidence. Slight increases in confidence also resulted from use of WoFS guidance during events in which Metwatch meteorologists were uncertain of how an event would evolve, thus the added information from WoFS guidance supported the forecast process.

Reasoning for Metwatch meteorologists' confidence increasing greatly from use of WoFS guidance overlapped with slight increases in confidence, though there was greater emphasis placed on events when it outperformed other high-resolution model guidance while strongly matching observations throughout an event. The overall impact on confidence was tied to Metwatch meteorologists' subjective impressions of the WoFS performance, with better-predicted events associated with greater increases in confidence.

c. Model performance

To capture the aspects of WoFS guidance that were well forecasted for each event, performance ratings were requested for separate attributes of the forecast (Fig. 7). Metwatch meteorologists rated WoFS forecasts of convective mode most highly, with a primary excellent rating. The primary rating for the four other attributes was good, with the next best forecasted attribute being intensity, followed by coverage, timing, and then location. Overall, the excellent rating was selected 176 times across the five attributes for the *N*=85 responses. These ratings include four days in which an excellent rating was given to all five of the forecast attributes. For one of these days that resulted in across-the-board excellent ratings, a Metwatch meteorologist explained that "*the higher probabilities of heavy rain (3") was close to observed heavy rain and flash flood warnings*," and the coverage, timing, and convective mode being "*very close*" or "*exact*" to observations.

By comparison, a very poor rating was assigned to a forecast attribute on 11 occasions. Though these ratings were comparably infrequent, they highlight instances in which WoFS guidance could not support the forecast process. One example in which WoFS performed very poorly was on 18 July 2019. The Metwatch meteorologist described a "*stationary boundary draped across the upper midwest with ripples of mid-level forcing moving overhead*." For this event, WoFS forecasted too much convection and in the wrong location. Four of the five attributes were rated as very poor, with only convective mode rated as acceptable. In this instance, the Metwatch meteorologist chose to discard the WoFS guidance entirely, and as a result, reported no impact of WoFS use on their workload or confidence.

The overall comparative analysis of cross-attribute average (median) performance ratings with workload and confidence ratings shows some interesting

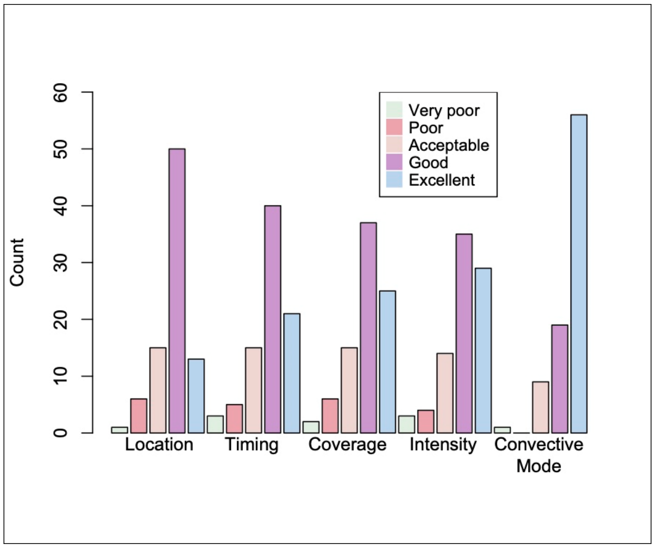

Figure 7. Performance ratings for the five WoFS guidance attributes.

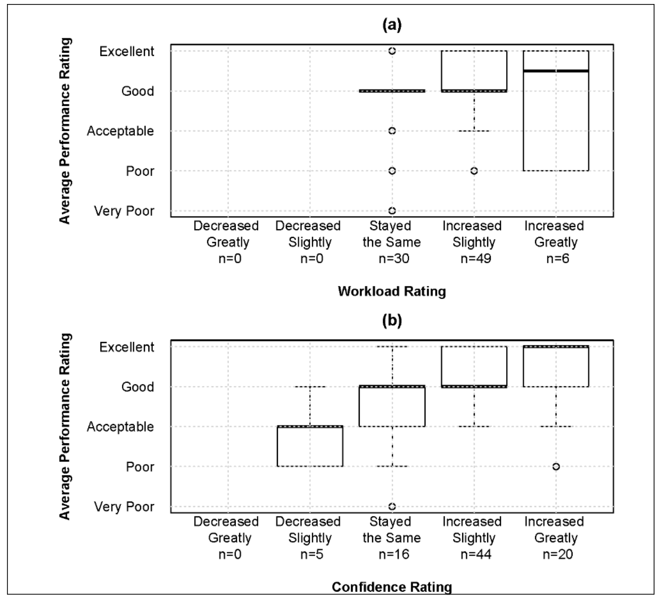

Figure 8. Distribution of cross-attribute average WoFS performance ratings compared to workload (a) and confidence ratings (b).

variation. Whereas the average performance rating did not deviate far beyond good (Fig. 8a), there was considerably more spread in performance ratings as workload impacts increased. Though a small sample size, those experiencing greatly increased workload from use of WoFS did so during events in which WoFS guidance was rated as good as well as when it was rated as poor. The impact of WoFS guidance on workload was therefore not solely a function of forecasters' perceptions of WoFS performance.

Compared to workload, a stronger tie was observed between WoFS performance and Metwatch meteorologists' confidence. A stepwise increase in average performance ratings with increased confidence is evident (Fig. 8b). Only a handful of instances occurred when confidence decreased slightly, and in those instances, WoFS performance was rated on average as acceptable, with the lower quartile extending below to poor (Fig. 8b). In contrast, confidence increased greatly when average WoFS performance was excellent. Performance was demonstrably more important for the impact on forecasters' confidence than it was for the impact on their workload.

4. Discussion and summary

A two-year longitudinal evaluation of Metwatch meteorologists' real-time use of WoFS has provided a unique data set to learn about the ways in which experimental storm-scale ensemble forecast guidance is used in short-term flash flood prediction. In collaboration with WPC, this real-time approach enabled an assessment of WoFS use alongside the opportunities and challenges that meteorologists encounter during operations. The realities of executing the decision-making process under time pressures, with competing responsibilities, deadlines, and inevitable distractions and disruptions, are impossible to replicate in a testbed setting.

The post-event online survey captured the experiences of Metwatch meteorologists as related to their situational awareness. In addition to recording which guidance products were most heavily used, Metwatch meteorologists highlighted the importance of run-to-run consistency and agreement between WoFS and their conceptual models, other numerical weather prediction models, and observations. The information acquired from WoFS guidance was important for building conceptual understanding of the event in real time, and Metwatch meteorologists demonstrated an ability and willingness to mentally adjust WoFS guidance in instances where it was not perfect but still deemed valuable. Further, WoFS guidance was reported to support MPD decision making, including the design of MPDs, such as the placement and spatial extent, the provided contextual specificity of the flash flood threat, and the accompanying forecast rationale.

The impacts of WoFS use on Metwatch meteorologists' workload and confidence were also documented. While WoFS usage tended to either keep workload the same or increase it slightly, it also tended to increase confidence. Qualitative reasonings, alongside ratings of WoFS performance, suggest that two different factors drove these experiences. First, Metwatch meteorologists' strategies for implementing WoFS into their workflows was influential to their workload, with greater prior WoFS experience resulting in more efficient use of WoFS and lower impact on workload. Second, Metwatch meteorologists' perceived performance of WoFS influenced changes to confidence, with better predicted events resulting in increases in forecast confidence. Increases in confidence were particularly notable in events when WoFS was perceived to have outperformed other available model guidance. Together, these findings demonstrate that human factors and meteorological drivers are both important to how WoFS guidance will influence the forecast process and experiences of NWS meteorologists.

The impact of WoFS use on decision accuracy was not quantified in this study for a couple of reasons. First, the real-time approach meant that setting up a control group to compare decision accuracy was not possible. A follow-on NOAA Hazardous Weather Testbed Warn-on-Forecast experiment was conducted in 2021 and will shed light on differences in MPD decision-making for those that do and do not have access to WoFS guidance. Second, as of the writing of this paper, MPD products are not objectively verified. Without such performance metrics, it is difficult to assess whether WoFS use results in more accurate or skillful MPDs. However, research is underway at WPC to develop objective performance metrics, and so it is likely that the impact of WoFS use can be assessed in an objective manner in the near future.

The ecological validity obtained through this real-time approach has resulted in a dataset and group of findings that are extremely representative of forecasters' real experiences, and is a notable strength of this study. This approach, however, also limited our ability to control the diversity of data collected, both in terms of the regional variety of weather events captured and the spread of feedback obtained from forecasters. For example, the inactive Southwest Monsoon season resulted in only a handful of WoFS cases in the western region. Additionally, although 12 Metwatch meteorologists used WoFS in real time and responded to the survey, 70% of the collected responses were from just two Metwatch meteorologists who frequently work the Metwatch desk (i.e., greater opportunity to use WoFS) and who were active collaborators on this project (i.e., more engaged and aware of opportunities to use WoFS).

These limitations are being addressed in our current and future research efforts through continued collaboration with WPC meteorologists (as well as with other NWS national centers and WFOs), and through an expanded ability to run WoFS for more events year-round. One recent advancement enabling this expanded real-time ability is the transition to cloud-based WoFS. This transition has resulted in a system that is simpler to run and is not dependent on research resources that have traditionally been shared with real-time experimental work. An increase in real-time runs will provide NWS meteorologists across the United States an opportunity to use WoFS for a greater variety of weather events and across more diverse locations. We anticipate this broadening in WoFS exposure and feedback will further inform our expectations for how WoFS guidance will aid the forecast process for different weather hazards.

Acknowledgments. We are grateful to the Weather Prediction Center's leadership and Metwatch meteorologists for committing to this operational WoFS test and evaluation. Many thanks to Steven Martinaitis, Kodi Berry, and Patrick Burke for providing reviews of this manuscript. Funding for this work was supported by the FY2018 Joint Technology Transfer Initiative Program within NOAA/OAR Weather Program Office under Grant No. NA18OAR4590361 and NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA21OAR4320204, U.S. Department of Commerce.

APPENDIX A

Survey Questions

Situational Awareness

1. Please list what information you specifically saw in the WoFS guidance (including names of products and corresponding model cycles).
2. Please explain how this information influenced your forecast rationale.
3. Please describe how this information impacted your forecast decisions. (Please provide links to any products you issued).

Workload and Confidence

Q1a. How did the use of WoFS guidance impact your overall workload level during this event? [Decreased greatly; Decreased slightly; Stayed the same; Increased slightly; Increased greatly]

Q1b. In 1–3 sentences, please describe what specific aspect of your workload was impacted by WoFS guidance and why.

Q2a. How did the use of WoFS guidance impact your overall confidence during this event? [Decreased greatly; Decreased slightly; Stayed the same; Increased slightly; Increased greatly]

Q2b. In 1–3 sentences, please describe how the WoFS guidance played a role in altering your confidence.

Subjective Evaluation of WoFS Performance

Q1. Please rate the performance of the WoFS guidance according to these attributes: location, coverage, timing, intensity, and convective mode. [Very poor; Poor; Acceptable; Good; Excellent]

Q2. In 1–2 sentences, why did you choose this rating for: location, coverage, timing, intensity, and convective mode?

REFERENCES

Ashley, S. T.and W. S. Ashley, 2008: Flood fatalities in the United States. *J. Appl. Meteor. Climatol*., **47**, 805–818, CrossRef.

Barthold, F. E., T. E. Workoff, B. A. Cosgrove, J. J. Gourley, D. R. Novak, and K. M. Mahoney, 2015: Improving flash flood forecasts: The HMT-WPC Flash Flood and Intense Rainfall Experiment. *Bull. Amer. Meteor. Soc*., **96**, 1859–1866, CrossRef.

Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. Mon. Wea. Review, 144, 1669–1694, CrossRef.

Bowden, K. A., and P. L. Heinselman, 2016: A qualitative analysis of NWS forecasters' use of phased-array radar data during severe hail and wind events. *Wea. Forecasting*, **31**, 43–55, CrossRef.

Braun, V., and V. Clarke, 2006: Using thematic analysis in psychology. *Qualitative Research in Psychology*, **3:2**, 77–101 CrossRef.

Burke, P. C., A. Lamers, G. Carbin, M. J. Erickson, M. Klein, M. Chenard, J. McNatt, and L. Wood, 2023: The excessive rainfall outlook at the Weather Prediction Center: Operational definition, construction, and real-time collaboration. *Bull. Amer. Meteor. Soc*., **104**, E542–E562, CrossRef.

Clark, A. J., and Coauthors, 2020: A real-time, simulated forecasting experiment for advancing the prediction of hazardous convective weather. *Bull. Amer. Meteor. Soc*., **101**, E2022–E2024, CrossRef.

Dowell, D., 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, CrossRef.

Endsley, M. R., 1995: Toward a theory of situation awareness in dynamic systems. *Human Factors*, **37**, 32–64, CrossRef.

Erickson, M. J., B. Albright, and J. A. Nelson, 2021: Verifying and redefining the Weather Prediction Center's Excessive Rainfall Outlook forecast product. *Wea. Forecasting*, **36**, 325–340, CrossRef.

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, CrossRef.

Guerra, J. E., P. S. Skinner, A. Clark, M. Flora, B. Matilla, K. Knopfmeier, and A. E. Reinhart, 2022: Quantification of NSSL Warn-on-Forecast System accuracy by storm age using object-based verification. *Wea. Forecasting*, **37**, 1973–1983, CrossRef.

Jones, T. A., K. Knopfmeier, D. Wheatley, G. Creager, P. Minnis, and R. Palikonda, 2016: Storm-scale data assimilation and ensemble forecasting with the NSSL Experimental Warn-on-Forecast System. Part II: Combined radar and satellite data experiments. *Wea. Forecasting*, **31**, 297–327, CrossRef.

Kalina, E. A., I. Jankov, T. Alcott, J. Olson, J. Beck, J. Berner, D. Dowell, and C. Alexander, 2021: A progress report on the development of the High-Resolution Rapid Refresh Ensemble. *Wea. Forecasting*, **36**, 791–804, CrossRef.

Lawson, J. R., J. S. Kain, N. Yussouf, D. C. Dowell, D. M. Wheatley, K. H. Knopfmeier, and T. A. Jones, 2018: Advancing from convection-allowing NWP to Warn-on-Forecast: Evidence of progress. *Wea. Forecasting*, **33**, 599–607, CrossRef.

Likert, R., 1932: A technique for the measurement of attitudes. *Archives of Psychology*, **22**, 55pp.

Martinaitis, S. A., and Coauthors, 2022: A path towards short-term probabilistic flash flood prediction. *Bull. Amer. Meteor.*, **104**, E585–E605, CrossRef.

NOAA, 2021: Weather related fatality and injury statistics. Accessed 29 August 2022, CrossRef.

____, 2022a: National Weather Service glossary. Accessed 29 August 2022, CrossRef.

____, 2022b: National Weather Service Jetstream. Accessed 14 November 2022, CrossRef.

Schumacher, R. S., 2017: Heavy rainfall and flash flooding. Invited contribution, Oxford Research Encyclopedia of Natural Hazard Science.

Skinner, P. S., and Coauthors, 2018: Object-based verification of a prototype Warn-on-Forecast System. *Wea. Forecasting*, **33**, 1225–1250, CrossRef.

Terti, G., I. Ruin, S. Anquetin, and J. J. Gourley, 2017: A situation-based analysis of flash flood fatalities in the United States. *Bull. Amer. Meteor. Soc.*, **98**, 333–345, CrossRef.

Yussouf, N., and K. H. Knopfmeier, 2019: Application of the Warn-on-Forecast system for flash-flood-producing heavy convective rainfall events. *Quarterly Journal of the Royal Meteorological Society*, **145**, 2385–2403, CrossRef.

____, J. S. Kain, and A. J. Clark, 2016: Short-term Probabilistic Forecasts of the 31 May 2013 Oklahoma Tornado and Flash Flood Event Using a Continuous-Update-Cycle Storm-scale Ensemble System. *Wea. Forecasting*, **31**, 957–983, CrossRef.

____, T. A. Jones, and P. S. Skinner, 2020a: Probabilistic high impact rainfall forecasts from landfalling tropical cyclones using Warn-on-Forecast System. *Quarterly Journal of Royal Meteorological Society*, **146**, 2050–2065, CrossRef .

____, K. A. Wilson, S. M. Martinaitis, H. Vergara, P. L. Heinselman, and J. J. Gourley, 2020b: The coupling of NSSL Warn-on-Forecast and FLASH systems for probabilistic flash flood prediction. *J. Hydrometeor.*, **21**, 123–141, CrossRef.