

Analysis of End User Access of Warn-on-Forecast Guidance Products during an Experimental Forecasting Task

KATIE A. WILSON,^{a,b} BURKELY T. GALLO,^{a,c} PATRICK SKINNER,^{a,b} ADAM CLARK,^b PAMELA HEINSELMAN,^b AND JESSICA J. CHOATE^d

^a Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma

^b NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

^c NOAA/NWS/Storm Prediction Center, Norman, Oklahoma

^d University of Illinois at Urbana–Champaign, Urbana, Illinois

(Manuscript received 18 December 2020, in final form 14 July 2021)

ABSTRACT: Convection-allowing model ensemble guidance, such as that provided by the Warn-on-Forecast System (WoFS), is designed to provide predictions of individual thunderstorm hazards within the next 0–6 h. The WoFS web viewer provides a large suite of storm and environmental attribute products, but the applicability of these products to the National Weather Service forecast process has not been objectively documented. Therefore, this study describes an experimental forecasting task designed to investigate *what* WoFS products forecasters accessed and *how* they accessed them for a total of 26 cases (comprising 13 weather events, each worked by two forecasters). Analysis of web access log data revealed that, in all 26 cases, product accesses were dominated in the reflectivity, rotation, hail, and surface wind categories. However, the number of different product types viewed and the number of transitions between products varied in each case. Therefore, the Levenshtein (edit distance) method was used to compute similarity scores across all 26 cases, which helped to identify what it meant for relatively similar versus dissimilar navigation of WoFS products. The Spearman's rank correlation coefficient R results found that forecasters working the same weather event had higher similarity scores for events that produced more tornado reports and for events in which forecasters had higher performance scores. The findings from this study will influence subsequent efforts for further improving WoFS products and developing an efficient and effective user interface for operational applications.

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

1. Introduction

The process for forecasting and nowcasting convective storms spans a timeline beginning days prior to storm occurrence and ending with real-time assessment of imminent or ongoing hazardous weather. Along this timeline, numerical weather prediction forecasts and observations provide essential information that guide the issuance of National Weather Service (NWS) outlook, watch, and warning products. In recent years, the development and operationalization of high-resolution, convection-allowing models (CAMs) has enhanced deterministic model guidance available for short-term thunderstorm prediction (Benjamin et al. 2019). Additionally, research has been under way to explore CAM ensemble systems. Unlike deterministic systems that provide a single forecast solution, ensemble systems provide forecast uncertainty information and thus give insight into the likelihood and potential severity of severe weather occurrence. Effective visualization of CAM ensemble uncertainty information is required if it is to provide meaningful and useful guidance to forecasters.

New and innovative methods for postprocessing and visualizing CAM ensemble guidance are necessary and are being explored within the modeling community. In 2015, real-time National Center for Atmospheric Research (NCAR) ensemble

guidance products were provided for public viewing on a website. This website had notable impact and motivated the continuation of NCAR's project for several years. It brought together members of the educational, research, and operational meteorological community to experience a year-round demonstration of a real-time formally designed CAM ensemble over the contiguous United States. Feedback from these users was documented in an informal survey (Schwartz et al. 2019). Since then, the High Resolution Ensemble Forecast, version 2 (HREFv2), system has become the first operational CAM ensemble system. The HREFv2 employs spatial neighborhood smoothing techniques in the postprocessing of its output and uses a variety of visualizations to display products (e.g., “paintball” plots, ensemble maximums, and “postage stamps”; Roberts et al. 2019).

How this type of CAM ensemble guidance might aid the convective forecast process has been assessed through several approaches, including survey tools, observations, and interviews (e.g., Wilson et al. 2019a; Demuth et al. 2020). Additionally, annual testing and evaluation of experimental CAM ensembles has taken place in the National Oceanic and Atmospheric Administration (NOAA) Hazardous Weather Testbed Spring Forecasting Experiment (SFE) since 2007 (Clark et al. 2012, 2018, 2020a; Gallo et al. 2017). Participants attending the SFE include scientists from academic and research institutions, NWS forecasters, and graduate students. Each year, participants assess CAM ensembles to provide subjective ratings of

Corresponding author: Katie Wilson, katie.wilson@noaa.gov

DOI: 10.1175/WCAS-D-20-0175.1

© 2021 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

model output and to issue experimental outlooks. Feedback from participants is then used to drive future model improvements.

One experimental CAM ensemble that has been a focus of SFE activities in recent years is the Warn-on-Forecast System (WoFS). Developed by the NOAA National Severe Storms Laboratory (NSSL), WoFS is a convective-scale, frequently cycled, 36-member WRF-based ensemble analysis and forecast system that provides probabilistic 18-member forecasts of individual storm hazards within the next 0–6 h (e.g., [Stensrud et al. 2009](#); [Wheatley et al. 2015](#); [Jones et al. 2016](#); [Skinner et al. 2018](#); [Yussouf et al. 2020](#)). During the SFE, the WoFS domain is centered over the region with the greatest severe weather threat and covers a 900-km-square area with 3-km grid spacing. The WoFS predictions cover the spatial and temporal scales that span the typical NWS severe thunderstorm and tornado watch and warning products. Given that current operational CAM ensemble systems are not designed to provide frequent guidance on this spatiotemporal scale, WoFS is expected to provide forecasters with a more continuous flow of higher-temporal-resolution probabilistic weather information and be available with less latency than other CAM ensembles. Therefore, WoFS guidance is expected to enhance situational awareness for next-hour convective activity and enable enhanced communication of weather threats to NWS core partners and the public ([Rothfusz et al. 2018](#)). However, for WoFS to positively impact the forecast process, forecasters must first be able to understand and interpret the guidance correctly.

In an effort to establish a baseline of meteorologists' current understanding of storm-scale ensemble-based forecast guidance, a survey of the 2017 SFE participants was issued to query their interpretations of numerous WoFS graphics ([Wilson et al. 2019b](#)). Findings from this survey highlighted the types of probability and percentile concepts that were understood consistently across participants, as well as concepts that proved more challenging. While training needs for using probabilistic forecast guidance have already been noted in numerous reports (e.g., [NRC 2006](#); [Novak et al. 2008](#)), the findings from this study were important for identifying training needs specific to WoFS guidance. Since conducting this survey, the results have been used to develop informal training for subsequent SFE participants and collaborative NWS partners, with an end goal to eventually produce an official WoFS training package. Furthermore, these findings can be used to improve WoFS products, such that visualizations are designed to aid interpretation and understanding, especially for products that proved to be particularly challenging.

While [Wilson et al.'s \(2019b\)](#) study was useful for identifying meteorologists' strengths and weaknesses in understanding guidance from a specific CAM ensemble system, exploring whether this type of guidance is what users need and want is also critically important in the research and development process of CAM ensemble systems. [Demuth et al. \(2020\)](#) recently undertook a study to examine this topic. Using a mix of observational and semistructured interview methods, [Demuth et al. \(2020\)](#) collected information on NWS forecasters' interpretations of prototype CAM ensemble guidance products, along with their specific information needs from CAM ensembles.

This user-centered approach is most ideal for ensuring that model developers create information that meet the needs and wants of operational forecasters, and it should be employed more widely in the development of CAM ensemble guidance, including that of WoFS guidance.

The availability of a large variety of WoFS products brings to question what information users seek when interacting with the WoFS web viewer during a forecasting task. Both subjective feedback and objective log data analysis are important for learning about user web viewing behavior ([Kuniavsky 2003](#); [Dumais et al. 2014](#)). For example, assessments of users' weather-related information seeking behaviors have shown to be useful for learning about how to tailor information based on Google search patterns of hurricane forecast information ([Sherman-Morris et al. 2011](#)) and for improving message suitability to support hurricane evacuation decision-making ([Cahyanto et al. 2016](#)).

In prior SFE experiments, informal observation and feedback of participants' use of WoFS products has provided some sense of what information forecasters seek most frequently. However, users' experiences when interacting with the WoFS web viewer were not logged, and thus forecasters' interactive behaviors over time were not objectively or accurately captured. Therefore, the study presented herein addresses this research gap by providing a thorough documentation and assessment of NWS forecasters' WoFS product usage during a designated experimental forecasting task. More specifically, this study uses WoFS product access log data collected during the 2019 SFE to examine the following research questions: 1) What WoFS products are accessed and with what frequency? 2) What are forecasters' WoFS product access patterns and how do they compare to each other? 3) In what ways do WoFS product access patterns relate to task performance and event type?

Addressing these research questions is important for a variety of reasons. First, while WoFS guidance is currently experimental, only a subset of products will be available once the model is operationalized, and, as such, this subset should contain the products that are most frequently used. Second, the identification of effective approaches for examining WoFS products will enable future users to apply WoFS guidance in the ways that best support their forecast process. Third, knowledge of how users access WoFS products will help to guide the design of a user-friendly interface that is efficient at delivering the information they want. Fourth, this study demonstrates an interdisciplinary approach to exploring user-focused research by bringing together concepts from meteorology, model development, forecasting, and human factors to examine questions that sit at the intersection of research and operations. Thus, while this research is focused on WoFS guidance, the data collection and analysis approaches described herein can be more broadly applied for learning about forecasters' use of other CAM ensembles as well as different types of forecast information. Additionally, the findings from this particular study will further our knowledge on forecasters' styles of data interrogation and how those styles relate to product issuance decisions and forecast performance.

2. Methods

a. Participants

The annual NOAA Hazardous Weather Testbed SFE brings together the operations and research communities to test, evaluate, and document advancements in experimental forecast guidance, with a recent focus on convection-allowing ensembles (Gallo et al. 2017). Since 2017, this experiment has included a 1-h, late-afternoon forecasting activity designed to evaluate use of WoFS guidance. However, since convective storms often initiate and develop into the evening hours, a 4-h evening activity was conducted during the 2019 SFE to more fully assess the potential operational utility of WoFS guidance as convective events unfold (Clark et al. 2020a).

Two different NWS forecasters participated in the evening activity during each of the five weeks of the 2019 SFE (29 April–31 May 2019). The 10 participating forecasters each represented a different local Weather Forecast Office and consisted of 3 females and 7 males. The activity took place Monday through Thursday [1200–2000 central daylight time (CDT)], and each Friday participants provided feedback on the experiment during an end-of-week discussion. This human-subject study was approved under the University of Oklahoma Institutional Review Board 5395. All participants provided informed consent prior to their participation.

b. Procedure

All 10 NWS forecasters underwent a similar research protocol during their participation in this study. Each week, the two participating forecasters were assigned a pseudonym of either F1 or F2. Upon arrival each day, forecasters first received a weather briefing from a retired Storm Prediction Center (SPC) forecaster. On Mondays, forecasters met with research scientists to build familiarity with what WoFS guidance is, learn to use the WoFS web interface (Fig. 1a) and product drawing tool, and discuss the experimental tasks that would be completed during the evening activity. The two NWS forecasters then joined all other SFE participants to 1) complete an online training module focused on storm-scale ensemble-based WoFS guidance concepts and 2) receive a brief overview of the WoFS forecasting task.

The WoFS forecasting task included the issuance of three probabilistic severe outlooks (i.e., to encompass the union of severe hail, severe wind, and/or tornado threats). The first outlook was valid for 1 h (short), the second was valid for 4 h (long), and the third was valid for a targeted 1-h period (2000–2100 CDT). These three outlooks were issued every hour, with the first series issued during a regular SFE WoFS small group activity during 1500–1600 CDT. Each NWS forecaster was assigned to one of two groups to complete the first series of outlooks jointly with other SFE participants (this task is not included in the analysis). The evening activity then began at 1600 CDT, at which time the NWS forecasters continued to issue the outlooks independently through 2000 CDT (Fig. 1b). A research scientist was on hand each evening to supervise and provide experimental support. However, during the Monday evening activity, the supporting researcher found it helpful to provide further discussion on aspects of WoFS guidance and to

demonstrate the experimental product issuance task in real time. Therefore, Monday's evening activity is treated as a familiarization spinup to the experiment and is not included in this study's analysis.

A web-based drawing tool was used to create the outlooks. This tool allowed participants to draw contours with assigned probabilities over available WoFS guidance products. Probabilities were considered coverage probabilities, comparable to the SPC's forecasts of severe weather as occurring within ~40 km (25 mi) of a point and for the duration of the outlook (i.e., either 1 or 4 h). All convective hazards (tornadoes, wind, and hail) were grouped together, so probabilities indicated the chance of any of the three hazards occurring within 25 mi of a point. Forecasters could draw the same contour levels as found in SPC outlooks for any type of severe hazard: 5%, 15%, 30%, 45%, and 60%. Prior outlooks were able to be loaded into the web-based drawing tool and modified, allowing forecasters to adjust previously issued products rather than draw entirely new products each hour.

To document WoFS product usage during the evening activity (which is the focus of this study), both NWS forecasters were provided with their own dedicated web viewer. Each hour (half hour) during the experiment, a 6-h (3 h) WoFS ensemble forecast was initialized and displayed on the web viewers. The NWS forecasters used these WoFS updates to assess storm development and inform outlook issuance each hour. Additionally, they freely viewed web-based observations (e.g., radar, satellite, and SPC mesoanalysis) to maintain situational awareness and to evaluate WoFS performance in real time.

c. Data and analysis

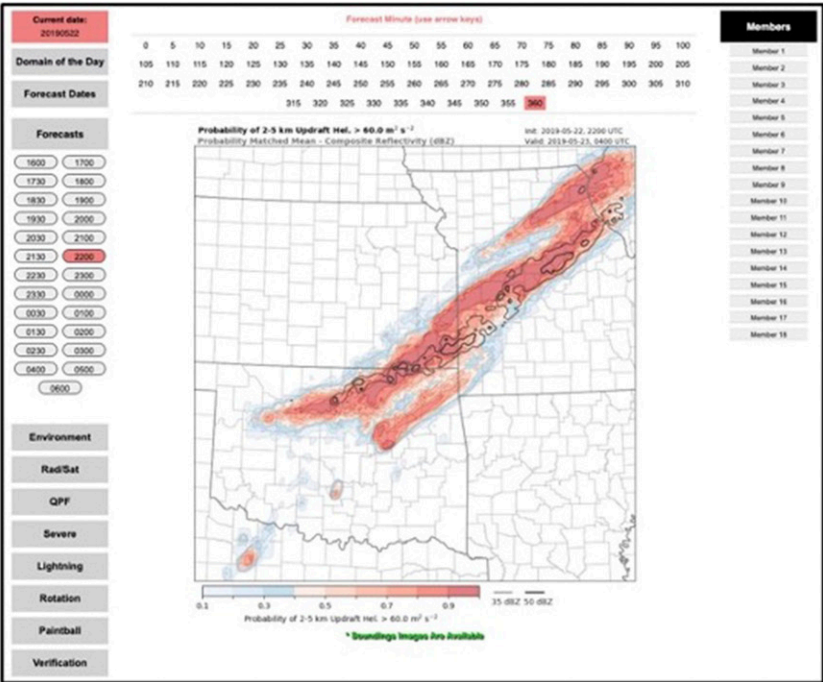
The data analyzed in this study were collected over 13 experiment days (totaling 26 cases because of two forecasters' participation each week, F1 and F2) during the 5-week 2019 SFE. Reference to individual cases in the results will follow a format of month, day, and forecaster for example, May22F1. Eleven days of the 2019 SFE were excluded from this analysis because of Monday's evening activity being treated as a familiarization spinup, no evening activity being held on Fridays, and WoFS availability issues that occurred on several experiment days. The data collected during the forecast task include the WoFS product access logs and experimental outlooks. A description of these datasets and the methods used to analyze them follows.

1) WOFS ACCESS LOGS

The NOAA NSSL access logs were obtained for the two WoFS websites created for F1 and F2's use. Although the forecast task began at 1600 CDT, oftentimes there was a brief overlap as the two NWS forecasters transitioned from the group SFE activity to the independent evening activity. Therefore, to remove F1's and F2's WoFS use during this brief overlap, data analyzed included access logs from 1615 CDT onward.

Product information was extracted from the access logs to reveal the name of the WoFS products accessed by F1 or F2. While timestamp information was also documented in the access logs, it did not accurately reflect the duration for which

(a) WoFS Web Viewer



(b) Experimental Outlooks

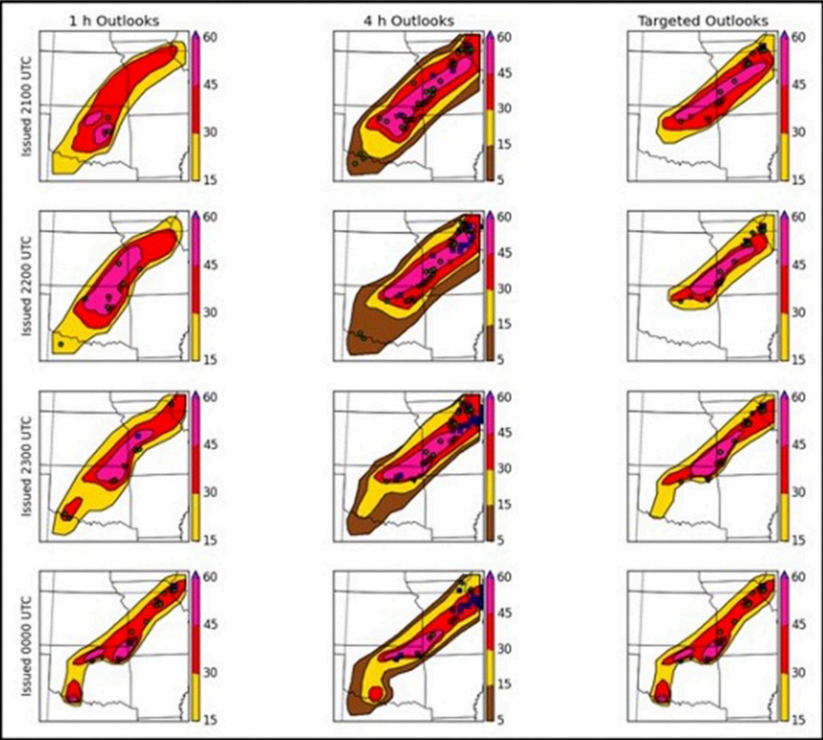


FIG. 1. Examples of (a) the WoFS web viewer interface (with a UH probability-of-exceedance product plotted for the 22 May 2019 event) and (b) a forecasters' experimental outlooks issued for the same event.

Category Letter Codes: a=reflectivity; b=hail; c=rotation; d=vertical motion; e=surface wind; f=member viewer; g=satellite; h=gpf; i=mlcape; j=mlcin; k=storm motion; l=shear; m=srh; n=stp; o=temperature			
Iteration 1 (Initial)	Iteration 2	... Iteration 21 (Final)	Compute the similarity score of two strings for each iteration of the moving window (i.e., blue portion of the longer string and the entire shorter string). Total Iterations = (Longer String Length - Shorter String Length) + 1 <i>E.g., (93 - 73) + 1 = 21 iterations in this example</i> Similarity Score = (Longest String Length - Number of Edits) / (Longest String Length) Find the final similarity score by averaging the similarity scores from all iterations.
May01F1 (93 product accesses) "accebaaaiaaekiaefaciabcna beaacabcbcaaceabeeeeeeee caecaccbefefdcceaeaaaafaa eeaiaieadeaeiaai"	May01F1 (93 product accesses) "accebaaaiaaekiaefaciabcna beaacabcbcaaceabeeeeeeee caecaccbefefdcceaeaaaafaa eeaiaieadeaeiaai"	May01F1 (93 product accesses) "aecebaaaiaaekiaefaeiabcna beaacabcbcaaceabeeeeeeee caecaccbefefdcceaeaaaafaa eeaiaieadeaeiaai"	
May01F2 (73 product accesses) "calkceaoboaaceeeacbeco aicebjioaciebaeiacabiaccaa eaoaecaenaeeeeeankcca"	May01F2 (73 product accesses) "calkceaoboaaceeeacbeco aicebjioaciebaeiacabiaccaa eaoaecaenaeeeeeankcca"	May01F2 (73 product accesses) "calkceaoboaaceeeacbeco aicebjioaciebaeiacabiaccaa eaoaecaenaeeeeeankcca"	

FIG. 2. A schematic detailing the Levenshtein (edit distance) moving-window approach, including the letters assigned to each product category and an example of the similarity score computation for strings representative of May01F1's and May01F2's WoFS product use.

F1 or F2 paid attention to a product (e.g., they may leave a product open while attending to another data source). Although duration is not considered in this analysis, the order in which F1 or F2 accessed products represents a temporal component to product accesses and is considered an important part of this analysis.

Although most WoFS products were reliably available throughout the evening activity, the products designed to provide verification were frequently intermittent or unavailable. Therefore, all visits to verification-related products were removed from the access logs. Additionally, the member viewer consists of simulated composite reflectivity guidance for each of the 18 WoFS ensemble members. To prompt the display of this guidance, forecasters would typically need to click back and forth between the members. However, given that each of the 18 members was treated as a separate product in the access logs, forecasters' viewing behavior of the member viewer was exaggerated. Therefore, to more accurately represent forecasters' visits to the member viewer, the 18 individual members were collapsed to represent a single product.

A total of 131 WoFS products were accessed over the 26 cases included in this analysis (after removing access attempts to verification products). While summary statistics are provided at this granular level for all 10 NWS forecasters' WoFS product use, product categorization was subsequently applied during the analysis process. The groupings were based on meteorological information and on the design of the WoFS web viewer (all products can be viewed at <https://wof.nssl.noaa.gov/retro/>). Groups were defined by 15 categories: reflectivity ($n = 20$ products), hail ($n = 31$), rotation ($n = 35$), vertical motion ($n = 5$), surface wind ($n = 19$), member viewer ($n = 1$), satellite ($n = 3$), quantitative precipitation forecasts (qpf; $n = 3$), mixed-layer convective available potential energy (mlcape; $n = 2$),

mixed-layer convective inhibition (mlcin; $n = 1$), storm motion ($n = 1$), vertical wind shear ($n = 2$), storm relative helicity (srh; $n = 2$), significant tornado parameter (stp; $n = 1$), and temperature ($n = 5$).

Bulk analysis of category visits per case was first conducted, but the notable advantage of grouping products into one of 15 categories was that it more easily enabled a comparison of WoFS product usage patterns across the 26 cases. This comparison was intended to extend our understanding beyond knowing *what* product categories are most used to knowing *how* forecasters extract WoFS guidance information. To perform a comparison of forecasters' WoFS guidance use, the access log information was converted into a string of letters, with each letter representing a specific product category that was accessed. The extent of similarity between forecasters' accesses to product categories (with order of accesses maintained) was computed for all possible case comparisons using the Levenshtein distance algorithm available in the R software package (Van der Loo 2014).

Also known as edit distance, the Levenshtein distance computes the shortest possible string distance between two character vectors (Fig. 2). Edits to a string can be made by inserting, deleting, or substituting a character. The number of edits required to transform one string into another is counted. The similarity score of two strings is then calculated as the difference between the longest string length and the number of edits, which is then divided by the longest string length. The similarity score ranges between values of 0 and 1, with 1 indicating an exact match between two strings, and decreasing values indicating decreasing similarity.

The number of products visited for each of the 26 cases varied, and thus two strings of different lengths are used to calculate the similarity score. The default Levenshtein distance

in R handles string length differences by recycling the shorter string until it is the length of the longer string. However, this approach misrepresents forecasters' product viewing behaviors by artificially inserting product category visits. Truncating the longer string to the length of the shorter string was explored as an alternative solution, but this approach was still limited because it did not fully encompass all forecasters' product viewing behavior. Therefore, for each case comparison, the shorter string was compared with a moving truncated window of the longer string. Similarity scores were computed for every possible iteration of the string comparison, and the final similarity score was taken as the average from all of the iterations. This approach supported a complete comparison of the entirety of two strings and was used to quantify the degree of similarity for WoFS product usage between all 26 cases.

2) EXPERIMENTAL OUTLOOKS

Participants were able to save their outlook multiple times throughout the forecast generation process, resulting in 378 total outlooks collected during the 13 experiment days. Only the latest outlook issued for each type and hour was retained for verification, under the assumption that the final outlook issued prior to the deadline was the most representative of the forecasters' thinking. This left a dataset of 11 outlooks per forecaster per day to examine, or 286 outlooks. Outlooks were stored at NSSL in Geo JavaScript Object Notation (GeoJSON) format and gridded to the WoFS domain prior to objective verification. No interpolation was performed between the contours, maintaining the stepwise nature of the probabilities.

Outlooks were evaluated using two types of objective verification metrics: area under the receiver operating characteristic (ROC) curve (ROCa; Mason 1982) and a probabilistic version of the fractions skill score (FSS; Roberts and Lean 2008). The FSS was calculated using a binary observation field following Roberts et al. (2020), and similar to the approach of Schwartz et al. (2010). Specifically, the FSS calculated here is closer to a Brier skill score (Brier 1950), but with a reference forecast that reflects the worst possible forecast that could be made with a given set of fractional values in the forecast probabilities and observed binary fields. This calculation frequently results in lower values of FSS relative to the original FSS formulation but avoids the issue of conflating the neighborhood and smoothing length scales pointed out by Schwartz and Sobash (2017) by not smoothing the observations. Thus, this calculation is essentially a neighborhood maximum ensemble probability-based FSS.

To perform the verification, filtered local storm reports (LSRs) from the SPC were regridded to the WoFS domain, and expanded such that all points within 25 mi of a report was considered a "hit." The ROC area scores range from 0.5, indicating that a forecast has the same skill as a random forecast, to 1.0, indicating a perfect forecast. An FSS of 1.0 also indicates a perfect forecast but has a lower limit of 0.0. To assess performance across the full period of the experiment, these statistics were then aggregated across the 11 forecasts produced by each forecaster on each day and included forecasts and observations valid over both 1- and 4-h time periods.

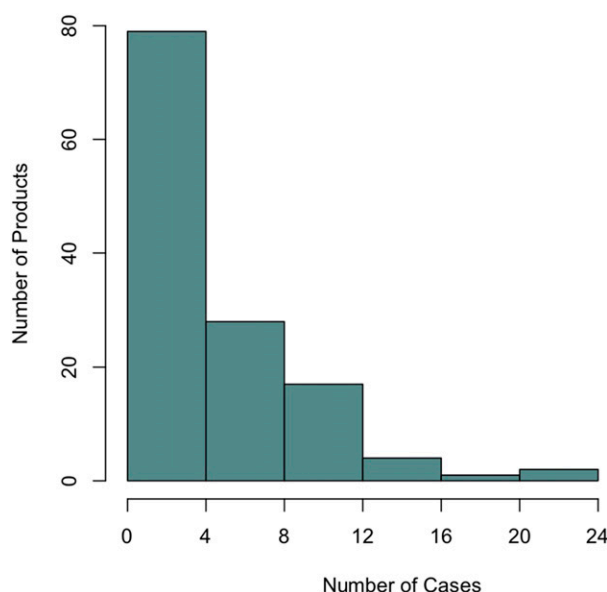


FIG. 3. Histogram depicting the number of WoFS products accessed per bin of number of cases (e.g., the 0–4 bin interval includes up to 3 cases, and the 4–8 bin interval includes 4–7 cases).

The Spearman's rank correlation coefficient R and corresponding significance value was used to examine the relationship(s) between the aggregated performance results, types of hazards produced during an event, and forecasters' WoFS product usage. The Spearman's rank correlation coefficient ranges from -1 to $+1$, and values indicate the strength of a monotonic relationship between two variables (Stowell 2014). This nonparametric measure was chosen since the Shapiro–Wilks normality test found that many of the datasets used in this analysis are nonnormally distributed.

3. Results

a. Product accesses

To understand overall WoFS product use, forecasters' total product accesses were reviewed. Together, the 10 NWS forecasters accessed a total of 131 WoFS products. However, not all of these products were accessed in all of the 26 cases included in this analysis. Over one-half of the logged products were accessed in fewer than four cases, while only six of the recorded products were accessed in more than one-half of the cases (Fig. 3). This result suggests that although the NWS forecasters had access to a large array of products, only a small subset of them were consistently accessed during the experiment. The most consistently accessed WoFS products across the 26 cases were the reflectivity paintball plot (24 cases), ensemble mean mlcape (24 cases), and 2–5-km updraft helicity (UH) paintball plot (17 cases). Additionally, three versions of the 2–5 km-UH probability products (9-km neighborhood radius at 5-min and 1-h intervals, and 15-km neighborhood radius at a 1-h interval) and the 90th-percentile surface wind value were accessed in more than one-half of the cases (16, 15, and 17 cases, respectively).

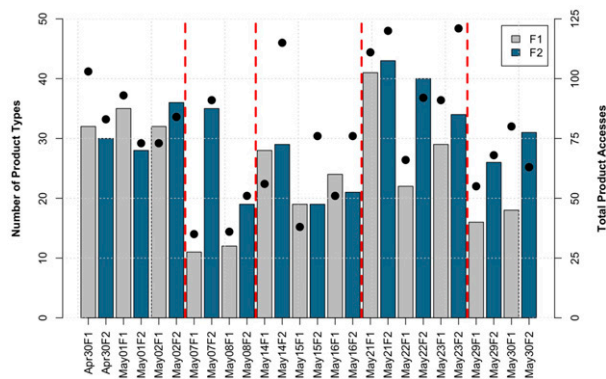


FIG. 4. The total product type count (bars) and product accesses (black dots) for all 26 cases, with vertical red dashed lines separating experiment weeks and the corresponding changeover in F1 (gray) and F2 (teal) participants.

To further understand WoFS product access behavior, the number of individual products accessed (i.e., product type count) and the number of transitions between products (i.e., product accesses) was considered on a per-case basis (Fig. 4). On average, forecasters accessed 29 (standard deviation = 9, minimum = 11, and maximum = 43) different products per case, with a total of 76 (standard deviation = 25, minimum = 35, and maximum = 121) product accesses. In general, those who accessed fewer types of products also made fewer transitions between products. However, there were instances in which participants made a relatively high number of product transitions in comparison with the number of product types they accessed (e.g., May14F2 and May22F1). In contrast, there were also instances in which a relatively low number of product transitions were made in comparison with the number of product types accessed (e.g., May14F1 and May30F2). This finding suggests different styles for viewing WoFS guidance, with varying degrees of exploration across the WoFS web viewer and varying rates at which forecasters sought new information by transitioning from one product type to another.

b. Product categories

In totality, the distribution of accesses to the 15 product categories across the whole experiment shows that the reflectivity and rotation categories were accessed in all of the 26 cases and each accounted for, on average, 30% of the total product category accesses within a case (Fig. 5). The next most accessed product categories were hail and surface wind, each accounting for an average of 10% of the total product category accesses within a case (Fig. 5). Given the higher number of individual products included in each of these categories than in the other categories, this result demonstrates that forecasters spent more time viewing and extracting information from the many products presented within these particular categories. However, we did not find that the number of products within a category was proportional to the average number of category accesses. The remaining categories accounted for very little of the total accesses made to product categories in this experiment (Fig. 5).

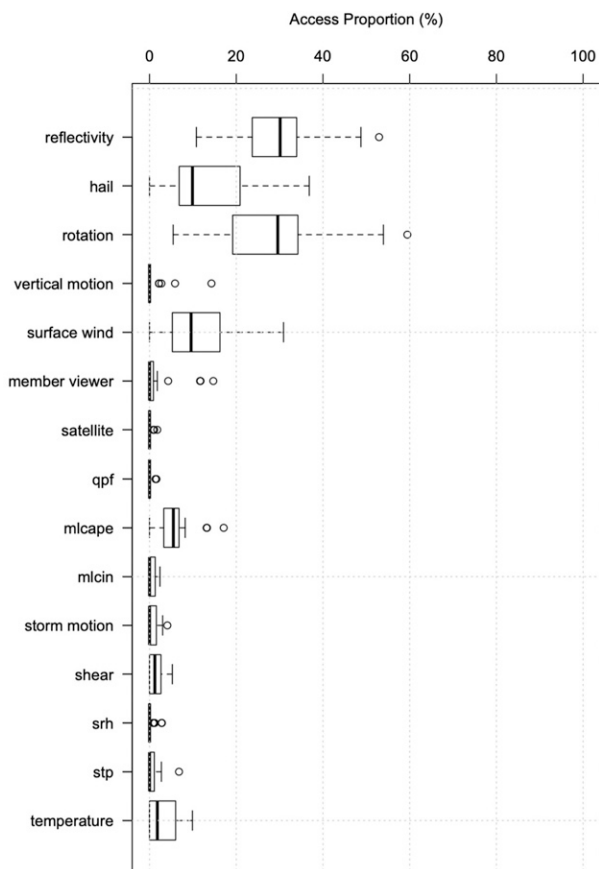


FIG. 5. The distribution of access proportions (%) per case for all 15 product categories. The thick line in the box indicates the median value, and the ends of the box represent the interquartile range (25th–75th percentile). The whiskers extend up to 1.5 times the lower (25th percentile) and upper (75th percentile) quartiles, and outliers are given by the open circles.

The specific products accessed for each of these four main product categories were examined next. For the reflectivity category, the reflectivity paintball plot was overwhelmingly most popular. Forecasters also tended to access the probability-of-exceedance products more often than percentile products, along with 5-min output produced using the smaller neighborhoods. For rotation, the 2–5-km UH paintball plot was accessed much more often than any other rotation product. Forecasters accessed the hourly interval products more so in this category than in others and used the 2–5-km UH products more frequently than the 0–2-km UH products. This finding may be due to a focus on midlevel UH in the early literature that first explored applications of UH for diagnosing severe weather threats and to the ongoing demonstrations of its usefulness for forecasting applications (e.g., Kain et al. 2008; Sobash et al. 2011). Unlike in the reflectivity and rotation categories, the 90th-percentile and maximum products were more popular in the hail and surface wind categories than the probability threshold products. This preference is likely due to forecasters using this information to directly quantify the potential severity of these convective hazards.

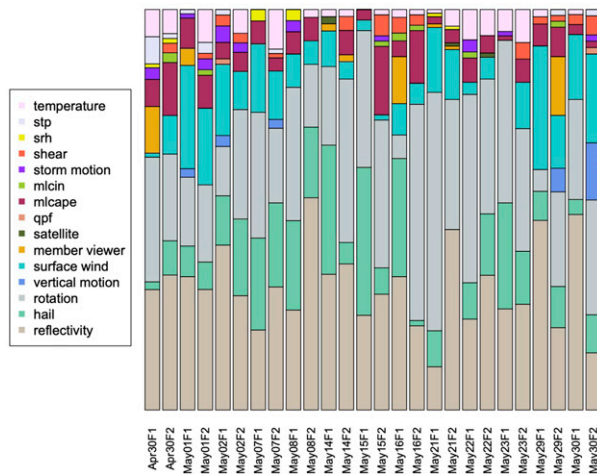


FIG. 6. The proportion (%) of each product category accessed per case relative to the total number of product category accesses for that same case.

While a review of the product categories accessed from the whole experiment provides an overall sense for what type of guidance was sought most, analyzing product categories accessed on a per-case basis is important for identifying WoFS product usage as it relates to individual forecasters and specific weather hazards. As indicated in the distributions of product category access proportions (Fig. 5), the product categories accessed in each case are typically dominated by the reflectivity, rotation, hail, and surface wind categories (Fig. 6). However, the visualization of case-by-case product category access reveals that product category preference also varied (e.g., May15F1 and May15F2; Fig. 6). This result suggests that it is possible that there were different approaches to how forecasters extracted information during the forecasting task, even for cases where the dominant product categories were the same. Therefore, the order (or pattern) in which the product categories were accessed is considered next. This analysis builds on the investigation of *what* product categories were accessed by considering *how* forecasters accessed them.

c. Product category access patterns

The order in which WoFS products were accessed during the forecasting task matters because it is representative of forecasters' information seeking behaviors and related processes as events unfolded. Even if forecasters viewed similar products, they may have accessed the information differently during the forecasting task. A comparison of the product access patterns helps to highlight similarities and differences in the ways forecasters extracted information, and thus gives insight into the variety of forecast styles and approaches adopted in this forecasting task.

Similarity scores were computed for all possible case comparisons using the Levenshtein (edit distance) algorithm as described in section 2c(1). Removing same-case comparisons, the distribution of similarity scores ranged from 0.16 to 0.49, with a median value of 0.33 (Figs. 7 and 8). To contextualize these similarity scores to forecasters' information seeking

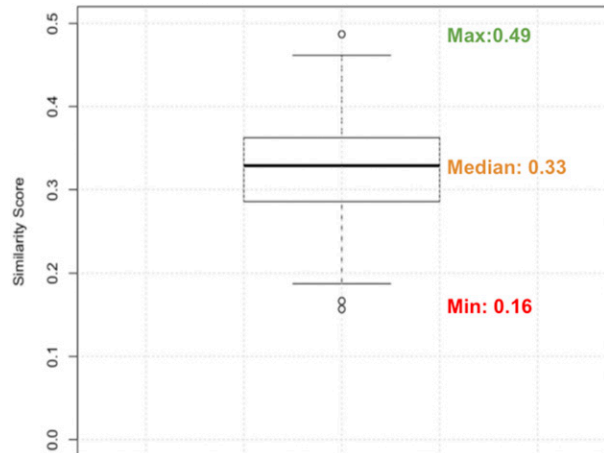


FIG. 7. The distribution of all similarity scores. The boxplot description is as in Fig. 5.

behaviors, examples of the cases resulting in maximum, median, and minimum similarity scores are reviewed.

In the maximum similarity score (0.49) example, May16F2 and May21F1 product accesses were predominantly in the reflectivity and rotation categories, with a small proportion of accesses mostly in the hail, surface wind, and mlcape categories (Fig. 9a). The time series of products accessed shows that May16F2 and May21F1 frequently transitioned between products within the rotation category (Fig. 10a). In both cases, transitions between the rotation and reflectivity categories were also made. A difference is that while May16F2 also transitioned from the rotation to mlcape category, May21F1

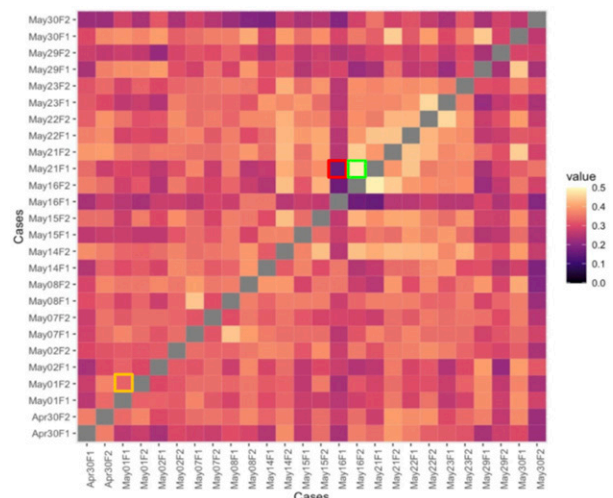


FIG. 8. Similarity scores of product category access patterns for all possible 26 case comparisons. Higher values represent greater similarity between product category access patterns. The similarity scores for same-case comparisons are removed from the analysis, as represented by the gray diagonal line. Maximum, median, and minimum similarity scores are shown in the green, orange, and red boxes and correspond to the examples shared in Fig. 10, below.

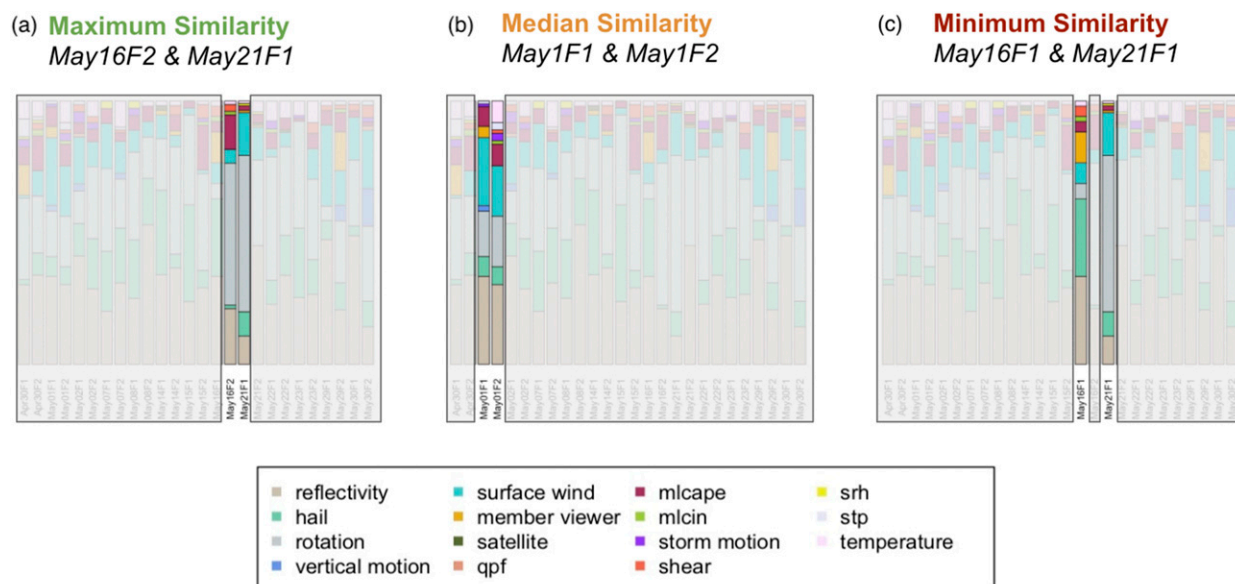


FIG. 9. As in Fig. 6, but with cases highlighted to demonstrate the product category access pattern comparisons resulting in (a) maximum, (b) median, and (c) minimum similarity scores. The highlighted cases follow the order indicated in the labels above each image.

instead transitioned from the rotation to surface wind category (Fig. 10a). However, these transitions were much less frequent than the other more dominant product category accesses.

Two cases representing the median similarity score (0.33) are May1F1 and May1F2 (Fig. 9b). Both forecasters were working the same weather event on this day. The proportion of accesses made to the reflectivity, hail, rotation, surface wind, and mlcape categories is strikingly similar. Therefore, in totality, both forecasters extracted similar types of information. However, with a greater variety of product categories accessed in this case than in the maximum similarity score example, there is also greater variation in how May1F1 and May1F2 extracted the WoFS product information (Fig. 10b).

In the minimum similarity score (0.16) example, the product categories accessed are drastically different between cases (Fig. 9c). Whereas May16F1's accesses are dominated by the reflectivity and hail categories, May21F1's accesses are largely dominated by the rotation category. Both May16F1 and May21F1 visit other categories but at a much lower rate than their dominant categories. The lack of similarity between these two cases is evident in the time series of product accesses, such that May16F1 typically transitioned from the reflectivity to the hail categories, as well as to other categories including the member viewer and surface wind (Fig. 10c). In contrast, May21F1 transitioned within the rotation category often, as well as from rotation to surface wind categories (Fig. 10c).

The minimum and maximum similarity score examples demonstrate two extremes of product category access pattern comparisons in this experiment. However, the median similarity score example is more representative of the extent to which forecasters' information seeking behavior overlapped. This example shows that even when forecasters predominantly access the same product categories with similar proportions,

there can still be notable variety in *how* they transition between products. While lengthy matches in product category transitions were not evident in the median similarity score example shared, there were some shorter recurring transitions in both forecasters' product accesses. The following analysis looks to identify these types of patterns in all 26 cases.

d. Common transition patterns

The analysis of transitions between products belonging to the most popular categories is constrained to those accounting for an average of at least 10% of participants' total product accesses. As discussed in section 2b, these categories include reflectivity, rotation, hail, and surface wind. All possible patterns of transitions between products using two (e.g., reflectivity to rotation; $n = 16$), three (e.g., reflectivity to rotation to hail; $n = 64$), or four (e.g., reflectivity to rotation to hail to reflectivity; $n = 256$) permutations of these product categories was investigated, and recurring patterns (i.e., occurring more than once within a case) were identified in each of the 26 cases.

Eight transition patterns were found to recur across the 26 cases (Fig. 11). These patterns consisted mostly of transitions between two products, and no recurring transition patterns of four products were identified. Most commonly, forecasters transitioned from products within the reflectivity category to those within the rotation category. The next most common types of product transitions were from the rotation category to the reflectivity category, and then within the reflectivity category itself (Fig. 11). Recurring transition patterns of three products were found between the reflectivity and rotation categories also, although their rates of occurrence were generally lower. The remaining three transition patterns also had lower rates of occurrence, but they captured forecasters'

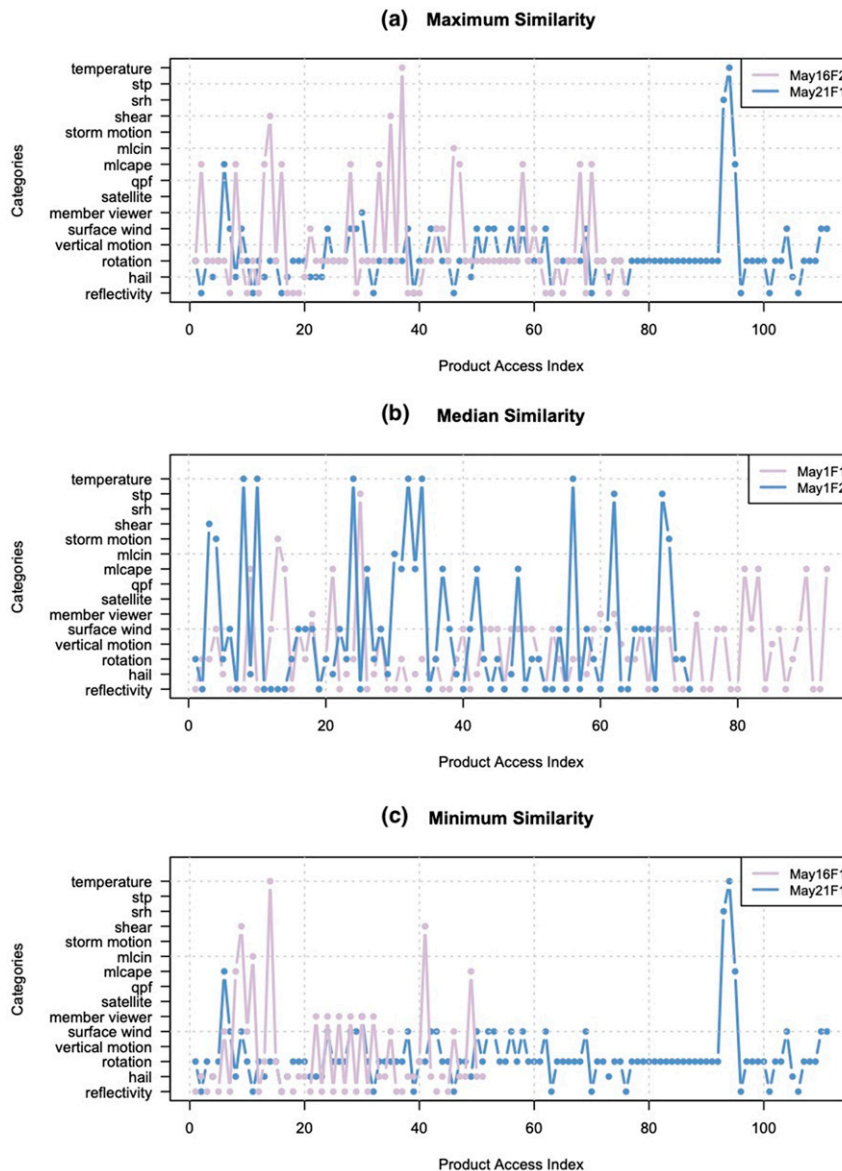


FIG. 10. Time series of product category accesses for case comparisons resulting in (a) maximum, (b) median, and (c) minimum similarity scores.

repeated conjoined use of rotation and reflectivity guidance with hail and surface wind guidance (Fig. 11).

Of the eight transition patterns plotted, only the first pattern (reflectivity to rotation) was identified in all 26 cases. Therefore, forecasters generally did not make the same types of product transitions. Furthermore, while the median values of pattern counts were relatively modest, outlier data points highlight cases in which transition patterns were frequently used. To explore the frequency of the eight transition patterns in each of the 26 cases, the individual data points were plotted in Fig. 12. In most experiment weeks, the pattern counts fell to values predominantly below 10. However, in some cases, and especially during week four, pattern counts spread to higher

values, meaning that these forecasters accessed WoFS products in the same order more often. These higher pattern counts occurred mostly for the reflectivity to rotation, rotation to rotation, and rotation to reflectivity category transitions, and are generally associated with cases in which forecasters accessed a relatively high number of products (i.e., exceeding 100 transitions; Fig. 4).

e. Product usage relative to performance and hazard type

The analysis in sections 3a–3d describes what WoFS products forecasters viewed, at what frequency, and in what order while completing an experimental forecasting task. A final step in the analysis relates the product usage findings to forecasters'

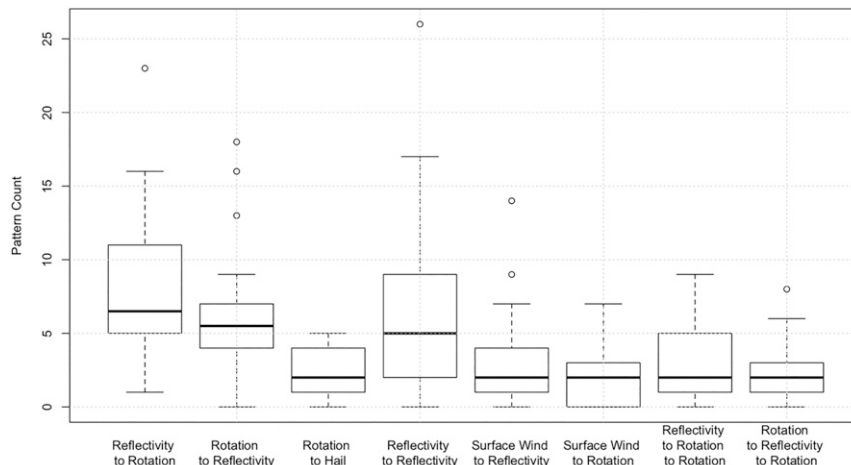


FIG. 11. The distribution of eight product category transition patterns for the 26 cases. The boxplot description is as in Fig. 5.

performance during the experimental task and the type of weather hazards produced in each event.

The FSS and ROCa was computed for each of the 26 cases, over all forecasts issued for that day by each forecaster. FSS values ranged from 0.03 to 0.51 and the ROCa values ranged from 0.59 to 0.97 (Fig. 13). In general, forecasters' performance results were similar for those working the same day. Of the 13 events, the number of tornado, hail, and wind LSRs varied such that 8 produced tornadoes, 12 produced severe wind, and all produced severe hail (Fig. 14).

How forecasters' performance related to the type of hazards produced during an event, and how both performance and event hazards related to forecasters' WoFS product usage, was investigated using the Spearman's rank correlation coefficient R and corresponding significance value. In total, 91 correlations were computed. These correlations were computed between all performance measures (including FSS, ROCa, and median FSS and ROCa values for the same event), tornado/hail/wind LSR counts, and the product usage analysis {including number of product types viewed; total product accesses; number of visits

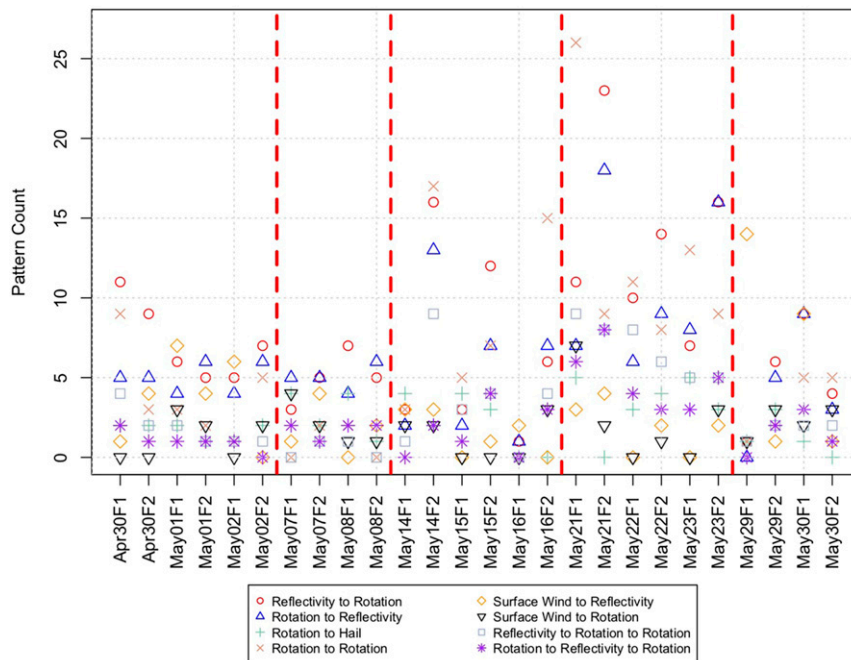


FIG. 12. Pattern counts for eight product category transitions plotted for all cases. Vertical red dashed lines separate the five experiment weeks.

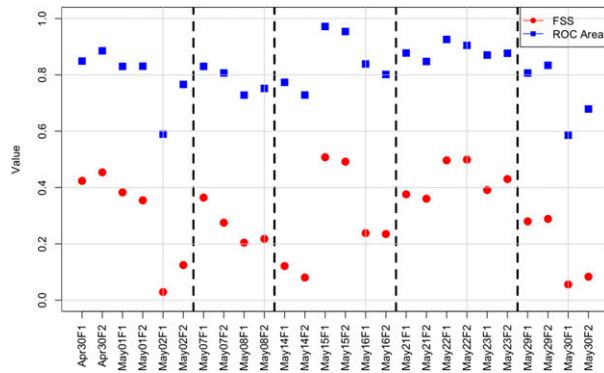


FIG. 13. The fractions skill score (red circles) and ROCa (blue squares) values for all 26 cases. Vertical dashed lines separate experiment weeks.

to each of the 15 product categories; same-event similarity scores [see Fig. 1 and section 2c(1) for details on similarity scores]; and number of occurrences for each of the identified eight product transition patterns].

Of the 91 correlations computed, 8 statistically significant moderate or strong positive associations were found (i.e., moderate $R \geq \pm 0.5$, strong $R \geq \pm 0.7$, and p value < 0.05 ; Hinkle et al. 2003). Of these eight associations, three moderate relationships were found between performance and event type (Figs. 15a–c). Specifically, there was a positive relationship between FSS and the number of tornado and severe hail LSRs, as well as between ROCa and the number of severe hail LSRs. It is possible that this relationship is a result of weather events producing a higher number of storm reports being more predictable, such that their environments were more conducive to producing severe weather. A moderate relationship was also established between the number of tornado LSRs and number of visits to products in both the SRH and temperature categories (Figs. 15d,e). The temperature category primarily included visits to the 2-m temperature and 2-m dewpoint products. The only strong relationship identified in the correlation calculations was between the same-event similarity scores and the number of tornado LSRs, meaning that participants had more similar WoFS product analysis patterns for events that produced more tornadoes (Fig. 15f). Two moderate relationships with the same-event similarity score were also discovered. These relationships were with the same-event median FSS and median ROCa values, meaning that forecasters with more similar WoFS product analysis patterns also had higher degrees of performance during an event (Figs. 15g,h). Of the remaining variable correlations, the Spearman's rank correlation coefficients were weak ($R < \pm 0.5$) and had almost entirely corresponding insignificant p values.

4. Discussion, limitations, and future work

The results presented in this study highlight what WoFS products forecasters used and the extent of similarity in forecasters' navigation of those WoFS products during an

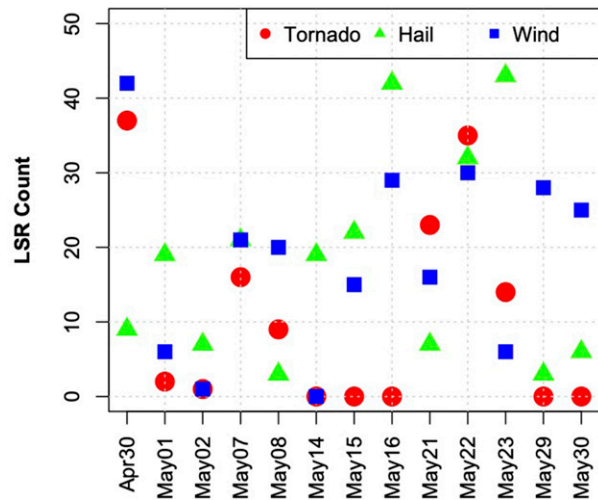


FIG. 14. The tornado (red circle), hail (green triangle), and wind (blue square) LSR counts for each of the 13 events.

experimental forecasting task. Although 131 WoFS products were accessed at some point during the 26 cases, less than half of these products were accessed somewhat consistently across the different cases. Furthermore, forecasters varied in the number of different types of products viewed per case and the number of times they transitioned from one product to another. Despite differences in these bulk measures, when products were grouped by category, a uniform finding across the 26 cases was that the majority of products viewed were in the reflectivity, rotation, hail, and surface wind categories. This finding was not surprising since these products directly relate to a severe weather forecasting task. Exploring these categories further, eight common transition patterns emerged, with the reflectivity to rotation transition pattern occurring in all 26 cases. These findings can help inform numerous future development and implementation strategies for WoFS guidance use in NWS operations. For example, this research helps to identify products that should be prioritized for use in severe weather forecasting, as well as ways in which products can be further improved to meet the needs of NWS forecasters. Products could be developed such that information available in currently separate products but concurrently viewed are integrated into one single visualization. Prior efforts to plot UH with composite reflectivity in the NCAR and HREF ensemble member viewers provide examples of ways to combine product information. Additionally, these findings can help guide user interface design such that products are organized in a manner that supports efficient transition between products.

The similarity score findings showed examples of what it means for forecasters to have relatively high, median, or low similarity in their accessing behavior of WoFS products in comparison with their forecaster counterpart for the same event, and these analyses highlight two important findings. First, same-event similarity scores were found to have significant moderate or strong relationships with the number of

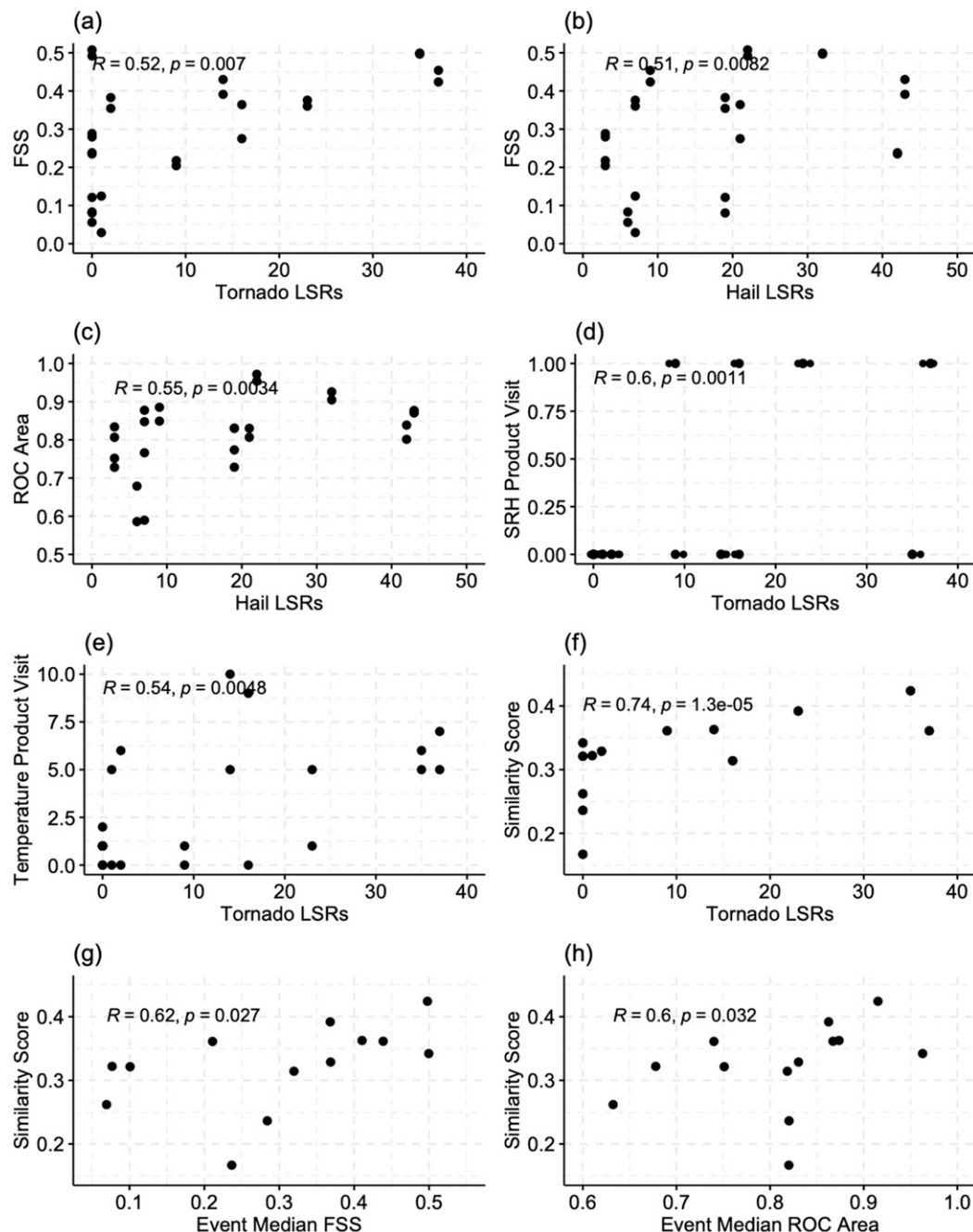


FIG. 15. Scatterplots of variables with a moderate or strong monotonic relationship as determined by Spearman's rank correlation coefficients R and their corresponding p values.

tornado LSRs and with performance (FSS and ROCa values). It is possible that the more similar viewing pattern of WoFS products in these two instances is a result of forecasters working events that are inherently less uncertain and easier to predict, thus driving a more predictable pattern of information seeking. The significant moderate correlation between performance and tornado LSRs supports this explanation, such that the more notable events were better, and perhaps more

easily, forecast. However, since we did not collect data on forecasters' expectations for event hazards, we cannot confirm if forecasters' expectations influenced their WoFS product use. Future research should include questions to examine this aspect of information-seeking behavior. Second, most of this study's other WoFS product use analyses had an insignificant weak association with forecasters' performance and event hazard type. This finding is important because it suggests that

there are numerous approaches to viewing the WoFS products, and these approaches do not necessarily relate to forecast performance or to the number of severe hail or wind LSRs produced in a given event. Likewise, this finding suggests that how forecasters accessed the WoFS products had less of an impact on forecast skill than the nature of the event itself did. However, there are limitations in using just LSRs to characterize the meteorological conditions, and it is therefore possible that an analysis of other measures of storm mode and predictability with WoFS product use may detect additional relationships between forecaster behavior and weather events.

While this study reports on findings to help guide future developments and operationalization of WoFS guidance, here we review them with the study's limitations in mind. First, the findings are a function of the experimental forecast task conducted with only 10 NWS forecasters and for a relatively small sample of weather events. It is possible that for either a different forecast task or for a task tested with many more forecasters over many more events, we may learn something new about what products are important to forecasters or how their navigation strategies compare. Second, the experimental forecast task was focused on total severe weather hazards (i.e., severe hail, severe wind, and/or tornadoes). It is therefore difficult to separate out WoFS product use for these individual hazards. However, the SFE is beginning to explore issuing experimental forecast products for individual severe weather hazards (Clark et al. 2020b); a natural extension of this work would therefore be to examine participants' WoFS product use in these subsequent experiments. Additionally, the focus on severe weather hazards means that these results are not applicable to other hazards, such as flash flooding. A result of the severe hazard focus, for example, was that forecasters barely accessed the quantitative precipitation forecast products during this experiment. Future work should collect access log data for other user groups and compare WoFS product use across a more varied spectrum of forecast challenges.

Last, in the interest of providing a suitable degree of ecological validity, this study took an exploratory approach and therefore did not constrain forecasters' access to other types of guidance during the experiment nor were forecasters' interactions with one another prevented. While this approach promoted an environment that is more similar to forecasters' operational experiences, it introduced confounding variables that are difficult to measure the impact of. Confounding variables include the extent to which forecasters depended on other sources of guidance (e.g., real-time radar and satellite imagery and SPC mesoanalysis), the influence of the experimental forecasts drawn in a group activity that preceded this experiment each day, and the influence of forecasters' interactions on their WoFS product use and forecast performance. Future research should incorporate questions to assess how forecasters are weighting different pieces of information alongside their use of WoFS guidance, as well as how they are weighting the initial consensus outlooks formed during group activities. Furthermore, increased data collection on topics like forecast expectation prior to working an event would help to provide insight on how

forecasters' initial conceptual models impact downstream WoFS product use.

5. Conclusions

This study presented a first objective analysis of forecasters' WoFS product use for an experimental forecasting task. The analysis of WoFS web viewer access log data enabled the types, frequency, and transition between WoFS products accessed to be investigated. Additionally, forecasters' patterns of WoFS product use were assessed using a string similarity scoring method. This method demonstrated an approach for capturing, quantifying, and comparing forecasters' information seeking behavior, and highlighted instances when forecasters' use of the WoFS web viewer was very similar and very dissimilar. The analyses of forecasters' WoFS product use were assessed for their association to forecast performance and event LSR records using the Spearman's rank correlation coefficient. This assessment showed that forecasters working the same event had more similar WoFS product use for events that produced more tornado LSRs, and for events that they had forecast better. These results suggest that greater diversity in WoFS product use behavior may be expected for events producing less tornado activity and for events that are more difficult to forecast. However, despite the overall lower performance in some of these cases, the findings suggest that there are multiple paths for navigating WoFS guidance products that lead to forecasts of similar skill. This topic should be further explored in future experimentation that would increase the sample size of participants, allow for an assessment of WoFS product use for individual hazard forecasts, and more tightly constrain or measure access to and influence of other weather information.

The findings from this study will directly influence subsequent efforts for further developing and improving WoFS products for severe weather forecasting. For example, this study provides insight that guides the prioritization of WoFS products for further development (e.g., combining information from multiple products, such as information depicted in the frequently accessed probability swath and paintball plot products) and operational implementation (e.g., into the Advanced Weather Interactive Processing System 2). This study also provides a foundation for further exploring, quantifying, and comparing WoFS product use for other forecast challenges (e.g., flash flooding events and mixed rainfall/severe weather threat events such as hurricanes). Through these research efforts, we hope to provide operational WoFS guidance in a manner that maximizes its friendly, efficient, and effective use, thus supporting and enhancing the forecast decision-making processes of NWS forecasters across the United States.

Acknowledgments. This study was conducted as part of the 2019 NOAA Hazardous Weather Testbed Spring Forecasting Experiment and was therefore possible because of the efforts and skills of many colleagues. We are extremely grateful to these colleagues, including Kent Knopfmeier for running the real-time WoFS, Brett Roberts for designing the outlook drawing tool, Joe Matthews and Vicki Farmer for coordinating the experimental WoFS web viewers and obtaining the access

log data, the numerous Warn-on-Forecast researchers who provided experimental support to participants during the evening activity, and the 10 NWS forecasters who participated in this experiment. We also thank Kodi Berry, Brett Roberts, and three anonymous reviewers for providing comments on this paper. Funding was provided by the NOAA/Office of Oceanic and Atmospheric Research Warn-on-Forecast Program and under NOAA–University of Oklahoma Cooperative Agreement NA11OAR4320072, U.S. Department of Commerce. The contents of this paper do not necessarily reflect the views or official position of any organization of the U.S. government.

Data availability statement. The nonidentifiable data collected in this study (i.e., the experimental outlook forecasts and NSSL web access logs) will be made available upon request and free of charge following a reasonable period of time for data analysis and publishing (approximately two years).

REFERENCES

- Benjamin, S. G., J. M. Brown, G. Brunet, P. Lynch, K. Saito, and T. W. Schlatter, 2019: 100 years of progress in forecasting and NWP applications. *A Century of Progress in Atmospheric and Related Sciences: Celebrating the American Meteorological Society Centennial*, Meteor. Monogr., No. 59, Amer. Meteor. Soc., <https://doi.org/10.1175/AMSMONOGRAPHS-D-18-0020.1>.
- Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3, [https://doi.org/10.1175/1520-0493\(1950\)078<0001:VOFEIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2).
- Cahyanto, I., L. Pennington-Gray, B. Thapa, S. Srinivasan, J. Villegas, C. Matyas, and S. Kioussis, 2016: Predicting information seeking regarding hurricane evacuation in the destination. *Tour. Manage.*, **52**, 264–275, <https://doi.org/10.1016/j.tourman.2015.06.014>.
- Clark, A. J., and Coauthors, 2012: An overview of the 2010 Hazardous Weather Testbed Experimental Forecast Program Spring Experiment. *Bull. Amer. Meteor. Soc.*, **93**, 55–74, <https://doi.org/10.1175/BAMS-D-11-00040.1>.
- , and Coauthors, 2018: The Community Leveraged Unified Ensemble (CLUE) in the 2016 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment. *Bull. Amer. Meteor. Soc.*, **99**, 1433–1448, <https://doi.org/10.1175/BAMS-D-16-0309.1>.
- , and Coauthors, 2020a: A real-time, simulated forecasting experiment for advancing the prediction of hazardous convective weather. *Bull. Amer. Meteor. Soc.*, **101**, E2022–E2024, <https://doi.org/10.1175/BAMS-D-19-0298.1>.
- , and Coauthors, 2020b: The Spring Forecasting Experiment 2020: Preliminary findings and results. NOAA Summary Rep., 77 pp, https://hwt.nssl.noaa.gov/sfe/2020/docs/HWT_SFE_2020_Prelim_Findings_FINAL.pdf.
- Demuth, J. L., 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>.
- Dumais, S., R. Jeffries, D. M. Russell, D. Tang, and J. Teevan, 2014: Understanding user behavior through log data and analysis. *Ways of Knowing in HCI*, J. S. Olson and W. A. Kellogg, Eds., Springer, 349–372.
- Gallo, B. A., 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>.
- Hinkle, D. E., W. Wiersma, and S. G. Jur, 2003: *Applied Statistics for the Behavioral Sciences*. Houghton Mifflin, 756 pp.
- 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, <https://doi.org/10.1175/WAF-D-15-0107.1>.
- Kain, J. S., and Coauthors, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting*, **23**, 931–952, <https://doi.org/10.1175/WAF2007106.1>.
- Kuniavsky, M., 2003: Log files and customer support. *Observing the User Experience: A Practitioner's Guide to User Research*, E. Goodman, M. Kuniavsky, and A. Moed, Eds., Elsevier, 395–418.
- Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303.
- Novak, D. R., D. R. Bright, and M. J. Brennan, 2008: Operational forecaster uncertainty needs and future roles. *Wea. Forecasting*, **23**, 1069–1084, <https://doi.org/10.1175/2008WAF2222142.1>.
- NRC, 2006: *Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts*. National Academies Press, 124 pp.
- Roberts, B., I. J. Jirak, A. J. Clark, S. J. Weiss, and J. S. Kain, 2019: Postprocessing and visualization techniques for convection-allowing ensembles. *Bull. Amer. Meteor. Soc.*, **100**, 1245–1258, <https://doi.org/10.1175/BAMS-D-18-0041.1>.
- , B. T. Gallo, I. L. Jirak, A. J. Clark, D. C. Dowell, X. Wang, and Y. Wang, 2020: What does a convection-allowing ensemble of opportunity buy us in forecasting thunderstorms? *Wea. Forecasting*, **35**, 2293–2316, <https://doi.org/10.1175/WAF-D-20-0069.1>.
- Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, <https://doi.org/10.1175/2007MWR2123.1>.
- 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>.
- Schwartz, C. S., and R. A. Sobash, 2017: Generating probabilistic forecasts from convection-allowing ensembles using neighborhood approaches: A review and recommendations. *Mon. Wea. Rev.*, **145**, 3397–3418, <https://doi.org/10.1175/MWR-D-16-0400.1>.
- , and Coauthors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, **25**, 263–280, <https://doi.org/10.1175/2009WAF2222267.1>.
- , G. S. Romine, R. A. Sobash, K. R. Fossell, and M. L. Weisman, 2019: NCAR's real-time convection-allowing ensemble project. *Bull. Amer. Meteor. Soc.*, **100**, 321–343, <https://doi.org/10.1175/BAMS-D-17-0297.1>.
- Sherman-Morris, K., J. Senkbeil, and R. Carver, 2011: Who's googling what? What internet searches reveal about hurricane information seeking. *Bull. Amer. Meteor. Soc.*, **92**, 975–985, <https://doi.org/10.1175/2011BAMS3053.1>.
- Skinner, P. S., and Coauthors, 2018: Object-based verification of a prototype Warn-on-Forecast System. *Wea. Forecasting*, **33**, 1225–1250, <https://doi.org/10.1175/WAF-D-18-0020.1>.

- Sobash, R. A., J. S. Kain, D. R. Bright, A. R. Dean, M. C. Coniglio, and S. J. Weiss, 2011: Probabilistic forecast guidance for severe thunderstorms based on the identification of extreme phenomena in convection-allowing model forecasts. *Wea. Forecasting*, **26**, 714–728, <https://doi.org/10.1175/WAF-D-10-05046.1>.
- Stensrud, D., 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>.
- Stowell, S., 2014: Summary statistics for continuous variables. *Using R for Statistics*, Apress, 59–72.
- Van der Loo, M. P. J., 2014: The stringdist package for approximate string matching. *R J.*, **6**, 111–122, <https://doi.org/10.32614/RJ-2014-011>.
- Wheatley, D. M., K. H. Knopfmeier, T. A. Jones, and G. J. Creager, 2015: Storm-scale data assimilation and ensemble forecasting with the NSSL Experimental Warn-on-Forecast System. Part I: Radar data experiments. *Wea. Forecasting*, **30**, 1795–1817, <https://doi.org/10.1175/WAF-D-15-0043.1>.
- Wilson, K. A., and Coauthors, 2019a: Exploring applications of storm-scale probabilistic warn-on-forecast guidance in weather forecasting. *International Conference on Human-Computer Interaction (HCI 2019): Virtual, Augmented and Mixed-Reality, Applications and Case Studies*, J. Chen and G. Fragomeni, Eds., Lecture Notes in Computer Science, Vol. 11575, Springer, 577–572.
- , P. L. Heinselman, P. S. Skinner, J. J. Choate, and K. E. Klockow-McClain, 2019b: 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>.
- 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>.