Skinner, P. S. and Coauthors, 2023: Interpreting guidance from the Warn-on-Forecast System, Part I: Review of probabilistic guidance concepts, product design, and best practices. *J. Operational Meteor.*, **11** (9), 110-131, doi: https://doi.org/10.15191/nwajom.2023.1109.

Journal of Operational Meteorology
Article

Interpreting Warn-on-Forecast System Guidance, Part I: Review of Probabilistic Guidance Concepts, Product Design, and Best Practices

PATRICK S. SKINNER[1,2,3], KATIE A. WILSON[4], BRIAN C. MATILLA[1,2], BRETT ROBERTS[5], NUSRAT YUSSOUF[1,2,3], PATRICK BURKE[2], PAMELA L. HEINSELMAN[2,3], BURKELY T. GALLO[6], THOMAS A. JONES[1,2,3], KENT H. KNOPFMEIER[1,2], MONTGOMERY L. FLORA[1,2], JOSHUA MARTIN[1,2], JORGE E. GUERRA[7], T. TODD LINDLEY[8], CHAD GRAVELLE[9], AND STEPHEN W. BIEDA III[10]

[1]Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma,
[2]NOAA/OAR/National Severe Storms Laboratory, and
[3]School of Meteorology, University of Oklahoma, Norman, Oklahoma
[4]The RAND Corporation, Chicago, Illinois
[5]CoreLogic Inc., Irvine, California
[6]USAF, 16th Weather Squadron, Offutt AFB, Nebraska
[7]Project Canary, Denver, Colorado
[8]NOAA/National Weather Service, Norman, Oklahoma
[9]NOAA/National Weather Service, Southern Region Headquarters, Fort Worth, Texas
[10]NOAA/National Weather Service Headquarters, Silver Spring, Maryland

(Manuscript received 6 January 2023; review completed 8 August 2023)

ABSTRACT

The Warn-on-Forecast System (WoFS) is a convection-allowing ensemble prediction system designed to primarily provide guidance on thunderstorm hazards from the meso-beta to storm-scale in space and from several hours to less than one hour in time. This article describes unique aspects of WoFS guidance product design and application to short-term severe weather forecasting. General probabilistic forecasting concepts for convection allowing ensembles, including the use of neighborhood, probability of exceedance, percentile, and paintball products, are reviewed and the design of real-time WoFS guidance products are described. Recommendations for effectively using WoFS guidance for severe weather prediction include evaluation of the quality of WoFS storm-scale analyses, interrogating multiple probabilistic guidance products to efficiently span the envelope of guidance provided by ensemble members, and application of conceptual models of convective storm dynamics and interaction with the broader mesoscale environment. Part II of this study provides specific examples where WoFS guidance can provide useful or potentially misleading guidance on convective storm likelihood and evolution.

1. Introduction

A key practical challenge in ensemble forecasting is finding ways to distill information from an ensemble that may include dozens (or more) individual members into guidance products that provide insight into specific forecast problems. In global ensembles used for medium-range forecasting, products are often produced for continuous fields (e.g., 500 hPa geopotential height) and provide information on the most likely solution (e.g., ensemble mean), forecast uncertainty (e.g., ensemble spread), and feature-based aggregates of individual member forecasts, such as "spaghetti" plots of specific contours in individual members (Sivillo et al. 1997). Similar information on likelihood, severity, and uncertainty is conveyed in convection-allowing[1] ensemble (CAE) products; however, differences in the typical forecast applications of CAEs and global ensembles result in differences in ensemble product design. For example, deterministic

[1] Convection-allowing refers to models with sufficient resolution to preclude the need for cumulus parameterization, which is typically considered to be horizontal grid spacing at or below 4 km (e.g., Schwartz and Sobash 2017).

Corresponding author address: Patrick Skinner, 120 David L. Boren Blvd, Norman, OK 73072
E-mail: patrick.skinner@noaa.gov

convection-allowing models and CAEs are often used for short-term prediction of convective storms using discrete output variables, such as accumulated rainfall, simulated reflectivity (Koch et al. 2005), or updraft helicity (UH; Kain et al. 2008), which require specialized post-processing techniques to extract and visually communicate the most relevant forecast information (Schwartz and Sobash 2017; Roberts et al. 2019; Schwartz et al. 2019).

Over the past decade, most real-time guidance produced by CAEs has focused on predicting high-impact convective phenomena in the "next-day" (0–36 hr) time frame (Xue et al. 2007; Jirak et al. 2012; Schwartz et al. 2015; Johnson et al. 2017; Clark et al. 2018; Gallo et al. 2020; Roberts et al. 2021; Dowell et al. 2022). These forecasts employ neighborhood methods (Schwartz and Sobash 2017) to provide probabilistic guidance of convective hazards over space and time scales analogous to watches or daily convective outlooks issued by the Storm Prediction Center (SPC; Clark et al. 2009; Sobash et al. 2011, 2016; Schwartz et al. 2014, 2015; Gallo et al. 2016, 2018, 2019; Loken et al. 2017, 2019; Roberts et al. 2019, 2020).

In contrast to these next-day CAEs, the Warn-on-Forecast System (WoFS; Stensrud et al. 2009, 2013) is designed for "next-hour" (0–6 hr) prediction of storm-scale phenomena at lead times between current typical National Weather Service (NWS) convective watch and warning products (e.g., Wheatley et al. 2015; Yussouf et al. 2015, 2016; Jones et al. 2016, 2018a,b, 2019, 2020; Lawson et al. 2018; Skinner et al. 2018, Flora et al. 2019, Yussouf and Knopfmeier 2019). A key distinction in the primary functions of WoFS and next-day CAEs stems from this difference in timescales: Next-day CAEs are designed to provide probabilistic guidance of *regional convective threats* and WoFS is designed to provide probabilistic guidance for hazards associated with *individual convective storms*. The focus of WoFS on predicting the evolution of individual convective storms on space and time scales roughly between a watch and warning offers the potential for the system to serve as a bridge between observations and next-day CAE guidance, filling a gap in probabilistic hazard guidance and supporting the Forecasting A Continuum of Environmental Threats (FACETs; Rothfusz et al. 2018) paradigm.

The first demonstration of a prototype WoFS [then known as the NSSL Experimental Warn-on-Forecast System for ensembles (NEWS-e)] in an operational environment occurred during the 2017 Hazardous Weather Testbed Spring Forecasting Experiment (SFE; Clark et al. 2012; Gallo et al. 2017, Clark et al. 2022), where WoFS output was used in two separate experiments.

In the first experiment, SFE participants and an expert lead forecaster collectively used real-time WoFS guidance to issue four 1-hr probabilistic outlooks of total severe weather (Choate et al. 2018). At the end of the experiment, the lead forecaster prepared a summary report that identified needs for improving the utility of real-time WoFS guidance. This report identified forecaster training as the top priority for advancing WoFS into operations, specifically mentioning the need for examples of effective ways to use WoFS guidance and descriptions of the strengths and weaknesses of the available probabilistic guidance products, D. Imy (2017, personal communication). The importance of developing training material for WoFS guidance has since been reiterated by another expert forecaster in the 2018 SFE, J. Hales (2018, personal communication), and through numerous interactions with National Weather Service (NWS) meteorologists in National Centers and Local Forecast Offices (Wilson et al. 2019a; Burke et al. 2022).

The second WoFS experiment conducted during the 2017 SFE utilized a survey to assess participant knowledge of different probabilistic forecasting concepts (Wilson et al. 2019b). The results of this survey found that a majority of respondents correctly interpreted probabilistic guidance products. However, a sizable minority (up to 40% of respondents for some questions) misinterpreted the guidance, with incorrect responses provided by participants from both the operational and research meteorology communities. The frequency of misinterpretation demonstrates the presence of a knowledge gap in interpreting uncertainty information from CAEs. Specific misinterpretations included the following: 1) inferring specific accumulated rainfall values from a probability of exceedance product; 2) interpreting an ensemble percentile product as an aggregation of values rather than a rank within a probability density function; and 3) inconsistent attribution of the impact of increasing ensemble spread with time on probability of exceedance products for different meteorological variables. For example, increasing spread was more often correctly identified for a presumably more familiar variable (accumulated rainfall) than for a less familiar one (2–5 km UH). Each of these misinterpretations involve application of deterministic forecasting concepts to probabilistic guidance and underscore the need for development of training material describing the methodology, strengths, and limitations of different probabilistic forecasting products.

The need for development of training material for probabilistic forecast guidance is well established and not limited to convection-allowing scales [e.g., National Research Council (NRC) 2006; Novak et al. 2008; AMS 2008; Demuth et al. 2020]. This initial paper in the two-part series attempts to partially address this community need by providing a brief review of probabilistic forecasting concepts in CAEs followed by a description of WoFS guidance product design and motivation. Unique aspects of WoFS probabilistic guidance and general recommendations for WoFS usage in short-term severe weather prediction are then discussed. Part II of this study examines the evolution of WoFS guidance in severe weather and flash flooding case studies and provides specific examples of scenarios where WoFS has provided either accurate or inaccurate storm-scale forecasts.

2. Probabilistic forecasting in convection-allowing ensembles

a. *General concepts, communication, and potential value of probabilistic forecasts*

The meteorological community has long recognized the limitations inherent in deterministic weather forecasts for decision making, which has led to numerous calls for the use of probabilistic forecast information. Identification of these limitations is not a new topic of discussion. For example, after demonstrating limitations of categorical weather forecasts for decision makers who have a variety of operational responsibilities, Thompson (1952) suggested providing an estimate of the probability of occurrence of hazardous weather as a practical solution for overcoming the limitations. This suggestion was further supported in a study by Murphy (1977), where the cost-loss ratio concept was used to illustrate the added value of reliable probabilistic forecasts for decision making when compared to the value provided by climatological and categorical forecasts.

In more recent years, the NRC (2006) and AMS (2008) updated the current state, use, and challenges related to probabilistic weather forecasting. These reports continue to reinforce that economic and social benefits are an expected outcome from the use of probabilistic information, and the potential benefits of such forecasts has been demonstrated in numerous studies. For example, during experimental forecasting tasks, participants were found to make improved forecast decisions related to wind and road temperature hazards when uncertainty information is provided (Joslyn et al. 2007; Joslyn and LeClerc 2012). The use of intervals in temperature forecasts was also shown to aid decision making, and participants were found to have less overall forecast uncertainty when uncertainty information was available than when only deterministic guidance was available, resulting in better identification of when a hazardous weather event will occur (Savelli and Joslyn 2013). Furthermore, a variety of user groups, including broadcast meteorologists, the general United States public, and emergency managers have expressed an overall preference and willingness to use probabilistic forecast information (e.g., Morss et al. 2008; Demuth et al. 2009; Fundel et al. 2019).

Despite scientific support for incorporating probabilistic information in weather forecasting, uncertainty information is oftentimes not communicated to users. From a numerical weather prediction perspective, providing a user base of operational meteorologists with quantitative information on forecast uncertainty often requires an ensemble. Ensemble forecasts both quantify uncertainty and provide improved skill over deterministic forecasts (e.g., Epstein 1969; Leith 1974), including at convection-allowing scales (e.g., Clark et al. 2011; Loken et al. 2017; Schwartz et al. 2017); however, effective ensemble post-processing and visualization techniques are needed to realize these potential benefits (AMS 2008; Hirschberg et al. 2011; Kaye et al. 2012).

A number of challenges for effective ensemble post-processing, visualization, and dissemination remain (NRC 2006). In a review of the social and behavioral science needs existing within the weather enterprise, the National Academy of Science (NAS; 2018) identified research topics that would help overcome these challenges, including "*examination of ways to display probabilistic forecast information for accurate interpretation*". Furthermore, effective probabilistic guidance products must serve the forecast needs of end users (Roebber et al. 2004) and may include a combination of deterministic and probabilistic guidance (Demuth et al. 2020). In the context of short-range severe weather prediction within the NWS, this requirement prioritizes guidance products that can be rapidly interrogated by end users and provide information on multiple relevant forecast properties (e.g., storm location, mode, and propagation) within a single visualization.

b. Forecast product generation in convection allowing ensembles

As prediction of convective storm hazards is a primary function of both next-day and next-hour CAEs, many probabilistic guidance products are similar in design. However, the varying time and space scales of interest result in importantces in the design, application, and interpretation of probabilistic guidance products. General descriptions of common post-processing techniques and probabilistic guidance products used at WoFS timescales are provided below. However, many techniques are analogous to next-day applications and readers are referred to recent studies by Schwartz and Sobash (2017), Roberts et al. (2019), and Schwartz et al. (2019) for thorough discussions pertaining to next-day CAEs.

1) Neighborhood Methods

Despite the relatively fine horizontal grid spacing employed by CAEs, probabilistic guidance products are typically not presented at the grid scale (e.g., Schwartz and Sobash 2017; Roberts et al. 2019). The reasoning behind this choice is that small displacement errors may reduce the usefulness of probabilistic forecasts at the grid scale. For example, small variations in the location of a small-scale feature, such as a mesocyclone, in different ensemble members may result in low grid-scale probabilities of feature occurrence within a region, even if every ensemble member has predicted a mesocyclone. These same small displacement errors are responsible for the “double penalty”, where a close forecast results in a contingency table classification of a predicted event as a “miss” and “false alarm” rather than a “hit”, when applying traditional forecast verification measures to convection-allowing scales. Therefore, neighborhood approaches are commonly used for probabilistic forecast product generation (e.g., Schwartz et al. 2010) and verification (e.g., Ebert 2009; Gilleland et al. 2009).

Current real-time WoFS guidance products are generated as the neighborhood maximum ensemble probability (Schwartz and Sobash 2017), which spreads and smooths ensemble member forecasts of relevant variables for forecasting hazards within convective storms. The value at each WoFS gridbox is replaced with the maximum value from a square neighborhood of variable size surrounding the gridbox. The likelihood of exceeding a prescribed threshold is calculated using the resulting fields and then smoothed using a 15x15 km convolution kernel, producing the final probabilistic product.

The primary difference in neighborhood methods applied to next-day CAEs and WoFS is that much smaller neighborhoods are used by WoFS to prevent smoothing scales associated with individual thunderstorms (Flora et al. 2019; Fig. 1). Many next-day CAEs employ a 40-km radius for finding the neighborhood maximum, which matches convective outlook products from the SPC (Schwartz and Sobash 2017; Roberts et al. 2019). Short forecast lead times and rapidly-cycled data assimilation in WoFS systems can accurately analyze and skillfully predict individual thunderstorms at spatial scales similar to NWS warnings (Skinner et al. 2018; Flora et al. 2019; Miller et al. 2022; Guerra et al. 2022), which allows much smaller neighborhood radii to be used. Current WoFS real-time guidance employs neighborhood radii of 13.5, 7.5, and 4.5 km (Fig. 1), which are chosen to provide probabilities of hazards within individual thunderstorms over an area roughly the size of a typical NWS warning product.

2) Ensemble Probability Of Exceedance and Percentile Products

Probabilistic guidance products are created using the distribution of forecast solutions provided by an ensemble. The ensemble probability density function (Fig. 2) at each gridpoint within the ensemble domain can be used to provide geographic visualizations of event likelihood (i.e., probability of exceedance) or severity (i.e., ensemble percentiles).

Uncertainty in weather forecasts has been expressed as a probability for well over a century (Murphy 1998). Given their history and familiarity to most forecasters, probability of exceedance products are unsurprisingly a staple of both global (Sivillo et al. 1997) and convection-allowing (Roberts et al. 2019; Schwartz et al. 2019) ensembles. These products provide a measure of the likelihood of an event, such as measurable precipitation at a given location, and can provide limited information on event severity as well. For example, nonzero probabilities of 2–5-km layer UH values >200 $m^2\ s^{-2}$ imply the potential for a strong mesocyclone (Sobash et al. 2016). However, specific measures of severity that

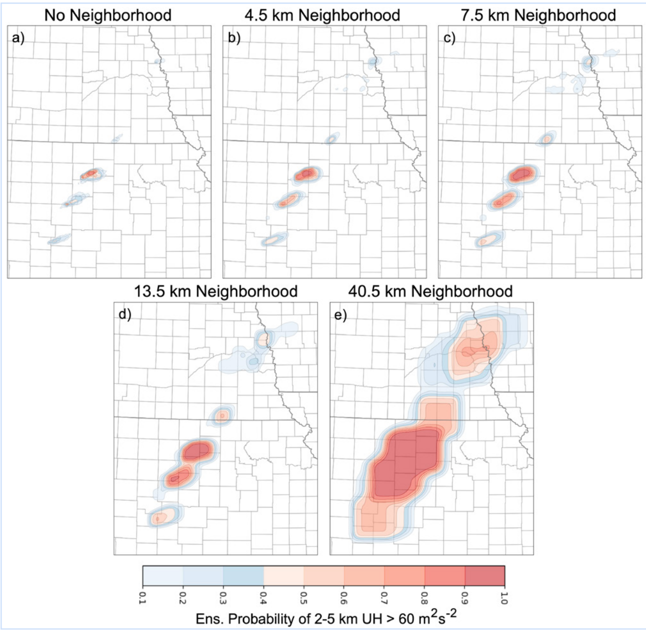

Figure 1. Example Neighborhood Maximum Ensemble Probability products for a 1-h 2-5 km UH forecast initialized at 2300 UTC on 1 May 2018 using (a) no neighborhood, (b) a 4.5-km radius neighborhood, (c) a 7.5-km radius neighborhood, (d) a 13.5-km radius neighborhood, and (e) a 40.5-km radius neighborhood similar to next-day CAE guidance. (*Figure adapted from Flora et al. 2019*). *Click image for an external version; this applies to all figures hereafter.*

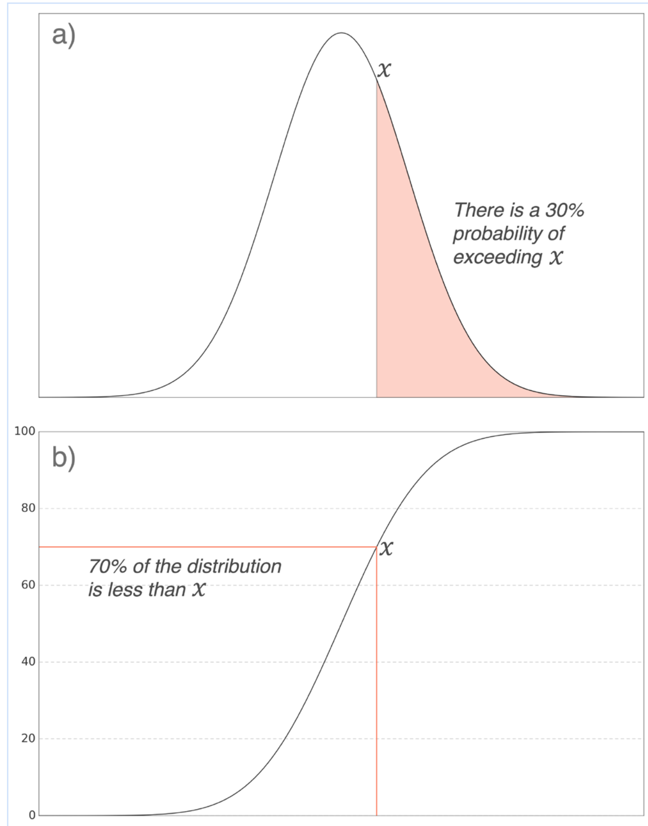

Figure 2. Schematic of an idealized (a) probability density function and (b) cumulative distribution function provided by an ensemble. The position of a forecast variable with fixed intensity 'X' within the distribution can be used to calculate the ensemble probability of exceeding 'X'. Alternatively, the value of 'X' at a fixed position (percentile) within the ensemble can be used as a measure of severity (*Figure reproduced from Wilson et al. 2019*).

span the range of ensemble solutions are desirable to forecasters (Novak et al. 2008; Evans et al. 2014) and are not easily extracted from probability of exceedance products.

Specific measures of severity can be found using values at a fixed position within the ensemble distribution, represented by a percentile, as opposed to finding the proportion of the ensemble exceeding a fixed value (Fig. 2b). The most intuitive, and likely familiar, percentile product is the ensemble maximum, which typically represents a worst-case forecast scenario. As will be discussed below, a limitation to using the ensemble maximum is that it includes information from all member forecasts, even if only a single member in the ensemble is predicting an event. Therefore, percentiles that apply some probability threshold for including intensity information, such as the 10th and 90th percentiles, are often used to supplement the ensemble maximum (Novak et al. 2014).

In next-day CAEs, the ensemble maximum is typically used to define the *boundaries of a region* where convective hazards are possible (Roberts et al. 2019). However, WoFS is often used for predicting *hazards associated with individual storms* that have been accurately initialized within most ensemble members. Use of the ensemble maximum in these forecasts can potentially lead to overpredictions of storm coverage and severity, as all ensemble solutions are included. Use of percentiles between the ensemble median and maximum can limit this potential for overprediction by providing a worst-case forecast scenario conditional on a minimum probability of exceedance. For example, WoFS products include the 90th percentile value, which represents the maximum value that is predicted by at least 10% of ensemble members, as well as the ensemble maximum.

3) Ensemble Statistical And Probability Matched Mean

The statistical mean of an ensemble[2] is possibly the most familiar ensemble product and provides a more skillful forecast than individual ensemble members when averaged over many forecasts (Leith 1974). The improved skill in the statistical mean is derived from smoothing low-confidence events in a forecast while retaining higher-confidence features. However, though ensemble mean fields are often plotted for continuous environmental fields in WoFS and other CAEs, these systems are primarily interested in providing guidance on high-impact events with limited predictability. Discrete, high-impact events, such as thunderstorms, are often predicted with lower confidence and smoothed out in the statistical mean (Surcell et al. 2014; Clark 2017). The probability matched mean (PMM) is a post-processing technique that restores characteristic amplitudes of ensemble members to the statistical mean field (Ebert 2001) and can improve skill in forecasts of discrete fields such as accumulated rainfall (Ebert 2001; Clark et al. 2009). Though the technique can be useful for regional forecasts and simplifying visualization of ensemble output, it has multiple limitations. First, the increased skill in PMM over individual ensemble members can arise from substituting higher amplitudes into smoothed values in the statistical mean (Surcel et al. 2014). In this case, the increased skill results from smoothing of smaller spatial scales in the statistical mean and the PMM does not necessarily retain information on low-confidence, high-impact events that may be of interest to forecasters. Secondly, if the PMM is calculated over an entire model domain (i.e., the contiguous United States), values in member forecasts may be substituted into grid boxes in the statistical mean thousands of kilometers away (Clark 2017). This limitation can be mitigated using the localized PMM described in Clark 2017; however, large differences between the amplitudes of the PMM and individual ensemble members may still occur at finer spatial scales in the localized PMM (c.f., Clark 2017; their Fig. 6). These small-scale differences in the localized PMM are more likely to produce misleading guidance in WoFS than a next-day CAE because WoFS is interested in providing guidance for storm-scale phenomena using discrete forecast variables such as simulated reflectivity and UH.

[2] This field is typically referred to as the ensemble mean. We use statistical mean throughout this section to more easily distinguish the field from the probability matched mean.

Figure 3 provides a simplified example of how both the statistical and probability matched mean can misrepresent forecast information present in individual ensemble members. In Fig. 3, the hypothetical ensemble members each produce similar forecasts of an arbitrary, non-continuous variable in the upper left portion of the domain. However, less confident predictions of

Figure 3. Schematic showing the method for calculating the probability matched mean. Members from a 3-member ensemble with nine grid boxes are shown on the left. Values in each grid box represent some forecast variable of interest and were chosen in this example to reflect an ensemble forecast with a high confidence of moderate values in the four grid boxes in the upper left portion of the domain and a low confidence of high values elsewhere. The probability matched mean is calculated by sorting values from each grid box in each ensemble member and substituting values from members into the corresponding position in a list of sorted ensemble mean values. It can be seen that neither the statistical mean nor probability matched mean produce a forecast similar to individual members in this example. The statistical mean produces similar values to members for the high-confidence region in the upper left of the domain, but smooths low-confidence, high-amplitude forecasts in the rest of the domain to a broad region of low amplitudes. In the probability matched mean, spatial translation of the highest values in the ensemble from low-confidence regions to the high-confidence region in the upper left results in introduction of a positive amplitude bias. Additionally, low amplitude values are substituted into regions where individual forecasts produced the highest values.

higher variable amplitudes are made elsewhere in the domain. The distribution of values in this example was designed to be analogous to an accumulated rainfall or reflectivity forecast for an event with a wide region of stratiform precipitation trailing a region of convective storms.

It can be seen that the statistical mean provides an accurate representation of the values in the upper left portion of the domain where there is little ensemble spread. However, the low-confidence, high amplitude values in the remainder of the domain are smoothed into a forecast field with broader coverage of lower amplitude values than any ensemble member. In this example, the PMM provides an equally misleading forecast product. The highest amplitudes in member forecasts are substituted into the gridboxes with the highest statistical mean values. This substitution introduces an intensity bias into the upper left corner of the domain. As grid boxes outside the upper left corner have the lowest statistical mean values, nonzero member values substituted into this region are of lower amplitude than forecasts from any forecast member. Though this is a highly simplified schematic, similar misrepresentations are seen in WoFS PMM fields.

4) Pseudo-Deterministic Products

While probabilistic guidance products efficiently condense information within the ensemble and provide measures of uncertainty, they provide limited information of the physical processes responsible for the evolution of model forecasts. Therefore, there is still a need, and strong demand (Novak et al 2008; Roberts et al. 2019; Demuth et al. 2020; Wilson et al. 2021), for guidance products within CAEs that provide an efficient means of examining deterministic guidance from ensemble members. The most straightforward way to provide this deterministic information is to visualize ensemble output so that different member solutions can be easily compared. The most familiar of these visualizations is a “Postage Stamp” plot that provides guidance from each ensemble member on a single plot. Postage stamps provide users with deterministic solutions from individual members; however, they sacrifice readability to the point of being impractical for large ensembles. Alternatively, web-based ensemble member viewers can provide interactive features that permit rapid interrogation of individual member solutions (Roberts et al. 2019; Schwartz et al. 2019) while preserving output readability.

A second method for displaying deterministic aspects of an ensemble forecast is to provide limited information from each ensemble member on a single plot. While these visualizations sacrifice the complexity provided by full deterministic products, they allow forecasters to rapidly assess ensemble spread in features of interest. The most familiar of these feature-based[3] visualizations (Obermaier and Joy 2014; Rautenhaus et al. 2018) is the spaghetti plot, which provides specific contours of a given field for each ensemble member (Sivillo et al. 1997). Spaghetti plots are typically employed to provide information on ensemble spread for features in a continuous field, for example shortwaves in a 500 hPa geopotential height field or airmass boundaries in a near-surface dewpoint field. WoFS products do not include traditional spaghetti plots of continuous, spatial variables, but the visualization technique is applied to forecast soundings and hodographs, where each member is plotted on top of the ensemble mean. Automated detection of features associated with specific phenomena may also be used to produce analogous visualizations to spaghetti plots for features like frontal boundaries (Hewson and Titley 2010; Chipilski et al. 2018; Lagerquist et al. 2019), tropical cyclone tracks (Hamill et al. 2012), or thunderstorm proxies (Schwartz et al. 2015). In particular, CAEs frequently use feature-based “paintball” plots to display ensemble information of thunderstorm and mesocyclone positions in simulated reflectivity and UH guidance (Schwartz et al. 2015; Roberts et al. 2019; Schwartz et al. 2019).

Although paintball plots provide valuable information on CAE prediction of convective hazards, a limitation is that they require a strict threshold to be prescribed for identifying feature boundaries (e.g., 40 dBZ in simulated reflectivity). The number and size of features identified in a field will be sensitive to the value chosen for this threshold (Wolff et al. 2014; Weniger and Friederichs 2016; Skinner et al. 2018). Furthermore, values of derived quantities such as simulated reflectivity and UH will be sensitive to multiple aspects of forecast system configuration, including but not limited to, physical parameterizations, dynamic core, and horizontal grid spacing (e.g., Morrison and Milbrandt 2015; Potvin et al. 2019). Therefore, it is important to account for the sensitivities of features to a given threshold when creating paintball plots (threshold sensitivity may also affect probability

[3] “Features” in forecasts may alternatively be referred to as “objects” or “events”.

of exceedance plots), particularly for multi-core, multi-physics CAEs such as the High-Resolution Ensemble Forecast System (HREF; Roberts et al. 2019, 2020). One method for mitigating feature sensitivity to the prescribed threshold is to set the threshold according to a percentile value in a climatology of model runs with a given configuration (e.g., Mittermaier and Roberts 2010; Sobash et al. 2016; Dawson et al. 2017; Skinner et al. 2018). Use of percentile thresholds will result in the same total area (though not necessarily the same number) of identified features in each ensemble member when considered across many forecasts and helps normalize feature identification in a multi-core or multi-physics ensemble.

3. Description of WoFS guidance product design

a. General description of WoFS configurations

The prototype WoFS has been used to provide real-time guidance since 2016[4]. Specific system configurations have changed across years and different experiments, but have remained generally similar to the original design described by Wheatley et al. 2015 and Jones et al. 2016. Subsequent variations in system configuration are described in Lawson et al. 2018, Skinner et al. 2018, Jones et al. 2018a,b, 2020, and Yussouf and Knopfmeier 2019. A general description of the current WoFS design is provided below and readers are referred to the preceding references for thorough descriptions of WoFS configurations in different years.

Real-time WoFS simulations are initialized using the High-Resolution Rapid Refresh Data Assimilation System (HRRRDAS; Dowell et al. 2022). The WoFS domain is 900x900 km and is re-locatable to cover any region within the CONUS[5] where high-impact weather is expected. The analysis component of WoFS consists of 36 members and produces ensemble analyses every 15 min through ensemble Kalman filter-based assimilation (Houtekamer and Zhang 2016) of remotely sensed and in situ observations. Most observations WoFS assimilates in each 15-min cycle are remotely sensed, including Multi-Radar Multi-Sensor (MRMS) gridded WSR-88D radar reflectivity (Smith et al. 2016) and radial velocity data (Wheatley et al. 2015) and GOES total water path and clear-sky water vapor observations (Jones et al. 2016, 2018a). When available, WoFS additionally assimilates Oklahoma Mesonet (Brock et al., 1995) surface observations every 15 min and ASOS observations each hour. As will be discussed in section 4, these rapid, 15-min assimilation cycles are a unique and important aspect of WoFS as they allow small-scale phenomena to be accurately analyzed within forecast initial conditions. In other words, given a sufficient number of observations and data assimilation cycles, WoFS will accurately analyze the location and intensity of ongoing convective storms at the beginning of a forecast (Fig. 4). An accurate analysis of an ongoing storm will result in a more accurate prediction of that storm out to at least 3 h of lead time (Guerra et al. 2022).

WoFS real-time forecasts consist of 18 members, are issued at 30- or 60-min intervals, are run for 3–6 h of forecast lead time, and provide output every 5 min. Each forecast member employs 3-km horizontal grid spacing and the NSSL two-moment microphysical parameterization (Mansell et al. 2010). A diverse set of three planetary boundary layer parameterizations and two short- and longwave radiation schemes are utilized to increase ensemble spread (Potvin et al. 2020).

[4] Though WoFS produced real-time guidance in 2016, it was configured differently (Skinner et al. 2018) and was not evaluated by NWS Meteorologists or within the SFE.

[5] The HRRRDAS and potential WoFS domains were limited to the eastern two-thirds of CONUS from 2016 through 2018.

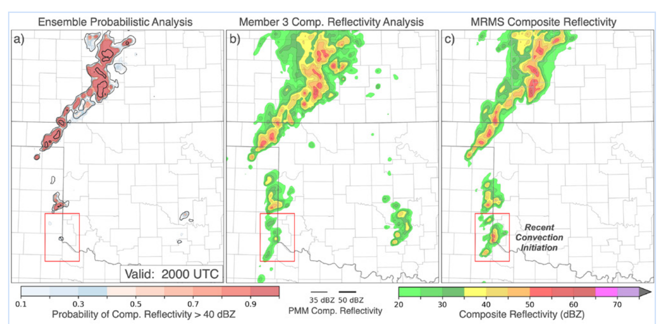

Figure 4. WoFS analyses valid at 2000 UTC 2 May 2018 of (a) probability of composite reflectivity exceeding 40 dBZ and (b) composite reflectivity in ensemble member 3. MRMS composite reflectivity observations are provided in (c) for comparison. Black contours in (a) show the 35 and 50 dBZ probability matched mean of composite reflectivity and the red rectangle in (a–c) denotes newly initiated storms that have not been accurately analyzed by the ensemble. A 600x510 km subdomain of the full 750x750 km WoFS domain is plotted to improve clarity.

b. Overview of WoFS post-processing, visualization, and dissemination

The principal challenges in post-processing, visualizing, and disseminating WoFS guidance stem from a goal of providing rapidly updating, short-term guidance for individual convective storms. The difficulty of creating useful numerical weather prediction products at these time and space scales was identified by Lilly (1990) as one of four technological challenges to be overcome in order to produce real-time numerical prediction of thunderstorms and summarized as: "The short timescales and expected rapid perishability of convective storm predictions compel almost instant transmission of output, preferably in some form which is visually acceptable both to forecasters and the general public." In practice, the duration of potentially useful information provided by WoFS is equal to the length of the forecast minus the time needed to visualize and disseminate guidance. For a short-term (e.g., 3-h) forecast, any latency in visualization results in loss of a significant proportion of usable forecast time. Furthermore, owing to the short predictability limits of convective storms (e.g., Zhang et al. 2016; Potvin et al. 2017; Weyn and Durran 2017; Flora et al. 2018), the potential value of WoFS guidance decreases relative to interrogation of current observations as forecast latency increases. As a simple example, if it takes 2 h to disseminate a 3-h WoFS forecast, the potential value is limited by both the short duration of usable guidance and by increasing forecast error during the initial 2 h of the forecast period. These challenges strongly motivate development of post-processing and visualization strategies for WoFS that provide guidance products with as little latency as possible (Fig. 5).

Currently, approximately 17 WoFS forecasts are issued each day the system is run and produce over four terabytes of model output. To reduce the memory requirements for guidance product generation, the large, three-dimensional fields in raw model output files are converted to two-dimensional summary products and visualized via custom python code based in part on the SHARPpy (Blumberg et al. 2017) and MetPy (May et al. 2022) libraries. Resulting visualizations are then disseminated using a web interface (e.g., Oakley and Daudert 2016; Roberts et al. 2019; Sobash et al. 2020). In total, more than 250 guidance products and over 10 000 (20 000) individual images are generated and made available on the web interface within approximately 30 (45) min for a single 3-h (6-h) forecast (Fig. 5).

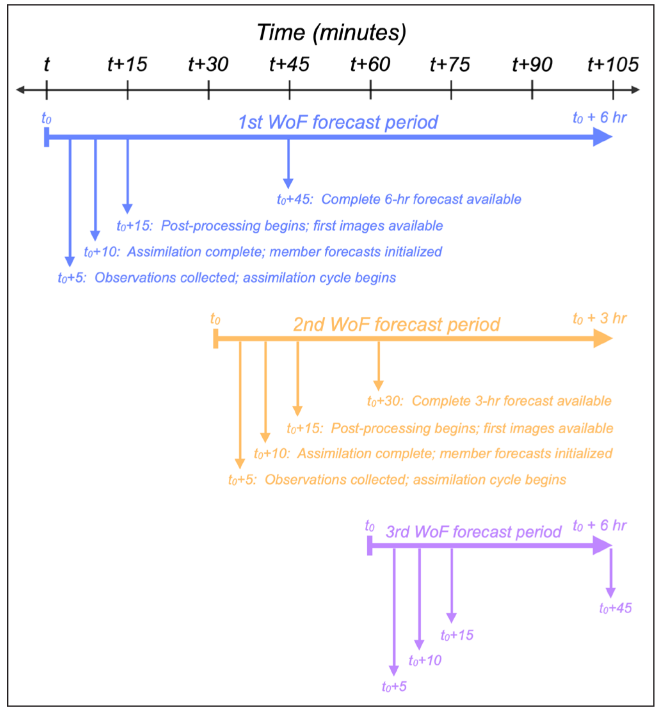

Figure 5. Timeline of WoFS forecast issuance over a 105-min period for a configuration where 6-hr forecasts are issued at the top of each hour and 3-h forecasts at the bottom of each hour. This configuration matches WoFS real-time forecasts issued in the spring of 2022. The approximate timing of selected aspects of each forecast are annotated, with time indicated relative to forecast initialization (t_0).

Guidance products are broadly categorized as products containing continuous fields related to the storm environment [e.g., 2-m temperature, Convective Available Potential Energy (CAPE), precipitable water, etc.] and products containing discrete, storm-scale information (e.g., simulated composite reflectivity, accumulated rainfall, UH, etc.). In order to preserve storage space, minimize latency, and limit the number of available guidance products on the website, only the ensemble mean of environmental products is provided . However, multiple probabilistic products are produced for storm-scale variables using different neighborhoods and exceedance thresholds. The neighborhood and exceedance thresholds available for each storm-scale product will vary based on the characteristics of the product.

With over 250 different products available, real-time interrogation of every available product in every available forecast is obviously not possible, but because relevant environmental and storm-scale guidance products will vary from case-to-case and across individual thunderstorms within a single forecast, each guidance product has the potential to provide valuable information. In practice it has been found that users develop a suite of a few broadly applicable guidance products that are checked regularly (e.g., the composite reflectivity paintball and 90th percentile of 2–5 km UH products) and supplement with additional products on a case-by-case basis (Wilson et al. 2021). This need for guidance products covering a wide variety of forecast scenarios can conflict with the requirement that guidance is disseminated with minimal latency and motivates careful design of specific guidance visualizations.

c. Considerations for visualization of WoFS guidance

The compressed timeline of short-term WoFS forecasts creates an additional challenge for operational meteorologists using WoFS[6] guidance. Namely, meteorologists issuing short-term forecasts of hazards associated with convective storms may be experiencing an excessive cognitive workload stemming from a need to condense a large amount of observational data on rapidly-evolving phenomena into forecast products for the public (Wilson et al. 2017; Karstens et al. 2018; James et al. 2020). User workload is an important consideration for designing new guidance products as workload levels that are too high (overload) can degrade human performance (Young et al. 2015). Integration of short-term, ensemble WoFS guidance, which provides hundreds of individual products every 30 min into an already demanding workflow, has the potential to exacerbate task demands on meteorologists and limit the utility of WoFS. It is therefore crucial that WoFS guidance be provided in a way that does not negatively impact users' cognitive workload. This consideration drives design of both the WoFS web interface and individual product visualizations.

The basic challenge in designing a WoFS guidance visualization is to provide as much salient information as possible in a manner that is easily accessible. As improvements to a forecast visualization can both improve the quality of forecast information communicated to end users and reduce the time needed for users to acquire that information (Hegarty et al. 2010; Ling et al. 2015; Klockow-McClain et al. 2020; Calvo et al. 2022), product design is an important component of successful use of real-time WoFS guidance. Successful product design requires consideration of tradeoffs between information density, aesthetic, and ease of interpretation, as well as product accessibility and interactivity across a spectrum of end users.

An initial challenge for probabilistic product design is that there are not straightforward, standardized methods for visualizing uncertainty in numerical weather prediction guidance (Kaye et al. 2012), and synthesizing probabilistic and deterministic aspects of guidance, although desirable to forecasters (Demuth et al. 2020), can result in overcomplicated products that are difficult to interpret. As an example, a forecaster interested in storm-scale severe weather guidance from WoFS could be interested in measures of likelihood of severe thunderstorms, potential severity of storms, and pseudo-deterministic information on storm location and evolution. These measures can all be visualized in a single product; however, even with alterations to colormaps, contour intervals, and opacity, the resulting visualization will be excessively complicated, take longer to accurately interpret, and potentially mask useful information (Hegarty et al. 2010; Kaye et al. 2012). Therefore, WoFS storm-scale guidance products visualize a single component of probabilistic information (i.e., likelihood or potential severity; Fig. 6a) with interactive options employed in the web viewer to allow end users to create custom overlays (Fig. 6b, c).

A second aspect of product design for WoFS is to use color, contour intervals, and opacity that maximize accessibility and highlight what are expected to be the most meteorologically relevant features. The first requirement in choosing color palettes for visualizations is that they are accessible to a broad population, including those with color vision deficiency. All color schemes used in WoFS guidance products are derived from the ColorBrewer2 repository and maintain color specification for users with colorblindness (Brewer et al. 2003). A second consideration is choosing a visualization aesthetic that draws viewer attention to the most meteorologically relevant features. To accomplish this goal, WoFS guidance products employ linearly segmented contour intervals and colormaps rather than continuous intervals and perceptually uniform colormaps. This choice is made because most meteorological fields are not interpreted continuously (users likely don't need to discern the difference between 83ºF and 84ºF) and changes between two specific values may be more relevant to the forecast

[6] WoFS guidance and product listing for past cases from 2017 through 2021 are available at wof.nssl.noaa.gov/realtime/, and guidance from a cloud-based WoFS system run in 2022 is available at https://cbwofs.nssl.noaa.gov/Forecast.

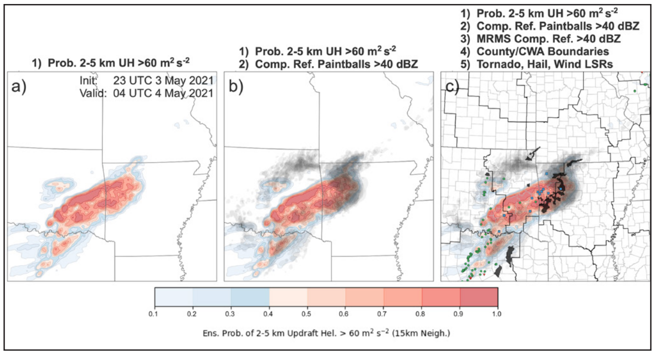

Figure 6. Examples of interactive features of WoFS guidance products using a 5-hr forecast initialized at 2300 on 3 May 2021. The base product (a) accumulated probability of 2–5 km UH exceeding 60 $m^2\ s^{-2}$ in a 15-km neighborhood is combined with (b) reflectivity paintballs exceeding 40 dBZ (gray) at 0400 UTC 4 May 2021 and (c) county (light gray) and NWS County Warning Area (dark gray) boundaries, MRMS reflectivity paintballs exceeding 40 dBZ (black), and Local Storm Reports of tornadoes (red triangles), severe hail (green circles), and severe straight-line winds (blue squares). MRMS reflectivity observations are available as they are collected in real time and Local Storm Report information is added the day following a WoFS run.

than an identical interval across different values. As an example, meteorologists often roughly categorize values of CAPE as "low" (1–1000 J Kg^{-1}), "moderate" (1000–2000 J Kg^{-1}), "high" (2000–3000 J Kg^{-1}), or "extreme" (>3000 J Kg^{-1}) for spring severe weather prediction in the central United States. Using a linear segmented colormap to visualize a CAPE field can highlight transitions between these values and facilitate rapid interpretation by end users (Fig. 7).

d. Considerations for real-time use of WoFS guidance

The previous section describes reasoning behind design choices for a single visualization of a WoFS guidance product valid at a single time. However, considerations must also be made for the generation and real-time display of these products. Effective use of real-time WoFS guidance requires users to be able to rapidly navigate between different forecast products and valid times. This navigation is accomplished through a web-based viewer that hosts individual guidance product images (https://cbwofs.nssl.noaa.gov/Forecast; Fig. 8). The design of the WoFS web viewer follows best practices for web display of atmospheric data (Oakley and Daudert 2016) and is based on other web viewers used for dissemination of numerical weather prediction guidance, particularly for the NCAR convection-allowing ensemble (Schwartz et al. 2019; Sobash et al. 2020) and the High Resolution Ensemble Forecast (HREF; Roberts et al. 2019). The web interface allows users to select specific dates, forecast initialization times, and guidance products from a series of drop-down menus. Forecast valid times are navigated using slider bars at the top of the viewer, keyboard shortcuts, or an animation feature. The ability to rapidly animate all possible forecast times (most easily accomplished using the "<" and ">" keyboard shortcuts) is particularly useful for WoFS guidance as the 5-min output frequency allows users to view more details of forecast evolution than is possible with the more common 1-h output frequency in most next-day CAEs (see animation of Fig. 7).

Forecasters demonstrate diverse product usage for predicting different severe weather threats (Wilson et al. 2021) and changes in storm mode or coverage can negatively impact readability for fixed combinations of fields. Therefore, an important function of the WoFS web viewer is to allow end users to customize products for specifics of a given case[7] with a goal of enhancing the baseline visualizations of guidance products (McInerny et al. 2014; Calvo et al. 2022). Current interactive elements in the WoFS viewer permit customizable overlays of geopolitical boundaries, pseudo-deterministic WoFS products, current observations from the MRMS system (Smith et al. 2016), or non-real-time verification information (available the following day) from NWS warnings and local storm reports (Fig. 6). In particular, the ability to overlay pseudo-deterministic guidance is valuable for creating "combination" probabilistic-deterministic guidance products (Demuth et al. 2020) that provide information on both the structure and evolution of individual storm features and uncertainty of the predicted evolution (Fig. 6b). Additionally, as Doppler radar observations are a foundation of short-term severe weather forecasting, real-time overlays of MRMS observations facilitate rapid diagnosis of storms where WoFS either contains large errors in the analysis or forecast, or where an

[7] Interactivity in the WoFS web viewer is limited by an inability to incorporate multiple different meteorological datasets. Ingestion of WoFS guidance into standalone software such as the Advanced Weather Interactive Processing System (AWIPS2) is needed to realize a fully interactive interrogation of varied observational and NWP datasets.

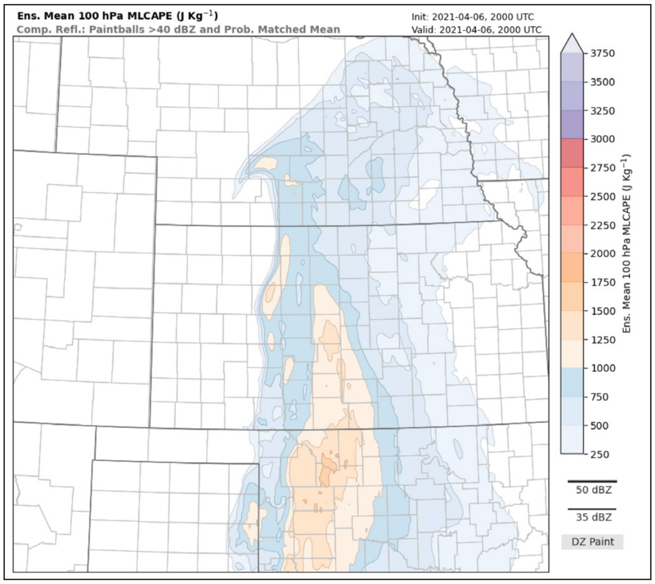

Figure 7. WoFS ensemble mean analysis of mixed layer CAPE (J Kg^{-1}) for the lowest 100 hPa of the atmosphere at 2000 UTC on 6 April 2021. Click for animation of the 6-hr forecast of mixed layer CAPE with an overlay of WoFS reflectivity paintballs exceeding 40 dBZ.

accurate analysis makes an accurate resulting forecast more likely (Guerra et al. 2022).

This section has described the current state of the WoFS post-processing and visualization software for real-time cases run during the spring of 2022. However, the 2022 WoFS product suite and web viewer have evolved significantly from initial versions introduced in 2016. The year-by-year changes to the viewer were intended to improve usefulness to end users (Roebber et al. 2004) and were driven by collaboration with NWS partners (Tables 1, 2; Wilson et al. 2019a,b, 2021; Burke et al. 2022). The Operation-to-Research flow of information resulting from WoFS end user research follows the principles of user-centered design (Abras et al. 2004) and has improved the information content, functionality, and ease of use of the WoFS guidance products and web viewer (Table 1).

Responses to survey tools provided to NWS meteorologists who have used WoFS guidance in real time provide preliminary, anecdotal evidence that the WoFS guidance products and web viewer are achieving their goals of providing useful information for thunderstorm hazard forecasting without adding significantly to the cognitive workload of forecasters. Table 2 includes example feedback on WoFS impacts on forecaster confidence and workload received from four different Weather Forecast Offices (WFOs) in Texas and Oklahoma, as well as SPC for two severe weather events in 2020 (28 April and 7 May). In general, respondents felt that WoFS increased confidence in their forecasts (nine of ten respondents said that WoFS guidance greatly or slightly increased their forecast confidence) without significantly impacting their workload (four of ten forecasters found that WoFS slightly increased their workload while five of ten did not notice a change and one forecaster found their workload decreased slightly). Though analysis of user feedback from many more cases is needed to confidently assess the impacts of WoFS guidance on the forecast process, these preliminary results indicate its potential value to end users without a correspondingly large increase in cognitive workload.

4. Interpreting WoFS guidance for severe weather forecasting

We have thus far reviewed probabilistic guidance products for CAEs and described the motivation behind post-processing and visualization methods for WoFS guidance. In this next section, we will describe novel aspects of WoFS relative to next-day CAEs and provide general recommendations for interpretation of WoFS guidance. Specific examples of the concepts discussed in this section are included in Part II.

a. Impacts of rapidly cycled data assimilation

WoFS design does not differ significantly from other CAEs; it shares a dynamic core, physical parameterizations, and data assimilation methods with next-day CAEs and is initialized within the HRRRDAS (Dowell et al. 2022). Despite these similarities, WoFS is designed to provide guidance for a fundamentally different forecast problem than next-day CAEs: Prediction of the structure, intensity, and evolution of individual convective storms. This distinction between providing guidance for specific convective storms rather than regions of convective storms drives differences in configuration choices between WoFS and next-day CAEs. The predictability limits of convective storms (e.g., Zhang et al. 2016; Weyn and Durran 2017; Potvin et al. 2017; Flora et al. 2018) constrain useful lead times for predicting individual storms to periods less than approximately 6 hrs and motivate the selection of short forecast durations in WoFS. Similarly, accurate prediction of individual storms requires an accurate analysis of those storms in model initial conditions, which motivates the use of rapidly cycled (15 min) assimilation of Doppler radar and satellite observations.

Table 1. Summary of user-suggested changes to WoFS guidance products or web viewer. The suggested change, initial suggesting agency [either the Storm Prediction Center (SPC), Weather Prediction Center (WPC), or local Weather Forecast Office (WFO)], and year the change was implemented are provided.

Suggested Change	NWS Collaborator(s)	Year Implemented
Hourly and 4-Hourly Probabilistic “Swath” Products	SPC	2017
Extension of Forecast Duration to 6-hr	SPC	2018
Addition of Heavy Rainfall Environmental and Probabilistic Products	WPC	2018
Remove 21 km Radius Neighborhood, add 7.5 km Neighborhood	WFO	2019
Addition of Forecast Soundings	SPC, WPC, WFO	2019
Redesign of Web Viewer similar to HREF Viewer	SPC, WPC, WFO	2020
Interactive Overlay of Reflectivity Paintballs or PMM	WFO	2021
Non-Real Time Overlay of Warning and LSR Products	SPC, WFO	2021
Ability to toggle between prior hour and current forecast	SPC, WPC, WFO	2021
Keyboard shortcut to cycle between probabilistic products	WFO	2022

This configuration choice results in a primary distinction between WoFS guidance and next-day CAEs:

- The quality of WoFS analyses and resulting forecasts for a specific convective storm is a function of the number of radar and satellite observations that have been assimilated, with greater accuracy achieved for more mature storms (Guerra et al. 2022; Fig. 4). Therefore, *the quality of WoFS guidance is a function of storm age and will vary from storm to storm in any given forecast*.

A consequence of this distinction is that each WoFS forecast can be thought of as a collection of forecasts for each storm within the domain rather than a single forecast comprising the entire forecast domain. Successful interrogation of these forecasts requires users to correctly identify storms where WoFS has produced accurate analyses of convective storms and their surrounding environments and where it has not. This distinction is not a trivial task, but is facilitated by comparison to current radar and satellite observations, such as real-time MRMS observations within the WoFS viewer (Fig. 6c), which allows users to rapidly diagnose storms that have been accurately initialized.

A second unique aspect to WoFS guidance that stems from the correlation between storm age and analysis accuracy is that WoFS guidance will evolve across forecast initializations more rapidly than CAEs with less frequent data assimilation, particularly

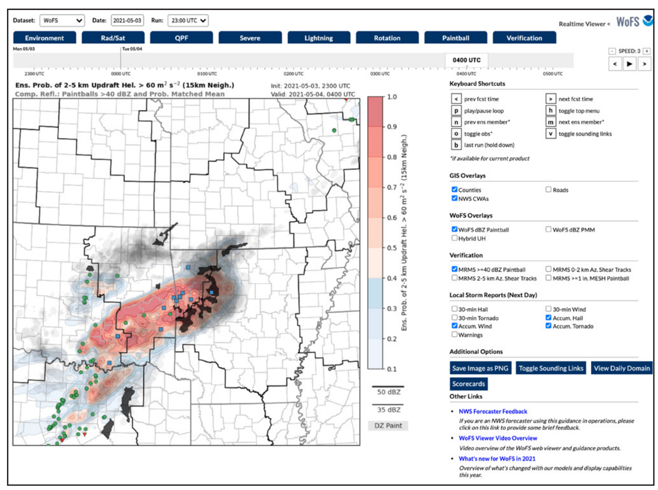

Figure 8. Screen capture of the WoFS web viewer for a 5-hr forecast initialized at 2300 UTC on 3 May 2021. Date and guidance product navigation features are located at the top of the viewer, with the product visualization display at the lower left. The right side of the viewer features navigation instructions, interactive overlays, display options, and external links from top to bottom. The product displayed matches options used in Fig. 6c.

Table 2. Sample of user feedback on real-time WoFS usage on 28 April and 7 May 2020 from NWS forecasters representing four WFOs and SPC. Representative comments on how WoFS impacted forecaster confidence and workload from five of the ten responses received from those cases are provided.

WoFS Impact on Forecast Confidence
"The event was fairly predictable to begin with. WoF gave one more piece of data confirming our expectations from mesoanalysis. It also added some confidence in timing of the event as well as timing for increases in specific threats."
"While I was confident on the type of event we were forecasting, I was unsure whether there would be discrete cells across the far western part of our CWA [County Warning Area]. The WoF guidance hinted at the possibility of discrete cells, which did actually develop across our far western counties."
"There was a bit of uncertainty introduced into the forecast as the WoF guidance did not show the weaker, elevated convection that had developed earlier in the west."
"I was skeptical of the hail forecast where the mode was more linear. And it appeared to have a more linear mode over more of the area that was occurring. So the forecast of hail was viewed with skepticism. I felt that locations that saw more cellular convection would see more hail, and that did occur, but that wasn't reflected in the guidance."
The WoFS matched exactly how I expected the event to play out from the environment and the synoptic setup. Because I had guidance showing what I expected to happen, it increased my confidence to run with that solution."
WoFS Impact on Forecaster Workload
"once the storm moved in the workload increased slightly mostly because I was trying to determine whether the storm would approach the Metroplex …, and the WoFS was an additional piece of guidance to look at."
"I spent more time looking at [WoFS] and less time looking at the HRRR."
"Working remotely gave me more time to continually investigate the sub-hourly guidance so workload was increased naturally by that versus looking solely at satellite, radar and hourly HRRR output."
Given how we've been using the guidance for the last several years, it wasn't much of a workload issue."
"This was my first time using the guidance so I was looking through the site, but I expect the workload shouldn't increase too much as I become familiar with the website and know what I would like to see."

immediately after convection initiation (CI). Following CI, the assimilation of spatially dense radar and satellite cloud water path observations will typically allow WoFS to initialize ongoing storms not present in pre-CI forecasts following about 3–6 assimilation cycles (45–90 min; Guerra et al. 2022). This transition in WoFS guidance following CI represents a shift from predicting regional, mesoscale threats of severe weather hazards pre-CI similar to next-day CAEs, to predicting the track and intensity of specific storms near the warning scale (Fig. 9). This shift generally results in rapid increases in the confidence and specificity of predicted storm tracks, and for severe hazards, it often marks a transition from a "low confidence, high impact" forecast of severe thunderstorms to a "high confidence, high impact" forecast (Galarneau et al. 2022).

This evolution of WoFS guidance with rapidly-cycled assimilation of satellite and radar observations can also allow WoFS guidance to "catch up" to and correct initially poor forecasts. For example, Part II will consider an example where WoFS predicts a predominately cellular storm mode in early forecasts and a linear storm mode develops. Later WoFS forecasts that assimilate observations of the ongoing mesoscale convective system (MCS) into their initial conditions provide more accurate guidance than earlier forecasts. This potential for rapid evolution in WoFS forecasts,

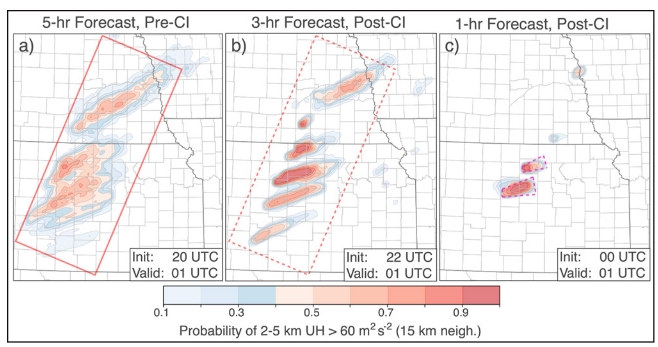

Figure 9. WoFS accumulated swaths of the probability of 2–5 km UH exceeding 60 m^2 s^{-2} on 1 May 2018. WoFS forecasts are initialized at (a) 20 UTC, (b) 22 UTC, and (c) 00 UTC and all forecasts are valid at 01 UTC. The evolution of predicted UH coverage from pre-CI to post-CI is illustrated using red (magenta) polygons as hypothetical convective watch (convective warning) areas.

coupled with the rapid issuance of forecasts, suggests that users should not discount WoFS guidance for an event if initially poor storm-scale forecasts are issued.

Finally, a special situation where rapidly-cycled data assimilation influences the usability of WoFS forecasts occurs during the period immediately following CI (i.e., for approximately the first hour of a storm's lifecycle). During this period, data assimilation will be improving the analysis of the newly initiated storm, but large errors in the analyses may still persist, with inconsistent error magnitudes across different ensemble members. This inconsistency results in a scenario where a subset of the ensemble may produce an accurate analysis and forecast, but the majority of the ensemble will not. In this situation, probability of exceedance guidance using small neighborhoods will likely be unreliable and biased towards lower likelihoods of event occurrence because the majority of the ensemble did not produce an accurate analysis. However, the ensemble members that have produced an accurate analysis may still provide useful information. Therefore, pseudo-deterministic products such as paintball plots or a member viewer are particularly useful in this scenario as they allow users to quickly isolate members that have produced an accurate analysis and interrogate them in a deterministic manner.

The distinctive characteristics of WoFS guidance, including increased forecast accuracy with storm age and the ability to adapt to varying predictability of evolving storm systems, are analogous to the human forecast process. For example, WoFS would be expected to provide a more accurate and confident forecast for a discrete supercell within a favorable environment than for developing multicellular convection within a marginal environment for supercell development (e.g., Lawson 2019). This variability often results in the quality of WoFS forecasts scaling with the perceived difficulty of a forecast by a human forecaster. It is therefore important to recognize that variation in forecast quality does not have a one-to-one correspondence with potential forecast value (Murphy 1993) when using WoFS (or any other) guidance. In other words, an accurate WoFS forecast of an ongoing discrete supercell in a favorable environment may be less useful than a low confidence forecast of a mesovortex developing within an MCS in a low-CAPE, high shear environment.

b. *Failure modes for accurate storm-scale prediction*

This paper is primarily focused on describing probabilistic, storm-scale guidance in WoFS so predictions of the near-storm environment have not been extensively discussed. However, because the near-storm environment modulates aspects of the mode, intensity, and propagation of convection (e.g., Weisman and Klemp 1982; Thompson et al. 2012; Markowski and Richardson 2014), knowledge of the accuracy of the predicted mesoscale environment in WoFS is vital. As will be shown in Part II, many situations where WoFS produces accurate analyses of convection but poor predictions of those storms are driven by errors in the near-storm environment. One dramatic example of environmental errors inducing storm-scale prediction errors in WoFS is driven by spatial errors in the position of airmass boundaries. For example, a small, eastward displacement in the prediction of a dryline can result in radar and satellite observations of developing convection being assimilated into dry air with 0 J Kg^{-1} of CAPE. Data assimilation may introduce convection into the WoFS initial conditions, but with no potential instability to support it, it immediately becomes outflow dominant and dissipates in the forecast. In these situations, what would normally be considered small errors in the environment (i.e., airmass boundary location errors of 10–20 km) can have large, negative consequences for predicting storms initiating along the misplaced boundary.

A second common error source in WoFS storm-scale forecasts results from inadequate suppression of spurious convection within the system[8]. WoFS suppresses spurious convection by assimilating "0" value observations of radar reflectivity and cloud water path where no storms are observed (Tong and Xue 2005); however, these observations are thinned compared to observations assimilated from within ongoing storms to save computational expense and can be insufficient to fully suppress erroneously predicted storms. Furthermore, in cases where WoFS predicts CI before it occurs, assimilation of "0" value observations may only suppress convection for a short time into the forecast. In these instances, early CI may be predicted in several successive WoFS forecasts, with the timing of CI delayed only by the frequency of WoFS forecast issuance (i.e., 30 min). In addition to introducing false alarm signals into the forecast, spurious storms in WoFS

[8] Specific examples of the errors discussed in this section are included in Part II.

will introduce errors into the mesoscale environment that can negatively impact predictions of ongoing storms. A common example of this is cold pool development from spurious convection upstream of ongoing convection, which can eliminate the potential instability available to the ongoing storms resulting in WoFS erroneously predicting dissipation.

Finally, errors in WoFS storm-scale predictions can stem from resolution limitations in the system. Although 3-km horizontal grid spacing is sufficient to partially resolve and accurately predict many supercells (Potvin and Flora 2015), many smaller-scale storms, or embedded features such as QLCS mesovortices, may be insufficiently resolved to predict accurately. For example, comparisons between WoFS forecasts run with 3-km horizontal grid spacing and those run with higher resolution have found similar skill in predicting thunderstorm reflectivity objects but improved skill at reduced grid spacing for predictions of UH objects representative of mesocyclone tracks (Lawson et al. 2021; Miller et al. 2022). This increased skill in predicting the smaller-scale mesocyclone objects, particularly in low-CAPE, high-shear environments (Lawson et al. 2021) with reduced grid spacing, suggests WoFS may fail to resolve some mesocyclones using 3-km horizontal grid spacing.

c. Recommendations for using WoFS guidance

In the discussion above we have outlined some principles for effectively using WoFS guidance, which lead to general recommendations that include:

- Independently assess the quality of WoFS analyses for each storm within the domain,
- Utilize current observations to identify storms that are accurately initialized by WoFS,
- Use expectations of the primary convective threats and forecast difficulty to guide expectations for WoFS performance and which guidance products may be most useful,
- Continue to check WoFS forecasts through the duration of an event as they are expected to improve with time, and
- Assess the quality of WoFS predictions of the near-storm environment.

Following these recommendations is a meteorologically intensive exercise that requires acquiring and combining knowledge of current observations, the mesoscale environment, and WoFS storm-scale predictions, followed by application of conceptual models for the anticipated mode, intensity, and evolution of convection. Performing this extensive analysis of each WoFS forecast conflicts with the priority to provide WoFS guidance in a manner that minimally impacts the cognitive workload of end users. One finding from real-time collaboration with NWS WFOs that mitigates this conflict is that WoFS guidance is best suited for meteorologists serving in a mesoanalyst role (Runk et al. 2020). As a mesoanalyst will be particularly knowledgeable of the storm scale environment, observations, and expected evolution prior to incorporation of WoFS guidance, it facilitates more seamless integration of WoFS into their workflow (Table 2). Additionally, it has been observed that users typically develop a “baseline” workflow for rapidly interrogating broad characteristics of WoFS guidance (Burke et al. 2022). For example, animations of the reflectivity paintball, 90th percentile of 2–5 km UH, and probability of 2–5 km UH exceeding 60 $m^2 s^{-2}$ in a 15-km neighborhood may be viewed in a few minutes or less and provide information on the expected evolution, potential severity, and likelihood of most severe storms within the forecast domain. In particular, *evaluating ensemble percentile and probability of exceedance products in tandem is an efficient way of spanning the majority of forecast scenarios predicted by the ensemble*. This baseline workflow may then be extended on a case-by-case basis depending on the expected convective hazards or mode (Wilson et al. 2021).

5. Summary, recommendations, and future work

In this paper we have provided a framework for interpreting thunderstorm hazard guidance from WoFS. Probabilistic forecasting concepts and guidance products have been reviewed with a focus on application to CAEs and differentiation between products designed for next-day (0–36 hr) and next-hour (0–6 h) lead times. The application of these concepts to the rationale and design of post-processing and visualization methods for WoFS guidance have also been described. Finally, general concepts and guidelines for effective use of WoFS guidance for severe weather prediction have been discussed and can be summarized as:

- The usefulness of WoFS guidance is dependent on rapid recognition of storms within the forecast domain that are accurately analyzed or not (Fig. 4). The quality of these analyses will likely determine where WoFS will be more likely to provide value over next-day CAEs and influence which guidance products are most appropriate to use.

- Knowledge of errors in WoFS predictions of the mesoscale environment, anticipation of the expected convective evolution and predictability, and application of conceptual models of the dynamics of severe convection are needed for effective use of WoFS guidance. These requirements make WoFS a natural fit with NWS mesoanalyst roles.

- Guidance from WoFS can be efficiently interrogated using different probabilistic products (i.e., probability of exceedance and ensemble percentile) in combination with deterministic/pseudo-deterministic products and current observations (Fig. 8).

This paper has provided general recommendations for effectively utilizing WoFS guidance. Specific examples of the concepts discussed from past real-time case studies are examined in Part II. WoFS is still an experimental system undergoing extensive active research on optimal system configurations. For example, the potential value of prototype WoFS systems with reduced horizontal grid spacing (on the order of 1 km) is being tested and evaluated against increased computational expense (Miller et al. 2022; Wang et al. 2022). Additionally, machine-learning post-processing models are being developed that can provide explicit hazard prediction and mitigate some of WoFS limitations (Flora et al. 2021; Clark and Loken 2022), and end-user applications for severe weather prediction are being refined through testbed experiments (Wilson et al. 2021, Burke et al. 2022, Gallo et al. 2022). Finally, the discussion herein has focused on prediction of convective storm hazards; however, WoFS is a numerical weather prediction system and can provide guidance for short-term prediction of any meteorological phenomenon, including fire (Jones et al. 2022; Lindley et al. 2023), winter weather, and aviation hazards (Avey et al. 2023). As with past refinement of WoFS guidance product design and application, future system development will be driven by collaboration with operational partners.

Acknowledgments. This work was supported by the NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement NA16OAR4320115 with additional funding provided by NOAA/Office of Oceanic and Atmospheric Research Joint Technology Transfer Initiative Grant NA18OAR4590361 and the Warn-on-Forecast Project. The authors are indebted to numerous National Weather Service Meteorologists and Hazardous Weather Testbed participants who have evaluated real-time WoFS guidance and provided feedback, much of which forms the basis of this article. In particular, multiyear collaborations with the NWS Southern Region, Weather Prediction Center, and Storm Prediction Center have been instrumental in guiding WoFS development and defining best practices for real-time interrogation of WoFS guidance. Additionally, we thank David Dowell, Gerry Creager, Brett Morrow, Karen Cooper, and Anthony Reinhart for contributions to real-time WoFS runs; Adam Clark and Israel Jirak for leading WoFS experiments within the Hazardous Weather Testbed Spring Forecasting Experiment; and Chauncy Schultz and two anonymous reviewers for providing thoughtful feedback that improved this article. Figures in this manuscript have been produced using the freely provided Anaconda Python distribution and the matplotlib, SciPy, and basemap libraries.

REFERENCES

Abras, C., and Coauthors, 2004: User-centered design. In Bainbridge, *W. Encyclopedia of Human-Computer Interaction*. Thousand Oaks: Sage Publications, **37**, 445–456.

AMS, 2008: Enhancing weather information with probability forecasts. Amer. Meteor Soc., CrossRef.

Avey, S., P. C. Burke, A. Cross, R. Hepper, H. Reeves, and P. S. Skinner, 2023: Exploring high-impact weather communication across time scales for route planning through the Aviation Weather Testbed. *Bull. Amer. Meteor. Soc.*, **104**, E810 – E814, CrossRef.

Blumberg, W. G., K. T. Halbert, T. A. Supinie, P. T. Marsh, R. L. Thompson, and J. A. Hart, 2017: SHARPpy: An open-source sounding analysis toolkit for the atmospheric sciences. *Bull. Amer. Meteor. Soc.*, **98**, 1625–1636, CrossRef.

Brewer, C. A., G. W. Hatchard, and M. A. Harrower, 2003: ColorBrewer in Print: A catalog of color Schemes for maps. *Cartogr. Geogr. Inf. Sci*., **30**, 5–32, CrossRef.

Brock, F. V., K. C. Crawford, R. L. Elliott, G. W. Cuperus, S. J. Stadler, H. L. Johnson, and M. D. Eilts, 1995: The Oklahoma Mesonet: A technical overview. *J. Atmos. Ocean. Technol*., **12**, 5–19, CrossRef.

Burke, P. C. and Coauthors, 2022: The virtual 2021 Warn-on-Forecast testbed experiment. *31st Conf. on Weather Analysis and Forecasting/27th Conf. on Numerical Weather Prediction*, Houston, TX, Amer. Meteor. Soc., J15B.1. [Available at https://ams.confex.com/ams/102ANNUAL/meetingapp.cgi/Paper/393718]

Calvo, L, I. Christel, M. Terrado, F. Cucchietti, and M. Perez-Montoro, 2022: Users' cognitive load: A key aspect to successfully communicate visual climate information. *Bull. Amer. Meteor. Soc*., **103**, E1 – E16, CrossRef.

Chipilski, H. G., X. Wang, and D. B. Parsons, 2018: Object-Based algorithm for the identification and tracking of convective outflow boundaries in numerical models. *Mon. Wea. Rev*., **146**, 4179–4200, CrossRef.

Choate, J. J., A. J. Clark, P. L. Heinselman, D. A. Imy, and P. S. Skinner, 2018: First demonstration of the NSSL Experimental Warn-on-Forecast System as part of the 2017 Spring Experiment. *98th American Meteorological Society Annual Meeting*, Austin, TX, Amer. Meteor. Soc. [Available online at ams.confex.com/ams/98Annual/meetingapp.cgi/Paper/335289]

Clark, A. J., 2017: Generation of ensemble mean precipitation forecasts from convection-allowing ensembles. *Wea. Forecasting*., **32**, 1569–1583, CrossRef.

____ and E. D. Loken, 2022: Machine learning-derived severe weather probabilities from a Warn-on-Forecast System. *Wea. Forecasting*., **37**, 1721 – 1740, CrossRef.

____, W. A. Gallus Jr., M. Xue, and F. Kong, 2009: A comparison of precipitation forecast skill between small convection-allowing and large convection-parameterizing ensembles. *Wea. Forecasting*., **24**, 1121–1140, CrossRef.

____, and Coauthors, 2011: Probabilistic precipitation forecast skill as a function of ensemble size and spatial scale in a convection-allowing ensemble. *Mon. Wea. Rev*., **139**, 1410–1418, CrossRef.

____, and Coauthors, 2012: An overview of the 2010 Hazardous Weather Testbed Experimental Forecast Program Spring Experiment. *Bull. Amer. Meteor. Soc*., **93**, 55–74, CrossRef.

____, and Coauthors, 2018: The Community Leveraged Unified Ensemble (CLUE) in the 2016 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment. *Bull. Am. Meteor. Soc*., **99**, 1433–1448, CrossRef.

____, and Coauthors, 2022: The second real-time, virtual Spring Forecasting Experiment to advance severe weather prediction. *Bull Amer. Meteor. Soc*., **103**, E1114 – E1116, CrossRef.

Dawson, L. C., G. S. Romine, R. J. Trapp, and M. E. Baldwin, 2017: Verifying supercellular rotation in a convection-permitting ensemble forecasting system with radar-derived rotation track data. *Wea. Forecasting*., **32**, 781–795, CrossRef.

Demuth, J. L., B. H. Morrow, and J. K. Lazo, 2009: Weather forecast uncertainty information. *Bull. Amer. Meteor. Soc*., **90**, 1614–1618, CrossRef.

____, R. E. Morss, I. Jankov, T. I. Alcott, C. R. Alexander, D. Nietfeld, T. L. Jensen, D. R. Novak, and S.G. Benjamin, 2020: Recommendations for developing useful and usable convection-allowing model ensemble information for NWS forecasters. *Wea. Forecasting*, **35**, 1381-1406, CrossRef.

Dowell, D. C., and coauthors, 2022. The High-Resolution Rapid Refresh (HRRR): An hourly updating convection-permitting forecast model. Part I: Motivation and system description. *Wea. Forecasting*, **37**, 1371–1395, CrossRef.

Ebert, E. E., 2001: Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon. Wea. Rev*., **129**, 2461–2480, CrossRef.

____, 2009: Neighborhood verification: A strategy for rewarding close forecasts. *Wea. Forecasting*., **24**, 1498–1510.

Epstein, E. S., 1969: Stochastic dynamic prediction. *Tellus*, **21**, 739–759, CrossRef.

Evans, C., D. F. Van Dyke, and T. Lericos, 2014: How do forecasters utilize output from a convection-permitting ensemble forecast system? Case study of a high-impact precipitation event. *Wea. Forecasting*., **29**, 466–486, CrossRef.

Flora, M. L., C. K. Potvin, and L. J. Wicker, 2018: Practical predictability of supercells: Exploring ensemble forecast sensitivity to initial condition spread. *Mon. Wea. Rev*., **146**, 2361–2379, CrossRef.

____, P. S. Skinner, C. K. Potvin, A. E. Reinhart, T. A. Jones, N. Yussouf, and K. H. Knopfmeier, 2019: Object-based verification of short-term, storm-scale probabilistic mesocyclone guidance from an experimental Warn-on-Forecast system. *Wea. Forecasting*, **34**, 1721 – 1739, CrossRef.

____, C. K. Potvin, P. S. Skinner, S. Handler, and A. McGovern, 2021: Using machine learning to generate storm-scale probabilistic guidance of severe weather hazards in the Warn-on-Forecast System. *Mon. Wea. Rev*., **149**, 1535–1557, CrossRef.

Fundel, V. J., N. Fleischhut, S. M. Herzog, M. Göber, and R. Hagedorn, 2019: Promoting the use of probabilistic weather forecasts through a dialogue between scientists, developers and end-users. *Quart. J. Roy. Meteor. Soc*., **145**, 210–231, CrossRef.

Galarneau, T. J. Jr., L. J. Wicker, K. H. Knopfmeier, W. J. S. Miller, P. S. Skinner, and K. A. Wilson, 2022: Short-term prediction of a nocturnal significant tornado outbreak using a convection-allowing ensemble. *Wea. Forecasting*, **37**, 1027–1047, CrossRef.

Gallo, B. T., A. J. Clark, and S. R. Dembek, 2016: Forecasting tornadoes using convection-permitting ensembles. *Wea. Forecasting*., **31**, 273–295, CrossRef.

____, 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.

____, A. J. Clark, B. T. Smith, R. L. Thompson, I. Jirak, and S. R. Dembek, 2018: Blended probabilistic tornado forecasts: Combining climatological frequencies with NSSL–WRF ensemble forecasts. *Wea. Forecasting*., **33**, 443–460, CrossRef.

____, A. J. Clark, B. T. Smith, R. L. Thompson, I. Jirak, and S. R. Dembek, 2019: Incorporating UH occurrence time to ensemble-derived tornado probabilities. *Wea. Forecasting*., **34**, 151–164, CrossRef.

____, and Coauthors, 2020: Scorecards for convection-allowing models. *Bull. Amer. Meteor. Soc*., **101**, 59–64, CrossRef.

____, and Coauthors, 2022: Exploring the watch-to-warning space: Experimental outlook performance during the 2019 Spring Forecasting Experiment in NOAA's Hazardous Weather Testbed. *Wea. Forecasting*., **37**, 617–637, CrossRef.

Gilleland, E., D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert, 2009: Intercomparison of spatial forecast verification methods. *Wea. Forecasting*, **24**, 1416–1430, CrossRef.

Guerra, J. E., P. S. Skinner, A. J. 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.

Hamill, T. M., M. J. Brennan, B. Brown, M. DeMaria, E. N. Rappaport, and Z. Toth, 2012: NOAA's future ensemble-based hurricane forecast products. *Bull. Amer. Meteor. Soc*., **93**, 209-220, CrossRef.

Hegarty, M., M. S. Canham, and S. I. Fabrikant, 2010: Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *J. Exp. Psychol. Learn. Mem. Cogn*., **36**, 37–53, CrossRef.

Hewson, T. D., and H. A. Titley, 2010: Objective identification, typing and tracking of the complete life-cycles of cyclonic features at high spatial resolution. *Meteorol. Appl*., **17**, 355–381, CrossRef.

Hirschberg, P. A., and Coauthors, 2011: A weather and climate enterprise strategic implementation plan for generating and communicating forecast uncertainty information. *Bull. Amer. Meteor. Soc*., **92**, 1651–1666, CrossRef.

Houtekamer, P. L., and F. Zhang, 2016: Review of the ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev*., **144**, 4489–4532, CrossRef.

James, J. J. J., C. Ling, C. D. Karstens, J. Correia Jr, K. Calhoun, T. Meyer, and D. LaDue, 2020: Forecasters' cognitive task analysis and mental workload analysis of issuing Probabilistic Hazard Information (PHI) during FACETs PHI prototype experiment. *Wea. Forecasting*., **35**, 1505–1521, CrossRef.

Jirak, I. L., S. J. Weiss, and C. J. Melick, 2012: The SPC storm-scale ensemble of opportunity: Overview and results from the 2012 Hazardous Weather Testbed Spring Forecasting Experiment. *26th Conf. on Severe Local Storms*, Nashville, TN, Amer. Meteor. Soc., p. 137.

Johnson, A., X. Wang, and S. Degalia, 2017: Design and implementation of a GSI-based convection-allowing ensemble-based data assimilation and forecast system for the PECAN field experiment. Part II: Overview and evaluation of a real-time system. *Wea. Forecasting*, **32**, 1227–1251, 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. and Forecasting*, **31**, 297-327, CrossRef.

____, X. Wang, P. Skinner, A. Johnson, and Y. Wang, 2018: Assimilation of GOES-13 imager clear-sky water vapor (6.5 μm) radiances into a Warn-on-Forecast system. *Mon. Wea. Rev*., **146**, 1077 – 1107, CrossRef.

____, P. Skinner, K. H. Knopfmeier, E. Mansell, P. Minnis, R. Palikonda, and W. Smith Jr., 2018: Comparison of cloud microphysics schemes in a Warn-on-Forecast system using synthetic satellite objects. *Wea. Forecasting*, **33**, 1681–1708, CrossRef.

____, ____, N. Yussouf, K. Knopfmeier, A. Reinhart, and D. Dowell, 2019: Forecasting high-impact weather in landfalling tropical cyclones using a Warn-on-Forecast system. *Bull. Amer. Meteor. Soc*., **100**, 1405–1417, CrossRef.

____, and Coauthors, 2020: Assimilation of GOES-16 radiances and retrievals into the Warn-on-Forecast system. *Mon. Wea. Rev*., **148**, 1829-1859, CrossRef.

____, R. Ahmadov, E. James, G. Pereira, S. Freitas, and G. Grell, 2022: Prototype of a Warn-on-Forecast System for smoke (WoFS-Smoke). *Wea. Forecasting*, **37**, 1191–1209, CrossRef.

Joslyn, S.L., K. Pak, D. Jones, J. Pyles, and E. Hunt, 2007: The effect of probabilistic information on threshold forecasts. *Wea. Forecasting*., **22**, 804–812, CrossRef.

____, and J. E. LeClerc, 2012: Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error. *J. Exp. Psychol. Appl*., **18**, 126–140, CrossRef.

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, CrossRef.

Karstens, C. D., and Coauthors, 2018: Development of a human–machine mix for forecasting severe convective events. *Wea. Forecasting*, **33**, 715–737, CrossRef.

Kaye, N. R., A. Hartley, and D. Hemming, 2012: Mapping the climate: guidance on appropriate techniques to map climate variables and their uncertainty. *Geosci. Model Dev*., **5**, 245-256, CrossRef.

Klockow-McClain, K. E., R. A. McPherson, and R. P. Thomas, 2020: Cartographic design for improved decision making: Trade-offs in uncertainty visualization for tornado threats. *Annals of the American Association of Geographers*, **110**, 314–333, CrossRef.

Koch, S., B. Ferrier, M. Stoelinga, E. Szoke, S. Weiss, J. Kain 2005. The use of simulated radar reflectivity fields in the diagnosis of mesoscale phenomena from high-resolution WRF model forecasts. *11th Conf. on Mesoscale Processes*, Albuquerque, NM, Amer. Meteor. Soc., J4J.7. [Available online at ams.confex.com/ams/32Rad11Meso/webprogram/Paper97032.html]

Lagerquist, R., A. McGovern, and D. J. Gagne II, 2019: Deep learning for spatially explicit prediction of synoptic-scale fronts. *Wea. Forecasting*., **34**, 1137–1160, CrossRef.

Lawson, J. R., 2019: Predictability of idealized thunderstorms in buoyancy-shear space. *J. Atmos. Sci*., **76**, 2653–2672, CrossRef.

____, 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.

_____, C. K. Potvin, P. S. Skinner, and A. E. Reinhart, 2021: The vice and virtue of increased horizontal resolution in ensemble forecasts of tornadic thunderstorms in low-CAPE, high-shear environments. *Mon. Wea. Rev*., **149**, 921–944, CrossRef.

Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev*., **102**, 409–418, CrossRef.

Lilly, D. K., 1990: Numerical prediction of thunderstorms – has its time come? *Quart. J. Roy. Meteor. Soc*., **116**, 779-798, CrossRef.

Lindley, T. T., A. B. Zwink, R. R. Barnes, G. P. Murdoch, B. C. Ancell, P. C. Burke, and P. S. Skinner, 2023: Preliminary use of convection-allowing models in fire weather. *J. Oper. Meteor*., **11**, 72-81, CrossRef.

Ling, C., L. Hua, C. D. Karstens, G. J. Stumpf, T. M. Smith, K. M. Kuhlman, and L. Rothfusz, 2015: A Comparison between WarnGen System and Probabilistic Hazard Information System for severe weather forecasting. *Proc. Hum. Fact. Ergon. Soc. Annu. Meet*., **59**, 1791–1795, CrossRef.

Loken, E.D., A.J. Clark, M. Xue, and F. Kong, 2017: Comparison of next-day probabilistic severe weather forecasts from coarse- and fine-resolution CAMs and a convection-allowing ensemble. *Wea. Forecasting*, **32**, 1403-1421, CrossRef.

Loken, E. D., A. J. Clark, M. Xue, and F. Kong, 2019: Spread and skill in mixed- and single-physics convection-allowing ensembles, *Wea. Forecasting*, **34**, 305–330, CrossRef.

Mansell, E. R., C. L. Ziegler, and E. C. Bruning, 2010: Simulated electrification of a small thunderstorm with two-moment bulk microphysics. *J. Atmos. Sci*., **67**, 171–194, CrossRef.

Markowski, P. M., and Y. P. Richardson, 2014: The influence of environmental low-level shear and cold pools on tornadogenesis: Insights from idealized simulations. *J. Atmos. Sci*., **71**, 243–275, CrossRef.

May, R. M., and Coauthors, 2022: MetPy: A meteorological Python library for data analysis and visualization. *Bull. Amer. Meteor. Soc*., **103**, E2273 – E2284, CrossRef.

McInerny, G. J., and Coauthors, 2014: Information visualisation for science and policy: engaging users and avoiding bias. *Trends Ecol. Evol*., **29**, 148–157, CrossRef.

Miller, W. J. S., and Coauthors, 2022: Exploring the usefulness of downscaling free forecasts from the Warn-on-Forecast System. *Wea. Forecasting*, **37**, 181–203, CrossRef.

Mittermaier, M., and N. Roberts, 2010: Intercomparison of spatial forecast methods: Identifying skillful spatial scales using the fractions skill score. *Wea. Forecasting*, **25**, 343–354, CrossRef.

Morss, R. E., J. L. Demuth, and J. K. Lazo, 2008: Communicating uncertainty in weather forecasts: A survey of the U.S. public. *Wea. Forecasting*., **23**, 974–991, CrossRef.

Morrison, H., and J. A. Milbrandt, 2015: Parameterization of cloud microphysics based on the prediction of bulk ice particle properties. Part I: Scheme description and idealized tests. *J. Atmos. Sci*., **72**, 287–311, CrossRef.

Murphy, A. H., 1977: The value of climatological, categorical and probabilistic forecasts in the cost-loss ratio situation. *Mon. Wea. Rev*., **105**, 803–816, CrossRef.

____, 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281–293, CrossRef.

____, 1998: The early history of probability forecasts: Some extensions and varifications. *Wea. Forecasting*., **13**, 5–15, CrossRef.

National Academy of Science, 2018: *Integrating Social and Behavioral Sciences within the Weather Enterprise*. National Academies Press.

National Research Council, 2006: *Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts.* National Academies Press.

Novak, D. R., D. R. Bright, and M. J. Brennan, 2008: Operational forecaster uncertainty needs and future roles. *Wea. Forecasting*., **23**, 1069–1084, CrossRef.

____, K. A. Brill, and W. A. Hogsett, 2014: Using percentiles to communicate snowfall uncertainty. *Wea. Forecasting*, **29**, 1259–1265, CrossRef.

Oakley, N. S., and B. Daudert, 2016: Establishing best practices to improve usefulness and usability of web interfaces providing atmospheric data. *Bull. Amer. Meteor. Soc*., **97**, 263–274, CrossRef.

Obermaier, H., and K. I. Joy, 2014: Future challenges for ensemble visualization. *IEEE Comput. Graph. Appl*., **34**, 8–11, CrossRef.

Potvin, C. K., and M. L. Flora, 2015: Sensitivity of idealized supercell simulations to horizontal grid spacing: Implications for Warn-on-Forecast. *Mon. Wea. Rev*., **143**, 2998–3024, CrossRef.

____, E. M. Murillo, M. L. Flora, and D. M. Wheatley, 2017: Sensitivity of supercell simulations to initial-condition resolution. *J. Atmos. Sci*., **74**, 5–26, CrossRef.

____, and Coauthors, 2019: Systematic comparison of convection-allowing models during the 2017 NOAA HWT Spring Forecasting Experiment. *Wea. Forecasting*, **34**, 1395–1416, CrossRef.

____, and Coauthors, 2020: Assessing systematic impacts of PBL schemes on storm evolution in the NOAA Warn-on-Forecast System. *Mon. Wea. Rev*., **148**, 2567–2590, CrossRef.

Rautenhaus, M., M. Böttinger, S. Siemen, R. Hoffman, R. Kirby, M. Mirzargar, N. Röber, and R. Westermann, 2018: Visualization in meteorology — A survey of techniques and tools for data analysis tasks. *IEEE Trans. Vis. Comput. Graph*., **24**, 3268–3296, CrossRef.

Roberts, B., I. L. Jirak, A. J. Clark, S. J. Weiss, and J. S. Kain, 2019: Post-processing and visualization techniques for convection-allowing ensembles. *Bull. Amer. Meteor. Soc*., **100**, 1245–1258, CrossRef.

____, 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, CrossRef.

Roebber, P. J., D. M. Schultz, B. A. Colle, and D. J. Stensrud, 2004: Toward improved prediction: High-resolution and ensemble modeling systems in operations. *Wea. Forecasting*, **19**, 936–949, CrossRef.

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

Runk, K. J., M. Foster, C. M. Gravelle, J. M. Laflin, A. E. Cohen, R. L. Thompson, and K. L. Crandall, 2020: Evaluating the role of the mesoanalyst in severe weather impacts-based decision support services. Part III – Messaging focus. *8th Symposium on Building a Weather-Ready Nation*, Amer. Meteor. Soc., Boston, MA. 6.3. [Available online at ams.confex.com/ams/2020Annual/meetingapp.cgi/Paper/366884]

Savelli, S., and S. Joslyn, 2013: The advantages of predictive interval forecasts for non-expert users and the impact of visualizations. *Appl. Cogn. Psychol*., **27**, 527–541, CrossRef.

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, CrossRef.

____, and Coauthors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*., **25**, 263–280, CrossRef.

____, G. S. Romine, K. R. Smith, and M. L. Weisman, 2014: Characterizing and optimizing precipitation forecasts from a convection-permitting ensemble initialized by a mesoscale ensemble Kalman Filter. *Wea. Forecasting*., **29**, 1295–1318, CrossRef.

____, ____, R. A. Sobash, K. R. Fossell, and M. L. Weisman, 2015: NCAR's experimental real-time convection-allowing ensemble prediction system. *Wea. Forecasting*, **30**, 1645–1654, CrossRef.

____, ____, K. R. Fossell, R. A. Sobash, and M. L. Weisman, 2017: Toward 1-km ensemble forecasts over large domains. *Mon. Wea. Rev*., **145**, 2943–2969, CrossRef.

____, ____, 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, CrossRef.

Sivillo, J. K., J. E. Ahlquist, and Z. Toth, 1997: An ensemble forecasting primer. *Wea. Forecasting*, **12**, 809–818, CrossRef.

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

Smith, T. M., and et al., 2016: Multi-Radar Multi-Sensor (MRMS) severe weather and aviation products: Initial operating capabilities. *Bull. Amer. Meteor. Soc*., **97**, 1617–1630, CrossRef.

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, CrossRef.

____, C. S. Schwartz, G. S. Romine, K. R. Fossell, and M. L. Weisman, 2016: Severe weather prediction using storm surrogates from an ensemble forecasting system. *Wea. Forecasting*, **31**, 255–271, CrossRef.

____, B. Roberts, and P. S. Skinner, 2020: Novel web-based tools for the visualization of high-impact weather forecasts with convection-allowing ensembles. *36th Conference on Environmental Information Processing Technologies, Amer. Meteor. Soc.*, Boston, MA 7B.2. [Available online at ams.confex.com/ams/2020Annual/meetingapp.cgi/Paper/369784]

Stensrud, D. J., and Coauthors, 2009: Convective-scale Warn-on-Forecast System: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1500, CrossRef.

____, and Coauthors, 2013: Progress and challenges with Warn-on-Forecast. *Atmos. Res.*, **123**, 2–16, CrossRef.

Surcel, M., I. Zawadzki, and M. K. Yau, 2014: On the filtering properties of ensemble averaging for storm-scale precipitation forecasts. *Mon. Wea. Rev.*, **142**, 1093–1105, CrossRef.

Thompson, J. C., 1952: On the operational deficiencies in categorical weather forecasts. *Bull. Amer. Meteor. Soc.*, **33**, 223–226, CrossRef.

Thompson, R. L., B. T. Smith, J. S. Grams, A. R. Dean, and C. Broyles, 2012: Convective modes for significant severe thunderstorms in the contiguous United States. Part II: Supercell and QLCS tornado environments. *Wea. Forecasting*., **27**, 1136–1154, CrossRef.

Tong, M. and M. Xue, 2005: Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSS experiments. *Mon. Wea. Rev.*, **133**, 1789–1807, CrossRef.

Wang, Y., N. Yussouf, C. A. Kerr, D. R. Stratman, and B. C. Matilla, 2022: An experimental 1-km Warn-on-Forecast System for hazardous weather events. *Mon. Wea. Rev*. **150**, 3081 – 3102, CrossRef.

Weisman, M. L., and J. B. Klemp, 1982: The dependence of numerically simulated convective storms on vertical wind shear and buoyancy. *Mon. Wea. Rev.*, **110**, 504–520, CrossRef.

Weyn, J. A., and D. R. Durran, 2017: The dependence of the predictability of mesoscale convective systems on the horizontal scale and amplitude of initial errors in idealized simulations. *J. Atmos. Sci.*, **74**, 2191–2210, CrossRef.

Weniger, M., and P. Friederichs, 2016: Using the SAL technique for spatial verification of cloud processes: A sensitivity analysis. *J. Applied Meteor. Clim.*, **55**, 2091–2108, CrossRef.

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, CrossRef.

Wilson, K. A., P. L. Heinselman, C. M. Kuster, D. M. Kingfield, and Z. Kang, 2017: Forecaster performance and workload: Does radar update time matter? *Wea. Forecasting*, **32**, 253–274, CrossRef.

____, and Coauthors, 2019a: Exploring applications of storm-scale probabilistic warn-on-forecast guidance in weather forecasting. *International Conference on Human–Computer Interaction (HCII 2019): Virtual, Augmented and Mixed-Reality, Applications and Case Studies*, J. Chen and G. Fragomeni, Eds., 577–572, CrossRef.

_____, 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, CrossRef.

_____, B. T. Gallo, P. S. Skinner, A. Clark, P. Heinselman, and J. J. Choate, 2021: Analysis of end user access of Warn-on-Forecast guidance products during an experimental forecasting task. *Wea. Climate Soc.*, **13**, 859–874, CrossRef.

Wolff, J. K., M. Harrold, T. Fowler, J. H. Gotway, L. Nance, and B. G. Brown, 2014: Beyond the basics: Evaluating model-based precipitation forecasts using traditional, spatial, and object-based methods. *Wea. Forecasting*., **29**, 1451–1472, CrossRef.

Xue, M., and Coauthors, 2007: CAPS realtime storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2007 spring experiment. *22nd Conf. Weather Analysis and Forecasting/18th Conf. Numerical Weather Prediction*. Park City, UT, Amer. Meteor. Soc. 3B.1. [Available online at ams.confex.com/ams/22WAF18NWP/techprogram/paper_124587.htm]

Young, M. S., K. A. Brookhuis, C. D. Wickens, and P. A. Hancock, 2015: State of science: mental workload in ergonomics. *Ergonomics*, **58**, 1–17, CrossRef.

Yussouf, N., D. C. Dowell, L. J. Wicker, K. Knopfmeier, and D. M. Wheatley, 2015: Storm-scale data assimilation and ensemble forecasts for the 27 April 2011 severe weather outbreak in Alabama. *Mon. Wea. Rev*. **143**, 3044-3066, 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.

____, and K. H. Knopfmeier, 2019: Application of the Warn-on-Forecast system for flash-flood-producing heavy convective rainfall events. *Q. J. R. Meteorol. Soc.*, **145**, 2385-2403, CrossRef.

Zhang, Y., F. Zhang, D. J. Stensrud, and Z. Meng, 2016: Intrinsic predictability of the 20 May 2013 tornadic thunderstorm event in Oklahoma at storm scales. *Mon. Wea. Rev.*, **144**, 1273–1298, CrossRef.