Warn-on-Forecast System: From Vision to Reality

PAMELA L. HEINSELMAN,[a,c] PATRICK C. BURKE,[a] LOUIS J. WICKER,[a,c] ADAM J. CLARK,[a,c] JOHN S. KAIN,[a,j] JIDONG GAO,[a,c] NUSRAT YUSSOUF,[a,b,c] THOMAS A. JONES,[a,b,c] PATRICK S. SKINNER,[a,b,c] COREY K. POTVIN,[a,c] KATIE A. WILSON,[a,b,g] BURKELY T. GALLO,[b,d,h] MONTGOMERY L. FLORA,[a,b,c] JOSHUA MARTIN,[a,b] GERRY CREAGER,[a,b] KENT H. KNOPFMEIER,[a,b] YUNHENG WANG,[a,b] BRIAN C. MATILLA,[a,b] DAVID C. DOWELL,[e] EDWARD R. MANSELL,[a] BRETT ROBERTS,[a,b,i] KIMBERLY A. HOOGEWIND,[a,b] DEREK R. STRATMAN,[a,b] JORGE GUERRA,[a,b,k] ANTHONY E. REINHART,[a] CHRISTOPHER A. KERR,[a,b] AND WILLIAM MILLER[a,b,f]

[a] *NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma*
[b] *Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma, Norman, Oklahoma*
[c] *School of Meteorology, University of Oklahoma, Norman, Oklahoma*
[d] *NOAA/NWS/NCEP Storm Prediction Center, Norman, Oklahoma*
[e] *NOAA/Earth System Research Laboratories/Global Systems Laboratory, Boulder, Colorado*
[f] *Earth System Science Interdisciplinary Center, University of Maryland, College Park, College Park, Maryland*
[g] *Rand Corporation, Santa Monica, California*
[h] *16th Weather Squadron, Offutt Air Force Base, Bellevue, Nebraska*
[i] *CoreLogic, Irvine, California*
[j] *IBSS Corp., Silver Spring, Maryland*
[k] *Project Canary, Denver, Colorado*

(Manuscript received 18 August 2023, in final form 26 October 2023, accepted 30 October 2023)

ABSTRACT: In 2009, advancements in NWP and computing power inspired a vision to advance hazardous weather warnings from a *warn-on-detection* to a *warn-on-forecast* paradigm. This vision would require not only the prediction of individual thunderstorms and their attributes but the likelihood of their occurrence in time and space. During the last decade, the warn-on-forecast research team at the NOAA National Severe Storms Laboratory met this challenge through the research and development of 1) an ensemble of high-resolution convection-allowing models; 2) ensemble- and variational-based assimilation of weather radar, satellite, and conventional observations; and 3) unique postprocessing and verification techniques, culminating in the experimental Warn-on-Forecast System (WoFS). Since 2017, we have directly engaged users in the testing, evaluation, and visualization of this system to ensure that WoFS guidance is usable and useful to operational forecasters at NOAA national centers and local offices responsible for forecasting severe weather, tornadoes, and flash floods across the watch-to-warning continuum. Although an experimental WoFS is now a reality, we close by discussing many of the exciting opportunities remaining, including folding this system into the Unified Forecast System, transitioning WoFS into NWS operations, and pursuing next-decade science goals for further advancing storm-scale prediction.

SIGNIFICANCE STATEMENT: The purpose of this research is to develop an experimental prediction system that forecasts the probability for severe weather hazards associated with individual thunderstorms up to 6 h in advance. This capability is important because some people and organizations, like those living in mobile homes, caring for patients in hospitals, or managing large outdoor events, require extended lead time to protect themselves and others from potential severe weather hazards. Our results demonstrate a prediction system that enables forecasters, for the first time, to message probabilistic hazard information associated with individual severe storms between the watch-to-warning time frame within the United States.

KEYWORDS: Forecast verification/skill; Numerical weather prediction/forecasting; Probability forecasts/models/distribution; Short-range prediction; Data assimilation; Decision support

1. Introduction

The development of NWP-based probabilistic forecast guidance enabling forecasters to communicate actionable high-impact weather information to a broad range of end-users is an ongoing scientific and societal challenge. A frequently updating ensemble analysis and prediction system is crucial for formulating and communicating this information, which includes the most likely path, intensity, and timing for hazards like tornadoes, high winds, large hail, and flash floods, at lead times from minutes to a few hours. Stensrud et al. (2009) envisioned the development of such a forecast system that, by 2020, would enable a transition from the current "warn-on-detection" paradigm to a "warn-on-forecast" paradigm within NWS operations. The motivation and societal need for the warn-on-forecast concept can also be found in several National Research Council (NRC) reports

Denotes content that is immediately available upon publication as open access.

Gerry Creager: Retired.

Corresponding author: Pamela Heinselman, pam.heinselman@noaa.gov

DOI: 10.1175/WAF-D-23-0147.1

(e.g., NRC 1999, 2003, 2006, 2012), and it fits within a newly developing framework termed FACETs (Forecasting a Continuum of Environmental Threats; Rothfusz et al. 2018), which aims to transition the delivery of severe weather information in the United States from rigid, deterministic products to a more continuously updating flow of probabilistic hazard information (Trujillo-Falcón et al. 2022). Both warn-on-forecast and FACETs align with an ongoing shift in the NWS toward less determinism, more communication of probabilities, and a focus on the "last mile" of warning communication to get specific forecasts to decision makers at greater lead time (Uccellini and Ten Hoeve 2019).

Providing probabilistic information at watch-to-warning time scales (i.e., 0–6 h) also has the potential for significant economic benefits. For example, analysis of a survey conducted by Howard et al. (2021) of about 500 firms in Dallas–Fort Worth, Texas, estimated annual cost avoidance of $2.3–$7.6 billion (U.S. dollars) from using probabilistic information instead of deterministic warnings for weather events affecting this area. These benefits are not surprising, as single deterministic NWP forecasts have little chance of consistent, accurate prediction of thunderstorms that evolve rapidly and are highly sensitive to environmental and internal storm processes (e.g., Stensrud and Gao 2010). Hence, both the social science and physical science point to the necessity of probabilistic prediction of storm-scale thunderstorm hazards.

While these ideas are common in 2023, in 2009 NWP capabilities were very limited for models that could depict individual thunderstorms and their hazards, and operational forecasters mainly relied on relatively coarse models that only depicted large-scale storm environments. Stensrud et al. (2009) outlined several scientific, technical, and sociological challenges to address this forecast challenge. For example, high-resolution initial conditions (ICs), from which accurate predictions of individual storms can be made, are foundational for warn-on-forecast. Generating these ICs was a scientific and technical challenge that required research on data assimilation (DA) methods, as well as the impacts of different observational datasets (e.g., radar, satellite, surface observations, etc.) on forecast quality. Another challenge involved how to most accurately model storms. Stensrud et al. (2009) recognized that multimoment schemes would be needed to reduce model error and sensitivity related to microphysics, and that developing efficient and accurate multimoment schemes that can be used in real time would require research and development. Finally, a key requirement for operational systems is that forecasts are produced quickly enough to be useful. Thus, efficiency is an integral component of all the technical and scientific challenges. Other technical hurdles included developing verification strategies and finding what observational datasets are needed to verify predictions of individual storms. Sociological challenges for warn-on-forecast included optimizing visualizations of probabilistic information that would best assist forecasters without overwhelming them, as well as learning about how different end-users apply and interpret warn-on-forecast guidance in different situations.

Since 2009, the NOAA NSSL and their partners have worked together to transform Stensrud et al. (2009)'s vision from basic research to reality. Namely, we created and maintain a *frequently updating, high-resolution, storm-scale*[1] *ensemble analysis and prediction system* to extend warning lead time for severe weather and flash floods, which is named the Warn-on-Forecast System (WoFS). Building WoFS required innovation in several areas and, as of 2023, steady successes have laid the groundwork for operational implementation in the next few years. In this paper, we describe the scientific and technical approaches employed within the 2023 WoFS, unique postprocessing, visualization, and verification approaches, and provide a summary of the user-focused research conducted iteratively with forecasters, and other end users, to address the technical and sociological challenges for system use.

2. WoFS ensemble data assimilation and prediction system

a. *WoFS evolution and development*

When the warn-on-forecast program started at NSSL in 2009, the ARW-WRF (hereafter WRF; Skamarock et al. 2008) was chosen as the driving model for WoFS. At the time, WRF was the state-of-the-art regional modeling system in the United States. WRF was an ideal choice for WoFS because: 1) the nonhydrostatic solver facilitates simulating convective storms, 2) WRF was designed for community modeling with many different options for physical parameterizations, 3) through National Science Foundation funding, the NCAR provided comprehensive model support, 4) adaptable regional domains were easy to configure, and 5) many NSSL scientists were already using WRF for convective storms research. Furthermore, WRF was used for the HRRR (Benjamin et al. 2016; Dowell et al. 2022), which was in development at NOAA's Global Systems Laboratory (GSL) and became operational in 2014. Matching the HRRR model core enabled close collaboration with GSL, a core partner of NSSL, and we envisioned that WoFS initial and lateral boundaries would eventually come from the HRRR. With NOAA's transition to the Unified Forecast System (UFS; https://ufscommunity.org) and NCAR ending their support of the WRF Model, future versions of WoFS will use a different dynamic core.

Initial versions of the WoFS DA system used NCAR's Data Assimilation Research Testbed software (DART; Anderson 2008). Ensemble DA approach is essential for effective assimilation of radar observations owing to its ability to include both flow dependence and multivariate cross covariances in the background. DART was the first and only available community ensemble DA framework that was designed to assimilate a wide variety of observations, including the capability to directly assimilate Doppler reflectivity and velocity

[1] "Storm scale" will generally refer to the scale of individual thunderstorms, consistent with the meso-gamma scale defined by Orlanski (1975).

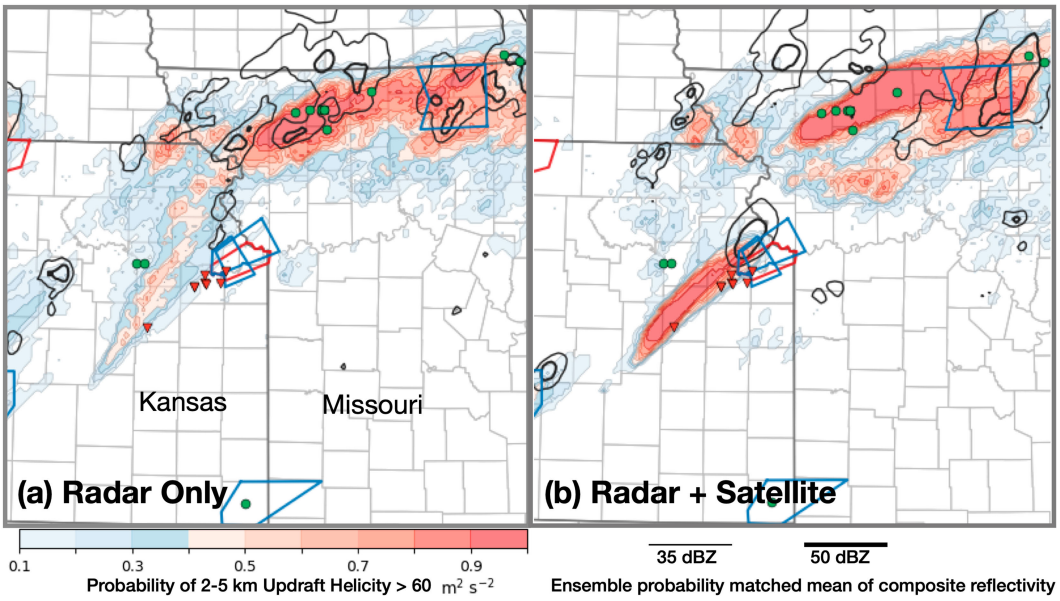

FIG. 1. Probability of 2–5-km AGL UH ≥ 60 m^2 s^{-2} valid at 2100–0000 UTC 28–29 May 2019 (shaded), and probability matched mean (PMM) simulated composite reflectivity (contours) valid at 0000 UTC 29 May from the 2100 UTC WoFS initializations using (a) radar DA only and (b) radar and satellite DA. Local storm reports during this period are indicated by red inverted triangles for tornadoes and green circles for hail. Tornado warnings (red polygons) and severe thunderstorm warnings (blue polygons) that included 0000 UTC 29 May are also shown.

data within high-resolution models via ensemble Kalman filter (EnKF) algorithms. This capability likely accelerated WoFS development by several years. EnKF is well suited for WoFS because its ensemble approach results directly in perturbed initial conditions, and research beginning in the mid- to late-2000s found that it was very effective for assimilating radar reflectivity and radial velocity data (e.g., Zhang et al. 2004; Tong and Xue 2005; Dowell et al. 2004; Gao and Xue 2008; Aksoy et al. 2009, 2010; Potvin and Wicker 2013). In 2016, NSSL began assimilating radar reflectivity from its Multi-Radar, Multi-Sensor system (MRMS; Smith et al. 2016), which was important due to its quality control and low latency. WoFS also assimilates radial velocity information, after thinning, from the raw Level-II radial velocity files (Dowell et al. 2011; Yussouf et al. 2013, 2015, 2016; Jones et al. 2013a, 2014, 2015, 2016; Wheatley et al. 2015). Controlled experiments, with and without radar DA, show that frequent radar DA (5–15-min cycles) is critical for short-term storm prediction (e.g., Jones et al. 2013a, 2014, 2015; Yussouf et al. 2016). Furthermore, Guerra et al. (2022) used a storm tracking algorithm to show that when observed storms depicted by WoFS have been through just 1 h of DA cycles, their probabilities of detection are on average about 80 percentage points higher than observed storms that initiated after model initialization. This improved performance is due, in large part, to the radar DA (e.g., Galarneau et al. 2022).

Additional major advancements in WoFS were made by assimilating satellite data, like cloud water-path and water vapor radiances (Jones et al. 2018a). When assimilated with radar data, satellite data provide additional forecast improvements for cloud properties, convective initiation, and the near-storm environment (Jones et al. 2013a,b, 2014, 2015, 2016). The launch of *GOES-16* in 2017 and *GOES-17* in 2018, provided satellite data with higher spatiotemporal resolution than previously available. The assimilation of these higher-resolution satellite observations results in further forecast improvements (Jones et al. 2020). For example, Fig. 1 shows 3-h probability swaths of updraft helicity (UH) > 60 m^2 s^{-2} (a proxy for storm rotation; Kain et al. 2010) associated with a long-track, violent tornado that approached Kansas City suburbs on 28 May 2019. The WoFS run using both radar and satellite DA (Fig. 1b) depicts much higher probabilities and better matches the observed storm track compared to simulations without satellite DA (Fig. 1a). Newer, derived satellite products, such as atmospheric motion vectors (AMVs), layered precipitable water, and lightning flash counts from the global lightning mapper, have also been tested in WoFS with promising results (e.g., Mallick and Jones 2020; Pan et al. 2018, 2021; Wang et al. 2021; Zhao et al. 2021a,b; Hu et al. 2021a,b). In the future, we expect to assimilate additional satellite products, such as fire radiative power and aerosol optical depth, as applications of WoFS expand (Jones et al. 2022). In 2019, when the EnKF capabilities available in DART were added to NOAA's operational the Community Gridpoint Statistical Interpolation (GSI; Wu et al. 2002; Hu et al. 2018)-based EnKF (Whitaker et al. 2008; Zhu et al. 2013) DA system, the WoFS DA system was updated from DART to GSI.

Another key research area early in WoFS development was microphysics parameterizations. NSSL was well positioned for testing and development in this area, as scientists therein had developed the multimoment variable density NSSL

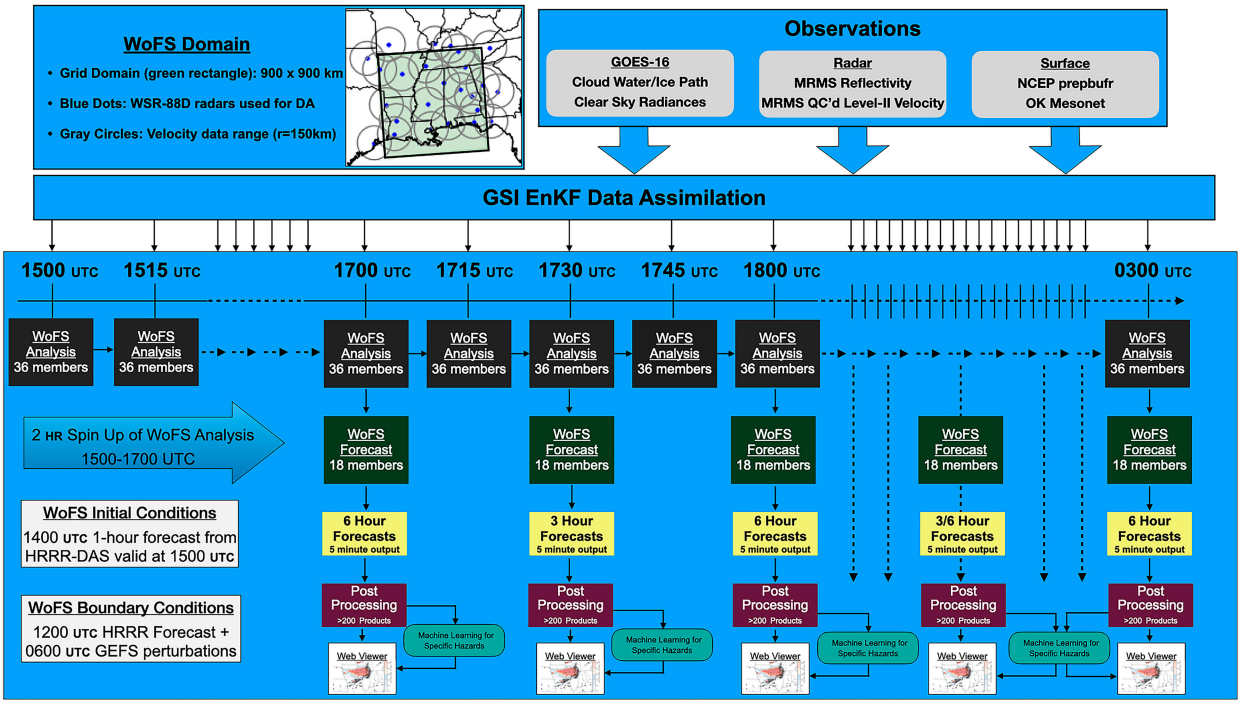

FIG. 2. Flowchart showing the workflow for WoFS. The WoFS analysis system is continuously cycled every 15 min up to 12–15 h (depending on the weather). The initial analysis, obtained using NCEP's HRRR-DAS, starts at 1500 UTC and spins up using 2 h of cycling. Ensemble forecasts are then launched every 30 min. (top left) An example domain from the south-central United States. (top right) The observations used in the GSI EnKF analysis cycling.

microphysics scheme (Mansell et al. 2010; Ziegler 1985; Mansell et al. 2020). Many studies in the 2010s showed the advantages of multimoment over single-moment schemes in depicting a range of convective modes (e.g., Dawson et al. 2010, 2012; Yussouf et al. 2013; Putnam et al. 2014, 2017a,b; Igel et al. 2015). The upgrade of the WSR-88D radar network to dual polarization in 2013 meant that additional information on particle size distributions, like differential reflectivity (Z_{DR}) and specific differential phase (K_{DP}), could be incorporated in DA to reduce errors. Directly assimilating dual polarization variables has limited utility without multimoment schemes (e.g., Jung et al. 2012). The direct assimilation of real Z_{DR} data is demonstrated by Putnam et al. (2019) and Putnam et al. (2021). Studies examining the importance of multimoment schemes in reproducing observed dual polarization signatures include Dawson et al. (2014), Putnam et al. (2014, 2017a,b; Putnam et al. 2019; Putnam et al. 2021), and Johnson et al. (2019). Also, Labriola et al. (2019a,b) examined multimoment microphysics schemes for explicit hail prediction. Because of its utility in accurately depicting a range of convective modes at high resolution, flexibility with up to three-moment capability, and community availability within the WRF Model and NOAA's Common Community Physics Package (Zhang et al. 2022), the current WoFS version uses NSSL microphysics.

b. 2023 WoFS configuration

The 2023 WoFS configuration is a 3-km grid spacing, 36-member ensemble with a regional domain (900 km × 900 km) centered around the area of expected high-impact weather for a particular day. Figure 2 depicts WoFS workflow. WoFS uses continuous 15-min DA intervals, and assimilates data from WSR-88D radars, GOES, and conventional observations (see Table 1). High-resolution radar and satellite-derived observations are resampled to a 5-km grid prior to assimilation. While the DA and cycling use all 36 members, forecasts are generated from the first 18 members and extend to 6- and 3-h lead times for top-of-the-hour and bottom-of-the-hour initializations, respectively. WoFS also has an experimental companion system known as WoF Hybrid 3DEnVAR (Gao and Stensrud 2014; Wang et al. 2019) that uses the hybrid ensemble–three dimensional (3D) variational data assimilation technique. This system incorporates the WoFS EnKF ensemble for flow-dependent background error covariances. It runs in a one-way coupling between the ensemble and the deterministic system to generate a single deterministic analysis and forecast component (not shown). Similar to WoFS, WoF Hybrid 3DEnVAR assimilates WSR-88D radar (reflectivity and velocity) and *GOES-16* and *GOES-17* satellite data (Table 1), and completes a DA cycle every 15 min (Wang et al. 2019; Pan et al. 2021). This approach is being looked into to develop a full 3D atmospheric severe thunderstorm analysis system for severe weather detection and NWP model verification.

Initial conditions (ICs) for WoFS are obtained from the 36-member HRRR-DAS (HRRR Data Assimilation System; Dowell et al. 2022). Until 2021, lateral boundary conditions (LBCs) for WoFS were generated from the first 9 members,

TABLE 1. Configuration comparison between the HRRR and WoFS ensemble.

Configuration	HRRR (state-of-the-art single run)	WoFS (first-of-its-kind ensemble)
WRF-based; RUC LSM; 36-member GSI-ENKF analysis; 3-km horizontal grid spacing; 50 vertical levels	Yes	Yes
Designed as forecast ensemble	No	Yes
Radar reflectivity assimilated	Yes	Yes
Radar velocity assimilated	No	Yes
GOES clear-sky radiances assimilated	No	Yes
GOES CWP assimilated	No	Yes
Multiple PBL schemes (YSU, MYJ, MYNN)	No	Yes
Multiple radiation schemes (RRTMG, RRTM, Dudhia)	No	Yes
Microphysics	Thompson aerosol aware	NSSL double moment
Data assimilation cadence	60 min	15 min
Temporal resolution of output	15 min	5 min

0–24-h experimental HRRR ensemble forecasts running on NOAA's Jet supercomputer at GSL (Dowell et al. 2016). Those forecasts included stochastic approaches applied to the HRRR physics suite (Jankov et al. 2019). This configuration yielded a unique combination of IC/LBCs, boundary layer parameterization, and radiation scheme for each WoFS member, and helped to maintain ensemble spread through the cycling period (Wheatley et al. 2015; Stensrud et al. 2000). In 2022, the experimental HRRR ensemble forecast was no longer available, so a new strategy for generating the 36 LBCs was needed. Instead, the operational HRRR forecast is combined with large-scale perturbations from 18 members of the GEFS (Zhou et al. 2022). Testing showed that this new method worked as well as or better than the old method and eliminates WoFS dependence on nonoperational systems.

During each DA cycle, WoFS uses the prior adaptive inflation technique developed by Anderson (2009). Horizontal and vertical localizations are applied using the Gaspari and Cohn (1999) method and vary as a function of observation type. The localizations are used to reduce the impact of spurious covariances due to sampling error from limited ensemble size. Horizontal localization length is longest for conventional observations (~500 km) and shortest for high-density radar and satellite data (~20 km) (Jones et al. 2020). Additive noise is applied to temperature, humidity, and wind variables where reflectivity observations are greater than 35 dBZ and reflectivity mean innovations are greater than 10 dBZ (Dowell and Wicker 2009; Sobash and Wicker 2015). Finally, outlier thresholds are applied to reduce the impact of potentially nonrepresentative observations on the system.

c. *Compute resources required*

WoFS runs as an on-demand system and choices on grid spacing, membership, forecast length, and domain size are based on 1) minimum NWP requirements for the ensemble analysis and forecasts, 2) forecaster needs, and 3) available computing resources that could be dedicated to a continuously cycling DA and forecast system that typically runs 9–12 h each day during active severe weather. On the Cray XC30, a WoFS configuration requires about 2800 cores (across 120 24-core nodes) dedicated to the analysis cycling, ensemble forecasts, and postprocessing. For the analysis cycling with GSI-EnKF, each of the 36 ensemble members runs on a single node. The GSI-EnKF also runs within that 36-node partition. For the 18-member WRF forecasts, 54 nodes are dedicated (3 nodes for each member) to complete a 6-h forecast within 30 wall-clock minutes. The remaining nodes are used for the preprocessing of radar and satellite observations, as well as the postprocessing of forecast output. Each ensemble member outputs data at 5-min resolution through the 3- or 6-h forecast period. The ensemble postprocessing accounts for 20% of the total compute costs, with over 20 000 images[2] per 6-h forecast generated for the web interface. Each day generates approximately 5 TB of data that is stored locally. Because of the large data volumes, WoFS has been utilized as a use case for an EarthCube Research Networks project on determining best practices for archiving and reproducibility of geoscience model output (Mullendore et al. 2021; https://modeldatarcn.github.io).

Since WoFS is designed to be an "on-demand" NWP system, with the need to potentially run distinct simultaneous domains for different threats in different locations, NSSL decided to move WoFS to a computing platform that could 1) scale with the needs of users, 2) provide a complete user interface to launch and control runs, and 3) provide visualization of WoFS output, all within a unified software approach. In the spring 2020, a software engineering project to port WoFS to the Microsoft Azure cloud began within NSSL. By early 2022, cloud-based (cb)-WoFS replaced the "Cray-based" WoFS for all real-time computation and postprocessing; only the observational preprocessing remained local. In spring 2023, all of WoFS' pre- and postprocessing scripts, web-based visualization code, the WRF model, and the DA software is cloud-based. Cb-WoFS enabled seven dual-domain runs in 2023, which included days with two simultaneous Moderate Risk Outlooks issued by the SPC. Replacing the original shell

[2] For 6-h forecasts, 10 000 of these images are forecast soundings covering a 20 × 20 grid with output available every 15 min. Because of the inordinate amount of resources required to generate forecast sounding images, cb-WoFS generates them upon user request rather than pregenerating the full suite of images. This change results in increased loading time for sounding images on the web viewer but significantly mitigates the postprocessing computational expense.

scripts, the cb-WoFS workflow was developed into a software stack that uses Azure SDK's to execute cb-WoFS components on Azure platform and infrastructure services. This software controls the entire analysis, forecast, and pre and postprocessing tasks without intervention over a period of 12–15 h.

3. Visualizations of WoFS guidance

Beyond computing and model development challenges, it is crucial that WoFS output be rapidly disseminated to end users with products and visualization strategies that do not prohibitively increase forecasters' cognitive workload (e.g., Wilson et al. 2017, 2019b; Karstens et al. 2018; Demuth et al. 2020). Compared to short- and medium-range ensemble guidance products, the products for 0–6-h lead times produced by WoFS differ significantly and have not been fully explored in NWS operations. Because of the relatively small spatial scales and limited intrinsic predictability of thunderstorms (e.g., Potvin et al. 2017; Flora et al. 2018), storm locations in ensemble members drift apart with increasing forecast lead time. Thus, the appearance of a storm in the ensemble-mean simulated reflectivity field often becomes increasingly broad, smooth, and less intense with lead time due to spread in storm placement, giving the misleading impression that the storm is weakening. Therefore, storm-based WoFS guidance products have been designed to preserve information on the location, timing, motion, likelihood, and severity of discrete features (Skinner et al. 2023). WoFS probabilistic guidance products can be classified using three general categories: 1) probability of exceedance, 2) ensemble percentile, and 3) "paintball" products. These are produced for several different severe weather proxies, and examples from a forecast on 7 May 2020 are provided in Fig. 3.

a. Probability of exceedance

Probability of exceedance products are created by calculating the percentage of ensemble members that exceed a prescribed threshold within 4.5-, 7.5-, and 13.5-km radii of each grid point (e.g., Schwartz and Sobash 2017; Roberts et al. 2019). Using different radii helps account for spatial uncertainty and provides information on the likelihood of an event at different spatial scales. To illustrate, Fig. 3 shows probabilities of UH $\geq$ 60 m^2 s^{-2} within each radius, which reveals a trade-off between spatial precision and accounting for spatial uncertainty. In this case, most WoFS members predict mesocyclone development across the northeastern Texas Panhandle; however, only a few members overlap precisely, resulting in low probabilities using the 4.5-km radius (Fig. 3a). In contrast, WoFS members have high overlap (i.e., high confidence) for the track of a mesocyclone through northwest Texas using the same 4.5-km neighborhood. If a larger, 13.5-km radius neighborhood is used (Fig. 3c), it spreads out the information from the members. In this case, storms in the northeastern Texas Panhandle have increased overlap, resulting in a more confident and representative forecast given that most members predicted storms in that area. The larger neighborhood also spreads information in the more confidently predicted mesocyclone track in western northern Texas, resulting in a loss of precision and an overly broad prediction of the track. Therefore, an appropriate neighborhood for viewing ensemble probabilities of exceedance will be highly case dependent and vary with user needs and changes in ensemble spread. WoFS also computes probability of exceedance products for rainfall, severe hail and winds, and lightning flash extent density.

b. Percentiles

Complementing exceedance probabilities, ensemble *percentiles* are the value exceeded by a specified percentage of the ensemble members (Fig. 3d). For example, the 50th percentile is the ensemble median while the ensemble maximum is the largest value the ensemble predicts at each grid box. The percentile products are useful for quantifying potential severity (e.g., Novak et al. 2014; Demuth et al. 2020). The ensemble maximum, which is the most familiar percentile product, is often used to provide a worst-case scenario, as well as predict the region where an event is possible since it contains information from all members (Fig. 3e). Including information from every ensemble member, however, can overly emphasize unlikely events predicted by a single member and provide misleading guidance on event coverage. Therefore, the ensemble 90th percentile value is often used to provide a "reasonable" worst-case scenario (Fig. 3d).

c. Paintballs

Despite the utility of exceedance probabilities and percentiles, they do not provide member-specific information on thunderstorm characteristics like size, motion, and evolution, which forecasters often use to make short-term thunderstorm forecasts. Owing to time constraints, it is often not feasible to examine deterministic forecasts from each ensemble member in operations. To provide an ensemble guidance product comparable to radar analysis, paintball guidance products are created by plotting limited information from each ensemble member on the same plot (Schwartz et al. 2015). These plots are analogous to "spaghetti" plots often used in medium-range forecasting (Sivillo et al. 1997), except applied to discontinuous fields. An example is provided in Fig. 3f where simulated composite reflectivity $\geq$ 45 dBZ is shown with different colors assigned to each member. When these plots are animated, they can quickly provide useful information on storm timing, coverage, motion, and likelihood (based on how many paintballs overlap in space). The size and shape of the paintballs informs users on likely storm modes (e.g., linear versus cellular), and paintball motion can reveal instances of deviant motion associated with left- and right-moving supercells (Rotunno and Klemp 1982). The breadth of information and ease of interpretability provided by paintball plots have made them the most popular WoF guidance product during past real-time experiments (Wilson et al. 2021).

Each of the guidance products described above is produced as a single image for a given lead time. WoFS currently uses a web interface to serve experimental real-time guidance to end users (https://cbwofs.nssl.noaa.gov/forecast), and the interface enables users to rapidly cycle through or animate guidance

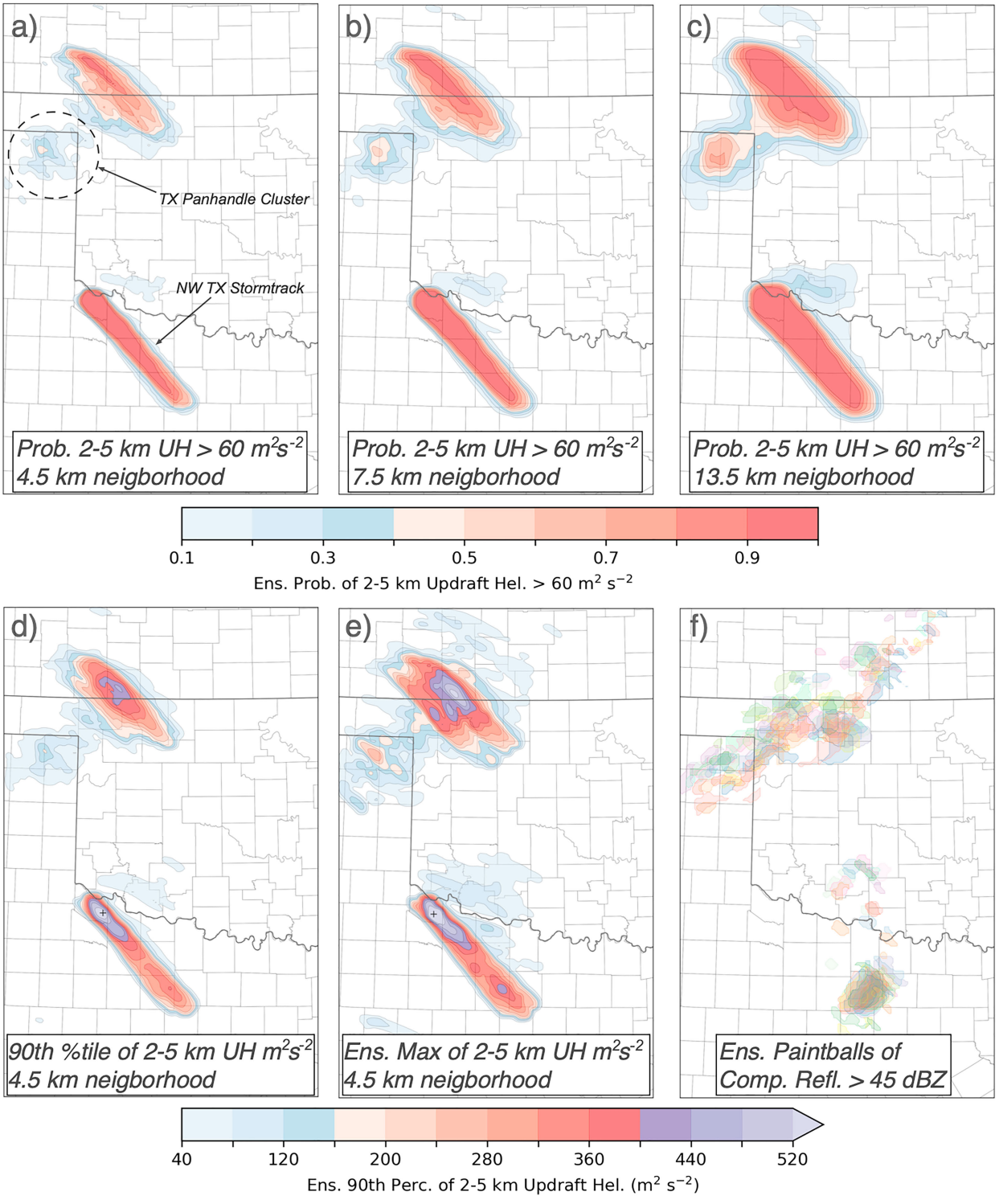

FIG. 3. Examples of WoFS 3-h probabilistic guidance products for the 2330 UTC 7 May 2020 forecast, valid at 0230 UTC 8 May. (top) The neighborhood maximum ensemble probability of 2–5-km UH ≥ 60 m^2 s^{-2} for (a) 4.5-, (b) 7.5-, and (c) 13.5-km neighborhood radii accumulated over the 3-h forecast. (bottom) The (d) ensemble 90th percentile and (e) ensemble maximum values of 2–5-km UH (m^2 s^{-2}) accumulated over the 3-h forecast and (f) simulated composite reflectivity paintballs exceeding 45 dB*Z* valid at 0230 UTC 8 May. A cluster of storms in northwest Texas and a supercell track through north Texas discussed in the text are annotated in (a).

products for each available lead time. Additionally, the web interface allows for implementation of interactive features, such as toggled overlays of geographical information and current real-world observations, and the ability to combine probabilistic and deterministic WoFS output. Several of these interactive features are unique compared to other model viewers.

A complication of rapidly cycled radar and satellite DA is that the quality of WoFS initial conditions will be correlated with the number of observations assimilated for each storm in the domain (e.g., Yussouf and Stensrud 2010; Stratman et al. 2020), resulting in varying accuracy of ensemble analyses and forecasts for different storms (Guerra et al. 2022). To determine how much confidence to have in WoFS guidance, users need an efficient means of assessing which storms are accurately analyzed in WoFS initial conditions, and which are not. Overlays of real-time radar and rainfall observations with WoFS guidance help to fill this need by enabling users to rapidly compare differences between the simulated and observed storms. Additionally, NWS forecasters have expressed a need for a combination of deterministic and probabilistic guidance to effectively utilize ensemble forecasts (Demuth et al. 2020). The WoFS viewer provides capabilities to overlay deterministic forecasts, such as UH swaths from individual members or WoF Hybrid, with any probabilistic product. This combination of probabilistic and deterministic guidance allows users to interrogate multiple aspects of forecast likelihood, severity, and evolution using a single animation. Beginning in 2022, select real-time WoFS probabilistic guidance products were demonstrated within AWIPS2 (Advanced Weather Interactive Processing System), providing access to WoFS in a fully interactive and integrated suite of operational observations, analyses, model outputs, and forecast tools.

d. *Explicit severe weather probabilities through machine learning*

Despite WoFS's usefulness for severe weather forecasting, grid spacing, and physical parameterization constraints limit explicit hazard guidance. Machine learning (ML) methods can enhance WoFS by using multiple predictors to establish multivariate relationships for skillful, calibrated probabilistic guidance. ML models have recently been implemented to calibrate WoFS 0–3-h object-based, probabilistic guidance for severe wind, severe hail, and tornadoes (Flora et al. 2021; known as WoFS-ML-Severe). WoFS-ML-Severe has run since spring 2021 and was included in a 2022 Hazardous Weather Testbed Spring Forecasting Experiment (HWT-SFE) forecasting activity. In this study, participants with access to the ML guidance produced better severe weather outlooks than those without (Clark et al. 2023).

4. Verification approaches for WoFS guidance

As with any modeling system, thorough forecast verification of WoFS is critical for quantifying skill, identifying areas for improvement, and monitoring progress as improvements are implemented. A big challenge for WoFS is that the severe weather hazards WoFS aims to predict are not observed well. The primary severe weather database for the United States, *Storm Data*, which is maintained by NOAA's National Centers for Environmental Information (NCEI) and SPC (http://www.ncdc.noaa.gov/stormevents/), has significant limitations, biases, and nonmeteorological influences unique to each hazard (e.g., Brooks et al. 2003; Trapp et al. 2006; Potvin et al. 2019, 2022). Despite these limitations, local storm reports (LSRs) in *Storm Data* are the most commonly used verification dataset for severe thunderstorm NWP (e.g., Sobash et al. 2011, 2016; Schwartz and Sobash 2017; Roberts et al. 2019). While remotely sensed observations can address issues with LSRs, such as undersampling and population biases, they have other limitations including limited representation (e.g., Doppler radars only measure one wind component) and varying data quality with geography (e.g., radar beam blockage). Another verification challenge is that WoFS does not explicitly predict severe weather hazards (e.g., tornadoes, hail size, and wind gusts). Thus, simulated severe hazard proxies that are known to be related to observed hazards are often used.

Another challenge for WoFS verification is accounting for "double penalties" that result when traditional, point-based metrics penalize small displacement errors once for a miss and again for a false alarm. In these situations, skill metrics can indicate poor performance when the forecasts are subjectively very skillful. Therefore, bulk verification of WoFS and other convection-allowing ensembles is generally accomplished using the spatial verification methods (e.g., Gilleland et al. 2009, 2010) described below.

a. *Neighborhood verification*

The neighborhood methods used to generate probabilistic WoFS guidance may also be used to generate verification fields for evaluating deterministic (e.g., Roberts and Lean 2008) and probabilistic (e.g., Sobash et al. 2011, 2016; Schwartz and Sobash 2017) guidance. Neighborhood verification has typically been applied to WoFS forecasts of heavy rainfall and severe thunderstorm hazards, and used to quantify differences in skill between WoFS and other forecast systems. For example, Lawson et al. (2018) quantified improvements in WoFS skill for 0–3-h deterministic forecasts of heavy rainfall over corresponding forecasts from two configurations of the HRRR model. The WoFS research team also found that WoFS could accurately predict the location and timing of heavy rainfall (Yussouf et al. 2016; Yussouf and Knopfmeier 2019) or tropical cyclone hazards (damaging winds, heavy rainfall, or tornadoes) with up to 6 h of lead time (Yussouf et al. 2020a).

Similarly, neighborhood verification has quantified improvements in probabilistic simulated reflectivity forecasts from WoFS relative to the HRRR Time-Lagged Ensemble (Roberts et al. 2020), and been used to evaluate probabilistic severe thunderstorm outlooks issued by participants in the Hazardous Weather Testbed (HWT) Spring Forecasting Experiment (SFE) using WoFS guidance (Wilson et al. 2021; Gallo et al. 2022). Finally, multiple neighborhood verification metrics may be aggregated for varying forecast quantities, lead times, and neighborhood sizes. Comparison of these

metrics between different forecast systems may be accomplished using a "scorecard" visualization that provides at-a-glance information about the strengths and weaknesses of different systems (Gallo et al. 2019). Scorecards of WoFS simulated reflectivity guidance can be produced relative to other CAM ensembles following each WoFS run (Matilla et al. 2021), providing measures of the bulk forecast skill and day-to-day variations in skill between systems. This scorecard is generated using the Model Evaluation Tools suite of products (METplus; http://dtcenter.org/community-code/metplus).

b. *Object-based verification*

A limitation of neighborhood verification is that it, by definition, spreads and smooths forecast data, which can obscure forecasts of individual small-scale events like thunderstorms (Flora et al. 2019). As a fundamental goal of WoFS is providing short-term predictions of hazards in individual thunderstorms, some method for verifying WoFS guidance on a storm-to-storm basis is needed. Object-based verification is a natural fit for this requirement as it identifies discrete events in forecast fields and matches them to observational proxies (e.g., Davis et al. 2006a,b).

As an example, objects representing thunderstorms in WoFS simulated reflectivity may be verified against corresponding objects in reflectivity observations from the WSR-88D network. Although in this example both the forecast and verification dataset include the same field, it is not an apples-to-apples comparison. Complicating this comparison are 1) resolution differences between the two datasets and 2) variation between reflectivity simulated by a microphysical parameterization in a numerical model and observed by radar. Therefore, objects are typically identified in the varying datasets using matched percentiles from the full distribution of values across all cases being considered (e.g., Mittermaier and Roberts 2010; Sobash et al. 2016) to help limit the impact of bias between different proxies for a given phenomenon. Once objects have been identified, they are matched according to desired criteria for measuring forecast accuracy. For example, composite reflectivity objects in WoFS are often matched to corresponding objects in MRMS data if they occur within the time and space scale of a typical NWS warning product (Skinner et al. 2018).

Once object pairs have been matched, a count of objects considered "hits," "misses," and "false alarms" can be produced; then verification metrics, derived from a 2×2 contingency table, can be used to evaluate the forecast quality. For example, Fig. 4 shows a performance diagram (Roebber 2009) of the daily mean scores of 45–75-min WoFS reflectivity object forecasts for each SFE case from 2017 to 2020. For the majority of cases, the mean probability of detection is higher than the false alarm ratio during this period, indicating that most thunderstorm objects in WoFS are predicted to be near the observed location of storms. Beyond these traditional verification metrics, object-based verification can produce extensive diagnostic information that can be used to characterize specific errors in the forecast (Wolff et al. 2014), such as quantifying an increase in accuracy for "older" storms relative to convection initiation,

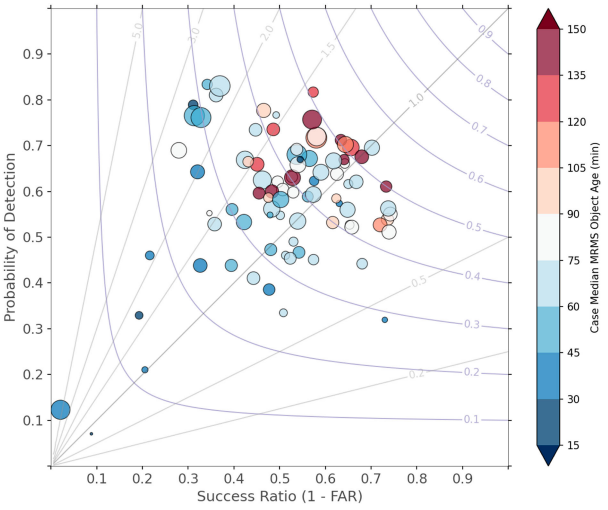

FIG. 4. Performance diagram (Roebber 2009) showing the daily, ensemble mean accuracy of WoFS 45–75-min forecasts of composite reflectivity objects across 87 SFE cases between 2017 and 2020. Reflectivity objects are identified in WoFS and matched to MRMS observations on spatial scales similar to a typical severe weather warning (approximately 32-km maximum displacement between objects) using a similar methodology to Skinner et al. (2018). Each marker on the diagram is the ensemble mean value for all forecasts and ensemble members for a given case, marker size is scaled by the total number of objects for a case. Markers are shaded by the median observed thunderstorm object age from convection initiation at WoFS initialization. Cases with older median storm ages are generally seen to result in more accurate forecasts owing to improved storm-scale WoFS initial conditions produced through radar and satellite assimilation (Guerra et al. 2022).

where multiple cycles of radar and satellite assimilation result in improved initial conditions and resulting forecasts (Fig. 4; Guerra et al. 2022).

Object-based verification provided the first bulk analysis for the skill of WoFS simulated reflectivity and UH forecasts, which were used as a proxy for thunderstorms and mesocyclones, respectively. Skinner et al. (2018) found that WoFS could skillfully predict both phenomena out to lead times of 3 h. This study was extended to probabilistic guidance for mesocyclones by Flora et al. (2019), who found that WoFS could produce reliable probabilities of mesocyclone occurrence in an object-based framework. Additionally, object-based verification has been used to establish skill for predicting specific events (Skinner et al. 2016; Yussouf et al. 2016; Pan and Gao 2022), quantify forecast improvements resulting from configuration changes (Jones et al. 2018a, 2020; Stratman et al. 2020; Kerr et al. 2021; Lawson et al. 2021; Miller et al. 2022), verify WoFS forecasts of upper-level clouds (Jones et al. 2018b) or tropical cyclone hazards (Jones et al. 2019), and create composite storm objects used to evaluate the accuracy of WoFS predictions of the near-storm environment (Britt et al. 2020; Laser et al. 2022; Potvin et al. 2020). While verification of WoFS forecast skill is critical to understanding the system's performance, as important is forecasters' evaluations of the usability and usefulness of WoFS 0–6-h forecast guidance in operations.

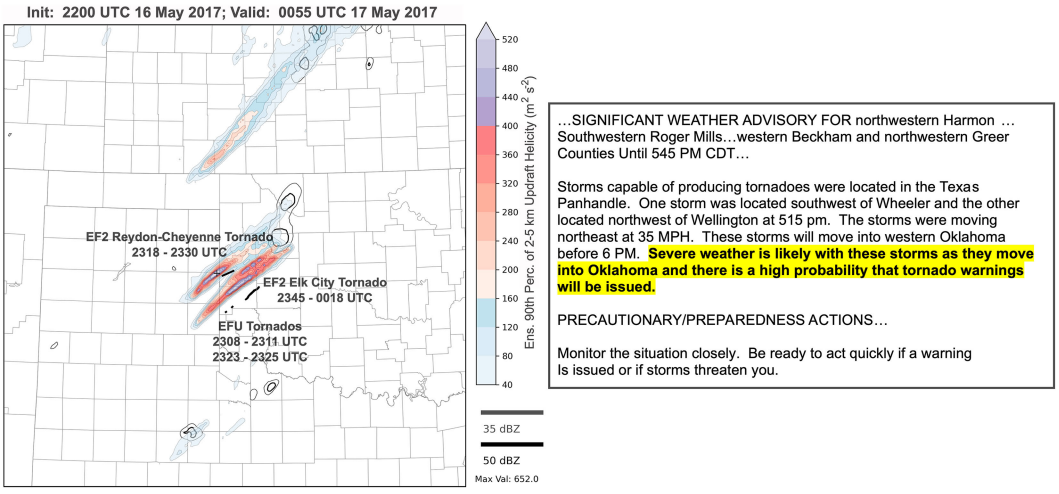

FIG. 5. First official use of WoFS in operations. (left) 90th percentile values of 2–5-km UH, valid 2200–0055 UTC 16 May 2017. Annotated are the paths of two EF2 and locations of two unrated, very short-lived tornadoes (EFUs) in west-central Oklahoma. (right) Significant Weather Advisory issued by Norman WFO, valid 2215–2245 UTC 16 May 2017.

5. User-focused research to address sociological challenges

Visualizations of WoFS guidance and incorporation of interdisciplinary expertise enabled the user-focused research that is a crucial and growing component of WoFS development. Since 2017, we have applied two approaches for conducting collaborative research between researchers and forecasters: 1) through HWT experiments, including the 2017–23 SFEs (Wilson et al. 2019a,b; Clark et al. 2021, 2022, 2023; Gallo et al. 2020; Burke et al. 2022; Skinner et al. 2023), Hydrometeorology Testbed Flash Flood and Intense Rainfall (HMT-FFaIR) experiment in 2018 (Weather Prediction Center 2018) and 2018/19 Hydrometeorology Testbed Multi-Radar Multi-Sensor Hydrology experiments (Yussouf et al. 2020b; Martinaitis et al. 2023), and 2) through experimental use in NWS operations, including the Storm Prediction Center (SPC), Weather Prediction Center (WPC) (Wilson et al. 2023), and Weather Forecast Offices (WFOs) (Burke et al. 2022; Skinner et al. 2023). HWT experiments test WoFS applications and usability in a pseudo-operational environment, whereas operational use tests the impact of WoFS guidance on decision making in the real world. Both approaches provide user-feedback to developers that enable the enhancement and addition of WoFS visualizations to best meet their forecast information needs and identify forecast challenges that need to be addressed. A summary of these outcomes is found in Skinner et al. (2023). These approaches also provide information on forecasters' understanding of WoFS guidance that informs training materials (e.g., Wilson et al. 2019a,b; Burke et al. 2022). This paper will next focus on how WoFS has been used in operations to provide messaging to the public and core partners on the timing, location, and intensity of significant weather hazards between Watch and Warning time scales.

a. *First use for operational decision-making*

The first official use of WoFS in operational decision making occurred on 16 May 2017 during real-time severe weather operations at the WFO Norman, OK. Therein, the joint interrogation of WoFS probabilistic guidance by WoFS researchers and WFO forecasters led to the issuance of a Significant Weather Advisory (Fig. 5). In this case, successive WoFS runs consistently forecast a high probability of two intense mesocyclones developing and then moving into western Oklahoma, which when coupled with the existing environment, suggested a high probability for severe weather and tornadoes. Two EF2 tornadoes occurred just south of the two swaths of noteworthy 2–5-km UH values per the WoFS 90th percentile of 2–5-km UH forecast (Fig. 5), beginning 78 min after the initial time of the ensemble forecast shown here. These included a long-track tornado that caused an estimated $25 million in property damage and killed one person in Elk City, Oklahoma (NCEI 2021). The language in the Significant Weather Advisory was unique, stating "Severe weather is likely with these storms as they move into Oklahoma and there is a high probability that tornado warnings will be issued." This statement indicated greater confidence and specificity than is traditionally achieved during Watches, and stopped just short of the immediate call to action that would be presented in eventual Tornado Warnings. This "watch-to-warning gap" is an area of active research, and is potentially the most relevant space for a 3-km WoFS to consistently make a positive impact on severe weather communication. In this case, Elk City Emergency Manager Lonnie Risenhoover reported, *"Based on the information [Significant Weather Advisory] from the National Weather Service, we were able to activate outdoor warning sirens about 30 min ahead of the tornado."*

b. *Use by SPC and a WFO for a large hail event*

Since 2018, WoFS has been available to SPC forecasters in their native product generation environment, NCEP AWIPS (NAWIPS). Display of WoFS guidance in NAWIPS allows the graphical component of products, such as the convective

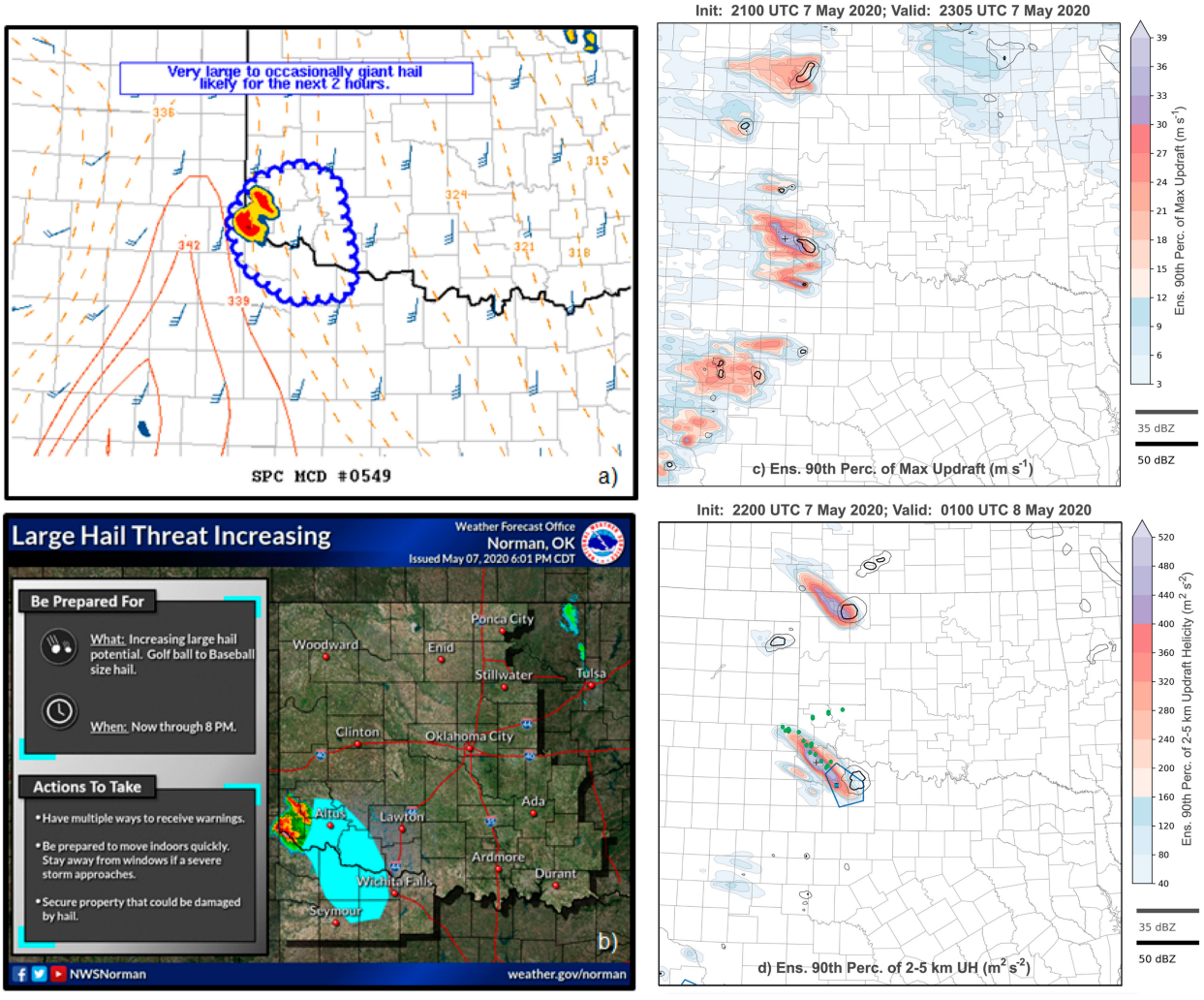

FIG. 6. (a) Mesoscale Convective Discussion issued by the Storm Prediction Center at 2304 UTC 7 May 2020, valid through 0000 UTC 8 May 2020. (b) Graphic issued by WFO Norman, OK, at 2301 UTC 7 May 2020. (c) WoFS percentile values of maximum updraft (m s^{-1}), initialized at 2100 UTC 7 May 2020 and valid through 2305 UTC 7 May 2020. Overlaid are the WoFS probability-matched mean composite reflectivity contours (30 and 50 dB*Z*) at 2305 UTC. (d) WoFS percentile values of UH (m^2 s^{-2}), initialized at 2200 UTC 7 May 2020 and valid through 0100 UTC 8 May 2020. Overlaid are severe hail reports valid at 2200–0100 UTC, a severe thunderstorm warning valid at 0100 UTC (blue), and WoFS probability-matched mean composite reflectivity contours (30 and 50 dB*Z*) at 0100 UTC.

outlook or mesoscale discussion (MD), to be drawn while using WoFS guidance as an underlay. WoFS guidance has been formally expressed to users by SPC as a part of forecast reasoning in Convective Outlooks and especially MDs, including confident identification of specific corridors of imminent severe weather potential and maximum hazard intensity. Of particular use to SPC are high-impact, low-predictability events well forecast by WoFS. An example of note is the use of WoFS guidance for post and downstream-watch decisions on 7 May 2020 (Fig. 6). Within the MD number 0549 issued at 2304 UTC 7 May 2020 (Fig. 6a), an SPC forecaster stated, "*A supercell near Childress has recently split with the left split moving northeastward in southwest Oklahoma. This storm split was well advertised by the NSSL Warn-on-Forecast system* (Fig. 6c) *while most other short-term guidance including the HRRR and HRRR-P suggest storm initiation should just be getting started now. In addition, the WoFS suggests the strongest updraft helicity may occur in the next 1 to 2 h* (Fig. 6d). *Current observations and short-term mesoscale evolution of the environment (strengthening LLJ and northward moisture advection) also support this, indicating that hail even larger than the 3.25 that has been reported so far, may be possible as the supercell further intensifies.*" During that evening, the WFO Norman, Oklahoma, also used WoFS guidance to message this supercell. At 2301 UTC, they issued a graphic forecast communicating an increasing potential for large hail, golf ball to baseball size, through 0100 UTC 8 May 2020. The orientation of the forecast corridor from Altus, Oklahoma, to Seymour, Texas, to west of Wichita Falls, Texas, indicated the issuing forecaster's trust in the predicted path of the right-turning supercell (Fig. 6b). A comparison of the forecast swath of WoFS 90th percentile 2–5-km UH values with

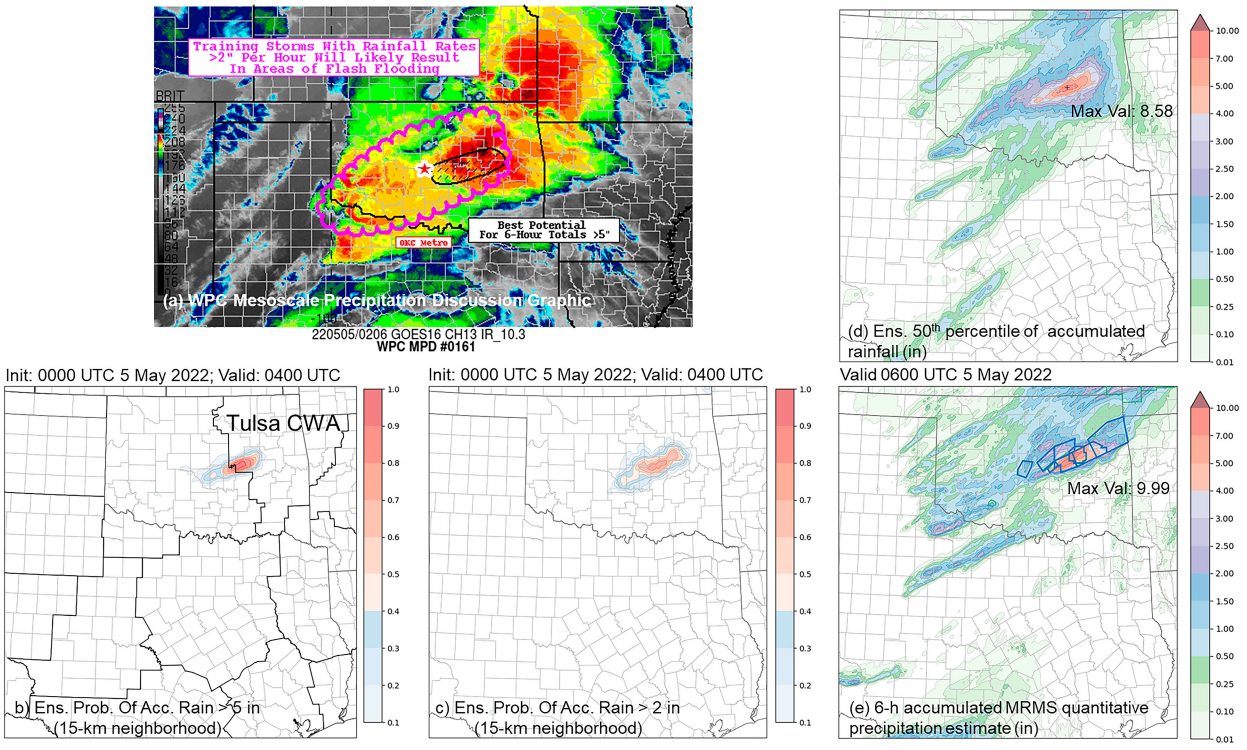

FIG. 7. (a) Mesoscale Precipitation Discussion image issued by the Weather Prediction Center at 0200 UTC 5 May 2022, valid through 0730 UTC 5 May 2020. (b) WoFS ensemble probability of accumulated rainfall > 5 in., initialized at 0000 UTC 5 May 2022 and valid through 0400 UTC 5 May 2022. (c) WoFS hourly probability of accumulated rainfall > 2 in., initialized at 0000 UTC 5 May 2022 and valid at 0400 UTC. (d) WoFS 50th percentile values of accumulated precipitation (in.), initialized at 0000 UTC 5 May 2022 and valid through 0600 UTC 5 May 2022. (e) The 6-h accumulated Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimate, valid at 0600 UTC 5 May 2022. Overlaid are flash flood warnings valid during the forecast period (blue).

recorded hail reports (NCEI 2021) shows close temporal and spatial correspondence (Fig. 6d). Furthermore, the overlap of the severe thunderstorm warning active at 0100 UTC with the WoFS probability matched mean 30- and 50-dBZ contours at the same time suggests potential utility of WoFS guidance for warning applications.

c. *Use by WPC and a WFO for a flash flood event*

While collaborative interactions with SPC and WFO forecasters have explored the usefulness of WoFS to real-time severe storms forecasting, collaborative interactions with WPC and WFO forecasters have tested applications of WoFS to flash flood forecasting. WoFS was run in real time during the summer months of 2017–20, in regions with a heightened risk of flash flooding per the WPC Excessive Rainfall Outlook (Burke et al. 2023). During the summers of 2019 and 2020, a longitudinal survey was issued after each event ($N = 85$) to query how WoFS impacted forecasters' situational awareness, confidence, workload, and product issuance decisions (Wilson et al. 2023). Forecasters reported that while use of WoFS guidance increased their workload in 55% of events, its use also increased their forecast confidence in 75% of events. Among WoFS guidance attributes, convective mode was rated most highly, followed by the intensity, coverage, timing, and then location of thunderstorms.

WoFS guidance has been formally expressed to users as a part of forecast reasoning in Mesoscale Precipitation Discussions issued by WPC, especially when contributing to confident identification of storm mode and persistent high probabilities of high rainfall rates over a specific area. One example is the use of WoFS guidance for heavy rainfall and flash flooding decisions on 4 May 2022 (Fig. 7). Within the Mesoscale Precipitation Discussion 0161 issued at 0200 UTC 5 May 2022 (Fig. 7a), a WPC forecaster stated, "*The experimental 00Z WoFS showed a series of training 40 dBZ paintballs across the midsection of Oklahoma with the area seeing the longest residency time being east of OKC.*"

"*Remarkably, the QPF 50th percentile of the 00Z WoFS between 00–06Z included a maximum of 8" east of OKC* (Fig. 7d) *with the 90th percentile even higher* (not shown). *It also identified a > 60% chance for WoFS ensemble probabilities of rainfall rates >2"/hr east of OKC this evening between 02–05Z)* (Fig. 7c)." WFO Tulsa, Oklahoma, also used WoFS guidance during this event. Roughly coincident with the issuance of their first flash flood warning, at 0101 UTC 5 May 2022, a forecaster at WFO Tulsa noted in NWSChat, "*NSSL Warn on Forecast data has high probabilities of exceeding 5 inches of rain in much of Okfuskee and Okmulgee Co.s* (Fig. 7b)*… this area has seen heavy rain over the last 3 days… serious flash flooding is likely as a series of storms move through this evening.*" During this event, the

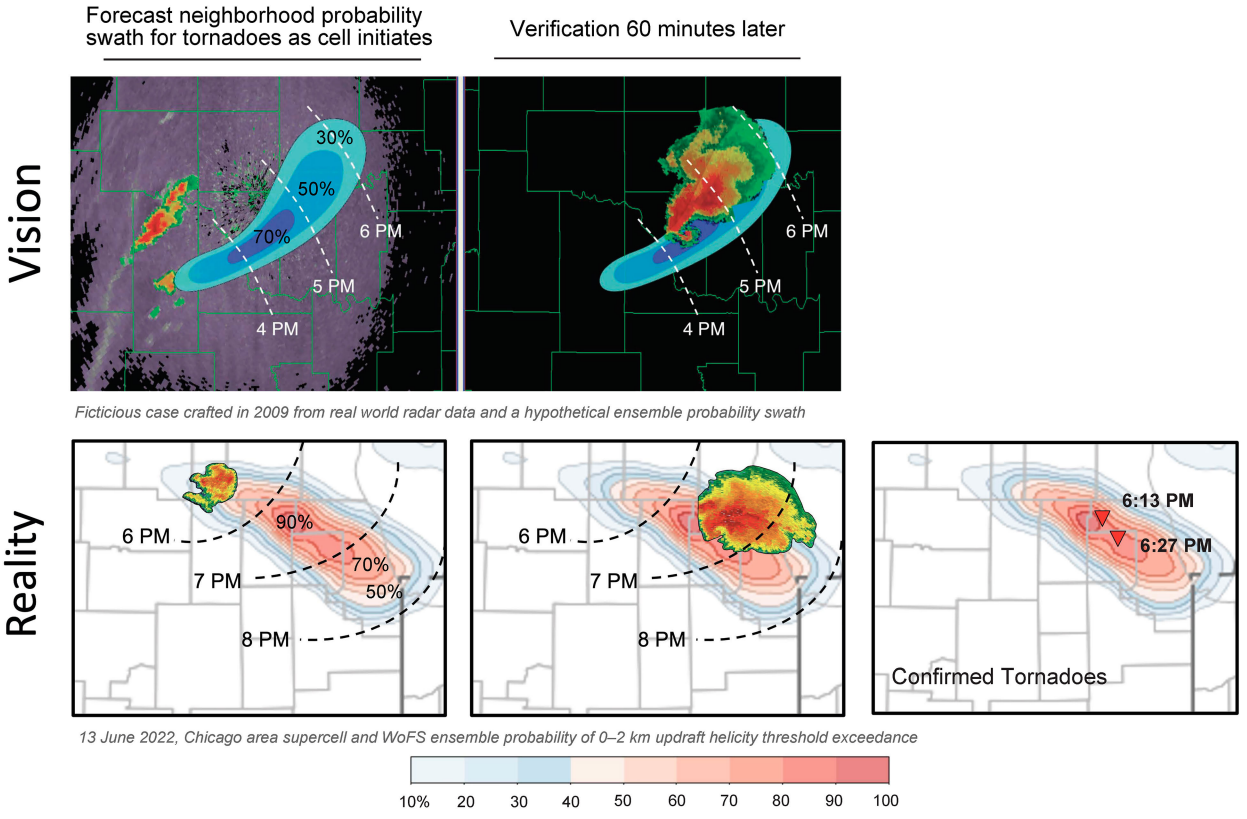

FIG. 8. (top) Original schematic for the Warn-on-Forecast vision from Stensrud et al. (2009), (left) depicting a hypothetical ensemble predicted swath of tornado probability (by proxy) based on assimilation of observations from a developing thunderstorm followed by (right) the verifying radar image of a mature tornadic thunderstorm 60 min later overlain on the forecast swath. (bottom) Data from a real-world case on 13 Jun 2022, depicting the ensemble probability of 0–2-km UH ≥ 20 m^2 s^{-2} (15-km neighborhood) over a 2.5-h period. The likelihood of significant 0–2-km UH was predicted to peak at >90% in northwest suburbs of Chicago, and two EF0 tornadoes occurred within the swath approximately 40–60 min after the forecast was initialized.

availability of a mesoanalyst enabled ongoing monitoring, verification, and communication of WoFS frequently updating probabilistic forecast information to the warning forecaster. Postevent feedback from WFO Tulsa to NSSL reported that, in addition to HRRR forecasts, the WoFS increasing probabilities of 2-, 3-, and 5-in. accumulations and the 50th percentile values were used as guidance during the flash flood warning decision process. A comparison of the WoFS 6-h forecast 50th percentiles of accumulated rainfall valid at 0600 UTC (Fig. 7d), with the accumulated MRMS quantitative precipitation estimate valid at 0600 UTC (Fig. 7e), shows spatial correspondence with the extreme rainfall amounts and flash flood warnings issued over a narrow swath within northeast Oklahoma.

d. Realization of warn-on-forecast vision

In 2009, the capability to produce ensemble forecasts of individual thunderstorms and their hazards in real time was a vision illustrated within Stensrud et al. (2009). As shown in Fig. 8, Stensrud et al. (2009) envisioned a forecast system that, at storm initiation, could predict individual thunderstorms, like supercells, and associated probability swaths for high-impact weather, like tornadoes, up to an hour in advance. A primary goal of NSSL's warn-on-forecast project has been to conduct the research and development to realize this vision. One example of note is the WoFS prediction of the 13 June 2022 tornadic supercell near Chicago, Illinois (Fig. 8). The WoFS ensemble probability of 0–2-km UH (15-km neighborhood), a proxy for tornado prediction, was predicted to peak at >90% in the northwest suburbs of Chicago. Two EF0 tornadoes occurred within the swath approximately 40–60 min after the forecast was initialized. This event is one of many demonstrating the realization of the warn-on-forecast vision.

6. Conclusions and future work

More than a decade of work has brought us to an era in which a 3-km version of WoFS influences the lead-time, specificity, and uncertainty expression of hazard information on the watch-to-warning scale. WoFS has demonstrated its potential in achieving the original, idealized vision of a system whose probabilistic swaths, when viewed alongside observations and traditional datasets, enable NWS meteorologists to "warn-on-forecast."

The capabilities arise from WoFS design: a frequently updating, high-resolution, storm-scale ensemble prediction system that can consistently provide accurate guidance of

thunderstorm coverage, mode, and evolution over the course of a severe weather event (Skinner et al. 2018; Flora et al. 2019; Miller et al. 2022; Guerra et al. 2022). In experimental real-time runs, WoFS has shown that rapidly cycled (15 min) assimilation of WSR-88D and *GOES-16* and *GOES-17* observations offer forecasters, for the first time, probabilistic NWP guidance on the storm scale, depicting weather evolution at 5-min resolution. Real-time WoFS output (e.g., probabilities, forecast timing, convective mode) adjusts through time with much the same cadence experienced by NWS meteorologists—who also mentally assimilate these observations in their workflow during severe weather operations. Development of interactive visualizations unique to WoFS have been informed by continual communication between researchers/developers and operational forecasters, in the hopes of streamlining the ability to browse NWP content that is relaunched every half hour within cognitively demanding NWS watch-to-warning operations (Wilson et al. 2021; Skinner et al. 2023; Wilson et al. 2023).

Beyond its original mission of providing accurate, short-term, probabilistic guidance of convective storm hazards, WoFS has proven beneficial in unexpected ways. First, a forecaster in Kentucky documented a case in which the system accurately predicted a swath of high probability of severe surface winds (>50 kt; 1 kt $\approx$ 0.51 m s^{-1}) driven by evaporative cooling as forward-anvil precipitation descended into a pocket of dry air ahead of thunderstorms in central Tennessee (Burke et al. 2022). This forecast example and other traits of WoFS guidance may prove applicable to aviation interests (Avey et al. 2023). Second, WoF applications extend beyond predicting continental convection. WoFS forecasts of landfalling tropical cyclones have accurately predicted the location, intensity and spatial distribution of intense rainfall and thunderstorm-scale low-level rotation (Jones et al. 2019; Yussouf et al. 2020a). WoFS has also shown the potential for forecasting landfalling tropical cyclone eyewall wind speeds along the coast and after the eye is fully inland (Jones et al. 2019). Third, WoFS is being extended to fire and air-quality applications through real-time experimental use in NWS and state forestry operations (Lindley et al. 2023), and through assimilating satellite-derived aerosol properties and fire detections using technology developed for the operational HRRR-smoke system (Jones et al. 2022). Fourth, the development of cb-WoFS is enabling multiple, discrete WoFS domains to run simultaneously when high-impact weather threats are anticipated over different geographic regions of the United States. Fifth, we continue to investigate potential improvements to the WoFS ensemble through the assimilation of additional boundary layer observations and the implementation of new reflectivity and dual-polarimetric forward operators developed for S-band WSR-88D radars (Zhang et al. 2021; Hu et al. 2023).

An active and promising avenue for improving WoFS guidance is using machine learning (ML) to postprocess the ensemble forecast output. ML algorithms are becoming increasingly common in the severe weather prediction community, ranging from nowcasting to next-day forecasting and beyond (e.g., Mecikalski et al. 2015; Gagne et al. 2017; Lagerquist et al. 2017, 2020; McGovern et al. 2017, 2019a,b, 2023; Burke et al. 2020; Cintineo et al. 2020; Hill et al. 2020; Steinkruger et al. 2020; Sobash et al. 2020; Loken et al. 2020). As discussed above, ML products currently present in the WoFS are providing utility to operational forecasters. We are actively developing ML-based approaches including grid-based probabilities for severe weather hazards (Clark and Loken 2022) and predicting the situational accuracy of WoFS forecasts (Potvin et al. 2023) to help forecasters calibrate their confidence in WoFS guidance per event, and thereby issue more accurate products. Finally, a project is underway that aims to merge WoFS dynamical forecast data with ProbSevere output using an ML algorithm (Loken et al. 2022). ProbSevere operates in a storm-centric framework producing hazard probabilities every two minutes for individual storms using data from MRMS, satellites, and RAP analyses (Cintineo et al. 2020). Since WoFS is initialized every 30 min with latency around 20 min, incorporating the more frequently updating ProbSevere information via ML methods will be critical for using WoFS in warning operations (i.e., 0–90-min lead times).

In the future, we envision WoFS technologies for subhourly cycling DA being eventually applied in the forthcoming Rapid Refresh Forecast System (RRFS; Carley et al. 2021), a convection-allowing ensemble covering North America, which is intended to replace current regional modeling capabilities (e.g., RAP/HRRR, NAM, SREF, and HREF) as part of NOAA's Unified Forecast System (UFS; https://ufscommunity.org/) initiative. As RRFS capabilities advance, WoFS will advance in parallel to always provide enhanced prediction capabilities over limited domains embedded within RRFS. In fact, the WoFS team is already engaged in research and development for next-generation WoFS capabilities, which include potentially replacing the WRF dynamic core with the Model for Prediction Across Scales (MPAS; Skamarock et al. 2012, 2018), increased resolution (i.e., 1-km grid spacing and below), and upgraded DA using the JEDI (Joint Effort for Data assimilation Integration; Auligne 2021) framework.

Acknowledgments. The authors thank the following people and organizations for their immeasurable help and support given to the Warn on Forecast project. WoFS would not exist except for the vision of our former NSSL director Dr. James Kimpel and former Deputy Director Kevin Kelleher, who first articulated this vision in 2003, and then persevered to obtain the funding for the initial WoF project in 2009 from NOAA and OAR. Dr. David Stensrud initially directed the program and provided strong scientific and technical leadership during the early years. Dr. Steven Koch was instrumental in obtaining additional NOAA support for the program in 2015, which enabled real-time demonstration. We are deeply grateful for the long-term support from our National Weather Center partners, including WFO Norman, Oklahoma (Michael Foster, David Andra, and Todd Lindley); the Storm Prediction Center (Dr. Russell Schneider, Steven Weiss, and Israel Jirak); CAPS (Dr. Ming Xue); and the Multi-scale Analysis and Prediction (MAP) group (Dr. Xuguang Wang). NOAA's Global Systems

Laboratory, led by Dr. Curtis Alexander, has partnered closely with the WoF team and contributed immensely to our real-time success for the past six years. In-house support from NSSL's Weather Research and Forecasting Division, especially the Multi-Radar Multi-Sensor team, enhanced our input data quality and aided our verification efforts tremendously. Support for satellite data products (SATCORPS) was kindly provided by the NASA Langley Research Center (Dr. Patrick Minnis, Dr. Kris Bedka, Dr. William Smith, and Dr. Rabindra Palikonda). Last but not least, we want to thank other members of the warn on forecast team both past and present for their enthusiastic contributions. Special thanks go to Dr. Vanna Chmielewski, Jessica Choate, Dr. JunJun Hu, Dr. John Lawson, Dr. Swapan Mallick, Sijie Pan, Dr. Dustan Wheatley, and Dr. Yaping Wang. And we thank Dr. Kodi Berry and Dr. Thomas Galarneau for their insightful reviews. This research was and continues to be supported by the NOAA Warn-on-Forecast project; it was also funded under the NOAA–University of Oklahoma Cooperative Agreement NA16OAR4320115.

Data availability statement. The WoFS output images used in this paper are available from https://wof.nssl.noaa.gov/. Official forecast images are available from https://spc.noaa.gov, https://wpc.noaa.gov, and https://mesonet.agron.iastate.edu/. Additional model data and system configuration information are available upon request.

REFERENCES

Anderson, J. L., 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus*, **61A**, 72–83, https://doi.org/10.1111/j.1600-0870.2008.00361.x.

Aksoy, A., D. C. Dowell, and C. Snyder, 2009: A multicase comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Mon. Wea. Rev.*, **137**, 1805–1824, https://doi.org/10.1175/2008MWR2691.1.

——, ——, and ——, 2010: A multicase comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part II: Short-range ensemble forecasts. *Mon. Wea. Rev.*, **138**, 1273–1292, https://doi.org/10.1175/2009MWR3086.1.

Auligne, T., 2021: JCSDA's vision of a community data assimilation for research and operations. *WCRP–WWRP Symp. on Data Assimilation and Reanalysis*, online, Deutscher Wetterdienst, O1-3, https://symp-bonn2021.sciencesconf.org/data/357293.pdf.

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, https://doi.org/10.1175/BAMS-D-22-0273.1.

Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, https://doi.org/10.1175/MWR-D-15-0242.1.

Britt, K. C., P. S. Skinner, P. L. Heinselman, and K. H. Knopfmeier, 2020: Effects of horizontal grid spacing and inflow environment on forecasts of cyclic mesocyclogenesis in NSSL's Warn-on-Forecast System (WoFS). *Wea. Forecasting*, **35**, 2423–2444, https://doi.org/10.1175/WAF-D-20-0094.1.

Brooks, H. E., C. A. Doswell III, and M. P. Kay, 2003: Climatological estimates of local daily tornado probability for the United States. *Wea. Forecasting*, **18**, 626–640, https://doi.org/10.1175/1520-0434(2003)018<0626:CEOLDT>2.0.CO;2.

Burke, A., N. Snook, D. J. Gagne II, S. McCorkle, and A. McGovern, 2020: Calibration of machine learning–based probabilistic hail predictions for operational forecasting. *Wea. Forecasting*, **35**, 149–168, https://doi.org/10.1175/WAF-D-19-0105.1.

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

——, A. Lamers, G. Carbin, M. J. Erickson, M. Klein, M. Chenard, J. McNatt, and L. Wood, 2023: The excessive rainfall outlook at the Weather Prediction Center: Operational definition, construction, and real-time collaboration. *Bull. Amer. Meteor. Soc.*, **104**, E542–E562, https://doi.org/10.1175/BAMS-D-21-0281.1.

Carley, J. R., and Coauthors, 2021: Status of NOAA's next generation convection-allowing ensemble: The Rapid Refresh forecast system. *WAF Symp. General Session*, online, Amer. Meteor. Soc., 12.8, https://ams.confex.com/ams/101ANNUAL/meetingapp.cgi/Paper/378383.

Cintineo, J. L., M. J. Pavolonis, J. M. Sieglaff, L. Cronce, and J. Brunner, 2020: NOAA ProbSevere v2.0—ProbHail, ProbWind, and ProbTor. *Wea. Forecasting*, **35**, 1523–1543, https://doi.org/10.1175/WAF-D-19-0242.1.

Clark, A. J., and E. D. Loken, 2022: Machine learning–derived severe weather probabilities from Warn-on-Forecast System. *Wea. Forecasting*, **37**, 1721–1740, https://doi.org/10.1175/WAF-D-22-0056.1.

——, and Coauthors, 2021: A real-time, virtual spring forecasting experiment to advance severe weather prediction. *Bull. Amer. Meteor. Soc.*, **102**, E814–E816, https://doi.org/10.1175/BAMS-D-20-0268.1.

——, and Coauthors, 2022: The second real-time, virtual Spring Forecasting Experiment to advance severe weather prediction. *Bull. Amer. Meteor. Soc.*, **103**, E1114–E1116, https://doi.org/10.1175/BAMS-D-21-0239.1.

——, and Coauthors, 2023: The first hybrid NOAA Hazardous Weather Testbed Spring Forecasting Experiment for advancing severe weather prediction. *Bull. Amer. Meteor. Soc.*, https://doi.org/10.1175/BAMS-D-23-0275.1, in press.

Davis, C., B. Brown, and R. Bullock, 2006a: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784, https://doi.org/10.1175/MWR3145.1.

——, ——, and ——, 2006b: Object-based verification of precipitation forecasts. Part II: Application to convective rain systems. *Mon. Wea. Rev.*, **134**, 1785–1795, https://doi.org/10.1175/MWR3146.1.

Dawson, D. T., II, M. Xue, J. A. Milbrandt, and M. K. Yau, 2010: Comparison of evaporation and cold pool development between single-moment and multimoment bulk microphysics schemes in idealized simulations of tornadic thunderstorms. *Mon. Wea. Rev.*, **138**, 1152–1171, https://doi.org/10.1175/2009MWR2956.1.

——, L. J. Wicker, E. R. Mansell, and R. L. Tanamachi, 2012: Impact of the environmental low-level wind profile on ensemble forecasts of the 4 May 2007 Greensburg, Kansas, tornadic

storm and associated mesocyclones. *Mon. Wea. Rev.*, **140**, 696–716, https://doi.org/10.1175/MWR-D-11-00008.1.

——, E. R. Mansell, Y. Jung, L. J. Wicker, M. R. Kumjian, and M. Xue, 2014: Low-level Z_{DR} signatures in supercell forward flanks: The role of size sorting and melting of hail. *J. Atmos. Sci.*, **71**, 276–299, https://doi.org/10.1175/JAS-D-13-0118.1.

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

Dowell, D. C., and L. J. Wicker, 2009: Additive noise for storm-scale ensemble data assimilation. *J. Atmos. Oceanic Technol.*, **26**, 911–927, https://doi.org/10.1175/2008JTECHA1156.1.

——, F. Zhang, L. J. Wicker, C. Snyder, and N. A. Crook, 2004: Wind and temperature retrievals in the 17 May 1981 Arcadia, Oklahoma, supercell: Ensemble Kalman filter experiments. *Mon. Wea. Rev.*, **132**, 1982–2005, https://doi.org/10.1175/1520-0493(2004)132<1982:WATRIT>2.0.CO;2.

——, L. J. Wicker, and C. Snyder, 2011: Ensemble Kalman filter assimilation of radar observations of the 8 May 2003 Oklahoma City supercell: Influences of reflectivity observations on storm-scale analyses. *Mon. Wea. Rev.*, **139**, 272–294, https://doi.org/10.1175/2010MWR3438.1.

——, and Coauthors, 2016: Development of a High-Resolution Rapid Refresh Ensemble (HRRRE) for severe weather forecasting. *28th Conf. on Severe Local Storms*, Portland, OR, Amer. Meteor. Soc., 8B.2, https://ams.confex.com/ams/28SLS/webprogram/Paper301555.html.

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

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, https://doi.org/10.1175/MWR-D-17-0374.1.

——, 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, https://doi.org/10.1175/WAF-D-19-0094.1.

——, 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, https://doi.org/10.1175/MWR-D-20-0194.1.

Gagne, D. J., II, A. McGovern, S. E. Haupt, R. A. Sobash, J. K. Williams, and M. Xue, 2017: Storm-based probabilistic hail forecasting with machine learning applied to convection-allowing ensembles. *Wea. Forecasting*, **32**, 1819–1840, https://doi.org/10.1175/WAF-D-17-0010.1.

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, https://doi.org/10.1175/WAF-D-21-0160.1.

Gallo, B. T., and Coauthors, 2019: Initial development and testing of a convection-allowing model scorecard. *Bull. Amer. Meteor. Soc.*, **100**, ES367–ES384, https://doi.org/10.1175/BAMS-D-18-0218.1.

——, and Coauthors, 2020: Scorecards for convection-allowing models. *Bull. Amer. Meteor. Soc.*, **101**, 59–64, https://doi.org/10.1175/BAMS-D-18-0218.A.

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

Gao, J., and M. Xue, 2008: An efficient dual-resolution approach for ensemble data assimilation and tests with assimilated Doppler radar data. *Mon. Wea. Rev.*, **136**, 945–963, https://doi.org/10.1175/2007MWR2120.1.

——, and D. J. Stensrud, 2014: Some observing system simulation experiments with a hybrid 3DEnVAR system for storm-scale radar data assimilation. *Mon. Wea. Rev.*, **142**, 3326–3346, https://doi.org/10.1175/MWR-D-14-00025.1.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757, https://doi.org/10.1002/qj.49712555417.

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, https://doi.org/10.1175/2009WAF2222269.1.

——, D. A. Ahijevych, B. G. Brown, and E. E. Ebert, 2010: Verifying forecasts spatially. *Bull. Amer. Meteor. Soc.*, **91**, 1365–1376, https://doi.org/10.1175/2010BAMS2819.1.

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

Hill, A. J., G. R. Herman, and R. S. Schumacher, 2020: Forecasting severe weather with random forests. *Mon. Wea. Rev.*, **148**, 2135–2161, https://doi.org/10.1175/MWR-D-19-0344.1.

Howard, S. P., K. E. Klockow-McClain, A. P. Boehmer, and K. M. Simmons, 2021: Firm behavior in the face of severe weather: Economic analysis between probabilistic and deterministic warnings. *Wea. Forecasting*, **36**, 757–767, https://doi.org/10.1175/WAF-D-20-0107.1.

Hu, J., and Coauthors, 2021a: Evaluation of an experimental Warn-on-Forecast 3DVAR analysis and forecast system on quasi-real-time short-term forecasts of high impact weather events. *Quart. J. Roy. Meteor. Soc.*, **147**, 4063–4082, https://doi.org/10.1002/qj.4168.

——, and Coauthors, 2021b: Assessment of storm-scale real time assimilation of GOES-16 GLM lightning-derived water vapor mass on short term precipitation forecasts during the 2020 Spring Forecast Experiment. *J. Geophys. Res. Atmos.*, **126**, e2021JD034603, https://doi.org/10.1029/2021JD034603.

——, J. Gao, C. Liu, G. Zhang, P. Heinselman, and J. T. Carlin, 2023: Test of power transformation function to hydrometeor and water vapor mixing ratios for direct variational assimilation of radar reflectivity data. *Wea. Forecasting*, **38**, 1995–2010, https://doi.org/10.1175/WAF-D-22-0158.1.

Hu, M., G. Ge, C. Zhou, D. Stark, H. Shao, K. Newman, J. Beck, and X. Zhang, 2018: Grid-point Statistical Interpolation (GSI) user's guide version 3.7. Developmental Testbed Center, 147 pp., http://www.dtcenter.org/com-GSI/users/docs/index.php.

Igel, A. L., M. R. Igel, and S. C. van den Heever, 2015: Make it a double? Sobering results from simulations using single-moment microphysics schemes. *J. Atmos. Sci.*, **72**, 910–925, https://doi.org/10.1175/JAS-D-14-0107.1.

Jankov, I., J. Beck, J. Wolff, M. Harrold, J. B. Olson, T. Smirnova, C. Alexander, and J. Berner, 2019: Stochastically perturbed parameterizations in an HRRR-based ensemble. *Mon. Wea. Rev.*, **147**, 153–173, https://doi.org/10.1175/MWR-D-18-0092.1.

Johnson, M., Y. Jung, J. A. Milbrandt, H. Morrison, and M. Xue, 2019: Effects of the representation of rimed ice in bulk microphysics schemes on polarimetric signatures. *Mon. Wea. Rev.*, **147**, 3785–3810, https://doi.org/10.1175/MWR-D-18-0398.1.

Jones, T., D. J. Stensrud, P. Minnis, and R. Palikonda, 2013a: Evaluation of a forward operator to assimilate cloud water path into WRF-DART. *Mon. Wea. Rev.*, **141**, 2272–2289, https://doi.org/10.1175/MWR-D-12-00238.1.

——, J. A. Otkin, D. J. Stensrud, and K. Knopfmeier, 2013b: Assimilation of satellite infrared radiances and Doppler radar observations during a cool season observing system simulation experiment. *Mon. Wea. Rev.*, **141**, 3273–3299, https://doi.org/10.1175/MWR-D-12-00267.1.

——, ——, ——, and ——, 2014: Forecast evaluation of an observing system simulation experiment assimilating both radar and satellite data. *Mon. Wea. Rev.*, **142**, 107–124, https://doi.org/10.1175/MWR-D-13-00151.1.

——, D. Stensrud, L. Wicker, P. Minnis, and R. Palikonda, 2015: Simultaneous radar and satellite data storm-scale assimilation using an ensemble Kalman filter approach for 24 May, 2011. *Mon. Wea. Rev.*, **143**, 165–194, https://doi.org/10.1175/MWR-D-14-00180.1.

——, K. Knopfmeier, D. Wheatley, G. Creager, P. Minnis, and R. Palikonda, 2016: Storm-scale data assimilation and ensemble forecasting with the NSSL experimental Warn-on-Forecast System. Part II: Combined radar and satellite data experiments. *Wea. Forecasting*, **31**, 297–327, https://doi.org/10.1175/WAF-D-15-0107.1.

——, X. Wang, P. Skinner, A. Johnson, and Y. Wang, 2018a: 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, https://doi.org/10.1175/MWR-D-17-0280.1.

——, P. Skinner, K. Knopfmeier, E. Mansell, P. Minnis, R. Palikonda, and W. Smith Jr., 2018b: Comparison of cloud microphysics schemes in a Warn-on-Forecast System using synthetic satellite objects. *Wea. Forecasting*, **33**, 1681–1708, https://doi.org/10.1175/WAF-D-18-0112.1.

——, ——, 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, https://doi.org/10.1175/BAMS-D-18-0203.1.

——, and Coauthors, 2020: Assimilation of *GOES-16* radiances and retrievals into the Warn-on-Forecast System. *Mon. Wea. Rev.*, **148**, 1829–1859, https://doi.org/10.1175/MWR-D-19-0379.1.

——, R. Ahmadov, E. James, G. Pereiria, S. Freitas, and G. Grell, 2022: Prototype of a Warn-on-Forecast System for Smoke (WoFS-Smoke). *Wea. Forecasting*, **37**, 1191–1209, https://doi.org/10.1175/WAF-D-21-0143.1.

Jung, Y., M. Xue, and M. Tong, 2012: Ensemble Kalman filter analyses of the 29–30 May 2004 Oklahoma tornadic thunderstorm using one- and two-moment bulk microphysics schemes, with verification against polarimetric radar data. *Mon. Wea. Rev.*, **140**, 1457–1475, https://doi.org/10.1175/MWR-D-11-00032.1.

Kain, J. S., S. R. Dembek, S. J. Weiss, J. L. Case, J. J. Levit, and R. A. Sobash, 2010: Extracting unique information from high-resolution forecast models: Monitoring selected fields and phenomena every time step. *Wea. Forecasting*, **25**, 1536–1542, https://doi.org/10.1175/2010WAF2222430.1.

Karstens, C. D., and Coauthors, 2018: Development of a human–machine mix for forecasting severe convective events. *Wea. Forecasting*, **33**, 715–737, https://doi.org/10.1175/WAF-D-17-0188.1.

Kerr, C. A., L. J. Wicker, and P. S. Skinner, 2021: Updraft-based adaptive assimilation of radial velocity observations in a Warn-on-Forecast System. *Wea. Forecasting*, **36**, 21–37, https://doi.org/10.1175/WAF-D-19-0251.1.

Labriola, J., N. Snook, Y. Jung, and M. Xue, 2019a: Explicit ensemble prediction of hail in 19 May 2013 Oklahoma City thunderstorms and analysis of hail growth processes with several multimoment microphysics schemes. *Mon. Wea. Rev.*, **147**, 1193–1213, https://doi.org/10.1175/MWR-D-18-0266.1.

——, ——, M. Xue, and K. W. Thomas, 2019b: Forecasting the 8 May 2017 severe hail storm in Denver, Colorado, at convection-allowing resolution: Understanding rimed ice treatments in multimoment microphysics schemes and their effect on hail size forecasts. *Mon. Wea. Rev.*, **147**, 3045–3068, https://doi.org/10.1175/MWR-D-18-0319.1.

Lagerquist, R., A. McGovern, and T. Smith, 2017: Machine learning for real-time prediction of damaging straight-line convective wind. *Wea. Forecasting*, **32**, 2175–2193, https://doi.org/10.1175/WAF-D-17-0038.1.

——, ——, C. R. Homeyer, D. J. Gagne II, and T. Smith, 2020: Deep learning on three-dimensional multiscale data for next-hour tornado prediction. *Mon. Wea. Rev.*, **148**, 2837–2861, https://doi.org/10.1175/MWR-D-19-0372.1.

Laser, J. L., M. C. Coniglio, P. S. Skinner, and E. N. Smith, 2022: Doppler lidar and mobile radiosonde observation-based evaluation of Warn-on-Forecast System predicted near-supercell environments during TORUS 2019. *Wea. Forecasting*, **37**, 1783–1804, https://doi.org/10.1175/WAF-D-21-0190.1.

Lawson, J. R., J. S. Kain, N. Yussouf, D. C. Dowell, D. M. Wheatley, K. H. Knopfmeier, and T. A. Jones, 2018: Advancing from convection-allowing NWP to Warn-on-Forecast: Evidence of progress. *Wea. Forecasting*, **33**, 599–607, https://doi.org/10.1175/WAF-D-17-0145.1.

——, 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, https://doi.org/10.1175/MWR-D-20-0281.1.

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, https://doi.org/10.15191/nwajom.2023.1106.

Loken, E. D., A. J. Clark, and C. D. Karstens, 2020: Generating probabilistic next-day severe weather forecasts from convection-allowing ensembles using random forests. *Wea. Forecasting*, **35**, 1605–1631, https://doi.org/10.1175/WAF-D-19-0258.1.

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

Mallick, S., and T. A. Jones, 2020: Assimilation of GOES-16 satellite derived winds into the Warn-on-Forecast System. *Atmos. Res.*, **245**, 105131, https://doi.org/10.1016/j.atmosres.2020.105131.

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, https://doi.org/10.1175/2009JAS2965.1.

——, D. T. Dawson II, and J. M. Straka, 2020: Bin-emulating hail melting in three-moment bulk microphysics. *J. Atmos. Sci.*, **77**, 3361–3385, https://doi.org/10.1175/JAS-D-19-0268.1.

Martinaitis, S. M., and Coauthors, 2023: A path toward short-term probabilistic flash flood prediction. *Bull. Amer. Meteor. Soc.*, **104**, E585–E605, https://doi.org/10.1175/BAMS-D-22-0026.1.

Matilla, B. C., N. Yussouf, P. S. Skinner, K. Knopfmeier, G. J. Creager, and G. P. McCabe, 2021: Implementation and assessment of METplus verification for the NSSL experimental Warn-on-Forecast System. *Special Symp. on Global and Mesoscale Models*, online, Amer. Meteor. Soc., 916, https://ams.confex.com/ams/101ANNUAL/meetingapp.cgi/Paper/379903.

McGovern, A., K. L. Elmore, D. J. Gagne II, S. E. Haupt, C. D. Karstens, R. Lagerquist, T. Smith, and J. K. Williams, 2017: Using artificial intelligence to improve real-time decision-making for high-impact weather. *Bull. Amer. Meteor. Soc.*, **98**, 2073–2090, https://doi.org/10.1175/BAMS-D-16-0123.1.

——, C. D. Karstens, T. Smith, and R. Lagerquist, 2019a: Quasi-operational testing of real-time-storm-longevity prediction via machine learning. *Wea. Forecasting*, **34**, 1437–1451, https://doi.org/10.1175/WAF-D-18-0141.1.

——, R. Lagerquist, D. J. Gagne II, G. E. Jergensen, K. L. Elmore, C. R. Homeyer, and T. Smith, 2019b: Making the black box more transparent: Understanding the physical implications of machine learning. *Bull. Amer. Meteor. Soc.*, **100**, 2175–2199, https://doi.org/10.1175/BAMS-D-18-0195.1.

——, R. J. Chase, M. Flora, D. J. Gagne II, R. Lagerquist, C. K. Potvin, N. Snook, and E. Loken, 2023: A review of machine learning for convective weather. *Artif. Intell. Earth Syst.*, **2**, e220077, https://doi.org/10.1175/AIES-D-22-0077.1.

Mecikalski, J. R., J. K. Williams, C. P. Jewett, D. Ahijevych, A. Leroy, and J. R. Walker, 2015: Probabilistic 0–1-h convective initiation nowcasts that combine geostationary satellite observations and numerical weather prediction model data. *J. Appl. Meteor. Climatol.*, **54**, 1039–1059, https://doi.org/10.1175/JAMC-D-14-0129.1.

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, https://doi.org/10.1175/WAF-D-21-0079.1.

Mittermaier, M., and N. Roberts, 2010: Intercomparison of spatial forecast verification methods: Identifying skillful spatial scales using the fractions skill score. *Wea. Forecasting*, **25**, 343–354, https://doi.org/10.1175/2009WAF2222260.1.

Mullendore, G. L., M. S. Mayernik, and D. C. Schuster, 2021: Open science expectations for simulation-based research. *Front. Climate*, **3**, 763420, https://doi.org/10.3389/fclim.2021.763420.

NCEI, 2021: Storm events database. NOAA, accessed 9 July 2021, https://www.ncdc.noaa.gov/stormevents/.

NRC, 1999: *A Vision for the National Weather Service: Road Map of the Future*. National Academies Press, 88 pp., https://doi.org/10.17226/6434.

——, 2003: *Communicating Uncertainties in Weather and Climate Information: A Workshop Summary*. National Academies Press, 68 pp., https://doi.org/10.17226/10597.

——, 2006: *Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts*. National Academies Press, 124 pp., https://doi.org/10.17226/11699.

——, 2012: *Weather Services for the Nation: Becoming Second to None*. National Academies Press, 86 pp., https://doi.org/10.17226/13429.

Novak, D. R., K. F. Brill, and W. A. Hogsett, 2014: Using percentiles to communicate snowfall uncertainty. *Wea. Forecasting*, **29**, 1259–1265, https://doi.org/10.1175/WAF-D-14-00019.1.

Orlanski, I., 1975: A rational subdivision of scales for atmospheric processes. *Bull. Amer. Meteor. Soc.*, **56**, 527–530, https://doi.org/10.1175/1520-0477-56.5.527.

Pan, S., and J. Gao, 2022: A method for assimilating pseudo dewpoint temperature as a function of GLM flash extent density in GSI-based EnKF data assimilation system—A proof of concept study. *Earth Space Sci.*, **9**, e2022EA002378, https://doi.org/10.1029/2022EA002378.

——, ——, D. J. Stensrud, X. Wang, and T. Jones, 2018: Assimilation of radar radial velocity and reflectivity, satellite cloud water path, and total precipitable water for convective-scale NWP in OSSEs. *J. Atmos. Oceanic Technol.*, **35**, 67–89, https://doi.org/10.1175/JTECH-D-17-0081.1.

——, ——, T. A. Jones, Y. Wang, X. Wang, and J. Li, 2021: The impact of assimilating satellite-derived layered precipitable water, cloud water path, and radar data on short-range thunderstorm forecasts. *Mon. Wea. Rev.*, **149**, 1359–1380, https://doi.org/10.1175/MWR-D-20-0040.1.

Potvin, C. K., and L. J. Wicker, 2013: Assessing ensemble forecasts of low-level supercell rotation within an OSSE framework. *Wea. Forecasting*, **28**, 940–960, https://doi.org/10.1175/WAF-D-12-00122.1.

——, 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, https://doi.org/10.1175/JAS-D-16-0098.1.

——, C. Broyles, P. S. Skinner, H. E. Brooks, and E. Rasmussen, 2019: A Bayesian hierarchical modeling framework for correcting reporting bias in the U.S. tornado database. *Wea. Forecasting*, **34**, 15–30, https://doi.org/10.1175/WAF-D-18-0137.1.

——, 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, https://doi.org/10.1175/MWR-D-19-0389.1.

——, C. Broyles, P. S. Skinner, and H. E. Brooks, 2022: Improving estimates of U.S. tornado frequency by accounting for unreported and underrated tornadoes. *J. Appl. Meteor. Climatol.*, **61**, 909–930, https://doi.org/10.1175/JAMC-D-21-0225.1.

——, M. Flora, P. Skinner, and A. Reinhart, 2023: Using machine learning to predict forecast skill in the NSSL Warn-on-Forecast System. *13th Conf. on Transition of Research to Operations*, Denver, CO, Amer. Meteor. Soc., 5.3, https://ams.confex.com/ams/103ANNUAL/meetingapp.cgi/Paper/412883.

Putnam, B. J., M. Xue, Y. Jung, N. Snook, and G. Zhang, 2014: The analysis and prediction of microphysical states and polarimetric radar variables in a mesoscale convective system using double-moment microphysics, multinetwork radar data, and the ensemble Kalman filter. *Mon. Wea. Rev.*, **142**, 141–162, https://doi.org/10.1175/MWR-D-13-00042.1.

——, ——, ——, G. Zhang, and F. Kong, 2017a: Simulation of polarimetric radar variables from 2013 CAPS Spring Experiment storm-scale ensemble forecasts and evaluation of microphysics schemes. *Mon. Wea. Rev.*, **145**, 49–73, https://doi.org/10.1175/MWR-D-15-0415.1.

——, ——, ——, N. Snook, and G. Zhang, 2017b: Ensemble probabilistic prediction of a mesoscale convective system and

associated polarimetric radar variables using single-moment and double-moment microphysical schemes and EnKF radar data assimilation. *Mon. Wea. Rev.*, **145**, 2257–2279, https://doi.org/10.1175/MWR-D-16-0162.1.

——, ——, ——, ——, and ——, 2019: Ensemble Kalman filter assimilation of polarimetric radar observations for the 20 May 2013 Oklahoma tornadic supercell case. *Mon. Wea. Rev.*, **147**, 2511–2533, https://doi.org/10.1175/MWR-D-18-0251.1.

——, Y. Jung, N. Yussouf, D. Stratman, T. A. Supinie, M. Xue, C. Kuster, and J. Labriola, 2021: The impact of assimilating Z_{DR} observations on storm-scale ensemble forecasts of the 31 May 2013 Oklahoma storm event. *Mon. Wea. Rev.*, **149**, 1919–1942, https://doi.org/10.1175/MWR-D-20-0261.1.

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

——, ——, B. T. Gallo, A. J. Clark, K. H. Knopfmeier, and P. S. Skinner, 2020: Comparison of the Warn-on-Forecast System and a High-Resolution Rapid Refresh time-lagged ensemble for forecasting short-term convective evolution. *10th Conf. on Transition of Research to Operations*, Boston, MA, Amer. Meteor. Soc., 1485, https://ams.confex.com/ams/2020Annual/meetingapp.cgi/Paper/368157.

Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, https://doi.org/10.1175/2007MWR2123.1.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608, https://doi.org/10.1175/2008WAF2222159.1.

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, https://doi.org/10.1175/BAMS-D-16-0100.1.

Rotunno, R., and J. B. Klemp, 1982: The influence of the shear-induced pressure gradient on thunderstorm motion. *Mon. Wea. Rev.*, **110**, 136–151, https://doi.org/10.1175/1520-0493(1982)110<0136:TIOTSI>2.0.CO;2.

Schwartz, C. S., and R. A. Sobash, 2017: Generating probabilistic forecasts from convection-allowing ensembles using neighborhood approaches: A review and recommendations. *Mon. Wea. Rev.*, **145**, 3397–3418, https://doi.org/10.1175/MWR-D-16-0400.1.

——, G. S. Romine, 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, https://doi.org/10.1175/WAF-D-15-0103.1.

Sivillo, J. K., J. E. Ahlquist, and Z. Toth, 1997: An ensemble forecasting primer. *Wea. Forecasting*, **12**, 809–818, https://doi.org/10.1175/1520-0434(1997)012<0809:AEFP>2.0.CO;2.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., https://doi.org/10.5065/D68S4MVH.

——, J. B. Klemp, M. G. Duda, L. D. Fowler, S.-H. Park, and T. D. Ringler, 2012: A multiscale nonhydrostatic atmospheric model using centroidal Voronoi tesselations and C-grid staggering. *Mon. Wea. Rev.*, **140**, 3090–3105, https://doi.org/10.1175/MWR-D-11-00215.1.

——, M. G. Duda, S. Ha, and S.-H. Park, 2018: Limited-area atmospheric modeling using an unstructured mesh. *Mon. Wea. Rev.*, **146**, 3445–3460, https://doi.org/10.1175/MWR-D-18-0155.1.

Skinner, P. S., L. J. Wicker, D. M. Wheatley, and K. H. Knopfmeier, 2016: Application of two spatial verification methods to ensemble forecasts of low-level rotation. *Wea. Forecasting*, **31**, 713–735, https://doi.org/10.1175/WAF-D-15-0129.1.

——, and Coauthors, 2018: Object-based verification of a prototype Warn-on-Forecast System. *Wea. Forecasting*, **33**, 1225–1250, https://doi.org/10.1175/WAF-D-18-0020.1.

——, and Coauthors, 2023: Interpreting Warn-on-Forecast System guidance. Part I: Review of probabilistic guidance concepts, product design, and best practices. *J. Oper. Meteor.*, in press.

Smith, T. M., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) severe weather and aviation products: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 1617–1630, https://doi.org/10.1175/BAMS-D-14-00173.1.

Sobash, R. A., and L. J. Wicker, 2015: On the impact of additive noise in storm-scale EnKF experiments. *Mon. Wea. Rev.*, **143**, 3067–3086, https://doi.org/10.1175/MWR-D-14-00323.1.

——, J. S. Kain, D. R. Bright, A. R. Dean, M. C. Coniglio, and S. J. Weiss, 2011: Probabilistic forecast guidance for severe thunderstorms based on the identification of extreme phenomena in convection-allowing model forecasts. *Wea. Forecasting*, **26**, 714–728, https://doi.org/10.1175/WAF-D-10-05046.1.

——, 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, https://doi.org/10.1175/WAF-D-15-0138.1.

——, G. S. Romine, and C. S. Schwartz, 2020: A comparison of neural-network and surrogate-severe probabilistic convective hazard guidance derived from a convection-allowing model. *Wea. Forecasting*, **35**, 1981–2000, https://doi.org/10.1175/WAF-D-20-0036.1.

Steinkruger, D., P. Markowski, and G. Young, 2020: An artificially intelligent system for the automated issuance of tornado warnings in simulated convective storms. *Wea. Forecasting*, **35**, 1939–1965, https://doi.org/10.1175/WAF-D-19-0249.1.

Stensrud, D. J., and J. Gao, 2010: Importance of horizontally inhomogeneous environmental initial conditions to ensemble storm-scale radar data assimilation and very short-range forecasts. *Mon. Wea. Rev.*, **138**, 1250–1272, https://doi.org/10.1175/2009MWR3027.1.

——, J.-W. Bao, and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.*, **128**, 2077–2107, https://doi.org/10.1175/1520-0493(2000)128<2077:UICAMP>2.0.CO;2.

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

Stratman, D. R., N. Yussouf, Y. Jung, T. A. Supinie, M. Xue, P. S. Skinner, and B. J. Putnam, 2020: Optimal temporal frequency of NSSL phased array radar observations for an experimental Warn-on-Forecast System. *Wea. Forecasting*, **35**, 193–214, https://doi.org/10.1175/WAF-D-19-0165.1.

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, https://doi.org/10.1175/MWR2898.1.

Trapp, R. J., D. M. Wheatley, N. T. Atkins, R. W. Przybylinski, and R. Wolf, 2006: Buyer beware: Some words of caution on the use of severe wind reports in postevent assessment and

research. *Wea. Forecasting*, **21**, 408–415, https://doi.org/10.1175/WAF925.1.

Trujillo-Falcón, J. E., J. Reedy, K. E. Klockow-McClain, K. L. Berry, G. J. Stumpf, A. V. Bates, and J. G. LaDue, 2022: Creating a communication framework for FACETs: How probabilistic hazard information affected warning operations in NOAA's Hazardous Weather Testbed. *Wea. Climate Soc.*, **14**, 881–892, https://doi.org/10.1175/WCAS-D-21-0136.1.

Uccellini, L. W., and J. E. Ten Hoeve, 2019: Evolving the National Weather Service to build a Weather-Ready Nation: Connecting observations, forecasts, and warnings to decision-makers through impact-based decision support services. *Bull. Amer. Meteor. Soc.*, **100**, 1923–1942, https://doi.org/10.1175/BAMS-D-18-0159.1.

Wang, Y., J. Gao, P. S. Skinner, K. Knopfmeier, T. A. Jones, G. Creager, P. L. Heinselman, and L. J. Wicker, 2019: Test of a weather-adaptive dual-resolution hybrid Warn-on-Forecast analysis and forecast system for several severe weather events. *Wea. Forecasting*, **34**, 1807–1827, https://doi.org/10.1175/WAF-D-19-0071.1.

——, N. Yussouf, E. R. Mansell, and B. C. Matilla, 2021: Impact of assimilating GOES-R geostationary lightning mapper flash extent density data on severe convection forecasts in a Warn-on-Forecast System. *Mon. Wea. Rev.*, **149**, 3217–3241, https://doi.org/10.1175/MWR-D-20-0406.1.

Weather Prediction Center, 2018: 2018 Flash Flood and Intense Rainfall Experiment final report. NOAA Tech. Rep., 97 pp., https://www.wpc.ncep.noaa.gov/hmt/2018_FFaIR_final_report.pdf.

Wheatley, D. M., K. H. Knopfmeier, T. A. Jones, and G. J. Creager, 2015: Storm-scale data assimilation and ensemble forecasting with the NSSL experimental Warn-on-Forecast System. Part I: Radar data experiments. *Wea. Forecasting*, **30**, 1795–1817, https://doi.org/10.1175/WAF-D-15-0043.1.

Whitaker, J. S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev.*, **136**, 463–482, https://doi.org/10.1175/2007MWR2018.1.

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, https://doi.org/10.1175/WAF-D-16-0157.1.

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

——, and Coauthors, 2019b: Exploring applications of storm-scale probabilistic Warn-on-Forecast guidance in weather forecasting. *HCII 2019: Virtual, Augmented and Mixed Reality. Applications and Case Studies*, J. Chen and G. Fragomeni, Eds., Lecture Notes in Computer Science, Vol. 11575, Springer, 557–572, https://doi.org/10.1007/978-3-030-21565-1_39.

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

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

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, https://doi.org/10.1175/WAF-D-13-00135.1.

Wu, W.-S., R. J. Purser, and D. F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905–2916, https://doi.org/10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2.

Yussouf, N., and D. J. Stensrud, 2010: Impact of phased-array radar observations over a short assimilation period: Observing system simulation experiments using an ensemble Kalman filter. *Mon. Wea. Rev.*, **138**, 517–538, https://doi.org/10.1175/2009MWR2925.1.

——, and K. H. Knopfmeier, 2019: Application of the Warn-on-Forecast System for flash-flood-producing heavy convective rainfall events. *Quart. J. Roy. Meteor. Soc.*, **145**, 2385–2403, https://doi.org/10.1002/qj.3568.

——, E. R. Mansell, L. J. Wicker, D. M. Wheatley, and D. J. Stensrud, 2013: The ensemble Kalman filter analyses and forecasts of the 8 May 2003 Oklahoma City tornadic supercell storm using single- and double-moment microphysics schemes. *Mon. Wea. Rev.*, **141**, 3388–3412, https://doi.org/10.1175/MWR-D-12-00237.1.

——, D. C. Dowell, L. J. Wicker, K. H. 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, https://doi.org/10.1175/MWR-D-14-00268.1.

——, 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, https://doi.org/10.1175/WAF-D-15-0160.1.

——, T. A. Jones, and P. S. Skinner, 2020a: Probabilistic high-impact rainfall forecasts from landfalling tropical cyclones using Warn-on-Forecast System. *Quart. J. Roy. Meteor. Soc.*, **146**, 2050–2065, https://doi.org/10.1002/qj.3779.

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

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253, https://doi.org/10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2.

Zhang, G., J. Gao, and M. Du, 2021: Parameterized forward operators for simulation and assimilation of polarimetric radar data with numerical weather predictions. *Adv. Atmos. Sci.*, **38**, 737–754, https://doi.org/10.1007/s00376-021-0289-6.

Zhang, M., L. Bernardet, G. Firl, D. Heinzeller, and W. Li, 2022: CCPP v6.0.0 Physics Scientific Documentation. Developmental Testbed Center, accessed 23 April 2023, https://dtcenter.ucar.edu/GMTB/v6.0.0/sci_doc/index.html.

Zhao, J., J. Gao, T. A. Jones, and J. Hu, 2021a: Impact of assimilating high-resolution atmospheric motion vectors on convective scale short-term forecasts: 1. Observing System Simulation Experiment (OSSE). *J. Adv. Model. Earth Syst.*, **13**, e2021MS002484, https://doi.org/10.1029/2021MS002484.

——, ——, ——, and ——, 2021b: Impact of assimilating high-resolution atmospheric motion vectors on convective scale short-term forecasts: 2. Assimilation experiments of GOES-16 satellite derived winds. *J. Adv. Model. Earth Syst.*, **13**, e2021MS002486, https://doi.org/10.1029/2021MS002486.

Zhou, X., and Coauthors, 2022: The development of the NCEP global ensemble forecast system version 12. *Wea. Forecasting*, **37**, 1069–1084, https://doi.org/10.1175/WAF-D-21-0112.1.

Zhu, K., Y. Pan, M. Xue, X. Wang, J. S. Whitaker, S. G. Benjamin, S. S. Weygandt, and M. Hu, 2013: A regional GSI-based ensemble Kalman filter data assimilation system for the Rapid Refresh configuration: Testing at reduced resolution. *Mon. Wea. Rev.*, **141**, 4118–4139, https://doi.org/10.1175/MWR-D-13-00039.1.

Ziegler, C. L., 1985: Retrieval of thermal and microphysical variables in observed convective storms. Part I: Model development and preliminary testing. *J. Atmos. Sci.*, **42**, 1487–1509, https://doi.org/10.1175/1520-0469(1985)042<1487:ROTAMV>2.0.CO;2.