

Breaking New Ground in Severe Weather Prediction: The 2015 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment

BURKELY T. GALLO,^a ADAM J. CLARK,^b ISRAEL JIRAK,^c JOHN S. KAIN,^b STEVEN J. WEISS,^c MICHAEL CONIGLIO,^b KENT KNOPFMEIER,^{d,b} JAMES CORREIA JR.,^{d,b} CHRISTOPHER J. MELICK,^e CHRISTOPHER D. KARSTENS,^{d,b} ESWAR IYER,^a ANDREW R. DEAN,^c MING XUE,^{a,f} FANYOU KONG,^f YOUNGSUN JUNG,^f FEIFEI SHEN,^f KEVIN W. THOMAS,^f KEITH BREWSTER,^f DEREK STRATMAN,^{a,f} GREGORY W. CARBIN,^g WILLIAM LINE,^{d,c} REBECCA ADAMS-SELIN,^h AND STEVE WILLINGTONⁱ

^a School of Meteorology, University of Oklahoma, Norman, Oklahoma

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

^c NOAA/Storm Prediction Center, Norman, Oklahoma

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

^e 557th Weather Wing/16th Weather Squadron, Offutt AFB, Nebraska

^f Center for Analysis and Prediction of Storms, Norman, Oklahoma

^g NOAA/Weather Prediction Center, College Park, Maryland

^h Atmospheric and Environmental Research, Inc., Lexington, Massachusetts

ⁱ Met Office, Exeter, Devon, United Kingdom

(Manuscript received 10 October 2016, in final form 31 May 2017)

ABSTRACT

Led by NOAA's Storm Prediction Center and National Severe Storms Laboratory, annual spring forecasting experiments (SFEs) in the Hazardous Weather Testbed test and evaluate cutting-edge technologies and concepts for improving severe weather prediction through intensive real-time forecasting and evaluation activities. Experimental forecast guidance is provided through collaborations with several U.S. government and academic institutions, as well as the Met Office. The purpose of this article is to summarize activities, insights, and preliminary findings from recent SFEs, emphasizing SFE 2015. Several innovative aspects of recent experiments are discussed, including the 1) use of convection-allowing model (CAM) ensembles with advanced ensemble data assimilation, 2) generation of severe weather outlooks valid at time periods shorter than those issued operationally (e.g., 1–4 h), 3) use of CAMs to issue outlooks beyond the day 1 period, 4) increased interaction through software allowing participants to create individual severe weather outlooks, and 5) tests of newly developed storm-attribute-based diagnostics for predicting tornadoes and hail size. Additionally, plans for future experiments will be discussed, including the creation of a Community Leveraged Unified Ensemble (CLUE) system, which will test various strategies for CAM ensemble design using carefully designed sets of ensemble members contributed by different agencies to drive evidence-based decision-making for near-future operational systems.

1. Introduction

Annual spring forecasting experiments (SFEs) conducted in the National Oceanic and Atmospheric Administration's (NOAA) Hazardous Weather Testbed (HWT) provide opportunities for testing new tools and techniques in forecasting severe thunderstorms. Jointly run by the National Severe Storms Laboratory (NSSL)

and the Storm Prediction Center (SPC), SFEs provide a two-way research-to-operations/operations-to-research pathway for enhanced understanding and problem-solving regarding severe thunderstorm forecasting. The real-time SFE takes place during the spring severe weather season, providing realistic operational pressure for participants as each day presents a unique set of conditions regarding severe weather potential.

Formal SFEs began in 2000; Kain et al. (2003) emphasize that collaboration is the crux of the SFEs,

Corresponding author: Burkely T. Gallo, burkely.twiest@noaa.gov

noting that “the interaction between forecasters and numerical modelers was the most rewarding part of (the) Spring Program.” This collaboration has created greater forecaster understanding of numerical models and greater researcher understanding of operational challenges (Kain et al. 2003). Clark et al. (2012a) further emphasize the SFE’s collaborative aspects, detailing the extension of severe thunderstorm forecasts issued during SFE 2010 to aviation and heavy-precipitation interests.

While SFEs involve real-time forecasting, daily evaluation exercises are another key aspect of SFEs (Clark et al. 2012a). Evaluating cutting-edge techniques such as experimental severe weather guidance derived from convection-allowing models (CAMs) allows participants to grasp the strengths and weaknesses of each technique and assess its readiness for operational adoption. Subjective evaluations illustrate the impressions participants have, while objective evaluations often take place after the SFEs, when time permits a thorough examination of the large volume of data (e.g., Johnson et al. 2013; Smith et al. 2014; Surcel et al. 2014; Duda et al. 2014).

Since 2007, the Center for Analysis and Prediction of Storms (CAPS) at the University of Oklahoma has provided a real-time continental United States (CONUS) forecast at 4-km grid spacing from a multi-model storm-scale ensemble forecast (SSEF) system to the SFE (Kong et al. 2015 and references therein). This system was reduced to 3-km grid spacing for SFE 2015. SFE 2015 also included five other unique CAM ensembles. Multiple organizations contributed numerical weather prediction (NWP) forecasts, including the Environmental Modeling Center (EMC), the Earth Science Research Laboratory’s Global Systems Division (ESRL/GSD), NSSL, CAPS, the National Center for Atmospheric Research (NCAR), and the 557th Weather Wing [formerly the U.S. Air Force Weather Agency (AFWA)]. Experimental deterministic guidance was also featured during SFE 2015, particularly three versions of the Unified Model (UM; Davies et al. 2005) from the Met Office and the Model for Prediction Across Scales (MPAS; Skamarock et al. 2012) from NCAR.

SFE 2015 pursued a number of goals consistent with the visions of both the Forecasting a Continuum of Environmental Threats (FACETs; Rothfus et al. 2014) and Warn-on-Forecast (WoF; Stensrud et al. 2009) initiatives. These programs aim to generate probabilistic hazard information (PHI), to go beyond the current binary paradigm of products such as watches, warnings, and advisories. Under a probabilistic paradigm, forecasters can give users more

specific, understandable information that they can assess to take action based on their individual needs. Developing probabilistic guidance to support this new paradigm requires cooperation between the operational forecasting and research communities, making the SFEs optimal for exploration of probabilistic forecasts. SFE 2015’s goals fall into two categories consistent with the visions of FACETs and WoF: 1) operational product and service improvements and 2) applied science activities. The operational product and service improvements goals focused on model guidance-driven forecast *generation* by participants, while applied science activities focused on the *evaluation* of new forecasting tools and forecast types, including new numerical guidance and postprocessing techniques. Numerical guidance characterization supported both types of goals by determining how to incorporate guidance into the forecasts and evaluating model output fields such as simulated reflectivity and hail-size estimates.

Introduced in SFE 2014 and continued in SFE 2015 is the incorporation of individual participant forecasts, essentially forming an “ensemble” of participant forecasts (Coniglio et al. 2014). Prior SFEs solely issued group forecasts, reaching a consensus on the placement of the day’s probability contours. While group discussion and consensus forming remained an integral part of the day 1 full period forecasting process, individuals then created higher time frequency forecasts. These forecasts tested the feasibility of operationally issuing more forecasts, each covering a shorter time window, and the subsequent increase in forecaster workload. Individuals’ forecasts also illustrated a variety of forecasting approaches, with differing reliance on observations, model guidance, and prior forecaster experience.

Also new to SFE 2015 are the evaluation capabilities of participants using laptops with Internet connectivity. Previously, evaluations were also consensus based. However, laptop usage enabled approximately five independent forecaster ratings per day for each evaluation. Although the SFE leaders had documented previous experiments’ discussions, enabling individuals to comment on products provided a more complete record of opinions, suggestions, and reflections on each product’s operational potential than in previous experiments.

This paper provides a broad overview of SFE 2015 and its innovations, which advance the two-way research-to-operations/operations-to-research pathway inherent to SFEs. Section 2a of this paper describes the NWP systems utilized throughout SFE 2015, and section 2b elaborates upon the daily activities of the

TABLE 1. NSSL-WRF ensemble specifications. All members use the WSM6 (Hong and Lim 2006), the MYJ (Mellor and Yamada 1982; Janjić 1994, 2002) PBL scheme, and the Noah (Chen and Dudhia 2001) LSM. For radiation, all members use the RRTM (Mlawer et al. 1997; Iacono et al. 2008) longwave radiation and Dudhia (Dudhia 1989) shortwave radiation schemes.

Ensemble member	No. of vertical levels	Initial conditions	Lateral boundary conditions	Microphysics	PBL
Cn	35	0000 UTC NAM	0000 UTC NAM	WSM6	MYJ
2	35	0000 UTC GFS	0000 UTC GFS	WSM6	MYJ
3	35	2100 UTC em_ctl	2100 UTC em_ctl	WSM6	MYJ
4	35	2100 UTC nmb_ctl	2100 UTC nmb_ctl	WSM6	MYJ
5	35	2100 UTC nmb_p1	2100 UTC nmb_p1	WSM6	MYJ
6	35	2100 UTC nmm_ctl	2100 UTC nmm_ctl	WSM6	MYJ
7	35	2100 UTC nmm_n1	2100 UTC nmm_n1	WSM6	MYJ
8	35	2100 UTC nmm_p1	2100 UTC nmm_p1	WSM6	MYJ
9	35	2100 UTC nmb_n1	2100 UTC nmb_n1	WSM6	MYJ
10	35	2100 UTC nmb_p2	2100 UTC nmb_p2	WSM6	MYJ

SFE. Section 3 highlights preliminary results from the SFE, including subjective and objective evaluations. Finally, section 4 provides a summary and evaluation of SFE 2015, along with plans and directions for future SFEs.

2. Experiment description

a. Experimental numerical guidance

SFE 2015 focused on experimental probabilistic forecast generation informed by a suite of experimental NWP forecasts. Four of the six experimental ensembles extended into the day 2 period, allowing for exploration of longer-range CAM forecasts. All models detailed below produced hourly maximum fields (Kain et al. 2010) of explicit storm attributes such as simulated reflectivity and updraft helicity (UH) for forecasting and evaluation purposes.

1) NSSL-WRF AND NSSL-WRF ENSEMBLE

Since the fall of 2006, SPC forecasters have used output from an experimental 4-km grid-spacing version of the Advanced Research version of the WRF Model (WRF-ARW; Skamarock et al. 2008) produced by the NSSL (Kain et al. 2010). Currently, this model runs twice daily at 0000 and 1200 UTC over a full-CONUS domain, with forecasts to 36 h [Table 1, ensemble member cn (control)]. Nine additional 4-km WRF-ARW members are run at 0000 UTC to 36 h by varying the initial conditions and lateral boundary conditions of the control to compose the 10-member NSSL-WRF ensemble (Table 1; Gallo et al. 2016). These members use the 0000 UTC National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) analysis or the 3-h Short-Range Ensemble Forecast (SREF; Du et al. 2014) system forecasts initialized at 2100 UTC for their initial conditions and corresponding GFS or SREF member forecasts as lateral boundary

conditions. The physics parameterizations among all members are identical.

2) CAPS STORM-SCALE ENSEMBLE FORECAST SYSTEMS

CAPS provided two ensembles to SFE 2015. The 20-member SSEF system included 12 members that accounted for as many sources of forecast error as possible (e.g., initial conditions, boundary conditions, multiphysics; Table 2). These members were used to generate probabilities of severe convective hazards. The eight remaining members tested physics sensitivities. WSR-88D data were used for data assimilation along with available surface and upper-air observations employing the Advanced Regional Prediction System (ARPS) three-dimensional variational data assimilation (3DVAR)/cloud-analysis system (Xue et al. 2003; Hu et al. 2006) to produce the control member. The 0000 UTC North American Mesoscale Forecast System (NAM) model analysis on a 12-km grid was used as a background for the analysis, and NAM forecasts provided boundary conditions. Perturbed members applied initial condition and boundary condition perturbations drawn from the SREF to the control analyses and forecasts. The CAPS forecasts were run with 3-km grid spacing and extended to 60 h, supporting day 2 forecasts.

A separate 12-member ensemble of 60-h forecasts was also produced on the same 3-km domain as the prior SSEF system (Table 3) using XSEDE supercomputing facilities (Towns et al. 2014). Rather than 3DVAR, the ensemble Kalman filter (EnKF; Evensen 1994, 2003) data assimilation (DA) method was used, specifically the CAPS EnKF DA system (Xue et al. 2006; Wang et al. 2013) that has been directly interfaced with the WRF Model. Specifically, 40-member ensemble forecasts were launched from NAM analysis plus SREF perturbations at 1800 UTC and run to 2300 UTC. The configuration of this ensemble involved both initial

TABLE 2. SSEF ensemble specifications. All members use RRTMG radiation schemes. Microphysics schemes used include Thompson (Thompson et al. 2004), Predicted Particle Properties (P3; Morrison and Milbrandt 2015), Milbrandt and Yau (M–Y; Milbrandt and Yau 2005), and Morrison (Morrison and Pinto 2005, 2006). Member 18 uses microphysics with two-category ice; all other P3 members use one-category ice. Planetary boundary layer schemes include Yonsei University (YSU; Hong et al. 2006), Thompson-modified YSU (YSU-T), and MYNN (Nakanishi and Niino 2004, 2006). Member 16 (Thompson ICLLOUD=3) accounts for the subgrid-scale clouds in the RRTMG radiation scheme based on research by G. Thompson. Italicized members compose the HWT baseline SSEF.

Ensemble member	No. of vertical levels	Initial conditions	Lateral boundary conditions	Microphysics	PBL
<i>cn</i>	51	0000 UTC ARPSa	0000 UTC NAMf	<i>Thompson</i>	<i>MYJ</i>
<i>c0</i>	51	0000 UTC ARPSa	0000 UTC NAMf	Thompson	MYJ
<i>m3</i>	51	<i>cn + nmmb-p2_pert</i>	<i>2100 UTC SREF nmmb-p2</i>	<i>P3</i>	<i>MYNN</i>
<i>m4</i>	51	<i>cn + nmmb-n2_pert</i>	<i>2100 UTC SREF nmmb-n2</i>	<i>M–Y</i>	<i>YSU</i>
<i>m5</i>	51	<i>cn + nmm-p1_pert</i>	<i>2100 UTC SREF nmm-p1</i>	<i>Morrison</i>	<i>MYNN</i>
<i>m6</i>	51	<i>cn + nmmb-n1_pert</i>	<i>2100 UTC SREF nmmb-n1</i>	<i>M–Y</i>	<i>MYJ</i>
<i>m7</i>	51	<i>cn + nmmb-p1_pert</i>	<i>2100 UTC SREF nmmb-p1</i>	<i>P3</i>	<i>YSU</i>
<i>m8</i>	51	<i>cn + em-n1_pert</i>	<i>2100 UTC SREF em-n1</i>	<i>P3</i>	<i>MYJ</i>
<i>m9</i>	51	<i>cn + em-p2_pert</i>	<i>2100 UTC SREF em-p2</i>	<i>M–Y</i>	<i>MYNN</i>
<i>m10</i>	51	<i>cn + nmmb-n3_pert</i>	<i>2100 UTC SREF nmmb-n3</i>	<i>Morrison</i>	<i>YSU</i>
<i>m11</i>	51	<i>cn + nmmb-p3_pert</i>	<i>2100 UTC SREF nmmb-p3</i>	<i>Thompson</i>	<i>YSU</i>
<i>m12</i>	51	<i>cn + nmm-n3_pert</i>	<i>2100 UTC SREF nmm-n3</i>	<i>Thompson</i>	<i>MYNN</i>
<i>m13</i>	51	<i>cn + nmm-p2_pert</i>	<i>2100 UTC SREF nmm-p2</i>	<i>Morrison</i>	<i>MJ</i>
<i>m14</i>	51	0000 UTC ARPSa	0000 UTC NAMf	Thompson	MYNN
<i>m15</i>	51	0000 UTC ARPSa	0000 UTC NAMf	Thompson	YSU-T
<i>m16</i>	51	0000 UTC ARPSa	0000 UTC NAMf	Thompson ICLLOUD=3	YSU-T
<i>m17</i>	51	0000 UTC ARPSa	0000 UTC NAMf	M–Y	MYJ
<i>m18</i>	51	0000 UTC ARPSa	0000 UTC NAMf	P3-cat2	MYJ
<i>m19</i>	51	0000 UTC ARPSa	0000 UTC NAMf	P3	MYJ
<i>m20</i>	51	0000 UTC ARPSa	0000 UTC NAMf	Morrison	MYJ

perturbations and mixed-physics options, to provide a variety of inputs for the EnKF analysis. Each member used the WRF single-moment 6-class (WSM6; Hong and Lim 2006) microphysics scheme with different intercept parameter settings for rain and graupel and the density of graupel, and included relatively small random perturbations (0.5 K for potential temperature and 5% for relative humidity) with recursive filtering of approximately 20-km horizontal correlations scales. EnKF cycling utilizing radar data was performed every 15 min from 2300 to 0000 UTC, using the 40-member ensemble

as background. In addition to radar data, only Meteorological Assimilation Data Ingest System (MADIS; Miller et al. 2005, 2007) surface observations, profiler, and radiosondes were assimilated at 2300 and/or 0000 UTC. A 12-member ensemble forecast to 60 h followed, using the last EnKF analyses at 0000 UTC (Table 3).

3) SPC STORM-SCALE ENSEMBLE OF OPPORTUNITY

The Storm-Scale Ensemble of Opportunity (SSEO; Jirak et al. 2012b) is a seven-member, multimodel,

TABLE 3. SSEF EnKF ensemble specifications.

Ensemble member	No. of vertical levels	Initial conditions	Lateral boundary conditions	Microphysics	PBL
enkf_cn	51	enk_m1a	0000 UTC NAMf	Thompson	MYJ
enkf_m6	51	enk_m2a	2100 UTC SREF nmmb-n1	M–Y	MYJ
enkf_m9	51	enk_m6a	2100 UTC SREF em-p2	M–Y	MYNN
enkf_m10	51	enk_m8a	2100 UTC SREF nmmb-n3	Morrison	YSU
enkf_m5	51	enk_m10a	2100 UTC SREF nmm-p1	Morrison	MYNN
enkf_m4	51	enk_m12a	2100 UTC SREF nmmb-n2	M–Y	YSU
enkf_m3	51	enk_m17a	2100 UTC SREF nmmb-p2	P3	MYNN
enkf_m8	51	enk_m23a	2100 UTC SREF em-n1	P3	MYJ
enkf_m7	51	enk_m26a	2100 UTC SREF nmmb-p1	P3	YSU
enkf_m12	51	enk_m37a	2100 UTC SREF nmm-n3	Thompson	MYNN
enkf_m11	51	enk_m39a	2100 UTC SREF nmmb-p3	Thompson	YSU
enkf_mn_thom	51	enfamean_thom	0000 UTC NAMf	Thompson	MYJ
enkf_mn_wsm6	51	enfamean_wdm6	0000 UTC NAMf	WSM6	MYJ
enkf_3dvar_thom	51	3dvar_thom	0000 UTC NAMf	Thompson	MYJ
enkf_3dvar_wsm6	51	3dvar_wdm6	0000 UTC NAMf	WSM6	MYJ

TABLE 4. SSEO specifications as of 12 Aug 2014.

Ensemble member	No. of vertical levels	Initial conditions	Lateral boundary conditions	Microphysics	PBL	Grid spacing (km)
NSSL WRF-ARW	35	NAM	NAM	WSM6	MYJ	4
EMC HRW WRF-ARW	40	RAP	GFS	WSM6	YSU	4.2
EMC HRW WRF-ARW; 12-h time lag	40	RAP	GFS	WSM6	YSU	4.2
EMC HRW NMMB	40	RAP	GFS	Ferrier updated	MYJ	3.6
EMC HRW NMMB; 12-h time lag	40	RAP	GFS	Ferrier updated	MYJ	3.6
EMC CONUS WRF-NMM	35	NAM	NAM	Ferrier	MYJ	4
EMC CONUS NAM NEST	60	NAM	NAM	Ferrier–Aligo	MYJ	4

multiphysics ensemble consisting of deterministic CAMs that is available year-round to the SPC (Table 4). Individual members include one model produced by NSSL, and six members produced by EMC. The ensemble has been utilized in SPC operations since 2011 as a practical alternative to a formal storm-scale ensemble (Jirak et al. 2012b), which is planned for implementation in the next few years (G. Dimego 2016, personal communication). Forecasts are initialized from the operational NAM with no additional data assimilation and are generated twice daily to 36 h, starting at 0000 and 1200 UTC. These members differ slightly in their grid spacing (3.6–4.2 km), vertical levels, and length, with 36-, 48-, and 60-h forecasts. Microphysics schemes of the members include WSM6, Ferrier (Ferrier 1994), and Ferrier–Aligo (Aligo et al. 2014).

4) U.S. AIR FORCE WEATHER AGENCY 4-KM ENSEMBLE

The U.S. Air Force 557th Weather Wing at Offutt Air Force Base (USAF), Nebraska, ran a real-time 10-member, 4-km WRF-ARW model ensemble (AFWA; Kuchera et al. 2014) over the CONUS for SFE 2015 to 60 h (Table 5). Forecasts were initialized twice daily, at 0000 and 1200 UTC, using 6- or 12-h forecasts from three

global models: the Met Office UM, the NCEP GFS, and the Canadian Meteorological Center Global Environmental Multiscale (GEM) model. Member microphysics and boundary layer parameterizations varied, and no data assimilation was performed during initialization.

5) NCAR ENKF-BASED ENSEMBLE

In SFE 2015, NCAR provided a new 10-member, 3-km grid-spacing ensemble with a CONUS domain (Schwartz et al. 2015). EnKF data assimilation occurred every 6 h with 15-km grid spacing using the following observational sources: the Aircraft Communications Addressing and Reporting System (ACARS), MADIS surface observations, METARs and radiosondes, NCEP marine data (MARINE), Cooperative Institute for Meteorological Satellite Studies (CIMSS) cloud-track winds (Menzel 2001), and the Oklahoma Mesonet stations. From this mesoscale background, 10 downscaled 3-km forecasts were initialized daily at 0000 UTC using consistent physics with the data assimilation system, sans cumulus parameterization. The first 10 members of the analysis were selected after random shuffling between analyses and, therefore, differed daily. Each selection of 10 members was equally representative of the ensemble mean analysis and perturbations, and unique

TABLE 5. AFWA ensemble specifications. Land initial conditions for each member are from the NASA Goddard Space Flight Center Land Information System (LIS; Kumar et al. 2006, 2008). The PBL schemes include the BouLac approach (Bougeault and Lacarrère 1989) and the updated asymmetric convective model (ACM2). Members use either the Noah or the Rapid Update Cycle (RUC; Smirnova et al. 1997, 2000, 2015) LSM. Microphysics schemes include the WRF double-moment microphysics (WDM6; Lim and Hong 2010) and the WRF single-moment 5-class microphysics (WSM5; Hong et al. 2004).

Ensemble member	No. of vertical levels	Initial conditions	Lateral boundary conditions	Microphysics	PBL
1	27	UM	UM	WSM5	YSU
2	27	GFS	GFS	Morrison	BouLac
3	24	GEM	GEM	WDM6	YSU
4	21	GEM	GEM	Ferrier	BouLac
5	21	UM	UM	WDM6	ACM2
6	24	GFS	GFS	Thompson	ACM2
7	24	GEM	GEM	Morrison	YSU
8	24	GFS	GFS	Ferrier	YSU
9	27	UM	UM	Thompson	ACM2
10	21	GFS	GFS	WSM5	ACM2

lateral boundary condition perturbations were member dependent but used random draws from global background error covariances. (Schwartz et al. 2014). Both the data assimilation scheme and the forecasts used Thompson microphysics (Thompson et al. 2008), the Rapid Radiative Transfer Model (RRTM; Mlawer et al. 1997) for global climate models (RRTMG; Iacono et al. 2008), Mellor–Yamada–Janjić (MYJ; Mellor and Yamada 1982; Janjić 1994, 2002) planetary boundary layer (PBL) parameterization, and the Noah land surface model (LSM) (Chen and Dudhia 2001). The analysis system contained 50 members of constant physics that were continuously cycled using the ensemble adjustment Kalman filter (EAKF; Anderson 2001, 2003) within NCAR’s Data Assimilation Research Testbed (DART; Anderson et al. 2009) software. The analyses provided initial conditions for the daily forecasts, which were run to 48 h. Both the analyses and the forecasts had 40 vertical levels.

6) UKMET CONVECTION-ALLOWING MODEL RUNS

The Met Office provided three nested, limited-area high-resolution versions of the UM to SFE 2015: two at 2.2-km grid spacing and one at 1.1-km grid spacing. The operational 2.2-km version incorporated the UM specifications currently run in the Met Office’s operational 1.5-km grid length, U.K.-centered model (McBeath et al. 2014; Mittermaier 2014). The operational 2.2-km version provided for SFE 2015 had 70 vertical levels across a domain ranging from just west of the Rocky Mountains to the western border of Maine. Initial and lateral boundary conditions were taken from the 0000 UTC 17-km global version of the UM without additional data assimilation, and forecasts extended to 48 h.

A unique aspect of the UM models was the configuration of the turbulence parameterization. The operational run used a 3D turbulent mixing scheme consisting of a locally scale-dependent blending of Smagorinsky (Smagorinsky 1963) and boundary layer mixing schemes, wherein stochastic perturbations were made to the low-level resolved-scale temperature field in conditionally unstable regimes to encourage the transition from subgrid- to resolved-scale flows (Clark et al. 2015). This turbulent mixing scheme differs from that of WRF, which utilizes 3D Smagorinsky turbulence closure to determine eddy viscosities in the absence of a PBL scheme (Skamarock et al. 2008). The operational 2.2-km run had single-moment microphysics (Wilson and Ballard 1999) and diagnosed partial cloudiness assuming a triangular moisture distribution whose width is a function of height only.

The parallel version of the 2.2-km UM used an experimental parameterization of partial cloudiness, expanding upon the prognostic scheme used in the Met Office global UM. The parallel scheme includes an additional parameterization of subgrid moisture variability linked to the boundary layer turbulence. This version was also run to 48 h and was otherwise identical to the operational 2.2-km version of the UM.

Finally, the 1.1-km horizontal resolution UM centered on Oklahoma ran over a 1300 km × 1800 km domain nested within the 2.2-km model. The initial and lateral boundary conditions were taken from hour 3 of the 0000 UTC 2.2-km run to reduce spinup time and run to 33 h. The 1.1-km run was otherwise identical to the 2.2-km operational run, thereby testing the horizontal resolution effects.

7) MODEL FOR PREDICTION ACROSS SCALES

Another new deterministic modeling system provided to SFE 2015 was the MPAS, which produced daily 0000 UTC initialized forecasts at 3-km grid spacing over the CONUS. Forecasts from MPAS extended to 120 h (5 days), allowing for a unique glimpse into the long-range capabilities of convection-allowing models. The MPAS horizontal mesh is based on spherical centroidal Voronoi tessellations (SCVTs; Satoh et al. 2008), allowing for quasi-uniform discretization of the sphere and local refinement with smoothly varying mesh spacing between regions with differing resolutions. The smoothly varying mesh eliminates major problems regarding transitions between the differing resolutions of nests (Skamarock et al. 2012). MPAS has 55 vertical levels, and the “scale aware” physics allows for the output of explicit storm attributes for those regions at convection-allowing resolution. Physics parameterizations include the MYJ PBL scheme and the WSM6 microphysics.

8) PARALLEL OPERATIONAL CAMS

During SFE 2015, SPC had access to parallel versions of NAM and the High-Resolution Rapid Refresh (HRRR; Alexander et al. 2010), which contains improvements over the operational versions of these models (Table 6). The parallel versions were candidates for operational implementation by NCEP. Parallel high-resolution window (HRW) WRF-ARW and Non-hydrostatic Multiscale Model on the B grid (NMMB; Janjić and Gall 2012) runs included slight changes such as increasing the number of vertical levels from 40 to 50, updating the WRF version used, and modifying the microphysics scheme in the WRF-ARW to decrease the amount of falling graupel. The parallel HRRR included changes in the physics to improve an afternoon warm,

TABLE 6. Parallel operational (op) CAM specifications. Initial and lateral boundaries include the GFS and the Rapid Refresh (RAP) model (Benjamin et al. 2016).

Model	No. of vertical levels	Initial conditions	Lateral boundary conditions	Microphysics	PBL
ESRL/GSD HRRR (op)	51	RAP version 2 (v2)	RAP v2	Modified Thompson	MYNN
ESRL/GSD HRRR (parallel)	51	RAP v3	RAP v3	Thompson–Eidhammer	Modified MYNN
EMC HRW WRF-ARW (op)	40	RAP v3	GFS	WSM6	YSU
EMC HRW WRF-ARW (parallel)	50	RAP v3	GFS	Modified WSM6	YSU
EMC HRW WRF-NMMB (op)	40	RAP v3	GFS	Ferrier updated	MYJ
EMC HRW WRF-NMMB (parallel)	50	RAP v3	GFS	Ferrier updated	MYJ
EMC CONUS NAM Nest (op)	60	GFS	NAM	Ferrier–Aligo	MYJ
EMC CONUS NAM Nest (parallel)	60	GFS	NAM	Ferrier–Aligo	MYJ

dry bias in the operational HRRR that had resulted in the overprediction of convective initiation (Alexander et al. 2015). These changes included updating the microphysics to the Thompson–Eidhammer scheme (Thompson and Eidhammer 2014) and modifying the Mellor–Yamada–Nakanishi–Niino (MYNN) PBL scheme. Changes to the NAM nest included reducing the grid spacing to 3 km in the parallel version, as opposed to the 4-km operational version. The parent NAM providing the boundary conditions was also updated.

b. Daily activities

SFE 2015 was conducted weekdays from 4 May through 5 June 2015, excepting the Memorial Day holiday on 25 May 2015, for a total of 24 days. Each day, participants completed the same activities, separated broadly into experimental forecasts and evaluations.

1) EXPERIMENTAL FORECASTS

Daily activities were split between two “desks,” led by SPC forecasters. Each desk focused on different experimental forecasts and evaluations, and participants rotated through desks during the week to gain exposure to all of the experimental products. Besides generating forecasts, participants at each desk evaluated prior forecasts and experimental numerical guidance. Activities took place at roughly the same time each day (Table 7) and mainly occurred over regions of the United States that had the greatest potential for severe weather during a given day.

Participants at the “individual hazards” desk issued daily probabilistic forecasts of severe hail, damaging wind, and tornadoes within 25 mi (40 km) of a point, consistent with the SPC’s definition of a severe convective hazard, valid from 1600 to 1200 UTC the following day. Meanwhile, participants at the “total severe” desk forecasted the risk of any severe hazard following the SPC’s operational day 2 convective outlook format, valid over the day 1 time period.

Participants at both desks then refined their day 1 forecasts into higher temporal resolution forecasts, with the individual hazards desk issuing hail, wind, and tornado forecasts for two 4-h periods: 1800–2200 and 2200–0200 UTC. Individual hazard forecasters could use temporally disaggregated first-guess probabilities generated from the full-period hazard outlook to constrain and scale the magnitude and spatial extent of the SSEO neighborhood probabilities of proxy variables (i.e., UH for tornadoes, updraft speed for hail, and 10-m wind speed for wind), ensuring consistency among the 24- and 4-h forecasts (Jirak et al. 2012a).

At the total severe desk, probabilistic forecasts were manually stratified by participants and the desk lead forecaster into 1-h periods valid starting at 1800–0000 UTC. The 2100–0000 UTC forecasts were updated each afternoon, with two additional hourly forecasts issued from 0100 to 0200 UTC and from 0200 to 0300 UTC. This approach was first attempted in 2014 and was continued in 2015. Reliability diagrams computed after the SFE 2014, when hourly forecasts were issued from 1800 to 0300 UTC (Coniglio et al. 2014; Fig. 1), showed that when verified on a 40-km grid (~20-km neighborhood), participants and the desk lead forecaster issued reliable hourly probabilistic forecasts, but overforecasted severe weather when verified on a 20-km grid (~10-km neighborhood). These hourly forecasts were verified by gridding local storm reports (LSRs) and grid points of the NSSL’s Multi-Radar/Multi-Sensor maximum estimated size of hail (MESH; Witt et al. 1998) ≥ 29 mm (following Cintineo et al. 2012), aggregated over the nine hourly periods initially forecast (Fig. 1a) and the six afternoon update hours (Fig. 1b).

Hourly probabilistic forecasts were tested with the goal of introducing probabilistic severe weather forecasts on time scales that are currently addressed only as needed operationally (e.g., severe thunderstorm/tornado watches). Breaking down a full-period outlook into hourly probabilities also tested seamlessly merging

TABLE 7. Daily activity schedule during the SFE in local time [central daylight time (CDT)].

Individual hazards desk	Total severe desk
0800–0845 CDT: Evaluation of experimental forecasts and guidance Subjective rating relative to radar evolution/characteristics, warnings, and preliminary reports, and objective verification using preliminary reports and MESH	
<ul style="list-style-type: none"> • Day 1 and 2 full-period probabilistic forecasts of tornado, wind, and hail • Day 1 four-hour period forecasts and guidance for tornado, wind, and hail 	<ul style="list-style-type: none"> • Days 1–3 full-period probabilistic forecast of total severe • Day 1 one-hour period forecasts and guidance for total severe
0845–1115 CDT: Day 1 convective outlook generation Hand analysis of 1200 UTC upper-air maps and surface charts	
<ul style="list-style-type: none"> • Day 1 full-period probabilistic forecasts of tornado, wind, and hail valid 1600–1200 UTC over the mesoscale area of interest • Day 1 four-hour probabilistic forecasts of tornado, wind, and hail valid 1800–2200 and 2200–0200 UTC^a 	<ul style="list-style-type: none"> • Day 1 full-period probabilistic forecast of total severe valid 1600–1200 UTC over the mesoscale area of interest • Day 1 one-hour probabilistic forecasts of total severe valid 1800–0000 UTC^a
1115–1130 CDT: Break Prepare for map discussion and discuss relationship/translation from probabilities to watch	
1130–1200 CDT: Map discussion Overview and discussion of today’s forecast challenges and products Highlight interesting findings from previous days	
1200–1300 CDT: Lunch Brief EWP participants at 1245	
1300–1400 CDT: Day 2 convective outlook generation	
<ul style="list-style-type: none"> • Day 2 full-period probabilistic forecasts of tornado, wind, and hail valid 1200–1200 UTC over mesoscale area of interest 	<ul style="list-style-type: none"> • Day 2 or 3 full-period probabilistic forecasts of total severe valid 1200–1200 UTC over mesoscale area of interest
1400–1500 CDT: Scientific evaluations	
<ul style="list-style-type: none"> • Convection-allowing ensemble comparison (reflectivity and hourly maximum fields): SSEO, AFWA, NSSL, SSEF, SSEF EnKF, and NCAR EnKF • EMC parallel CAM comparison (reflectivity): NAM nest, High-Resolution Window (HiResW), and HRRR 	<ul style="list-style-type: none"> • Met Office CAMs: vertical resolution • SSEF 3DVar vs EnKF comparison: impact on first few hours of control forecast • Model forecasts of explicit hail size: HAILCAST and Thompson • MPAS
1500–1600 CDT: Short-term outlook	
<ul style="list-style-type: none"> • Update 4-h probabilistic forecasts of tornado, wind, and hail valid 2200–0200 UTC^a • Generate 1-h probabilistic forecasts of tornado valid 2200–0200 UTC 	<ul style="list-style-type: none"> • Update and generate 1-h probabilistic forecasts of total severe valid 2100–0200 UTC^a

^a Forecast that is also made by participants using the PHI tool on laptops.

probabilistic severe weather outlooks to probabilistic severe weather warnings, consistent with the visions of FACETs (Rothfus *et al.* 2014) and WoF (Stensrud *et al.* 2009).

All participants individually generated the hourly forecasts using a web-based PHI tool (Karstens *et al.* 2014, 2015) to draw hazard probability contours (Fig. 2). Five laptops were available at each desk. If there were more participants than laptops at a desk, some participants worked in pairs to generate forecasts. Individual forecasts (Figs. 2a–e) were later compared with those issued by the desk lead (Fig. 2f). Participants subjectively evaluated the previous day’s short-term

forecast issued by the desk lead on a 1–10 scale, with 10 being the highest rating, compared to a “practically perfect” forecast (Brooks *et al.* 1998; Hitchens *et al.* 2013), which is analogous to probabilities a forecaster would issue with prior perfect knowledge of the LSR distribution (Fig. 2g). While preliminary LSRs provided the largest component of ground truth, MESH, watches, warnings, and observed composite reflectivity were also considered.

The individual hazard desk’s day 2 outlooks explored the feasibility of issuing individual hazard forecasts beyond day 1, utilizing experimental extended CAM guidance. Currently, individual hazard forecasts are

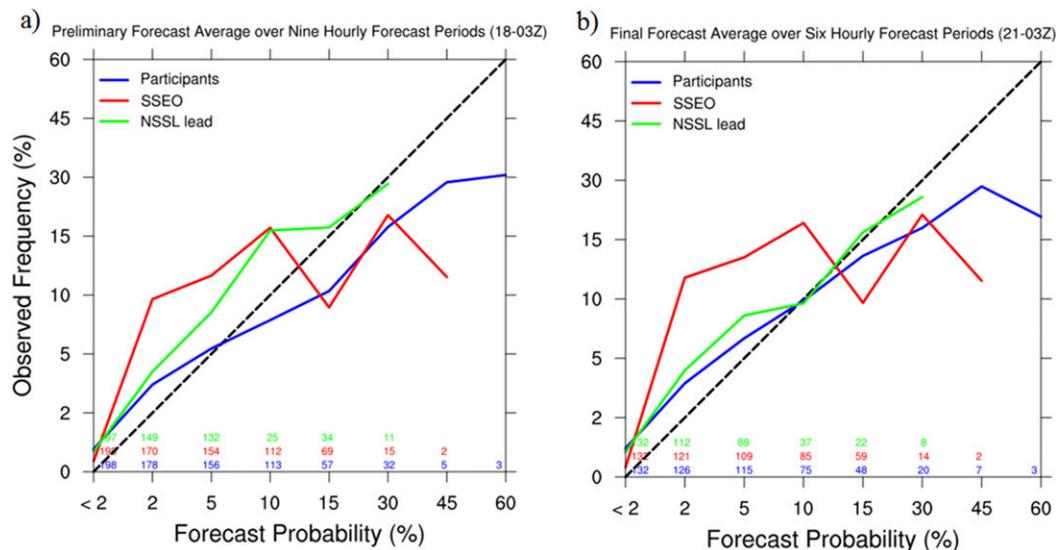


FIG. 1. Reliability diagrams generated for SFE 2014 hourly probabilistic forecasts for (a) the nine initial hourly forecasts and (b) the six afternoon updates. The black dashed line indicates perfect reliability, and the colored numbers along the x axis correspond to the number of forecasts with at least one forecast of that probability magnitude. [From Coniglio et al. (2014).]

limited to day 1 in SPC operations. The total severe desk also generated day 2 forecasts, as is done operationally by SPC, but informed by experimental CAM guidance. Day 3 forecasts were occasionally issued by the total severe desk, depending on time constraints and the anticipated severity of day 3. MPAS often heavily informed these extended forecasts, particularly because two prior runs encompassed a day 3 outlook, allowing consideration of run-to-run consistency.

The final forecasting activity of each day was an update to the earlier, participant-drawn forecasts informed by group discussion and updated data. Individual hazard participants updated their 2200–0200 UTC period, and the total severe participants updated their hourly forecasts from 2100 to 0000 UTC. Total severe participants also issued new hourly probabilities for 0000–0200 UTC.

While issuing forecasts, participants had access to high temporal resolution satellite imagery. The 1-min visible and infrared satellite imagery from *GOES-14* was made available experimentally to participants during SFE 2015 from 18 May to 11 June. This special 1-min imagery, known as Super Rapid Scan Operations for GOES-R (SRSOR), helps to prepare users for the very high temporal resolution sampling capability of the GOES-R Advanced Baseline Imagery (Line et al. 2016). SFE 2015 participants primarily utilized the 1-min satellite imagery to identify and track boundaries, assess cumulus cloud trends, and diagnose areas of convective initiation.

2) EVALUATIONS

In addition to forecasting activities, each participant performed multiple evaluations of the previous day's forecasts and model guidance. Participants rated the desk lead's forecasts and numerical guidance on a scale from 1 (very poor) to 10 (very good) and commented on particular strengths and weaknesses. This evaluation subjectively assessed the skill of the first-guess guidance and the human-generated forecasts for all periods (i.e., each hourly forecast at the total severe desk was assigned a rating). Model evaluations focused on the accuracy of the forecasts in predicting severe convective threats (including considerations such as the mode and timing of convective initiation) by comparing forecasts of hourly maximum fields (i.e., UH) relative to LSRs, maximum MESH, and radar observations across the previous day's domain. For ensembles extending to the day 2 period, participants compared day 2 guidance to day 1 guidance, to examine if the ensembles improved with shorter lead times.

New experimental fields were also evaluated, such as hail guidance available in the WRF-ARW (Adams-Selin et al. 2014), tornado probabilities generated from the NSSL-WRF ensemble (Gallo et al. 2016), and pre-convective, model-generated environmental soundings from the UM and the NSSL-WRF. In SFE 2015, the WRF-HAILCAST algorithm was implemented in the CAPS ensembles to predict hail size (Adams-Selin and

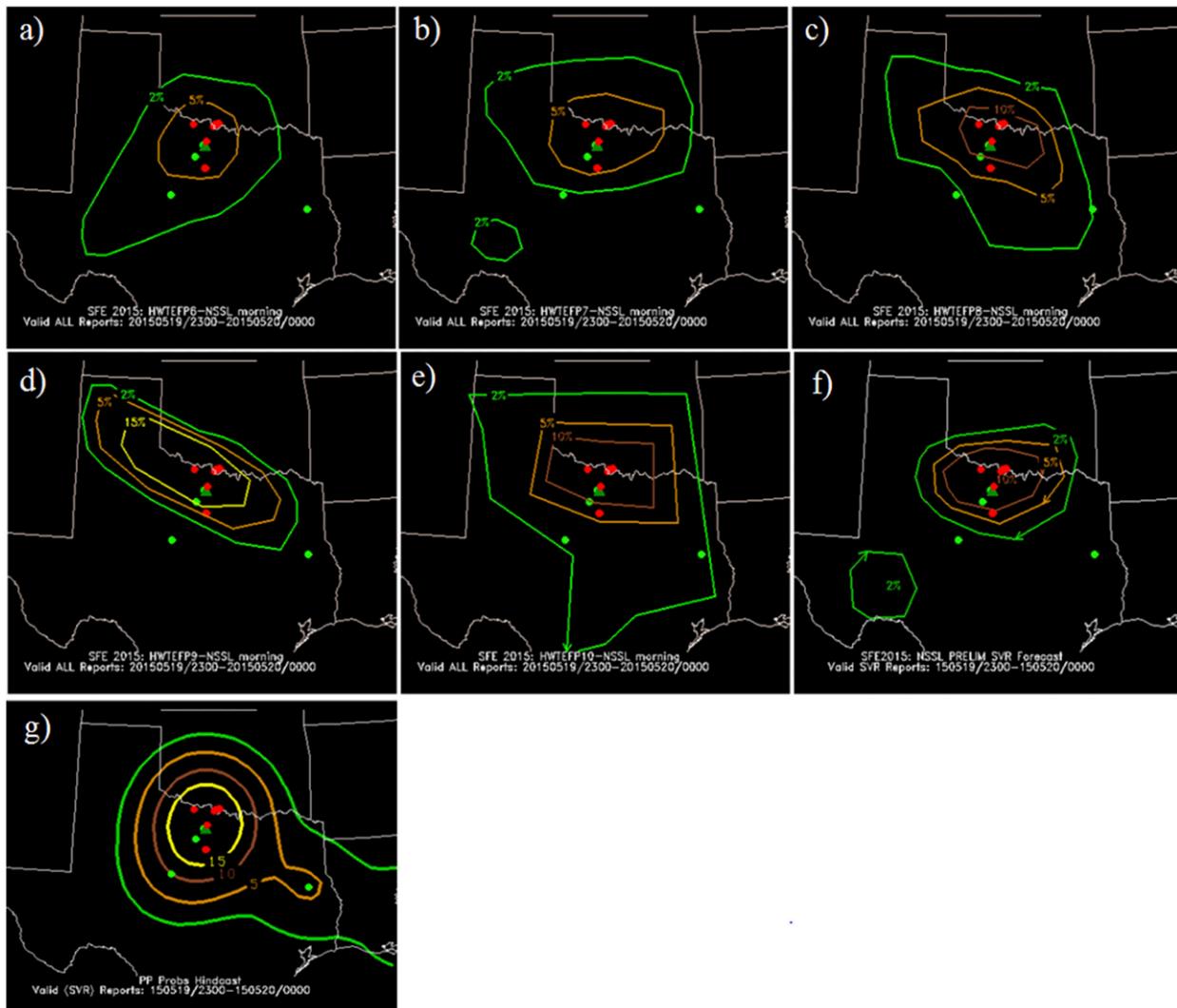


FIG. 2. (a)–(e) Five participant forecasts, (f) one SPC forecaster forecast, and (g) the practically perfect forecast valid 2300 UTC 19 May–0000 UTC 20 May 2015. Probabilistic contours indicate the likelihood of any type of severe weather (tornado, wind, or hail) during the forecast period. Overlaid red dots are tornado LSRs, green dots are hail LSRs, and dark green triangles are significant hail (hail diameter ≥ 2 in.) LSRs.

Ziegler 2016). This algorithm is a modified version of the coupled cloud and hail model found in Brimelow et al. (2002) and Jewell and Brimelow (2009), which forecasts the maximum expected hail diameter at the surface using a profile of nearby atmospheric temperature, moisture, and winds. The WRF-HAILCAST model uses WRF-generated convective cloud and updraft attributes coupled with a physical model of hail growth to determine hail growth from five predetermined initial embryo sizes. Another hail size diagnostic, derived directly from the microphysical parameterizations and developed by G. Thompson (Skamarock et al. 2008), was new to SFE 2015 and was output by the NCAR ensemble.

During SFE 2015, probabilistic tornado forecasts were generated from the NSSL-WRF ensemble using 2–5-km $UH \geq 75 \text{ m}^2 \text{ s}^{-2}$ as a proxy for tornadoes, with varying environmental constraints on probability generation. The environmental constraints required the probabilities to reflect UH only at grid points where certain environmental criteria were met in the previous hour: lifted condensation level < 1500 m, ratio of surface-based CAPE to most unstable CAPE ≥ 0.75 (Clark et al. 2012b), and significant tornado parameter (STP) ≥ 1 (Thompson et al. 2003). Gallo et al. (2016) elaborate on the probability generation details. Tornado reports from the LSR database were overlaid on the forecast probabilities for subjective evaluation, which considered the entire CONUS.

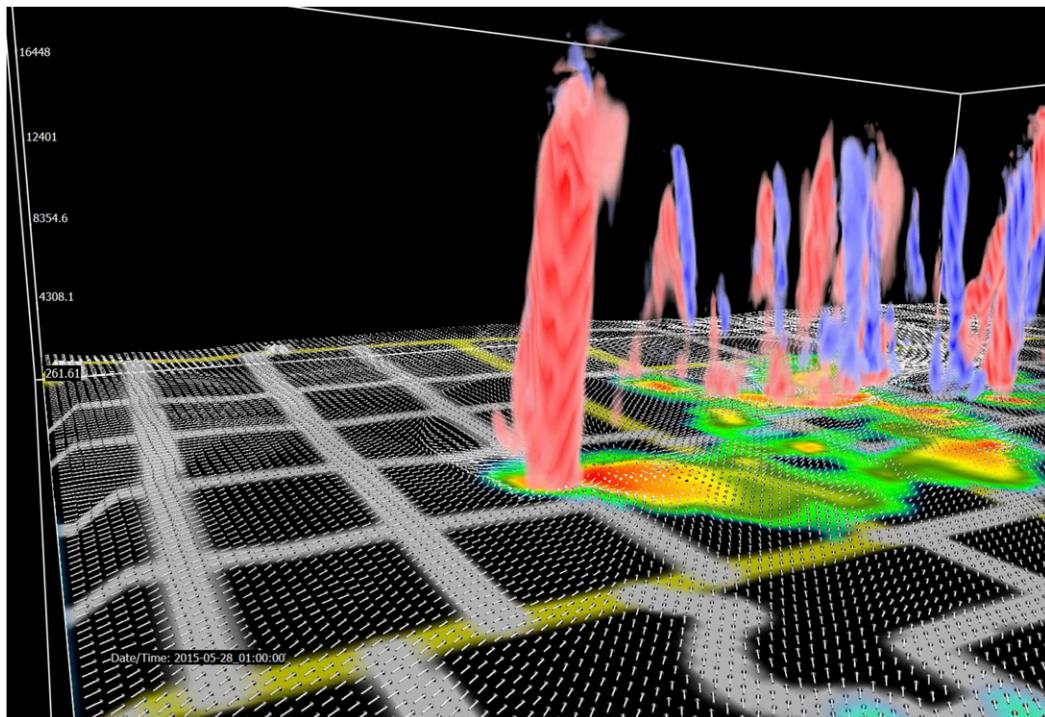


FIG. 3. 3D visualization of forecasted storms valid at 0100 UTC 28 May 2015, looking to the northwest from western OK and showing near-surface wind vectors (white), near-surface radar reflectivity (2D color-shaded field), and UH (red, positive; blue, negative). County boundaries are in white and state boundaries are in yellow.

Introduced in SFE 2014 and enhanced in SFE 2015 were three-dimensional animations of CAM output (Clyne et al. 2007). The 3D images were generated from a $600 \text{ km} \times 600 \text{ km}$ subdomain of the CAPS control forecast chosen daily based on prior forecasts and 2D output fields. Selected 3D animations were shown to participants during the daily weather briefing, allowing a deeper investigation of the processes that lead to potential severe weather threats. These four-dimensional depictions showed features such as local UH [calculated at each volume rendered in the visualization; Brewster et al. (2016)], near-surface radar reflectivity, and near-surface wind vectors (Fig. 3). Deep columns of UH indicated supercellular storms, while animation of the images showed the longevity of such columns: long-lived UH columns often indicated heightened tornado risk.

The experimental forecasts for individual severe hazards were objectively evaluated in near-real time for SFE 2015, a continuation of efforts that had started in SFE 2014 (Melick et al. 2014). For the probabilistic hail forecasts, side-by-side spatial plots and corresponding forecast verification metrics for both LSR and MESH were provided daily, allowing participants to test the usefulness of alternative verifying data sources. For the current work, comparisons of MESH and LSR observational datasets for hail verification were made using the area under the receiver

operating characteristic curve (ROC curve; Mason 1982), estimated using a triangular approach, which measures the ability of a forecast to discriminate between events (i.e., hail occurrence) and nonevents (i.e., no hail occurrence). ROC area values range from 0 to 1, with 1 indicating perfect discrimination and 0.5 indicating no forecast skill.

The experimental, probabilistic hail forecasts for the day 1 and 2 full periods and the 4-h periods of 1800–2200 UTC and 2200–0200 UTC were verified using practically perfect forecasts, formed from the LSRs by applying a two-dimensional Gaussian smoother (Brooks et al. 2003) to reports within 40 km of a $40 \text{ km} \times 40 \text{ km}$ grid box. For effective comparison against LSRs, similar practically perfect forecasts for MESH were produced by applying the same smoother to a separate set of derived severe hail events created by determining if $\text{MESH} \geq 29 \text{ mm}$ (Cintineo et al. 2012) at each grid point. To avoid inclusion of spurious hourly MESH tracks, the presence of at least one cloud-to-ground lightning flash detected by the National Lightning Detection Network (Cummins et al. 1998) within a 40-km radius of influence (ROI) was also required. A 40-km ROI neighborhood maximum was then applied to the final analyses. These quality control measures are similar in nature to those outlined in Melick et al. (2014). The components of the POD and the probability of false detection (POFD) were aggregated

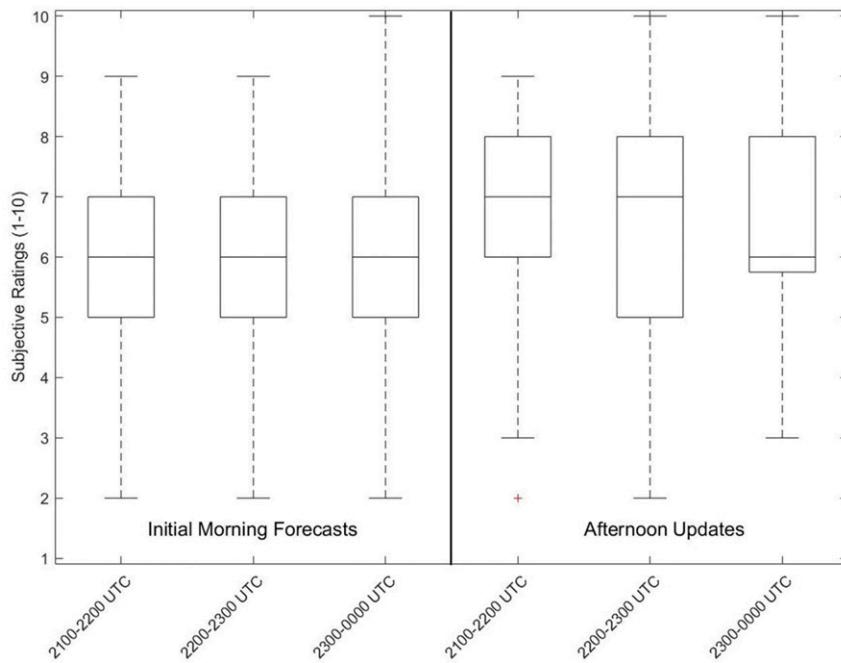


FIG. 4. Distribution of subjective ratings (1–10) for (left) the preliminary hourly experimental forecasts (2100–0000 UTC) issued at 1600 UTC compared with (right) the final experimental forecasts (valid 2100–0000 UTC) issued at 2100 UTC. The boxes compose the IQR of the distributions and the whiskers extend to the 10th and 90th percentiles. Outliers are indicated by red plus symbols.

over the subdomains that had the highest severe weather potential for the given day across the experiment. In addition to the objective verification, participants commented on using MESH compared to LSRs for verifying probabilistic severe hail forecasts.

In addition to the evaluation of severe convective hazards, objective evaluation of the ensemble mean quantitative precipitation forecasts (QPFs) also took place during SFE 2015. The ensemble means were computed using the probability matching technique (Ebert 2001) over a domain encompassing approximately the eastern two-thirds of the CONUS. This technique assumes that the best spatial representation of the precipitation field is given by the ensemble mean, and that the best probability density function of rain rates is given by the ensemble member QPFs of all n ensemble members.

Objective evaluation of these mean fields used the equitable threat score (ETS; Schaefer 1990) for four QPF thresholds. This analysis encompassed five of the six ensembles within the experiment. The ETS measures the fraction of observed and/or forecast events that were correctly predicted, adjusted for correct yes forecasts associated with random chance. The ETS was calculated using contingency table elements computed every 3 h (from forecast hour 3 through forecast hour 36) from each grid point in the ensemble mean analysis domain,

using NCEP Stage IV precipitation data as truth. Forecasts and observations were regridded to a common 4-km grid prior to evaluation. An ETS of 1 is perfect, and a negative score represents no forecast skill. Probabilities of exceeding each threshold were computed by using the ratio of members that exceeded the specified threshold to the total number of members. These forecasts were evaluated using the ROC area, with probability thresholds ranging from 0.05 to 0.95 in increments of 0.05.

3. Preliminary findings and results

a. Evaluation of short-term severe forecasts

1) 1-H TOTAL SEVERE FORECASTS

Participants generally rated the hourly total severe forecasts highly (Fig. 4), with the updated afternoon forecasts garnering higher ratings than the corresponding preliminary morning forecasts. These ratings encompass all individual hourly forecast ratings and, therefore, include timing, placement, and magnitude errors. Afternoon updates allowed forecasters to shift both the magnitude and the location of the probabilities, which produced mixed subjective results in SFE 2015. As stated by a 4 May participant: “21–22Z improved from morning due to pulling the probabilities southward.

However, an increase in probs was not appropriate.” Though generally the afternoon updates occurred closer to the event, participants had difficulty forecasting on days when the convective mode was not yet apparent: “Shorter lead time no help in anticipating messy storm evolution.” (4 May). On other days, there was some evidence that the convective mode was more apparent by the time of update issuance: “Definitely an improvement from earlier. Convective mode was forecasted more accurately...” (7 May). According to participants, the variability within the ensemble of participant forecasts mostly came from varying probability magnitudes, rather than varying locations. Some participants mention the difficulty of calibrating themselves to issue appropriate 1-h forecast probabilities as a potential cause for the variability. Also, the afternoon updates to the forecasts often narrowed the envelope of participant forecasts, as ongoing convection often removed the convective initiation forecast problem.

The mode forecasting problem was perhaps partially illustrated by the widening of the interquartile range (IQR) of the forecast ratings during the afternoon updates (Fig. 4). Difficulty in convective mode forecasting increases the ratings’ variability, as it is difficult to discern to the hour when and if individual supercells will grow upscale into an organized mesoscale convective system (MCS). SFE 2015 also encompassed many days with complex, mixed-mode convection, leading to difficulty of forecasting on an hourly basis. A 4 May participant reflected: “There were also questions early about whether or not convection would occur across the entire frontal boundary, and this question did not seem fully resolved by the afternoon update.” Ultimately, overall afternoon forecast improvement was also subjectively noted: “The afternoon updates were able to trim false alarm areas and refine the major regions for higher probabilities” (3 June).

2) 4-H INDIVIDUAL HAZARD FORECASTS

Participants rated the preliminary 4-h individual hazard forecasts and the disaggregated first-guess hazard probabilities for 1800–2200 and 2200–0200 UTC. During the earlier period, experimental forecasts and the first-guess guidance were often rated similarly, with a median rating difference of 0 for tornadoes and wind, and +1 for hail on a scale from –3 to +3 (Fig. 5). While the evening period experimental forecasts improved upon the earlier, first-guess guidance, most of these ratings reflected marginal improvement (i.e., from 0 to +1). Participant comments also supported only marginal improvement, partially as a result of having relatively little updated model information available: “It was difficult to justify substantial updates to the

afternoon forecast given a modicum of new information (i.e., the new information we had, small in nature compared to the larger set of data from the 0000 UTC cycle), did not warrant changes” (26 May).

b. Comparison of convection-allowing ensembles

SFE 2015 provided the unique opportunity to compare multiple CAM ensemble designs of varying complexity. The 3-h ETS scores of QPF for each ensemble across the experiment were positive, indicating that all ensembles showed positive forecast skill at each threshold and hour. The lowest QPF threshold (Fig. 6a) overall had the highest ETS scores, with the SSEF 3DVAR performing better than all of the other ensembles at all forecast hours, though the difference typically only showed significance for the first few hours, and then again at approximately 24 h from initialization. At the highest precipitation threshold (Fig. 6d), the ETS score differences among the ensembles was largest in the first 12 h of the forecast period and had essentially vanished by forecast hour 18. The ROC areas at each threshold (Fig. 7) show a similar trend at all precipitation thresholds, although the dominance of the SSEF 3DVAR is less pronounced. Interestingly, however, these ROC area differences between the SSEF 3DVAR and the other ensembles were often significant, particularly at the lower thresholds. At the 0.10-in. (Fig. 7a) and 0.25-in. (Fig. 7b) exceedance thresholds, all ensembles (with the exception of the NCAR ensemble 3-h forecast) maintain skillful ROC areas. At higher thresholds the ensembles were less skillful, with the NCAR and SSEF EnKF ensembles having ROC areas less than 0.7 for most forecast hours when considering at least 0.50 in. of precipitation (Fig. 7c) and only a handful of forecast hours for each ensemble system having skillful ROC areas at the 0.75-in. threshold (Fig. 7d). EnKF-analyzed reflectivity was noted to be too low, suggesting that there may have been an error in the EnKF configuration. Additionally, differing ensemble background and data assimilation may have affected the score; for example, only limited sets of conventional observations were assimilated in the SSEF EnKF compared to other ensembles. ROC areas tended to decrease later in the forecast period at low thresholds and had a slight decrease in the middle of the forecast period at higher thresholds. Overall, the SSEF 3DVAR generally scored highest in the objective QPF metrics.

Subjectively, the participants’ ratings of the day 1 ensemble forecasts hourly maximum fields were again rather similar between ensembles (Fig. 8), excepting the SSEF EnKF, which was clearly the lowest-rated ensemble. The top-performing ensembles had a mean rating above six for the SFE, indicating that they

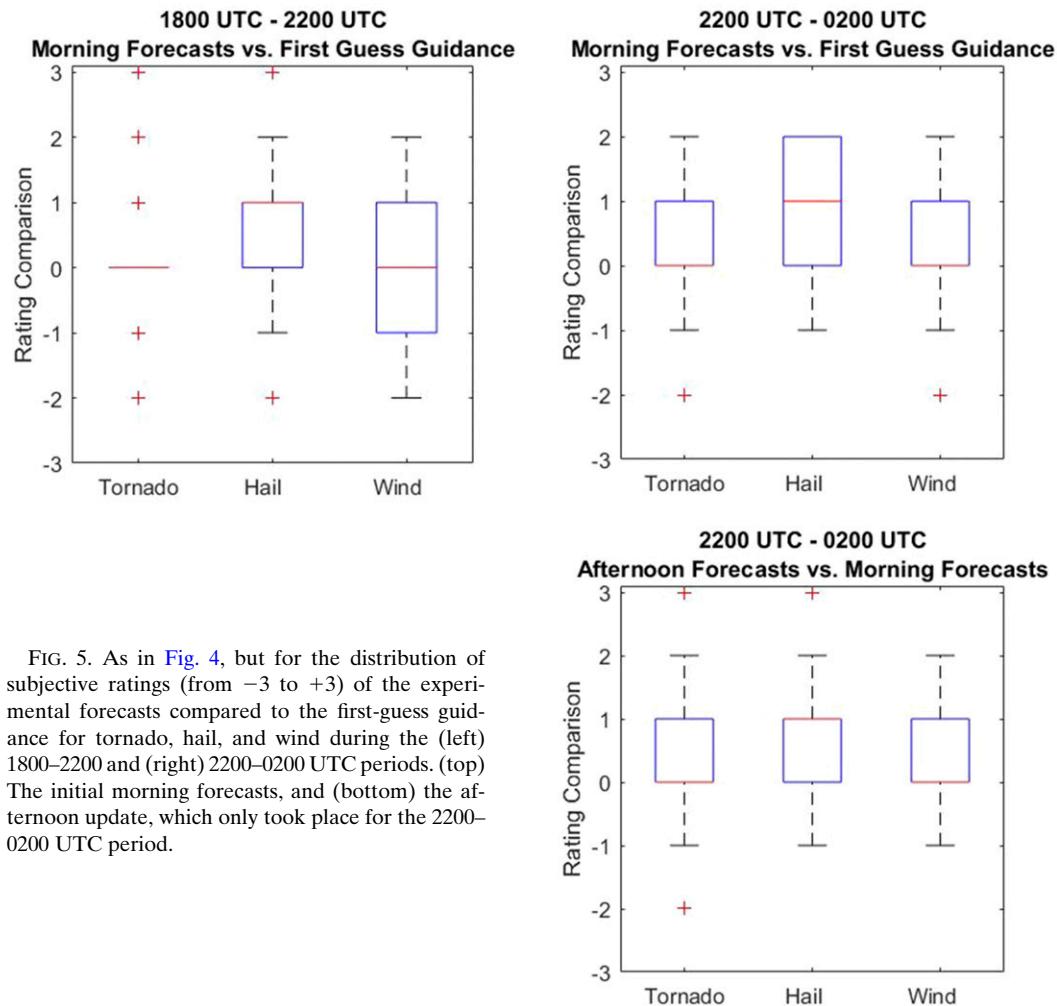


FIG. 5. As in Fig. 4, but for the distribution of subjective ratings (from -3 to $+3$) of the experimental forecasts compared to the first-guess guidance for tornado, hail, and wind during the (left) 1800–2200 and (right) 2200–0200 UTC periods. (top) The initial morning forecasts, and (bottom) the afternoon update, which only took place for the 2200–0200 UTC period.

provided useful severe weather guidance more often than not. As one participant commented, “Mostly agreeing forecasts which all did reasonably well. Some modest discrimination based on amount of false alarm” (14 May). Of the six CAM ensembles, the NSSL ensemble had a slightly higher mean and median rating than the other ensembles, which was significantly higher than the SSEF, SSEF EnKF, and the AFWA ensembles, as determined by a paired-sample t test. The AFWA and NCAR ensembles had lower mean ratings than the SSEO, NSSL, and SSEF, but the difference did not reach the point of being considered significant. The only other significant difference between the mean ratings was that the SSEF EnKF was rated significantly lower than the NSSL, AFWA, and SSEO ensembles.

The day 2 period (forecast hours 36–60) was less frequently objectively evaluated than the day 1 period as a result of computational and data constraints, but the preliminary subjective results provide some insights. The AFWA and NCAR ensembles were more

likely to have day 2 forecasts rated similar to or better than their day 1 ratings compared to the SSEF 3DVAR or the SSEF EnKF, as illustrated by the AFWA ensemble on 21 May 2015 (Figs. 9a–g). For this case, the day 1 forecasts (Figs. 9b–d) placed the majority of the UH-based ensemble neighborhood severe probabilities too far north and offshore, away from the verifying LSRs, whereas the day 2 forecasts (Figs. 9e–g) encompass all LSRs. However, specificity of the day 2 probabilities was also occasionally problematic: “One issue with the longer range forecasts is that areas seem more joined rather than separate, which is reasonable (expected) but still makes it not as good as the day 1” (3 June). Another participant stated that “at least for this date the ensemble sets not assimilating radar data do better from the Day 2 forecast over the Day 1 forecast. I’m guessing this would be more likely for cases in which convection is ongoing and the non-radar assimilating ensembles

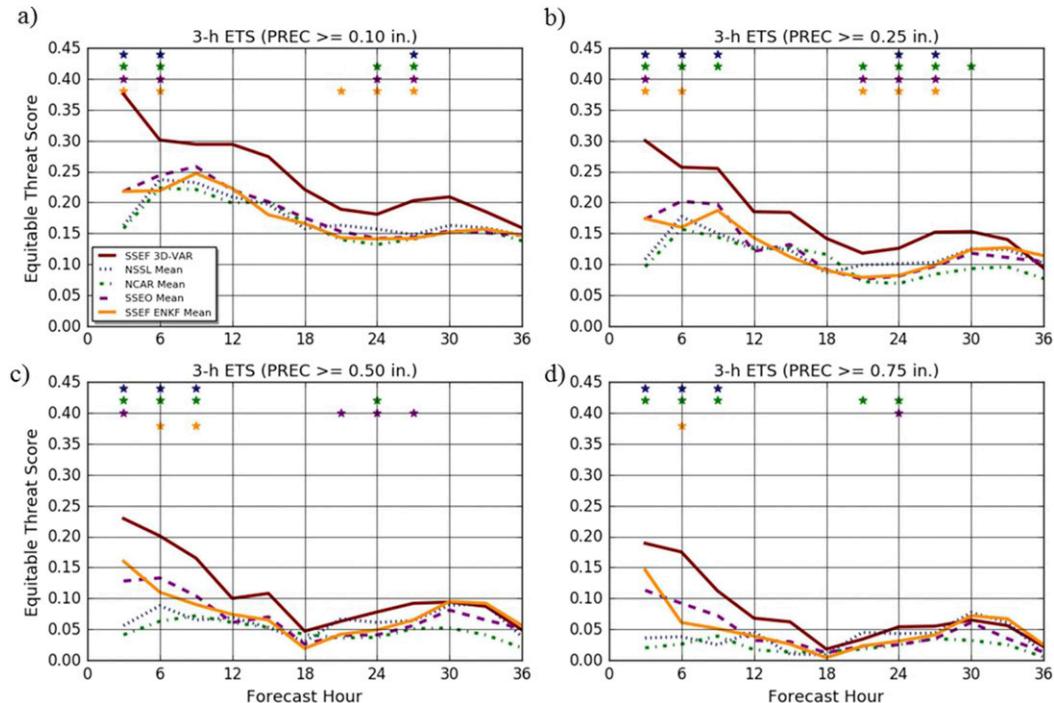


FIG. 6. ETS scores for 3-h ensemble probability-matched mean fields at four QPF exceedance thresholds: (a) 0.10, (b) 0.25, (c) 0.50, and (d) 0.75 in. Different colored lines represent the different models, and colored stars indicate a significant difference between the SSEF 3DVAR ensemble and the ensemble corresponding to that color.

serve more utility as a medium 24–48 range forecast” (14 May). Overall, the extended CAM ensembles provided useful day 2 severe weather guidance, although poor depiction of day 1 convection can detract from the day 2 forecasts.

c. Comparison and evaluation of convection-allowing deterministic models

1) PARALLEL OPERATIONAL CAMS

The parallel versions of both the NAM nest and the HRRR showed subjective improvements over the operational versions, while the parallel and operational NAM runs were given similar subjective ratings (not shown). The parallel HRRR results showed a reduction in the warm, dry, afternoon bias compared with the operational HRRR, resulting in improved convective initiation forecasts (e.g., Fig. 10). The parallel HRRR became operational on 23 August 2016, displacing the operational version used during the SFE 2015 time frame, and the parallel NAM nest became operational on 8 September 2015.

2) MET OFFICE UM

Participants compared the operational UM to the NSSL-WRF daily in SFE 2015. In addition to the 12–

36-h forecasts, the 1–11-h forecasts were compared between the modeling systems to test which system better handled convective spinup. Out of 133 responses, 55% rated the UM better than the NSSL-WRF, 23% rated the UM worse than the NSSL-WRF, and 22% said that they were the same in the first 12 h of the forecast. These percentages were roughly the same when considering the 12–36-h period (132 total responses), with a slightly larger percentage (26% of responses) reporting that they were the same. Overall, the parallel UM (122 responses) was generally worse than (46%) or the same as (30%) the operational UM, and the 1.1-km UM (104 responses) was typically the same as (43%) or worse than (32%) the 2.2-km version.

Sounding comparisons between the NSSL-WRF and the operational UM (Fig. 11) often showed striking differences. Throughout SFE 2015, capping inversions in the operational UM were consistently more sharply defined than in the NSSL-WRF, more closely matching the observational soundings and consistent with the examples shown in Kain et al. (2017). Out of 89 total participant responses, 60 expressed that the UM soundings were better than the NSSL-WRF, while 19 felt the two were the same. Only 10 responses rated the NSSL-WRF soundings better than the UM. The structure and sharpness of the strong capping inversions were

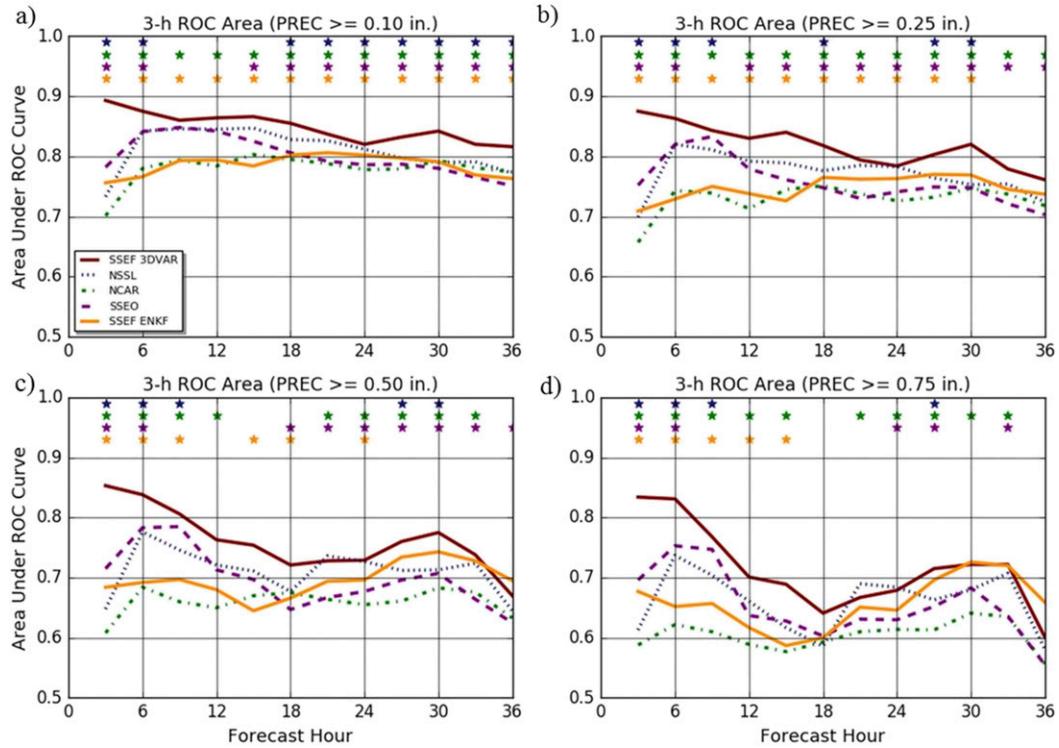


FIG. 7. As in Fig. 6, but for ROC area scores.

subjectively noted by participants as being much better depicted in the UM than the NSSL-WRF: “UKMET is better. Depicts inversion temperature profile perfectly. This is the biggest difference” (2 June). Although the UKMET has nearly double the number of vertical levels of the NSSL-WRF, Kain et al. (2017) state that merely increasing the vertical

resolution of the NSSL-WRF does not negate this tendency.

3) MPAS

While no formal evaluation of the MPAS forecasts took place, the guidance was examined on a daily basis and used during the forecasting process. Two cases

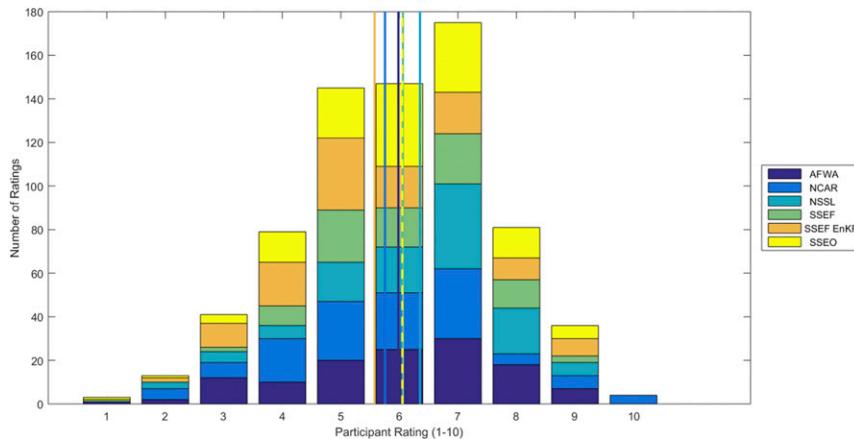


FIG. 8. Distribution of subjective ratings (1–10) for the ensemble hourly maximum field forecasts compared to LSRs for each ensemble. Mean subjective ratings are indicated by a vertical line. The dashed line indicates the mean of both the SSEF (3DVAR) and the SSEO subjective ratings.

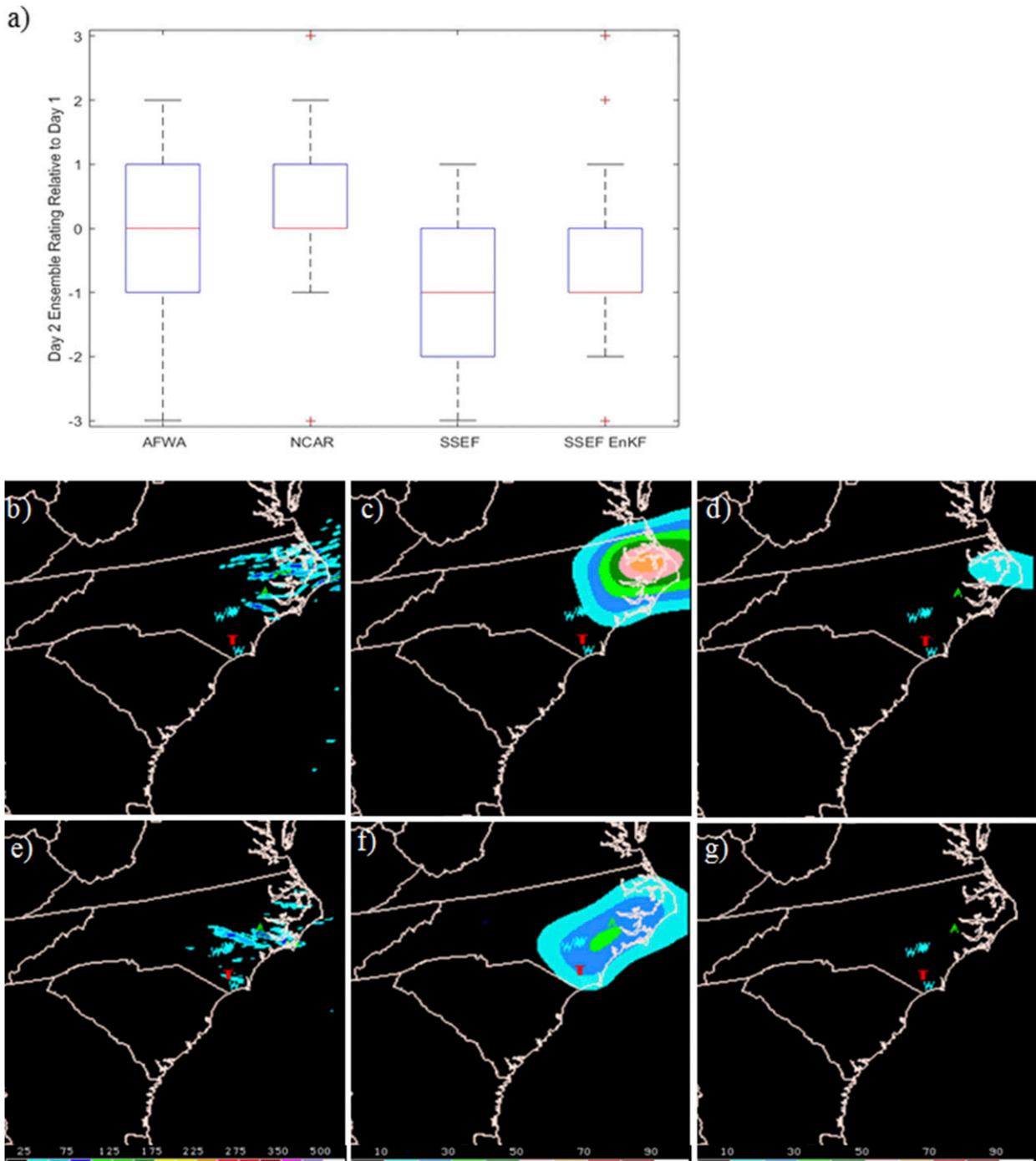


FIG. 9. (a) As in Fig. 4, but for the distribution of subjective ratings (from -3 to $+3$) for the day 2 ensemble forecasts compared with the day 1 forecasts, valid for the same time period. As an example, the AFWA (middle) day 1 and (bottom) day 2 forecasts of the (b),(e) 4-h ensemble maximum UH, (c),(f) ensemble neighborhood probability of UH $\geq 25 \text{ m}^2 \text{ s}^{-2}$, and (d),(g) ensemble neighborhood probability of UH $\geq 100 \text{ m}^2 \text{ s}^{-2}$ valid at 1800–2200 UTC 21 May 2015. The severe reports during this 4-h period are plotted as letters in each panel (T for tornado, W for wind, and A for hail).

where useful convective-scale guidance to day 3 and beyond are presented here, as a preliminary indication of the usefulness of MPAS in forecasting severe convection at longer time scales than most current

convection-allowing guidance. Both days provided similar synoptic patterns conducive to a severe weather outbreak across the southern plains, with the eventual outcome heavily dependent on the presence of

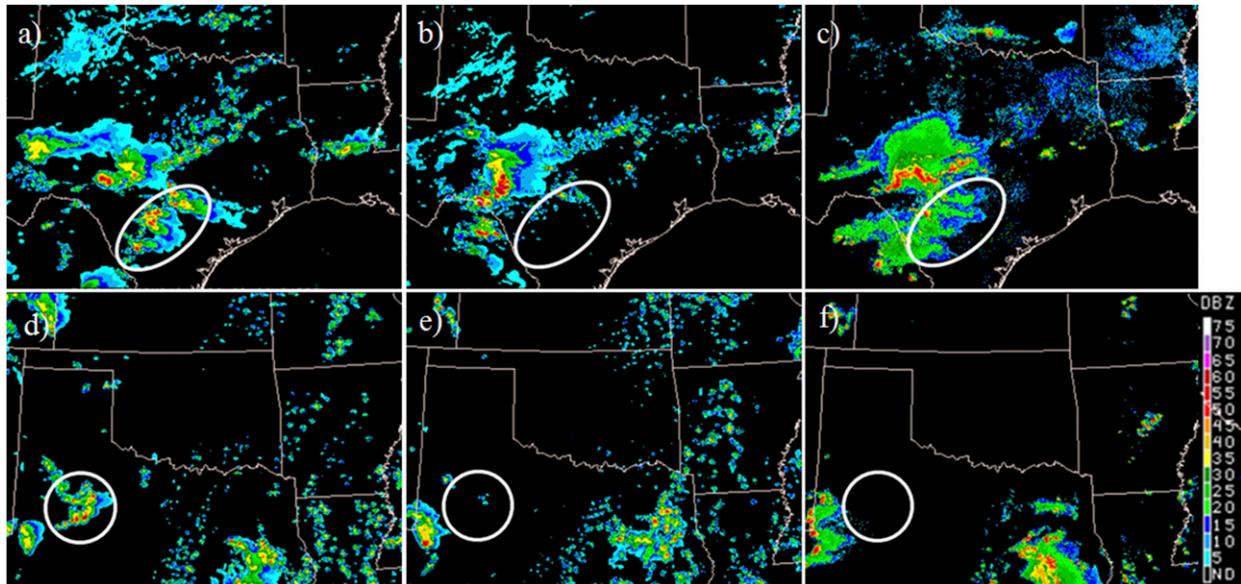


FIG. 10. Simulated reflectivity forecasts valid at 0300 UTC 21 May 2015 from the (a) 1500 UTC operational HRRR, (b) 1500 UTC parallel HRRR, and (c) observed reflectivity. Simulated reflectivity forecasts valid at 2200 UTC 14 May 2015 from the (d) 1500 UTC operational HRRR, (e) parallel HRRR, and (f) observed reflectivity.

morning convection, related to the strength of the capping inversion.

Several days in advance of 9 May 2015, the SPC day 3 convective outlook outlined an area across Oklahoma and Kansas as having a moderate risk for severe storms. In reality, during the late morning of 9 May, strong forcing for ascent combined with a weak capping inversion led to widespread convection and associated cloud cover across much of western Oklahoma and Kansas, inhibiting afternoon destabilization. The early convection led to minimal CAPE ($<1000 \text{ J kg}^{-1}$) across much of Oklahoma and Kansas (Fig. 12a). Although severe storms did occur from Texas into western Kansas, because of the early storms the event as a whole ended up being less significant than what some earlier model guidance had suggested. While forecasting a synoptic-scale pattern favorable for widespread severe weather 3 days in advance of 9 May, the MPAS forecasts also indicated that widespread convection would develop early in the day on 9 May. The impact of this early convection manifested itself in reduced CAPE simulated across Oklahoma and Kansas (Fig. 12c). Thus, the scenario depicted by MPAS 3 days in advance was consistent with what occurred.

The second case with a favorable synoptic pattern for severe weather in which MPAS provided useful extended-range guidance was on 16 May 2015. Similar to 9 May, the extent and intensity of the severe weather threat was uncertain, because it was not clear how much early convection would inhibit heating and destabilization in the warm sector. Despite a shallow layer

of clouds, a lack of widespread early convection allowed enough destabilization (Fig. 12b) to support a significant severe weather event and several long-lived tornadic supercells across the Texas Panhandle, Oklahoma, and Missouri. The forecasts from MPAS 3 days in advance were consistent with this scenario, maintaining CAPE through early convection (Fig. 12d) and matching quite well the observed range of CAPE values. Furthermore, the MPAS forecasts depicted intense supercells forming in the warm sector around 2100 UTC beginning with the 93-h, day 4 forecast (Fig. 13e) and continuing through the day 3 (Fig. 13d), day 2 (Fig. 13c), and day 1 (Fig. 13b) forecasts. The location of the storms was initially too far east compared with observations (Fig. 13a). Additionally, the timing of upscale growth was also well depicted as far as 4 days in advance (Fig. 13k), clearly showing the squall line over central Oklahoma at 0355 UTC (Fig. 13g). The overall forecast scenario corresponded well to the observations, particularly regarding the mode and timing of mode evolution and again would have provided useful extended-range convective-scale guidance to forecasters.

d. Evaluation of new diagnostics

1) HAIL DIAGNOSTICS

Three days' worth of WRF-HAILCAST results were formally evaluated in SFE 2015, precluding robust conclusions. Compatibility issues resulted in the Thompson method only being available in the NCAR ensemble, and thus a direct comparison to the WRF-HAILCAST

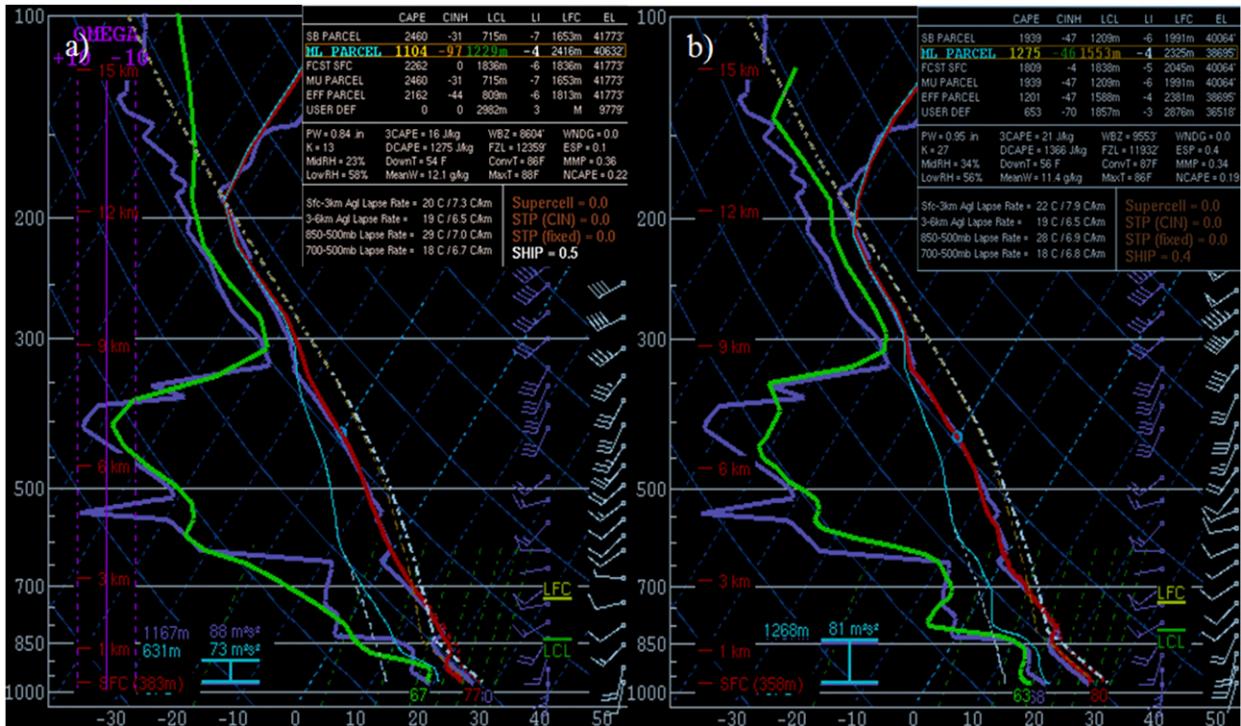


FIG. 11. The 24-h forecast soundings valid 15 May 2015 for the OUN station from (a) the NSSL-WRF control member and (b) the UKMET 2.2-km model. The observed sounding is plotted in purple in each panel.

implemented in the SSEF system was impossible. However, participants unanimously agreed that across the three cases the hail size forecasts provided additional useful information relative to more commonly used hourly maximum fields such as UH, prompting the inclusion of the new hail diagnostics in future SFEs.

2) TORNADO DIAGNOSTICS

The distributions of subjective ratings assigned to the 24-h tornado probabilities by the individual participants suggest that incorporating environmental information results in an improved forecast over solely using UH (Fig. 14). None of the environmental filters (LCL, CAPE, STP, or combined) clearly stood out as the best method; however, they all generally improved upon the UH-only guidance. Participants often noted that the incorporation of environmental information helped focus the area of interest and reduce the number of false alarms. However, they often felt that the probabilities were too high on a given day to directly translate into the current operational convective outlook categories (i.e., tornado probabilities of 30% on a day that SPC forecasters would not consider a “moderate” risk given the environment).

e. Hail verification comparisons

When participants evaluated MESH as verification for probabilistic severe hail forecasts, rather than LSRs,

the responses were generally positive. A participant said on 4 May: “Assuming that MESH is reasonably representative of what actually occurred, it definitely helps fill in areas between local storm reports.” Many participants commented that MESH provided verification in low-population-density areas such as eastern Colorado (Figs. 15a,b), where obtaining even a single report to verify a warning may be difficult. Participants “liked the spatial and temporal details much better” (7 May), and noted that in these locations when reports did occur, MESH often also diagnosed large hail (Figs. 15c,d). However, participants were unsure of directly comparing LSRs and MESH, stating: “Hard to say how well it does in verifying when not comparing hail sizes in MESH to actual LSR observed hail sizes. . . .” Ortega et al. (2009) performed a concentrated verification of MESH tracks, but a larger-scale verification database does not yet exist. Wilson et al. (2009) found that MESH performs best at values greater than 19 mm, which would include all severe hail, although they advise against using MESH alone as a form of synthetic verification; MESH has also been found to overforecast hail size (Wilson et al. 2009; Cintineo et al. 2012). Cintineo et al. (2012) find that Heidke skill scores are maximized in a comparison of MESH with high-resolution ground-truth reports of severe hail when a threshold of 29 mm is used. Further, Melick et al. (2014) have suggested that

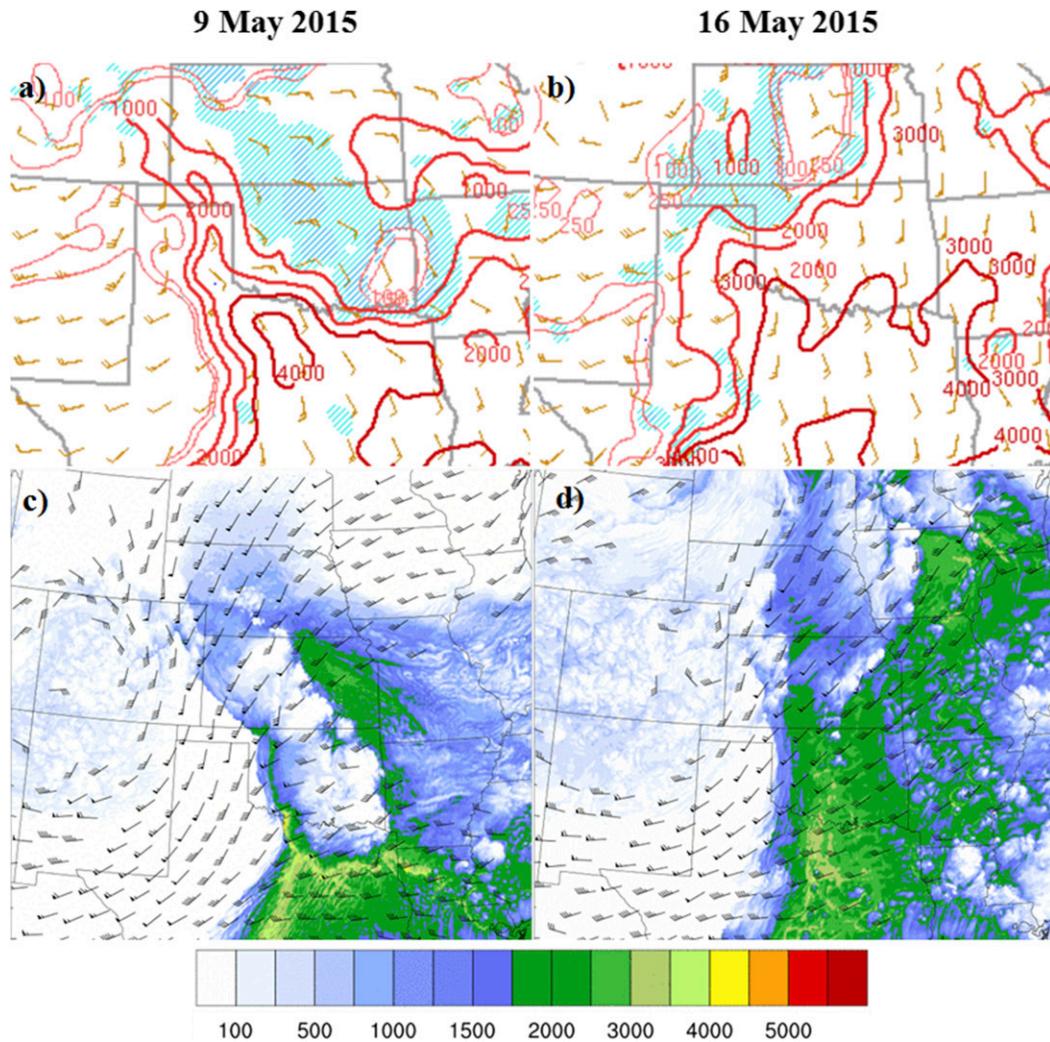


FIG. 12. CAPE and CIN from SPC's mesoanalysis valid at (a) 2100 UTC 9 May and (b) 2100 UTC 16 May 2015. CAPE contour levels (red) are 100, 250, 500, and 1000 J kg^{-1} and then are spaced every 1000 J kg^{-1} . Light blue shading indicates CIN less than -25 J kg^{-1} , and dark blue shading indicates CIN less than -100 J kg^{-1} . The 69-h MPAS forecasts of CAPE and 0–6-km shear vectors beginning at 30 kt (where 1 kt = 0.51 m s^{-1}), valid at (c) 2100 UTC 9 May and (d) 2100 UTC 16 May 2015.

MESH tracks can be useful as an independent dataset to supplement hail LSRs. Consequently, the positive response from participants recommends an objective look at MESH verification over the daily subdomains.

Objective verification of the experimental hail forecasts with practically perfect forecasts generated by MESH (17 cases) and LSR (23 cases) at different periods via ROC area (Fig. 15e) showed that whether MESH or LSR verified the forecasts best was dependent on the time period examined. Looking at the full-period day 1 forecasts, LSRs had a higher POD and approximately the same POFD as the MESH, leading to a larger ROC area. Conversely, the day 2 full-period forecasts show both higher POD and higher POFD when verified

using the MESH results, rather than the LSRs. The 4-h outlooks generally performed better than the daily outlooks in both verifications. This is particularly evident in the 2200–0200 UTC time frame, when convective initiation was less of a forecast problem. These results suggest that the hail forecasts are typically able to distinguish the area of hail. However, ROC areas do not take into consideration the reliability of the forecasts, which was a large factor in participants' subjective ratings of the verification methods. Indeed, participants noted that the higher “practically perfect” probabilities were often generated using the MESH tracks (Fig. 15d) compared with the LSRs (Fig. 15b): “Practically perfect probabilities from MESH seemed overestimated

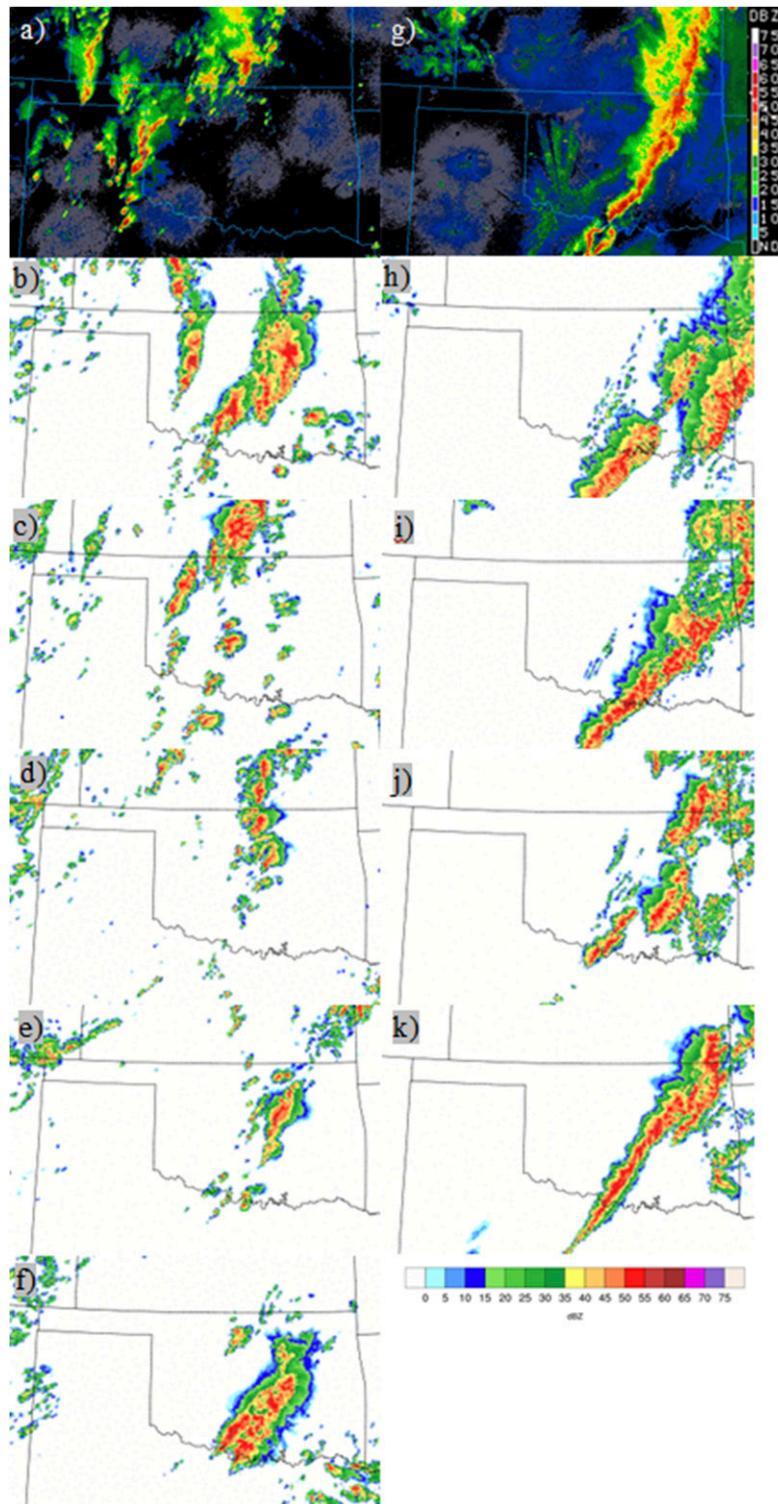


FIG. 13. Composite reflectivity observations at (a) 2100 UTC 16 May and (g) 0400 UTC 17 May 2015. MPAS (b) 21-, (c) 45-, (d) 69-, (e) 93-, and (f) 117-h composite reflectivity forecasts valid at 2300 UTC 16 May 2015 and (h) 28-, (i) 52-, (j) 76-, and (k) 100-h composite reflectivity forecasts valid at 0500 UTC 17 May 2015.

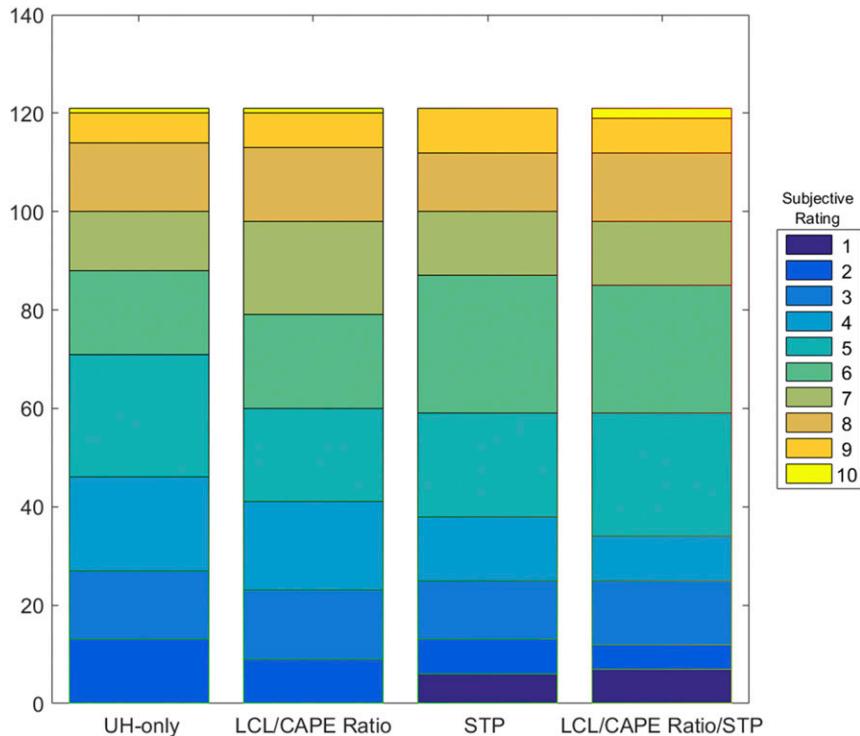


FIG. 14. Subjective ratings of 24-h tornado probabilities generated from the NSSL-WRF ensemble requiring four different environmental criteria, along with UH $\geq 75 \text{ m}^2 \text{ s}^{-2}$. Each set of probabilities received 121 ratings total. [Adapted from Gallo et al. (2016).]

compared to the report probabilities” (19 May). This may be because participants are not used to seeing MESH-derived practically perfect probabilities. However, these higher probabilities did not seem to dampen the participants’ enthusiasm for using MESH as a verification metric. One participant on 27 May stated, “Even if a slight oververification [sic] given its construction, the use of MESH for verification seems to be an improvement on this day.”

4. Summary and discussion

Overall, SFE 2015 succeeded in testing new forecast products and modeling systems to address relevant issues in predicting hazardous convective weather. The sheer volume of daily numerical weather guidance examined throughout SFE 2015 was unprecedented, and the real-time, operational nature of the experiment emphasized the need for tools that forecasters can use to summarize large volumes of information when forecasting severe convective weather. The innovative nature of the experiment gave participants access to cutting-edge, operationally relevant research from multiple institutions, evaluating six CAM ensembles,

three deterministic Met Office CAMs, a deterministic CAM with forecasts extending out to 5 days (MPAS), parallel versions of current operational models, and new diagnostic techniques for hail size and tornado occurrence. The experiment found that parallel versions of the HRRR and the NAM nest improved upon the current operational versions, providing strong evidence to support implementation of the experimental parallel modeling systems. Additionally, CAMs were found useful when issuing day 2 forecasts, providing mode insight for medium-range severe convective forecasts. Day 2 forecasts occasionally rated more highly than the corresponding day 1 forecasts, although participants noted that day 2 forecasts started from ensembles assimilating radar data can be affected if the day 1 convection is poorly handled, essentially relying on them as a medium-range forecast. The SFE also helped to determine that applying environmental filters to explicit UH diagnostics improved guidance for probabilistic tornado forecasting compared to using UH only with the NSSL-WRF ensemble.

Increased participant interaction was a key component of SFE 2015. Using laptops in the experiment allowed participants to submit individual, rather than

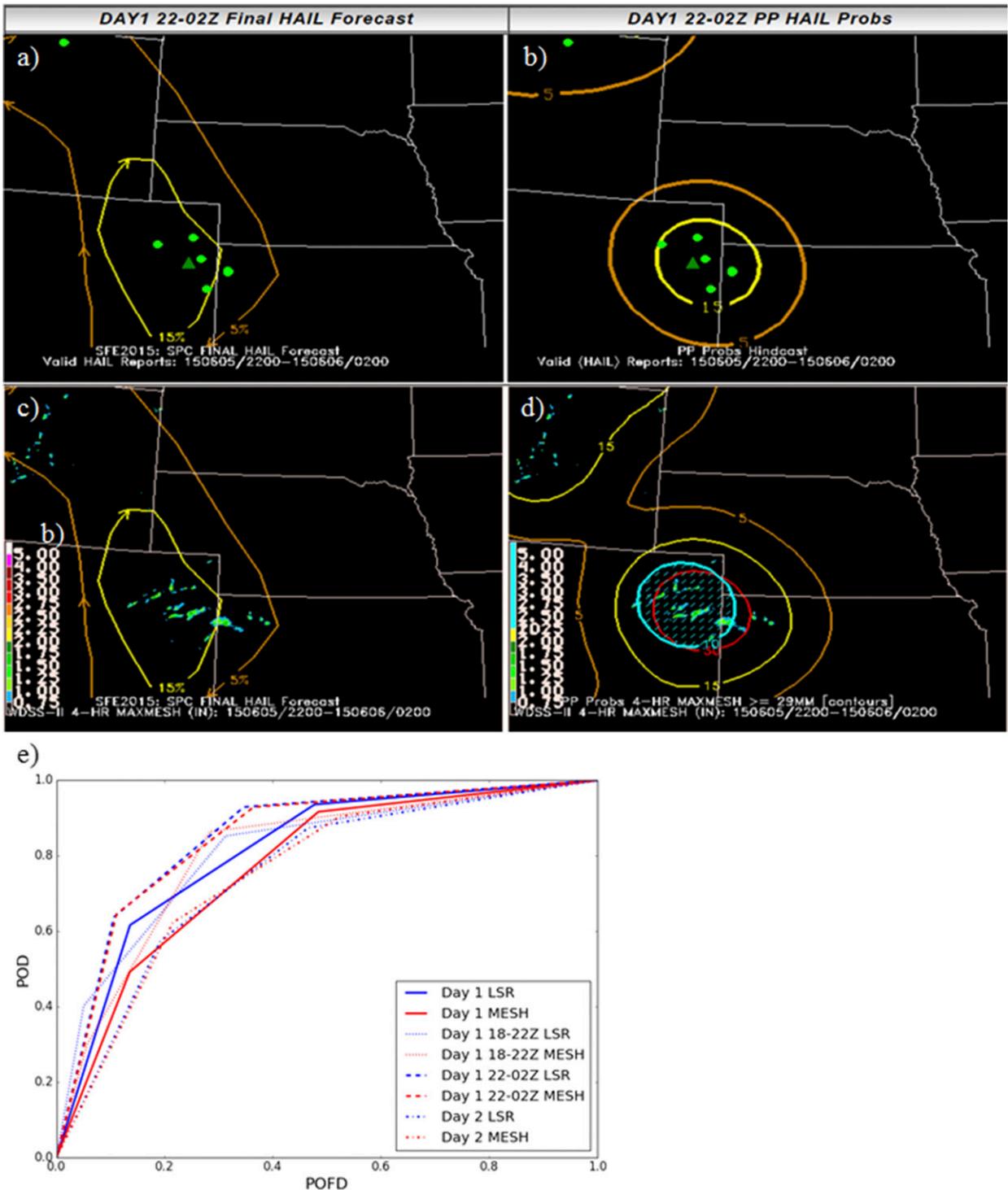


FIG. 15. Individual hazard desk SPC forecaster’s hail forecasts from 2200 UTC 5 May to 0200 UTC 6 May 2015 (a),(c) verified against practically perfect forecasts generated using (b) hail LSRs (green dots) and significant hail LSRs (dark green triangles) and (d) MESH tracks. Full periods encompass 1600–1200 UTC the following day. The blue hatched area is indicative of severe hail (≥ 2 in.). (e) ROC curves showing the accumulated verification results for all of SFE 2015 using LSRs and MESH.

group consensus, evaluations and allowed for personalized feedback. The usage of the PHI tool on individual laptops allowed for more participant engagement, as they drew their own short-term forecasts. These short-term forecasts performed well objectively and subjectively, suggesting that moving these products into operations is feasible, fulfilling an operational product and service improvement goal. These forecasts benefited greatly from the availability of the CAM guidance, particularly the hourly forecasts of total severe. To make such reliable forecasts without CAM guidance would have been difficult.

Annual SFEs in the HWT have a long history of impacting National Weather Service operations, but oftentimes one has to consider a multiyear period to get a full measure of these impacts. For example, SFE 2010 contained one CAM ensemble provided by CAPS, and was just beginning to evaluate hourly maximum fields such as UH and simulated 1 km AGL reflectivity. These fields tested in SFE 2010 are now considered key output parameters in operational CAMs and are used worldwide, showing how the SFEs succeed in research-to-operations efforts. Since that SFE, grid spacing has decreased, and the number and availability of CAM ensembles has greatly increased. SFE 2015 allowed its participants to study the behavior of these ensembles, bolstering their knowledge of the latest forecasting techniques. SFE 2015 also provided researchers with knowledge of how the many NWP guidance options provided to forecasters are perceived, in addition to information about how comparable these ensembles are at the height of the spring convective season.

SFE 2015 highlighted areas requiring future study through verification efforts in conjunction with the NOAA applied science activities goals. Participant comments on using MESH in addition to LSRs for hail verification suggest that MESH tracks may be a good future verification source, albeit after a larger comparison database is compiled between MESH and LSRs. The tendency of hail guidance to either overforecast (WRF-HAILCAST) or underforecast (Thompson) hail sizes, and the overforecasting tendency of the tornado probabilities noted by participants, highlight that more work is needed regarding individual hazard diagnostics. Future work focusing on individual hazard diagnostics is planned to compare the diagnostics between ensembles and to current SPC forecasts for individual hazards. Finally, the striking difference between the Met Office CAMs and the NSSL-WRF in representing strong vertical gradients in temperature and moisture near capping inversions demonstrates that work is still needed to hone the accuracy of the vertical profiles.

With SFE 2015 complete, future SFEs can build upon the lessons learned therein. Surprisingly, though the six ensembles in SFE 2015 were configured differently, the ensembles' performance according to both objective and subjective measures was quite similar. This result led to a focus in SFE 2016 on uncovering how differences in ensemble configuration affect model performance with regard to severe convective weather using the recently developed Community Leveraged Unified Ensemble (CLUE; Clark et al. 2016). CLUE consisted of 65 members provided by a number of institutions, all of which had the same domain, grid spacing, and output fields. These members were divided into a number of subexperiments for directly comparing configuration strategies (i.e., multicore versus single core, multiphysics versus single physics, 3DVAR versus EnKF, ensemble size sensitivity). By minimizing as many differences as possible between the members, it is hoped that CLUE will help inform key ensemble configuration decisions, providing valuable guidance for operational CAM ensemble design.

Acknowledgments. First, the authors thank all of the participants and contributors to the annual spring forecasting experiments, whose work, insight, and excitement make the SFEs possible. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant DGE-1102691, Project A00-4125. AJC, KK, JC, CJM, CDK and WL were provided support by NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreement NA11OAR4320072, U.S. Department of Commerce. AJC also received support from a Presidential Early Career Award for Scientists and Engineers. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation Grant ACI-1053575. MX, KB, and FK received support from NOAA CSTAR Grant NWSPO-2010-201696. Thanks also go to Scott Rentschler for providing information regarding the AFWA ensemble. Finally, the authors thank two anonymous reviewers and Philip N. Schumacher for their careful consideration and comments, which greatly improved the clarity of the manuscript.

REFERENCES

- Adams-Selin, R., and C. Ziegler, 2016: Forecasting hail using a one-dimensional hail growth model within WRF. *Mon. Wea. Rev.*, **144**, 4919–4939, doi:10.1175/MWR-D-16-0027.1.
- , —, and A. J. Clark, 2014: Forecasting hail using a one-dimensional hail growth model inline within WRF. *27th Conf. on Severe Local Storms*, Madison, WI, Amer. Meteor. Soc.,

- 11B.2. [Available online at <https://ams.confex.com/ams/27SLS/webprogram/Paper255184.html>.]
- Alexander, C. R., S. S. Weygandt, T. G. Smirnova, S. Benjamin, P. Hofmann, E. P. James, and D. A. Koch, 2010: High Resolution Rapid Refresh (HRRR): Recent enhancements and evaluation during the 2010 convective season. *25th Conf. on Severe Local Storms*, Denver, CO, Amer. Meteor. Soc., 9.2. [Available online at https://ams.confex.com/ams/25SLS/techprogram/paper_175722.htm.]
- , and Coauthors, 2015: The 2015 operational upgrades to the Rapid Refresh (RAP) and High-Resolution Rapid Refresh (HRRR). *27th Conf. on Weather Analysis and Forecasting/23rd Conf. on Numerical Weather Prediction*, Chicago, IL, Amer. Meteor. Soc., 2A.2. [Available online at <https://ams.confex.com/ams/27WAF23NWP/webprogram/Paper273721.html>.]
- Aligo, E., B. S. Ferrier, J. Carley, E. Rogers, M. Pyle, S. J. Weiss, and I. L. Jirak, 2014: Modified microphysics for use in high-resolution NAM forecasts. *27th Conf. on Severe Local Storms*, Madison, WI, Amer. Meteor. Soc., 16A.1. [Available online at <https://ams.confex.com/ams/27SLS/webprogram/Paper255732.html>.]
- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903, doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.
- , 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, **131**, 634–642, doi:10.1175/1520-0493(2003)131<0634:ALLSFF>2.0.CO;2.
- , T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Avellano, 2009: The Data Assimilation Research Testbed: A community facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296, doi:10.1175/2009BAMS2618.1.
- Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, doi:10.1175/MWR-D-15-0242.1.
- Bougeault, P., and P. Lacarrère, 1989: Parameterization of orography-induced turbulence in a mesobeta-scale model. *Mon. Wea. Rev.*, **117**, 1872–1890, doi:10.1175/1520-0493(1989)117<1872:POOITI>2.0.CO;2.
- Brewster, K. A., D. R. Stratman, and R. Hepper, 2016: 4D visualization of storm-scale forecasts using VAPOR in the Hazardous Weather Testbed Spring Forecasting Experiment. *28th Conf. on Severe Local Storms*, Portland, OR, Amer. Meteor. Soc., 15B.6. [Available online at <https://ams.confex.com/ams/28SLS/webprogram/Paper301696.html>.]
- Brimelow, J. C., G. W. Reuter, and E. R. Poolman, 2002: Modeling maximum hail size in Alberta thunderstorms. *Wea. Forecasting*, **17**, 1048–1062, doi:10.1175/1520-0434(2002)017<1048:MMHSIA>2.0.CO;2.
- Brooks, H. E., M. Kay, and J. A. Hart, 1998: Objective limits on forecasting skill of rare events. Preprints, *19th Conf. on Severe Local Storms*, Minneapolis, MN, Amer. Meteor. Soc., 552–555.
- , 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, doi:10.1175/1520-0434(2003)018<0626:CEOLDT>2.0.CO;2.
- Chen, F., and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model description and implementation. *Mon. Wea. Rev.*, **129**, 569–585, doi:10.1175/1520-0493(2001)129,0569:CAALSH.2.0.CO;2.
- Cintineo, J. L., T. M. Smith, V. Lakshmanan, H. E. Brooks, and K. L. Ortega, 2012: An objective high-resolution hail climatology of the contiguous United States. *Wea. Forecasting*, **27**, 1235–1248, doi:10.1175/WAF-D-11-00151.1.
- Clark, A. J., and Coauthors, 2012a: An overview of the 2010 Hazardous Weather Testbed Experimental Forecast Program Spring Experiment. *Bull. Amer. Meteor. Soc.*, **93**, 55–74, doi:10.1175/BAMS-D-11-00040.1.
- , J. S. Kain, P. T. Marsh, J. Correia, M. Xue, and F. Kong, 2012b: Forecasting tornado pathlengths using a three-dimensional object identification algorithm applied to convection-allowing forecasts. *Wea. Forecasting*, **27**, 1090–1113, doi:10.1175/WAF-D-11-00147.1.
- , and Coauthors, 2015: Spring Forecasting Experiment 2015 program overview and operations plan. NOAA/NSSL, 24 pp. [Available online at https://hwt.nssl.noaa.gov/Spring_2015/HWT_SFE_2015_OPS_plan_final.pdf.]
- , and Coauthors, 2016: Spring Forecasting Experiment 2016 program overview and operations plan. NOAA/NSSL, 30 pp. [Available online at https://hwt.nssl.noaa.gov/Spring_2016/HWT_SFE2016_operations_plan_final.pdf.]
- Clyne, J., P. Mininni, A. Norton, and M. Rast, 2007: Interactive desktop analysis of high resolution simulations: Application to turbulent plume dynamics and current sheet formation. *New J. Phys.*, **9**, 301, doi:10.1088/1367-2630/9/8/301.
- Coniglio, M. C., D. A. Imy, C. D. Karstens, A. J. Clark, J. Correia Jr., and C. J. Melick, 2014: Evaluation of one-hour probabilistic severe weather forecasts issued during the 2014 NOAA Hazardous Weather Testbed Spring Forecasting Experiment. *27th Conf. on Severe Local Storms*, Madison, WI, Amer. Meteor. Soc., 47. [Available online at <https://ams.confex.com/ams/27SLS/webprogram/Paper255617.html>.]
- Cummins, K. L., M. J. Murphy, E. A. Bardo, W. L. Hiscox, R. B. Pyle, and A. E. Pifer, 1998: A combined TOA/MDF technology upgrade of the U.S. National Lightning Detection Network. *J. Geophys. Res.*, **103**, 9035–9044, doi:10.1029/98JD00153.
- Davies, T., M. J. P. Cullen, A. J. Malcolm, M. H. Mawson, A. Staniforth, A. A. White, and N. Wood, 2005: A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Quart. J. Roy. Meteor. Soc.*, **131**, 1759–1782, doi:10.1256/qj.04.101.
- Du, J., and Coauthors, 2014: NCEP regional ensemble update: Current systems and planned storm-scale ensembles. *26th Conf. on Weather Analysis and Forecasting/22nd Conf. on Numerical Weather Prediction*, Atlanta, GA, Amer. Meteor. Soc., J1.4. [Available online at <https://ams.confex.com/ams/94Annual/webprogram/Paper239030.html>.]
- Duda, J. D., X. Wang, F. Kong, and M. Xue, 2014: Using varied microphysics to account for uncertainty in warm-season QPF in a convection-allowing ensemble. *Mon. Wea. Rev.*, **142**, 2198–2219, doi:10.1175/MWR-D-13-00297.1.
- Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107, doi:10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2.
- Ebert, E. E., 2001: Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon. Wea. Rev.*, **129**, 2461–2480, doi:10.1175/1520-0493(2001)129<2461:AOAPMS>2.0.CO;2.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162, doi:10.1029/94JC00572.

- , 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, doi:10.1007/s10236-003-0036-9.
- Ferrier, B. S., 1994: A double-moment multiple-phase four-class bulk ice scheme. Part I: Description. *J. Atmos. Sci.*, **51**, 249–280, doi:10.1175/1520-0469(1994)051<0249:ADMMPF>2.0.CO;2.
- Gallo, B. T., A. J. Clark, and S. R. Dembek, 2016: Forecasting tornadoes using convection-permitting ensembles. *Wea. Forecasting*, **31**, 273–295, doi:10.1175/WAF-D-15-0134.1.
- Hitchens, N. M., H. E. Brooks, and M. P. Kay, 2013: Objective limits on forecasting skill of rare events. *Wea. Forecasting*, **28**, 525–534, doi:10.1175/WAF-D-12-00113.1.
- Hong, S.-Y., and J.-O. J. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Korean Meteor. Soc.*, **42**, 129–151.
- , J. Dudhia, and S. H. Chen, 2004: A revised approach to ice microphysical processes for the bulk parameterization of cloud and precipitations. *Mon. Wea. Rev.*, **132**, 103–120, doi:10.1175/1520-0493(2004)132<0103:ARATIM>2.0.CO;2.
- , S. Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341, doi:10.1175/MWR3199.1.
- Hu, M., M. Xue, and K. Brewster, 2006: 3DVAR and cloud analysis with WSR-88D level-II data for the prediction of Fort Worth tornadic thunderstorms. Part I: Cloud analysis and its impact. *Mon. Wea. Rev.*, **134**, 675–698, doi:10.1175/MWR3092.1.
- Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, doi:10.1029/2008JD009944.
- Janjić, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927–945, doi:10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2.
- , 2002: Nonsingular implementation of the Mellor–Yamada level 2.5 scheme in the NCEP Meso Model. NCEP Office Note 437, 61 pp. [Available online at http://www2.mmm.ucar.edu/wrf/users/phys_refs/SURFACE_LAYER/eta_part4.pdf.]
- , and R. Gall, 2012: Scientific documentation of the NCEP Nonhydrostatic Multiscale Model on the B grid (NMMB). Part 1: Dynamics. NCAR/TN-489+STR, 75 pp. [Available online at <https://opensky.ucar.edu/islandora/object/technotes%3A502>.]
- Jewell, R., and J. Brimelow, 2009: Evaluation of Alberta hail growth model using severe hail proximity soundings from the United States. *Wea. Forecasting*, **24**, 1592–1609, doi:10.1175/2009WAF2222230.1.
- Jirak, I. L., C. J. Melick, A. R. Dean, S. J. Weiss, and J. Correia Jr., 2012a: Investigation of an automated temporal disaggregation technique for convective outlooks during the 2012 Hazardous Weather Testbed Spring Forecasting Experiment. *26th Conf. on Severe Local Storms*, Nashville, TN, Amer. Meteor. Soc., 10.2. [Available online at <https://ams.confex.com/ams/26SLS/webprogram/Paper211733.html>.]
- , S. J. Weiss, and C. J. Melick, 2012b: The SPC storm-scale ensemble of opportunity: Overview and results from the 2012 Hazardous Weather Testbed Spring Forecasting Experiment. *26th Conf. on Severe Local Storms*, Nashville, TN, Amer. Meteor. Soc., 137. [Available online at <https://ams.confex.com/ams/26SLS/webprogram/Paper211729.html>.]
- Johnson, A., X. Wang, F. Kong, and M. Xue, 2013: Object-based evaluation of the impact of horizontal grid spacing on convection-allowing forecasts. *Mon. Wea. Rev.*, **141**, 3413–3425, doi:10.1175/MWR-D-13-00027.1.
- Kain, J. S., P. R. Janish, S. J. Weiss, M. E. Baldwin, R. S. Schneider, and H. E. Brooks, 2003: Collaboration between forecasters and research scientists at the NSSL and SPC: The Spring Program. *Bull. Amer. Meteor. Soc.*, **84**, 1797–1806, doi:10.1175/BAMS-84-12-1797.
- , 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, doi:10.1175/2010WAF2222430.1.
- , and Coauthors, 2017: Collaborative efforts between the United States and the United Kingdom to advance prediction of high-impact weather. *Bull. Amer. Meteor. Soc.*, **98**, 937–948, doi:10.1175/BAMS-D-15-00199.1.
- Karstens, C. D., T. M. Smith, K. M. Kuhlman, A. J. Clark, C. Ling, G. J. Sutmpf, and L. P. Rothfus, 2014: Prototype tool development for creating probabilistic hazard information for severe convective phenomena. *Second Symp. on Building a Weather-Ready Nation: Enhancing Our Nation's Readiness, Responsiveness, and Resilience to High Impact Weather Events*. Atlanta, GA, Amer. Meteor. Soc., 2.2. [Available online at <https://ams.confex.com/ams/94Annual/webprogram/Paper241549.html>.]
- , and Coauthors, 2015: Evaluation of a probabilistic forecasting methodology for severe convective weather in the 2014 Hazardous Weather Testbed. *Wea. Forecasting*, **30**, 1551–1570, doi:10.1175/WAF-D-14-00163.1.
- Kong, F., and Coauthors, 2015: An overview of CAPS storm-scale ensemble forecast for the 2015 NOAA HWT Spring Forecasting Experiment. *27th Conf. Weather Analysis and Forecasting/23rd Conf. on Numerical Weather Prediction*, Chicago, IL, Amer. Meteor. Soc., 32. [Available online at <https://ams.confex.com/ams/27WAF23NWP/webprogram/Paper273814.html>.]
- Kuchera, E., S. Rentschler, G. Creighton, and J. Hamilton, 2014: The Air Force weather ensemble prediction suite. *15th Annual WRF Users' Workshop*, Boulder, CO, UCAR–NCAR. [Available online at <http://www2.mmm.ucar.edu/wrf/users/workshops/WS2014/ppts/2.3.pdf>.]
- Kumar, S. V., and Coauthors, 2006: Land Information System: An interoperable framework for high resolution land surface modeling. *Environ. Modell. Software*, **21**, 1402–1415, doi:10.1016/j.envsoft.2005.07.004.
- , C. D. Peters-Lidard, J. L. Eastman, and W.-K. Tao, 2008: An integrated high-resolution hydrometeorological modeling testbed using LIS and WRF. *Environ. Modell. Software*, **23**, 169–181, doi:10.1016/j.envsoft.2007.05.012.
- Lim, K.-S. S., and S.-Y. Hong, 2010: Development of an effective double-moment cloud microphysics scheme with prognostic cloud condensation nuclei (CCN) for weather and climate models. *Mon. Wea. Rev.*, **138**, 1587–1612, doi:10.1175/2009MWR2968.1.
- Line, W. E., T. J. Schmit, D. T. Lindsey, and S. J. Goodman, 2016: Use of geostationary super rapid scan satellite imagery by the Storm Prediction Center. *Wea. Forecasting*, **31**, 483–494, doi:10.1175/WAF-D-15-0135.1.
- Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303.
- McBeath, K., P. R. Field, and R. J. Cotton, 2014: Using operational weather radar to assess high-resolution numerical weather prediction over the British Isles for a cold air outbreak case-study. *Quart. J. Roy. Meteor. Soc.*, **140**, 225–239, doi:10.1002/qj.2123.

- Melick, C. J., I. L. Jirak, J. Correia Jr., A. R. Dean, and S. J. Weiss, 2014: Exploration of the NSSL maximum expected size of hail (MESH) product for verifying experimental hail forecasts in the 2014 Spring Forecasting Experiment. *27th Conf. on Severe Local Storms*, Madison, WI, Amer. Meteor. Soc., 76. [Available online at <https://ams.confex.com/ams/27SLS/webprogram/Paper254292.html>.]
- Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys. Space Phys.*, **20**, 851–875, doi:10.1029/RG020i004p00851.
- Menzel, W. P., 2001: Cloud tracking with satellite imagery: From the pioneering work of Ted Fujita to the present. *Bull. Amer. Meteor. Soc.*, **82**, 33–47, doi:10.1175/1520-0477(2001)082<0033:CTWSIF>2.3.CO;2.
- Milbrandt, J. A., and M. K. Yau, 2005: A multimoment bulk microphysics parameterization. Part I: Analysis of the role of the spectral shape parameter. *J. Atmos. Sci.*, **62**, 3051–3064, doi:10.1175/JAS3534.1.
- Miller, P. A., M. F. Barth, and L. A. Benjamin, 2005: An update on MADIS observation ingest, integration, quality control and distribution capabilities. *21st Int. Conf. on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology/14th Symp on Education*, San Diego, CA, Amer. Meteor. Soc., J7.12. [Available online at https://ams.confex.com/ams/Annual2005/techprogram/paper_86703.htm.]
- , —, —, R. S. Artz, and W. R. Pendergrass, 2007: MADIS support for UrbanNet. *14th Symp. on Meteorological Observation and Instrumentation/16th Conf. on Applied Climatology*, San Antonio, TX, Amer. Meteor. Soc., JP2.5. [Available online at https://ams.confex.com/ams/87ANNUAL/techprogram/paper_119116.htm.]
- Mittermaier, M. P., 2014: A strategy for verifying near-convection-resolving model forecasts at observing sites. *Wea. Forecasting*, **29**, 185–204, doi:10.1175/WAF-D-12-00075.1.
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682, doi:10.1029/97JD00237.
- Morrison, H., and J. O. Pinto, 2005: Mesoscale modeling of springtime Arctic mixed-phase stratiform clouds using a new two-moment bulk microphysics scheme. *J. Atmos. Sci.*, **62**, 3683–3704, doi:10.1175/JAS3564.1.
- , and —, 2006: Intercomparison of bulk microphysics scheme in mesoscale simulations of springtime Arctic mixed-phase stratiform clouds. *Mon. Wea. Rev.*, **134**, 1880–1900, doi:10.1175/MWR3154.1.
- , and J. A. Milbrandt, 2015: Parameterization of cloud microphysics based on the prediction of bulk ice particle properties. Part I: Scheme description and idealized tests. *J. Atmos. Sci.*, **72**, 287–311, doi:10.1175/JAS-D-14-0065.1.
- Nakanishi, M., and H. Niino, 2004: An improved Mellor–Yamada level-3 model with condensation physics: Its design and verification. *Bound.-Layer Meteor.*, **112**, 1–31, doi:10.1023/B:BOUN.0000020164.04146.98.
- , and —, 2006: An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Bound.-Layer Meteor.*, **119**, 397–407, doi:10.1007/s10546-005-9030-8.
- Ortega, K. L., T. M. Smith, K. L. Manross, K. A. Scharfenberg, A. Witt, A. G. Kolodziej, and J. J. Gourley, 2009: The Severe Hazards Analysis and Verification Experiment. *Bull. Amer. Meteor. Soc.*, **90**, 1519–1530, doi:10.1175/2009BAMS2815.1.
- Rothfus, L., C. D. Karstens, and D. Hilderbrand, 2014: Forecasting a continuum of environmental threats: Exploring next-generation forecasting of high impact weather. *Eos, Trans. Amer. Geophys. Union*, **95**, 325–326, doi:10.1002/2014EO360001.
- Satoh, M., T. Masuno, H. Tomita, H. Miura, T. Nasuno, and S. Iga, 2008: Nonhydrostatic icosahedral atmospheric model (NICAM) for global cloud resolving simulations. *J. Comput. Phys.*, **227**, 3486–3514, doi:10.1016/j.jcp.2007.02.006.
- Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting*, **5**, 570–575, doi:10.1175/1520-0434(1990)005<0570:TCSIAA>2.0.CO;2.
- Schwartz, C. S., G. S. Romine, K. R. Smith, and M. L. Weisman, 2014: Characterizing and optimizing precipitation forecasts from a convection-permitting ensemble initialized by a mesoscale ensemble Kalman filter. *Wea. Forecasting*, **29**, 1295–1318, doi:10.1175/WAF-D-13-00145.1.
- , —, 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, doi:10.1175/WAF-D-15-0103.1.
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., doi:10.5065/D68S4MVH.
- , J. B. Klemp, M. Duda, L. D. Fowler, S.-H. Park, and T. Ringler, 2012: A multiscale nonhydrostatic atmospheric model using centroidal Voronoi tessellations and C-grid staggering. *Mon. Wea. Rev.*, **140**, 3090–3105, doi:10.1175/MWR-D-11-00215.1.
- Smagorinsky, J., 1963: General circulation experiments with the primitive equations. *Mon. Wea. Rev.*, **91**, 99–164, doi:10.1175/1520-0493(1963)091<0099:GCEWTP>2.3.CO;2.
- Smirnova, T. G., J. M. Brown, and S. G. Benjamin, 1997: Performance of different soil model configurations in simulating ground surface temperature and surface fluxes. *Mon. Wea. Rev.*, **125**, 1870–1884, doi:10.1175/1520-0493(1997)125<1870:PODSMC>2.0.CO;2.
- , —, and D. Kim, 2000: Parameterization of cold-season processes in the 551 MAPS land-surface scheme. *J. Geophys. Res.*, **105**, 4077–4086, doi:10.1029/1999JD901047.
- , —, S. Benjamin, and J. Kenyon, 2015: Modifications to the Rapid Update Cycle land surface model (RUC LSM) available in the Weather Research and Forecasting (WRF) Model. *Mon. Wea. Rev.*, **144**, 1851–1865, doi:10.1175/MWR-D-15-0198.1.
- Smith, T. M., and Coauthors, 2014: Examination of a real-time 3DVAR analysis system in the Hazardous Weather Testbed. *Wea. Forecasting*, **29**, 63–77, doi:10.1175/WAF-D-13-00044.1.
- Stensrud, D. J., and Coauthors, 2009: Convective-scale warm-on-forecast system: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1499, doi:10.1175/2009BAMS2795.1.
- Surcel, M., I. Zawadzki, and M. K. Yau, 2014: On the filtering properties of ensemble averaging for storm-scale precipitation forecasts. *Mon. Wea. Rev.*, **142**, 1093–1105, doi:10.1175/MWR-D-13-00134.1.
- Thompson, G., and T. Eidhammer, 2014: A study of aerosol impacts on clouds and precipitation development in a large winter cyclone. *J. Atmos. Sci.*, **71**, 3636–3658, doi:10.1175/JAS-D-13-0305.1.
- , R. M. Rasmussen, and K. Manning, 2004: Explicit forecasts of winter precipitation using an improved bulk microphysics

- scheme. Part I: Description and sensitivity analysis. *Mon. Wea. Rev.*, **132**, 519–542, doi:10.1175/1520-0493(2004)132<0519:EFOWPU>2.0.CO;2.
- , P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, doi:10.1175/2008MWR2387.1.
- Thompson, R. L., R. Edwards, J. A. Hart, K. L. Elmore, and P. Markowski, 2003: Close proximity soundings within supercell environments obtained from the Rapid Update Cycle. *Wea. Forecasting*, **18**, 1243–1261, doi:10.1175/1520-0434(2003)018<1243:CPSWSE>2.0.CO;2.
- Towns, J., and Coauthors, 2014: XSEDE: Accelerating scientific discovery. *Comput. Sci. Eng.*, **16** (5), 62–74, doi:10.1109/MCSE.2014.80.
- Wang, Y., Y. Jung, T. A. Supinie, and M. Xue, 2013: A hybrid MPI–OpenMP parallel algorithm and performance analysis for an ensemble square root filter suitable for dense observations. *J. Atmos. Oceanic Technol.*, **30**, 1382–1397, doi:10.1175/JTECH-D-12-00165.1.
- Wilson, C. J., K. L. Ortega, and V. Lakshmanan, 2009: Evaluating multi-radar, multi-sensor hail diagnosis with high resolution hail reports. *25th Conf. on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, Phoenix, AZ, Amer. Meteor. Soc., P2.9. [Available online at https://ams.confex.com/ams/89annual/techprogram/paper_146206.htm.]
- Wilson, D. R., and S. P. Ballard, 1999: A microphysically based precipitation scheme for the UK Meteorological Office Unified Model. *Quart. J. Roy. Meteor. Soc.*, **125**, 1607–1636, doi:10.1002/qj.49712555707.
- Witt, A., M. D. Eilts, G. J. Stumpf, J. T. Johnson, E. D. Mitchell, and K. W. Thomas, 1998: An enhanced hail detection algorithm for the WSR-88D. *Wea. Forecasting*, **13**, 286–303, doi:10.1175/1520-0434(1998)013<0286:AEHDAF>2.0.CO;2.
- Xue, M., D.-H. Wang, J.-D. Gao, K. Brewster, and K. K. Droegemeier, 2003: The Advanced Regional Prediction System (ARPS), storm-scale numerical weather prediction and data assimilation. *Meteor. Atmos. Phys.*, **82**, 139–170, doi:10.1007/s00703-001-0595-6.
- , M. Hu, and M. Tong, 2006: Assimilation of radar data and short-range prediction of thunderstorms using 3DVAR, cloud analysis and ensemble Kalman filter methods. *12th Conf. on Aviation, Range, and Aerospace Meteorology*, Atlanta, GA, Amer. Meteor. Soc., 9.2. [Available online at <https://ams.confex.com/ams/Annual2006/webprogram/Paper105115.html>.]