Advancing from Convection-Allowing NWP to Warn-on-Forecast: Evidence of Progress

JOHN R. LAWSON

Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, and NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

JOHN S. KAIN[a]

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

NUSRAT YUSSOUF

Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, and NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

DAVID C. DOWELL

NOAA/OAR/Earth System Research Laboratory, Boulder, Colorado

DUSTAN M. WHEATLEY, KENT H. KNOPFMEIER, AND THOMAS A. JONES

Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, and NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

(Manuscript received 2 October 2017, in final form 17 February 2018)

ABSTRACT

The Warn-on-Forecast (WoF) program, driven by advanced data assimilation and ensemble design of numerical weather prediction (NWP) systems, seeks to advance 0–3-h NWP to aid National Weather Service warnings for thunderstorm-induced hazards. An early prototype of the WoF prediction system is the National Severe Storms Laboratory (NSSL) Experimental WoF System for ensembles (NEWSe), which comprises 36 ensemble members with varied initial conditions and parameterization suites. In the present study, real-time 3-h quantitative precipitation forecasts (QPFs) during spring 2016 from NEWSe members are compared against those from two real-time deterministic systems: the operational High Resolution Rapid Refresh (HRRR, version 1) and an upgraded, experimental configuration of the HRRR. All three model systems were run at 3-km horizontal grid spacing and differ in initialization, particularly in the radar data assimilation methods. It is the impact of this difference that is evaluated herein using both traditional and scale-aware verification schemes. NEWSe, evaluated deterministically for each member, shows marked improvement over the two HRRR versions for 0–3-h QPFs, especially at higher thresholds and smaller spatial scales. This improvement diminishes with forecast lead time. The experimental HRRR model, which became operational as HRRR version 2 in August 2016, also provides added skill over HRRR version 1.

1. Introduction

The Warn-on-Forecast program (WoF; Stensrud et al. 2009) is addressing the challenge of creating numerical weather prediction (NWP) models that can predict specific thunderstorm-induced hazards such as large hail, flood-producing rainfall, strong wind, and proxies for tornadogenesis. The vision for implementing WoF technologies in NWS operations calls for on-demand activation of horizontal grid-spacing WoF ensemble prediction systems (EPSs) of $O(1)$ km, within a much larger scale, for example, the continental U.S. (CONUS) convection-allowing model (CAM) EPS. These EPSs would be frequently updated with observations,

[a] Current affiliation: NCEP Environmental Modeling Center, NOAA/Center for Weather and Climate Prediction, College Park, Maryland.

Corresponding author: John R. Lawson, john.lawson@noaa.gov

notionally with hourly updates on the outer CAM domain and 15 min or less on the embedded WoF domain(s).

If successful, the WoF program will transform the National Weather Service's (NWS) ability to issue warnings for these hazards, by advancing numerical guidance capabilities to the point where they can meaningfully influence decisions during short-fuse NWS warning situations. While these advances will require many more years of research—particularly in the areas of data assimilation, social science, and ensemble design—the program has already borne fruit. Since its inception in 2009, over 70 papers have been published that document WoF-funded results (e.g., Stensrud and Gao 2010; Dowell et al. 2011; Potvin and Wicker 2012; Jones et al. 2013; Wheatley et al. 2014; Yussouf et al. 2013, 2015, 2016; Gao et al. 2016; Potvin et al. 2017). However, most WoF publications have focused on technical aspects of specialized topics and, perhaps, have neglected direct evidence of added value compared to current operational NWP models.

The purpose of the present paper is to provide such evidence, with a straightforward focus on the impact of advanced storm-scale data assimilation techniques on short-term thunderstorm precipitation forecasts. Our approach is to compare 0–3-h precipitation forecasts from a prototype WoF prediction system with those from current-generation operational CAMs, specifically the High Resolution Rapid Refresh (HRRR; Benjamin et al. 2016) modeling system. Both the HRRR version (hereafter HRRRv1) operational in spring 2016 and a parallel, experimental next-generation version (HRRRv2) were compared. We will demonstrate the added skill that WoF-related advances in data assimilation bring to short-term thunderstorm forecasts.

During the spring 2016 Hazardous Weather Testbed (e.g., Karstens et al. 2015), the HRRR team at the Global Systems Division (GSD) of the National Oceanic and Atmospheric Administration/Earth System Research Laboratory (NOAA/ESRL) and the WoF team at the NOAA/National Severe Storms Laboratory (NSSL) demonstrated a jointly developed prototype EPS comprising two components: an hourly updated ensemble over a relatively large, fixed domain and a nested, relocatable ensemble that was updated every 15 min. All NWP systems considered in this study (i.e., fixed and nested domains of the prototype EPS and the two versions of the deterministic HRRR) employed similar versions of the WRF-ARW model (Skamarock et al. 2008) with 3-km horizontal grid spacing. For comparison with HRRR forecasts, we focus on the nested component of the EPS updated every 15 min, hereafter denoted the 2016 prototype of the NSSL

Experimental WoF System for ensembles (NEWSe). A primary difference between the 2016 prototype NEWSe and HRRR is the data assimilation scheme, including data type, volume, update frequency, and total time that the 3-km grid is cycled. Thus, the specific purpose of this paper is to examine how the advanced data assimilation in the 2016 prototype WoF system affects deterministic forecast skill. Presumably, additional skill could also come from higher resolution and the ensemble forecasting approach, but assessment of these impacts is left for future studies, probabilistic or otherwise. Section 2 begins by outlining the systems and verification methods. Section 3 presents results from the impact of data assimilation on 0–3-h precipitation forecasts, and this research is summarized in section 4.

2. Data and methods

In addition to the following sections, a brief comparison of the 2016 prototype NEWSe system with both HRRR systems can be found in Table 1.

a. NEWSe system

The NEWSe system (Wheatley et al. 2015; Jones et al. 2016) assimilates Doppler radar and satellite observations using an ensemble Kalman (EnKF) filter approach. In this study, forecast integrations are performed with the WRF-ARW model, version 3.6.1, run at 3-km horizontal grid spacing to balance the computational cost of resolution and ensemble size. The relocatable domain used for this experiment comprised 249×249 grid points and was recentered daily based on expected regions of hazardous weather. Initialization each day at 1800 UTC involved the interpolation of initial and lateral boundary conditions (ICs and LBCs, respectively) from the fixed EPS domain; 18 different pairs of ICs and LBCs were thence created. NEWSe comprised 36 members, each with a different permutation of ICs, LBCs, and diversity in the planetary boundary layer [i.e., Yonsei University (YSU), MYJ, and Mellor–Yamada–Nakanishi–Niino (MYNN)], shortwave radiation [Dudhia and Rapid Radiative Transfer Model for GCMs (RRTMG)] and longwave radiation (RRTM and RRTMG) parameterization schemes (see Wheatley et al. 2015). Only the first 18 members (covering all IC–LBC pairs) are examined here. Assimilating observations every 15 min, NEWSe was cycled for a total of 9 h, from 1800 to 0300 UTC. Ensemble forecasts were produced hourly beginning at 1900 UTC.

b. HRRR systems

The HRRR configurations examined here comprise the operational HRRRv1, run by the National Centers

TABLE 1. A summary comparison of the NEWSe, HRRRv1, and HRRRv2 configurations.

	NEWSe	HRRRv1	HRRRv2
Model version	WRF-ARW v3.6+	WRF-ARW v3.4.1+	WRF-ARW v3.6+
Grid points	250 × 250 × 50	1799 × 1059 × 50	1799 × 1059 × 50
Grid spacing (km)	3	3	3
DA cycling	EnKF Data Assimilation Research Testbed (DART) every 15 min, 36 members	Hourly 3D-Var, deterministic	Hourly 3D-EnVar, deterministic
Observations	Doppler velocity from ~10 WSR-88D sites; MRMS radar reflectivity; satellite cloud water path	Conventional observations—T, q_v, u, v, and p—from rawinsonde, aircraft, surface (land and marine), profiler, GPS, AMSU/HIRS/Microwave Humidity Sounder (MHS), WindSat scatterometer, and MRMS radar reflectivity	Conventional observations—T, q_v, u, v, and p—from rawinsonde, aircraft, surface (land, mesonet, and marine), profiler, GPS, lightning, AMSU/HIRS/MHS, WindSat scatterometer, and MRMS radar reflectivity
Long- and shortwave radiation	Dudhia/RRTM and RRTMG/RRTMG	RRTM/Goddard	RRTMG/RRTMG
Microphysics	Thompson (aerosol aware)	Thompson v3.4.1	Thompson v3.6 (aerosol aware)
PBL	YSU, MYJ, or MYNN	MYNN v3.4.1	MYNN v3.6
LSM	RUC (Smirnova)	RUC (Smirnova) v3.4.1	RUC (Smirnova) v3.6

for Environmental Prediction (NCEP) at the time of this study, and an experimental HRRRv2 run by ESRL that became operational at NCEP three months after the time of this study. HRRRv1 forecasts were generated hourly with WRF-ARW v3.4.1 running at 3-km horizontal grid spacing, and LBCs were supplied by the parent Rapid Refresh (RAP) model (Benjamin et al. 2016). HRRR ICs were produced as follows: 1) interpolation of model fields from RAP to HRRR grids, 2) reflectivity data assimilation during a 1-h model integration, 3) three-dimensional variational data assimilation (3D-Var) of conventional observations with the NCEP Gridpoint Statistical Interpolation (GSI) package (Wu et al. 2002; Kleist et al. 2009), and 4) cloud and hydrometeor adjustment based on satellite and radar-reflectivity observations (Benjamin et al. 2016). The reflectivity assimilation method in step 2 is a low-cost latent-heating method (Weygandt et al. 2008), performed without a digital filter, that specifies microphysics temperature tendency during the 1-h initialization based on the reflectivity field. HRRRv2 differed from HRRRv1 by using version 3.6.1 of WRF-ARW, three-dimensional hybrid ensemble–variational data assimilation (3D-EnVar) of conventional observations, and refinements of the WRF physical parameterizations.

c. Data assimilation methods

Previous studies have demonstrated the benefits of frequently updated radar data assimilation for short-term (0–6 h) NWP, including improved skill of predicting low-level mesocyclones and other severe-storm features (e.g., Yussouf et al. 2015) and increased lead times for flash flooding (Yussouf et al. 2016). We therefore expect differences in radar data assimilation methods between NEWSe and HRRR to be substantial factors. The

TABLE 2. Cases from 2016 included in the present study. The date refers to the day 1 initialization (convection typically continues into the next UTC day).

Date	Location of interest	0–3-h forecast initialization time (UTC)
7 May	CO, WY	2000, 2100, 0300
8 May	OK, KS	1900–2200, 0000, 0100, 0300
9 May	OK	1900, 2100, 2200, 0100
10 May	KY	1900, 2100, 0100, 0300
16 May	TX Panhandle	1900–2200 0000–0300
17 May	TX	1900, 2100, 2200, 0000–0300
22 May	TX, OK, KS	1900, 2100, 0100, 0300
23 May	TX Panhandle	1900, 2100, 0100, 0300
24 May	KS	1900, 2100, 0300
25 May	KS	1900, 2100, 0100, 0300
26 May	KS	1900, 2100, 0100, 0300
27 May	OK, KS	1900
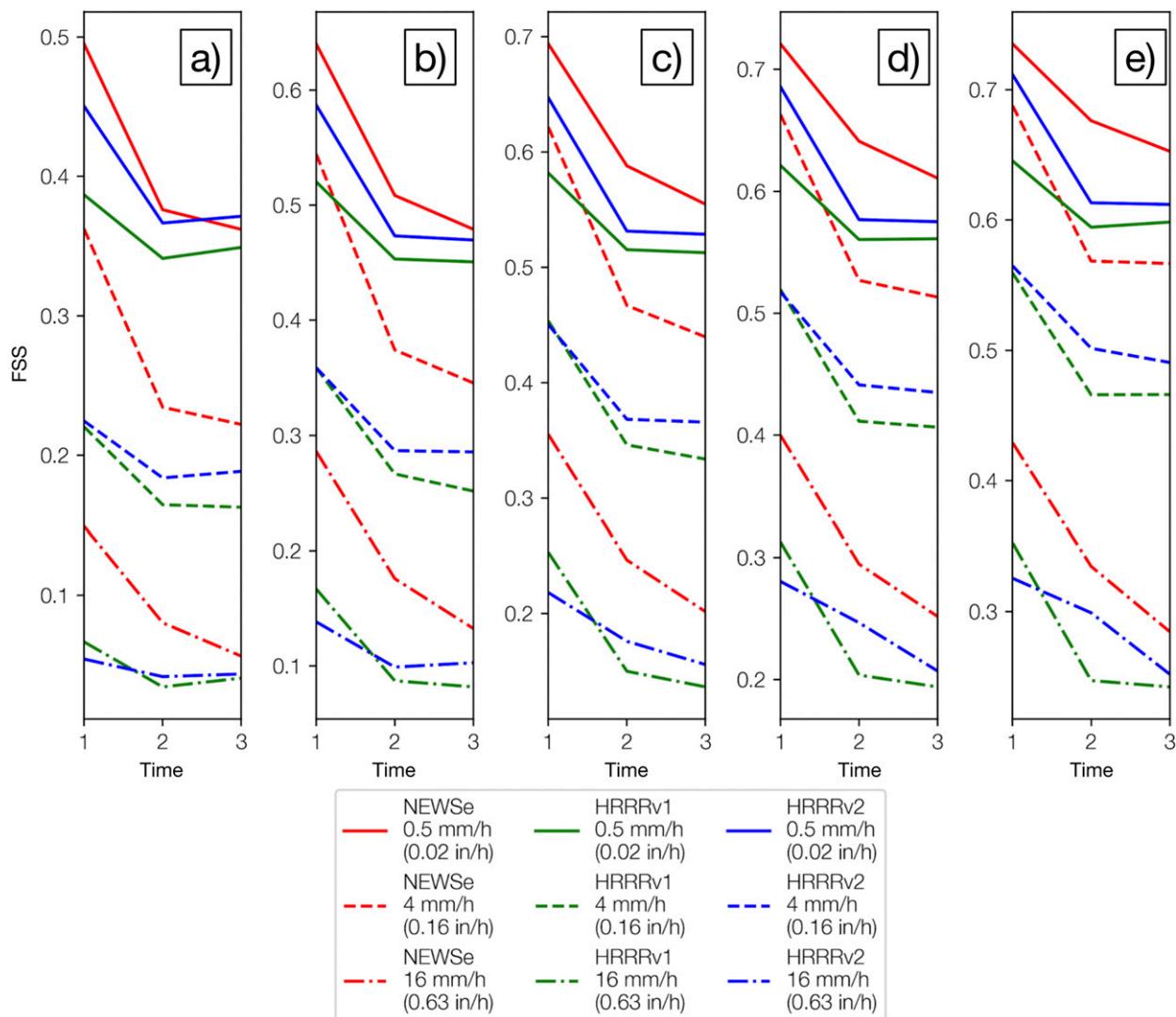
FIG. 1. FSS as a function of precipitation rate (line styles), model (color), and neighborhood size (increasing from the left to right panels). Note the different *y*-axis limits between panels, as FSS asymptotes to unity as the neighborhood size approaches the size of the domain. NEWSe (red lines) is the mean of the FSS values computed deterministically for each member. Higher FSS is better.

HRRR reflectivity assimilation occurs during 1 h, using two low-cost approaches with substantial fine-tuning and parameter adjustments for the latent heating and cloud analysis. In contrast, NEWSe uses an advanced but costlier EnKF technique that benefits from flow-dependent, cross-covariance information for updating unobserved analysis variables (Lorenc 2003). Furthermore, the EnKF assimilation in NEWSe occurs every 15 min during cycling for as long as 9 h.

d. Verification data

Hourly Stage IV precipitation estimates (Lin 2011) were obtained for each case. The Stage IV dataset is a standard for QPF verification, provided at 4.7-km horizontal grid spacing. Stage IV analyses are generated by NCEP from multiple sensor sources and are manually controlled for quality before distribution online.

e. Cases

Table 2 lists the 12 days in 2016 on which the NEWSe was run. Overall, there are 106 forecast times, although not all model forecasts were available for all times. Herein, only forecasts with the first 3 h common to all three datasets are examined, yielding 53 initialization times (i.e., 159 forecast times). For each case, all forecast and verification data were reprojected onto the NEWSe grid. Cases were chosen in real time if the Storm Prediction Center outlook contained at least an enhanced risk of severe weather. Storm modes observed during the cases spanned supercellular, MCSs, and upscale growth.
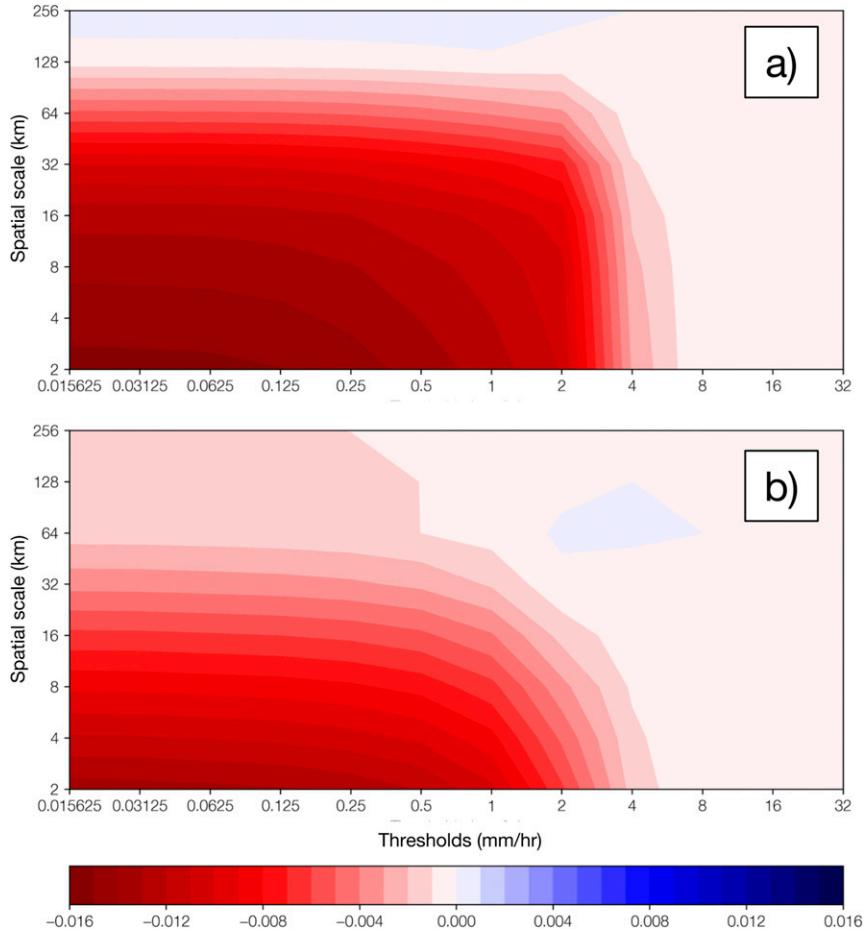
FIG. 2. The differences in 1-h QPFs in MSE between (a) NEWSe and HRRRv2, where red indicates that NEWSe is more skillful and blue indicates that HRRRv2 is more skillful, and (b) HRRRv2 and HRRRv1, where red indicates that HRRRv2 is more skillful and blue indicates that HRRRv1 is more skillful. The x axis depicts increasing precipitation-rate threshold, and the y axis shows increasing spatial scale (note both axes are plotted with a uniform scale). MSE was calculated after Haar wavelet decomposition for each of the three models (averaged over all cases and times, and in the case of NEWSe, over all members). Lower MSE is better.

f. Verification methods

1) TRADITIONAL SCORES

The skill of deterministic forecasts can be computed according to a 2 × 2 contingency table based on the occurrence and detection of an event. Herein, we evaluate each grid point in a given domain. From this contingency table, we then formulate the scores described below.

The critical success index (CSI; Jolliffe and Stephenson 2003) is the ratio of correct forecasts of a given event a to the total number of forecasts made (b represents forecasted events not observed; c represents observed events not forecasted). CSI represents the hit probability, given that the event was forecast and/or observed. While particularly suited to rare and extreme events, CSI is sensitive to the frequency of a given event, and by nature ignores events neither forecast nor observed d:

$$\mathrm{CSI} = \frac{a}{a+b+c}. \qquad (1)$$

The probability of detection (POD; Jolliffe and Stephenson 2003) is computed thus:

$$\mathrm{POD} = \frac{a}{a+c}. \qquad (2)$$

The success ratio (SR) is similar to the false alarm ratio (FAR), calculated as in Roebber (2009):

$$\mathrm{SR} = 1 - \mathrm{FAR} = 1 - \frac{b}{a+b}. \qquad (3)$$
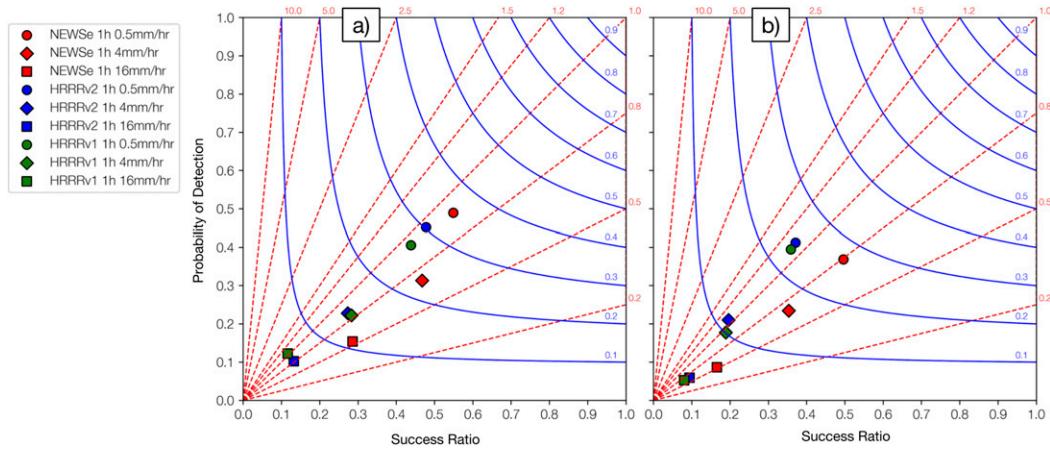
FIG. 3. Performance diagram for (a) 1- and (b) 3-h forecast times for hourly QPFs from NEWSe (red), HRRRv2 (blue), and HRRRv1 (green). The symbols represent different QPF thresholds: $0.5\,\mathrm{mm\,h^{-1}}$ (circles), $4\,\mathrm{mm\,h^{-1}}$ (diamonds), and $16\,\mathrm{mm\,h^{-1}}$ (squares). Dotted red lines show lines of constant BIAS; solid blue lines mark contours of CSI; the x and y axes show increasing success ratio ($1 - $ FAR) and POD, respectively. A perfect forecast lies in the top-right corner of the figure axes. The NEWSe points represent the mean of all members evaluated deterministically for a given lead time.

Finally, the frequency bias (BIAS; Jolliffe and Stephenson 2003) is gauged as

$$\mathrm{BIAS} = \frac{a+b}{a+c}. \qquad (4)$$

2) FRACTIONS SKILL SCORE

The fractions skill score (FSS; Roberts and Lean 2008) is generated by first converting forecast and observed fields to binary arrays, based on the threshold exceedance. Next, for each point, squares of increasing size (neighborhoods) are considered and the fraction of exceedance points in each neighborhood is computed, yielding arrays of fractional coverage centered at each grid point for each neighborhood size. Herein, this is performed using a square uniform filter, with the maximum square length (i.e., neighborhood size) equal to the domain size. Where the filter kernel lies outside of the domain, the fraction field is neglected. For each neighborhood size, the observed and forecast fraction fields are then differenced in a form of mean square error (MSE). This MSE is compared with a reference—essentially the largest possible MSE—to yield the FSS.

3) SCALE-AWARE (CASATI) MSE

We employ a second scale-aware method to evaluate the forecasts, based on a binary decomposition method (Casati et al. 2004), herein referred to as scale-aware MSE. First, forecast and observation fields are converted into binary arrays (based on threshold exceedance) then progressively decomposed

into wavelets with respect to increasing length scale (Haar wavelet decomposition). The waves represent the mean error and variation around the mean and are combined to yield a scale-aware, threshold-dependent MSE score.

3. Results

a. General results (QPFs)

We first present results from the verification of 1-h quantitative precipitation forecasts (QPFs). FSS was calculated over a range of thresholds and neighborhood sizes for all NEWSe members (i.e., all members are treated like deterministic forecasts) and both HRRR configurations. We compare performance through the mean FSS for each system (Figs. 1 and 2). We see that NEWSe offers higher skill at all QPF thresholds than both HRRRv1 and HRRRv2, though this advantage diminishes with lead time. This advantage for the NEWSe members likely stems from its more frequent, longer cycling EnKF radar data assimilation. According to the Student's paired t test (not shown), the advantage of NEWSe over both HRRR models is statistically significant at the 95% confidence level except at the largest thresholds (small sample size) and neighborhood sizes (where the advantage of NEWSe radar data assimilation is lost), and after 2 h of simulation time (same as for the largest neighborhood sizes).

The differences in scale-aware MSEs between NEWSe and HRRRv2, and the two HRRR models, are presented in Fig 2, showing the maximum error at the
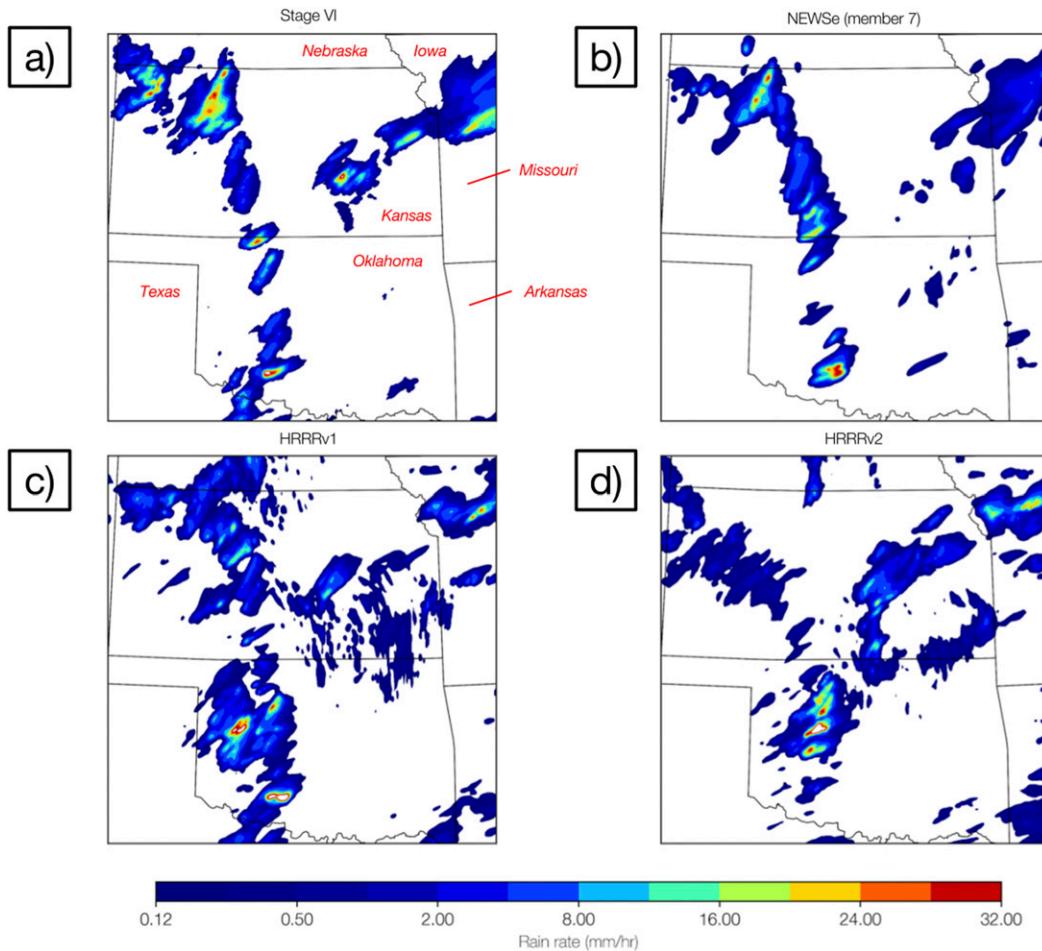
FIG. 4. One-hour precipitation valid at 0000 UTC 9 May 2016 of (a) Stage IV observations, (b) the NEWSe member closest to the mean (member 7), (c) HRRRv1, and (d) HRRRv2. Forecasts in (b)–(d) are 2-h lead times (i.e., initialized at 2200 UTC 8 May 2016). Rainfall is contoured according to the key; rates higher than 32 mm h^{-1} are colored white. States in (a) are labeled in red; some are referenced in the text.

smallest scales and thresholds. For a given threshold and scale, the error is largest in HRRRv1, decreases in HRRRv2, and is lowest in NEWSe. This result is constant for all 3-hourly time periods (not shown); again, the results are significant except at large scales and thresholds and in the 2–3-h period (not shown).

In Fig. 3, we combine CSI with BIAS, POD, and SR into a performance diagram (Roebber 2009) for all three configurations at 1 and 3 h, at three 1-h accumulated QPF thresholds (0.5, 4, and 16 mm h^{-1}). The NEWSe members are again treated deterministically. At 1 h (Fig. 3a), while the NEWSe forecasts have considerably higher CSI, POD, and SR values than both HRRR systems, they also have relatively large negative biases. HRRRv2 provides a more skillful forecast at the low threshold (0.5 mm h^{-1}) than HRRRv1 by all four measures. By 3 h (Fig. 3b), forecasts are less skillful than at

1 h, as expected by error-growth considerations and the reduced impact of radar-data assimilation. At the low threshold (0.5 mm h^{-1}), NEWSe has better SR, but worse POD, than the HRRR configurations. However, at high thresholds, NEWSe forecasts still score better in SR, POD, and CSI. NEWSe forecasts across all times and thresholds underestimate precipitation coverage (negative bias), while HRRR configurations are closer to the optimal bias of unity.

b. Representative example

These objective results are corroborated by a representative case presented in Fig. 4, depicting 2-h lead-time hourly QPFs valid at 0000 UTC 9 May 2016, from both HRRR models and the closest NEWSe member to the mean, along with corresponding Stage IV estimates. The NEWSe member reasonably captures the series of

observed QPF maxima extending in a broken line from northwestern Kansas to southwestern Oklahoma (Figs. 4a,b); this is likely due to the assimilation of storms associated with these maxima into NEWSe members at initialization. In contrast, a second precipitation maximum in southeastern Kansas is poorly represented in the NEWSe member (Fig. 4b) because it only began to form near the initialization time. Both HRRR models generate excessive precipitation in Oklahoma and too little in Kansas.

4. Synthesis and conclusions

We have demonstrated an increase in short-term QPF skill associated with advanced data assimilation methods currently being tested in NEWSe, a WoF prototype forecast system. The increase in skill is shown by comparing QPFs from NEWSe members to two different configurations of the HRRR model: HRRRv1, operational at the time, and HRRRv2, experimental for this study but later operational. The NEWSe prognostic model is nearly identical to this now-operational version of the HRRR, so the differences in skill between NEWSe members and either HRRR configuration are primarily a reflection of the advanced data assimilation methods currently being developed and tested in the 2016 prototype WoF system. These methods are computationally intensive and thus not practicable for current implementation in the CONUS-scale HRRR model, but may be viable for regional-scale domains currently envisioned for early WoF implementations.

Positive differences in QPF skill between NEWSe members and either of the HRRR configurations are largest (i.e., most improvement) near the start of the NEWSe 3-h forecast period and approach zero by its end; differences are maximized for length scales less than 100 km. The now-operational configuration of the HRRR exhibits substantially better skill than the older operational version over the same 3-h time period.

These results are encouraging for ongoing developments in storm-scale modeling and data assimilation. Specifically, they indicate that HRRR developers are making measurable progress with their WRF-ARW-based modeling system and that the advanced data assimilation methods envisioned for WoF implementations substantially enhance storm-scale predictive capabilities on 0–3-h time scales. These development efforts benefit greatly from close collaboration between NSSL (where the WoF project is anchored) and ESRL/GSD (where HRRR development and implementation is a priority). For example, NSSL scientists consult with GSD collaborators to maintain near-operational configurations of the HRRR model in the experimental NEWSe system, and GSD developers

entrain promising WoF-based data assimilation technologies into their complementary research, development, and implementation efforts. The two groups are also collaborating on the design of convection-allowing/resolving EPS systems. The success of these systems depends heavily on design strategies that are still the subject of considerable debate in the meteorological community.

REFERENCES

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

Casati, B., G. Ross, and D. B. Stephenson, 2004: A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteor. Appl.*, **11**, 141–154, https://doi.org/10.1017/S1350482704001239.

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

Gao, J., C. Fu, D. J. Stensrud, and J. S. Kain, 2016: OSSEs for an ensemble 3DVAR data assimilation system with radar observations of convective storms. *J. Atmos. Sci.*, **73**, 2403–2426, https://doi.org/10.1175/JAS-D-15-0311.1.

Jolliffe, I. T., and D. B. Stephenson, 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science.* John Wiley and Sons, 254 pp.

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

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

Karstens, C. D., and Coauthors, 2015: Evaluation of a probabilistic forecasting methodology for severe convective weather in the 2014 Hazardous Weather Testbed. *Wea. Forecasting*, **30**, 1551–1570, https://doi.org/10.1175/WAF-D-14-00163.1.

Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP Global

Data Assimilation System. *Wea. Forecasting*, **24**, 1691–1705, https://doi.org/10.1175/2009WAF2222201.1.

Lin, Y., 2011: GCIP/EOP surface: Precipitation NCEP/EMC 4KM Gridded Data (GRIB) Stage IV data. Earth Observing Laboratory, NCAR–UCAR, accessed 1 May 2017, https://data.eol.ucar.edu/dataset/21.093.

Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP—A comparison with 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **129**, 3183–3203, https://doi.org/10.1256/qj.02.132.

Potvin, C. K., and L. J. Wicker, 2012: Comparison between dual-Doppler and EnKF storm-scale wind analyses: Observing system simulation experiments with a supercell thunderstorm. *Mon. Wea. Rev.*, **140**, 3972–3991, https://doi.org/10.1175/MWR-D-12-00044.1.

——, E. M. Murillo, M. L. Flora, and D. M. Wheatley, 2017: Sensitivity of supercell simulations to initial-condition resolution. *J. Atmos. Sci.*, **74**, 5–26, https://doi.org/10.1175/JAS-D-16-0098.1.

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

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

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

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

——, and Coauthors, 2009: Convective-scale warn-on-forecast system: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1499, https://doi.org/10.1175/2009BAMS2795.1.

Weygandt, S. S., S. G. Benjamin, T. G. Smirnova, and J. M. Brown, 2008: Assimilation of radar reflectivity data using a diabatic digital filter within the Rapid Update Cycle. *12th Conf. on IOAS–AOLS*, New Orleans, LA, Amer. Meteor. Soc., 8.4, https://ams.confex.com/ams/88Annual/techprogram/paper_134081.htm.

Wheatley, D. M., N. Yussouf, and D. J. Stensrud, 2014: Ensemble Kalman filter analyses and forecasts of a severe mesoscale convective system using different choices of microphysics schemes. *Mon. Wea. Rev.*, **142**, 3243–3263, https://doi.org/10.1175/MWR-D-13-00260.1.

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

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

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

——, D. C. Dowell, L. J. Wicker, K. H. Knopfmeier, and D. M. Wheatley, 2015: Storm-scale data assimilation and ensemble forecasts for the 27 April 2011 severe weather outbreak in Alabama. *Mon. Wea. Rev.*, **143**, 3044–3066, https://doi.org/10.1175/MWR-D-14-00268.1.

——, J. S. Kain, and A. J. Clark, 2016: Short-term probabilistic forecasts of the 31 May 2013 Oklahoma tornado and flash flood event using a continuous-update-cycle storm-scale ensemble system. *Wea. Forecasting*, **31**, 957–983, https://doi.org/10.1175/WAF-D-15-0160.1.