Toward Unifying Short-Term and Next-Day Convection-Allowing Ensemble Forecast Systems with a Continuously Cycling 3-km Ensemble Kalman Filter over the Entire Conterminous United States

CRAIG S. SCHWARTZ,[a] GLEN S. ROMINE,[a] AND DAVID C. DOWELL[b]

[a] *National Center for Atmospheric Research, Boulder, Colorado*
[b] *NOAA/Earth System Research Laboratory, Boulder, Colorado*

(Manuscript received 2 July 2020, in final form 13 October 2020)

ABSTRACT: Using the Weather Research and Forecasting Model, 80-member ensemble Kalman filter (EnKF) analyses with 3-km horizontal grid spacing were produced over the entire conterminous United States (CONUS) for 4 weeks using 1-h continuous cycling. For comparison, similarly configured EnKF analyses with 15-km horizontal grid spacing were also produced. At 0000 UTC, 15- and 3-km EnKF analyses initialized 36-h, 3-km, 10-member ensemble forecasts that were verified with a focus on precipitation. Additionally, forecasts were initialized from operational Global Ensemble Forecast System (GEFS) initial conditions (ICs) and experimental ''blended'' ICs produced by combining large scales from GEFS ICs with small scales from EnKF analyses using a low-pass filter. The EnKFs had stable climates with generally small biases, and precipitation forecasts initialized from 3-km EnKF analyses were more skillful and reliable than those initialized from downscaled GEFS and 15-km EnKF ICs through 12–18 and 6–12 h, respectively. Conversely, after 18 h, GEFS-initialized precipitation forecasts were better than EnKF-initialized precipitation forecasts. Blended 3-km ICs reflected the respective strengths of both GEFS and high-resolution EnKF ICs and yielded the best performance considering all times: blended 3-km ICs led to short-term forecasts with similar or better skill and reliability than those initialized from unblended 3-km EnKF analyses and ~18–36-h forecasts possessing comparable quality as GEFS-initialized forecasts. This work likely represents the first time a convection-allowing EnKF has been continuously cycled over a region as large as the entire CONUS, and results suggest blending high-resolution EnKF analyses with low-resolution global fields can potentially unify short-term and next-day convection-allowing ensemble forecast systems under a common framework.

KEYWORDS: Data assimilation; Ensembles; Model evaluation/performance; Numerical weather prediction/forecasting

1. Introduction

Convection-allowing ensembles (CAEs) produce better precipitation and severe weather forecasts than coarser-resolution, convection-parameterizing ensembles (e.g., Clark et al. 2009; Duc et al. 2013; Iyer et al. 2016; Schellander-Gorgas et al. 2017), are operational at many weather forecasting offices (e.g., Gebhardt et al. 2011; Peralta et al. 2012; Hagelin et al. 2017; Raynaud and Bouttier 2017; Jirak et al. 2018; Klasa et al. 2018), and have proven useful and valuable for various meteorological applications around the world (e.g., Xue et al. 2007; Clark et al. 2012; Evans et al. 2014; Maurer et al. 2017; Zhang 2018; Cafaro et al. 2019; Porson et al. 2019; Schwartz et al. 2019). Thus, as computing power has increased, CAE domains have gradually enlarged, with operational global CAEs on the horizon.

While CAEs can be initialized by downscaling coarser-resolution, convection-parameterizing analyses, convection-allowing numerical weather prediction (NWP) models are typically best when initialized from corresponding convection-allowing analyses, particularly for short-term forecasts (e.g., Ancell 2012; Harnisch and Keil 2015; Johnson et al. 2015; Johnson and Wang 2016; Raynaud and Bouttier 2016; Schwartz 2016; Gustafsson et al. 2018). Therefore, to produce the best possible CAE forecasts over ever-expanding domains, convection-allowing data assimilation (DA) systems over large areas are needed to provide optimal initial conditions (ICs).

However, there are obstacles to implementing convection-allowing DA systems over domains large enough to resolve mesoalpha- to synoptic-scale features, especially when using state-of-the-science ensemble-based DA algorithms like the ensemble Kalman filter (EnKF; Evensen 1994; Houtekamer and Zhang 2016), which produces flow-dependent analysis ensembles and has become popular for initializing CAEs (e.g., Jones and Stensrud 2012; Melhauser and Zhang 2012; Schumacher and Clark 2014; Schwartz et al. 2014, 2015a,b, 2019). One challenge is simply computational expense, which grows directly with domain size,[1] and accordingly, most convection-allowing EnKFs and their associated CAE forecasts have relatively small domains centered on a single European country (e.g., Schraff et al. 2016; COSMO 2020) or a small portion of the conterminous United States (CONUS). For example, NOAA's experimental ''Warn-on-Forecast'' (WoF; Stensrud et al. 2009, 2013) system, initialized from 36-member 3-km EnKF analyses, covers less than

[1] Mixed-resolution DA systems (e.g., Gao and Xue 2008; Rainwater and Hunt 2013; Li et al. 2015) possessing both convection-allowing and convection-parameterizing resolution components can lessen costs and make large-domain convection-allowing analyses more feasible (e.g., Schwartz 2016; Rogers et al. 2017).

Corresponding author: Craig Schwartz, schwartz@ucar.edu

DOI: 10.1175/WAF-D-20-0110.1

1000 km × 1000 km (Wheatley et al. 2015; Jones et al. 2016, 2018, 2020; Skinner et al. 2018).

Fortunately, computing challenges can be overcome with increased resources, and recently, several studies initialized CAE forecasts from 40-member EnKF analyses with 3-km or finer horizontal grid spacing over the entire CONUS (Duda et al. 2019; Gasperoni et al. 2020; Johnson et al. 2020). Similarly, NOAA's real-time, experimental High-Resolution Rapid Refresh Ensemble (HRRRE) is initialized from CONUS-spanning, 3-km, 36-member EnKF analyses (Dowell et al. 2016; Ladwig et al. 2018). However, 36–40-member EnKFs are likely smaller than desirable, considering that operational global EnKFs run by the United States and Canada, respectively, have 80 and 256 members, and generally, EnKFs benefit from larger ensembles (e.g., Zhang et al. 2013; Houtekamer et al. 2014).

But, even with unlimited resources, there are fundamental scientific concerns that must be addressed to develop stable, high-quality, convection-allowing EnKFs over large regional domains, especially in continuously cycling limited-area EnKFs where external models are relegated to providing boundary conditions. In particular, model physics deficiencies can lead to accumulation of biases throughout EnKF DA cycles, potentially degrading analysis system performance and subsequent forecasts (e.g., Torn and Davis 2012; Romine et al. 2013; Cavallo et al. 2016; Wong et al. 2020). Although all continuously cycling limited-area EnKFs are prone to bias accumulation, this issue may be exacerbated as both model resolution and domain size increase: biases may accumulate more in high-resolution EnKFs than low-resolution EnKFs because of rapid small-scale error growth (e.g., Lorenz 1969; Zhang et al. 2003; Hohenegger and Schär 2007; Judt 2018), and EnKFs over large domains may suffer from bias accumulations more than EnKFs over small domains because of reduced influence from lateral boundaries provided by potentially less biased global models (e.g., Warner et al. 1997; Romine et al. 2014; Schumacher and Clark 2014).

Given these scientific and computing challenges, operational convection-allowing continuously cycling EnKFs and attendant CAEs over Europe have small domains (e.g., Schraff et al. 2016; COSMO 2020), while large-domain convection-allowing EnKFs over the CONUS (e.g., Duda et al. 2019; Gasperoni et al. 2020; Johnson et al. 2020; HRRRE) employ "partial cycling" strategies that periodically discard convection-allowing analysis cycles and replace them with coarser-resolution, large-scale external analyses in hopes of tempering bias accumulations (e.g., Hsiao et al. 2012; Benjamin et al. 2016; Wu et al. 2017). This partial cycling approach over the CONUS seems justified, as Schwartz et al. (2020) showed that a limited-area continuously cycling EnKF with convection-parameterizing resolution did not initialize better CAE precipitation forecasts over the CONUS than downscaled global analyses.

Nonetheless, as discussed at length by Schwartz et al. (2019), continuously cycling EnKFs have many attractive properties for CAE initialization, including the ability to diagnose model biases while simultaneously producing flow-dependent ICs that are dynamically consistent with and span all possible resolvable scales of the convection-allowing forecast model. Thus, despite formidable challenges, it is desirable to further explore and develop continuously cycling EnKFs over large geographic areas at convection-allowing resolutions for CAE initialization purposes.

Accordingly, we produced continuously cycling, 80-member, 3-km EnKF analyses with a 1-h cycling period for 4 weeks over a computational domain spanning the entire CONUS. EnKF analysis ensembles then initialized 36-h, 3-km, 10-member CAE forecasts. For comparison, 3-km CAE forecasts were also initialized by downscaling both 15-km EnKF analyses and global ICs produced for NCEP's operational Global Ensemble Forecast System (GEFS; Zhou et al. 2017). The impact of assimilating radar observations into the 3-km EnKF was also assessed. Relative to the EnKF described in Schwartz et al. (2020), our EnKFs used more advanced observation processing, an upgraded NWP model, and a shorter cycling period, and inclusion of 3-km EnKF DA was also new. To our knowledge, this work presents the first time convection-allowing continuously cycling EnKF analyses have been produced over the entire CONUS.

Results indicated benefits of EnKF-initialized forecasts with respect to GEFS-initialized forecasts diminished with forecast length, presumably because large-scale fields were better represented in GEFS ICs and became more important at longer forecast ranges. These findings motivated experimentation with a "blending" approach combining large-scale fields from an external (e.g., global) NWP model with small-scale fields from a limited-area model, which can be achieved by augmenting a variational cost function with a global model constraint (e.g., Guidard and Fischer 2008; Dahlgren and Gustafsson 2012; Vendrasco et al. 2016; Keresturi et al. 2019) or using filtering to perform scale separation (e.g., Yang 2005; Wang et al. 2011; Caron 2013; H. Wang et al. 2014; Y. Wang et al. 2014; Hsiao et al. 2015; Zhang et al. 2015; Feng et al. 2020); we used a low-pass filter to combine large scales from GEFS ICs with small scales from EnKF analyses. These previous studies collectively suggested blended limited-area ICs improved forecasts compared to those initialized from unblended limited-area ICs, including for a CAE within a perturbed-observation variational DA framework (Keresturi et al. 2019). However, our application of blending within the context of a large-domain convection-allowing continuously cycling EnKF was unique, and, as described below, blending global fields with high-resolution EnKF analyses can potentially unite short-term and next-day (18–36-h) CAE forecast systems under a common framework.

2. Model configurations, EnKF settings, and experimental design

a. Forecast model

All forecasts were produced by version 3.9.1.1 of the Advanced Research Weather Research and Forecasting (WRF) Model (Skamarock et al. 2008; Powers et al. 2017) over a nested computational domain (Fig. 1a). The horizontal grid spacing was 15 km in the outer domain and 3 km in the nest, and time steps were 60 and 12 s in the 15- and 3-km

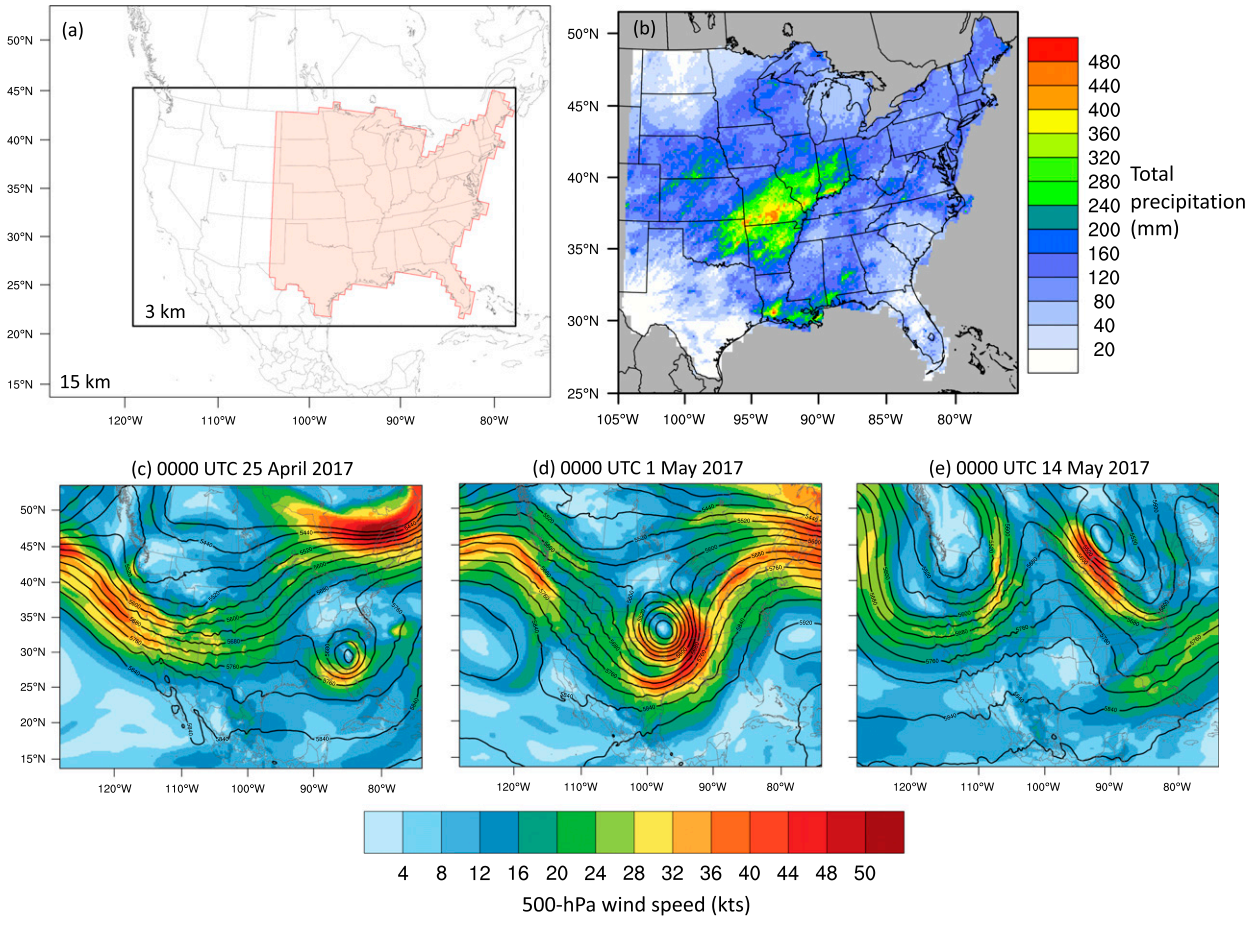

FIG. 1. (a) Computational domain. Horizontal grid spacing was 15 km in the outer domain (415 × 325 points) and 3 km in the nest (1581 × 986 points). Objective precipitation verification only occurred over the red shaded region of the 3-km domain (CONUS east of 105°W). (b) Total accumulated Stage IV (ST4) precipitation (mm) over the verification region between 0000 UTC 25 Apr and 1200 UTC 21 May 2017, which encompasses all possible valid times of the 36-h forecasts. (c)–(e) The 500-hPa wind speed (shaded; kt; 1 kt ≈ 0.51 m s^{-1}) and height (m; contoured every 40 m) from Global Forecast System analyses valid at 0000 UTC (c) 25 Apr, (d) 1 May, and (e) 14 May 2017.

domains, respectively. Both domains had 51 vertical levels distributed as in the Rapid Refresh model (Benjamin et al. 2016) with a 15-hPa top. Physical parameterizations were identical across the two domains (Table 1), except no cumulus parameterization was employed on the convection-allowing 3-km grid, and all ensemble members used common physics and dynamics options.

b. *EnKF DA systems*

1) ENKF EXPERIMENTS AND CONFIGURATIONS

Two primary DA experiments with 80-member ensembles were performed using an ensemble adjustment Kalman filter (Anderson 2001, 2003; Anderson and Collins 2007), a type of EnKF, as implemented in the Data Assimilation Research Testbed

TABLE 1. Physical parameterizations for all WRF Model forecasts. Cumulus parameterization was only used on the 15-km domain.

Physical parameterization	WRF Model option	References
Microphysics	Thompson	Thompson et al. (2008)
Longwave and shortwave radiation	Rapid Radiative Transfer Model for Global Climate Models (RRTMG) with ozone and aerosol climatologies	Mlawer et al. (1997); Iacono et al. (2008); Tegen et al. (1997)
Planetary boundary layer	Mellor–Yamada–Janjić (MYJ)	Mellor and Yamada (1982); Janjić (1994, 2002)
Land surface model	Noah	Chen and Dudhia (2001)
Cumulus parameterization	Tiedtke (15-km domain only)	Tiedtke (1989); Zhang et al. (2011)

TABLE 2. Summary of EnKF configurations.

Parameter	15-km EnKF	3-km EnKF
Ensemble size	80 members	
Updated WRF Model variables	Zonal and meridional wind components; perturbation geopotential height, potential temperature, and dry surface pressure; and water vapor, graupel, snow, and rain mixing ratios	
Localization function	Eq. (4.10) from Gaspari and Cohn (1999)	
Horizontal localization full width	1280 km	640 km, except 1280 km for rawinsonde observations
Vertical localization full width	1.0 scale height	
Inflation method	Posterior relaxation-to-prior-spread [RTPS; Whitaker and Hamill (2012)]	
Inflation factor (α)	1.06	
Sampling error correction	Anderson (2012)	
Horizontal thinning for aircraft and satellite-tracked wind observations	30 km	15 km
Vertical thinning for aircraft and satellite-tracked wind observations	25 hPa	

(DART; Anderson et al. 2009) software. The first EnKF experiment only produced analyses on the 15-km domain (Fig. 1a), and the 3-km domain was removed during WRF Model advances between EnKF analyses. Conversely, the second EnKF experiment produced separate, independent analyses on *both* the 15- and 3-km domains, with nested WRF Model forecasts between EnKF analyses. During these nested forecasts, which were ~45 times more expensive than the single-domain 15-km model advances, one-way feedback was employed such that the 15-km EnKF DA system was unaffected by the 3-km EnKF DA system (i.e., 15-km fields in the nested- and single-domain EnKF DA systems were identical), permitting a clean comparison of analysis and forecast sensitivity to EnKF resolution. The 15- and 3-km EnKFs updated identical state variables (Table 2), with hydrometers included in anticipation of experimentation with radar DA (section 4c).

Initial 80-member ensembles were produced by interpolating the 0.25° NCEP Global Forecast System (GFS) analysis at 0000 UTC 23 April 2017 onto the 15-km domain and adding random, correlated, Gaussian noise with zero mean (e.g., Barker 2005; Torn et al. 2006) drawn from background error covariances provided by the WRF Model's DA system (Barker et al. 2012). The randomly produced 15-km ensemble was then downscaled onto the 3-km grid to initialize the 3-km EnKF, ensuring initial 15- and 3-km ensembles were identical aside from interpolation errors. These randomly generated ensembles served as prior (before assimilation) ensembles for the first EnKF analyses, and the posterior (after assimilation) ensembles at 0000 UTC 23 April 2017 initialized 1-h, 80-member ensemble forecasts that became prior ensembles for the next EnKF analyses at 0100 UTC 23 April 2017. Analysis–forecast cycles with a 1-h period continued until 0000 UTC 20 May 2017 (649 total DA cycles). This experimental period (23 April–20 May 2017) was similar to that in Schwartz (2019), which featured several heavy precipitation episodes primarily driven by strong synoptic forcing, a broad overall precipitation maximum centered in Missouri (Fig. 1b), and a variety of flow patterns (Figs. 1c–e).

During EnKF cycles, soil states freely evolved for each member, sea surface temperature was updated daily from NCEP's 0.12° analyses (e.g., Gemmill et al. 2007), and identical randomly perturbed lateral boundary conditions (LBCs) were applied to the 15-km domain in each DA system, with perturbations for individual members generated using the same method to produce initial ensembles at 0000 UTC 23 April 2017. The first two days of cycling were regarded as spinup.

Spurious correlations due to sampling error were mitigated with a sampling error correction scheme (Anderson 2012) and covariance localization [Eq. (4.10) of Gaspari and Cohn (1999)]. Vertical localization limited analysis increments to ±1.0 scale height (in log pressure coordinates) away from an observation in both the 15- and 3-km EnKFs. However, horizontal localizations differed depending on EnKF resolution: 15-km EnKF analysis increments were forced to zero 1280 km from an observation, but to lessen expense and complete 3-km EnKF analyses quickly enough for operational applications, 3-km EnKF analysis increments were forced to zero 640 km from an observation, except rawinsonde observations could produce increments up to 1280 km away (Table 2). The vertical and 15-km EnKF horizontal localization distances were guided by previous experiences with DART (e.g., Romine et al. 2013, 2014; Schwartz et al. 2015a,b, 2019), and while our 3-km EnKF horizontal localization distances were similar to Johnson et al. (2015), they were larger than those in many other convection-allowing EnKFs (e.g., Harnisch and Keil 2015; Yussouf et al. 2015, 2016; Degelia et al. 2018; Gasperoni et al. 2020; Jones et al. 2020). However, these studies with smaller localization distances either used partial cycling strategies or only continuously cycled for a short period (days), and we believed that larger localization distances were necessary to provide stronger observational constraints in a large-domain continuously cycling 3-km EnKF.

EnKF spread was maintained by applying covariance inflation to posterior state-space perturbations about the ensemble mean following Whitaker and Hamill (2012)'s "relaxation-to-prior spread" algorithm with an inflation parameter $\alpha = 1.06$ in both the 15- and 3-km EnKFs. As noted by Schwartz and

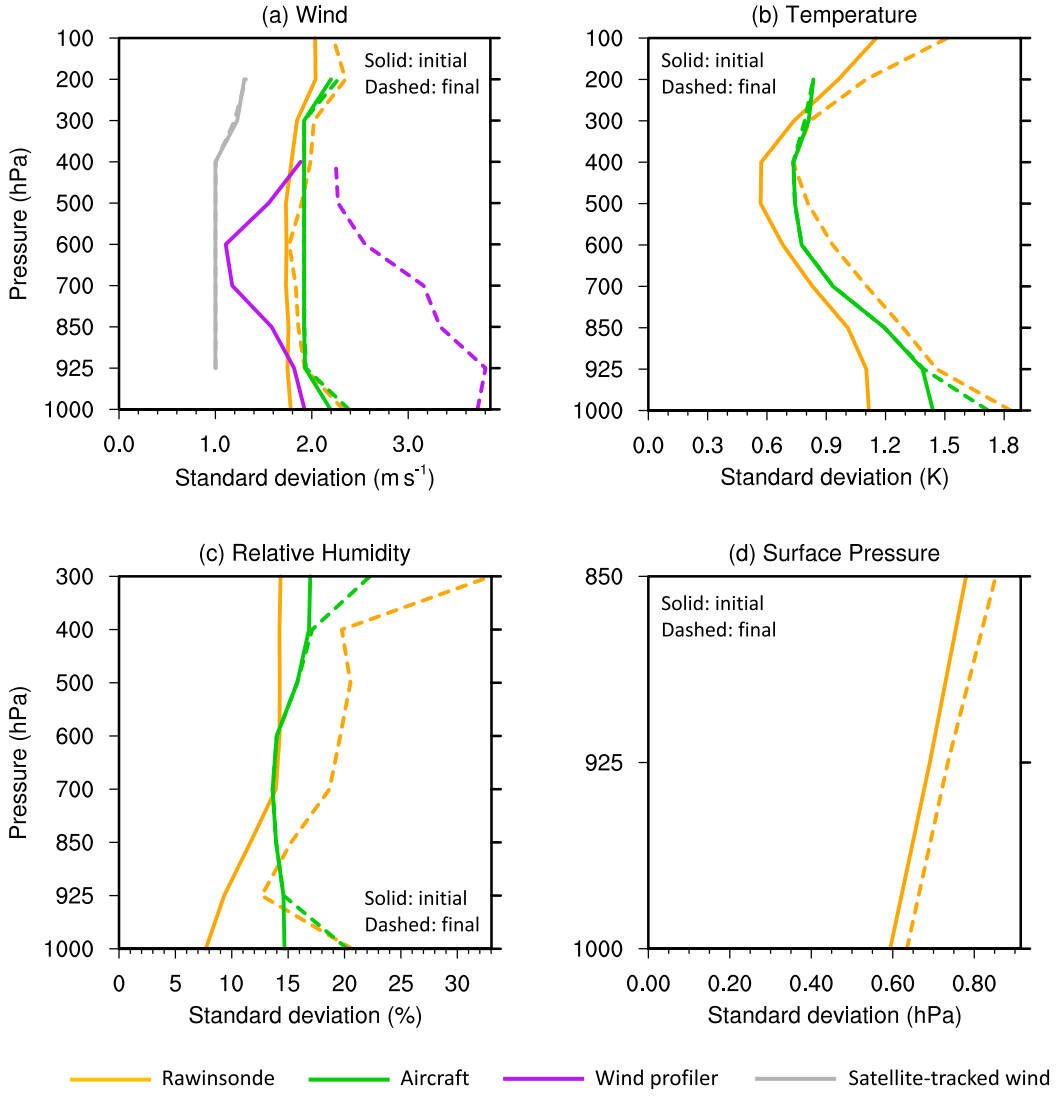

FIG. 2. Initially specified (solid lines) and final (after GSI adjustment; dashed lines) observation error standard deviations as a function of pressure for (a) wind ($m\,s^{-1}$), (b) temperature (K), (c) relative humidity (%), and (d) surface pressure (hPa) observations with vertically varying errors averaged over all observations assimilated between 0000 UTC 25 Apr and 0000 UTC 20 May 2017 (inclusive) by both the 15- and 3-km EnKFs. If a particular observation type was not assimilated at a certain pressure level, no value is plotted.

Liu (2014), $\alpha > 1$ meant inflated posterior spread was greater than prior spread, which, while counterintuitive, was necessary to maintain reasonable spread given absence of other spread-inducing methods like multiphysics ensembles, additive inflation, or stochastic physics. Several iterative weeklong trials with 15-km EnKFs were performed to settle on $\alpha = 1.06$, which provided acceptable prior observation-space statistics for the assumed observation errors (section 3).

2) OBSERVATIONS

Although DART has observation processing capabilities, we instead used NCEP's operational Gridpoint Statistical Interpolation (GSI) DA system (Kleist et al. 2009; Shao et al. 2016) for observation processing, which, relative to DART, has more sophisticated quality control, observation thinning, and observation error assignment capabilities. In addition, GSI's observation operators were used instead of DART's built-in observation operators to produce model-simulated conventional observations. Initially specified observation errors were based on the HRRRE and identical in the 15- and 3-km EnKFs (Fig. 2; Table 3); GSI adjusted these errors to produce "final" observation error standard deviations σ_o actually used in the assimilation, as described by several texts (e.g., Schwartz and Liu 2014; Developmental Testbed Center 2016; Johnson and Wang 2017). These adjustments often inflated initially specified observation errors (Fig. 2).

Time windows for the observation platforms varied and were based on Rapid Refresh model (Benjamin et al. 2016) and

TABLE 3. Conventional observations that were assimilated and their outlier check thresholds, time windows, and initially specified observation error standard deviations.

Observing platform	Observation type	Initial observation error	Outlier check threshold (a)	Time window (h)
Rawinsonde	Surface pressure	Fig. 2d	5	1.5
	Temperature	Fig. 2b	7	1.5
	Relative humidity	Fig. 2c	7	1.5
	Wind	Fig. 2a	10	1.5
Aircraft	Temperature	Fig. 2b	7	0.75
	Relative humidity	Fig. 2c	7	0.75
	Wind	Fig. 2a	10	0.75
Wind profiler	Wind	Fig. 2a	5	0.4
Global positioning system radio occultation (GPSRO)	Refractivity	1% of observation value	10	3.0
Infrared and water vapor channel satellite-tracked wind	Wind	Fig. 2a	2.5	1.5
Ship and buoy	Surface pressure	0.44 hPa	5	1.5
	Temperature	0.8 K	7	1.5
	Relative humidity	3.9%	7	1.5
	Wind	1.45 m s^{-1}	5	1.5
SYNOP and METAR	Surface pressure	0.54 hPa	5	0.25
	Temperature	2.3 K	5	0.25
	Relative humidity	3.4%	7	0.25
	Wind	1.2 m s^{-1}	5	0.25
Oklahoma and West Texas mesonet	Surface pressure	0.35 hPa	5	0.1
	Temperature	1.5 K	5	0.1
	Relative humidity	4%	7	0.1
	Wind	1.1 m s^{-1}	5	0.1

HRRRE settings, with generally smaller windows for frequently reporting, stationary platforms, like METAR observations (Table 3), and all observations were assumed valid at the analysis time. Moisture observations were initially processed as specific humidity, but because GSI requires moisture observation errors in terms of relative humidity, moisture observations were ultimately converted to and assimilated as relative humidity using the prior ensemble mean saturation specific humidity. Satellite-tracked wind and aircraft observations were thinned such that remaining observations were spaced 25 hPa apart vertically and 30 and 15 km apart horizontally in the 15- and 3-km EnKFs, respectively (Table 2); these different horizontal thinnings were chosen so the 15- and 3-km EnKFs had equal numbers of satellite-tracked wind and aircraft observations within their respective horizontal localization radii. Radiance observations were not assimilated since they generally yield small impacts over the CONUS (Lin et al. 2017) given the multitude of available conventional observations. Additionally, the EnKFs did not assimilate radar observations, although an auxiliary experiment was performed where radar observations were assimilated with a 3-km EnKF (section 4c).

Observations were subject to numerous quality control procedures, such as excluding observations from specific aircraft with known biases and applying an "outlier check" to reject observations whose ensemble mean innovations[2] were $>a\sigma_o$, where a varied from 2.5 to 10 depending on observation type and platform (Table 3). These a were generally fairly lenient and allowed most observations to pass the outlier check, which, along with our relatively large localization distances, reflected a philosophy that we wanted observations to heavily constrain the 1-h WRF Model forecasts between EnKF analyses. Overall, the EnKFs assimilated 30 000–100 000 conventional observations each cycle, with a relative dearth of overnight observations due to fewer commercial flights and maxima at 0000 and 1200 UTC reflecting the majority of rawinsonde launches (Fig. 3). Ultimately, GSI-provided observations, final observation errors, and prior model-simulated observations for each ensemble member were ingested directly into DART for use in EnKF DA.

3) FORECAST INITIALIZATION

EnKF analysis ensembles initialized 36-h 10-member ensemble forecasts over the nested computational domain (Fig. 1a) at 0000 UTC between 25 April and 20 May 2017 (inclusive; 26 forecasts). Although 80 EnKF analysis members were available, due to computing constraints, 36-h forecasts were only initialized from members 1–10; 10-member CAEs are sufficient to provide skillful and valuable probabilistic forecasts (e.g., Clark et al. 2011, 2018; Schwartz et al. 2014) and similar in size as the HRRRE and NCEP's operational High-Resolution Ensemble Forecast system (Jirak et al. 2018). Choosing members 1–10 was effectively the same as randomly selecting 10 members since all ensemble members had identical configurations (e.g., Schwartz et al. 2014). In principle, free forecasts could have been initialized every hour, but given finite resources, forecasts were solely initialized at 0000 UTC,

[2] The "innovation" is the difference between an observation and the prior model-simulated observation.

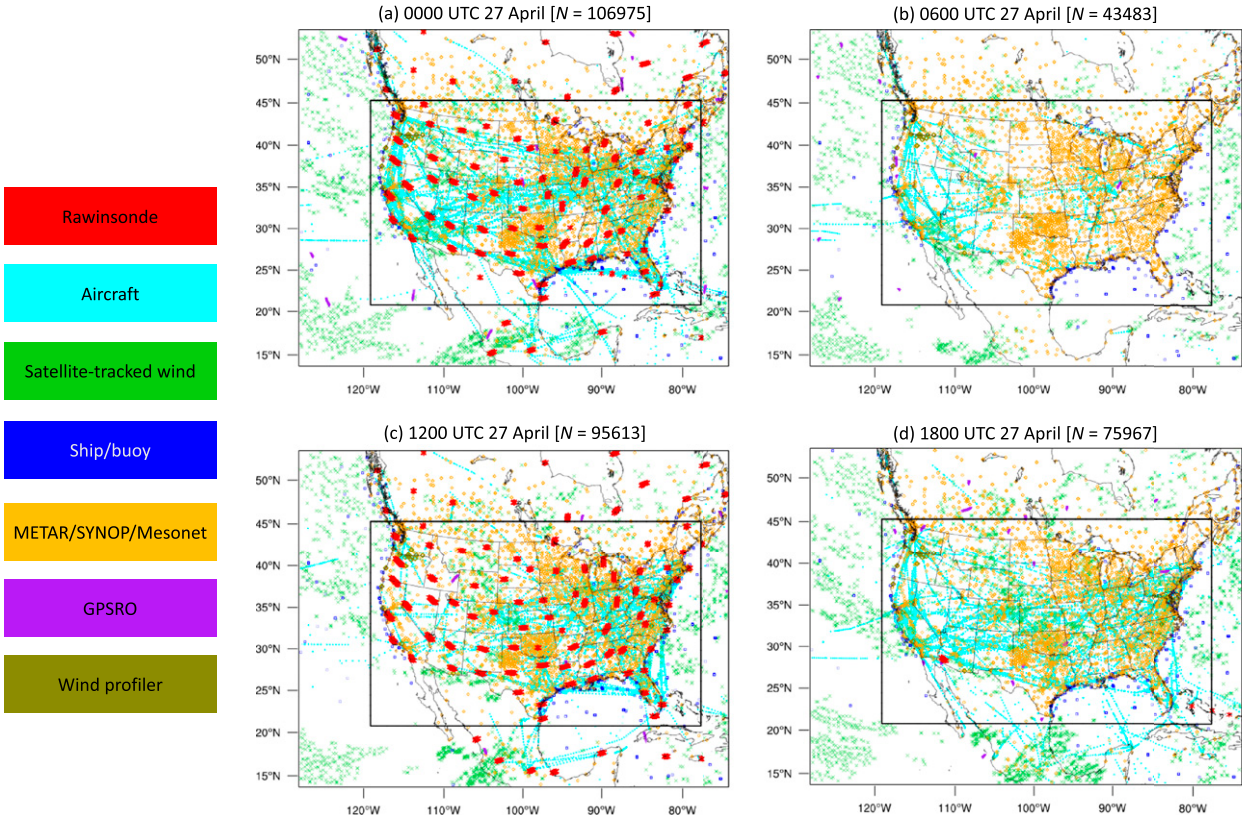

FIG. 3. Computational domain overlaid with observations assimilated by the 15-km EnKF during the (a) 0000, (b) 0600, (c) 1200, and (d) 1800 UTC 27 Apr 2017 analyses. Values of *N* in the headers indicate the number of assimilated observations. The inner box represents bounds of the 3-km domain; most observations located within the 3-km domain were also assimilated by the 3-km EnKF at these times.

which allowed us to focus on both short-term and next-day forecast periods featuring active convection.

When initializing 36-h forecasts from 15-km EnKF analyses, the 3-km nest was initialized by downscaling 15-km EnKF analyses onto the 3-km grid. Conversely, downscaling was unnecessary to initialize 36-h forecasts from the 3-km EnKF; 3-km ICs were simply 3-km EnKF analysis members. For both sets of EnKF-initialized 36-h forecasts, perturbation members 1–10 from the GEFS (Zhou et al. 2017) with 0.5° horizontal grid spacing provided LBCs at 3-h intervals for the 15-km domain, which in turn provided LBCs for the 3-km nest. While random LBCs could have been used for the 36-h forecasts as in the EnKF DA system, we believed it was more appropriate to use flow-dependent LBCs for these longer unconstrained forecasts.

c. Benchmark ensemble

To serve as a benchmark for the EnKF-initialized CAE forecasts, 36-h forecasts on the nested grid (Fig. 1a) with the configurations in section 2a were initialized by interpolating 0.5° ICs from perturbation members 1–10 of the GEFS onto the computational domain at 0000 UTC daily between 25 April and 20 May 2017 (inclusive), with LBCs provided by GEFS forecasts identically as in the EnKF-initialized CAEs. As described by Zhou et al. (2017), GEFS ICs were produced by adding 6-h forecast perturbations from a global EnKF DA system (Whitaker and Hamill 2002) to ''hybrid'' variational-ensemble analyses produced for NCEP's deterministic GFS (e.g., Wang and Lei 2014; Kleist and Ide 2015a,b). Relative to the limited-area EnKF analyses, GEFS ICs were much coarser but reflected assimilation of many more observations, including satellite radiances. Overall, comparison of GEFS- and EnKF-initialized CAE forecasts provides insight about whether the vastly more expensive EnKF initialization procedure was warranted.

d. Blending

Based on performance of the EnKF- and GEFS-initialized CAE forecasts (section 4b), additional ensemble ICs were created by ''blending'' small scales from EnKF analyses with large scales from GEFS ICs. Blending was solely performed at 0000 UTC between 25 April and 20 May 2017 (inclusive) immediately after EnKF DA and before initializing 36-h CAE forecasts; blending was not employed within the context of continuously cycling EnKF DA, as the blended 0000 UTC fields were not used to initialize 1-h WRF Model forecasts that served as priors for the next DA cycle.

Specifically, ICs from corresponding GEFS and EnKF ensemble members were blended on both the 15- and 3-km domains[3] to create new initial ensembles using

$$x^i_{\text{blend}} = (\text{EnKF}_i - \text{EnKF}_{\text{FILT},i}) + \text{GEFS}_{\text{FILT},i}, \quad (1)$$

where x^i_{blend} represents the blended ICs for the ith ensemble member, EnKF_i is the EnKF analysis for the ith member, and $\text{EnKF}_{\text{FILT},i}$ and $\text{GEFS}_{\text{FILT},i}$ are the low-pass filtered EnKF and GEFS ICs for the ith member, respectively, for $i = 1, \ldots, 10$. To perform the scale separation, a low-pass, sixth-order implicit tangent filter (e.g., Raymond 1988; Raymond and Garder 1991) as implemented by several studies (e.g., Yang 2005; H. Wang et al. 2014; Hsiao et al. 2015; Feng et al. 2020) and given by

$$H(L) = [1 + \tan^{-6}(\pi\Delta x/L_x)\tan^6(\pi\Delta x/L)]^{-1} \quad (2)$$

was employed (Fig. 4), where Δx is the horizontal grid spacing (either 15 or 3 km), L the wavelength, $H(L)$ the scale-dependent response function, and L_x a specified filter cutoff (km) physically representing the spatial scale (wavelength) where the blended ICs (e.g., x^i_{blend}) had equal contributions from GEFS and EnKF initial states [i.e., when $L = L_x$, $H(L) = 0.5$]. Blending was applied at all 51 vertical levels to zonal and meridional wind components; perturbation geopotential height, potential temperature, and dry surface pressure; and water vapor mixing ratio, and the cutoff length was height and variable invariant.

We produced blended ICs using filter cutoff lengths L_x of 640, 960, and 1280 km, guided by EnKF horizontal localization lengths and previous work suggesting values between 640 and 1280 km were appropriate (e.g., H. Wang et al. 2014; Hsiao et al. 2015; Feng et al. 2020). CAE forecasts initialized from these three sets of blended ICs objectively had similar skill, although $L_x = 960$ km yielded slightly better results. Therefore, results are shown only for the 960-km cutoff.

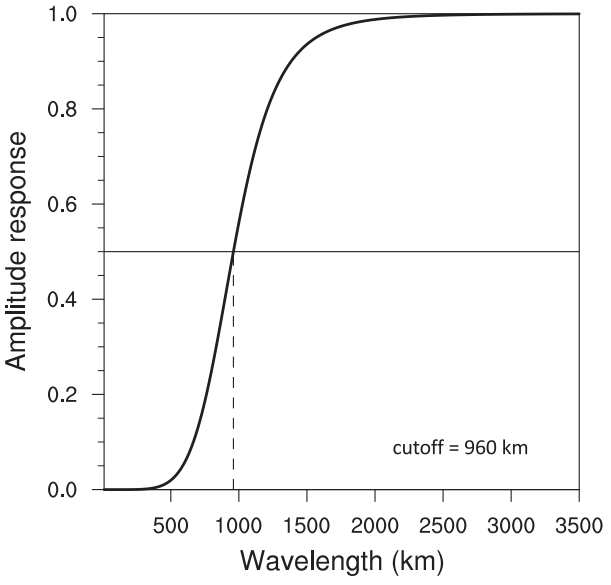

FIG. 4. Amplitude response (y axis) of a sixth-order implicit tangent filter as a function of wavelength (km) for a specified cutoff length of 960 km. In the context of this study, the curve denotes the contribution of GEFS ICs to blended ICs at a given wavelength (e.g., for wavelengths where the amplitude response is 1, 100% of the blended ICs at those wavelengths were from the GEFS). The dashed vertical and solid horizontal lines illustrate how the amplitude response is 0.5 at the cutoff length.

3. EnKF performance

To assess EnKF performance, we examined the observation-space bias and relationship between the prior ensemble mean root-mean-square error (RMSE) and "total spread," the square root of the sum of the observation error variance σ^2_o and ensemble variance of the simulated observations (Houtekamer et al. 2005). Ideally, the ratio of total spread to RMSE [termed the consistency ratio (CR; Dowell and Wicker 2009)] should be near 1.0. To fairly compare the 15- and 3-km EnKFs, we restricted this analysis solely to those observations assimilated by both EnKFs, although overall findings were unchanged when computing identical statistics with inhomogeneous samples. We focused on aircraft and rawinsonde observations because of their large impacts on springtime forecasts over the CONUS (James and Benjamin 2017).

Ensemble mean additive biases (model minus observations) and RMSEs aggregated over all prior ensembles (1-h forecasts) between 0000 UTC 25 April and 0000 UTC 20 May 2017 (inclusive) were similar in the 15- and 3-km EnKFs with respect to zonal wind and temperature observations at most levels (Figs. 5a,b,d,e), while biases and RMSEs for moisture were typically smaller in the 3-km EnKF (Figs. 5c,f). Magnitudes of temperature biases were typically <0.1 K, except near the surface and in the upper troposphere for rawinsonde observations (Fig. 5a); the latter is consistent with other continuously cycling EnKFs over the CONUS (e.g., Romine et al. 2013; Schumacher and Clark 2014; Schwartz and Liu 2014; Cavallo et al. 2016; Schwartz 2016) and likely due to closer fits to the more numerous aircraft observations that may have systematically warm biases compared to rawinsonde observations (Ballish and Krishna Kumar 2008). That upper-tropospheric temperature biases relative to aircraft observations (Fig. 5d) were smaller than and opposite the sign of temperature biases relative to rawinsonde observations (Fig. 5a) further supports this reasoning.

Prior total spreads were similar in both EnKFs (Fig. 5) and CRs were usually between 0.8 and 1.2, although CRs suggest

[3] It was unclear whether blending should be performed on just the 3-km domain or on both the 15- and 3-km domains. While the former perhaps enables a fairer comparison between forecasts initialized from blended and unblended 3-km ICs, the latter maintains consistency across both domains that intuitively seems desirable. So, we experimented with both scenarios, which yielded remarkably similar 36-h forecasts. Thus, forecast impacts of blending were due to changes in 3-km ICs and not attributable to modified LBCs for the 3-km domain provided by 15-km forecasts. All results regarding blending are for the scenario where blending occurred on both the 15- and 3-km domains.

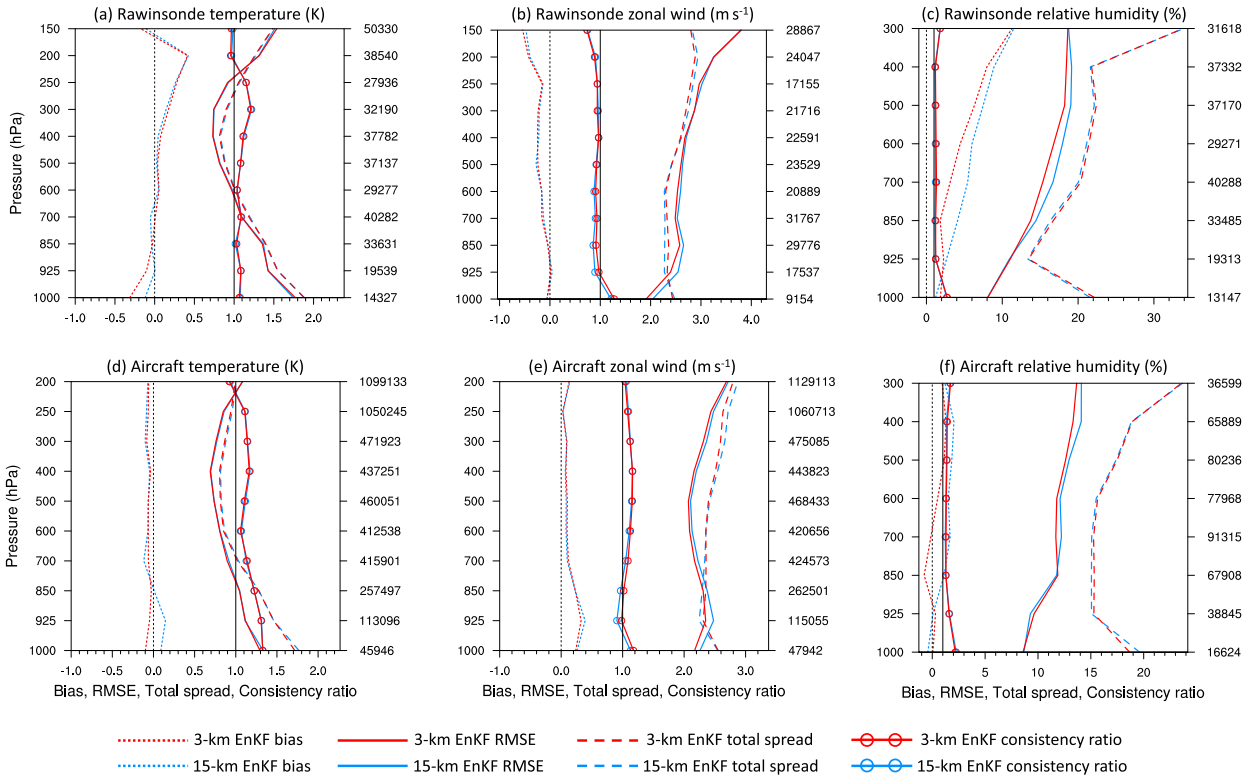

FIG. 5. Ensemble mean additive bias (model minus observations; short-dashed lines), ensemble mean RMSE (solid lines), total spread (long-dashed lines), and consistency ratio (CR; solid lines with circles) for (a) rawinsonde temperature (K), (b) rawinsonde zonal wind ($m\,s^{-1}$), (c) rawinsonde relative humidity (%), (d) aircraft temperature (K), (e) aircraft zonal wind ($m\,s^{-1}$), and (f) aircraft relative humidity (%) observations aggregated over all prior ensembles (1-h forecasts) between 0000 UTC 25 Apr and 0000 UTC 20 May 2017 (inclusive). These statistics were computed for those observations assimilated by both the 15- and 3-km EnKFs. Sample size at each pressure level is shown at the right of each panel. Vertical lines at $x = 0$ and $x = 1$ are references for biases and CRs, respectively.

moisture observation errors could potentially be decreased. While more spread may have been expected in the 3-km EnKF because small-scale errors grow rapidly upscale (e.g., Lorenz 1969; Zhang et al. 2003; Hohenegger and Schär 2007), cumulus parameterization in the 15-km DA system may have served as an error source that compensated for missing storm-scale structures, and assimilating copious observations each cycle (Fig. 3) with fairly large localization distances highly constrained the 15- and 3-km EnKFs, limiting spread growth during 1-h WRF Model integrations between analyses. In balance, these factors potentially contributed to the similar 15- and 3-km prior spreads.

Overall, systematic biases were usually small and EnKF performance appeared acceptable. Moreover, after the first two days, prior total spread and ensemble mean biases were steady throughout the cycles (Fig. 6), and observation rejection rates varied little with time (not shown). These results indicate the continuously cycling EnKFs maintained stable climates, which is particularly noteworthy for the 3-km EnKF, as it has not previously been demonstrated that a convection-allowing EnKF can be continuously cycled over a large domain without deleterious consequences like a drifting model climate or filter divergence [see appendix A of Houtekamer and Zhang (2016) for a succinct summary of filter divergence].

4. Precipitation forecast verification

Hourly accumulated precipitation forecasts were verified against Stage IV (ST4) analyses (Lin and Mitchell 2005) produced at NCEP considered as "truth." Objective evaluations were performed over the CONUS east of 105°W (hereafter the "verification region"; Fig. 1a), where ST4 analyses were most robust (e.g., Nelson et al. 2016). For metrics requiring a common grid for forecasts and observations, we used a budget algorithm (e.g., Accadia et al. 2003) to interpolate forecast precipitation to the ST4 grid (4.763-km horizontal grid spacing). Otherwise, metrics were computed from native grid output.

The following statistics were aggregated over all twenty-six 3-km forecasts initialized at 0000 UTC.

a. Precipitation climatologies

To assess precipitation climatologies, aggregate domain-total precipitation per grid point and fractional coverages of 1-h accumulated precipitation meeting or exceeding various

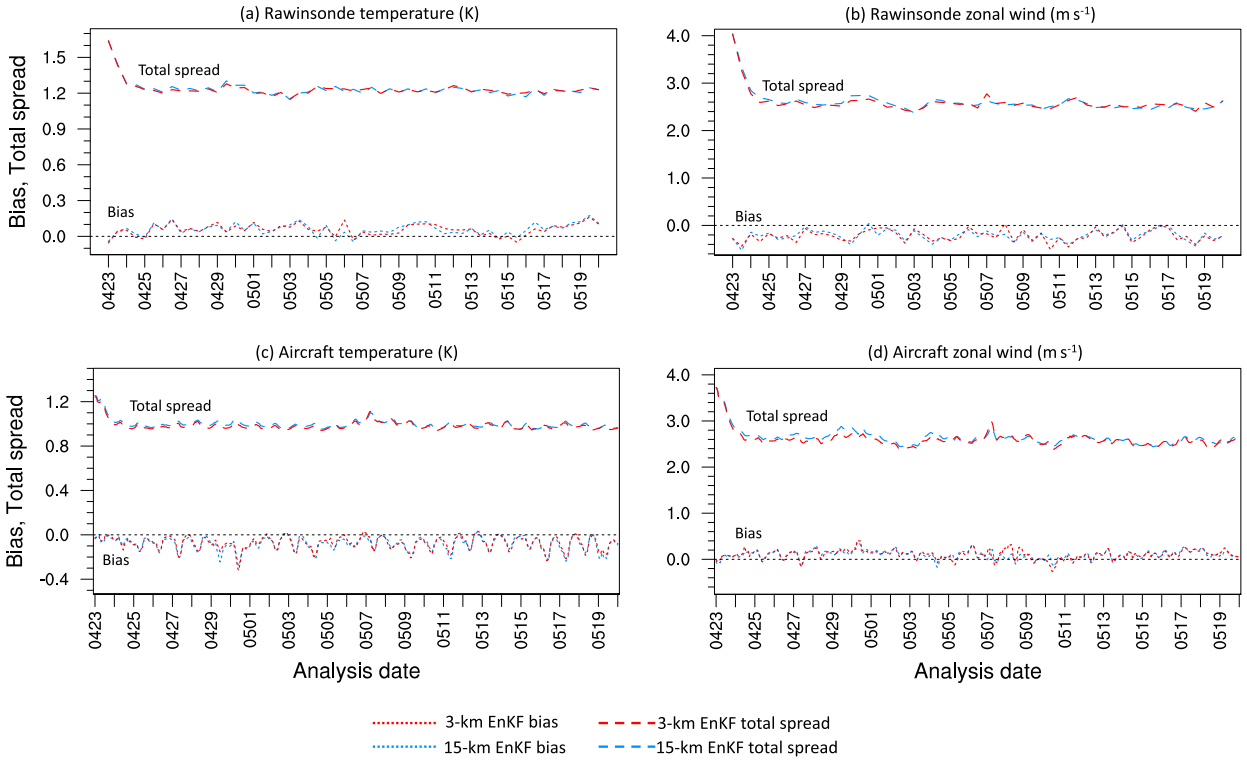

FIG. 6. Prior (1-h forecast) total spread (long-dashed lines) and ensemble mean additive bias (model minus observations; short-dashed lines) for (a) rawinsonde temperature (K), (b) rawinsonde zonal wind ($m\,s^{-1}$), (c) aircraft temperature (K), and (d) aircraft zonal wind ($m\,s^{-1}$) observations between 150 and 1000 hPa as a function of time. In (c) and (d) values are plotted every hour between 0000 UTC 23 Apr and 0000 UTC 20 May 2017 (inclusive) and smoothed with a 6-h running average, while in (a) and (b) values are plotted every 12 h between 0000 UTC 23 Apr and 0000 UTC 20 May 2017 (inclusive) without smoothing. These statistics were computed for those observations assimilated by both the 15- and 3-km EnKFs. The *x*-axis labels represent 0000 UTC for a specific month and day in 2017 (e.g., the marker for "0511" denotes 0000 UTC 11 May 2017). Dashed lines at $y = 0$ are for reference.

accumulation thresholds (e.g., $2.5\,mm\,h^{-1}$) were calculated on native grids over the verification region. Additionally, spatial patterns of total precipitation over all 26 forecasts were examined, which were similar in the various ensembles and generally agreed with observations (e.g., Fig. 1b), including the southwest–northeast-oriented maximum across Missouri and adjacent areas. Although magnitudes of these maxima differed across the ensembles, these differences were manifested by the following domain-average statistics, so spatial variations of precipitation climatologies are not discussed further.

1) IMPACT OF ANALYSIS RESOLUTION

Differences between ensembles were largest over the first 12 h, when GEFS-initialized forecasts were spinning up precipitation from coarse 0.5° ICs. While this spinup meant GEFS-initialized forecasts underpredicted total precipitation (Fig. 7) and areal coverages (Fig. 8) over the first 5 h, ultimately, the spinup process yielded too much 6–12-h total precipitation and excessive coverages $\geq 2.5\,mm\,h^{-1}$. Forecasts initialized from 15-km EnKF analyses also overpredicted total precipitation over the first 12 h, accompanied by excessive coverages for thresholds $\geq 5.0\,mm\,h^{-1}$.

Overall, forecasts initialized from unblended 3-km EnKF analyses had precipitation climatologies best matching observations through 12 h, but there were shortcomings. For example, although at 1 h, unblended 3-km EnKF analyses produced forecasts with areal coverages closest to observations (Fig. 8), coverages rapidly decreased between 2 and 3 h and were further from those observed between 2 and 12 h for the 1.0 and $2.5\,mm\,h^{-1}$ thresholds (Figs. 8a,b) compared to forecasts with 15-km or blended 3-km ICs, suggesting poor maintenance of stratiform precipitation regions after initialization. However, forecasts with unblended 3-km ICs had 6–12-h areal coverages at the $5.0\,mm\,h^{-1}$ threshold well-matching observations (Fig. 8c) and 2–6-h coverages at the 10.0–$50.0\,mm\,h^{-1}$ thresholds closer to observations than forecasts with GEFS and 15-km EnKF ICs (Figs. 8d–f). Furthermore, 2–12-h domain-total precipitation was clearly best in forecasts with unblended 3-km ICs (Fig. 7).

Despite differences between the ensembles through 12 h, domain-total precipitation and areal coverages were broadly similar between 18 and 36 h, with too much total precipitation (Fig. 7) and general underprediction and overprediction of areal coverages at the 1.0 and 10.0–$50.0\,mm\,h^{-1}$ thresholds, respectively (Figs. 8a,d–f). Collectively, for precipitation

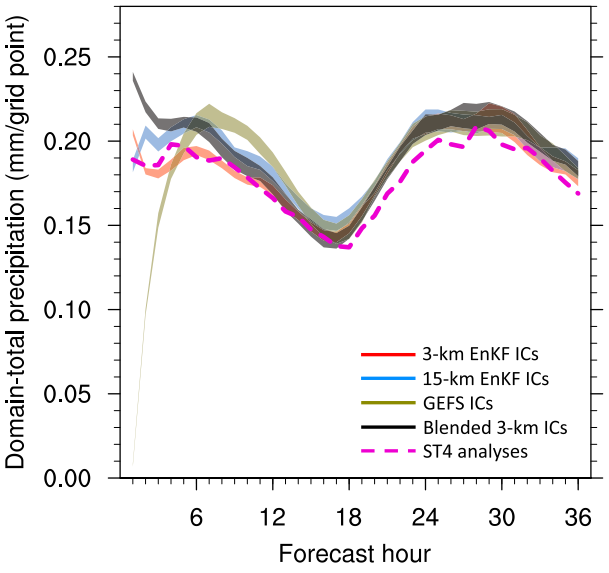

FIG. 7. Average 1-h accumulated precipitation (mm) per grid point over all twenty-six 3-km forecasts and the verification region (CONUS east of 105°W) computed on native grids as a function of forecast hour. Red, blue, gold, and black shadings represent envelopes of the 10 members comprising the ensembles with 3-km EnKF ICs, 15-km EnKF ICs, GEFS ICs, and blended 3-km ICs, respectively, and darker shadings indicate intersections of two or more ensemble envelopes. Values on the *x* axis represent ending forecast hours of 1-h accumulation periods (e.g., an *x*-axis value of 24 is for 1-h accumulated precipitation between 23 and 24 h). ST4 data during the 0–12- and 24–36-h forecast periods were identical except for 1 day (the former included data between 0000 and 1200 UTC 25 Apr–20 May while the latter instead included data between 0000 and 1200 UTC 26 Apr–21 May), and because domain-total ST4 precipitation between 0000–1200 UTC 21 May was much larger than that between 0000 and 1200 UTC 25 Apr, average 24–36-h domain-total ST4 precipitation was greater than average 0–12-h domain-total ST4 precipitation.

climatologies, these findings suggest benefits of convection-allowing analyses relative to convection-parameterizing analyses are primarily confined to short-term forecasts and heavier rainfall rates.

2) IMPACT OF BLENDING

With respect to forecasts initialized from unblended 3-km EnKF analyses, forecasts with blended 3-km ICs (using a 960-km cutoff) had similar 18–36-h areal coverages and total precipitation but higher domain-total precipitation and areal coverages over the first 6–12 h that typically compared worse to observations through 3 h (Figs. 7 and 8). Examination of individual forecasts indicated blended 3-km ICs mostly enhanced 1–3-h forecast precipitation within and near precipitation entities also predicted by forecasts with unblended 3-km ICs and that widespread spurious features did not cause the overprediction. This behavior is illustrated by the forecast initialized at 0000 UTC 1 May 2017, which had the largest difference of domain-total precipitation (e.g., Fig. 7) between member 1 in the CAEs with blended and unblended 3-km ICs across all twenty-six 36-h forecasts (Fig. 9). While both 1–3-h precipitation forecasts had similar spatial patterns, blended ICs led to more numerous cells in places with scattered rainfall, and these additional entities were usually erroneous compared to observations (black and gold circles in Fig. 9). Additionally, within features, the forecast with blended ICs had heavier rainfall maxima than ST4 observations and the forecast with unblended ICs (red circles in Figs. 9b,c,e,f,h,i).

Thus, overall, it appears blending did not improve short-term precipitation climatologies, likely due to imbalances created by blending (e.g., Yang 2005; H. Wang et al. 2014). Additional steps like digital filter initialization (DFI) applied to blended ICs (e.g., Yang 2005) may potentially lessen these imbalances, but DFI could result in spinups that are smoother than desirable for short-term high-resolution NWP model applications.

b. Ensemble precipitation verification

As in many studies, we used percentile thresholds to define events (e.g., the 95th percentile, which selects the top 5% of values), which removes bias and permits a thorough assessment of spatial performance given a model's climate (e.g., Roberts and Lean 2008; Mittermaier and Roberts 2010; Mittermaier et al. 2013; Dey et al. 2014; Gowan et al. 2018; Woodhams et al. 2018; Schwartz 2019). Our application of percentile thresholds exactly followed section 5a(1) of Schwartz (2019), where physical thresholds corresponding to percentile thresholds were obtained separately for observations and each ensemble member on the ST4 grid for each precipitation accumulation interval. These physical thresholds were ultimately used to determine forecast and observed event occurrence. To help interpret subsequent objective statistics, mean physical thresholds corresponding to specific percentile thresholds are provided in Fig. 10. As with areal coverages (Fig. 8), the largest differences among the ensembles' percentiles were over the first 6–12 h.

After interpolating precipitation forecasts to the ST4 grid, a ''neighborhood approach'' (e.g., Theis et al. 2005; Ebert 2008, 2009) was used to produce ''neighborhood ensemble probabilities'' (NEPs; Schwartz et al. 2010; Schwartz and Sobash 2017) that were ultimately verified. In short, NEPs were computed at the *i*th grid point by averaging point-based ensemble probabilities over all grid points within the neighborhood of the *i*th point, which incorporates spatial uncertainty and reflects the inherent inaccuracy of high-resolution NWP models at individual grid points. We produced NEPs for neighborhood length scales *r* between 5 and 200 km, which represented radii of circular neighborhoods. Please see section 2a of Schwartz and Sobash (2017) for more information about constructing and verifying NEPs and Eqs. (1)–(3) in Schwartz (2019), which explicitly describe NEP computation when using percentile thresholds.

Statistical significance testing followed section 5a(3) of Schwartz (2019). Specifically, a pairwise difference bootstrap technique with 10 000 resamples was used to determine whether aggregate differences between two ensembles' statistics were

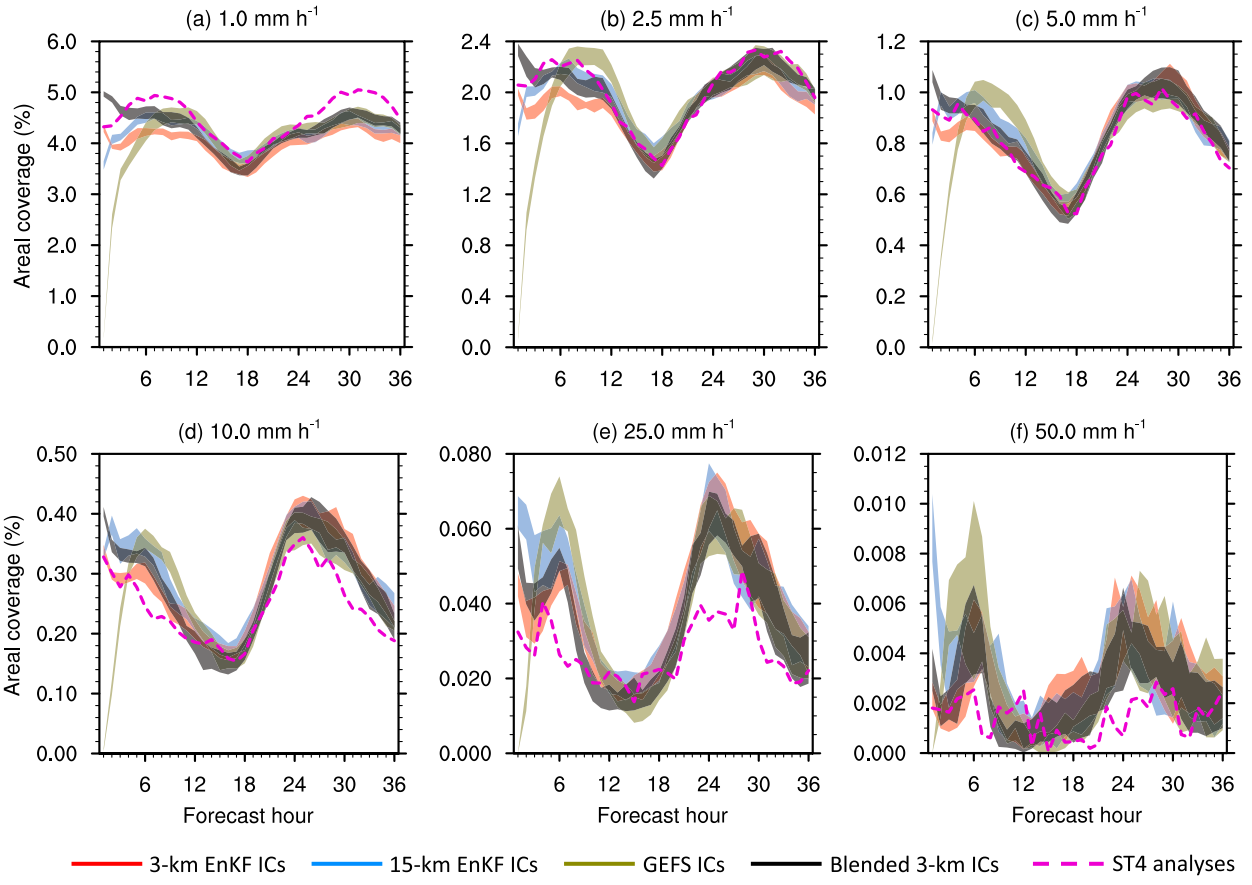

FIG. 8. Fractional areal coverage (%) of 1-h accumulated precipitation meeting or exceeding (a) 1.0, (b) 2.5, (c) 5.0, (d) 10.0, (e) 25.0, and (f) 50.0 mm h^{-1} over the verification region (CONUS east of 105°W), computed on native grids and aggregated over all twenty-six 3-km forecasts as a function of forecast hour. Red, blue, gold, and black shadings represent envelopes of the 10 members comprising the ensembles with 3-km EnKF ICs, 15-km EnKF ICs, GEFS ICs, and blended 3-km ICs, respectively, and darker shadings indicate intersections of two or more ensemble envelopes. Values on the *x* axis represent ending forecast hours of 1-h accumulation periods (e.g., an *x*-axis value of 24 is for 1-h accumulated precipitation between 23 and 24 h).

statistically significant at the 95% level (e.g., Hamill 1999; Wolff et al. 2014).

1) ATTRIBUTES STATISTICS AND RANK HISTOGRAMS

To assess calibration, attributes diagrams (Wilks 2011) were produced with forecast probability bins of 0%–5%, 5%–15%, 15%–25%, . . . , 85%–95%, and 95%–100%; curves on the diagonal indicate perfect reliability. Varying *r* changes sharpness and the resulting NEP distribution (Schwartz and Sobash 2017), which in turn impacts reliability. Over the 1–12- and 18–36-h forecast periods, the smallest *r* yielding near-perfect reliability for any experiment was $r = 90$ km and $r = 125$ km, respectively, so we focus on reliability computed with those *r*.

Over the first 12 h for $r = 90$ km, the ensemble initialized from unblended 3-km EnKF analyses was statistically significantly more reliable than the ensembles initialized from GEFS and 15-km EnKF ICs, with the GEFS-initialized ensemble having the worst reliability (Fig. 11). Conversely, between 18 and 36 h for $r = 125$ km, the GEFS-initialized ensemble was regularly statistically significantly more reliable than the ensembles with unblended 15- and 3-km EnKF ICs, and the ensemble with 15-km ICs usually had comparable or better reliability than the ensemble with unblended 3-km ICs (Fig. 12). Except for the 99.9% threshold, all ensembles had skill with respect to forecasts of sample climatology.

These findings suggest aspects of GEFS ICs were beneficial for next-day (18–36-h) forecasts, which motivated blending GEFS and EnKF initial states. Indeed, blended 3-km ICs led to 18–36-h forecasts with comparable or better reliability as GEFS-initialized forecasts and statistically significantly better reliability than the ensemble with unblended 3-km ICs (Fig. 12). Over the first 12 h, differences between the ensembles with blended and unblended 3-km ICs were also often statistically significant, suggesting that blending can additionally improve short-term forecast reliability (Fig. 11).

Rank histograms (e.g., Hamill 2001) based on domain-total precipitation (e.g., Schwartz et al. 2014, 2020) corroborated

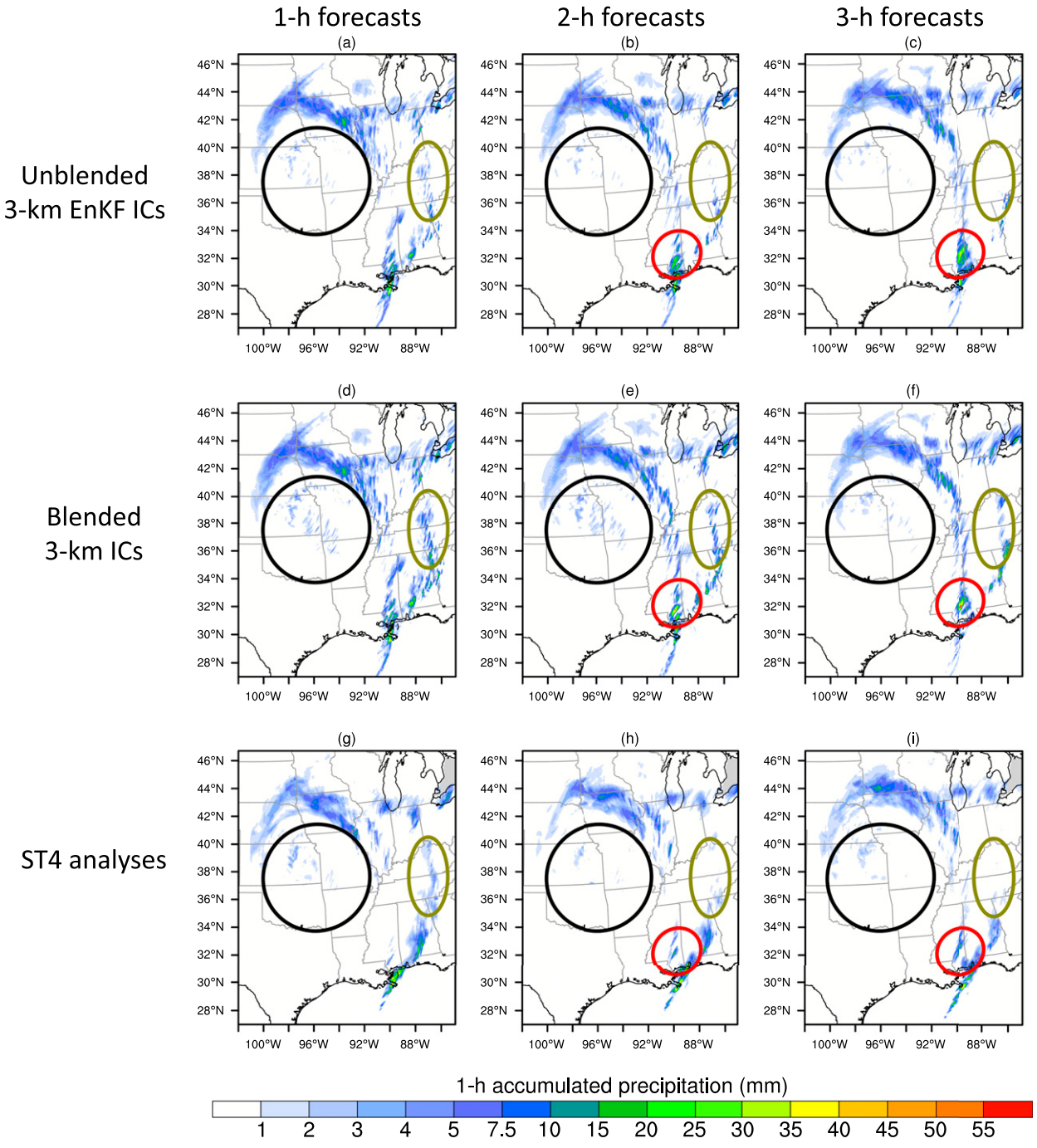

FIG. 9. The 1-h accumulated precipitation (mm) for (left) 1-, (center) 2-, and (right) 3-h forecasts initialized at 0000 UTC 1 May 2017 from member 1 of the 3-km ensembles with (a)–(c) unblended 3-km EnKF ICs and (d)–(f) blended 3-km ICs (using a 960-km cutoff length). (g)–(i) Corresponding ST4 analyses, with gray-shaded areas denoting no data. Annotated circles correspond to features noted in the text.

attributes statistics. Specifically, over the first 12 h, bin counts in the ensemble with unblended 3-km ICs were closer to optimal in most bins compared to those for the ensembles with GEFS and 15-km EnKF ICs (Fig. 13a), which was quantified by the smaller-is-better reliability index (RI; Delle Monache et al. 2006). Blended 3-km ICs yielded slightly lower 1–12-h RIs than unblended 3-km ICs, but the difference was small compared to that between 18 and 36 h (Fig. 13b), where rank histograms and RIs indicated more observations fell within the ensemble and dispersion was improved when GEFS initial states were either

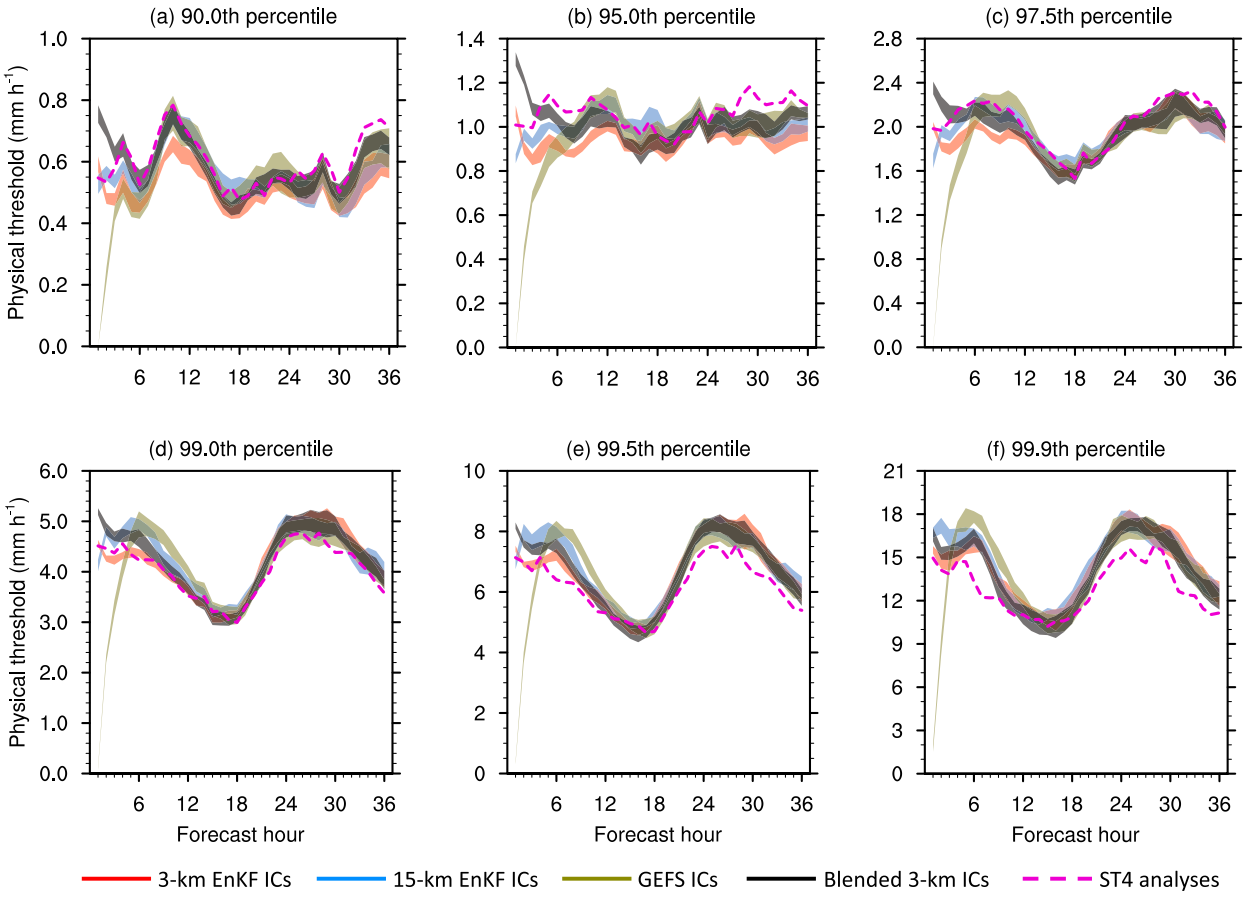

FIG. 10. Average physical thresholds (mm h^{-1}) over all twenty-six 3-km forecasts of 1-h accumulated precipitation corresponding to the (a) 90th, (b) 95th, (c) 97.5th, (d) 99th, (e) 99.5th, and (f) 99.9th percentile thresholds as a function of forecast hour. The physical thresholds were computed separately for each day and 1-h forecast period on the ST4 grid over the verification region (CONUS east of 105°W) and averaged to obtain the *y*-axis values. Red, blue, gold, and black shadings represent envelopes of the 10 members comprising the ensembles with 3-km EnKF ICs, 15-km EnKF ICs, GEFS ICs, and blended 3-km ICs, respectively, and darker shadings indicate intersections of two or more ensemble envelopes. Values on the *x* axis represent ending forecast hours of 1-h accumulation periods (e.g., an *x*-axis value of 24 is for 1-h accumulated precipitation between 23 and 24 h).

used as standalone ICs or combined with 3-km EnKF analyses through blending.

2) SPREAD AND SPECTRA

Improved reliability and rank histograms engendered by GEFS and blended 3-km ICs was associated with increased ensemble spread. In particular, the ensembles with GEFS and blended 3-km ICs had statistically significantly more 24–30-h precipitation spread compared to the ensembles with unblended EnKF ICs (Fig. 14). Additionally, blended 3-km ICs led to significantly more spread than unblended 3-km ICs over the first 6 h that may have improved reliability statistics and rank histograms, even though this enhanced spread reflected excessive early precipitation (e.g., Figs. 7–9). The greater spread through ~18 h in the ensembles with GEFS and 15-km ICs relative to that from the ensemble with unblended 3-km ICs may reflect a substantial contribution from the small, yet intense precipitation entities that were more numerously predicted when forecasts had downscaled, rather than 3-km, ICs (Figs. 8c–f and 10d–f).

To further understand spread characteristics, perturbation power spectra were computed with the discrete cosine transform (Denis et al. 2002), which is well suited for obtaining spectra from limited-area models. Perturbation spectra were determined with respect to the ensemble mean over the entire 3-km domain except for the 15 points nearest each lateral boundary. Final spectra were averaged over all 10 perturbations and 26 forecasts.

At 1 h, 500-hPa perturbation kinetic energy (PKE) in the ensemble with blended 3-km ICs broadly followed PKEs of the GEFS-initialized ensemble at scales > 500 km and the ensemble with unblended 3-km ICs at smaller scales, reflecting the blending procedure (Fig. 15a). Compared to the GEFS-initialized ensemble, the ensemble with unblended 3-km ICs had more 1-h forecast PKE at most scales (Fig. 15a), with enhanced large-scale power possibly reflecting upscale error growth with time through the continuous 3-km DA cycles. But, PKE in the GEFS-initialized ensemble grew fastest between 3 and 6 h (Figs. 15b,c) and was largest at all scales after 6 h (Figs. 15d–f),

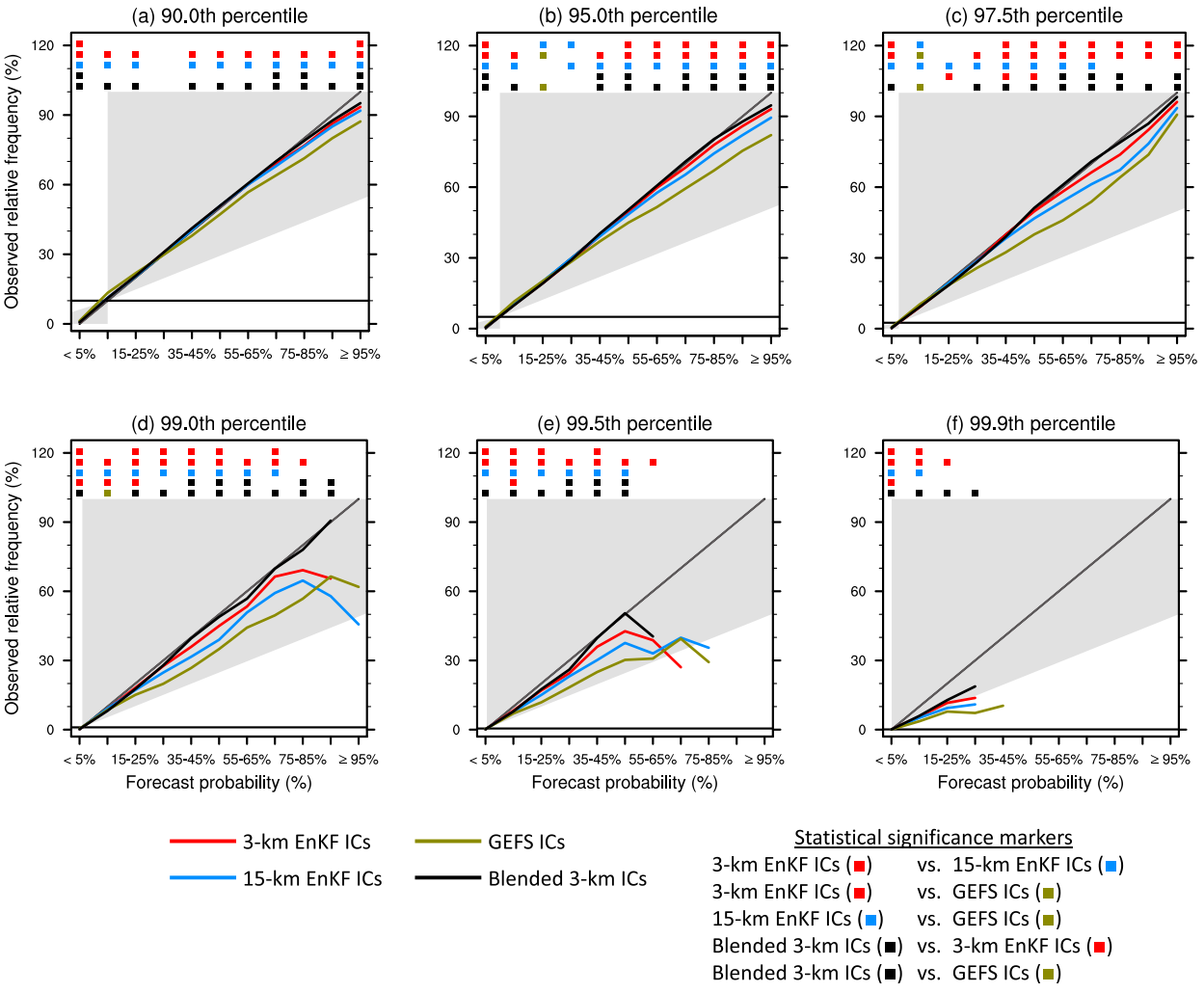

FIG. 11. Attributes statistics computed over the verification region (CONUS east of 105°W) with a 90-km neighborhood length scale aggregated over all twenty-six 1–12-h 3-km forecasts of 1-h accumulated precipitation for the (a) 90th, (b) 95th, (c) 97.5th, (d) 99th, (e) 99.5th, and (f) 99.9th percentile thresholds. Horizontal lines near the x axis represent observed frequencies of the event (sample climatology) and diagonal lines are lines of perfect reliability. Points lying in gray-shaded regions had skill compared to forecasts of sample climatology as measured by the Brier skill score (Brier 1950; Wilks 2011). Values were not plotted for a particular bin if fewer than 500 grid points had forecast probabilities in that bin over the verification region and all 26 forecasts. Symbols along the top axis denote those probability bins where differences between two ensembles were statistically significant at the 95% level, with the five rows of colored symbols corresponding to the five comparisons in the legend to denote which ensemble was statistically significantly closest to perfect reliability. For example, in the top row, red symbols indicate the ensemble with 3-km EnKF ICs had statistically significantly better reliability than the ensemble with 15-km EnKF ICs, while blue symbols indicate the ensemble with 15-km EnKF ICs had statistically significantly better reliability than the ensemble with 3-km EnKF ICs. Absence of a symbol means the differences were not statistically significant at the 95% level. Note that the attributes diagrams themselves stop at 100%; area above 100% was added to make room for statistical significance markers.

while unblended 3-km ICs yielded the least 12–36-h PKE at scales $>$ 100 km. Thus, more robust large-scale perturbation growth and kinetic energy in the GEFS-initialized ensemble was associated with its superior 18–36-h forecast reliability and rank histograms relative to the ensembles with unblended EnKF ICs. However, blending GEFS ICs with 3-km EnKF analyses promoted large-scale PKE growth after 6 h, and by 24–36 h, the ensembles initialized from GEFS and blended 3-km ICs had comparable large-scale PKEs, indicating blending successfully recovered these apparently favorable large-scale spectral characteristics that benefited reliability statistics and rank histograms.

3) FRACTIONS SKILL SCORES

Forecast skill was further evaluated with the fractions skill score (FSS; Roberts and Lean 2008), where FSS = 1 indicates a perfect forecast and FSS = 0 means no skill. We present FSSs for r = 100 km, although conclusions were unchanged when

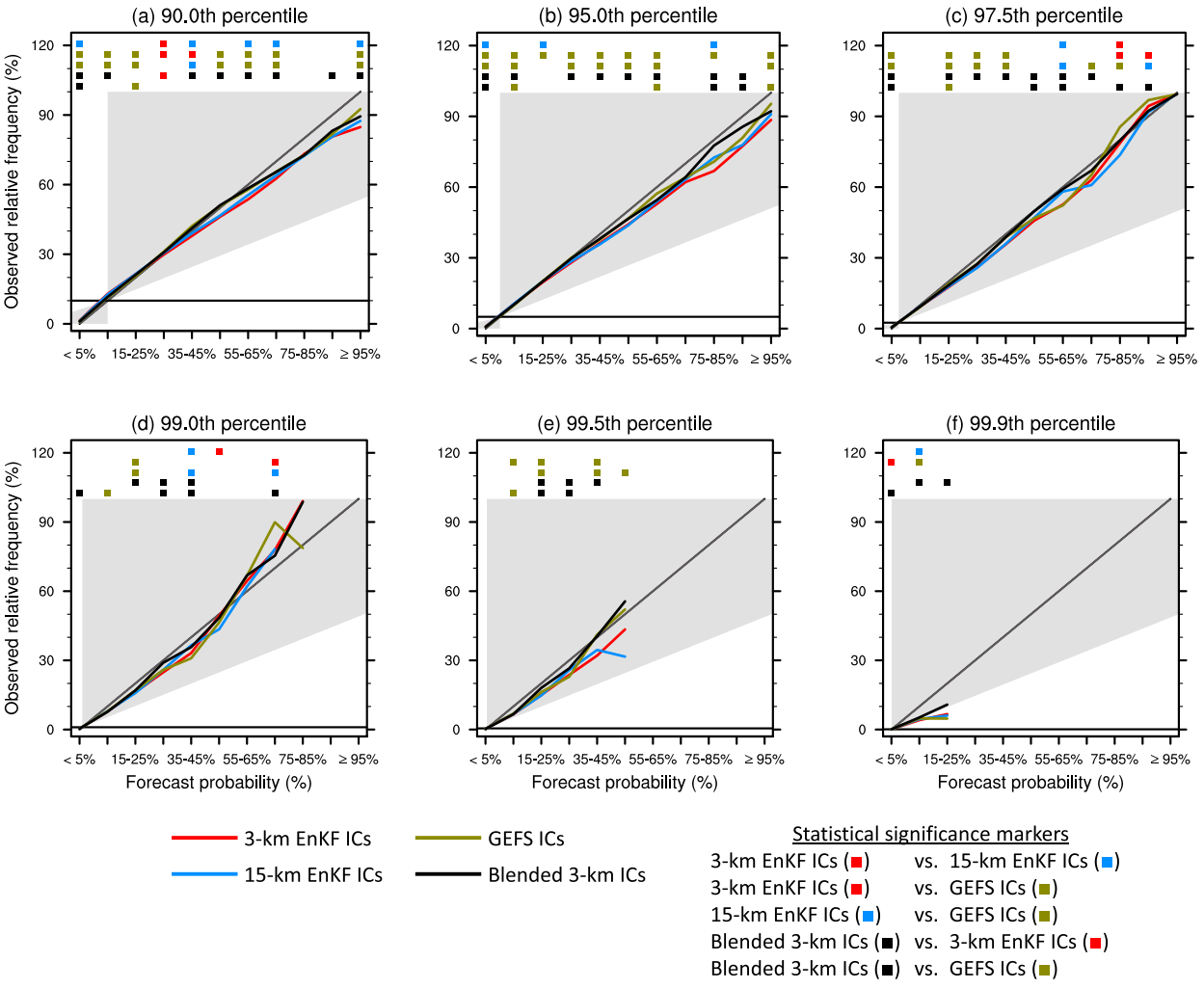

FIG. 12. As in Fig. 11, but statistics were aggregated over all twenty-six 18–36-h, 3-km forecasts of 1-h accumulated precipitation using a 125-km neighborhood length scale.

FSSs were computed with different neighborhood length scales. Moreover, areas under the relative operating characteristic curve (Mason 1982; Mason and Graham 2002) provided identical conclusions as FSSs and are not discussed further.

Forecasts initialized from unblended 3-km EnKF analyses had higher FSSs than those initialized from downscaled 15-km EnKF analyses through 6–12 h, both when aggregated over all forecasts (Fig. 16) and on an hour-by-hour basis (Figs. 17a–d), with many instances of significant differences. However, after 6–12 h, the ensembles with unblended 15- and 3-km EnKF ICs usually had statistically indistinguishable FSSs. Compared to the GEFS-initialized ensemble, the unblended EnKF-initialized ensembles had statistically significantly higher aggregate FSSs through 12–18 h but comparable or lower aggregate FSSs thereafter (Fig. 16), similar to attributes statistics. These 1–12-h forecast benefits from unblended 3-km EnKF ICs compared to GEFS ICs were evident for most hourly forecasts (Figs. 17i–l), while individual 1-h accumulated precipitation forecasts over the 18–36-h period from the ensemble with GEFS ICs were frequently comparable to or better than those from the ensemble with unblended 3-km EnKF ICs (Figs. 17m–p).

Blended 3-km ICs led to FSSs mirroring those from unblended 3-km EnKF ICs over the first 12–18 h (Fig. 16), indicating blending preserved short-term forecast benefits of increased analysis resolution for spatial placement. Furthermore, after 18–24 h, the ensemble with blended 3-km ICs had higher FSSs than the ensemble with unblended 3-km EnKF ICs both on an hourly basis (Figs. 17e–h) and in aggregate that were similar to or higher than FSSs from the GEFS-initialized ensemble.

4) SYNTHESIS

FSSs, attributes statistics, and rank histograms revealed clear benefits of convection-allowing analyses compared to convection-parameterizing analyses for 1–12-h precipitation forecasts, consistent with previous work (e.g.,

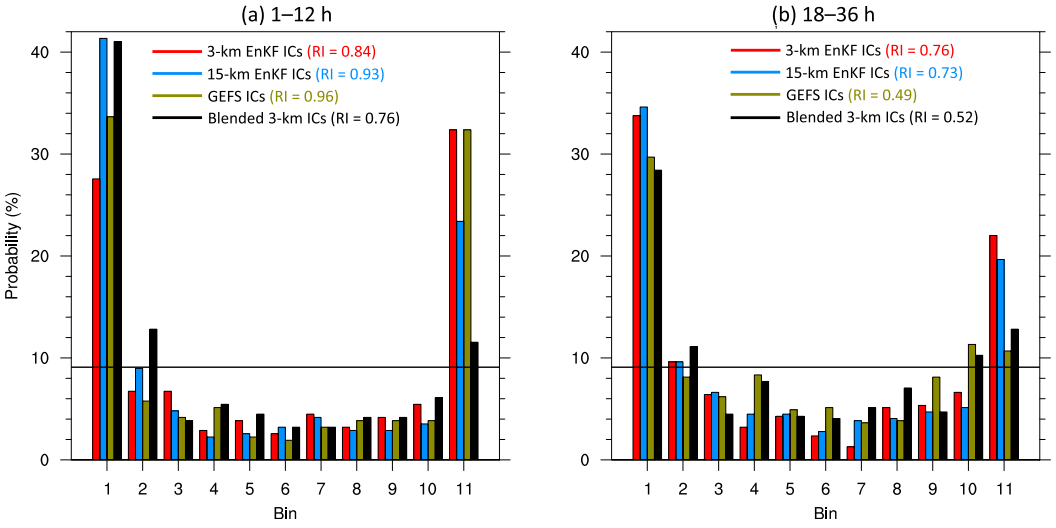

FIG. 13. Rank histograms containing all twenty-six 3-km (a) 1–12- and (b) 18–36-h forecasts of domain-total 1-h accumulated precipitation on the ST4 grid over the verification region (CONUS east of 105°W) for the various ensembles. Horizontal lines are optimal values, and the reliability index (RI; Delle Monache et al. 2006) is annotated for each ensemble in the legend; lower values are better and indicate flatter rank histograms.

Johnson et al. 2015; Johnson and Wang 2016; Schwartz 2016; Gustafsson et al. 2018). But, these improvements from convection-allowing ICs did not persist to next-day forecast ranges, where GEFS-initialized forecasts outperformed EnKF-initialized forecasts. However, blended 3-km ICs led to similar or better 18–36-h forecasts than GEFS ICs, suggesting that blending large-scale fields from a global model with convection-allowing EnKF analyses can improve next-day CAE forecast dispersion, skill, and reliability while preserving short-term forecast benefits of increased IC resolution. Thus, when considering all forecast ranges, blending yielded initial ensembles that produced the best probabilistic forecasts.

c. *Impact of hourly radar DA*

Because our 3-km EnKF was highly constrained, we wondered whether assimilating radar observations could realize meaningful analysis and forecast improvements. So, to assess the impact of assimilating radar reflectivity observations, another EnKF was configured exactly as the nested 15-/3-km EnKF DA system (section 2b), except reflectivity observations throughout the CONUS were assimilated into 3-km analyses along with conventional observations hourly from 1900 to 0000 UTC. Although reflectivity observations could easily be assimilated more frequently in our framework, hourly radar DA mimics the HRRRE configuration. Backgrounds for 1900 UTC radar-assimilating EnKF analyses were provided by 1-h forecasts initialized from 1800 UTC posterior ensembles from the nested 15-/3-km EnKF assimilating solely conventional observations. Thus, the impact of assimilating reflectivity observations was confined to a 6-h period each day. This approach was adopted primarily to avoid the expense of continuously cycling another 3-km EnKF over the entire 4-week period. However, assimilating radar observations for just a few hours was methodologically consistent with numerous other high-resolution DA systems, including the WoF system (e.g., Wheatley et al. 2015; Jones et al. 2016; Skinner et al. 2018), and 6 h of assimilating

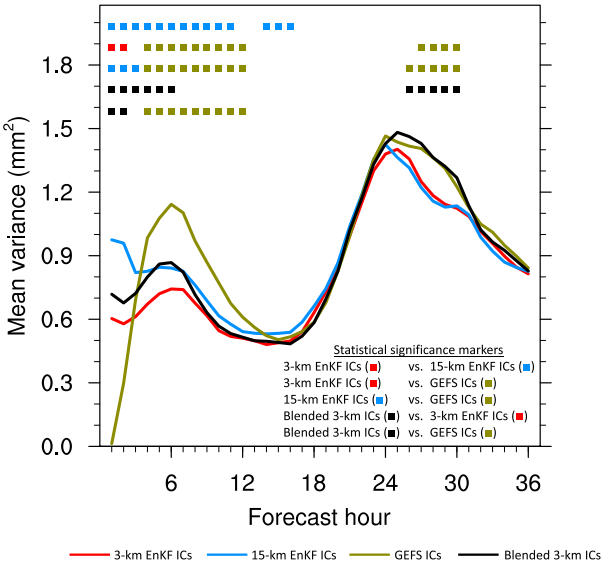

FIG. 14. Average ensemble variance (mm^2) over the verification region (CONUS east of 105°W) and all twenty-six 3-km forecasts of 1-h accumulated precipitation as a function of forecast hour. Values on the *x* axis represent ending forecast hours of 1-h accumulation periods (e.g., an *x*-axis value of 24 is for 1-h accumulated precipitation between 23 and 24 h). Symbols along the top axis indicate forecast hours when differences between two ensembles were statistically significant at the 95% level as in Fig. 11 and denote the ensemble with statistically significantly higher variance.

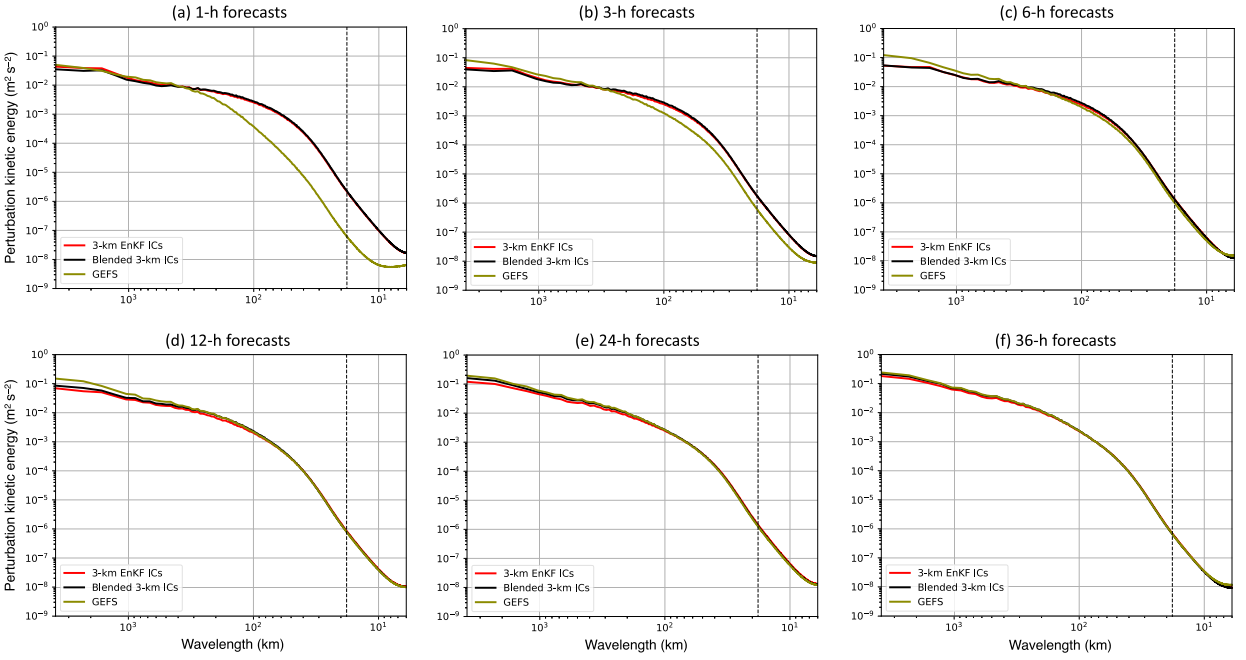

FIG. 15. Average 500-hPa perturbation kinetic energy ($m^2 s^{-2}$) as a function of wavelength (km) computed from all 26 (a) 1-, (b) 3-, (c) 6-, (d) 12-, (e) 24-, and (f) 36-h 3-km, 10-member ensemble forecasts over the entire 3-km domain, excluding the 15 grid points nearest each lateral boundary. Perturbations were computed with respect to the ensemble mean, and the spectra were averaged over all 10 perturbations and 26 forecasts. Dashed vertical lines denote 6 times the horizontal grid spacing (3 km), the approximate effective resolution of the forecasts (Skamarock 2004). The discrete cosine transform was used to perform the spectral analysis, and spectral variance binning employed the method of Ricard et al. (2013).

radar observations was more than sufficient to assess the data impact (e.g., Johnson and Wang 2017 and references therein).

Specific radar DA configurations mostly followed Duda et al. (2019) and references therein (Table 4), and like the other EnKFs, 0000 UTC analysis ensembles initialized 36-h, 3-km, 10-member CAE forecasts. Furthermore, to examine the interplay of blending and radar DA, we also created a set of ICs by blending GEFS ICs with radar-assimilating 3-km EnKF analyses using a 960-km filter cutoff.

Assimilating reflectivity observations generally improved FSSs over the first 3 h but had small impacts thereafter (Fig. 18), similar to other studies finding short-lived benefits of radar DA (e.g., Kain et al. 2010; Johnson et al. 2015; Fabry and Meunier 2020). Within the radar-assimilating experiments, blending boosted FSSs at later times, as with the nonradar DA experiments. Assimilating reflectivity observations negligibly impacted attributes statistics, although assimilation of 100 000–200 000 radar observations each cycle lessened precipitation spread over the first hour (not shown). While more frequent assimilation cycles could potentially realize additional improvements from radar DA, it is unlikely that the small-scale information from radar observations can consistently yield forecast improvements after the shortest forecast ranges, especially in an EnKF highly constrained by other observations. Nonetheless, these experiments suggest feasibility of performing radar-assimilating, WoF-like analyses over large domains in a continuously cycling EnKF DA framework.

5. Summary and conclusions

EnKF DA systems with 80 members and 15- and 3-km horizontal grid spacings were continuously cycled with a 1-h period for 4 weeks over a computational domain spanning the entire CONUS. Both the 15- and 3-km EnKFs had stable climates throughout the cycling period and acceptable prior observation-space statistics, demonstrating the viability of a convection-allowing continuously cycling EnKF over the CONUS. However, our EnKFs were highly constrained by observations, and whether convection-allowing EnKFs can be continuously cycled without deleterious consequences over large data-sparse domains is unclear.

At 0000 UTC, EnKF analyses initialized 36-h, 10-member CAE forecasts with 3-km horizontal grid spacing that were evaluated with a focus on precipitation. CAE forecasts were also initialized from NCEP's operational GEFS and ''blended'' ICs produced by using a low-pass filter to combine large scales from GEFS ICs with small scales from EnKF analysis members. Precipitation forecasts initialized from continuously cycling EnKF analyses outperformed GEFS-initialized forecasts through 12–18 h, and benefits from initializing 3-km forecasts from corresponding 3-km analyses, rather than downscaled 15-km analyses, were

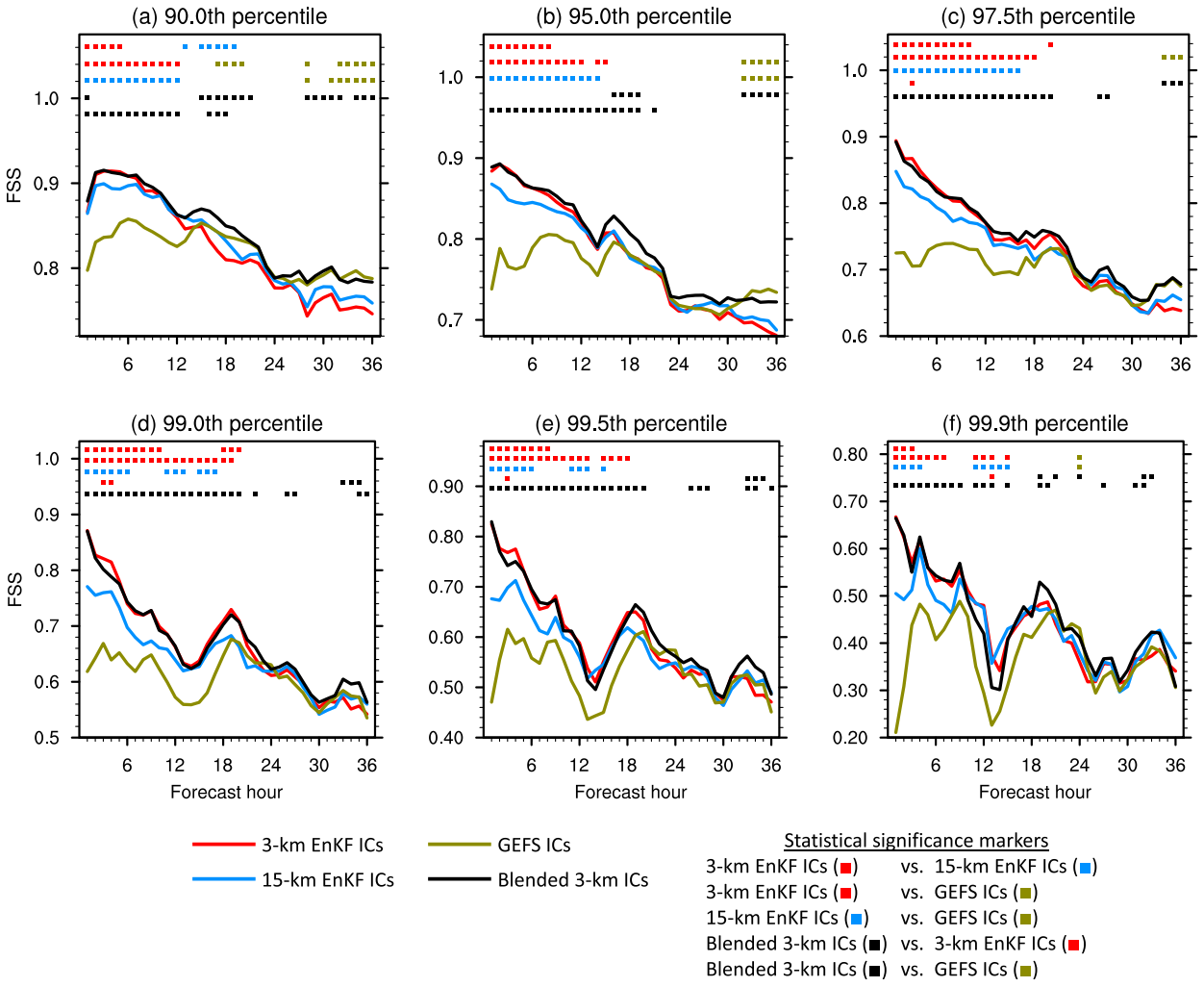

FIG. 16. Fractions skill scores (FSSs) over the verification region (CONUS east of 105°W) with a 100-km neighborhood length scale for the (a) 90th, (b) 95th, (c) 97.5th, (d) 99th, (e) 99.5th, and (f) 99.9th percentile thresholds aggregated over all twenty-six 3-km forecasts of 1-h accumulated precipitation as a function of forecast hour. Values on the *x* axis represent ending forecast hours of 1-h accumulation periods (e.g., an *x*-axis value of 24 is for 1-h accumulated precipitation between 23 and 24 h). The *y*-axis scales are different in each panel. Symbols along the top axis indicate forecast hours when differences between two ensembles were statistically significant at the 95% level as in Fig. 11 and denote the ensemble with statistically significantly higher FSSs. Note that the maximum FSS is 1.0; the area above 1.0 was added to make room for statistical significance markers.

realized through 6–12 h. But, after 18 h, GEFS-initialized forecasts were comparable to or better than EnKF-initialized forecasts, indicating limitations of limited-area continuously cycling EnKFs as initialization tools for next-day CAE precipitation forecasts, consistent with Schwartz et al. (2020). Benefits of assimilating radar reflectivity observations into the 3-km EnKF were confined solely to 1–3-h forecasts.

Although blending sometimes degraded precipitation climatologies over the first 12 h, forecasts initialized from blended 3-km ICs reflected the respective strengths of both GEFS and 3-km EnKF ICs. Specifically, through 12–18 h, forecasts initialized from blended 3-km ICs had similar or better skill, reliability, and dispersion than those initialized from unblended 3-km EnKF analyses, while after 18–24 h, forecasts with blended 3-km ICs were comparable to or better than those with GEFS ICs. Therefore, blending produced ICs yielding the best performance when considering the entire 36-h forecast, indicating how combining large-scale global fields with high-resolution, limited-area EnKF analyses can potentially unify short-term WoF-like and next-day CAE guidance systems under a common framework.

There are many avenues for additional research and improvements. For example, while using identical inflation factors and observation errors in the 15- and 3-km EnKFs provided reasonable results, these choices may have been suboptimal. In particular, because observation errors are the sum of measurement and representativeness errors and representativeness errors are resolution dependent (e.g., Ben Bouallegue et al. 2020), observation errors should arguably be

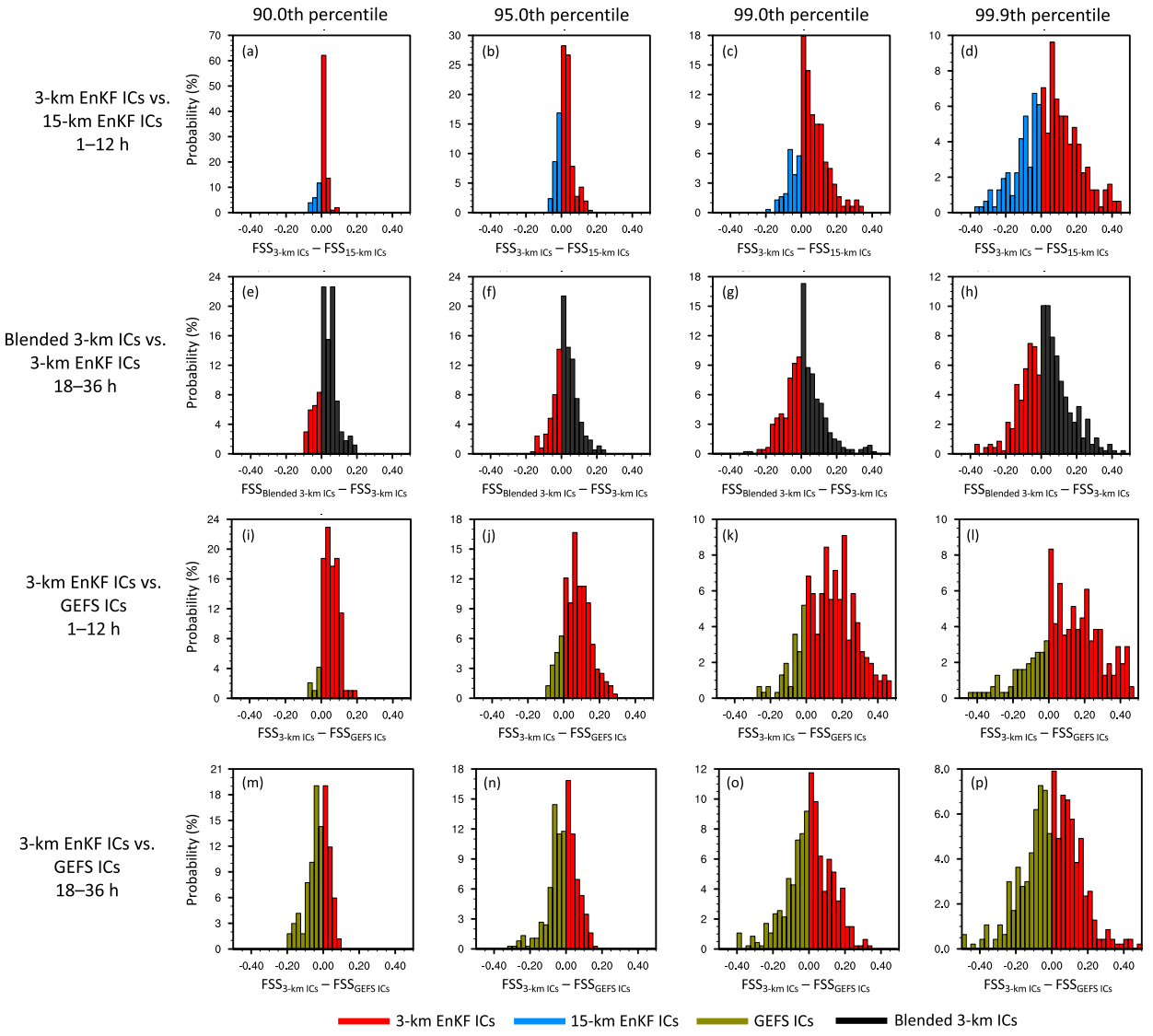

FIG. 17. (a)–(d) Histogram [expressed as probabilities (%)] of FSS differences with $r = 100$ km between the ensembles with 3-km EnKF ICs and 15-km EnKF ICs (3-km ICs minus 15-km ICs) computed from all twenty-six 0–1-, 1–2-, . . . , 10–11-, and 11–12-h 3-km forecasts of 1-h accumulated precipitation for the (a) 90th, (b) 95th, (c) 99th, and (d) 99.9th percentile thresholds. (e)–(h) As in (a)–(d), but for differences from all twenty-six 18–19-, 19–20-, . . . , 34–35-, and 35–36-h 3-km forecasts of 1-h accumulated precipitation between the ensembles with blended and unblended 3-km ICs (blended 3-km ICs minus unblended 3-km ICs). (i)–(l), (m)–(p) As in (a)–(d) and (e)–(h), respectively, but for differences between the ensembles with unblended 3-km EnKF and GEFS ICs (3-km ICs minus GEFS ICs). Values on the x axis denote the leftmost points of each bin, and bin widths were 0.025 (e.g., the bars with left edges at 0.05 are for bins spanning 0.05–0.075). Colors of the bars correspond to the legend and indicate the experiment with the higher FSS in that bin.

tuned for each domain, which, in turn, might require adjusting inflation factors. Thus, it may be possible to further improve our 3-km EnKF.

Additionally, our blending procedure did not impact the continuously cycling EnKF DA systems, and future work might assess whether incorporating large scales from global analyses into hourly limited-area DA cycles is beneficial. Furthermore, blending could potentially be optimized by dynamically determining the filter cutoff scale (e.g., Feng et al. 2020) or using height- and variable-specific cutoffs (e.g., Zhang et al. 2015), and efforts to mitigate blending-induced initial imbalances tailored for high-resolution models are needed. Moreover, next-day forecast benefits of blending suggest further exploring the value of mixed-resolution ensemble-based DA systems for convective applications may be worthwhile. Also, blending and partial cycling DA approaches should be compared; while both methods introduce external large-scale information into limited-area ICs, whether either method is

TABLE 4. Settings for assimilation of radar reflectivity observations.

Radar observation source	Three-dimensional Multi-Radar Multi-Sensor (MRMS; Smith et al. 2016) reflectivity mosaic valid at the top of the hour
Horizontal localization full width	18 km
Vertical localization full width	0.5 scale heights
Observation error standard deviation	5.0 dBZ
Outlier check	$3(\sigma_f^2 + \sigma_o^2)^{1/2}$, where σ_f is the prior ensemble standard deviation at the observation location and σ_o is the observation error standard deviation (5.0 dBZ)
Observation operator	Interpolate diagnosed reflectivity from the Thompson microphysics scheme to observation locations within DART
Excluded observations	0–10 dBZ
Assimilation of nonprecipitation observations	Reflectivity observations $<$ 0.0 dBZ reset to 0.0 dBZ and assimilated
Minimum allowed forward operator value	0.0 dBZ; priors $<$ 0.0 dBZ reset to 0.0 dBZ

preferable is unclear. It is also important to note that our blending methodology [Eq. (1)] changed the large-scale component of both the IC perturbations and the initial ensemble mean state, differing from an approach of blending *perturbations* derived from two different ensembles without changing the spectral representation of the initial ensemble mean (e.g., Caron 2013). Therefore, we cannot determine whether the 18–36-h forecast improvements from blending were due to altering the large-scale IC perturbations or large-scale initial ensemble mean, and it would be interesting to refine attribution in future work.

Finally, computing availability limited our cycling period to just 4 weeks, and additional experimentation is needed over longer periods, different seasons, and varied geographic

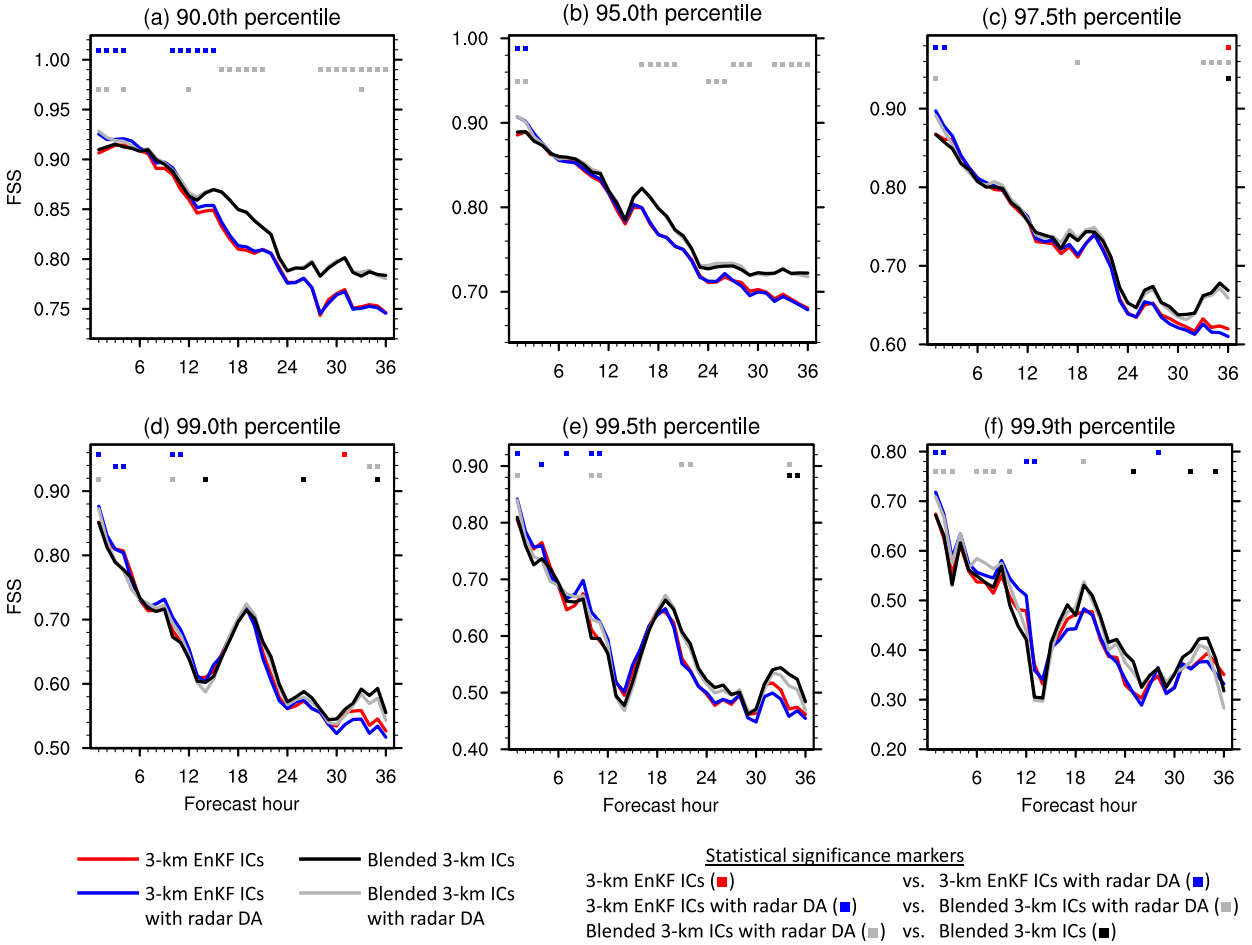

FIG. 18. As in Fig. 16, but for comparisons focusing on the impact of assimilating radar reflectivity observations and aggregated over 24, rather than 26, forecasts (necessitated due to missing radar observations that precluded radar data assimilation sensitivity experiments for the forecasts initialized at 0000 UTC 13 and 14 May 2017). Gray curves are often beneath the black curves, especially in (a) and (b).

regions to further understand large-domain convection-allowing continuously cycling EnKF performance and whether benefits of blending are regime- and location-dependent. Nonetheless, this work suggests a combination of blending and high-resolution EnKF DA may represent a promising pathway toward an operational ensemble-based convection-allowing analysis–forecast system suitable for both nowcasting and next-day prediction over the CONUS.

Acknowledgments. This work was partially funded by NCAR's Short-term Explicit Prediction (STEP) program and NOAA/OAR Office of Weather and Air Quality Grants NA17OAR4590182, NA17OAR4590114, and NA17OAR4590122. All forecasts were produced on NCAR's Cheyenne supercomputer (Computational and Information Systems Laboratory 2017). Thanks to three anonymous reviewers for their constructive comments that improved this paper. The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Data availability statement. Data supporting the findings of this work are available from the corresponding author. Specifically, raw precipitation forecasts from all numerical experiments are available from the corresponding author, and ST4 observations used for verification are available from https://doi.org/10.5065/D69Z93M3. Each of the 649 EnKF DA cycles required at least 250 GB of disk space, so data for each cycle were not saved. However, with GFS GRIB files (available from https://doi.org/10.5065/D65D8PWK) and model configuration and observation files (both available from the corresponding author), EnKF DA cycles can be reproduced. Files used to initialize the 36-h ensemble forecasts and the corresponding GEFS files used for LBCs and blending were saved and are available upon request.

REFERENCES

Accadia, C., S. Mariani, M. Casaioli, A. Lavagnini, and A. Speranza, 2003: Sensitivity of precipitation forecast skill scores to bilinear interpolation and a simple nearest-neighbor average method on high-resolution verification grids. *Wea. Forecasting*, **18**, 918–932, https://doi.org/10.1175/1520-0434(2003)018<0918:SOPFSS>2.0.CO;2.

Ancell, B. C., 2012: Examination of analysis and forecast errors of high-resolution assimilation, bias removal, and digital filter initialization with an ensemble Kalman filter. *Mon. Wea. Rev.*, **140**, 3992–4004, https://doi.org/10.1175/MWR-D-11-00319.1.

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903, https://doi.org/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, https://doi.org/10.1175/1520-0493(2003)131<0634:ALLSFF>2.0.CO;2.

——, 2012: Localization and sampling error correction in ensemble Kalman filter data assimilation. *Mon. Wea. Rev.*, **140**, 2359–2371, https://doi.org/10.1175/MWR-D-11-00013.1.

——, and N. Collins, 2007: Scalable implementations of ensemble filter algorithms for data assimilation. *J. Atmos. Oceanic Technol.*, **24**, 1452–1463, https://doi.org/10.1175/JTECH2049.1.

——, T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Arellano, 2009: The Data Assimilation Research Testbed: A community facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296, https://doi.org/10.1175/2009BAMS2618.1.

Ballish, B. A., and V. Krishna Kumar, 2008: Systematic differences in aircraft and radiosonde temperatures. *Bull. Amer. Meteor. Soc.*, **89**, 1689–1708, https://doi.org/10.1175/2008BAMS2332.1.

Barker, D. M., 2005: Southern high-latitude ensemble data assimilation in the Antarctic mesoscale prediction system. *Mon. Wea. Rev.*, **133**, 3431–3449, https://doi.org/10.1175/MWR3042.1.

——, and Coauthors, 2012: The Weather Research and Forecasting Model's community variational/ensemble data assimilation system: WRFDA. *Bull. Amer. Meteor. Soc.*, **93**, 831–843, https://doi.org/10.1175/BAMS-D-11-00167.1.

Ben Bouallegue, Z., T. Haiden, N. J. Weber, T. M. Hamill, and D. S. Richardson, 2020: Accounting for representativeness in the verification of ensemble precipitation forecasts. *Mon. Wea. Rev.*, **148**, 2049–2062, https://doi.org/10.1175/MWR-D-19-0323.1.

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

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3, https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.

Cafaro, C., T. H. A. Frame, J. Methven, N. Roberts, and J. Bröcker, 2019: The added value of convection-permitting ensemble forecasts of sea breeze compared to a Bayesian forecast driven by the global ensemble. *Quart. J. Roy. Meteor. Soc.*, **145**, 1780–1798, https://doi.org/10.1002/qj.3531.

Caron, J.-F., 2013: Mismatching perturbations at the lateral boundaries in limited-area ensemble forecasting: A case study. *Mon. Wea. Rev.*, **141**, 356–374, https://doi.org/10.1175/MWR-D-12-00051.1.

Cavallo, S. M., J. Berner, and C. Snyder, 2016: Diagnosing model errors from time-averaged tendencies in the Weather Research and Forecasting (WRF) Model. *Mon. Wea. Rev.*, **144**, 759–779, https://doi.org/10.1175/MWR-D-15-0120.1.

Chen, F., and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev.*, **129**, 569–585, https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2.

Clark, A. J., W. A. Gallus Jr., M. Xue, and F. Kong, 2009: A comparison of precipitation forecast skill between small convection-allowing and large convection-parameterizing ensembles. *Wea. Forecasting*, **24**, 1121–1140, https://doi.org/10.1175/2009WAF2222222.1.

——, and Coauthors, 2011: Probabilistic precipitation forecast skill as a function of ensemble size and spatial scale in a convection-allowing ensemble. *Mon. Wea. Rev.*, **139**, 1410–1418, https://doi.org/10.1175/2010MWR3624.1.

——, and Coauthors, 2012: An overview of the 2010 Hazardous Weather Testbed experimental forecast program spring experiment. *Bull. Amer. Meteor. Soc.*, **93**, 55–74, https://doi.org/10.1175/BAMS-D-11-00040.1.

——, and Coauthors, 2018: The Community Leveraged Unified Ensemble (CLUE) in the 2016 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment. *Bull. Amer. Meteor. Soc.*, **99**, 1433–1448, https://doi.org/10.1175/BAMS-D-16-0309.1.

Computational and Information Systems Laboratory, 2017: Cheyenne: HPE/SGI ICE XA System (NCAR Community Computing). National Center for Atmospheric Research, accessed 13 October 2020, https://doi.org/10.5065/D6RX99HX.

COSMO, 2020: MeteoSwiss operational applications within COSMO. Accessed 24 April 2020, http://www.cosmo-model.org/content/tasks/operational/meteoSwiss/default.htm.

Dahlgren, P., and N. Gustafsson, 2012: Assimilating host model information into a limited area model. *Tellus*, **64A**, 15836, https://doi.org/10.3402/tellusa.v64i0.15836.

Degelia, S. K., X. Wang, D. J. Stensrud, and A. Johnson, 2018: Understanding the impact of radar and in situ observations on the prediction of a nocturnal convection initiation event on 25 June 2013 using an ensemble-based multiscale data assimilation system. *Mon. Wea. Rev.*, **146**, 1837–1859, https://doi.org/10.1175/MWR-D-17-0128.1.

Delle Monache, L., J. P. Hacker, Y. Zhou, X. Deng, and R. B. Stull, 2006: Probabilistic aspects of meteorological and ozone regional ensemble forecasts. *J. Geophys. Res.*, **111**, D24307, https://doi.org/10.1029/2005JD006917.

Denis, B., J. Coté, and R. Laprise, 2002: Spectral decomposition of two-dimensional atmospheric fields on limited-area domains using the discrete cosine transform (DCT). *Mon. Wea. Rev.*, **130**, 1812–1829, https://doi.org/10.1175/1520-0493(2002)130<1812:SDOTDA>2.0.CO;2.

Developmental Testbed Center, 2016: Gridpoint statistical interpolation advanced user's guide version 3.5.0.0. Developmental Testbed Center, 119 pp., https://dtcenter.org/com-GSI/users/docs/users_guide/AdvancedGSIUserGuide_v3.5.0.0.pdf.

Dey, S. R., G. Leoncini, N. M. Roberts, R. S. Plant, and S. Migliorini, 2014: A spatial view of ensemble spread in convection permitting ensembles. *Mon. Wea. Rev.*, **142**, 4091–4107, https://doi.org/10.1175/MWR-D-14-00172.1.

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

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

Duc, L., K. Saito, and H. Seko, 2013: Spatial–temporal fractions verification for high-resolution ensemble forecasts. *Tellus*, **65A**, 18171, https://doi.org/10.3402/tellusa.v65i0.18171.

Duda, J. D., X. Wang, Y. Wang, and J. Carley, 2019: Comparing the assimilation of radar reflectivity using the direct GSI based ensemble-variational (EnVar) and indirect cloud analysis methods in convection-allowing forecasts over the continental United States. *Mon. Wea. Rev.*, **147**, 1655–1678, https://doi.org/10.1175/MWR-D-18-0171.1.

Ebert, E. E., 2008: Fuzzy verification of high resolution gridded forecasts: A review and proposed framework. *Meteor. Appl.*, **15**, 51–64, https://doi.org/10.1002/met.25.

——, 2009: Neighborhood verification: A strategy for rewarding close forecasts. *Wea. Forecasting*, **24**, 1498–1510, https://doi.org/10.1175/2009WAF2222251.1.

Evans, C., D. F. Van Dyke, and T. Lericos, 2014: How do forecasters utilize output from a convection-permitting ensemble forecast system? Case study of a high-impact precipitation event. *Wea. Forecasting*, **29**, 466–486, https://doi.org/10.1175/WAF-D-13-00064.1.

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, https://doi.org/10.1029/94JC00572.

Fabry, F., and V. Meunier, 2020: Why are radar data so difficult to assimilate skillfully? *Mon. Wea. Rev.*, **148**, 2819–2836, https://doi.org/10.1175/MWR-D-19-0374.1.

Feng, J., J. Sun, and Y. Zhang, 2020: A dynamic blending scheme to mitigate large-scale bias in regional models. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001754, https://doi.org/10.1029/2019MS001754.

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

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

Gasperoni, N. A., X. Wang, and Y. Wang, 2020: A comparison of methods to sample model errors for convection-allowing ensemble forecasts in the setting of multiscale initial conditions produced by the GSI-based EnVar assimilation system. *Mon. Wea. Rev.*, **148**, 1177–1203, https://doi.org/10.1175/MWR-D-19-0124.1.

Gebhardt, C., S. E. Theis, M. Paulat, and Z. Ben Bouallègue, 2011: Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. *Atmos. Res.*, **100**, 168–177, https://doi.org/10.1016/j.atmosres.2010.12.008.

Gemmill, W., B. Katz, and X. Li, 2007: Daily real-time, global sea surface temperature—High-resolution analysis: RTG_SST_HR. NOAA/NWS/NCEP/EMC/MMAB, Science Application International Corporation, and Joint Center for Satellite Data Assimilation Tech. Note 260, 39 pp., http://polar.ncep.noaa.gov/mmab/papers/tn260/MMAB260.pdf.

Gowan, T. M., W. J. Steenburgh, and C. S. Schwartz, 2018: Validation of mountain precipitation forecasts from the convection-permitting NCAR ensemble and operational forecast systems over the western United States. *Wea. Forecasting*, **33**, 739–765, https://doi.org/10.1175/WAF-D-17-0144.1.

Guidard, V., and C. Fischer, 2008: Introducing the coupling information in a limited-area variational assimilation. *Quart. J. Roy. Meteor. Soc.*, **134**, 723–735, https://doi.org/10.1002/qj.215.

Gustafsson, N., and Coauthors, 2018: Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. *Quart. J. Roy. Meteor. Soc.*, **144**, 1218–1256, https://doi.org/10.1002/qj.3179.

Hagelin, S., J. Son, R. Swinbank, A. McCabe, N. Roberts, and W. Tennant, 2017: The Met Office convective-scale ensemble, MOGREPS-UK. *Quart. J. Roy. Meteor. Soc.*, **143**, 2846–2861, https://doi.org/10.1002/qj.3135.

Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167, https://doi.org/10.1175/1520-0434(1999)014<0155:HTFENP>2.0.CO;2.

——, 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550–560, https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2.

Harnisch, F., and C. Keil, 2015: Initial conditions for convective-scale ensemble forecasting provided by ensemble data assimilation. *Mon. Wea. Rev.*, **143**, 1583–1600, https://doi.org/10.1175/MWR-D-14-00209.1.

Hohenegger, C., and C. Schär, 2007: Predictability and error growth dynamics in cloud-resolving models. *J. Atmos. Sci.*, **64**, 4467–4478, https://doi.org/10.1175/2007JAS2143.1.

Houtekamer, P. L., and F. Zhang, 2016: Review of the ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **144**, 4489–4532, https://doi.org/10.1175/MWR-D-15-0440.1.

——, H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.*, **133**, 604–620, https://doi.org/10.1175/MWR-2864.1.

——, X. Deng, H. L. Mitchell, S.-J. Baek, and N. Gagnon, 2014: Higher resolution in an operational ensemble Kalman filter. *Mon. Wea. Rev.*, **142**, 1143–1162, https://doi.org/10.1175/MWR-D-13-00138.1.

Hsiao, L.-F., D.-S. Chen, Y.-H. Kuo, Y.-R. Guo, T.-C. Yeh, J.-S. Hong, C.-T. Fong, and C.-S. Lee, 2012: Application of WRF 3DVAR to operational typhoon prediction in Taiwan: Impact of outer loop and partial cycling approaches. *Wea. Forecasting*, **27**, 1249–1263, https://doi.org/10.1175/WAF-D-11-00131.1.

——, and Coauthors, 2015: Blending of global and regional analyses with a spatial filter: Application to typhoon prediction over the western North Pacific Ocean. *Wea. Forecasting*, **30**, 754–770, https://doi.org/10.1175/WAF-D-14-00047.1.

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, https://doi.org/10.1029/2008JD009944.

Iyer, E. R., A. J. Clark, M. Xue, and F. Kong, 2016: A comparison of 36–60-h precipitation forecasts from convection-allowing and convection-parameterizing ensembles. *Wea. Forecasting*, **31**, 647–661, https://doi.org/10.1175/WAF-D-15-0143.1.

James, E. P., and S. G. Benjamin, 2017: Observation system experiments with the hourly updating Rapid Refresh model using GSI hybrid ensemble–variational data assimilation. *Mon. Wea. Rev.*, **145**, 2897–2918, https://doi.org/10.1175/MWR-D-16-0398.1.

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, https://doi.org/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., http://www.emc.ncep.noaa.gov/officenotes/newernotes/on437.pdf.

Jirak, I. L., A. J. Clark, B. Roberts, B. T. Gallo, and S. J. Weiss, 2018: Exploring the optimal configuration of the High Resolution Ensemble Forecast System. *25th Conf. on Numerical Weather Prediction*, Denver, CO, Amer. Meteor. Soc., 14B.6, https://ams.confex.com/ams/29WAF25NWP/webprogram/Paper345640.html.

Johnson, A., and X. Wang, 2016: A study of multiscale initial condition perturbation methods for convection-permitting ensemble forecasts. *Mon. Wea. Rev.*, **144**, 2579–2604, https://doi.org/10.1175/MWR-D-16-0056.1.

——, and ——, 2017: Design and implementation of a GSI-based convection-allowing ensemble data assimilation and forecast system for the PECAN field experiment. Part I: Optimal configurations for nocturnal convection prediction using retrospective cases. *Wea. Forecasting*, **32**, 289–315, https://doi.org/10.1175/WAF-D-16-0102.1.

——, ——, J. Carley, L. Wicker, and C. Karstens, 2015: A comparison of multiscale GSI-based EnKF and 3DVar data assimilation using radar and conventional observations for midlatitude convective-scale precipitation forecasts. *Mon. Wea. Rev.*, **143**, 3087–3108, https://doi.org/10.1175/MWR-D-14-00345.1.

——, ——, Y. Wang, A. Reinhart, A. J. Clark, and I. L. Jirak, 2020: Neighborhood- and object-based probabilistic verification of the OU MAP ensemble forecasts during 2017 and 2018 Hazardous Weather Testbeds. *Wea. Forecasting*, **35**, 169–191, https://doi.org/10.1175/WAF-D-19-0060.1.

Jones, T. A., and D. J. Stensrud, 2012: Assimilating AIRS temperature and mixing ratio profiles using an ensemble Kalman filter approach for convective-scale forecasts. *Wea. Forecasting*, **27**, 541–564, https://doi.org/10.1175/WAF-D-11-00090.1.

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

——, X. Wang, P. S. Skinner, A. Johnson, and Y. Wang, 2018: Assimilation of *GOES-13* imager clear-sky water vapor (6.5 μm) radiances into a Warn-on-Forecast system. *Mon. Wea. Rev.*, **146**, 1077–1107, https://doi.org/10.1175/MWR-D-17-0280.1.

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

Judt, F., 2018: Insights into atmospheric predictability through global convection-permitting model simulations. *J. Atmos. Sci.*, **75**, 1477–1497, https://doi.org/10.1175/JAS-D-17-0343.1.

Kain, J. S., and Coauthors, 2010: Assessing advances in the assimilation of radar data and other mesoscale observations within a collaborative forecasting–research environment. *Wea. Forecasting*, **25**, 1510–1521, https://doi.org/10.1175/2010WAF2222405.1.

Keresturi, E., Y. Wang, F. Meier, F. Weidle, C. Wittmann, and A. Atencia, 2019: Improving initial condition perturbations in a convection-permitting ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **145**, 993–1012, https://doi.org/10.1002/qj.3473.

Klasa, C., M. Arpagaus, A. Walser, and H. Wernli, 2018: An evaluation of the convection-permitting ensemble COSMO-E for three contrasting precipitation events in Switzerland. *Quart. J. Roy. Meteor. Soc.*, **144**, 744–764, https://doi.org/10.1002/qj.3245.

Kleist, D. T., and K. Ide, 2015a: An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the NCEP GFS. Part I: System description and 3D-hybrid results. *Mon. Wea. Rev.*, **143**, 433–451, https://doi.org/10.1175/MWR-D-13-00351.1.

——, and ——, 2015b: An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the NCEP GFS. Part II: 4DEnVar and hybrid variants. *Mon. Wea. Rev.*, **143**, 452–470, https://doi.org/10.1175/MWR-D-13-00350.1.

——, 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.

Ladwig, T. T., and Coauthors, 2018: Development of the High-Resolution Rapid Refresh Ensemble (HRRRE) toward an operational convective-allowing ensemble data assimilation and forecast system. *Sixth Symp. on the Weather, Water, and Climate Enterprise*, Austin, TX, Amer. Meteor. Soc., TJ1.2, https://ams.confex.com/ams/98Annual/webprogram/Paper334565.html.

Li, Z., J. C. McWilliams, K. Ide, and J. Farrara, 2015: A multiscale variational data assimilation scheme: Formulation and illustration. *Mon. Wea. Rev.*, **143**, 3804–3822, https://doi.org/10.1175/MWR-D-14-00384.1.

Lin, H., S. S. Weygandt, S. G. Benjamin, and M. Hu, 2017: Satellite radiance data assimilation within the hourly updated Rapid Refresh. *Wea. Forecasting*, **32**, 1273–1287, https://doi.org/10.1175/WAF-D-16-0215.1.

Lin, Y., and K. E. Mitchell, 2005: The NCEP stage II/IV hourly precipitation analyses: Development and applications. *19th Conf. on Hydrology*, San Diego, CA, Amer. Meteor. Soc., 1.2, http://ams.confex.com/ams/pdfpapers/83847.pdf.

Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307, https://doi.org/10.3402/tellusa.v21i3.10086.

Mason, I. B., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303.

Mason, S. J., and N. E. Graham, 2002: Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quart. J. Roy. Meteor. Soc.*, **128**, 2145–2166, https://doi.org/10.1256/003590002320603584.

Maurer, V., N. Kalthoff, and L. Gantner, 2017: Predictability of convective precipitation for West Africa: Verification of convection-permitting and global ensemble simulations. *Meteor. Z.*, **26**, 93–110, https://doi.org/10.1127/metz/2016/0728.

Melhauser, C., and F. Zhang, 2012: Practical and intrinsic predictability of severe and convective weather at the mesoscales. *J. Atmos. Sci.*, **69**, 3350–3371, https://doi.org/10.1175/JAS-D-11-0315.1.

Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys. Space Phys.*, **20**, 851–875, https://doi.org/10.1029/RG020i004p00851.

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

——, ——, and S. A. Thompson, 2013: A long-term assessment of precipitation forecast skill using the fractions skill score. *Meteor. Appl.*, **20**, 176–186, https://doi.org/10.1002/met.296.

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, https://doi.org/10.1029/97JD00237.

Nelson, B. R., O. P. Prat, D.-J. Seo, and E. Habib, 2016: Assessment and implications of NCEP Stage IV quantitative precipitation estimates for product intercomparisons. *Wea. Forecasting*, **31**, 371–394, https://doi.org/10.1175/WAF-D-14-00112.1.

Peralta, C., Z. B. Bouallègue, S. E. Theis, C. Gebhardt, and M. Buchhold, 2012: Accounting for initial condition uncertainties in COSMO-DE-EPS. *J. Geophys. Res.*, **117**, D07108, https://doi.org/10.1029/2011JD016581.

Porson, A. N., S. Hagelin, D. F. A. Boyd, N. M. Roberts, R. North, S. Webster, and J. C.-F. Lo, 2019: Extreme rainfall sensitivity in convective-scale ensemble modelling over Singapore. *Quart. J. Roy. Meteor. Soc.*, **145**, 3004–3022, https://doi.org/10.1002/qj.3601.

Powers, J. G., and Coauthors, 2017: The Weather Research and Forecasting Model: Overview, system efforts, and future directions. *Bull. Amer. Meteor. Soc.*, **98**, 1717–1737, https://doi.org/10.1175/BAMS-D-15-00308.1.

Rainwater, S., and B. Hunt, 2013: Mixed resolution ensemble data assimilation. *Mon. Wea. Rev.*, **141**, 3007–3021, https://doi.org/10.1175/MWR-D-12-00234.1.

Raymond, W. H., 1988: High-order low-pass implicit tangent filters for use in finite area calculations. *Mon. Wea. Rev.*, **116**, 2132–2141, https://doi.org/10.1175/1520-0493(1988)116<2132:HOLPIT>2.0.CO;2.

——, and A. Garder, 1991: A review of recursive and implicit filters. *Mon. Wea. Rev.*, **119**, 477–495, https://doi.org/10.1175/1520-0493(1991)119<0477:ARORAI>2.0.CO;2.

Raynaud, L., and F. Bouttier, 2016: Comparison of initial perturbation methods for ensemble prediction at convective scale. *Quart. J. Roy. Meteor. Soc.*, **142**, 854–866, https://doi.org/10.1002/qj.2686.

——, and ——, 2017: The impact of horizontal resolution and ensemble size for convective-scale probabilistic forecasts. *Quart. J. Roy. Meteor. Soc.*, **143**, 3037–3047, https://doi.org/10.1002/qj.3159.

Ricard, D., C. Lac, S. Riette, R. Legrand, and A. Mary, 2013: Kinetic energy spectra characteristics of two convection-permitting limited-area models AROME and Meso-NH. *Quart. J. Roy. Meteor. Soc.*, **139**, 1327–1341, https://doi.org/10.1002/qj.2025.

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.

Rogers, E., and Coauthors, 2017: Mesoscale modeling development at the National Centers for Environmental Prediction: Version 4 of the NAM forecast system and scenarios for the evolution to a high-resolution ensemble forecast system. *28th Conf. on Weather and Forecasting/24th Conf. on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., 3B.4, https://ams.confex.com/ams/97Annual/webprogram/Paper311212.html.

Romine, G. S., C. S. Schwartz, C. Snyder, J. L. Anderson, and M. L. Weisman, 2013: Model bias in a continuously cycled assimilation system and its influence on convection-permitting forecasts. *Mon. Wea. Rev.*, **141**, 1263–1284, https://doi.org/10.1175/MWR-D-12-00112.1.

——, ——, J. Berner, K. R. Fossell, C. Snyder, J. L. Anderson, and M. L. Weisman, 2014: Representing forecast error in a convection-permitting ensemble system. *Mon. Wea. Rev.*, **142**, 4519–4541, https://doi.org/10.1175/MWR-D-14-00100.1.

Schellander-Gorgas, T., Y. Wang, F. Meier, F. Weidle, C. Wittmann, and A. Kann, 2017: On the forecast skills of a convection-permitting ensemble. *Geosci. Model Dev.*, **10**, 35–56, https://doi.org/10.5194/gmd-10-35-2017.

Schraff, C., H. Reich, A. Rhodin, A. Schomburg, K. Stephan, A. Periáñez, and R. Potthast, 2016: Kilometre-scale ensemble data assimilation for the COSMO model (KENDA). *Quart. J. Roy. Meteor. Soc.*, **142**, 1453–1472, https://doi.org/10.1002/qj.2748.

Schumacher, R. S., and A. J. Clark, 2014: Evaluation of ensemble configurations for the analysis and prediction of

heavy-rain-producing mesoscale convective systems. *Mon. Wea. Rev.*, **142**, 4108–4138, https://doi.org/10.1175/MWR-D-13-00357.1.

Schwartz, C. S., 2016: Improving large-domain convection-allowing forecasts with high-resolution analyses and ensemble data assimilation. *Mon. Wea. Rev.*, **144**, 1777–1803, https://doi.org/10.1175/MWR-D-15-0286.1.

——, 2019: Medium-range convection-allowing ensemble forecasts with a variable-resolution global model. *Mon. Wea. Rev.*, **147**, 2997–3023, https://doi.org/10.1175/MWR-D-18-0452.1.

——, and Z. Liu, 2014: Convection-permitting forecasts initialized with continuously cycling limited-area 3DVAR, ensemble Kalman filter, and ''hybrid'' variational-ensemble data assimilation systems. *Mon. Wea. Rev.*, **142**, 716–738, https://doi.org/10.1175/MWR-D-13-00100.1.

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

——, and Coauthors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, **25**, 263–280, https://doi.org/10.1175/2009WAF2222267.1.

——, G. S. Romine, 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, https://doi.org/10.1175/WAF-D-13-00145.1.

——, ——, M. L. Weisman, R. A. Sobash, K. R. Fossell, K. W. Manning, and S. B. Trier, 2015a: A real-time convection-allowing ensemble prediction system initialized by mesoscale ensemble Kalman filter analyses. *Wea. Forecasting*, **30**, 1158–1181, https://doi.org/10.1175/WAF-D-15-0013.1.

——, ——, R. A. Sobash, K. R. Fossell, and M. L. Weisman, 2015b: NCAR's experimental real-time convection-allowing ensemble prediction system. *Wea. Forecasting*, **30**, 1645–1654, https://doi.org/10.1175/WAF-D-15-0103.1.

——, ——, ——, ——, and ——, 2019: NCAR's real-time convection-allowing ensemble project. *Bull. Amer. Meteor. Soc.*, **100**, 321–343, https://doi.org/10.1175/BAMS-D-17-0297.1.

——, M. Wong, G. S. Romine, R. A. Sobash, and K. R. Fossell, 2020: Initial conditions for convection-allowing ensembles over the conterminous United States. *Mon. Wea. Rev.*, **148**, 2645–2669, https://doi.org/10.1175/MWR-D-19-0401.1.

Shao, H., and Coauthors, 2016: Bridging research to operations transitions: Status and plans of community GSI. *Bull. Amer. Meteor. Soc.*, **97**, 1427–1440, https://doi.org/10.1175/BAMS-D-13-00245.1.

Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev.*, **132**, 3019–3032, https://doi.org/10.1175/MWR2830.1.

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

Skinner, P. S., and Coauthors, 2018: Object-based verification of a prototype warn-on-forecast system. *Wea. Forecasting*, **33**, 1225–1250, https://doi.org/10.1175/WAF-D-18-0020.1.

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

Stensrud, D. J., and Coauthors, 2009: Convective-scale warn-on-forecast system: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1500, https://doi.org/10.1175/2009BAMS2795.1.

——, and Coauthors, 2013: Progress and challenges with warn-on-forecast. *Atmos. Res.*, **123**, 2–16, https://doi.org/10.1016/j.atmosres.2012.04.004.

Tegen, I., P. Hollrig, M. Chin, I. Fung, D. Jacob, and J. Penner, 1997: Contribution of different aerosol species to the global aerosol extinction optical thickness: Estimates from model results. *J. Geophys. Res.*, **102**, 23 895–23 915, https://doi.org/10.1029/97JD01864.

Theis, S. E., A. Hense, and U. Damrath, 2005: Probabilistic precipitation forecasts from a deterministic model: A pragmatic approach. *Meteor. Appl.*, **12**, 257–268, https://doi.org/10.1017/S1350482705001763.

Thompson, G., 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, https://doi.org/10.1175/2008MWR2387.1.

Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Wea. Rev.*, **117**, 1779–1800, https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2.

Torn, R. D., and C. A. Davis, 2012: The influence of shallow convection on tropical cyclone track forecasts. *Mon. Wea. Rev.*, **140**, 2188–2197, https://doi.org/10.1175/MWR-D-11-00246.1.

——, G. J. Hakim, and C. Snyder, 2006: Boundary conditions for limited-area ensemble Kalman filters. *Mon. Wea. Rev.*, **134**, 2490–2502, https://doi.org/10.1175/MWR3187.1.

Vendrasco, E. P., J. Sun, D. L. Herdies, and C. F. De Angelis, 2016: Constraining a 3DVAR radar data assimilation system with large-scale analysis to improve short-range precipitation forecasts. *J. Appl. Meteor. Climatol.*, **55**, 673–690, https://doi.org/10.1175/JAMC-D-15-0010.1.

Wang, H., X.-Y. Huang, D. Xu, and J. Liu, 2014: A scale-dependent blending scheme for WRFDA: Impact on regional weather forecasting. *Geosci. Model Dev.*, **7**, 1819–1828, https://doi.org/10.5194/gmd-7-1819-2014.

Wang, X., and T. Lei, 2014: GSI-based four-dimensional ensemble–variational (4DEnsVar) data assimilation: Formulation and single-resolution experiments with real data for the NCEP Global Forecast System. *Mon. Wea. Rev.*, **142**, 3303–3325, https://doi.org/10.1175/MWR-D-13-00303.1.

Wang, Y., and Coauthors, 2011: The central European limited-area-ensemble forecasting system: ALADIN-LAEF. *Quart. J. Roy. Meteor. Soc.*, **137**, 483–502, https://doi.org/10.1002/qj.751.

——, M. Bellus, J.-F. Geleyn, X. Ma, W. Tian, and F. Weidle, 2014: A new method for generating initial condition perturbations in a regional ensemble prediction system: Blending. *Mon. Wea. Rev.*, **142**, 2043–2059, https://doi.org/10.1175/MWR-D-12-00354.1.

Warner, T. T., R. A. Peterson, and R. E. Treadon, 1997: A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. *Bull. Amer. Meteor. Soc.*, **78**, 2599–2617, https://doi.org/10.1175/1520-0477(1997)078<2599:ATOLBC>2.0.CO;2.

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

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924, https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.

——, and ——, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 3078–3089, https://doi.org/10.1175/MWR-D-11-00276.1.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 704 pp.

Wolff, J. K., M. Harrold, T. Fowler, J. H. Gotway, L. Nance, and B. G. Brown, 2014: Beyond the basics: Evaluating model-based precipitation forecasts using traditional, spatial, and object-based methods. *Wea. Forecasting*, **29**, 1451–1472, https://doi.org/10.1175/WAF-D-13-00135.1.

Wong, M., G. Romine, and C. Snyder, 2020: Model improvement via systematic investigation of physics tendencies. *Mon. Wea. Rev.*, **148**, 671–688, https://doi.org/10.1175/MWR-D-19-0255.1.

Woodhams, B. J., C. E. Birch, J. H. Marsham, C. L. Bain, N. M. Roberts, and D. F. Boyd, 2018: What is the added value of a convection-permitting model for forecasting extreme rainfall over tropical East Africa? *Mon. Wea. Rev.*, **146**, 2757–2780, https://doi.org/10.1175/MWR-D-17-0396.1.

Wu, W.-S., D. F. Parrish, E. Rogers, and Y. Lin, 2017: Regional ensemble–variational data assimilation using global ensemble forecasts. *Wea. Forecasting*, **32**, 83–96, https://doi.org/10.1175/WAF-D-16-0045.1.

Xue, M., and Coauthors, 2007: CAPS real-time storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2007 Spring Experiment. *22nd Conf. on Weather Analysis and Forecasting/18th Conf. on Numerical Weather Prediction*, Salt Lake City, UT, Amer. Meteor. Soc., 3B.1, http://ams.confex.com/ams/pdfpapers/124587.pdf.

Yang, X., 2005: Analysis blending using a spatial filter in grid-point model coupling. *HIRLAM Newsletter*, No. 48, Article 10, HIRLAM Programme, de Bilt, Netherlands, 49–55, http://www.hirlam.org/index.php/hirlam-documentation/doc_view/517-hirlam-newsletter-no-48-article10-yang.

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

Zhang, C., Y. Wang, and K. Hamilton, 2011: Improved representation of boundary layer clouds over the southeast Pacific in ARW-WRF using a modified Tiedtke cumulus parameterization scheme. *Mon. Wea. Rev.*, **139**, 3489–3513, https://doi.org/10.1175/MWR-D-10-05091.1.

Zhang, F., C. Snyder, and R. Rotunno, 2003: Effects of moist convection on mesoscale predictability. *J. Atmos. Sci.*, **60**, 1173–1185, https://doi.org/10.1175/1520-0469(2003)060<1173:EOMCOM>2.0.CO;2.

——, M. Zhang, and J. Poterjoy, 2013: E3DVar: Coupling an ensemble Kalman filter with three-dimensional variational data assimilation in a limited-area weather prediction model and comparison to E4DVar. *Mon. Wea. Rev.*, **141**, 900–917, https://doi.org/10.1175/MWR-D-12-00075.1.

Zhang, H., J. Chen, X. Zhi, Y. Wang, and Y. Wang, 2015: Study on multi-scale blending initial condition perturbations for a regional ensemble prediction system. *Adv. Atmos. Sci.*, **32**, 1143–1155, https://doi.org/10.1007/s00376-015-4232-6.

Zhang, X., 2018: Application of a convection-permitting ensemble prediction system to quantitative precipitation forecasts over southern China: Preliminary results during SCMREX. *Quart. J. Roy. Meteor. Soc.*, **144**, 2842–2862, https://doi.org/10.1002/qj.3411.

Zhou, X., Y. Zhu, D. Hou, Y. Luo, J. Peng, and D. Wobus, 2017: Performance of the new NCEP global ensemble forecast system in a parallel experiment. *Wea. Forecasting*, **32**, 1989–2004, https://doi.org/10.1175/WAF-D-17-0023.1.