Short-Term Convection-Allowing Ensemble Precipitation Forecast Sensitivity to Resolution of Initial Condition Perturbations and Central Initial States

CRAIG S. SCHWARTZ,[a,b] JONATHAN POTERJOY,[b,c] GLEN S. ROMINE,[a] DAVID C. DOWELL,[d] JACOB R. CARLEY,[e] AND JAMIE BRESCH[a]

[a] *National Center for Atmospheric Research, Boulder, Colorado*
[b] *University of Maryland, College Park, College Park, Maryland*
[c] *NOAA/Atlantic Oceanographic and Meteorological Laboratory, Miami, Florida*
[d] *NOAA/Earth System Research Laboratory, Boulder, Colorado*
[e] *NOAA/NWS/NCEP/Environmental Modeling Center, College Park, Maryland*

(Manuscript received 7 October 2021, in final form 22 March 2022)

ABSTRACT: Nine sets of 36-h, 10-member, convection-allowing ensemble (CAE) forecasts with 3-km horizontal grid spacing were produced over the conterminous United States for a 4-week period. These CAEs had identical configurations except for their initial conditions (ICs), which were constructed to isolate CAE forecast sensitivity to resolution of IC perturbations and central initial states about which IC perturbations were centered. The IC perturbations and central initial states were provided by limited-area ensemble Kalman filter (EnKF) analyses with both 15- and 3-km horizontal grid spacings, as well as from NCEP's Global Forecast System (GFS) and Global Ensemble Forecast System. Given fixed-resolution IC perturbations, reducing horizontal grid spacing of central initial states improved ~1–12-h precipitation forecasts. Conversely, for constant-resolution central initial states, reducing horizontal grid spacing of IC perturbations led to comparatively smaller short-term forecast improvements or none at all. Overall, all CAEs initially centered on 3-km EnKF mean analyses produced objectively better ~1–12-h precipitation forecasts than CAEs initially centered on GFS or 15-km EnKF mean analyses *regardless of IC perturbation resolution*, strongly suggesting it is more important for central initial states to possess fine-scale structures than IC perturbations for short-term CAE forecasting applications, although fine-scale perturbations could potentially be critical for data assimilation purposes. These findings have important implications for future operational CAE forecast systems and suggest CAE IC development efforts focus on producing the best possible high-resolution deterministic analyses that can serve as central initial states for CAEs.

SIGNIFICANCE STATEMENT: Ensembles of weather model forecasts are composed of different "members" that, when combined, can produce probabilities that specific weather events will occur. Ensemble forecasts begin from specified atmospheric states, called initial conditions. For ensembles where initial conditions differ across members, the initial conditions can be viewed as a set of small perturbations added to a central state provided by a single model field. Our study suggests it is more important to increase horizontal resolution of the central state than resolution of the perturbations when initializing ensemble forecasts with 3-km horizontal grid spacing. These findings suggest a potential for computational savings and a streamlined process for improving high-resolution ensemble initial conditions.

KEYWORDS: Ensembles; Forecast verification/skill; Numerical weather prediction/forecasting; Model evaluation/performance

1. Introduction

An ensemble of initial conditions (ICs) can be viewed as a set of IC perturbations added to a deterministic model solution. In this framework, the deterministic solution serves as a central initial state[1] for the IC ensemble. Theoretically, central initial states and IC perturbations can originate from disparate sources with different underlying physics, dynamics, and resolutions, which is common for convection-allowing ensemble (CAE) applications. For example, CAE ICs have regularly been constructed by adding perturbations derived from relatively coarse analyses or short-term forecasts to comparatively higher-resolution deterministic analyses (e.g., Xue

[1] This deterministic central state is often, but not necessarily, exactly the mean of the IC ensemble. For example, many analysis–forecast systems using ensemble Kalman filters (EnKFs) produce N-member analysis ensembles but only initialize "free forecasts" of interest from M members, where $M < N$ (e.g., Houtekamer et al. 2014; Schwartz et al. 2015a; Johnson et al. 2017; Zhou et al. 2017; Gasperoni et al. 2020). Through EnKF equations, all N posterior (after assimilation) ensemble members, including the subset of M members, are naturally *centered* on the ensemble mean of the N members ($\overline{x_N}$). However, if only initializing free forecasts from M members, the *mean* of the M-member IC ensemble ($\overline{x_M}$) is clearly not necessarily $\overline{x_N}$. Although differences between $\overline{x_N}$ and $\overline{x_M}$ are small in equally likely, single-physics, single-dynamics ensembles like those considered in this study, given the above technical considerations, we prefer the term "central initial state" instead of "mean initial state."

Corresponding author: Craig Schwartz, schwartz@ucar.edu

DOI: 10.1175/WAF-D-21-0165.1

et al. 2008; Peralta et al. 2012; Kühnlein et al. 2014; Tennant 2015; Raynaud and Bouttier 2016, 2017; Hagelin et al. 2017; Johnson and Wang 2020).

High-quality central initial states and IC perturbations are both critical for producing skillful and reliable probabilistic CAE forecasts. Central initial states establish an overall forecast trajectory about which IC perturbations evolve (e.g., Ancell 2013) and IC perturbations are important contributors to CAE spread, especially at short forecast ranges before lateral boundary condition (LBC) or physics perturbations generate appreciable forecast diversity (e.g., Hohenegger et al. 2008; Vié et al. 2011; Peralta et al. 2012; Kühnlein et al. 2014; Zhang 2019). Thus, to improve CAE forecasts, it is important to improve both central initial states and IC perturbations.

One way to potentially realize these improvements is to increase horizontal resolutions of central initial states and IC perturbations to convection-allowing scales, as numerous studies have indicated short-term (e.g., ~1–12-h) convection-allowing model forecasts are improved when initialized from corresponding convection-allowing analyses, rather than from coarser convection-parameterizing analyses (e.g., Ancell 2012; Johnson et al. 2015; Johnson and Wang 2016; Schwartz 2016; Lu et al. 2017; Gustafsson et al. 2018; Schwartz et al. 2021). Given that central initial state and IC perturbation resolutions can differ, it seems prudent to assess whether it is *necessary* for both central initial states and IC perturbations to possess convection-allowing horizontal grid spacing. In other words, are CAE forecasts degraded if one of the IC components possesses convection-parameterizing, rather than convection-allowing, resolution?

The answer to this question has important implications for how next-generation CAEs, like NCEP's ensemble-based Rapid Refresh Forecast System (RRFS; Carley et al. 2021), are designed. For example, substantial computational resources can potentially be saved if increasing central initial state resolution to convection-allowing scales significantly improves CAE forecasts but increasing IC perturbation resolution to convection-allowing scales has comparatively smaller impacts. In this case, RRFS development efforts can primarily be devoted to producing high quality, convection-allowing, deterministic central initial states about which relatively coarse and inexpensive IC perturbations are centered.[2] Conversely, if increasing IC perturbation resolution to convection-allowing scales is more important, a stronger emphasis should be placed on developing a pure ensemble-based convection-allowing data assimilation (DA) system for the RRFS.

Although previous work has not directly assessed the relative benefits of increasing IC perturbation resolution versus increasing central initial state resolution for CAE forecasting applications, several recent studies touched on issues concerning central initial states and IC perturbations for real-world CAEs. For example, Schwartz et al. (2020, hereafter S20) suggested CAE precipitation forecasts were more sensitive to central initial states than IC perturbations. However, the three sources of IC perturbations and two sources of central initial states considered by S20 all possessed convection-parameterizing resolution and reflected vastly different underlying numerical weather prediction (NWP) models and DA systems. Thus, it is unclear if S20's findings would hold for finer-scale ICs or in frameworks with more unified system configurations among central initial states and IC perturbations.

In addition, Schwartz et al. (2021, hereafter S21) showed that 3-km ensemble Kalman filter (EnKF; Evensen 1994; Houtekamer and Zhang 2016) analyses initialized better short-term 3-km CAE precipitation forecasts than downscaled 15-km EnKF analyses. S21's 3-km EnKF ICs can be viewed as 3-km IC perturbations centered on 3-km states (the mean of the 3-km EnKF analysis members), while their 15-km EnKF ICs can be conceived as 15-km IC perturbations centered on 15-km states (the mean of the 15-km EnKF analysis members). Therefore, S21's experiments could not disentangle precisely whether a specific component of the 3-km EnKF—its higher-resolution central initial state or higher-resolution IC perturbations—was responsible for yielding better short-term forecasts than the 15-km EnKF.

Furthermore, Johnson and Wang (2020, hereafter JW20) examined ten 18-h forecasts over a small (1200 km × 1200 km) domain from CAEs centered about common 3-km initial states but with different IC perturbations. Their results indicated that 3-km IC perturbations led to better CAE forecasts than IC perturbations derived from a 0.5° global ensemble. However, benefits of higher-resolution IC perturbations steadily decreased with forecast lead time and generally vanished beyond 12–15 h, potentially due to their small domain. These findings were similar to Raynaud and Bouttier (2016), who noted that 4-km IC perturbations yielded better 9–12-h CAE forecasts than 15-km IC perturbations given common 2.5-km central initial states. While these collective results suggest short-term CAE forecasts benefit from possessing convection-allowing IC perturbations, neither Raynaud and Bouttier (2016) nor JW20 concurrently examined CAE forecast sensitivity to central initial state resolution.

Thus, to assess whether it is more important for IC perturbations or central initial states to possess convection-allowing horizontal grid spacing, we designed, executed, and evaluated a series of CAE forecast experiments based upon S21's EnKFs. Sections 2 and 3 describe our experiments, while section 4 presents results. Although our focus is on IC resolution requirements for CAE forecasting applications, there are parallels between this topic and resolution requirements for dual-resolution ensemble-based DA systems, which we discuss in section 5. Our overall conclusions (section 6) provide guidance about how development efforts to improve CAE ICs might best proceed.

2. Model and EnKF configurations

Thirty-six-hour (36-h), 10-member CAE forecasts were initialized from 9 sets of ICs based on 3 sources of IC perturbations

[2] Potential caveats: 1) Processes to generate IC perturbations could be improved; and 2) even if IC perturbations can be coarsened without harming CAE forecasts, CAEs could be critical for providing flow-dependent background error covariances within convective-scale data assimilation systems. We discuss this possibility more thoroughly in section 5a.

TABLE 1. Summary of continuously cycling EnKF configurations. See S21 for justifications for these settings.

Parameter	Setting
EnKF algorithm	Ensemble adjustment Kalman filter (EAKF; Anderson 2001, 2003; Anderson and Collins 2007)
Ensemble size	80 members
Cycling period	1 h
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	For 15-km EnKF analyses, 1280 km for all observations; for 3-km EnKF analyses, 640 km, except 1280 km for rawinsonde observations
Vertical localization full width	1.0 scale height
Sampling error correction	Anderson (2012)
Inflation method	Posterior relaxation-to-prior spread (RTPS; Whitaker and Hamill 2012)
Inflation factor	1.06
Lateral boundary condition perturbations	Random perturbations based on Gaussian noise added to GFS analyses and forecasts (e.g., Torn et al. 2006)
Sea surface temperature updates	Daily updates from NCEP's 0.12° analyses (e.g., Gemmill et al. 2007)
Assimilated observations	Rawinsonde, aircraft, wind profiler, satellite-tracked wind, global positioning system radio occultation (GPSRO), and surface observations
Observation errors and time windows	Based on the High-Resolution Rapid Refresh Ensemble (HRRRE; Dowell et al. 2016, 2022)
Horizontal thinning for aircraft and satellite-tracked wind observations	30 km for 15-km EnKF analyses; 15 km for 3-km EnKF analyses
Vertical thinning for aircraft and satellite-tracked wind observations	25 hPa

and 3 sources of central initial states (section 3c). Eight of the IC sets leveraged analyses produced by the continuously cycling EnKF DA systems described by S21, who provided comprehensive details and justifications for specific EnKF DA settings (summarized in Table 1). Thus, only brief descriptions of EnKF and NWP model configurations are provided here, with a more thorough discussion reserved for how the nine sets of ICs were constructed (section 3).

Specifically, S21 performed two limited-area continuously cycling DA experiments using a square root form of the EnKF (Anderson 2001) implemented within the Data Assimilation Research Testbed software (Anderson et al. 2009). Both EnKFs used 80 ensemble members and produced analyses every hour between 0000 UTC 23 April and 0000 UTC 20 May 2017 (inclusive; 649 hourly cycles). As noted by S21, this experimental period featured a variety of flow patterns and several heavy precipitation episodes primarily driven by strong synoptic forcing. Both EnKFs assimilated approximately 30 000–100 000 observations each cycle, and the first two days of cycling (i.e., 23 and 24 April) were regarded as spinup where model solutions moved away from randomly perturbed initial states prescribed at 0000 UTC 23 April 2017. S21 demonstrated that two days was a sufficiently long spinup period.

One EnKF experiment produced analyses solely on a 15-km computational domain (Figs. 1 and 2a), while the second produced analyses on both 15- and 3-km domains (Figs. 1 and 2b). To advance the 80-member ensemble states between hourly analyses, version 3.9.1.1 of the Advanced Research version of the Weather Research and Forecasting (WRF) Model (Skamarock et al. 2008; Powers et al. 2017) was used. WRF Model physical parameterizations (Table 2) were identical across all 80 ensemble members and the two

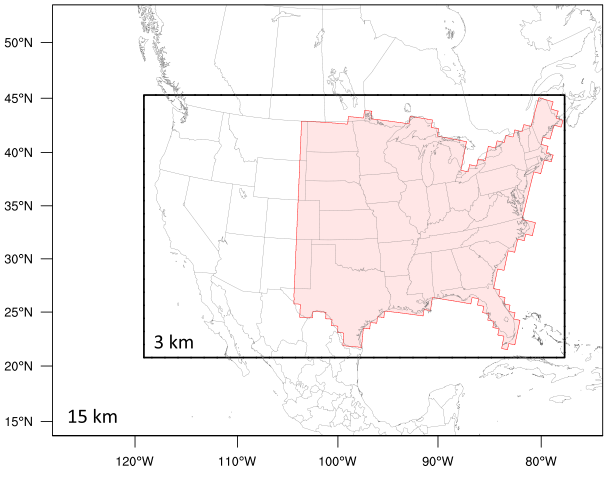

FIG. 1. Computational domain. Horizontal grid spacing was 15 km in the outer domain (415 × 325 points) and 3 km in the nest (1581 × 986 points). Lateral boundary conditions (LBCs) provided by global models were applied to the 15-km domain, which in turn provided LBCs for the 3-km domain. Objective precipitation verification only occurred within the red shaded region of the 3-km domain (CONUS east of 105°W).

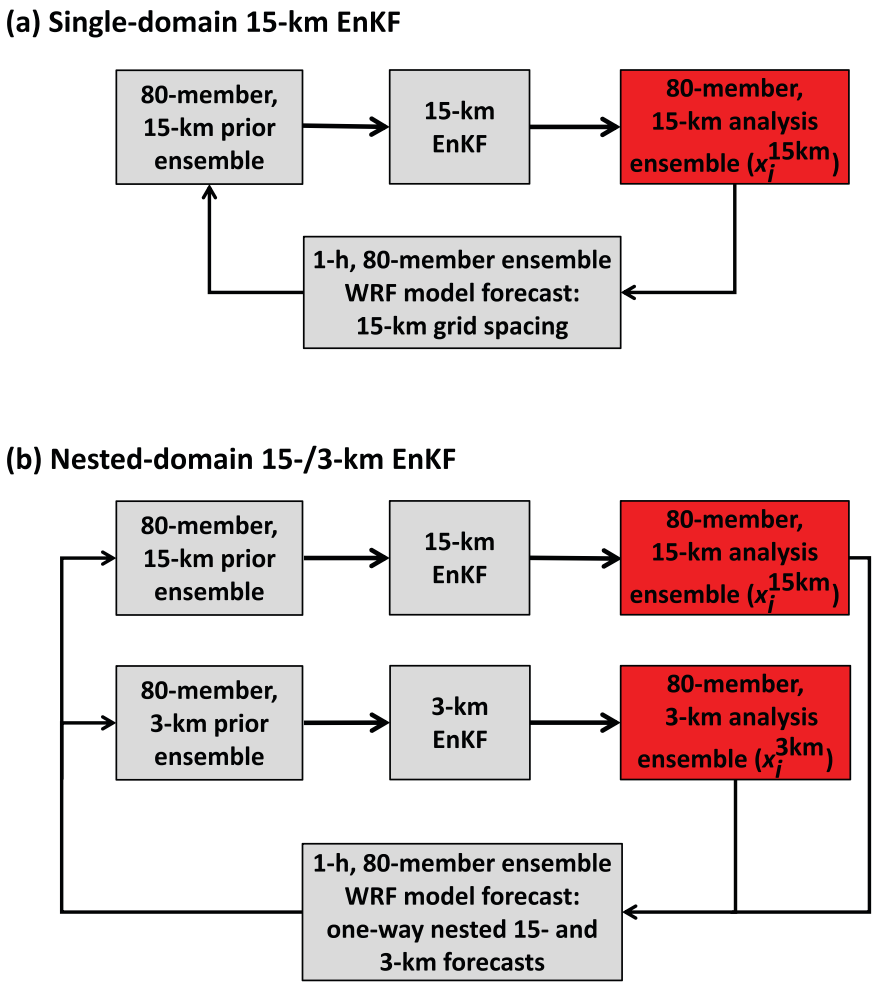

FIG. 2. Flowcharts of continuously cycling EnKF data assimilation systems over (a) solely the 15-km computational domain (i.e., outer domain in Fig. 1) and (b) both the 15- and 3-km computational domains (i.e., both domains in Fig. 1). Posterior ensembles (red shaded boxes) at 0000 UTC were used to construct initial conditions for CAE forecasts. In the nested-domain 15-/3-km EnKF data assimilation system, the 15-km domain provided lateral boundary conditions for the 3-km domain during WRF Model forecasts.

domains, except no cumulus parameterization was employed on the convection-allowing 3-km grid. In the nested 15-/3-km EnKF DA system (Fig. 2b), 1-h WRF Model forecasts between analyses were one-way nested, ensuring that the 15-km EnKF was unaffected by the 3-km EnKF (i.e., 15-km fields in the nested- and single-domain EnKF DA systems were identical). Both EnKFs performed well, yielding acceptable spread–error relationships (e.g., Houtekamer et al. 2005) and similar model climates that were stable throughout the 4 weeks of continuous cycling (see Figs. 5 and 6 in S21).

The 15- and 3-km EnKFs had very similar configurations, but they differed in terms of their horizontal localization for non-rawinsonde observations and horizontal thinning of satellite-tracked wind and aircraft observations (Table 1). As noted by S21, the smaller horizontal localization in the 3-km EnKF was used to lessen computational expense, and the different horizontal observation 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" (S21). Appendix A shows that differences regarding horizontal localization and observation thinning were not responsible for differences in 15- and 3-km EnKF performance.

3. Experimental design

Analysis ensembles from the 15- and 3-km EnKFs at 0000 UTC (i.e., red boxes in Fig. 2) were used to derive central initial states and IC perturbations for various CAEs. Operational global models were also used as sources of IC perturbations and central initial states.

Ultimately, 3-km precipitation forecasts initialized at 0000 UTC from various sets of ICs were verified over the conterminous

TABLE 2. WRF Model settings used for the EnKFs and CAE forecasts.

Parameter	WRF Model setting
Model version	Version 3.9.1.1 of the Advanced Research WRF Model
Horizontal grid spacing	15 and 3 km in the outer and inner domains, respectively
Time step	60 and 12 s in the 15- and 3-km domains, respectively
Number of vertical levels	51 (based on the Rapid Refresh model; Benjamin et al. 2016)
Model top	15 hPa
Microphysics parameterization	Thompson (Thompson et al. 2008)
Longwave and shortwave radiation parameterizations	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 parameterization	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)

United States (CONUS) east of the Rockies (Fig. 1), where validation data are most robust (discussed further in section 4a). Employing a nested WRF Model configuration provided a computationally affordable way to place this evaluation region far from lateral boundaries to minimize LBC impacts (e.g., Warner et al. 1997). However, using a nested configuration required ICs to be constructed for both the 15- and 3-km domains, which somewhat complicated the experimental design. While we offer some remarks about initializing the 15-km domain in upcoming subsections, the primary function of the 15-km domain was to provide LBCs for the 3-km domain, and the most important aspect of the experimental design concerned how ICs were produced for the 3-km domain. Note that ICs for the 3-km domain did not have to possess 3-km horizontal grid spacing, as coarser fields could be downscaled onto the 3-km grid, as described below.

a. *Central initial states for the 3-km domain*

Central initial states for the 3-km domain were provided by three sources. One source was 3-km EnKF mean analyses ($\overline{x^{3\,\mathrm{km}}}$), given by

$$\overline{x^{3\,\mathrm{km}}} = \frac{1}{80}\sum_{i=1}^{80} x_i^{3\,\mathrm{km}}, \tag{1}$$

and another source was 15-km EnKF mean analyses ($\overline{x^{15\,\mathrm{km}}}$), given by

$$\overline{x^{15\,\mathrm{km}}} = \frac{1}{80}\sum_{i=1}^{80} x_i^{15\,\mathrm{km}}, \tag{2}$$

where $x_i^{15\,\mathrm{km}}$ and $x_i^{3\,\mathrm{km}}$ denote 15- and 3-km EnKF analyses for the ith of 80 ensemble members, respectively.[3] Operational Global Forecast System (GFS) analyses (x^{GFS}) with 0.25° horizontal grid spacing[4] also served as central initial states for the 3-km domain. From a spectral perspective using the discrete cosine transform (Denis et al. 2002), all three sources of central initial states were usually similar on mutually resolvable scales, with the largest differences at scales $<$ 100 km that were not resolvable by all analyses (Fig. 3). In addition to providing central states for CAE ICs, $\overline{x^{3\,\mathrm{km}}}$, $\overline{x^{15\,\mathrm{km}}}$, and x^{GFS} were used to initialize deterministic forecasts (section 4b).

Differences between 15- and 3-km EnKF mean analyses were attributable to differences in horizontal grid spacing and associated representation of convection (parameterized versus explicit). Conversely, because GFS and limited-area EnKF analyses reflected entirely different NWP models and DA systems, differences between GFS and limited-area EnKF mean analyses were potentially due to many factors other than disparities in resolution and associated treatment of convection. Nonetheless, employing GFS analyses as a source of central initial states was useful for examining whether CAE forecast sensitivity to IC perturbations depends on the central initial state and provides insights about whether there are benefits of using limited-area EnKFs to generate central initial states relative to adopting those given by a well-tuned operational global analysis system.

While 3-km EnKF analyses were dynamically consistent with the 3-km CAE forecast model, 15-km EnKF analyses, and especially GFS analyses, were not. Thus, differences regarding dynamical consistency between the various central initial states and the CAE forecast model may impact results, in addition to differences in IC resolution. It is difficult to disentangle the influence of these two factors (dynamical consistency and IC resolution) in our experimental framework, in particular when comparing CAEs with GFS central initial states to CAEs with limited-area EnKF central initial states. However, the extent to which 15- and 3-km EnKF-based ICs were dynamically consistent with the CAE forecast model was a direct consequence of horizontal resolution (i.e., 15-km EnKF analyses were dynamically inconsistent with the 3-km CAE forecast model by virtue of their 15-km grid spacing).

[3] Although we refer to EnKF ensemble means (e.g., $\overline{x^{3\,\mathrm{km}}}$) as possessing identical resolutions to individual ensemble members (e.g., $x_i^{3\,\mathrm{km}}$), ensemble means are effectively coarser than individual ensemble members because of spatial smoothing inherent in ensemble averaging (e.g., Leith 1974; Surcel et al. 2014).

[4] The GFS had approximately 13-km horizontal grid spacing, but data available to us were coarsened to 0.25° by NCEP.

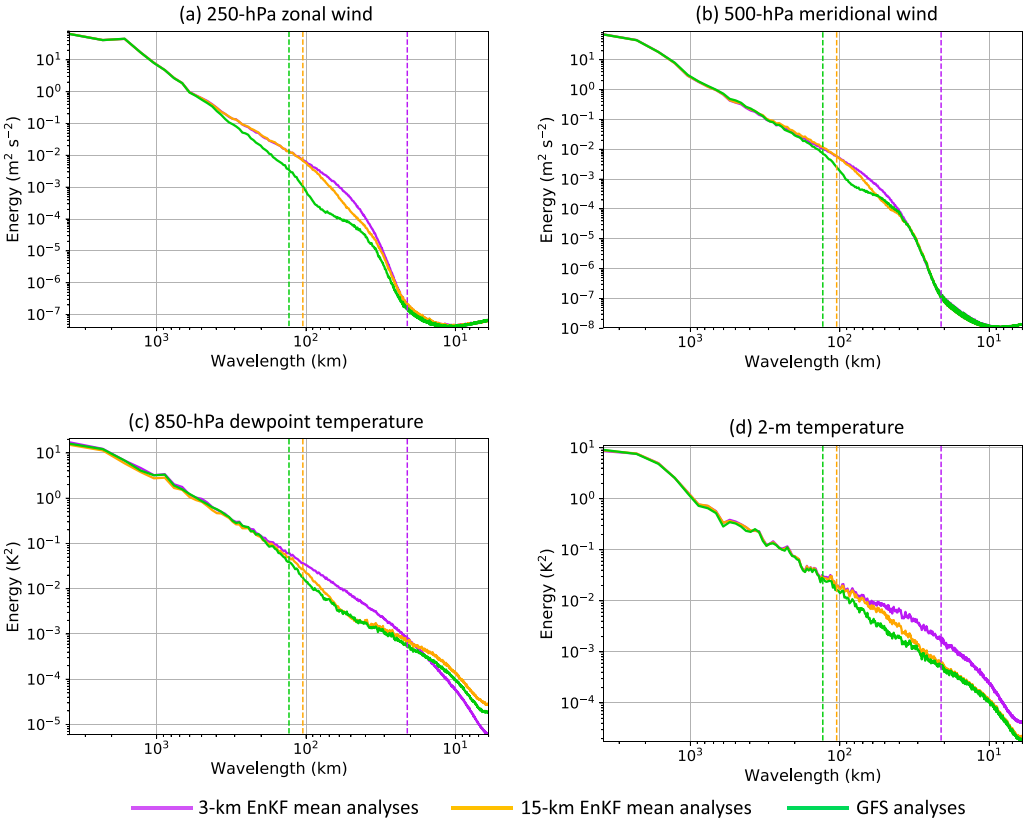

FIG. 3. Power spectra of 3-km EnKF mean analyses (purple), 15-km EnKF mean analyses (orange), and GFS analyses (green) as a function of wavelength (km) for (a) 250-hPa zonal wind (m^2 s^{-2}), (b) 500-hPa meridional wind (m^2 s^{-2}), (c) 850-hPa dewpoint temperature (K^2), and (d) 2-m temperature (K^2) averaged over all 0000 UTC analyses between 25 Apr and 20 May 2017 (inclusive). To compute spectra from the various datasets, all fields were interpolated onto the 3-km domain (Fig. 1). Spectra were then computed over the geographic area covered by the 3-km domain, excluding points within 45 km of each lateral boundary. The discrete cosine transform (Denis et al. 2002) was used to compute spectra and spectral variance binning employed the method of Ricard et al. (2013). Dashed purple, orange, and green vertical lines denote effective resolutions of 3-km EnKF mean analyses, 15-km EnKF mean analyses, and GFS analyses, respectively, which were approximately 7 times the horizontal grid spacing for the WRF-based EnKFs (e.g., Skamarock 2004) and approximately 10 times the horizontal grid spacing for the GFS (e.g., Ji et al. 2016). The *y*-axis values are different in each panel.

Therefore, we attribute differences between CAEs with common IC perturbations but 15- or 3-km EnKF-based central initial states to the different central initial state resolutions.

b. *IC perturbations for the 3-km domain*

IC perturbations for the 3-km domain were derived from three sources, including both 15- and 3-km EnKF analysis ensembles (e.g., Fig. 2), given by

$$\delta x_i^{15\text{km}} = x_i^{15\text{km}} - \overline{x^{15\text{km}}}, \text{and} \tag{3}$$

$$\delta x_i^{3\text{km}} = x_i^{3\text{km}} - \overline{x^{3\text{km}}}, \tag{4}$$

where $\delta x_i^{15\text{km}}$ and $\delta x_i^{3\text{km}}$, respectively, denote 15- and 3-km posterior (after assimilation) EnKF perturbations for the ith ensemble member. Although Eqs. (3) and (4) were valid for $i = 1, \ldots, 80$, IC perturbations from just members 1–10 were required because CAE forecasts only had 10 members (as explained in section 3d). Thus, there was some uncertainty about whether EnKF-based IC perturbations should be computed with respect to the mean of all 80 posterior members [as in Eqs. (3) and (4)] or just the mean of posterior members 1–10. However, auxiliary experiments revealed that CAE forecasts were insensitive to whether IC perturbations were computed with respect to the mean of posterior members 1–80 or 1–10 (not shown).

ICs from NCEP's operational 0.5° Global Ensemble Forecast System (GEFS; Zhou et al. 2017) provided the third source of IC perturbations for the 3-km domain.[5] Specifically, perturbations for the ith GEFS member (δx_i^{GEFS}) were given by

$$\delta x_i^{\text{GEFS}} = x_i^{\text{GEFS}} - \overline{x^{\text{GEFS}}}, \tag{5}$$

[5] The GEFS had approximately 34-km horizontal grid spacing, but data available to us were coarsened to 0.5° by NCEP.

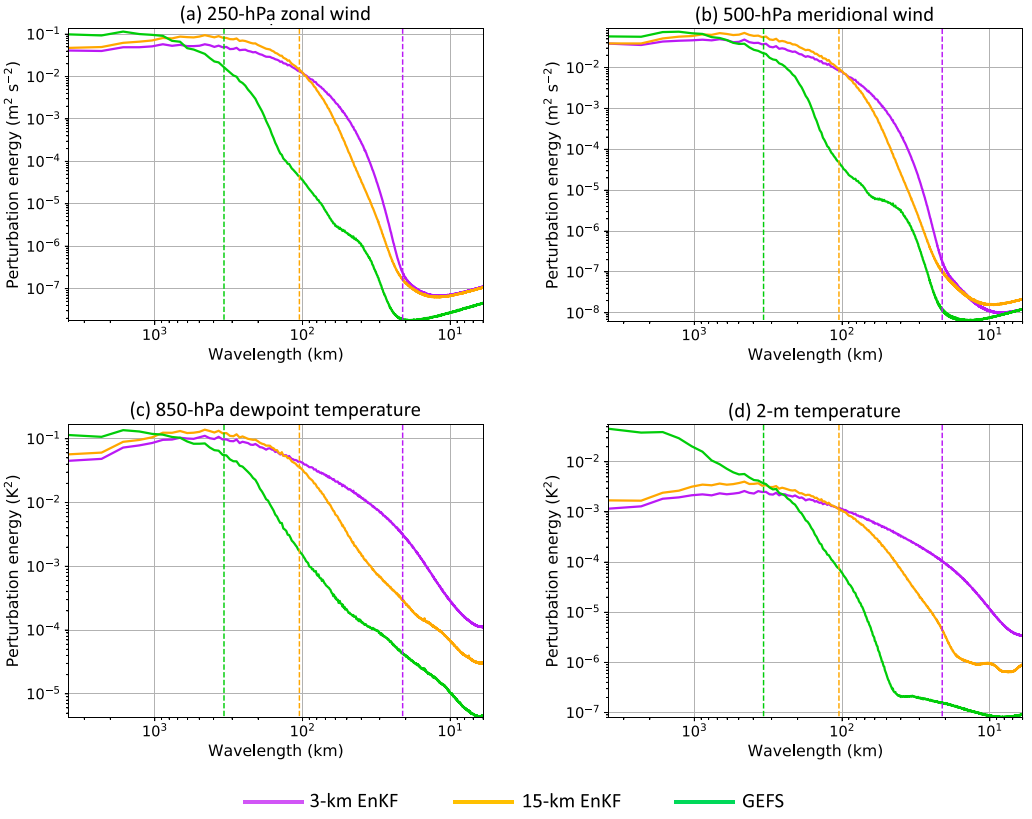

FIG. 4. Perturbation power spectra of 3-km EnKF analysis ensembles (purple), 15-km EnKF analysis ensembles (orange), and GEFS initial condition ensembles (green) as a function of wavelength (km) for (a) 250-hPa zonal wind (m^2 s^{-2}), (b) 500-hPa meridional wind (m^2 s^{-2}), (c) 850-hPa dewpoint temperature (K^2), and (d) 2-m temperature (K^2) averaged over all ensemble perturbations and all 0000 UTC analyses between 25 Apr and 20 May 2017 (inclusive). Perturbations were defined with respect to the ensemble mean. To compute spectra from the various datasets, all fields were interpolated onto the 3-km domain (Fig. 1). Spectra were then computed over the geographic area covered by the 3-km domain, excluding points within 45 km of each lateral boundary. The discrete cosine transform (Denis et al. 2002) was used to compute spectra, and spectral variance binning employed the method of Ricard et al. (2013). Dashed purple, orange, and green vertical lines denote effective resolutions of the 3-km ensemble, 15-km ensemble, and GEFS, respectively, which were approximately 7 times the horizontal grid spacing for the WRF-based ensembles (e.g., Skamarock 2004) and approximately 10 times the horizontal grid spacing for the GEFS (e.g., Ji et al. 2016). The *y*-axis values are different in each panel.

where x_i^{GEFS} denotes GEFS ICs for the ith ensemble member (for $i = 1, \ldots, 10$) and

$$\overline{x^{\mathrm{GEFS}}} = \frac{1}{10}\sum_{i=1}^{10} x_i^{\mathrm{GEFS}}. \tag{6}$$

While the GEFS had 20 perturbation members during our experimental period (April–May 2017), as with the limited-area EnKF perturbations, GEFS-based IC perturbations were derived for just perturbation members 1–10, again because subsequent CAE forecasts only had 10 members. Like the 15- and 3-km IC perturbations, GEFS IC perturbations were EnKF based, and, thus, flow dependent (Zhou et al. 2017).

The 15-km EnKF IC perturbations had more energy than 3-km EnKF IC perturbations at most mutually resolvable scales (Fig. 4), possibly due to the 15-km EnKF's use of cumulus parameterization that may impart a large-scale structure to the errors (e.g., Torn and Davis 2012; Romine et al. 2013; Mahoney 2016; Wong et al. 2020), whereas the 3-km EnKF was convection-allowing. However, 3-km IC perturbations had the most power at scales < 100 km. Differences between 15- and 3-km EnKF IC perturbation power spectra were typically small compared to their collective differences with respect to GEFS IC perturbation power spectra, which had the most perturbation energy at scales > 1000 km. Similarly, differences of domain-average spread between 15- and 3-km posterior ensembles were small compared to differences between EnKF and GEFS IC ensembles, the latter of which had uniformly smaller spreads except for low-level temperature[6] (Fig. 5).

[6] This enhanced low-level temperature spread from GEFS IC perturbations exclusively occurred over the high plains adjacent to the Rocky Mountains and other regions of the intermountain western CONUS for reasons that are unclear.

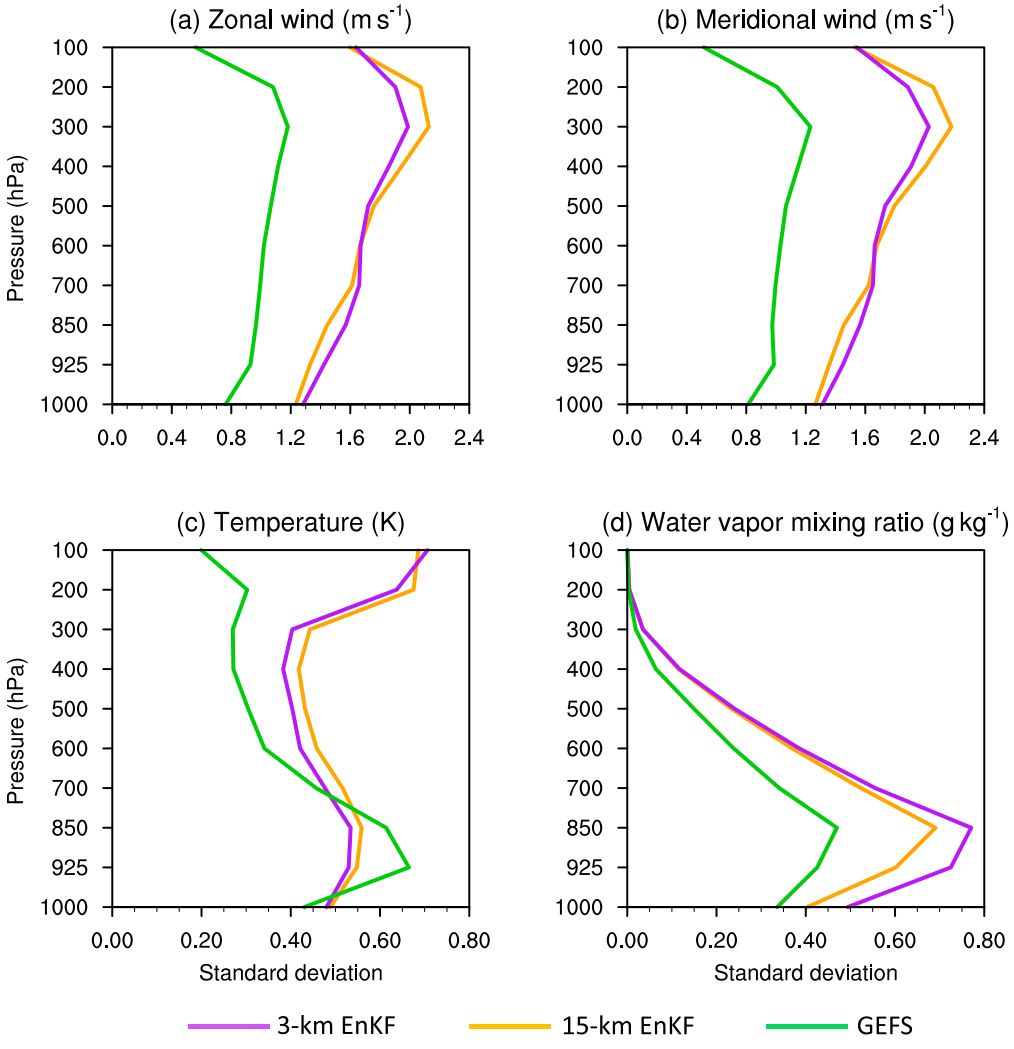

FIG. 5. Average standard deviation over the geographic area covered by the 3-km domain (Fig. 1) and all 0000 UTC 3-km EnKF analysis ensembles (purple), 15-km EnKF analysis ensembles (orange), and GEFS IC ensembles (green) between 0000 UTC 25 Apr and 0000 UTC 20 May 2017 (inclusive) for (a) zonal wind (m s^{-1}), (b) meridional wind (m s^{-1}), (c) temperature (K), and (d) water vapor mixing ratio (g kg^{-1}). The *x*-axis values are different in each row.

c. *IC construction*

At 0000 UTC daily between 25 April and 20 May 2017 (inclusive), each of the three sets of IC perturbations [Eqs. (3)–(5)] was re-centered about each of the three sets of central initial states [Eq. (1), Eq. (2), and GFS analyses (x^{GFS})], yielding nine sets of IC ensembles that differed by their IC perturbations and central initial states (Tables 3 and 4). While re-centering is a common CAE initialization technique (e.g., Xue et al. 2008; Kong et al. 2008, 2009; Peralta et al. 2012; Kühnlein et al. 2014; Tennant 2015; Johnson and Wang 2016; Raynaud and Bouttier 2016, 2017; Hagelin et al. 2017; JW20; S20), an alternative way of assessing forecast sensitivity to IC resolution would be to remove small-scale features from 3-km EnKF posterior ensembles with a low-pass filter to produce coarser central initial states and IC perturbations (e.g., Potvin et al. 2017; JW20). Although this method is elegant, we did not apply filtering to construct ICs with varied resolutions because re-centering has greater relevance within operational environments, where output from multiple modeling systems with different resolutions is typically available. Furthermore, re-centering is common within operational DA systems (e.g., Clayton et al. 2013; Wang et al. 2013).

Similar to S20 and Peralta et al. (2012), only zonal and meridional wind, potential temperature, water vapor mixing ratio, and perturbation geopotential and dry surface pressure[7] (U, V, θ, q_v, ϕ, and μ, respectively) were re-centered when

[7] Perturbation geopotential and dry surface pressure were defined relative to a base state per the WRF Model formulation (Skamarock et al. 2008).

TABLE 3. Experiment names and their corresponding central initial states (columns) and IC perturbations (rows) for the 3-km domain. All IC sets were constructed at 0000 UTC each day during the experimental period. Bolded experiments denote the four "EnKF-only" experiments, while italicized experiments denote the four "EnKF+Global" experiments.

		Central initial state for the 3-km domain		
		3-km EnKF mean analyses	15-km EnKF mean analyses downscaled onto the 3-km domain	0.25° GFS analyses downscaled onto the 3-km domain
IC perturbations for the 3-km domain	3-km EnKF analysis perturbations	**3kmCent_3kmPert**	**15kmCent_3kmPert**	*GFSCent_3kmPert*
	15-km EnKF analysis perturbations downscaled onto the 3-km domain	**3kmCent_15kmPert**	**15kmCent_15kmPert**	*GFSCent_15kmPert*
	0.5° GEFS IC perturbations downscaled onto the 3-km domain	*3kmCent_GEFSPert*	*15kmCent_GEFSPert*	GFSCent_GEFSPert

constructing ICs for the ith ensemble member. All other fields for the ith member's ICs, like hydrometeors, were provided by the corresponding member of either the 15- or 3-km posterior ensemble depending on the finest-resolution field used to construct a specific set of ICs. For example, for ICs with either 3-km central initial states or 3-km IC perturbations, initial microphysics states for the ith member were provided by the corresponding member of the 3-km posterior ensemble (i.e., $x_i^{3\text{km}}$). Similarly, for ICs where the finest resolution among central initial states and IC perturbations was 15 km, initial microphysics states for the ith member were provided by the corresponding member of the 15-km posterior ensemble (i.e., $x_i^{15\text{km}}$). We used variables from ensemble members for these fields to provide IC diversity, rather than forcing these other variables to common values provided by deterministic central initial states. However, initialization of these auxiliary fields likely has little forecast impact, as S20 suggested that precipitation forecasts over the CONUS east of the Rockies were more sensitive to IC perturbation characteristics of U, V, θ, q_v, ϕ, and μ than initial hydrometeor states.

TABLE 4. Mathematical expressions for the ith ensemble member's ICs for the 3-km domain in the various experiments. Term D represents a downscaling operator within the WRF Model that remaps fields with horizontal grid spacing coarser than 3 km (e.g., 15-km EnKF analyses) onto the 3-km computational domain. Terms $x_i^{15\text{km}}$ and $x_i^{3\text{km}}$ denote 15- and 3-km EnKF analyses for the ith ensemble member, respectively, and x^{GFS} denotes GFS analyses. All other terms are defined in Eqs. (1)–(5).

Experiment name	Expression for the ith ensemble member's ICs for the 3-km domain
3kmCent_3kmPert	$\overline{x^{3\text{km}}} + \delta x_i^{3\text{km}}$
3kmCent_15kmPert	$\overline{x^{3\text{km}}} + D\delta x_i^{15\text{km}}$
3kmCent_GEFSPert	$\overline{x^{3\text{km}}} + D\delta x_i^{\text{GEFS}}$
15kmCent_3kmPert	$D\overline{x^{15\text{km}}} + \delta x_i^{3\text{km}}$
15kmCent_15kmPert	$D\overline{x^{15\text{km}}} + D\delta x_i^{15\text{km}}$
15kmCent_GEFSPert	$D\overline{x^{15\text{km}}} + D\delta x_i^{\text{GEFS}}$
GFSCent_3kmPert	$Dx^{\text{GFS}} + \delta x_i^{3\text{km}}$
GFSCent_15kmPert	$Dx^{\text{GFS}} + D\delta x_i^{15\text{km}}$
GFSCent_GEFSPert	$Dx^{\text{GFS}} + D\delta x_i^{\text{GEFS}}$

1) ICs BASED SOLELY ON LIMITED-AREA ENKF ANALYSES

Four sets of ICs were based solely on 0000 UTC limited-area EnKF analyses and are henceforth collectively referred to as the "EnKF-only" experiments (bold experiments in Table 3). Because the nested-domain EnKF employed one-way nesting (section 2, Fig. 2b), ICs for the 15-km domain across all EnKF-only experiments were identical, namely, 0000 UTC 15-km EnKF analysis ensembles (i.e., $x_i^{15\text{km}}$). Thus, CAE forecast differences among the four EnKF-only experiments were solely due to how the 3-km domain (Fig. 1) was initialized at 0000 UTC.

The most straightforward EnKF-based ICs for the 3-km domain were provided by 3-km EnKF analysis members (i.e., $x_i^{3\text{km}}$), which are 3-km EnKF analysis perturbations centered on 3-km EnKF mean analyses ("3kmCent_3kmPert"; Tables 3 and 4). Initial states for the 3-km domain were also provided by downscaling 15-km EnKF analysis members (i.e., $x_i^{15\text{km}}$) onto the 3-km computational domain ("15kmCent_15kmPert"; Tables 3 and 4). The downscaling process, achieved through a monotone interpolation scheme (Smolarkiewicz and Grell 1992) per WRF Model defaults (Skamarock et al. 2008), does not add detail, so even though downscaled 15-km fields resided on the 3-km grid, the ICs still possessed spatial resolution associated with their native 15-km horizontal grid spacing.

The final two sets of ICs for the 3-km domain within the EnKF-only experiments had mixed resolutions of IC perturbations and central initial states. In one set, the 3-km domain was initialized by re-centering downscaled 15-km EnKF analysis perturbations [i.e., Eq. (3)] about 3-km EnKF mean analyses ("3kmCent_15kmPert"; Tables 3 and 4). The other set was produced by re-centering 3-km EnKF analysis perturbations [i.e., Eq. (4)] about downscaled 15-km EnKF mean analyses ("15kmCent_3kmPert"; Tables 3 and 4).

Because 3-km IC ensembles can be averaged to create 3-km central initial states [e.g., Eq. (1)], a configuration like 15kmCent_3kmPert would be unlikely within operational

environments. However, this experiment was helpful for elucidating whether it is more important for central initial states or IC perturbations to possess convection-allowing horizontal grid spacing, as CAE forecast differences between 15kmCent_3kmPert and 15kmCent_15kmPert were solely attributable to IC perturbation resolution for the 3-km domain and 15kmCent_3kmPert and 3kmCent_3kmPert only differed regarding central initial state resolution.

2) ICs based on a combination of limited-area EnKF analyses and global fields

The four EnKF-only experiments suffice to disentangle whether it is more critical for central initial states or IC perturbations to possess convection-allowing horizontal grid spacing. However, as noted earlier, initial states for the 3-km domain were also produced by leveraging GFS and GEFS fields, allowing us to assess whether ICs partially derived from readily available operational data can yield similar quality CAE forecasts as ICs derived solely from limited-area EnKFs. Comparing CAEs with GEFS and 3-km EnKF IC perturbations also replicates some of JW20's analyses, but with a larger sample size and evaluation domain. Furthermore, incorporating GFS and GEFS states into the experiments provides further insights about whether CAE forecast sensitivity to IC perturbation and central initial state resolution varies depending on whether global fields are a component of CAE ICs.

Specifically, another four sets of ICs for the 3-km domain were constructed from 0000 UTC fields that relied in part on the GFS or GEFS and are collectively referred to as the "EnKF+Global" experiments (italicized experiments in Table 3). Two sets of ICs for the 3-km domain were produced by re-centering downscaled perturbations derived from GEFS ICs [Eq. (5)] about 15- and 3-km EnKF mean analyses ("3kmCent_GEFSPert" and "15kmCent_GEFSPert"; Tables 3 and 4). The other two sets of ICs for the 3-km domain were produced by re-centering 15- and 3-km EnKF analysis perturbations [Eqs. (3) and (4)] about downscaled GFS analyses ("GFSCent_3kmPert" and "GFSCent_15kmPert"; Tables 3 and 4).

Introducing global fields poses somewhat of a dilemma: should GFS and GEFS fields participate in re-centering only for purposes of initializing the 3-km domain or for purposes of initializing *both* the 15- and 3-km domains (Fig. 1)? The former would mean all four EnKF+Global experiments have identical ICs for the 15-km domain that are equal to those of the EnKF-only experiments (i.e., $x_i^{15\,\mathrm{km}}$). Conversely, the latter would maintain consistency across both domains, which intuitively seems desirable. However, in this latter scenario, 3-km forecasts from the four EnKF+Global experiments could potentially be sensitive to LBCs inherited from different 15-km forecasts. Likewise, 3-km forecasts from the EnKF+Global experiments could potentially differ from those produced by the EnKF-only experiments because of different 15-km forecasts.

To address this conundrum, we performed several exploratory experiments where pairs of experiments solely differed by either their ICs for the 15-km domain or ICs for the 3-km domain. Findings revealed that 3-km precipitation forecasts were far more sensitive to ICs for the 3-km domain than ICs for the 15-km domain. Moreover, S21 arrived at similar conclusions (see their footnote 3). Therefore, ICs for the 15-km domain appeared to have little impact on 3-km forecasts, and differences between various 3-km CAE forecasts were attributed to different ICs for the 3-km domain. Ultimately, we chose to maintain consistency across both domains, so GEFS IC perturbations and GFS analyses were introduced to ICs for both the 15- and 3-km domains in the EnKF+Global experiments. For these experiments, expressions for the 15-km domain's ICs are analogous to those for the 3-km domain's ICs given in Table 4, except $\overline{x^{3\,\mathrm{km}}}$ and $\delta x_i^{3\,\mathrm{km}}$ are replaced by $\overline{x^{15\,\mathrm{km}}}$ and $\delta x_i^{15\,\mathrm{km}}$, respectively; GFS and GEFS fields are downscaled onto the 15-km domain; and there is no need to downscale 15-km fields.

3) ICs based solely on global fields

The final set of ICs for the 3-km domain was produced by re-centering downscaled GEFS IC perturbations about downscaled GFS analyses at 0000 UTC ("GFSCent_GEFSPert"; Tables 3 and 4). These ICs were independent of the limited-area EnKFs, and the CAE with these ICs served as a benchmark for the other CAEs whose ICs relied on limited-area EnKF analyses (Tables 3 and 4). As per the above discussion, ICs for the 15-km domain were also produced by re-centering GEFS IC perturbations about GFS analyses.

While the other eight sets of ICs had nonzero hydrometeor mixing ratios consistent with the WRF Model, microphysics variables produced by the GFS and GEFS were incompatible with the Thompson et al. (2008) microphysics scheme. Accordingly, GFSCent_GEFSPert ICs did not have hydrometeors, which is typical for WRF Model ICs provided by global analyses (i.e., initial hydrometeor mixing ratios were set to zero). Thus, a longer spinup relative to the other experiments was expected during model integration, which should be recalled when interpreting verification statistics.

d. *CAE forecasts*

At 0000 UTC daily between 25 April and 20 May 2017 (inclusive), members 1–10 from all nine sets of ICs (i.e., section 3c; Tables 3 and 4) initialized 36-h forecasts over the nested domain (Fig. 1); there were 26 10-member CAE forecasts per experiment. These 36-h forecasts employed identical WRF Model configurations as the nested 15-/3-km EnKF DA system (Table 2).

Although ICs for more than 10 ensemble members were available, computing constraints limited 36-h CAE forecasts to 10 members, which can be considered as sufficient to provide skillful and valuable probabilistic precipitation forecasts (e.g., Clark et al. 2011, 2018; Schwartz et al. 2014) and is comparable to ensemble sizes of other CAEs that operate regularly over the CONUS (e.g., Dowell et al. 2016, 2022; Roberts et al. 2020).

For all 36-h forecasts, perturbation members 1–10 from the GEFS provided LBCs for the 15-km domain, which in

turn provided LBCs for the 3-km nest. The 3-km forecasts were then verified with a focus on precipitation, as described next.

4. Precipitation forecast verification

a. Methods

As this study builds upon S21, overall methods for precipitation verification closely followed S21, which can be consulted for additional details. Specifically, forecasts of 1-h accumulated precipitation were objectively verified against NCEP's ~4.8-km Stage IV (ST4) analyses (Lin and Mitchell 2005) over the CONUS east of 105°W (Fig. 1), where ST4 analyses are most reliable (e.g., Nelson et al. 2016). Precipitation forecasts were interpolated to the ST4 grid with a budget algorithm that conserves total precipitation (Accadia et al. 2003) for comparison to ST4 analyses. As in S21, we focus on precipitation for brevity and because numerical precipitation forecasts depend on many physical processes, providing a good summary of model performance. Moreover, precipitation is commonly used to verify CAEs.

Following S21, we used percentile thresholds (e.g., the 95th percentile), rather than absolute thresholds (e.g., 1.0 mm h^{-1}), to define events. Using percentile thresholds removes bias, permitting a robust assessment of spatial placement within the context of a model's climate (e.g., Roberts and Lean 2008; Mittermaier and Roberts 2010; Dey et al. 2014; Gowan et al. 2018; Cafaro et al. 2021). We used the 95.0th, 97.5th, 99.0th, 99.5th, 99.75th, and 99.9th percentile thresholds to examine a range of precipitation intensities. To facilitate interpretation of subsequent verification statistics, Table 5 shows the average physical thresholds corresponding to the percentile thresholds over the first 12 h. Average physical thresholds for the CAEs were broadly similar to ST4 physical thresholds, although there was a tendency for overprediction at the 99.5–99.9th percentiles and underprediction at the 95.0th and 97.5th percentiles, with GFSCent_GEFSPert physical thresholds lowest due to its slower spinup.

Additionally, rather than verifying point-based probabilities, we used a neighborhood approach to derive and verify "neighborhood ensemble probabilities" (NEPs; Schwartz et al. 2010; Schwartz and Sobash 2017), which were computed by spatially averaging point-based probabilities within circular neighborhoods surrounding each grid point. NEPs are more appropriate for verifying CAE forecasts than point-based probabilities because they recognize that high-resolution NWP models are inaccurate at the grid scale (e.g., Theis et al. 2005; Ebert 2008, 2009). We constructed NEPs with neighborhood length scales (r) between 5 and 150 km.

Statistical significance was assessed with a bootstrap resampling approach (with replacement) applied to pairwise differences between two experiments (e.g., Hamill 1999; Wolff et al. 2014) using 10 000 resamples, which were randomly drawn from daily statistics as in S21. Significance levels were determined as the percentile where the distribution of resampled differences equaled zero (e.g., Davis et al. 2010; Schwartz 2016; Lu et al. 2017; Zhang 2021), with significance levels ≥ 90% regarded as statistically significant.

TABLE 5. Average physical thresholds (mm h^{-1}) corresponding to the 95.0th, 97.5th, 99.0th, 99.5th, 99.75th, and 99.9th percentile thresholds over all 26 1–12-h, 3-km forecasts of 1-h accumulated precipitation, computed over the CONUS east of 105°W. For the ensembles, percentiles were calculated separately for each member. Values in the table are averages over all 10 members.

		Percentile threshold					
		95.0	97.5	99.0	99.5	99.75	99.9
Expt	3kmCent_3kmPert	0.9	1.9	4.1	6.6	9.7	14.2
	3kmCent_15kmPert	0.9	1.9	4.2	6.6	9.7	14.4
	3kmCent_GEFSPert	0.9	1.9	4.2	6.7	9.9	14.5
	15kmCent_3kmPert	1.0	2.1	4.5	7.1	10.2	14.7
	15kmCent_15kmPert	1.0	2.0	4.5	7.1	10.3	15.0
	15kmCent_GEFSPert	1.0	2.1	4.5	7.2	10.4	15.2
	GFSCent_3kmPert	1.0	2.0	4.3	6.7	9.7	14.2
	GFSCent_15kmPert	1.0	2.0	4.2	6.7	9.8	14.4
	GFSCent_GEFSPert	0.8	1.7	3.9	6.3	9.3	13.8
	Stage IV analyses	1.1	2.2	4.2	6.3	8.9	13.0

b. Results

Probabilistic precipitation forecasts from the nine CAEs were evaluated with fractions skill scores [FSSs; Roberts and Lean (2008)], attributes diagrams (e.g., Wilks 2011), and areas under the relative operating characteristic (ROC) curve (Mason 1982; Mason and Graham 2002). Both FSSs and ROC areas range between 0 and 1, with higher values indicating more skill, while perfect reliability is achieved for curves lying on the diagonal of attributes diagrams. We also evaluated the Brier score (Brier 1950), which provided identical conclusions as FSSs and ROC areas and is not further discussed. However, the "reliability component" of the Brier score decomposition (BS_{REL}; Murphy 1973) was useful for elucidating aspects of reliability that were sometimes challenging to visually glean from attributes diagrams, so values of BS_{REL} are presented. Smaller values of BS_{REL} are better.

These metrics were computed for NEPs constructed with r = 5, 25, 50, 75, 100, 125, and 150 km, but varying r did not change overall conclusions, as shown in appendix B. Thus, we focus on statistics computed with r = 100 km, which was approximately the smallest r where, over the first 12 h, near-perfect reliability was obtained for some thresholds and FSSs achieved "useful" skill [per Roberts and Lean (2008)] at all thresholds for the best performing CAEs. To provide a holistic assessment, we present statistics aggregated over all 26 forecasts.

1) CAE FORECAST SENSITIVITY TO CENTRAL INITIAL STATES

Given identical IC perturbations, CAEs with 3-km central initial states typically had higher aggregate FSSs over the first ~12–21 h than CAEs with 15-km central initial states, which in turn had higher FSSs than CAEs with GFS central initial

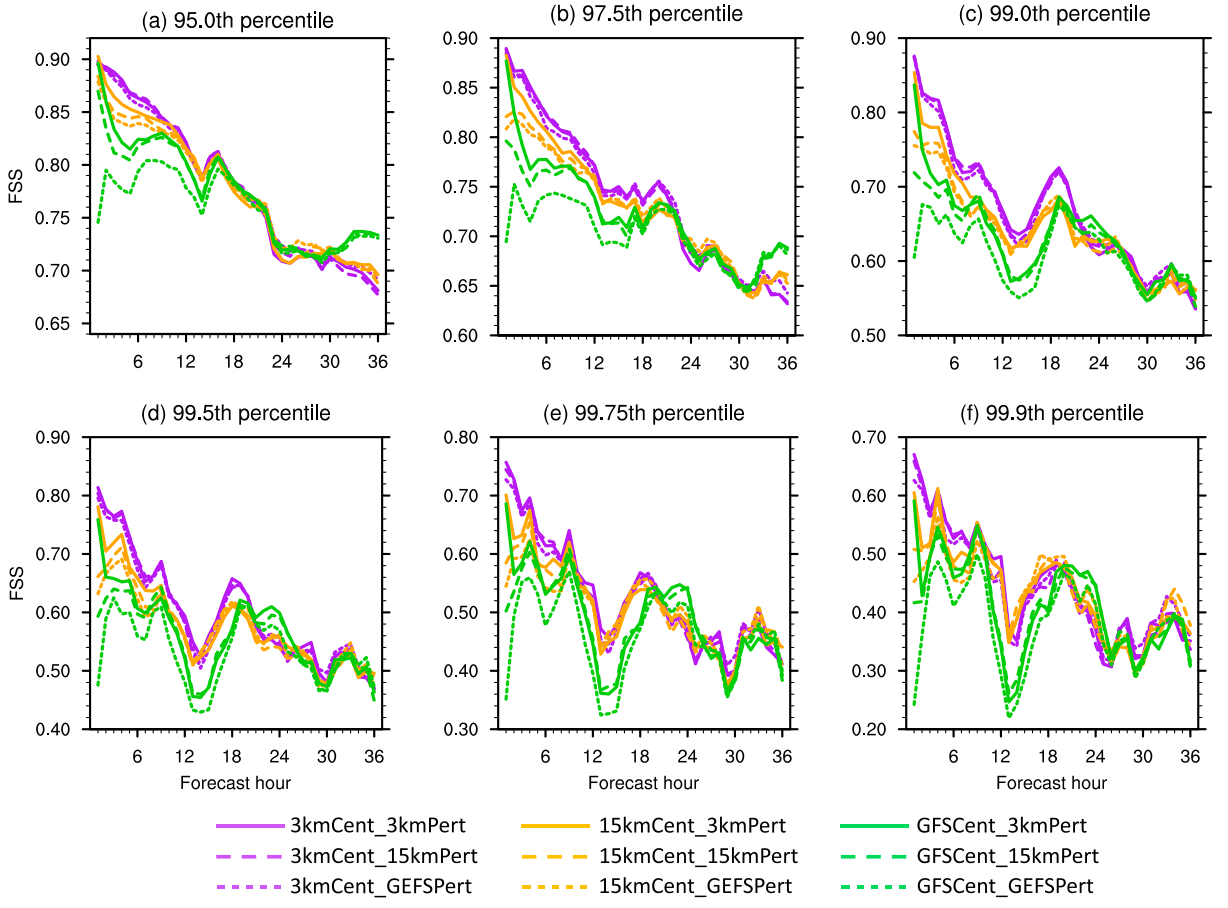

FIG. 6. Fractions skill scores (FSSs) over the CONUS east of 105°W (Fig. 1) with a 100-km neighborhood length scale for the (a) 95th, (b) 97.5th, (c) 99th, (d) 99.5th, (e) 99.75th, and (f) 99.9th percentile thresholds aggregated over all 26 3-km forecasts of 1-h accumulated precipitation as a function 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.

states (cf. purple, orange, and green curves with common line patterns in Fig. 6). These relationships held at all thresholds and differences were regularly statistically significant at the 90% confidence level, particularly those between CAEs with 15- and 3-km central initial states and CAEs with 3-km and GFS central initial states (Fig. 7). Aggregate ROC areas over the first 12 h yielded identical conclusions as FSSs (Fig. 8), and statistical significance between pairs of experiments for ROC areas echoed patterns in Fig. 7 (not shown). Moreover, for constant IC perturbations, CAEs with 3-km central initial states usually had better 1–12-h forecast reliabilities than CAEs with 15-km central initial states, and CAEs with GFS central initial states typically had the poorest reliabilities (Fig. 9, Table 6). All CAEs had comparable sharpness and were typically skillful compared to climatological forecasts (Fig. 9). The consistency of FSSs, ROC areas, and reliabilities through 12 h strongly suggests that short-term CAE precipitation forecasts are improved by using convection-allowing analyses as central initial states.

Conversely, after ~18–21 h, differences between CAEs were generally smaller, with the largest differences occurring after ~30 h at the 95th and 97.5th percentile thresholds, where CAEs with GFS central initial states exhibited a local increase in FSSs and were more skillful than CAEs with 15- and 3-km central initial states given fixed IC perturbations (Figs. 6a,b). Additionally, at higher thresholds, GFS central initial states typically led to the best ~22–26-h forecasts (Figs. 6c–f). ROC areas and attributes statistics after ~18 h yielded similar results as FSSs (not shown). As S20 found that CAE forecasts were sensitive to ICs through 48 h over an identical computational domain, we do not believe LBC influences were responsible for relatively smaller differences between the experiments after ~18–21 h.

Overall, benefits of convection-allowing central initial states were mostly confined to the first ~12 h, consistent with S21, who found that GEFS-initialized CAEs had better ~18–36-h precipitation forecasts than CAEs with 15- and 3-km EnKF ICs due to improved large-scale representation in GEFS ICs compared to limited-area EnKF analyses. We thus presume that large scales, which are more critical for next-day (i.e., ~18–36-h) forecasts than shorter-term forecasts, were better represented in GFS

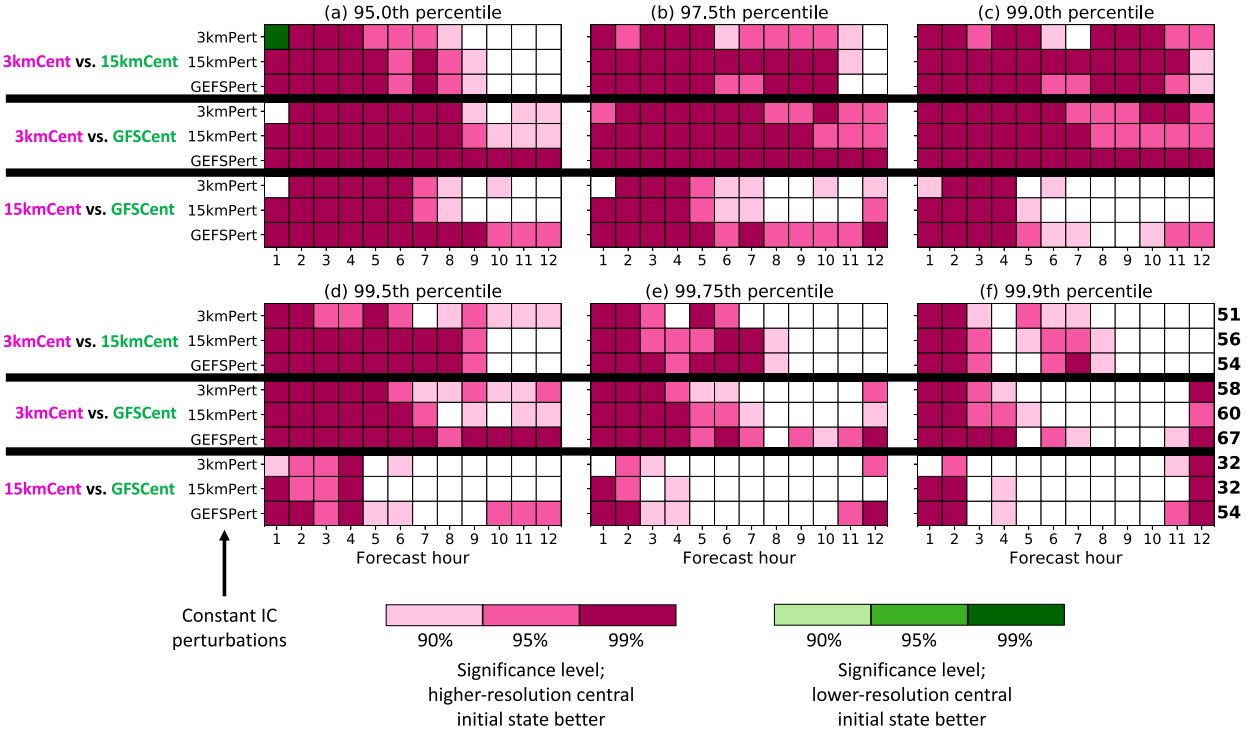

FIG. 7. Statistical significance levels of aggregate FSS differences (e.g., Fig. 6) between various experiments for the (a) 95th, (b) 97.5th, (c) 99th, (d) 99.5th, (e) 99.75th, and (f) 99.9th percentile thresholds for forecast hours 1–12 as determined through bootstrap resampling (section 4a). These comparisons assess the impact of changing central initial state resolution. Specifically, in each panel, a given row represents a fixed set of IC perturbations; "3kmPert," "15kmPert," and "GEFSPert" refer to IC perturbations provided by 3-km EnKF analysis ensembles, 15-km EnKF analysis ensembles, and GEFS ICs, respectively. Each panel is broken into thirds to represent different comparisons, and pink text denotes the experiment in each comparison with the higher-resolution central initial state. The top third compares experiments with 15- and 3-km central initial states ("3kmCent vs 15kmCent"). The middle third compares experiments with 3-km and 0.25° GFS central initial states ("3kmCent vs GFSCent"). The bottom third compares experiments with 15-km and 0.25° GFS central initial states ("15kmCent vs GFSCent"). Pink shadings indicate that higher-resolution central initial states led to statistically significantly higher FSSs for the fixed IC perturbations, while green shadings indicate that lower-resolution central initial states led to statistically significantly higher FSSs for the fixed IC perturbations. White cells indicate that aggregate FSSs of two experiments with varied central initial states but common IC perturbations were not statistically significantly different at the 90% confidence level or higher. Annotations to the right of (f) represent the number of occurrences where, for the given row, experiments with higher-resolution central initial states had statistically significantly higher FSSs than experiments with lower-resolution central initial states across all six percentile thresholds and all forecast hours (i.e., the total number of pink-shaded boxes in each row across all six panels).

analyses than in limited-area EnKF ICs. Because these general ~18–36-h forecast behaviors were exhaustively discussed by S21, we henceforth focus on ~1–12-h forecasts, where differences among the CAEs were usually the largest. However, some thoughts about how our findings enhance S21's conclusions about ~18–36-h forecast quality are provided in section 5d.

2) CAE FORECAST SENSITIVITY TO IC PERTURBATIONS

CAE precipitation forecast sensitivity to IC perturbations somewhat depended on the central initial state. For example, given identical 3-km central initial states, 3-km IC perturbations rarely yielded better CAE forecasts than 15-km IC perturbations over the first 12 h (cf. solid and long-dashed purple curves in Figs. 8–10; also see Table 6 and the top row of each panel in Fig. 11). But, when the CAEs had common GFS or 15-km central initial states, there were more instances where 3-km IC perturbations led to better forecasts than 15-km IC perturbations (Figs. 8–10; also see Table 6 and the second and third rows from the top of each panel in Fig. 11). Overall, any benefits of decreasing IC perturbation horizontal grid spacing from 15 to 3 km were reserved for situations where central initial states had convection-parameterizing horizontal grid spacing. There were no consistent benefits of 3-km IC perturbations relative to 15-km IC perturbations when CAEs had demonstrably preferable 3-km central initial states.

However, 1–12-h forecasts with 3-km IC perturbations were usually better than those with GEFS IC perturbations when holding central initial states constant (e.g., compare solid and short-dashed curves with common colors in Figs. 8 and 10; also note the middle third of each panel in Fig. 11 has more statistically significant differences than the top third). CAEs with

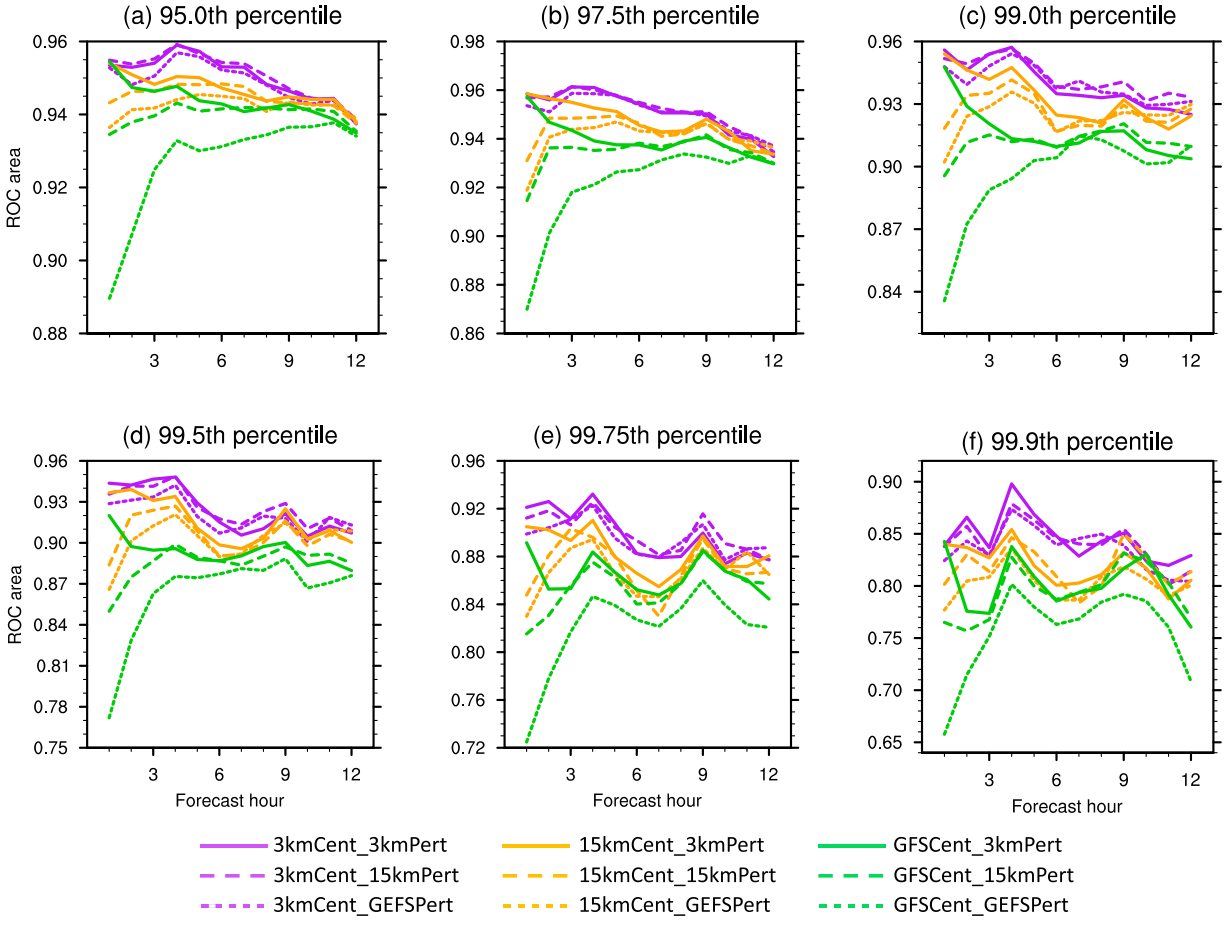

FIG. 8. As in Fig. 6, but for areas under the relative operating characteristic (ROC) curve computed using decision thresholds of 1%, 2%, 3%, 4%, 5%, 10%, 15%, … , 95%, and 100% and a trapezoidal method. Only forecast hours 1–12 are presented to zoom-in on the period with the largest systematic differences between experiments and to more easily visualize differences between CAEs with identical central initial states but different IC perturbations (which have common line colors but different line patterns). The *y*-axis scales are different in each panel.

15-km IC perturbations also typically outperformed CAEs with GEFS IC perturbations given common central initial states (Figs. 8 and 10; also see the bottom third of each panel in Fig. 11). The greatest benefit of 15- and 3-km EnKF IC perturbations compared to GEFS IC perturbations occurred when GFS analyses provided central initial states (e.g., see the sixth row from the top and bottom row of each panel in Fig. 11), illustrating that combining EnKF-based IC perturbations with GFS analyses is preferable to employing purely global ICs. Moreover, differences between CAEs with 15- and 3-km IC perturbations were generally smaller than differences between CAEs with EnKF- and GEFS-based IC perturbations, consistent with GEFS IC perturbations possessing very different spectral characteristics (Fig. 4) and spread (Fig. 5) than the two sets of EnKF IC perturbations, which resembled each other in many ways (Figs. 4 and 5).

Short-term precipitation forecast sensitivity to IC perturbations appeared to have some association with forecast evolution of small-scale perturbations, which exert greater control on short-term forecasts than large-scale perturbations. For instance, given 3-km central initial states, the CAE with 15-km IC perturbations quickly spunup fine-scale structures and had nearly identical perturbation spectra to the CAE with 3-km IC perturbations at scales < 100 km by 3 h (Figs. 12a,b). This fast spinup of small-scale structures from 15-km IC perturbations may be related to why CAEs with 15- and 3-km IC perturbations had similar short-term precipitation forecast skill given common 3-km central initial states (e.g., Figs. 8–11).

Similarly, the CAE with GEFS IC perturbations and 3-km central initial states also quickly spunup perturbations at scales < 100 km over the first 3 h. However, some differences between CAEs with GEFS and limited-area EnKF IC perturbations remained through 6 h (Figs. 12a–c), consistent with most statistically significant differences regarding precipitation forecast skill between CAEs with EnKF and GEFS IC perturbations occurring before 6 h (see the fourth and seventh rows from the top of each panel in Fig. 11). By 12 h, all three

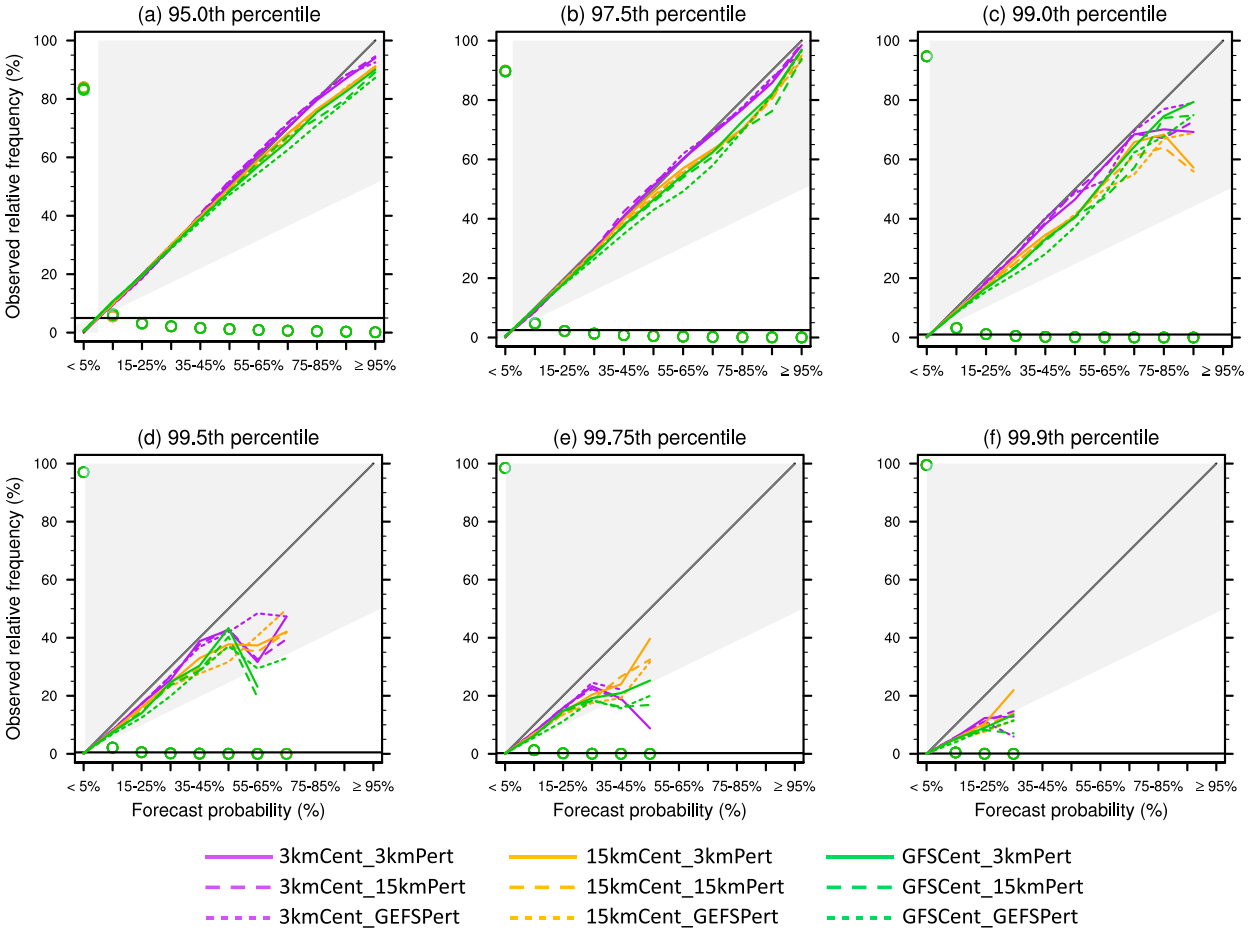

FIG. 9. Attributes diagrams computed over the CONUS east of 105°W (Fig. 1) with a 100-km neighborhood length scale aggregated over all 26 1–12-h 3-km forecasts of 1-h accumulated precipitation for the (a) 95th, (b) 97.5th, (c) 99th, (d) 99.5th, (e) 99.75th, and (f) 99.9th percentile thresholds. Horizontal lines near the *x* axis represent observed frequencies of the event, diagonal lines are lines of perfect reliability, and forecast frequencies (%) within each probability bin are shown as open circles (all nine CAEs had very similar probability distributions, so the circles lie atop each other). Points lying in gray-shaded regions had skill compared to climatological forecasts 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 CONUS east of 105°W and all 26 forecasts.

CAEs with 3-km central initial states had similar perturbation spectra at scales < 800 km (Fig. 12d), consistent with little precipitation forecast sensitivity to IC perturbations after 12 h (e.g., see purple curves in Fig. 6). Similar spectral evolutions occurred over the first 12 h for other meteorological variables (not shown).

These results echo previous studies indicating that small-scale (e.g., <100 km) perturbations quickly develop from relatively coarse ensemble ICs once high-resolution model integration commences (e.g., Harnisch and Keil 2015; Tennant 2015; Johnson and Wang 2016; Raynaud and Bouttier 2016; Potvin et al. 2017; JW20). These findings are also commensurate with hypotheses that large-scale perturbations are important drivers of error growth (e.g., Durran and Gingrich 2014), as the absence of small-scale perturbations in GEFS ICs (Fig. 4) did not seem to fundamentally limit error growth in CAEs with GEFS IC perturbations (Fig. 12).

3) OVERALL SENSITIVITIES OF SHORT-TERM CAE FORECASTS

Considering all nine CAEs, those with 3-km central initial states usually had better 1–12-h precipitation forecasts than CAEs with 15-km central initial states, which in turn were typically better than CAEs with GFS central initial states, *regardless of IC perturbations* (e.g., Figs. 8–10; Table 6; note that all purple curves are usually above all orange curves, which are usually above all green curves in Figs. 8 and 10). The only systematic exception occurred at forecast hour 1, where the five CAEs with at least one 3-km IC component had the five best forecasts, suggesting that information content at convection-allowing scales provided by either central initial states or IC perturbations is helpful for the shortest forecasts. However, skill in the CAEs with 3-km IC perturbations and GFS or 15-km EnKF central initial states (i.e., 15kmCent_3kmPert and GFSCent_3kmPert) diminished between hours 1–2, sometimes rapidly (Figs. 8 and 10).

TABLE 6. Reliability component of the Brier score decomposition (Murphy 1973; smaller is better) aggregated over all 26 1–12-h, 3-km forecasts of 1-h accumulated precipitation for various percentile thresholds, computed over the CONUS east of 105°W (Fig. 1) with a 100-km neighborhood length scale. These values correspond to the curves in Fig. 9.

		Percentile threshold					
		95.0	97.5	99.0	99.5	99.75	99.9
Expt	3kmCent_3kmPert	0.000 342	0.000 434	0.000 507	0.000 551	0.000 577	0.000 596
	3kmCent_15kmPert	0.000 359	0.000 442	0.000 505	0.000 548	0.000 575	0.000 595
	3kmCent_GEFSPert	0.000 351	0.000 433	0.000 504	0.000 543	0.000 573	0.000 596
	15kmCent_3kmPert	0.000 347	0.000 445	0.000 521	0.000 553	0.000 577	0.000 598
	15kmCent_15kmPert	0.000 345	0.000 448	0.000 526	0.000 555	0.000 577	0.000 597
	15kmCent_GEFSPert	0.000 348	0.000 446	0.000 527	0.000 561	0.000 582	0.000 600
	GFSCent_3kmPert	0.000 363	0.000 434	0.000 524	0.000 558	0.000 578	0.000 597
	GFSCent_15kmPert	0.000 364	0.000 450	0.000 529	0.000 562	0.000 581	0.000 597
	GFSCent_GEFSPert	0.000 415	0.000 496	0.000 565	0.000 582	0.000 591	0.000 602

Forecast skill characteristics over the first 2 h in 15kmCent_3kmPert and GFSCent_3kmPert appear related to precipitation spinup. For instance, given GFS and 15-km central initial states, 3-km IC perturbations led to the most domain-total precipitation at forecast hour 1 (Fig. 13). Adding 3-km perturbations to relatively coarser GFS and 15-km fields led to imbalances and gravity wave generation; the small-scale perturbations likely acted as noise that stimulated precipitation development. Although the greater domain-total precipitation resulting from 3-km IC perturbations did not always agree well with observed domain-total precipitation (e.g., see the solid orange line in Fig. 13), these enhanced precipitation elements were

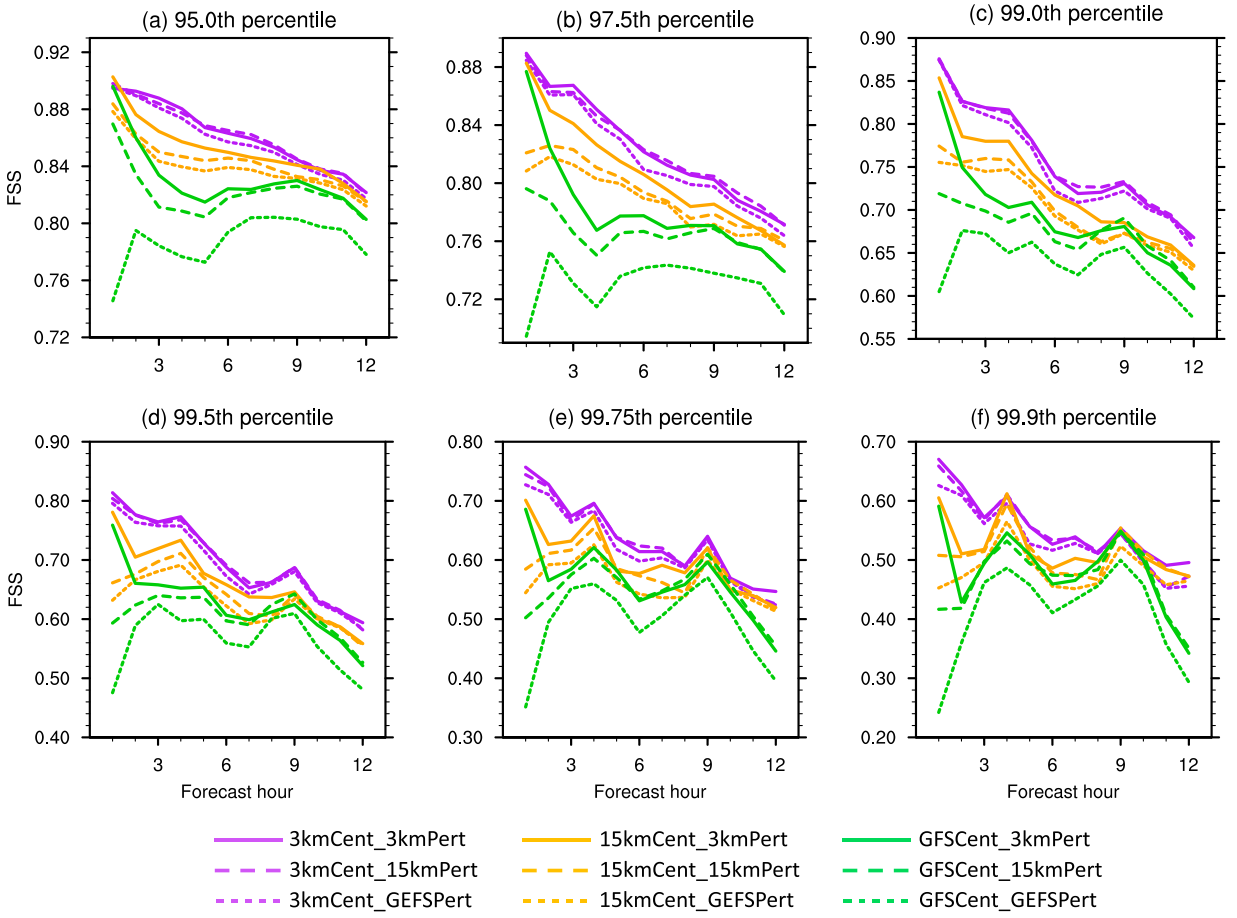

FIG. 10. As in Fig. 6, but zoomed-in over the first 12 forecast hours to more easily visualize differences between CAEs with identical central initial states but different IC perturbations (which have common line colors but different line patterns).

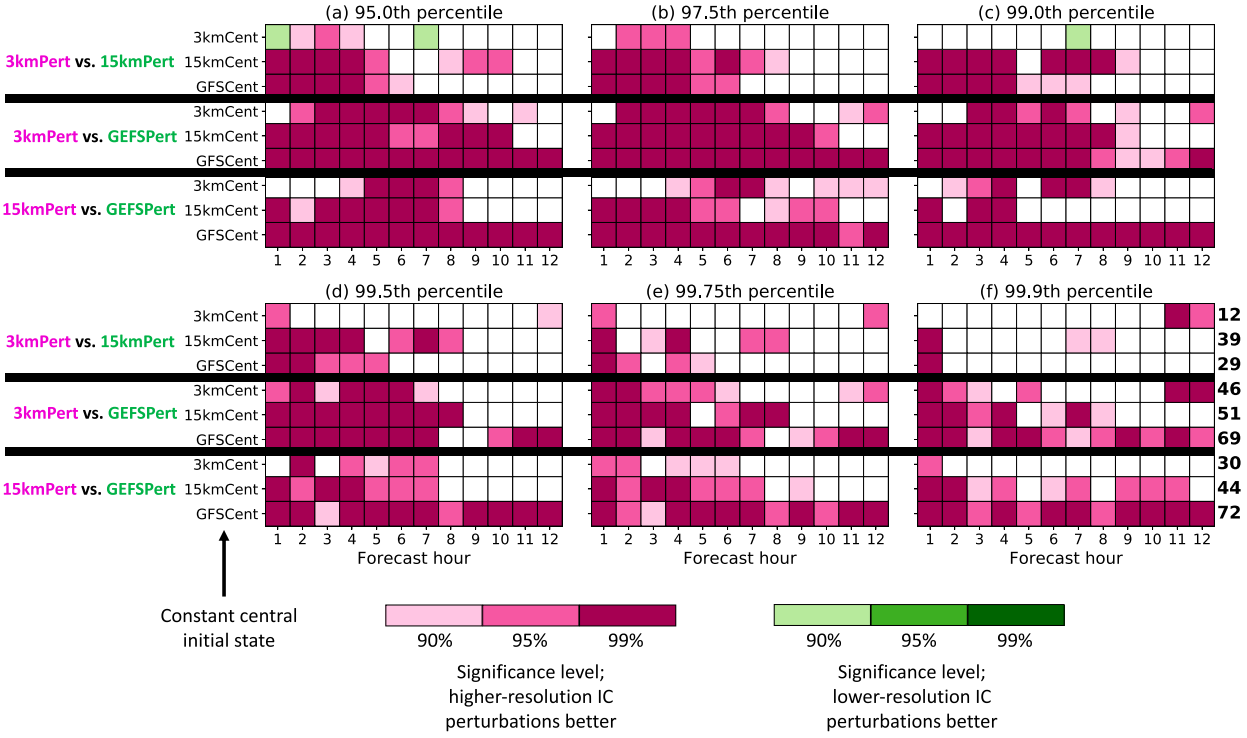

FIG. 11. As in Fig. 7, but for different comparisons that assess the impact of changing IC perturbation resolution. Specifically, in each panel, a given row represents a fixed set of central initial states; "3kmCent," "15kmCent," and "GFSCent" refer to central initial states provided by 3-km EnKF ensemble mean analyses, 15-km EnKF ensemble mean analyses, and GFS analyses, respectively. Each panel is broken into thirds to represent different comparisons, and pink text denotes the experiment in each comparison with the higher-resolution IC perturbations. The top third compares experiments with 15- and 3-km IC perturbations ("3kmPert vs 15kmPert"). The middle third compares experiments with 3-km and 0.5° GEFS IC perturbations ("3kmPert vs GEFSPert"). The bottom third compares experiments with 15-km and 0.5° GEFS IC perturbations ("15kmPert vs GEFSPert"). Pink shadings indicate that higher-resolution IC perturbations led to statistically significantly higher FSSs for the fixed central initial state, while green shadings indicate that lower-resolution IC perturbations led to statistically significantly higher FSSs for the fixed central initial state. White cells indicate that aggregate FSSs of two experiments with varied IC perturbations but common central initial states were not statistically significantly different at the 90% confidence level or higher. Annotations to the right of (f) represent the number of occurrences where, for the given row, experiments with higher-resolution IC perturbations had statistically significantly higher FSSs than experiments with lower-resolution IC perturbations across all six percentile thresholds and all forecast hours (i.e., the total number of pink-shaded boxes in each row across all six panels).

often placed correctly, given that ROC areas and FSSs were greatly improved at forecast hour 1 by adding 3-km IC perturbations, rather than GEFS or 15-km IC perturbations, to GFS and 15-km central initial states (Figs. 8 and 10). Between 1 and 2 h, domain-total precipitation in 15kmCent_3kmPert and GFSCent_3kmPert decreased as imbalances were resolved (Fig. 13), which is consistent with the sometimes precipitous decline of skill in these two experiments over this period (Figs. 8 and 10) and epitomizes the dominant influence of central initial states for short-term forecast evolution. Overall, after 3–6 h, domain-total precipitation provided similar conclusions as other metrics: for a fixed central initial state GEFS IC perturbations produced forecasts that were typically furthest from observations, 15- and 3-km IC perturbations added to 3-km central initial states yielded comparable performance, and CAEs with 3-km central initial states were usually closest to observations regardless of IC perturbations.

Furthermore, differences between CAEs with identical IC perturbations but different central initial states were statistically significant more often than differences between CAEs with identical central initial states but varied IC perturbations (cf. Figs. 7 and 11). Most importantly, the four EnKF-only experiments (bold experiments in Table 3) clearly revealed that CAEs with 3-km central initial states were statistically significantly better than CAEs with 15-km central initial states (given constant IC perturbations) more often than CAEs with 3-km IC perturbations were statistically significantly better than CAEs with 15-km IC perturbations (given constant central initial states; compare the top two rows of each panel in Figs. 7 and 11).

Therefore, collective findings strongly suggest it is more important that central initial states possess convection-allowing horizontal grid spacing than IC perturbations for short-term CAE precipitation forecasts. These results imply that small-scale structures in central initial states help to define a more accurate envelope within which ensemble members' short-term forecasts evolve.

A consequence of these findings is that deterministic forecasts initialized from central initial states (e.g., EnKF mean

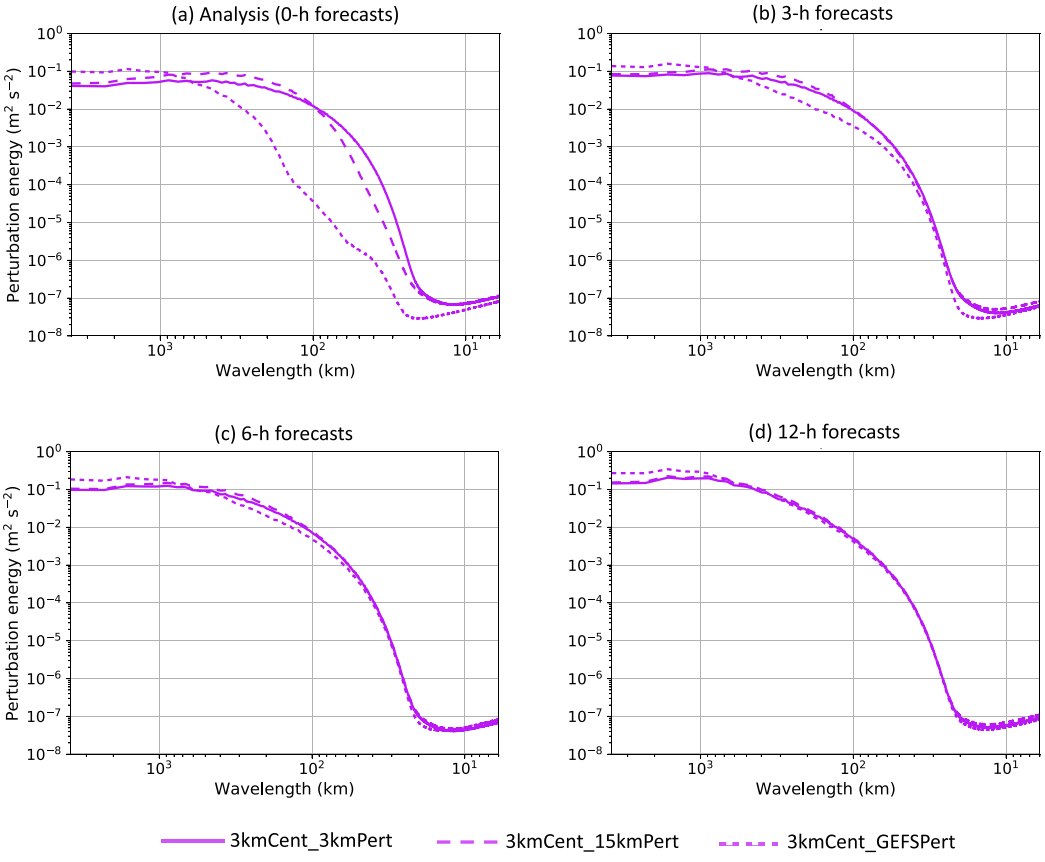

FIG. 12. Average 250-hPa zonal wind perturbation energy (m^2 s^{-2}) over all 26 3-km forecasts and all 10 ensemble perturbations as a function of wavelength (km) for the three CAEs with 3-km central initial states for (a) analyses (0-h forecasts) and (b) 3-, (c) 6-, and (d) 12-h forecasts. Perturbations were defined with respect to the ensemble mean, and spectra were computed over the entire 3-km domain (Fig. 1), excluding points within 45 km of each lateral boundary. The discrete cosine transform (Denis et al. 2002) was used to compute spectra and spectral variance binning employed the method of Ricard et al. (2013).

analyses) can potentially be used as proxies for CAE forecasts at 1/M the cost of an M-member CAE. As evidence of this possibility, comparison of deterministic forecasts initialized from GFS, 3-km EnKF mean, and 15-km EnKF mean analyses over the nested domain (Fig. 1) yielded identical conclusions as comparisons of 10-member CAEs solely differing by their central initial states: over the first ~12–21 h, 3-km EnKF mean analyses yielded the best forecasts and GFS analyses the worst, whereas forecasts initialized from GFS analyses were comparable to or better than those initialized from limited-area EnKF analyses after ~18–24 h (Fig. 14). Therefore, CAE developers may only need to initialize deterministic forecasts from central initial states during portions of experimentation, potentially saving resources and enabling trials over longer time periods.

5. Discussion

a. *Connection to dual-resolution ensemble–variational DA systems*

Our results are broadly consistent with, but not directly comparable to, previous work that examined deterministic forecasts initialized from ensemble–variational (EnVar) DA systems (e.g., Hamill and Snyder 2000; Lorenc 2003; Wang et al. 2008; Wang 2010) with "dual-resolution" configurations, where a comparatively low-resolution ensemble provides background error covariances (BECs) for a relatively higher-resolution deterministic background.[8] Specifically, several studies noted that deterministic forecasts were improved when increasing resolution of the deterministic background while holding ensemble perturbation resolution constant (e.g., Schwartz 2016; Lu et al. 2017; Pan et al. 2018; Wang et al. 2019). Conversely, studies isolating sensitivity to perturbation (i.e., BEC) resolution in dual-resolution EnVar DA systems

[8] Incremental four-dimensional variational (4DVAR) DA systems (Courtier et al. 1994) can also employ dual-resolution concepts by obtaining innovations from a high-resolution nonlinear model while inner loop minimizations use low-resolution formulations of tangent linear and adjoint models (e.g., Liu et al. 2020). Our discussion does not concern pure 4DVAR systems, but rather, solely EnVar DA systems where "dual resolution" refers to disparities in resolution among ensemble-based BECs and a deterministic background.

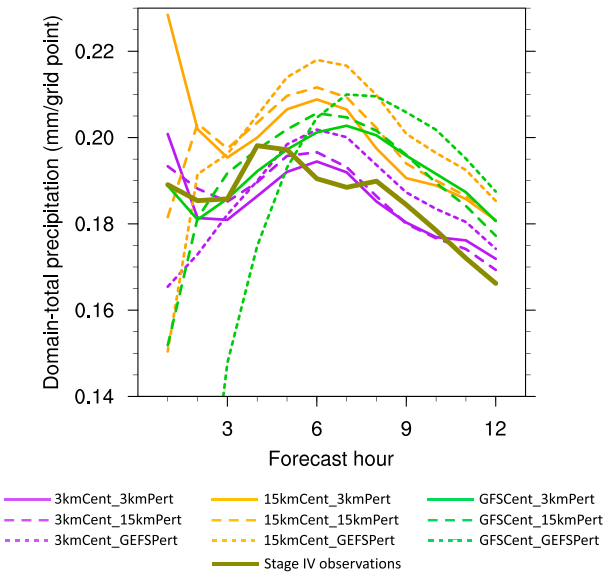

FIG. 13. Average 1-h accumulated precipitation (mm) per grid point over all 26 3-km forecasts and the CONUS east of 105°W (Fig. 1), computed on native grids as a function of forecast hour. These statistics were computed for all 10 ensemble members, but for readability, only ensemble means are shown. Values on the *x* axis represent ending forecast hours of 1-h accumulation periods (e.g., an *x*-axis value of 9 is for 1-h accumulated precipitation between 8 and 9 h). At forecast hours 1 and 2, GFSCent_GEFSPert domain-total precipitation was nonzero but below the *x* axis.

have yielded mixed results and provide scant overall evidence that BECs coarser than the deterministic background systematically degrade subsequent forecasts (e.g., Schwartz et al. 2015b; Schwartz 2016; Lei and Whitaker 2017; Bédard et al. 2018, 2020; Kay and Wang 2020). Furthermore, some ensemble-based DA systems use simpler procedures to update ensemble perturbations relative to methods for updating deterministic backgrounds, finding few adverse impacts from the simplifications and implicitly acknowledging the overriding importance of central states (e.g., Buehner et al. 2017; Lorenc et al. 2017; Bédard et al. 2018). Therefore, past research collectively suggests that ensemble perturbation resolution likely has secondary importance relative to resolution of deterministic backgrounds in dual-resolution EnVar DA systems, consistent with our results for short-term CAE forecasts.

However, most previous studies assessing sensitivity of dual-resolution EnVar analyses to BEC resolution focused on DA systems at convection-parameterizing scales, and when it comes to understanding BEC resolution requirements for convective-scale DA applications, there is substantial uncertainty. Nonetheless, there are likely situations where convection-parameterizing BECs cannot provide relevant spatiotemporal details about small-scale features represented in convection-allowing EnVar backgrounds, potentially leading to suboptimal analyses. Thus, while it may be unnecessary to initialize CAE "free forecasts" with convection-allowing IC perturbations, it is conceivable that convection-allowing BECs are in fact critical to producing optimal convection-allowing analyses. Future studies should investigate this topic to better understand whether BEC resolution can be degraded without also degrading convective-scale EnVar analyses.

b. *Theoretical aspects*

Our findings are limited by the effectiveness of Gaussian-based DA methodologies, like the EnKF, to effectively represent posterior means and perturbations at various spatial resolutions. For example, in a highly idealized scenario, Posselt and Bishop (2012) suggested that given non-Gaussian priors (states before assimilation), EnKFs reasonably represent posterior means (i.e., central initial states) but poorly estimate posterior covariances (i.e., IC perturbations). This deficiency can directly limit forecast performance of convective-scale NWP systems that use EnKFs (Poterjoy et al. 2017, 2019). Assuming that 3-km EnKF priors were more non-Gaussian than 15-km EnKF priors, our findings are consistent with past research: benefits of decreasing central initial state horizontal grid spacing from 15 to 3 km suggest reliable EnKF updates for the mean at all scales, while lack of benefits from decreasing IC perturbation horizontal grid spacing from 15 to 3 km suggests the 3-km EnKF may not properly represent convective-scale posterior covariances. These concepts are consistent with the discussion in section 5a and may provide a theoretical basis for using dual-resolution DA systems to initialize CAEs.

c. *Similarities with JW20*

In general, our results corroborate JW20's findings that 3-km IC perturbations lead to better short-term CAE forecasts than 0.5° GEFS IC perturbations given 3-km central initial states, and, together with JW20's conclusions, suggest that approximately 12 h represents an upper bound on forecast ranges for which convection-allowing IC perturbations are beneficial. As there were meaningful differences between JW20 and our study (e.g., JW20 used a 40-member partial cycling EnVar/EnKF DA system while we used an 80-member continuously cycling EnKF), these collective findings concerning 3-km versus 0.5° GEFS IC perturbations appear robust.

Additional experiments in JW20 suggested these results were due to missing small-scale structures in GEFS IC perturbations, rather than the myriad other differences between GEFS IC perturbations and IC perturbations provided by limited-area 3-km WRF-based DA systems, like physical parameterizations. However, our findings that 3-km IC perturbations did not lead to better precipitation forecasts than 15-km IC perturbations given common 3-km central initial states suggest a point of diminishing returns for increasing IC perturbation resolution.

d. *Further insights into S21*

Comparison of the EnKF-only experiments (bold experiments in Table 3) indicates that S21's 3-km EnKF initialized better 6–12-h forecasts than their 15-km EnKF due to increased resolution of its central initial state, not because of finer-resolution IC perturbations. Furthermore, S21 noted that ICs produced by "blending" small scales from 3-km

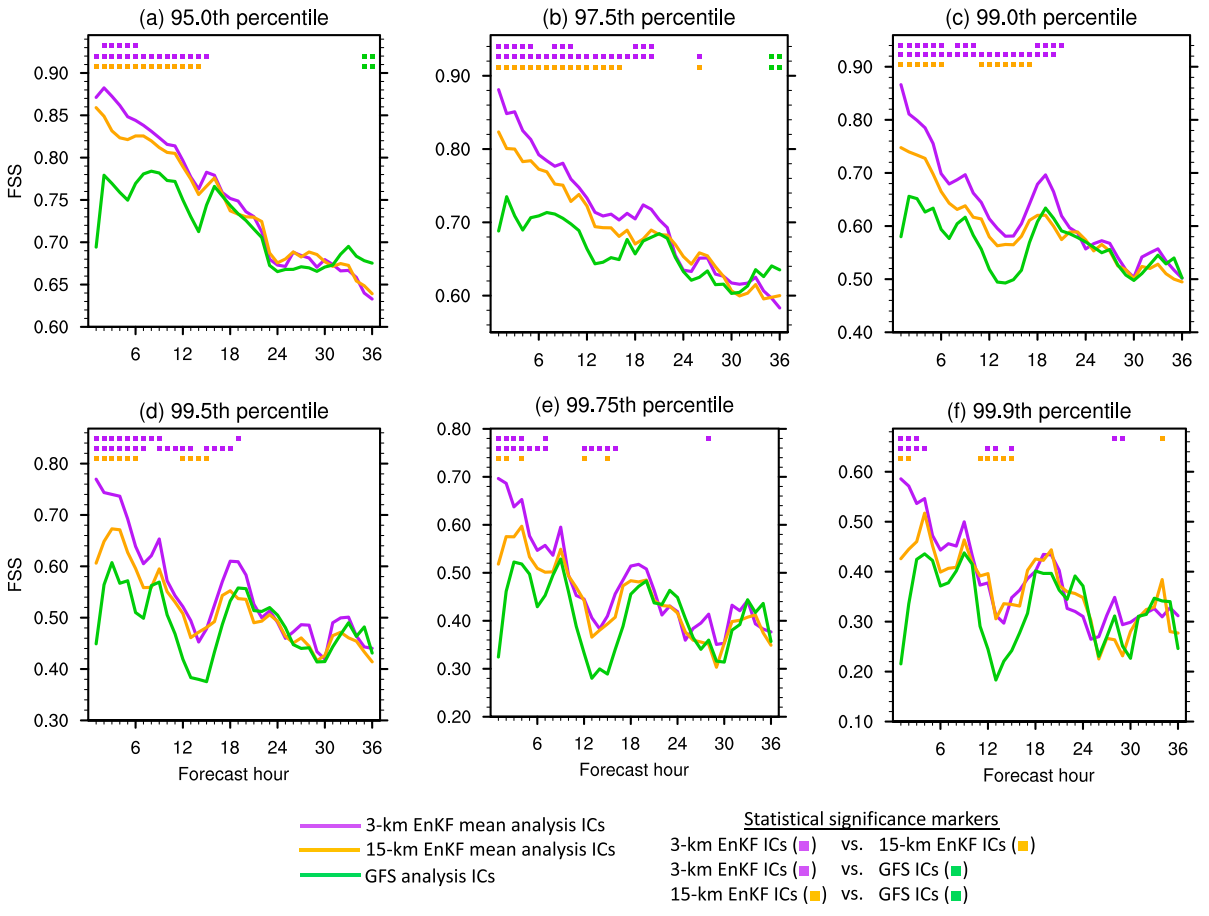

FIG. 14. Fractions skill scores (FSSs) over the CONUS east of 105°W (Fig. 1) with a 100-km neighborhood length scale aggregated over all 26 forecasts of 1-h accumulated precipitation for deterministic 3-km forecasts initialized from 3-km EnKF mean analyses (purple), 15-km EnKF mean analyses (orange), and GFS analyses (green) for the (a) 95th, (b) 97.5th, (c) 99th, (d) 99.5th, (e) 99.75th, and (f) 99.9th percentile thresholds 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 denote instances where differences between two forecasts were statistically significant at the 95% level as determined through bootstrap resampling (section 4a), with the three rows of colored symbols corresponding to the three comparisons in the legend to denote which forecast had statistically significantly higher FSSs. For example, in the middle row, purple symbols indicate the forecasts with ICs provided by 3-km EnKF mean analyses had statistically significantly higher FSSs than forecasts with ICs provided by GFS analyses, while green symbols indicate forecasts with ICs provided by GFS analyses had statistically significantly higher FSSs than forecasts with ICs provided by 3-km EnKF mean analyses. Absence of a symbol means the differences were not statistically significant at the 95% level. The *y*-axis scales are different in each panel.

EnKF analysis members with large scales from corresponding GEFS IC members led to better ~18–36-h CAE forecasts than ICs provided by unblended 3-km EnKF analysis ensembles (i.e., $x_i^{3\,\mathrm{km}}$). However, because individual members from 3-km EnKF and GEFS IC ensembles were blended (i.e., $x_i^{3\,\mathrm{km}}$ and x_i^{GEFS} were blended for the *i*th member to create new ICs), S21 could not assess whether changing the *large-scale central initial state* or the *large-scale IC perturbations* was responsible for the success of blending.

Although we did not perform experiments to explicitly examine the impact of modifying large-scale central initial states and IC perturbations, our experiments nonetheless provide some insights on S21's findings. Specifically, after ~30 h at the 95th and 97.5th percentile thresholds (Figs. 6a,b) and between ~22 and 26 h at higher thresholds (Figs. 6c–f), CAEs with GFS central initial states outperformed CAEs with EnKF central initial states regardless of IC perturbation resolution. Moreover, ~18–36-h forecast sensitivity to using global IC perturbations was relatively modest: given 3-km central initial states, while employing GEFS IC perturbations sometimes boosted 30–36-h FSSs compared to using 15- or 3-km IC perturbations, much bigger performance gains were realized by changing central initial states to GFS analyses (Figs. 6a,b).

Therefore, ~18–36-h forecasts were improved most by using central initial states provided by a global model. This finding suggests that S21's blended 3-km ICs yielded better next-day forecasts than unblended 3-km ICs because of

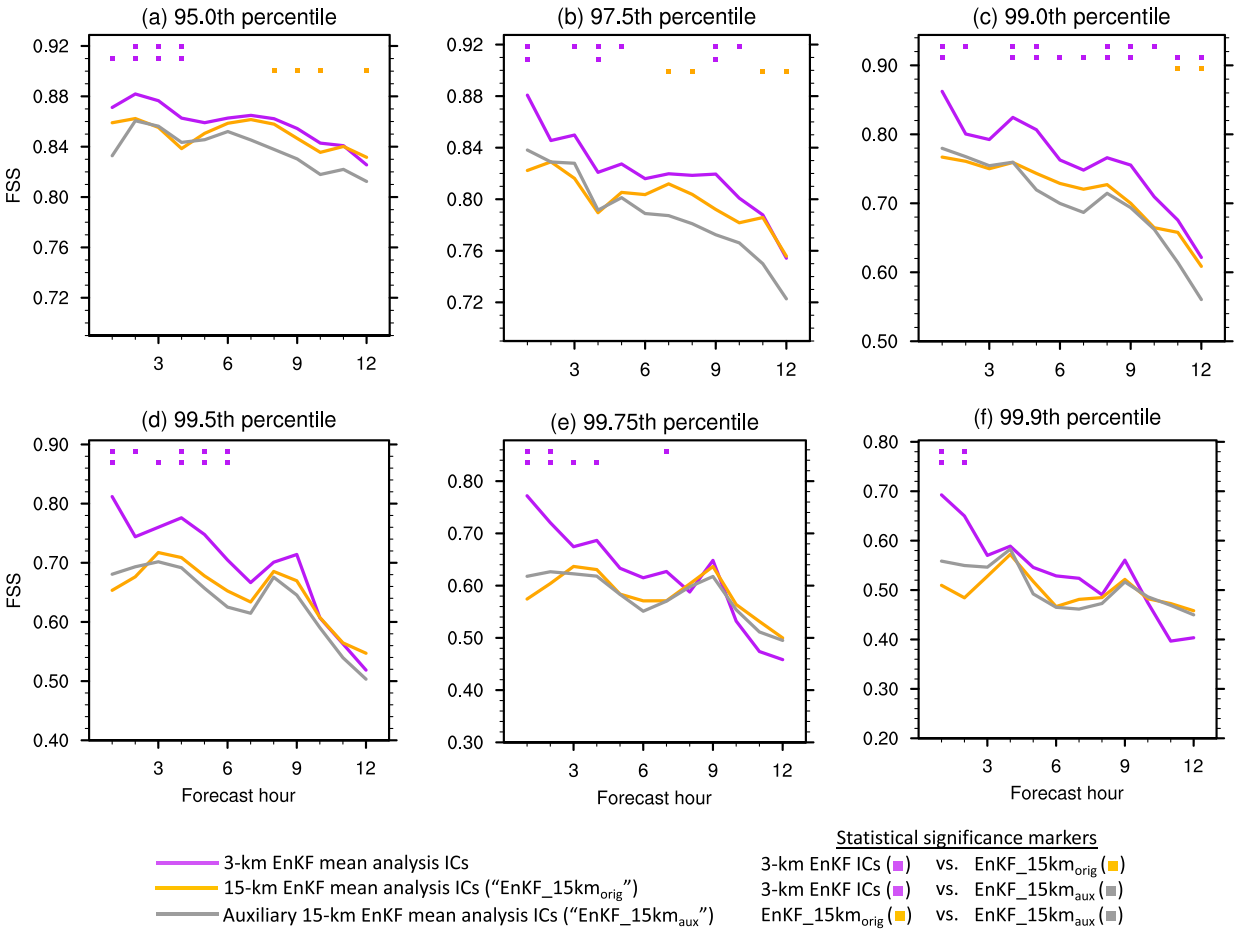

FIG. A1. As in Fig. 14, but for FSSs with a 100-km neighborhood length scale aggregated over deterministic 3-km forecasts of 1-h accumulated precipitation initialized from 0000 UTC EnKF mean analyses between 25 Apr and 7 May 2017 (inclusive), focusing on experiments designed to assess forecast sensitivity to horizontal localization and observation thinning in 15-km EnKFs.

forcing large-scale central initial states to global model large scales during blending, rather than forcing large-scale IC perturbations to those provided by a global model. It thus seems more critical to accurately depict large-scale central initial states than large-scale IC perturbations for next-day CAE forecasts.

6. Summary and conclusions

Nine sets of 36-h, 10-member CAE forecasts were produced over the CONUS for a 4-week period spanning April–May 2017. The various CAEs differed solely with regard to their central initial states and IC perturbations (Tables 3 and 4) and were verified with a focus on precipitation east of the Rockies.

When holding IC perturbations constant, CAE precipitation forecasts over the first ~12 h were best when central initial states were provided by 3-km EnKF mean analyses, rather than GFS or 15-km EnKF mean analyses. Thus, short-term CAE forecasts clearly benefited when central initial states possessed convection-allowing horizontal grid spacing. However, when holding these optimal 3-km central initial states constant and varying IC perturbations, there were no systematic benefits of decreasing IC perturbation horizontal grid spacing from 15 to 3 km, although 3-km IC perturbations typically led to better short-term CAE forecasts than GEFS IC perturbations.

Overall, considering all nine CAEs, in aggregate, the three with 3-km central initial states produced better short-term precipitation forecasts than CAEs with GFS or 15-km central initial states, *regardless of IC perturbations*. Therefore, for short-term CAE forecasting applications, while increasing IC perturbation resolution to convection-allowing scales can be helpful in some instances, it is far more important for central initial states than for IC perturbations to possess convection-allowing horizontal grid spacing. Consistent with these results, comparing deterministic forecasts initialized from various central initial states can provide identical conclusions as comparing CAEs solely differing by their central initial states, potentially saving resources (e.g., Fig. 14). Of course, our findings must be interpreted within the context of this study, which focused primarily on strongly forced events and used

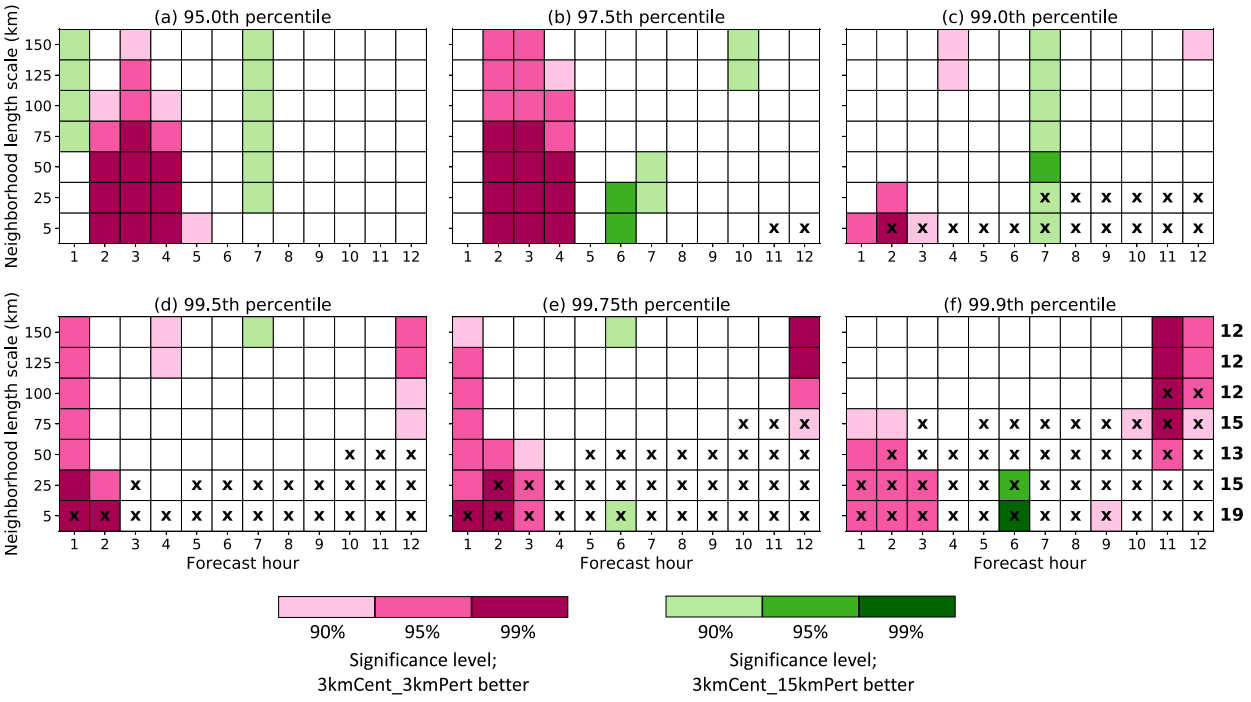

FIG. B1. Statistical significance levels of aggregate FSS differences between 3kmCent_3kmPert and 3kmCent_15kmPert for the (a) 95th, (b) 97.5th, (c) 99th, (d) 99.5th, (e) 99.75th, and (f) 99.9th percentile thresholds for forecast hours 1–12 for different neighborhood length scales (km) as determined through bootstrap resampling (section 4a). These comparisons assess the impact of changing IC perturbation resolution. Pink shadings indicate that 3kmCent_3kmPert had statistically significantly higher FSSs then 3kmCent_15kmPert, while green shadings indicate that 3kmCent_15kmPert had statistically significantly higher FSSs than 3kmCent_3kmPert. White cells indicate that aggregate FSSs of the two experiments were not statistically significantly different at the 90% confidence level or higher. Crosses indicate those forecast hours and neighborhood length scales where the highest aggregate FSS among 3kmCent_3kmPert and 3kmCent_15kmPert was lower than the FSS necessary to achieve "useful" skill, as defined by Roberts and Lean (2008). Annotations to the right of (f) represent the number of occurrences where, for the given row (i.e., a given neighborhood length scale), 3kmCent_3kmPert had statistically significantly higher FSSs than 3kmCent_15kmPert across all six percentile thresholds and all forecast hours (i.e., the total number of pink-shaded boxes in each row across all six panels).

EnKF-based IC perturbations limited by their various Gaussian assumptions. In weakly forced scenarios and with future advances in non-Gaussian DA methods and NWP models, conclusions could differ.

Nonetheless, given that it appears IC perturbations can be coarser than central initial states for CAE forecasting applications, dual-resolution EnVar DA systems may be prime candidates to initialize future CAEs because they can provide convection-allowing analyses while leveraging relatively coarse, cheap ensemble perturbations (e.g., Schwartz 2016; Lu et al. 2017). These relatively coarse ensembles could then be recentered about convection-allowing deterministic EnVar analyses to initialize CAE forecasts. Dual-resolution EnKFs (e.g., Rainwater and Hunt 2013) could also potentially be developed for CAE initialization. However, further work is needed to determine whether dual-resolution EnKFs or EnVar DA systems can produce similar quality analyses as those provided by single-resolution convection-allowing EnKFs. Thus, although convection-allowing IC perturbations appear unnecessary for CAE forecasts, paradoxically, ensemble-based BECs possessing convection-allowing horizontal grid spacing could conceivably be necessary to produce the best possible convection-allowing central initial states and to leverage the full potential of high-resolution observations like those derived from radars.

In conclusion, our results suggest scientists working on initialization of future operational CAEs like the RRFS primarily concentrate their energies on producing the best possible high-resolution deterministic analyses that can be used as central initial states for CAEs. A common focus on this aspect of CAE ICs across the community can potentially accelerate progress toward advancing CAE capabilities, thus leading to better probabilistic weather forecasts.

Acknowledgments. This material is based upon work supported by the National Center for Atmospheric Research (NCAR), which is a major facility sponsored by the National Science Foundation under Cooperative Agreement 1852977. Additional funding was provided by NCAR's Short-term Explicit Prediction (STEP) program and NOAA/OAR Office of Weather and Air Quality Grants NA17OAR4590182 and NA19OAR4590232. All forecasts were produced on NCAR's Cheyenne supercomputer (Computational and Information

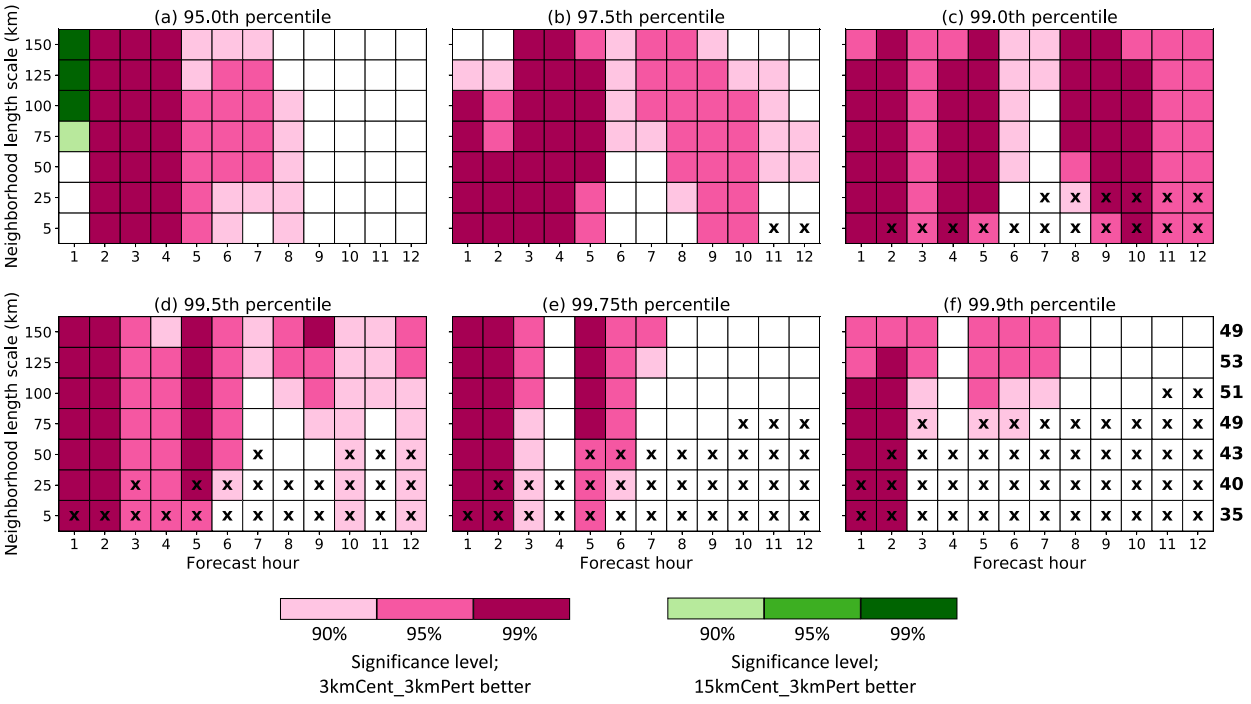

FIG. B2. As in Fig. B1, but comparing 3kmCent_3kmPert and 15kmCent_3kmPert to focus on the impact of changing central initial state resolution. Annotations to the right of (f) represent the number of occurrences where, for the given row (i.e., a given neighborhood length scale), 3kmCent_3kmPert had statistically significantly higher FSSs than 15kmCent_3kmPert across all six percentile thresholds and all forecast hours (i.e., the total number of pink-shaded boxes in each row across all six panels).

Systems Laboratory 2017). Thanks to Jonathan Guerrette (NCAR) and three anonymous reviewers for their constructive comments.

Data availability statement. 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. Data for each assimilation 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 were saved and are available upon request.

APPENDIX A

Sensitivity of 15-km EnKFs to Horizontal Localization and Observation Thinning

Our collective results indicate that centering IC perturbations about 3-km EnKF mean analyses led to better CAE forecasts than centering IC perturbations about 15-km EnKF mean analyses. To ensure that these findings were not attributable to different horizontal localization and observation thinning between the 15- and 3-km EnKFs (e.g., Table 1), we ran another continuously cycling 15-km EnKF. This auxiliary 15-km EnKF (hereafter referred to as "EnKF_15km$_{aux}$") was identical to the 15-km EnKF described in section 2 (hereafter referred to as "EnKF_15km$_{orig}$"), except it used smaller horizontal localization and observation thinning distances that matched those in the 3-km EnKF. This new EnKF (i.e., EnKF_15km$_{aux}$) was continuously cycled between 0000 UTC 23 April and 0000 UTC 7 May 2017 (inclusive) with a 1-h cycling period. Given that deterministic forecasts initialized from EnKF mean analyses yielded identical conclusions about forecast performance as 10-member CAEs [section 4b(3)], we only initialized 3-km forecasts from EnKF mean analyses produced by EnKF_15km$_{aux}$ to save computing resources.

Relative to EnKF_15km$_{orig}$, the smaller localization and assimilation of more aircraft and satellite-tracked wind observations in EnKF_15km$_{aux}$ led to closer fits to observations (not shown) and often increased FSSs at forecast hour 1 (Fig. A1). However, 1-h precipitation forecasts from EnKF_15km$_{aux}$ remained statistically significantly worse than those initialized from 3-km EnKF mean analyses. Moreover, beyond 1 h, EnKF_15km$_{aux}$ frequently had poorer precipitation forecasts than EnKF_15km$_{orig}$ (Fig. A1).

Therefore, different horizonal localization and observation thinning between EnKF_15km$_{orig}$ and the 3-km EnKF cannot explain why CAEs initially centered on 3-km EnKF mean analyses were better than those initially centered on 15-km EnKF mean analyses. Rather, fundamental differences between the 15- and 3-km EnKFs—horizontal resolution and associated

treatment of convection—were likely the dominant driver of their differences.

APPENDIX B

Statistical Significance of FSSs as a Function of Spatial Scale

To demonstrate that our overall conclusions were unaffected by neighborhood length scale (r), we present statistical significance of aggregate FSS differences between two pairs of experiments for r between 5 and 150 km (Figs. B1 and B2). Given fixed 3-km central initial states, 3-km IC perturbations only occasionally led to statistically significantly higher FSSs than 15-km IC perturbations through 12 h, regardless of r (Fig. B1). Conversely, given fixed 3-km IC perturbations, 3-km central initial states frequently led to statistically significantly higher FSSs than 15-km central initial states over the first 12 h, again, regardless of r (Fig. B2). Over the first 1–3 h at the 99.0–99.9th percentile thresholds, 3-km IC perturbations somewhat more regularly led to statistically significantly higher FSSs than 15-km IC perturbations given fixed 3-km central initial states when reducing r below 100 km (Figs. B1c–f). However, in most of these instances with statistically significant differences for $r < 100$ km, aggregate FSSs were below those associated with "useful" skill, as defined by Roberts and Lean (2008) [given by $(0.5 + f_0/2)$, where f_0 is the observed event frequency]. Ultimately, for all r, changing the central initial state led to more statistically significant differences than changing IC perturbations, echoing conclusions given by Figs. 7 and 11, and focusing on statistics computed with $r = 100$ km in the main text appropriately represents our findings.

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.

——, 2013: Nonlinear characteristics of ensemble perturbation evolution and their application to forecasting high-impact events. *Wea. Forecasting*, **28**, 1353–1365, https://doi.org/10.1175/WAF-D-12-00090.1.

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.

Bédard, J., M. Buehner, J.-F. Caron, S.-J. Baek, and L. Fillion, 2018: Practical ensemble-based approaches to estimate atmospheric background error covariances for limited-area deterministic data assimilation. *Mon. Wea. Rev.*, **146**, 3717–3733, https://doi.org/10.1175/MWR-D-18-0145.1.

——, J.-F. Caron, M. Buehner, S.-J. Baek, and L. Fillion, 2020: Hybrid background error covariances for a limited-area deterministic weather prediction system. *Wea. Forecasting*, **35**, 1051–1066, https://doi.org/10.1175/WAF-D-19-0069.1.

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.

Buehner, M., R. McTaggart-Cowan, and S. Heilliette, 2017: An ensemble Kalman filter for numerical weather prediction based on variational data assimilation: VarEnKF. *Mon. Wea. Rev.*, **145**, 617–635, https://doi.org/10.1175/MWR-D-16-0106.1.

Cafaro, C., and Coauthors, 2021: Do convection-permitting ensembles lead to more skillful short-range probabilistic rainfall forecasts over tropical East Africa? *Wea. Forecasting*, **36**, 697–716, https://doi.org/10.1175/WAF-D-20-0172.1.

Carley, J. R., and Coauthors, 2021: Status of NOAA's next generation convection-allowing ensemble: The Rapid Refresh Forecast System. *Special Symp. on Global and Mesoscale Models:Updates and Center Overviews WAF Symp. General Session*, online, Amer. Meteor. Soc., 12.8, https://ams.confex.com/ams/101ANNUAL/meetingapp.cgi/Paper/378383.

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., 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, 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.

Clayton, A. M., A. C. Lorenc, and D. M. Barker, 2013: Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quart. J. Roy. Meteor. Soc.*, **139**, 1445–1461, https://doi.org/10.1002/qj.2054.

Computational and Information Systems Laboratory, 2017: Cheyenne: HPE/SGI ICE XA System (NCAR Community Computing). National Center for Atmospheric Research, accessed 27 June 2022, https://doi.org/10.5065/D6RX99HX.

Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367–1387, https://doi.org/10.1002/qj.49712051912.

Davis, C., W. Wang, J. Dudhia, and R. Torn, 2010: Does increased horizontal resolution improve hurricane wind forecasts? *Wea. Forecasting*, **25**, 1826–1841, https://doi.org/10.1175/2010WAF2222423.1.

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.

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

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

Durran, D. R., and M. Gingrich, 2014: Atmospheric predictability: Why butterflies are not important. *J. Atmos. Sci.*, **71**, 2476–2488, https://doi.org/10.1175/JAS-D-14-0007.1.

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.

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.

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.

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.

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.

——, and C. Snyder, 2000: A hybrid ensemble Kalman filter–3D variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905–2919, https://doi.org/10.1175/1520-0493(2000)128<2905:AHEKFV>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., A. Walser, W. Langhans, and C. Schär, 2008: Cloud-resolving ensemble simulations of the August 2005 Alpine flood. *Quart. J. Roy. Meteor. Soc.*, **134**, 889–904, https://doi.org/10.1002/qj.252.

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.

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.

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.

Ji, M., and Coauthors, 2016: Dynamical core evaluation test report for NOAA's Next Generation Global Prediction System (NGGPS). NOAA Rep., 95 pp., https://repository.library.noaa.gov/view/noaa/18653.

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 ——, 2020: Interactions between physics diversity and multiscale initial condition perturbations for storm-scale ensemble forecasting. *Mon. Wea. Rev.*, **148**, 3549–3565, https://doi.org/10.1175/MWR-D-20-0112.1.

——, ——, 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.

——, ——, and S. Degelia, 2017: Design and implementation of a GSI-based convection allowing ensemble-based data assimilation and forecast system for the PECAN field experiment.

Part II: Overview and evaluation of a real-time system. *Wea. Forecasting*, **32**, 1227–1251, https://doi.org/10.1175/WAF-D-16-0201.1.

Kay, J., and X. Wang, 2020: A multiresolution ensemble hybrid 4DEnVar for global numerical prediction. *Mon. Wea. Rev.*, **148**, 825–847, https://doi.org/10.1175/MWR-D-19-0002.1.

Kong, F., and Coauthors, 2008: Real-time storm-scale ensemble forecast experiment—Analysis of 2008 spring experiment data. *24th Conf. on Severe Local Storms*, Savannah, GA, Amer. Meteor. Soc., 12.3., https://ams.confex.com/ams/pdfpapers/141827.pdf.

——, and Coauthors, 2009: A real-time storm-scale ensemble forecast system: 2009 spring experiment. *23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction*, Omaha, NE, Amer. Meteor. Soc., 16A.3., https://ams.confex.com/ams/pdfpapers/154118.pdf.

Kühnlein, C., C. Keil, G. C. Craig, and C. Gebhardt, 2014: The impact of downscaled initial condition perturbations on convective-scale ensemble forecasts of precipitation. *Quart. J. Roy. Meteor. Soc.*, **140**, 1552–1562, https://doi.org/10.1002/qj.2238.

Lei, L., and J. S. Whitaker, 2017: Evaluating the trade-offs between ensemble size and ensemble resolution in an ensemble-variational data assimilation system. *J. Adv. Model. Earth Syst.*, **9**, 781–789, https://doi.org/10.1002/2016MS000864.

Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409–418, https://doi.org/10.1175/1520-0493(1974)102<0409:TSOMCF>2.0.CO;2.

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.

Liu, Z., J. Ban, J.-S. Hong, and Y.-H. Kuo, 2020: Multi-resolution incremental 4D-Var for WRF: Implementation and application at convective scale. *Quart. J. Roy. Meteor. Soc.*, **146**, 3661–3674, https://doi.org/10.1002/qj.3865.

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

——, M. Jardak, T. Payne, N. E. Bowler, and M. A. Wlasak, 2017: Computing an ensemble of variational data assimilations using its mean and perturbations. *Quart. J. Roy. Meteor. Soc.*, **143**, 798–805, https://doi.org/10.1002/qj.2965.

Lu, X., X. Wang, M. Tong, and V. Tallapragada, 2017: GSI-based, continuously cycled, dual-resolution hybrid ensemble–variational data assimilation system for HWRF: System description and experiments with Edouard (2014). *Mon. Wea. Rev.*, **145**, 4877–4898, https://doi.org/10.1175/MWR-D-17-0068.1.

Mahoney, K. M., 2016: The representation of cumulus convection in high-resolution simulations of the 2013 Colorado Front Range flood. *Mon. Wea. Rev.*, **144**, 4265–4278, https://doi.org/10.1175/MWR-D-16-0211.1.

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.

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.

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 long-wave. *J. Geophys. Res.*, **102**, 16 663–16 682, https://doi.org/10.1029/97JD00237.

Murphy, A. H., 1973: A new vector partition of the probability score. *J. Appl. Meteor. Climatol.*, **12**, 595–600, https://doi.org/10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2.

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.

Pan, Y., M. Xue, K. Zhu, and M. Wang, 2018: A prototype regional GSI-based EnKF-variational hybrid data assimilation system for the Rapid Refresh forecasting system: Dual-resolution implementation and testing results. *Adv. Atmos. Sci.*, **35**, 518–530, https://doi.org/10.1007/s00376-017-7108-0.

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.

Posselt, D. J., and C. H. Bishop, 2012: Nonlinear parameter estimation: Comparison of an ensemble Kalman smoother with a Markov chain Monte Carlo algorithm. *Mon. Wea. Rev.*, **140**, 1957–1974, https://doi.org/10.1175/MWR-D-11-00242.1.

Poterjoy, J., R. A. Sobash, and J. L. Anderson, 2017: Convective-scale data assimilation for the Weather Research and Forecasting Model using the local particle filter. *Mon. Wea. Rev.*, **145**, 1897–1918, https://doi.org/10.1175/MWR-D-16-0298.1.

——, L. Wicker, and M. Buehner, 2019: Progress toward the application of a localized particle filter for numerical weather prediction. *Mon. Wea. Rev.*, **147**, 1107–1126, https://doi.org/10.1175/MWR-D-17-0344.1.

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

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.

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, B., B. T. Gallo, I. L. Jirak, A. J. Clark, D. C. Dowell, X. Wang, and Y. Wang, 2020: What does a convection-allowing

ensemble of opportunity buy us in forecasting thunderstorms? *Wea. Forecasting*, **35**, 2293–2316, https://doi.org/10.1175/WAF-D-20-0069.1.

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

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.

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.

——, 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.

——, ——, R. A. Sobash, K. R. Fossell, and M. L. Weisman, 2015a: 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.

——, Z. Liu, and X.-Y. Huang, 2015b: Sensitivity of limited-area hybrid variational-ensemble analyses and forecasts to ensemble perturbation resolution. *Mon. Wea. Rev.*, **143**, 3454–3477, https://doi.org/10.1175/MWR-D-14-00259.1.

——, 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.

——, G. S. Romine, and D. C. Dowell, 2021: 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. *Wea. Forecasting*, **36**, 379–405, https://doi.org/10.1175/WAF-D-20-0110.1.

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.

Smolarkiewicz, P. K., and G. A. Grell, 1992: A class of monotone interpolation schemes. *J. Comput. Phys.*, **101**, 431–440, https://doi.org/10.1016/0021-9991(92)90018-T.

Surcel, M., I. Zawadzki, and M. K. Yau, 2014: On the filtering properties of ensemble averaging for storm-scale precipitation forecasts. *Mon. Wea. Rev.*, **142**, 1093–1105, https://doi.org/10.1175/MWR-D-13-00134.1.

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.

Tennant, W., 2015: Improving initial condition perturbations for MOGREPS-UK. *Quart. J. Roy. Meteor. Soc.*, **141**, 2324–2336, https://doi.org/10.1002/qj.2524.

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.

Vié, B., O. Nuissier, and V. Ducrocq, 2011: Cloud-resolving ensemble simulations of Mediterranean heavy precipitation events: Uncertainty on initial conditions and lateral boundary conditions. *Mon. Wea. Rev.*, **139**, 403–423, https://doi.org/10.1175/2010MWR3487.1.

Wang, X., 2010: Incorporating ensemble covariance in the Gridpoint Statistical Interpolation (GSI) variational minimization: A mathematical framework. *Mon. Wea. Rev.*, **138**, 2990–2995, https://doi.org/10.1175/2010MWR3245.1.

——, D. M. Barker, C. Snyder, and T. M. Hamill, 2008: A hybrid ETKF–3DVAR data assimilation scheme for the WRF Model. Part I: Observing system simulation experiment. *Mon. Wea. Rev.*, **136**, 5116–5131, https://doi.org/10.1175/2008MWR2444.1.

——, D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar-based ensemble–variational hybrid data assimilation for NCEP Global Forecast System: Single-resolution experiments. *Mon. Wea. Rev.*, **141**, 4098–4117, https://doi.org/10.1175/MWR-D-12-00141.1.

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

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.

Whitaker, J. S., and T. M. Hamill, 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. Elsevier, 676 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.

Xue, M., and Coauthors, 2008: CAPS real-time storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2007 Spring Experiment. *24th Conf. on Severe Local Storms*, Savannah, GA, Amer. Meteor. Soc., 12.2, https://ams.confex.com/ams/24SLS/techprogram/paper_142036.htm.

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, X., 2019: Multiscale characteristics of different-source perturbations and their interactions for convection-permitting ensemble forecasting during SCMREX. *Mon. Wea. Rev.*, **147**, 291–310, https://doi.org/10.1175/MWR-D-18-0218.1.

——, 2021: Case dependence of multiscale interactions between multisource perturbations for convection-permitting ensemble forecasting during SCMREX. *Mon. Wea. Rev.*, **149**, 1853–1871, https://doi.org/10.1175/MWR-D-20-0316.1.

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.