Comparing Partial and Continuously Cycling Ensemble Kalman Filter Data Assimilation Systems for Convection-Allowing Ensemble Forecast Initialization

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

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

(Manuscript received 6 May 2021, in final form 3 August 2021)

ABSTRACT: Several limited-area 80-member ensemble Kalman filter (EnKF) data assimilation systems with 15-km horizontal grid spacing were run over a computational domain spanning the conterminous United States (CONUS) for a 4-week period. One EnKF employed continuous cycling, where the prior ensemble was always the 1-h forecast initialized from the previous cycle's analysis. In contrast, the other EnKFs used a partial cycling procedure, where limited-area states were discarded after 12 or 18 h of self-contained hourly cycles and reinitialized the next day from global model fields. "Blended" states were also constructed by combining large scales from global ensemble initial conditions (ICs) with small scales from limited-area continuously cycling EnKF analyses using a low-pass filter. Both the blended states and EnKF analysis ensembles initialized 36-h, 10-member ensemble forecasts with 3-km horizontal grid spacing. Continuously cycling EnKF analyses initialized ~1–18-h precipitation forecasts that were comparable to or somewhat better than those with partial cycling EnKF ICs. Conversely, ~18–36-h forecasts with partial cycling EnKF ICs were comparable to or better than those with unblended continuously cycling EnKF ICs. However, blended ICs yielded ~18–36-h forecasts that were statistically indistinguishable from those with partial cycling ICs. ICs that more closely resembled global analysis spectral characteristics at wavelengths > 200 km, like partial cycling and blended ICs, were associated with relatively good ~18–36-h forecasts. Ultimately, findings suggest that EnKFs employing a combination of continuous cycling and blending can potentially replace the partial cycling assimilation systems that currently initialize operational limited-area models over the CONUS without sacrificing forecast quality.

SIGNIFICANCE STATEMENT: Numerical weather prediction models (i.e., weather models) are initialized through a process called data assimilation, which combines real atmospheric observations with a previous short-term weather model forecast using statistical techniques. The overarching data assimilation strategy currently used to initialize operational regional weather models over the United States has several disadvantages that ultimately limit progress toward improving weather model forecasts. Thus, we suggest an alternative data assimilation strategy be adopted to initialize a next-generation, high-resolution (~3 km) probabilistic forecast system currently being developed. This alternative approach preserves forecast quality while fostering a framework that can accelerate weather model improvements, which in turn will lead to better weather forecasts.

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

1. Introduction

Limited-area convection-allowing ensembles (CAEs) have become increasingly popular over the past decade and are now operational at many numerical weather prediction (NWP) centers (e.g., Gebhardt et al. 2011; Peralta et al. 2012; Hagelin et al. 2017; Raynaud and Bouttier 2017; Klasa et al. 2018; Roberts et al. 2020). While CAEs can be initialized by simply downscaling operationally available coarse-resolution analyses and short-term forecasts onto the computational domain (e.g., Xue et al. 2007; Kong et al. 2008, 2009; Tennant 2015; Clark 2017; Schellander-Gorgas et al. 2017; Cafaro et al. 2019; Porson et al. 2019), as data assimilation (DA) methods have matured and computing has increased, CAE initial conditions (ICs) are now commonly produced by customized limited-area DA systems explicitly designed for CAE initialization (e.g., Jones and Stensrud 2012; Schumacher and Clark 2014; Schwartz et al. 2014, 2015, 2021; Harnisch and Keil 2015; Wheatley et al. 2015; Dowell et al. 2016, 2021, manuscript submitted to *Wea. Forecasting*; Raynaud and Bouttier 2016; Schraff et al. 2016; Johnson and Wang 2017; Gustafsson et al. 2018; Keresturi et al. 2019; Gasperoni et al. 2020; Johnson et al. 2020; COSMO 2021).

Over the conterminous United States (CONUS), NCEP's operational CAE, the High Resolution Ensemble Forecast (HREF; Roberts et al. 2020), currently lacks its own analysis system and instead is an ad hoc aggregation of independent deterministic convection-allowing model forecasts. However, within the Unified Forecast System framework, NCEP intends to replace the HREF with a Rapid Refresh Forecast System (RRFS; Carley et al. 2021) initialized from its own ensemble-based limited-area analyses. Thus, configurations for the RRFS's DA system must be carefully considered.

One design choice concerns DA cycling methodology, as two overarching strategies are possible: continuous cycling and partial cycling. In continuous cycling, the short-term forecast

Corresponding author: Craig S. Schwartz, schwartz@ucar.edu

DOI: 10.1175/WAF-D-21-0069.1

TABLE 1. WRF Model settings for all experiments.

Parameter	WRF Model setting
Model version	Version 3.9.1.1 of the Advanced Research version of WRF
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)

initialized from the previous cycle's analysis always serves as the background for the current analysis cycle, relegating the role of external models to supplying boundary conditions and yielding a self-contained limited-area DA system. Conversely, in partial cycling, limited-area analysis cycles are periodically discarded and replaced with coarser-resolution external analyses or short-term forecasts typically provided by a global NWP model.

Although CAE forecast sensitivity to cycling strategy has not been systematically examined, prior research at convection-parameterizing resolutions indicated partial cycling three-dimensional variational (3DVAR; e.g., Courtier et al. 1994; Lorenc et al. 2000) DA systems initialized better deterministic forecasts than continuously cycling 3DVAR DA systems (e.g., Rogers et al. 2009; Hsiao et al. 2012; Benjamin et al. 2016). While reasons for these findings are not completely understood, one possibility is that continuously cycling DA systems poorly represent large-scale features that may exert important controls on forecast evolution (e.g., Durran and Gingrich 2014; Durran and Weyn 2016; Weyn and Durran 2017), whereas partial cycling DA systems might possess smaller large-scale errors because they ingest global fields with a "better longwave representation not available via regional data assimilation unable to use the full global set of observations" (Benjamin et al. 2016). Another possible reason for the historical deficiencies of continuously cycling systems may be related to biases that can accumulate throughout continuous DA cycles; these biases likely arise from imperfect physical parameterizations and can eventually degrade analyses and subsequent forecasts. In contrast, the act of periodically replacing limited-area states with comparatively less biased global fields may limit how much bias can accumulate in partial cycling DA systems. For example, Hsiao et al. (2012) demonstrated that partial cycling 3DVAR analyses were substantially less biased than continuously cycling 3DVAR analyses and initialized commensurately better forecasts over Taiwan and its surroundings, and several studies employing continuous cycling over the CONUS and adjacent areas also documented bias accumulations (e.g., Torn and Davis 2012; Romine et al. 2013; Cavallo et al. 2016; Wong et al. 2020; Poterjoy et al. 2021).

Given the collective findings questioning the suitability of continuous cycling, NCEP's operational limited-area North American Mesoscale Forecast System (NAM), Rapid Refresh (RAP; Benjamin et al. 2016), and High-Resolution Rapid Refresh (HRRR; Benjamin et al. 2016; Dowell et al. 2021, manuscript submitted to *Wea. Forecasting*) models, as well as NOAA's experimental real-time CAEs, the HRRR-Ensemble (HRRRE; Dowell et al. 2016, 2021, manuscript submitted to *Wea. Forecasting*) and "Warn-on-Forecast" system (Stensrud et al. 2009, 2013; Wheatley et al. 2015; Jones et al. 2016), all employ partial cycling.[1] In addition, several research studies effectively used partial cycling approaches to initialize convection-allowing model forecasts over the CONUS (e.g., Schumacher and Clark 2014; Johnson et al. 2015, 2020; Johnson and Wang 2017; Gasperoni et al. 2020).

However, partial cycling DA systems have several limitations. For instance, while continuous cycling facilitates a straightforward diagnosis of model biases—such that they can be remedied—forecast errors in partial cycling systems reflect both the external and limited-area models, increasing the difficulty of pinpointing error sources or masking errors altogether (e.g., Poterjoy et al. 2021). Additionally, partial cycling DA system performance may depend on both characteristics of the external fields and frequency with which they are ingested, introducing extra sources of potential sensitivity compared to continuously cycling DA systems. Furthermore, partial cycling workflows can be complicated and require simultaneous execution of two limited-area DA systems, including a "primary" system and a "parallel" or "catch-up" system that essentially handles the periodic ingestion of external fields (e.g., Djalalova et al. 2016; Hu et al. 2017). Perhaps recognizing these shortcomings, Rogers et al. (2009) noted,

[1] Notably, operational European limited-area models, including CAEs, are initialized from continuously cycling DA systems (e.g., Schraff et al. 2016; Hagelin et al. 2017; Raynaud and Bouttier 2017; Keresturi et al. 2019; COSMO 2021). Although it is unclear whether this approach is optimal given the absence of studies intercomparing forecasts initialized from partial and continuously cycling DA systems over Europe, it is possible that continuously cycling DA systems spanning relatively large geographic areas like the CONUS may be more prone to the issue of bias accumulation than continuously cycling DA systems over comparatively small European domains, where more assertive lateral boundary conditions (e.g., Warner et al. 1997) may limit bias accumulations.

"It should be pointed out that the use of partial cycling in the [NAM DA system] is considered a temporary solution," and overall, relative to partial cycling DA systems, continuously cycling DA systems permit more rapid progress toward improving NWP models and are easier to maintain and upgrade. Accordingly, it would be preferable to initialize the future RRFS with a continuously cycling DA system, so long as it produces similar quality forecasts as other potential initialization methods like partial cycling.

Thus, for RRFS development purposes, it seems sensible to rigorously revisit partial versus continuous cycling for limited-area modeling applications, especially with modern DA systems incorporating flow-dependent background error covariances like the ensemble Kalman filter (EnKF; Evensen 1994; Houtekamer and Zhang 2016), contrasting previous systematic studies concerning partial and continuous cycling that employed inferior 3DVAR DA methodologies (e.g., Rogers et al. 2009; Hsiao et al. 2012; Benjamin et al. 2016). Moreover, limited-area continuously cycling EnKFs can perform well and initialize better convection-allowing model forecasts than downscaled global analyses over the CONUS (e.g., Schwartz and Liu 2014; Schwartz 2016), including for CAE applications (Schwartz et al. 2021). Finally, Schumacher and Clark (2014) suggested partial and continuously cycling EnKFs yielded similar caliber CAE forecasts, which is encouraging, but their study was limited by its small sample size of just 16 assimilation cycles over 4 days and specific experimental design choices, like initializing their partial cycling EnKF with randomly perturbed 36-h forecasts rather than flow-dependent analyses or shorter-term forecasts. Ultimately, it remains unclear whether continuously cycling EnKFs can systematically initialize comparable quality CAE forecasts as partial cycling EnKFs, as there has yet to be a study devoted to such an investigation.

To address this uncertainty about cycling strategy, this work directly compares CAE forecasts initialized from partial and continuously cycling EnKF DA systems over the CONUS for a 4-week period. In addition, this study investigates another method for CAE forecast initialization that, like partial cycling, entrains external information into limited-area ICs. Specifically, CAE forecasts were also initialized from "blended" states, where small scales provided by continuously cycling EnKF analyses were combined with large scales provided by global ensemble ICs. Our experiments offer insights about CAE ICs and guidance for how future CAEs like the RRFS should be initialized.

2. Model and data assimilation configurations

CAE forecast sensitivity to EnKF cycling procedure (i.e., partial or continuous cycling) was explored through several experiments. The following descriptions about experimental model and EnKF settings are brief, as configurations were identical to those described by Schwartz et al. (2021; hereafter S21). Despite this parallel, the current study fundamentally differs from S21, who focused on comparing CAE forecasts initialized from continuously cycling 15- and 3-km EnKFs and did not intercompare forecasts initialized from partial and continuously cycling EnKFs.

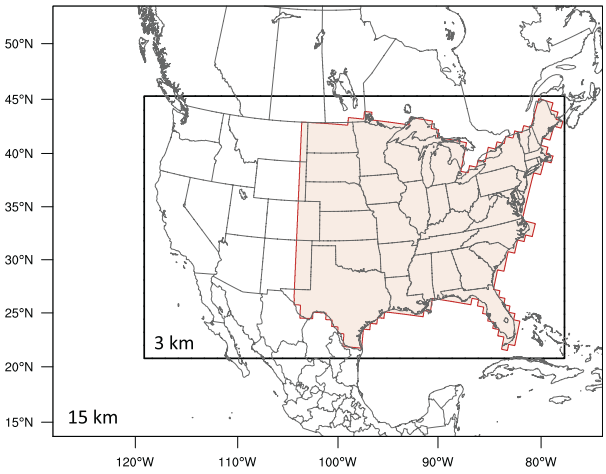

FIG. 1. Computational domain. Horizontal grid spacing was 15 km in the outer domain (415 × 325 points) and 3 km in the nest (1581 × 986 points). Objective precipitation verification only occurred over the red shaded region of the 3-km domain (CONUS east of 105°W).

a. *Model configurations*

All EnKF experiments used identical model configurations as S21 (Table 1). Specifically, 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) produced all forecasts over a nested computational domain with 15-km horizontal grid spacing in the outer domain and 3-km horizontal grid spacing in the nest (Fig. 1). The same physics options (Table 1) were used on both domains, except cumulus parameterization was not employed on the convection-allowing 3-km grid. All ensemble members used identical physical parameterizations.

b. *EnKF configurations*

Both the partial and continuously cycling EnKFs had identical configurations to the 15-km continuously cycling EnKF described by S21, who thoroughly documented and justified their settings (summarized in Table 2). Moreover, S21 showed their 15-km EnKF DA system had acceptable spread error statistics (e.g., Houtekamer et al. 2005), was stable from a climatological perspective, and initialized better short-term CAE precipitation forecasts than ICs provided by an operational global ensemble.

Specifically, using the Data Assimilation Research Testbed (DART; Anderson et al. 2009) software, 80-member EnKF analyses were produced hourly (i.e., hourly cycles) on solely the 15-km domain (Fig. 1); the 3-km domain was removed during 1-h, 80-member ensemble forecasts between EnKF analyses. As in S21, these 1-h ensemble forecasts employed perturbed lateral boundary conditions (LBCs) that were constructed by adding random, correlated, Gaussian noise with zero mean (e.g., Barker 2005; Torn et al. 2006) to Global Forecast System (GFS) analyses and forecasts; this approach was chosen for its simplicity and is commonly used to provide

TABLE 2. Summary of partial and continuously cycling EnKF configurations. See S21 for more details and 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	1280 km
Vertical localization full width	1.0 scale height
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
Assimilated observations	Rawinsonde, aircraft, wind profiler, satellite-tracked wind, global positioning system radio occultation (GPSRO), and surface observations
Moisture observations	Assimilated as relative humidity
Horizontal thinning for aircraft and satellite-tracked wind observations	30 km
Vertical thinning for aircraft and satellite-tracked wind observations	25 hPa

LBCs for limited-area EnKFs (e.g., Torn and Davis 2012; Romine et al. 2013; Schumacher and Clark 2014; Johnson et al. 2015; Schwartz et al. 2015, 2020; Zhu et al. 2019). Whereas S21 produced both 15- and 3-km EnKF analyses, we only produced the more affordable 15-km analyses to enable several 4-week experiments (section 3) given finite computing resources. Although future operational CAEs will likely be initialized from convection-allowing DA systems, as we further discuss in section 6, higher-resolution DA systems would probably not provide different conclusions about the comparative performance of partial and continuously cycling DA methodologies.

The EnKFs used sampling error correction (Anderson 2012) and covariance localization to mitigate spurious correlations, and EnKF spread was maintained with posterior inflation (Table 2). Approximately 30 000–100 000 conventional observations were assimilated each cycle (Table 2), all assumed to be valid at the analysis time. Radar-based observations were not assimilated. Furthermore, as in S21, radiance observations were not assimilated. There are two reasons for this choice: 1) consistency with S21; and 2) although assimilating radiances has shown promise for improving forecasts of specific events over small portions of the CONUS (e.g., Zou et al. 2011; Zhang et al. 2019; Jones et al. 2020), radiance observations historically have yielded only small impacts over the CONUS in *systematic studies* with limited-area DA systems (Lin et al. 2017a,b; Zhu et al. 2019), likely because of ample conventional observation coverage over the CONUS. Table 3 of S21 provides a complete list of assimilated observations.

Following S21, NCEP's operational Gridpoint Statistical Interpolation (GSI) DA system (Kleist et al. 2009; Shao et al. 2016) provided observation operators, performed observation quality control, thinned aircraft and satellite-tracked wind observations (Table 2), specified observation time windows, and assigned observation errors. GSI's observation-related output was then ingested into DART.

It is important to note that specific DA configurations (e.g., Table 2) were determined while developing the continuously cycling EnKF, and optimal settings for the partial cycling EnKFs may differ. Thus, a hypothetical operational partial cycling EnKF that has been exhaustively tuned may perform better than our partial cycling EnKFs. Nonetheless, fine-tuning partial cycling DA parameters is beyond the scope of this study, and all EnKFs used identical configurations to attribute differences between partial and continuously cycling EnKF analyses and subsequently initialized CAE forecasts to the external fields introduced during partial cycling EnKF initialization.

3. Experimental design

As in S21, EnKF experiments were performed between 23 April and 20 May 2017. This period featured several severe weather and heavy precipitation events over the CONUS.

a. Continuously cycling EnKF

The 80-member continuously cycling 15-km EnKF in S21 and used again here ("CC_{EnKF}"; Table 3; Fig. 2) was initialized by downscaling the 0000 UTC 23 April 2017 0.25° GFS analysis onto the 15-km domain (Fig. 1) and adding random, correlated, Gaussian noise with zero mean, akin to the method for generating perturbed LBCs (section 2b). This randomly generated ensemble served as the prior (before assimilation) ensemble for the first EnKF analysis, and the posterior (after assimilation) ensemble initialized a 1-h, 80-member ensemble forecast that became the prior for EnKF DA at 0100 UTC 23 April 2017.

Thereafter, continuous analysis–forecast cycles with a 1-h period were performed until 0000 UTC 20 May 2017 (inclusive; 649 hourly DA cycles). Land surface and microphysics states freely evolved for each member throughout the 4-week

TABLE 3. Summary of experiments. Also see Fig. 2.

Experiment name	Description
CC_{EnKF}	Continuously cycling EnKF initialized at 0000 UTC 23 Apr 2017 by adding random noise to GFS analyses. Hourly assimilation cycles were then performed until 0000 UTC 20 May 2017 (inclusive), and 0000 UTC analysis ensembles initialized 36-h, 10-member CAE forecasts.
PC_{12z}	Partial cycling EnKF initialized daily at 1200 UTC between 24 Apr and 19 May 2017 (inclusive) by recentering perturbations derived from 6-h GDAS-EnKF forecasts about 1200 UTC GFS analyses. The perturbations were inflated according to Fig. 3. Hourly self-contained assimilation cycles were then performed for 12 h until 0000 UTC, and 0000 UTC analysis ensembles initialized 36-h, 10-member CAE forecasts. After CAE forecast initialization, limited-area cycles were discarded.
PC_{06z}	Exactly the same as PC_{12z}, except the partial cycling EnKF was initialized 6 h earlier at 0600 UTC daily by recentering inflated perturbations derived from 6-h GDAS-EnKF forecasts about 0600 UTC GFS analyses. Hourly self-contained assimilation cycles were then performed for 18 h until 0000 UTC, and 0000 UTC analysis ensembles initialized 36-h, 10-member CAE forecasts.
PC_{12z_soil}	Exactly the same as PC_{12z}, except initial land surface states at 1200 UTC were taken from 1200 UTC continuously cycling EnKF (CC_{EnKF}) members.
CC_{EnKF_blend}	Exactly the same as CC_{EnKF}, except at 0000 UTC, small scales from CC_{EnKF} analysis members 1–10 were blended with large scales from corresponding GEFS IC members 1–10 using a 960-km filter cutoff (Fig. 4). These blended ICs then initialized 36-h, 10-member CAE forecasts. Blending did not impact continuous EnKF assimilation cycles.
GEFS	0000 UTC GEFS ICs were downscaled onto the computational domain to initialize 36-h, 10-member CAE forecasts.

cycling period, and sea surface temperatures were updated daily from NCEP's 0.12° analyses (e.g., Gemmill et al. 2007). S21 showed it took approximately two days for the EnKF to spin up from and effectively "forget" about the initially specified random noise (i.e., develop flow-dependent structures consistent with the WRF Model climate).

Members 1–10 from 0000 UTC posterior ensembles initialized 36-h forecasts over the nested domain (Fig. 1) between 25 April and 20 May 2017 (inclusive; 26 forecasts); the 10-member ensemble forecasts on the 3-km grid were the CAE forecasts of interest. Because only 15-km EnKF analyses were produced, the 3-km nest was initialized by downscaling 15-km posterior ensembles onto the 3-km grid. Although 80-member EnKF analyses were available, computing constraints limited CAE forecasts to just 10 members, which is sufficient to provide skillful and valuable probabilistic forecasts of precipitation and severe weather-related quantities (e.g., Clark et al. 2011, 2018; Schwartz et al. 2014) and is similar in size to the HRRRE and HREF. For these 36-h forecasts, LBCs provided by perturbation members 1–10 from NCEP's operational Global Ensemble Forecast System (GEFS; Zhou et al. 2017) with 0.5° horizontal grid spacing were applied to the 15-km domain, which in turn provided LBCs for the 3-km nest.

b. *Primary partial cycling EnKF*

The primary partial cycling EnKF ("PC_{12z}"; Table 3; Fig. 2) was initialized daily at 1200 UTC between 24 April and 19 May 2017 (inclusive). First, deterministic 1200 UTC 0.25° GFS analyses were interpolated onto the 15-km computational domain (Fig. 1). Then, flow-dependent perturbations of horizontal winds, temperature, water vapor mixing ratio, and surface pressure were derived from global, 6-h, 80-member ensemble forecasts valid at 1200 UTC; these 6-h global ensemble forecasts had T574 resolution (~34 km) and were initialized from operational EnKF analyses produced within NCEP's Global Data Assimilation System (GDAS; e.g., Whitaker and Hamill 2002; Whitaker et al. 2008; Wang et al. 2013). Finally, the GDAS-EnKF perturbations[2] were interpolated onto the 15-km grid and added to downscaled GFS analyses to construct 80-member ensembles that initialized the limited-area partial cycling EnKF. As the mean of GDAS-EnKF perturbations was zero, ensemble mean states at 1200 UTC in the partial cycling EnKF were identical to GFS analyses. Therefore, the partial cycling EnKF was influenced by radiance measurements assimilated within the GDAS, despite not assimilating these measurements directly. Moreover, in GDAS analyses, observations located outside the regional domain can influence locations within the regional domain, meaning 1200 UTC partial cycling EnKF states reflected observations outside the regional domain. Thus, from an observational perspective, the continuously cycling EnKF was somewhat disadvantaged with respect to the partial cycling EnKF, as the former was unable to implicitly benefit from additional observations through global analyses, aside from LBC influences.

Constructing initial ensembles by adding perturbations derived from GDAS-EnKF forecasts to GFS analyses is similar to HRRRE and GEFS initialization procedures (e.g., Zhou et al. 2017; Dowell et al. 2021, manuscript submitted to *Wea. Forecasting*). Additionally, perturbations were derived from 6-h ensemble forecasts, rather than from analysis ensembles, in recognition that using short-term forecasts to initialize partial cycling DA systems is common (e.g., Rogers et al. 2009; Benjamin et al. 2016; Djalalova et al. 2016; Hu et al. 2017; Wu et al. 2017; Dowell et al. 2021, manuscript submitted to *Wea. Forecasting*) given operational constraints sometimes

[2] Perturbations were defined with respect to the ensemble mean.

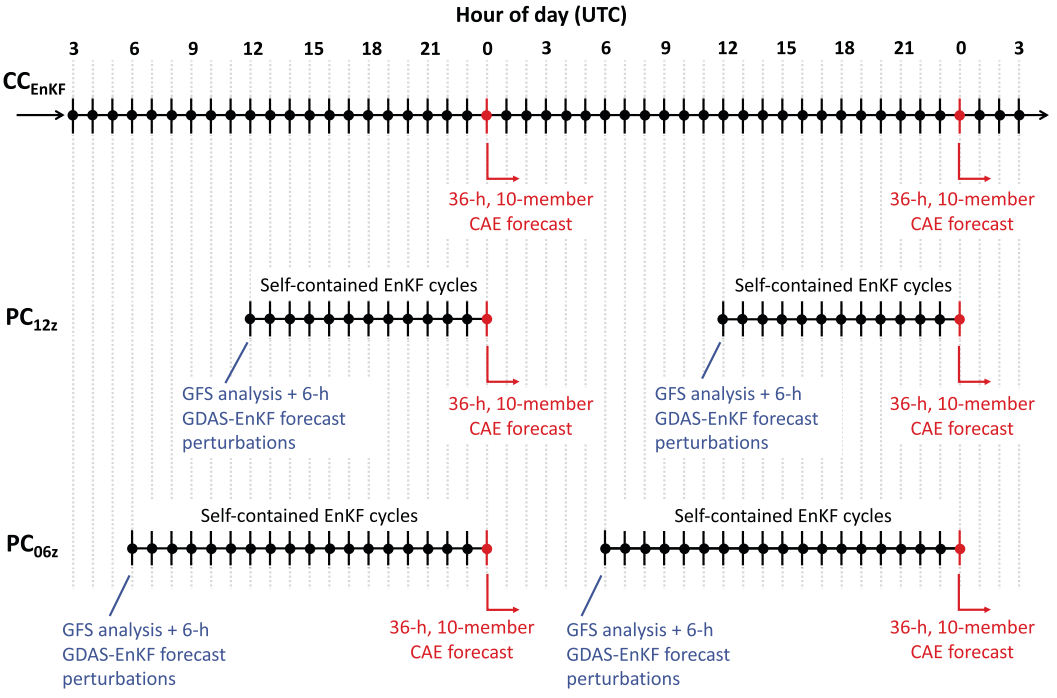

FIG. 2. Schematic diagram of CC_{EnKF} (top), PC_{12z} (middle), and PC_{06z} (bottom) cycling methodologies. Solid vertical lines with filled black circles represent EnKF analyses, and red denotes CAE forecast initialization times. Priors for EnKF analyses at 0600 and 1200 UTC in PC_{06z} and PC_{12z}, respectively, were 6-h GDAS-EnKF forecast perturbations recentered about GFS analyses.

requiring a modeling system to start before global ensemble analyses are available (e.g., Zhou et al. 2017).

The above procedure produced prior ensembles for 1200 UTC EnKF analyses, and 1200 UTC posterior ensembles initialized 1-h, 80-member ensemble forecasts that became priors for EnKF DA at 1300 UTC. Thereafter, self-contained hourly assimilation cycles were performed until 0000 UTC (i.e., 12 h of self-contained cycles) using identical configurations and assimilating the same observations as the continuously cycling EnKF (Table 2), with 1-h, 15-km, 80-member ensemble forecasts between analyses.

As with the continuously cycling EnKF, 0000 UTC posterior ensembles initialized 36-h, 10-member CAE forecasts between 25 April and 20 May 2017 (inclusive) that employed GEFS LBCs. After these CAE forecasts were initialized, 0000 UTC posterior ensembles were discarded and the partial cycling EnKF was initialized anew the next day (e.g., Fig. 2). Performing 12 h of self-contained cycles before initializing forecasts of interest was similar to Hsiao et al. (2012), the RAP (Benjamin et al. 2016; Hu et al. 2017), and previous versions of the NAM DA system (Wu et al. 2017).

c. *Intricacies of partial cycling initialization: Additional experiments and discussion*

Partial cycling EnKF initialization has several intricacies and subjectivities that warrant discussion and motivated additional experimentation. Notably, the following issues are irrelevant for continuously cycling EnKFs, illustrating how partial cycling EnKFs have more sources of potential sensitivity than continuously cycling EnKFs.

1) PARTIAL CYCLING DURATION

CAE forecasts initialized from partial cycling EnKFs could be sensitive to self-contained cycling length, and previous partial cycling systems employed between 6 and 24 h of self-contained cycles before initializing forecasts of interest (e.g., Johnson et al. 2015; Wu et al. 2017; Gasperoni et al. 2020). As determining the optimal self-contained cycling length for CAE forecast initialization was not a primary goal of this study, we did not experiment with a wide range of self-contained cycling lengths.

Nonetheless, some of our results suggested performing only 12 h of self-contained cycles before initializing CAE forecasts may have been insufficient. Therefore, we initialized another partial cycling EnKF at 0600 UTC daily between 24 April and 19 May 2017 (inclusive) that produced 18 h of self-contained cycles until 0000 UTC, when posterior ensembles initialized 36-h, 10-member CAE forecasts ("PC_{06z}"; Table 3; Fig. 2). Aside from their initialization times, PC_{06z} and PC_{12z} were identically configured and used the same GFS/GDAS-EnKF initialization method (e.g., section 3b).

2) INITIAL ENSEMBLE SPREAD

Six-hour GDAS-EnKF forecast spread is not tuned for limited-area WRF Model applications and is potentially insufficient at low levels (e.g., Zhou et al. 2017; Gehne et al. 2019).

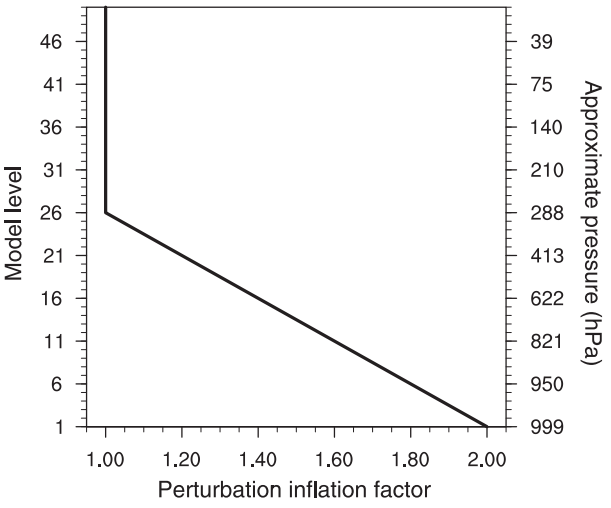

FIG. 3. Perturbation inflation factor as a function of model level (level 1 is nearest the ground). These inflation factors were applied to perturbations derived from 6-h GDAS-EnKF forecasts during partial cycling EnKF initialization. The approximate pressure (hPa) at each model level is given on the right axis.

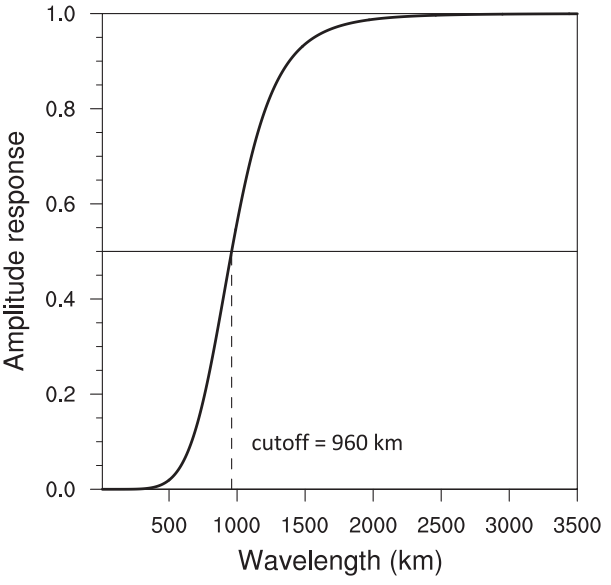

FIG. 4. Amplitude response (y axis) of a sixth-order implicit tangent filter as a function of wavelength (km) for a specified cutoff length of 960 km. In the context of this study, the curve denotes the contribution of GEFS ICs to blended ICs at a given wavelength (e.g., for wavelengths where the amplitude response is 1.0, 100% of the blended ICs at those wavelengths were from the GEFS). The dashed vertical and solid horizontal lines illustrate how the amplitude response is 0.5 at the specified cutoff length.

Thus, following the HRRRE, 6-h GDAS-EnKF forecast perturbations were inflated while initializing all partial cycling EnKFs (Table 3), with inflation factors linearly increasing from 1.0 (no inflation) at model level 26 to 2.0 at the lowest model level (Fig. 3). Although HRRRE developers found these tunings improved HRRRE forecast spread–error statistics compared to applying no inflation, these specific inflation factors may not be optimal for our model and DA settings (section 2). However, the ideal external ensemble spread for partial cycling EnKFs that eventually initialize CAE forecasts cannot be determined a priori, and finding this optimum is a potentially expensive tuning exercise that is well outside our scope and reflects one of the many challenges of working with two modeling systems in partial cycling EnKFs.

3) INITIAL LAND SURFACE STATES

There are multiple options for initializing land surface states (e.g., soil temperature and moisture) in partial cycling DA systems. For example, operational partial cycling DA systems continuously cycle land surface states and only ingest atmospheric fields from external models (e.g., Rogers et al. 2009; Hu et al. 2017; Wu et al. 2017). Conversely, some research studies used external models to initialize their partial cycling systems' land surface states (e.g., Hsiao et al. 2012; Johnson et al. 2015, 2020; Duda et al. 2019).

An additional complexity for ensemble-based partial cycling is initial land surface state spread. As 6-h GDAS-EnKF forecast perturbations of land surface variables were extremely small (e.g., Gehne et al. 2019), all 80 ensemble members in PC_{12z} were effectively initialized with identical GFS analysis land surface states at 1200 UTC. We believe this approach is satisfactory, as we surmised that top-level soil states would quickly adjust to diverse atmospheric forcings during self-contained DA cycles and expected initial atmospheric fields to impact EnKF analyses and subsequent forecasts more than initial land surface states. However, to both ensure that this method did not needlessly harm PC_{12z} and test our hypotheses, an additional experiment was performed. This new experiment was identical to PC_{12z}, except initial land surface states for the 80 members were taken from continuously cycling EnKF (i.e., CC_{EnKF}) members' land surface states at 1200 UTC each day, meaning diverse initial land surface states reflecting the continuously cycling EnKF's land surface climate ("PC_{12z_soil}"; Table 3). As described in the appendix, although PC_{12z} and PC_{12z_soil} had different 0000 UTC soil moisture characteristics, aggregate precipitation forecast skill was insensitive to land surface state initialization in the partial cycling EnKFs.

4) INITIAL MICROPHYSICS STATES

Like land surface states, microphysics initialization also requires consideration in partial cycling DA systems. During our experimental period (April–May 2017), NCEP's GDAS employed the Zhao and Carr (1997) microphysics scheme, which only produces total cloud ice and cloud water and is incompatible with the Thompson et al. (2008) microphysics scheme (Table 1) that predicts five liquid and ice species. Thus, using GFS and GDAS-EnKF fields to initialize microphysics variables in the partial cycling EnKFs was not possible, and rather than borrowing microphysics states from the continuously cycling EnKF (analogously to how PC_{12z_soil} borrowed land surface states from CC_{EnKF}), we simply set initial microphysics variables to zero in all ensemble members and expected

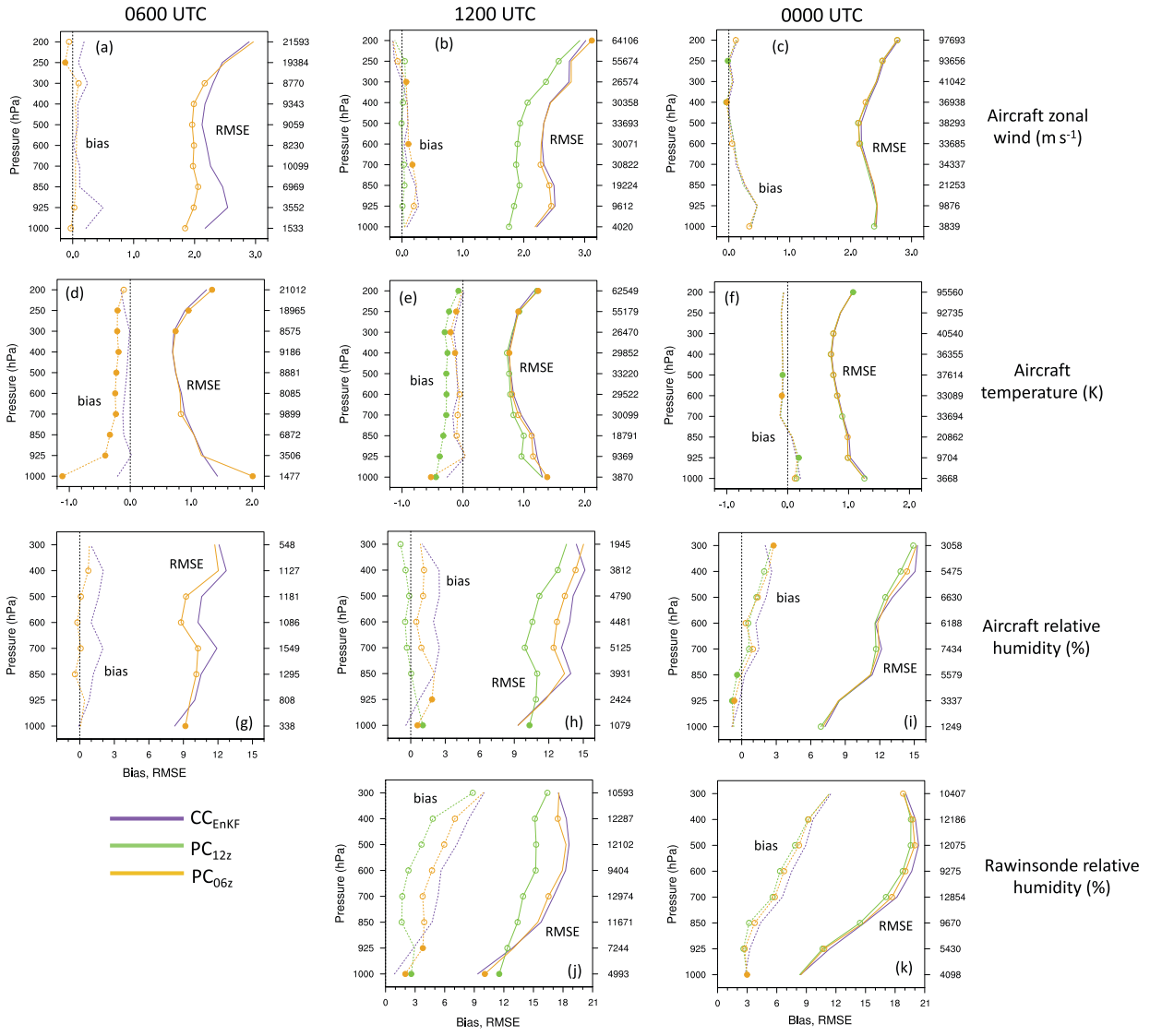

FIG. 5. Ensemble mean additive bias (model minus observations; short-dashed lines) and RMSE (solid lines) compared to (a)–(c) aircraft zonal wind (m s^{-1}), (d)–(f) aircraft temperature (K), (g)–(i) aircraft relative humidity (%), and (j),(k) rawinsonde relative humidity (%) observations aggregated over all prior ensembles valid at (left) 0600, (center) 1200, and (right) 0000 UTC between 0600 UTC 24 Apr and 0000 UTC 20 May 2017 (inclusive). The priors were 1-h forecasts except for PC_{06z} and PC_{12z} at 0600 and 1200 UTC, respectively, where prior ensemble mean statistics quantified GFS analysis fits to observations. Sample size at each pressure level is shown at the right of each panel. Vertical lines at $x = 0$ are references for biases. Circles on the PC_{12z} and PC_{06z} curves denote instances where differences between CC_{EnKF} and PC_{12z} and differences between CC_{EnKF} and PC_{06z} were statistically significant at the 95% level; open circles indicate PC_{12z} or PC_{06z} had statistically significantly better scores than CC_{EnKF}, while filled circles indicate CC_{EnKF} had statistically significantly better scores. Absence of a circle means differences were not statistically significant at the 95% level. Note that x-axis values differ in each row.

microphysics fields to rapidly adjust to dynamic and thermodynamic states during the self-contained cycling period.

d. *Blended ICs*

An alternative to partial cycling for introducing external (i.e., global) fields into limited-area DA systems is a "blending" approach, where large scales from a global model are combined with small scales from a limited-area analysis, which can improve subsequent limited-area forecasts (e.g., Yang 2005; H. Wang et al. 2014; Y. Wang et al. 2014; Hsiao et al. 2015; Zhang et al. 2015; Keresturi et al. 2019; Feng et al. 2020; S21). Our blending methodology was thoroughly detailed in section 2d of S21, so only a short description follows.

Specifically, new 10-member IC ensembles were created daily at 0000 UTC between 25 April and 20 May 2017 (inclusive) by blending small scales from members 1–10 of continuously cycling

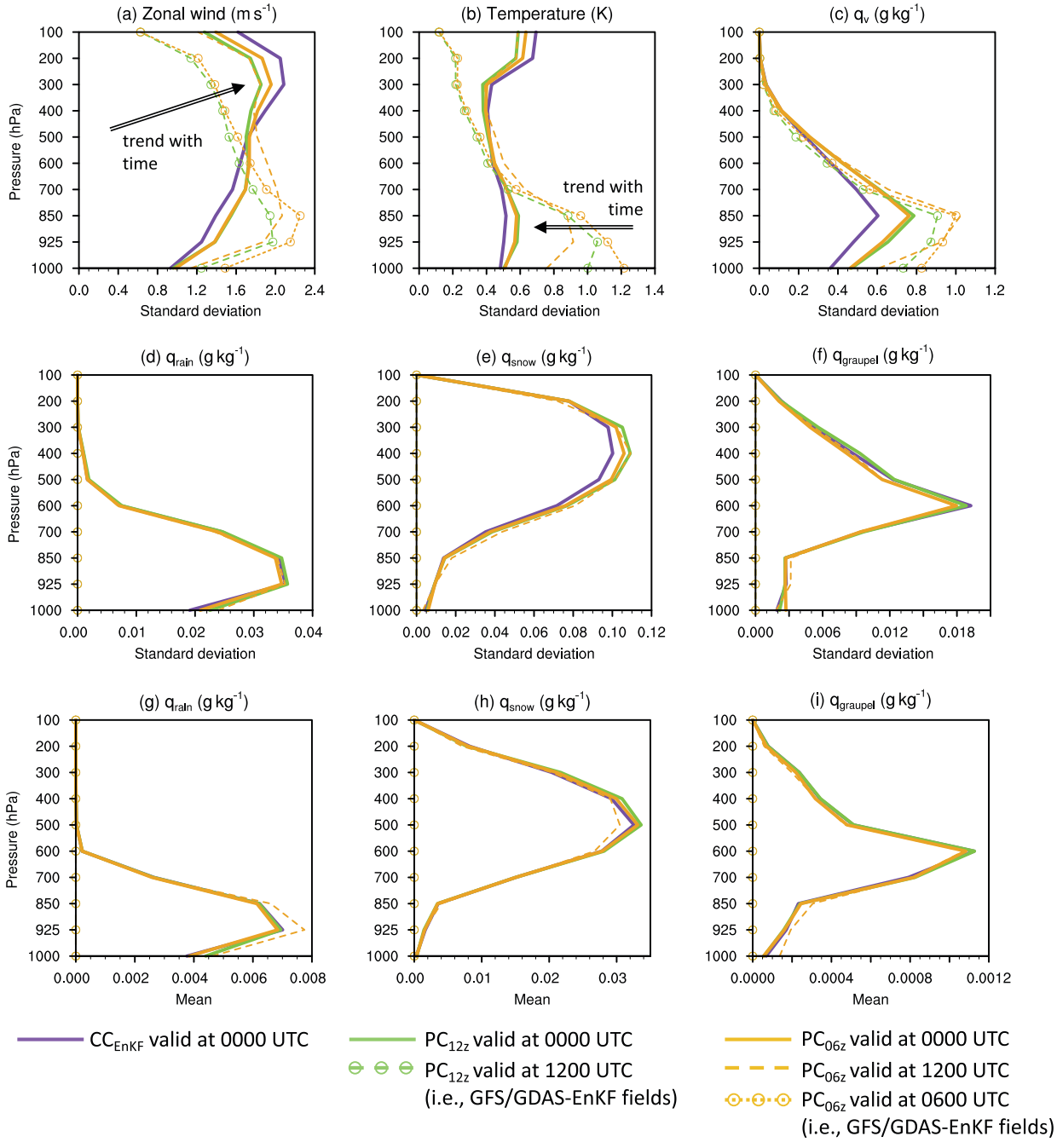

FIG. 6. Average standard deviation over land points within the portion of the 15-km domain collocated with the 3-km domain (Fig. 1) and all posterior ensembles between 0600 UTC 24 Apr and 0000 UTC 20 May 2017 (inclusive) at 0000 UTC (solid lines), 1200 UTC (long-dashed lines), and 0600 UTC (short-dashed lines) for (a) zonal wind (m s^{-1}), (b) temperature (K), (c) water vapor mixing ratio (q_v; g kg^{-1}), (d) rain mixing ratio (q_{rain}; g kg^{-1}), (e) snow mixing ratio (q_{snow}; g kg^{-1}), and (f) graupel mixing ratio ($q_{graupel}$; g kg^{-1}). Open circles denote those curves representing GFS/GDAS-EnKF statistics (i.e., PC_{06z} and PC_{12z} at 0600 and 1200 UTC, respectively). Annotations in (a) and (b) indicate how partial cycling EnKF statistics changed with time. (g)–(i) As in (d)–(f), but for domain-average means. In (d)–(i), open circles at $x = 0$ reflect how the partial cycling EnKFs had no hydrometeors at initialization. Note that x-axis values are different in each panel.

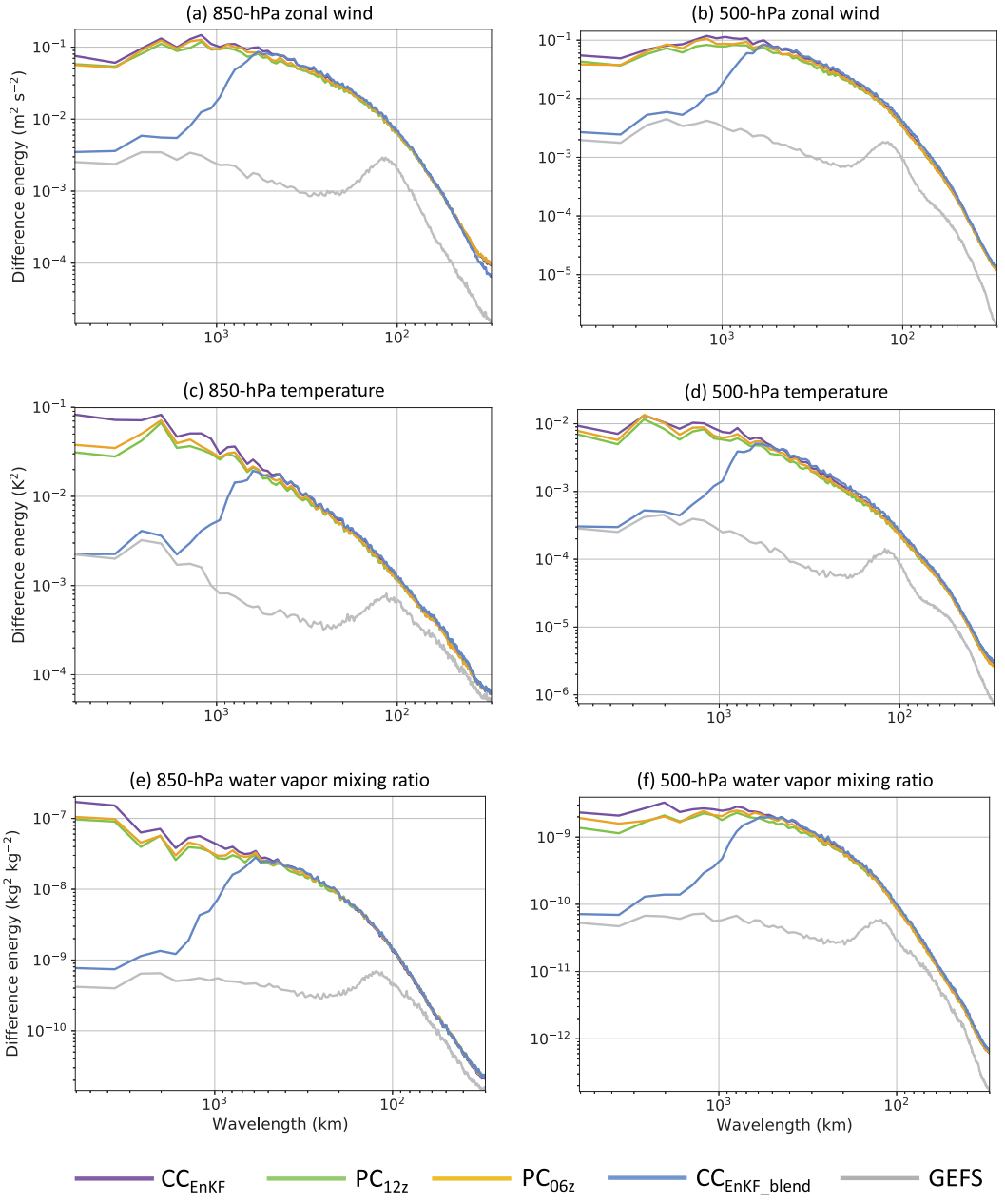

FIG. 7. Spectra for differences between GFS analyses and ensemble mean initial states for various experiments (see legend) as a function of wavelength (km) for (a) 850-hPa zonal wind (m^2 s^{-2}), (b) 500-hPa zonal wind (m^2 s^{-2}), (c) 850-hPa temperature (K^2), (d) 500-hPa temperature (K^2), (e) 850-hPa water vapor mixing ratio (kg^2 kg^{-2}), and (f) 500-hPa water vapor mixing ratio (kg^2 kg^{-2}), averaged over all 0000 UTC analyses between 25 Apr and 20 May 2017 (inclusive). The spectra were computed over the entire 15-km domain, excluding the 10 grid points nearest each lateral boundary, using the discrete cosine transform, and spectral variance binning employed the method of Ricard et al. (2013). Note that *y*-axis values are different in each panel.

EnKF analysis ensembles with large scales from corresponding members 1–10 of 0.5° GEFS IC ensembles using a low-pass, sixth-order implicit tangent filter (e.g., Raymond 1988; Raymond and Garder 1991), similar to several studies (e.g., Yang 2005; H. Wang et al. 2014; Hsiao et al. 2015; Feng et al. 2020; S21). This filter requires a specified cutoff, which, within the context of this work, represents the spatial scale (wavelength) where blended ICs had equal contributions from GEFS and continuously cycling EnKF initial states. S21 noted that ICs produced by blending GEFS ICs and 3-km EnKF analyses with a 960-km cutoff yielded slightly better CAE forecasts compared to using 640- and 1280-km cutoffs. Thus, we used a 960-km cutoff (Fig. 4).

The 0000 UTC blended states initialized 36-h, 10-member CAE forecasts like unblended EnKF analysis ensembles ("CC_{EnKF_blend}"; Table 3). Blending did not impact the continuously cycling EnKF itself, as blended ICs were solely used for purposes of CAE initialization and not incorporated into EnKF DA cycles. However, Feng et al. (2021) found that incorporating blending into 3DVAR DA cycles improved deterministic forecasts over China, and future work may assess whether integrating blending within continuous EnKF DA cycles is beneficial for CAE forecast initialization over the CONUS.

e. Benchmark ensemble

Finally, as in S21, 36-h CAE forecasts were initialized by interpolating 0.5° ICs from members 1–10 of NCEP's operational GEFS (Zhou et al. 2017) onto the nested computational domain (Fig. 1) daily at 0000 UTC between 25 April and 20 May 2017 (inclusive). These GEFS-initialized CAE forecasts (Table 3) used identical WRF Model configurations and LBCs as the EnKF-initialized CAE forecasts and served as a benchmark to assess whether experimental limited-area EnKF analyses could initialize better CAE forecasts than operational ICs. Unlike the other ICs that had nonzero hydrometeor fields at 0000 UTC, GEFS-initialized forecasts began without hydrometeors, so a long spinup was expected.

4. Partial and continuously cycling EnKF characteristics

While the continuously cycling EnKF (i.e., CC_{EnKF}) required two days to spin up from random noise, it was unclear how quickly the partial cycling EnKFs would move away from their flow-dependent GFS/GDAS-EnKF initial states and adjust to the WRF Model climate. Because 0000 UTC analyses initialized CAE forecasts, we wanted to understand properties of 0000 UTC partial cycling EnKF states, in particular, whether they resembled 0000 UTC CC_{EnKF} states or retained characteristics of their prescribed initial GFS/GDAS-EnKF states from 12 or 18 h earlier.

Thus, the following analyses were performed to elucidate the composition of 0000 UTC partial cycling EnKF states and their similarities with 0000 UTC CC_{EnKF} states. These analyses are also offered as evidence that partial cycling EnKF performance was acceptable given several subjective configuration choices (section 3c). As partial cycling EnKF spinup can largely be controlled through DA parameters like observation errors, the following statistics were purely diagnostic, and ultimately, we hoped to relate differences between various sets of 0000 UTC-initialized CAE forecasts to differences between their ICs.

a. Observation-space diagnostics

Prior ensemble mean additive biases (model minus observations) and root-mean-square errors (RMSEs) were computed with respect to rawinsonde and aircraft observations, the latter of which have particularly important influences in hourly updated DA systems over the CONUS (James and Benjamin 2017; James et al. 2020). Observation-space ensemble spreads were also assessed but are not presented, as state-space spreads yielded identical conclusions (section 4b). Given the partial cycling initialization procedure (sections 3b and 3c), PC_{06z} and PC_{12z} prior ensemble mean statistics at 0600 and 1200 UTC, respectively, quantified GFS analysis fits to observations, whereas PC_{06z} and PC_{12z} prior ensemble mean statistics at later hours (during self-contained cycling) measured how the partial cycling EnKFs were adjusting toward the WRF Model climate. Statistical significance of aggregate statistics at the 95% confidence level was assessed with a bootstrap resampling approach using 10 000 resamples (with replacement) applied to pairwise differences between two experiments (e.g., Hamill 1999; Wolff et al. 2014).

Compared to continuously cycling EnKF (i.e., CC_{EnKF}) prior ensemble means, GFS analyses more closely fit zonal wind and relative humidity (RH) observations[3] and were drier at most levels (cf. orange and purple lines in Figs. 5a,g and green and purple lines in Figs. 5b,h,j). Additionally, GFS analyses had significantly smaller 925–400-hPa RMSEs compared to temperature observations than CC_{EnKF} prior ensemble means at 1200 UTC (Fig. 5e), but not at 0600 UTC (Fig. 5d). However, at both 0600 and 1200 UTC, GFS analyses had significant cold biases (Figs. 5d,e), possibly due to GFS physics errors (e.g., Zheng et al. 2017). GFS analyses also had cold biases compared to rawinsonde observations between 925 and 300 hPa (not shown).

As self-contained cycles progressed, prior ensemble mean biases and RMSEs in the partial cycling EnKFs generally became more similar to those of CC_{EnKF} at most levels, suggesting the partial cycling EnKFs were behaving properly. For example, differences of zonal wind RMSEs and temperature biases between CC_{EnKF} and PC_{06z} decreased going from 0600 to 1200 to 0000 UTC (Figs. 5a–f), indicating PC_{06z} was moving away from GFS analyses. However, small, but often statistically significant, differences between the partial and continuously cycling EnKFs remained at 0000 UTC regarding temperature and zonal wind RMSEs (~0.01 K and ~0.01–0.05 m s^{-1} differences), which were lower in the partial cycling EnKFs.

Compared to zonal wind and temperature, RH adjustments appeared smaller, especially according to biases, which indicated 0000 UTC partial cycling EnKF prior ensemble means were regularly statistically significantly drier than CC_{EnKF} (Figs. 5i,k). This finding suggests that moisture fields had not fully moved away from GFS/GDAS-EnKF states assigned at partial cycling EnKF initialization even after 18 h of self-contained cycles.

b. State-space characteristics

State-space characteristics were also assessed to explore partial cycling EnKF evolution. Regarding ensemble spread,

[3] Evaluating continuously cycling EnKF posterior ensemble means lessens these differences. However, because GFS analyses indeed served as prior ensemble means for partial cycling EnKF initialization, comparing GFS analyses to continuously cycling EnKF prior ensemble means is the relevant comparison.

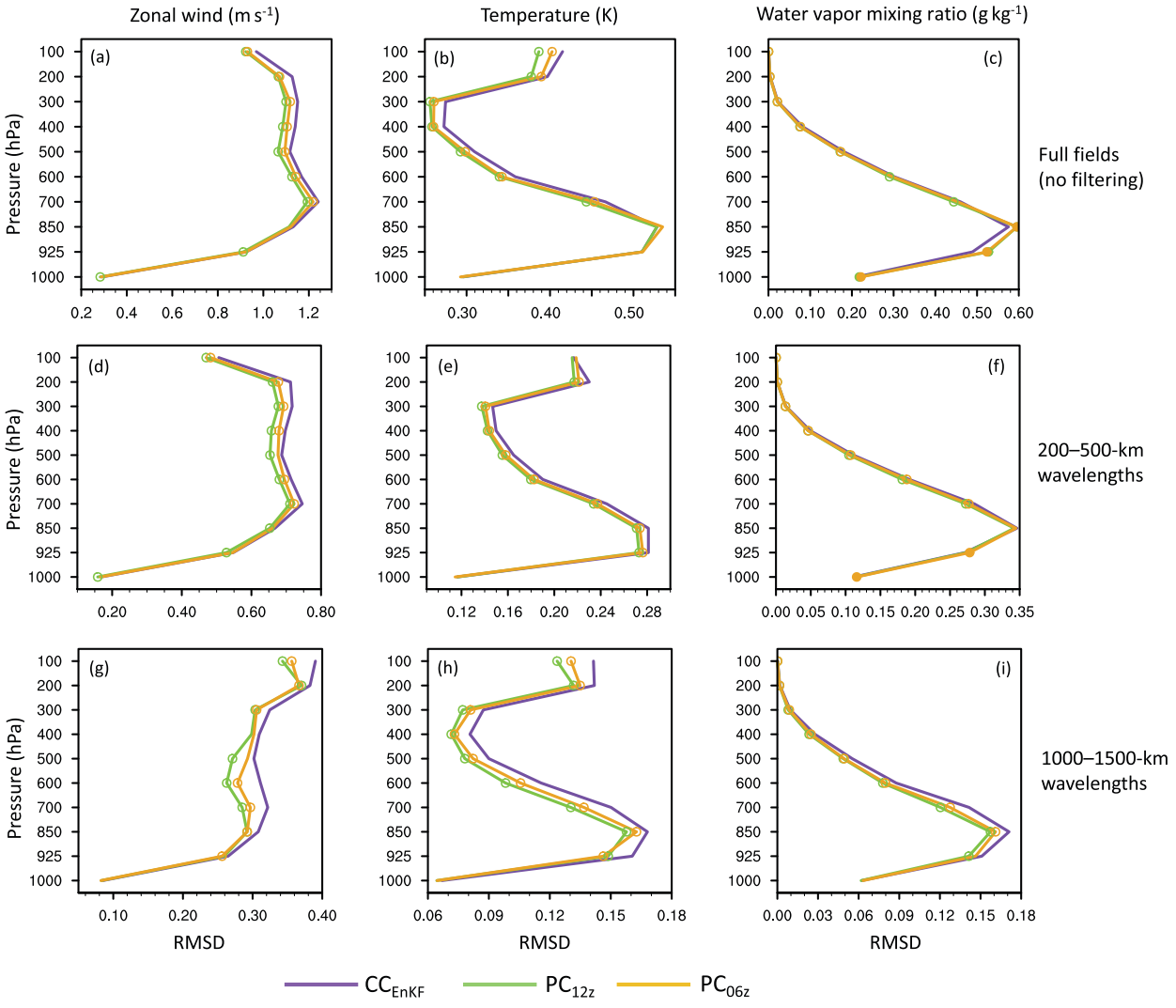

FIG. 8. Aggregate RMSDs between GFS and EnKF mean analyses [Eq. (1)] for (left) zonal wind (m s^{-1}), (center) temperature (K), and (right) water vapor mixing ratio (g kg^{-1}) over all 0000 UTC analyses between 25 Apr and 20 May 2017 (inclusive) for (a)–(c) full fields, (d)–(f) bandpass filtered fields for 200–500-km wavelengths, and (g)–(i) bandpass filtered fields for 1000–1500-km wavelengths. These statistics were computed over land points within the portion of the 15-km domain collocated with the 3-km domain. Statistically significant differences between CC_{EnKF} and PC_{12z} and between CC_{EnKF} and PC_{06z} at the 95% level are denoted as in Fig. 5. Note that *x*-axis values are different in each panel.

inflated 6-h GDAS-EnKF forecasts (i.e., PC_{06z} and PC_{12z} at 0600 and 1200 UTC, respectively) had lower domain-average standard deviations[4] than CC_{EnKF} for zonal wind, temperature, and water vapor mixing ratio above 600 hPa and higher standard deviations below (cf. purple and circular-marked dashed curves in Figs. 6a–c), reflecting initial variance inflation (Fig. 3). While patterns were similar at 0000 UTC, differences between the partial and continuously cycling EnKFs were smaller than at 0600 and 1200 UTC, indicating the partial cycling EnKFs had moved away from their inflated GFS/GDAS-EnKF initial states.

However, some noteworthy 0000 UTC differences remained. For example, above 500 hPa, while PC_{06z} had closer temperature and zonal wind spreads to CC_{EnKF} than PC_{12z} due to greater adjustment afforded by an extra 6 h of self-contained cycles (cf. solid lines in Figs. 6a,b), even 18 h of self-contained cycles was not enough for PC_{06z} spread to match the larger CC_{EnKF} spread, suggesting spinup was not complete. In addition, the partial cycling EnKFs had more moisture spread than CC_{EnKF} below 500 hPa (cf. solid lines in Fig. 6c). It is possible that moisture spread did not adjust as much as temperature and zonal wind spreads below 500 hPa in the partial cycling EnKFs

[4] The continuously cycling EnKF had a stable climate with only small diurnal spread variations, primarily in the planetary boundary layer. Thus, to foster readability, CC_{EnKF} domain-average spread is only shown at 0000 UTC, as its spread was approximately the same at 0600 and 1200 UTC.

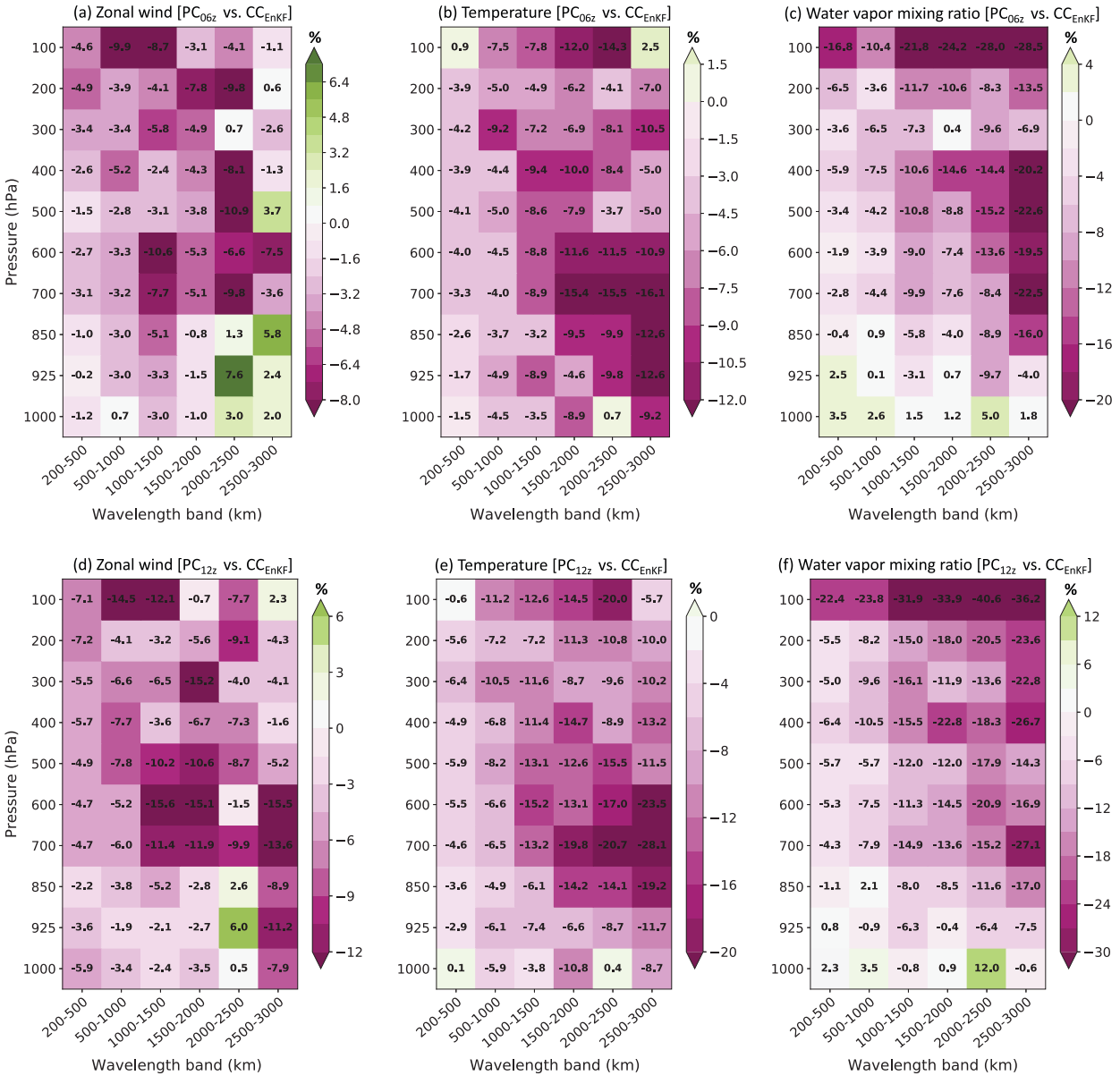

FIG. 9. Normalized RMSD reductions [%; Eq. (2)] between CC_{EnKF} and PC_{06z} for (a) zonal wind, (b) temperature, and (c) water vapor mixing ratio for various wavelength bands (km) and pressure levels (hPa) aggregated over all 0000 UTC analyses between 25 Apr and 20 May 2017 (inclusive). (d)–(f) As in (a)–(c), but for normalized RMSD reductions between CC_{EnKF} and PC_{12z}. Negative values indicate RMSDs with respect to GFS analyses [Eq. (1)] were smaller in PC_{06z} and PC_{12z} compared to CC_{EnKF}. Color bars and their ranges are different in each panel. These statistics were computed over land points within the portion of the 15-km domain collocated with the 3-km domain.

due to the relative scarcity of moisture observations to directly constrain EnKF analyses (see sample sizes on Fig. 5), although specific DA settings may also have played a role.

Despite initializing the partial cycling EnKFs without hydrometeors, 0000 UTC domain-average standard deviations (Figs. 6d–f) and means (Figs. 6g–i) of rain, snow, and graupel mixing ratios in PC_{06z} and PC_{12z} were comparable to or greater than those in the continuously cycling EnKF. These findings confirm that microphysics variables quickly respond to atmospheric states given our configurations and suggest that initializing partial cycling EnKFs without hydrometeors may be acceptable.

c. *Spectral analysis*

To examine how EnKF analyses represented different spatial scales, power spectra were computed using the discrete cosine transform (DCT; Denis et al. 2002). Power spectra of EnKF analysis perturbations reflected conclusions from Fig. 6 (there was typically more 0000 UTC perturbation power below 500 hPa and less above in the partial cycling EnKFs compared to

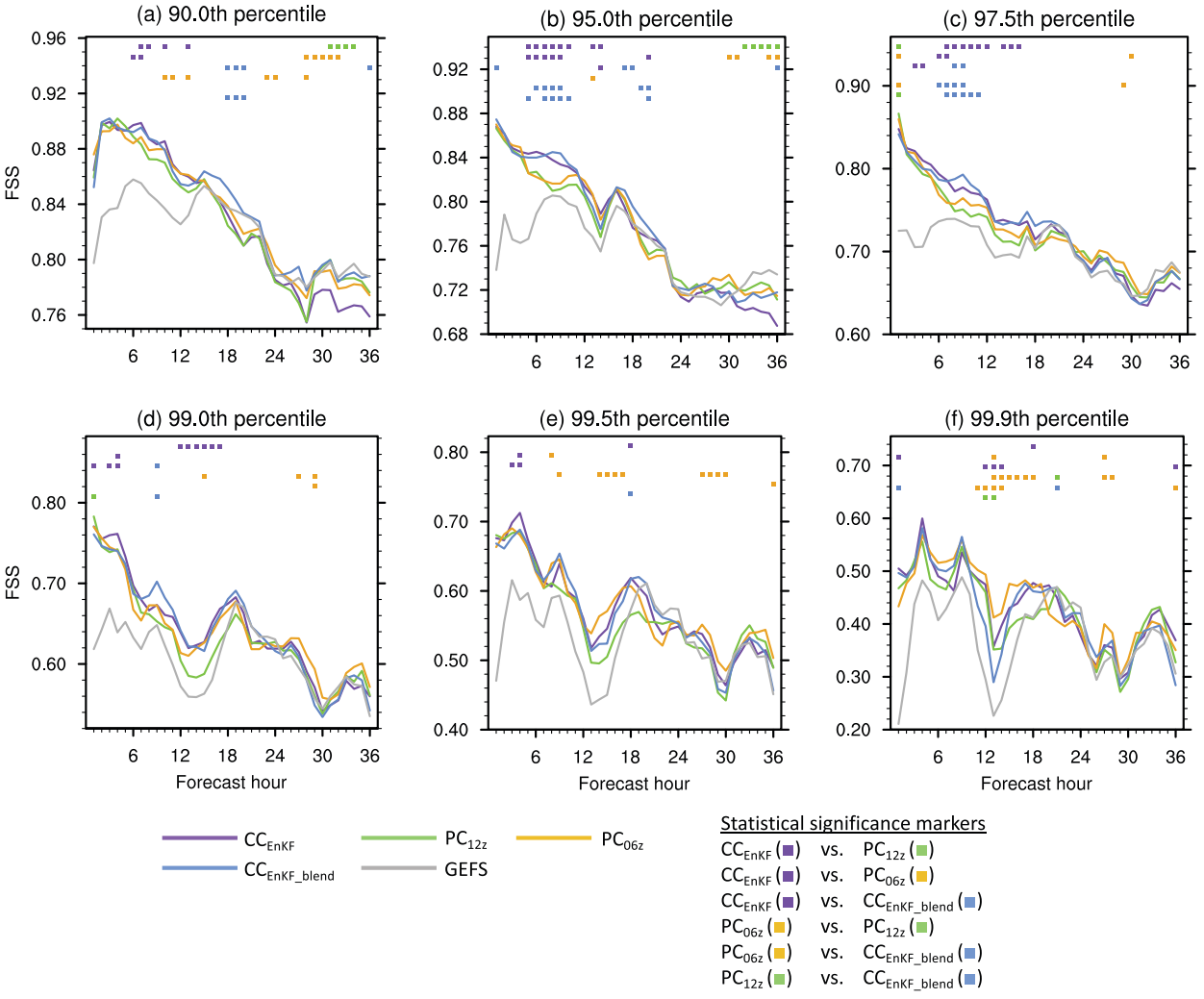

FIG. 10. FSSs over the CONUS east of 105°W (Fig. 1) with a 100-km neighborhood length scale for the (a) 90th, (b) 95th, (c) 97.5th, (d) 99th, (e) 99.5th, and (f) 99.9th percentile thresholds aggregated over all 26 3-km forecasts of 1-h accumulated precipitation as a function of forecast hour. Values 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. Symbols along the top axis denote instances where differences between two ensembles were statistically significant at the 95% level, with the six rows of colored symbols in each panel corresponding to the six comparisons in the legend (from top to bottom) to denote which ensemble had statistically significantly higher FSSs. For example, the top row of symbols in each panel compares CC_{EnKF} and PC_{12z}; purple symbols indicate CC_{EnKF} had statistically significantly higher FSSs than PC_{12z}, while green symbols indicate PC_{12z} had statistically significantly higher FSSs than CC_{EnKF} (see Table 3 for descriptions of the experiments). As another example, the bottom row of symbols in each panel compares PC_{12z} and CC_{EnKF_blend}; green symbols indicate PC_{12z} had statistically significantly higher FSSs than CC_{EnKF_blend}, while blue symbols indicate CC_{EnKF_blend} had statistically significantly higher FSSs than PC_{12z}. Absence of a symbol means the differences were not statistically significant at the 95% level.

CC_{EnKF}) and are not further discussed. Instead, we focus on understanding how 0000 UTC partial and continuously cycling EnKF mean analyses compared with global analyses across a range of scales, as mean IC states exert a strong influence on CAE forecast skill (e.g., Schwartz et al. 2020).

Specifically, power spectra of *differences* between EnKF mean and GFS analyses were computed, which indicated 0000 UTC partial cycling EnKF mean analyses more closely resembled GFS analyses than continuously cycling EnKF mean analyses for wavelengths > 200 km (Fig. 7). Moreover, the gap between difference spectra of the partial and continuously cycling EnKFs typically widened as wavelength increased, especially for temperature and moisture (Figs. 7c–f), suggesting the partial and continuously cycling EnKFs differed more at larger scales than smaller ones. For most scales > 1000 km, differences between the two partial cycling EnKFs were smaller than their collective differences with respect to CC_{EnKF}. Differences between GFS analyses and GEFS mean initial states reflected the link between the GFS and GEFS (Zhou et al. 2017) and were at least an order of magnitude smaller than limited-area EnKF difference spectra for scales > 200 km, and difference spectra of mean blended states affirmed the blending procedure.

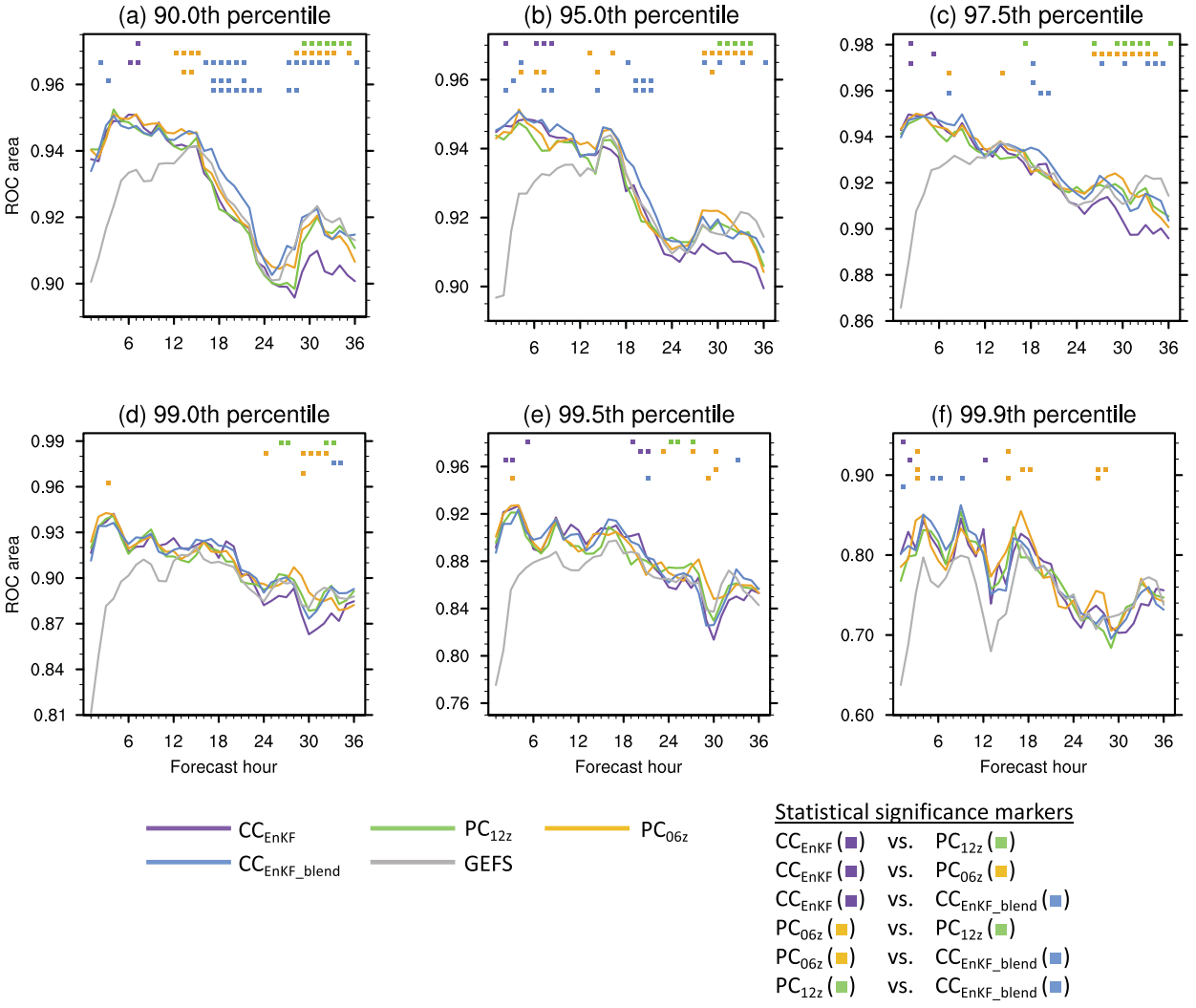

FIG. 11. As in Fig. 10, but for areas under the ROC curve computed using decision thresholds of 1%, 2%, 3%, 4%, 5%, 10%, 15%, …, 95%, and 100% and a trapezoidal method. Symbols along the top axis indicate forecast hours when differences between two ensembles were statistically significant at the 95% level as in Fig. 10 and denote the ensemble with statistically significantly higher ROC areas. The *y*-axis scales are different in each panel.

To further explore spectral differences, 0000 UTC GFS and EnKF mean analyses were filtered within various wavelength bands using the DCT and its inverse (e.g., Denis et al. 2002). These bandpass filtered fields were then directly compared to calculate root-mean-square differences (RMSDs) between GFS and EnKF mean analyses as a function of spatial scale using

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(\mathrm{GFS}_k - \overline{\mathrm{EnKF}}_k\right)^2}, \qquad (1)$$

where for the kth of N points, GFS_k is the GFS analysis and $\overline{\mathrm{EnKF}}_k$ the EnKF mean analysis for a particular experiment (e.g., Table 3). Additionally, normalized reductions of RMSDs between two experiments (D) were computed as

$$D = \frac{\mathrm{RMSD}_i - \mathrm{RMSD}_j}{\mathrm{RMSD}_j} \times 100\%, \qquad (2)$$

where RMSD_i and RMSD_j are RMSDs of the ith and jth experiments, respectively (Table 3). The term D is interpreted as "relative to experiment j, experiment i had a D% smaller or larger RMSD," where $D < 0$ indicates experiment i had the smaller RMSD (i.e., $\mathrm{RMSD}_i < \mathrm{RMSD}_j$) and was more similar to GFS analyses than experiment j.

Corroborating Fig. 7 and generally consistent with Figs. 5c,f,i,k, 0000 UTC partial cycling EnKF mean analyses usually had statistically significantly smaller aggregate RMSDs than continuously cycling EnKF mean analyses for full fields (no filtering; Figs. 8a–c) and in the 200–500- and 1000–1500-km wavelength bands (Figs. 8d–i). RMSDs were smaller in the 1000–1500-km band than the 200–500-km band, indicating

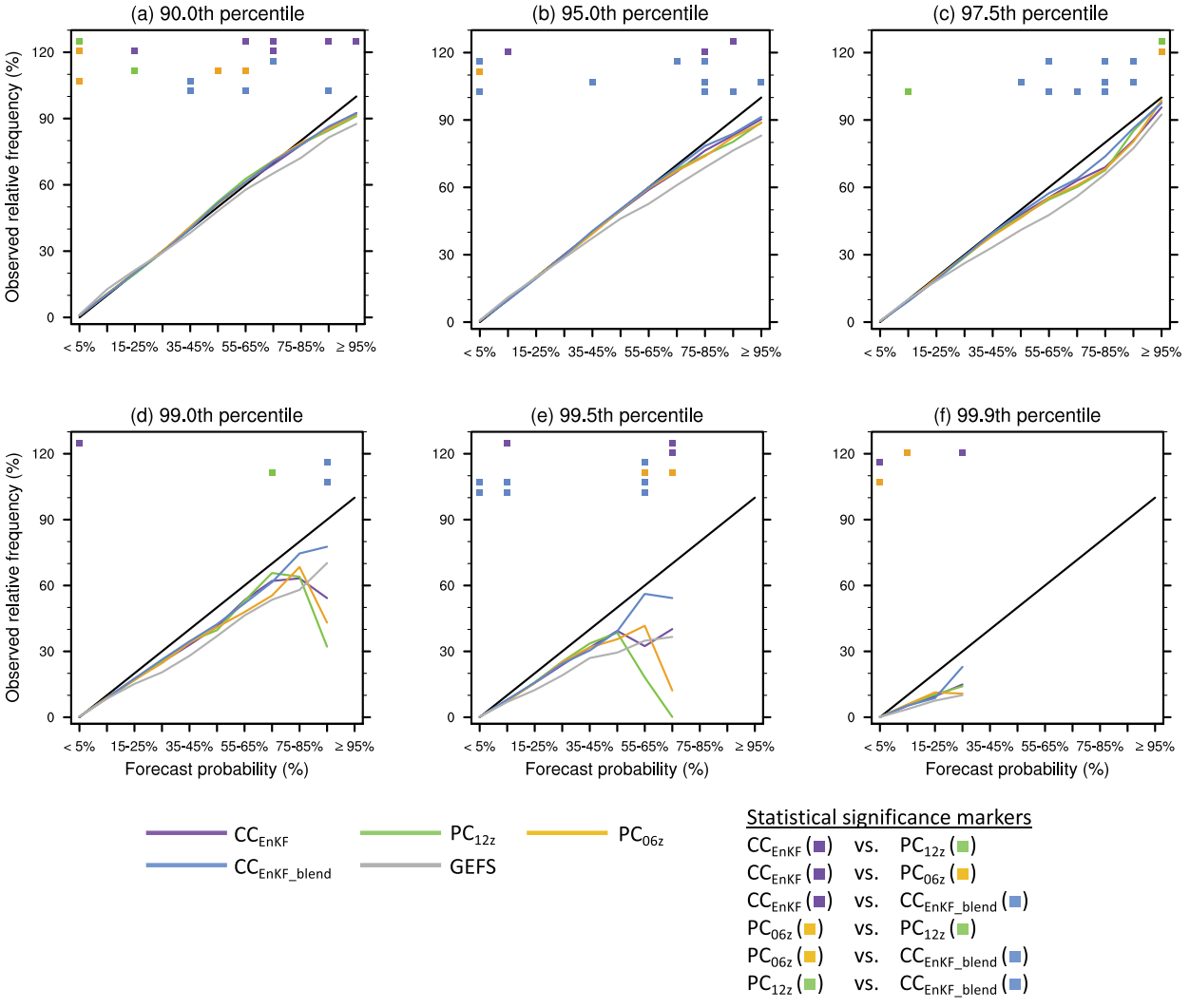

FIG. 12. Reliability diagrams computed over the CONUS east of 105°W (Fig. 1) with a 100-km neighborhood length scale aggregated over all 26 1–12-h 3-km forecasts of 1-h accumulated precipitation for the (a) 90th, (b) 95th, (c) 97.5th, (d) 99th, (e) 99.5th, and (f) 99.9th percentile thresholds. Diagonal lines are lines of perfect reliability. 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. Symbols along the top axis indicate probability bins where differences between two ensembles were statistically significant at the 95% level as in Fig. 10 and denote the ensemble with statistically significantly better reliability as determined by block bootstrapping. Note that the reliability diagrams themselves stop at 100%; area above 100% was added to make room for statistical significance markers.

EnKF mean and GFS analyses were more alike on larger, more predic table scales. However, when letting $RMSD_i$ and $RMSD_j$ represent aggregate 0000 UTC RMSDs of PC_{06z} and CC_{EnKF}, respectively, D [Eq. (2)] typically became more negative as spatial scale increased (Figs. 9a–c), especially for temperature and moisture, for which PC_{06z} had ~10%–20% smaller RMSDs than CC_{EnKF} for many wavelengths ≥ 1000 km but RMSD reductions of typically only ~5% or less for scales < 1000 km (Figs. 9b,c). Therefore, relative differences between PC_{06z} and CC_{EnKF} generally grew as wavelength increased.

Collectively, Figs. 7–9 indicated PC_{06z} and CC_{EnKF} mean analyses differed most at larger scales, where PC_{06z} mean analyses were closer to GFS analyses than CC_{EnKF} mean analyses. Thus, even after 18 h of self-contained cycles, PC_{06z} had "memory" of its most recent injection of GFS/GDAS-EnKF fields, especially at large scales; PC_{12z} unsurprisingly had an even stronger memory of and was more similar to GFS/GDAS-EnKF fields at 0000 UTC than PC_{06z} (e.g., Figs. 7, 8; also compare Figs. 9a–c and Figs. 9d–f). The next section shows how the large-scale differences between partial and continuously cycling EnKF ICs impacted subsequent CAE precipitation forecasts.

5. Precipitation forecast verification

a. Methods

Our precipitation verification methods were the same as in S21, who in turn followed section 5a of Schwartz (2019), so descriptions here are brief. Specifically, hourly accumulated

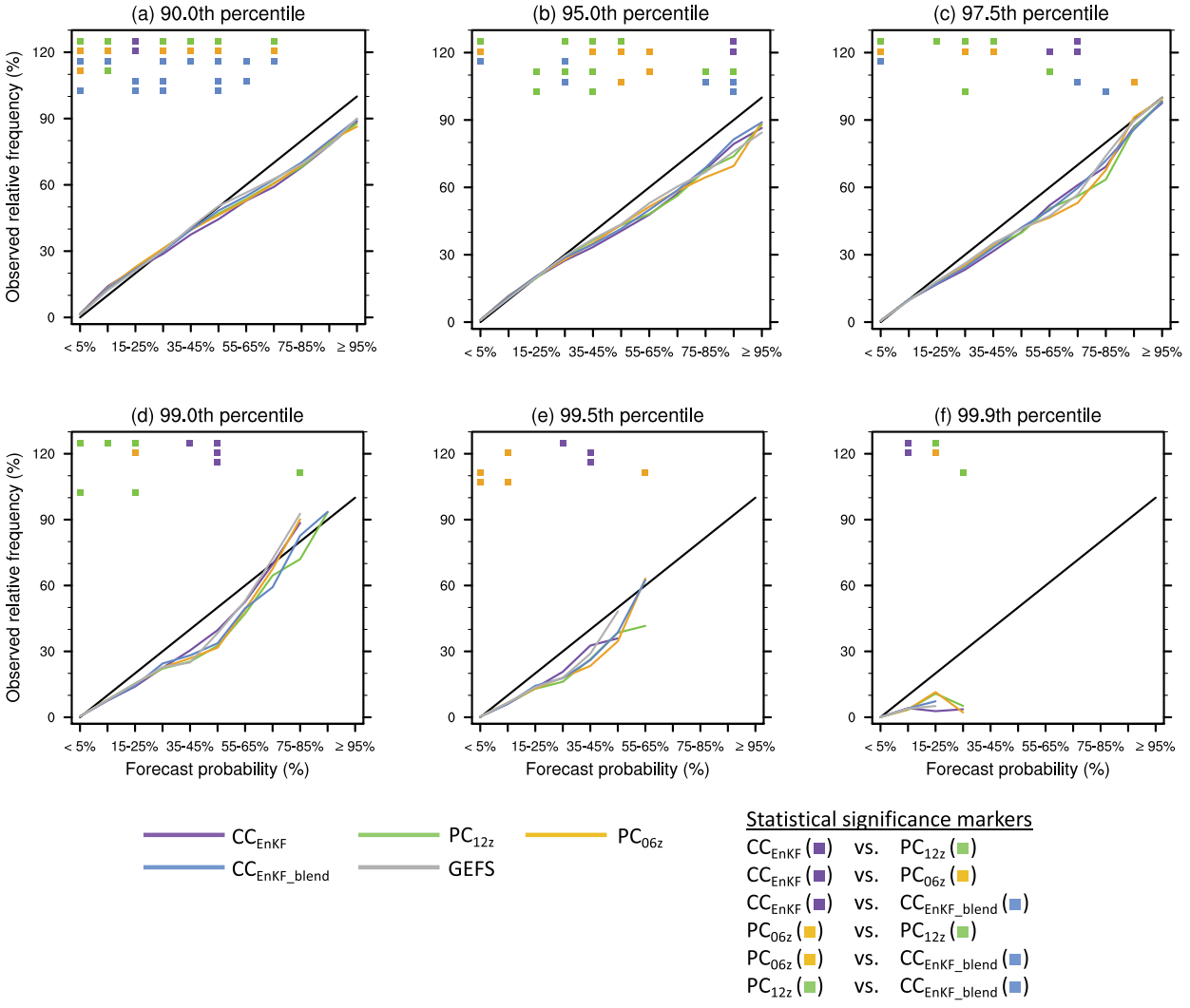

FIG. 13. As in Fig. 12, but statistics were aggregated over all 26 24–36-h 3-km forecasts of 1-h accumulated precipitation.

precipitation forecasts were objectively compared to NCEP's Stage IV (ST4) analyses (Lin and Mitchell 2005) over the CONUS east of 105°W (Fig. 1), where ST4 analyses were most robust (e.g., Nelson et al. 2016). Some metrics were computed from native 3-km output, while a budget algorithm (e.g., Accadia et al. 2003) was used to interpolate precipitation forecasts to the ~4.763-km ST4 grid to compute metrics requiring a common grid for forecasts and observations.

As in S21, event occurrence was determined using percentile thresholds (e.g., the 95th percentile, which selects the top 5% of events). This approach defines the same number of forecast and observed events, thus, removing bias and permitting a thorough assessment of spatial performance given a model's climate (e.g., Roberts and Lean 2008; Mittermaier and Roberts 2010; Dey et al. 2014; Woodhams et al. 2018; Schwartz 2019). We used percentiles between 90% and 99.9% to verify both broad precipitation features and localized, intense events.

Additionally, because convection-allowing models are inherently inaccurate at the grid-scale, a "neighborhood approach" (e.g., Theis et al. 2005; Ebert 2008, 2009) was applied to derive "neighborhood ensemble probabilities" (NEPs; Schwartz et al. 2010; Schwartz and Sobash 2017), which are smoothed ensemble probabilities within a designated neighborhood length scale (r) and more appropriate for verifying CAEs than point-based probabilities. Values of r between 5 and 150 km, which represented radii of circular neighborhoods, were used to construct NEPs that were ultimately verified. Pairwise difference bootstrapping was again used to assess statistical significance, and when bootstrap confidence intervals were obtained for statistics aggregated over multiple forecast hours, a circular block bootstrapping method (e.g., Politis and Romano 1992; Wilks 1997; Gilleland et al. 2018) was used with a 4-h block length to preserve autocorrelations.

b. *Results*

1) FRACTIONS SKILL SCORES AND ROC AREAS

To assess spatial skill, fractions skill scores (FSSs; Roberts and Lean 2008) and areas under the relative operating

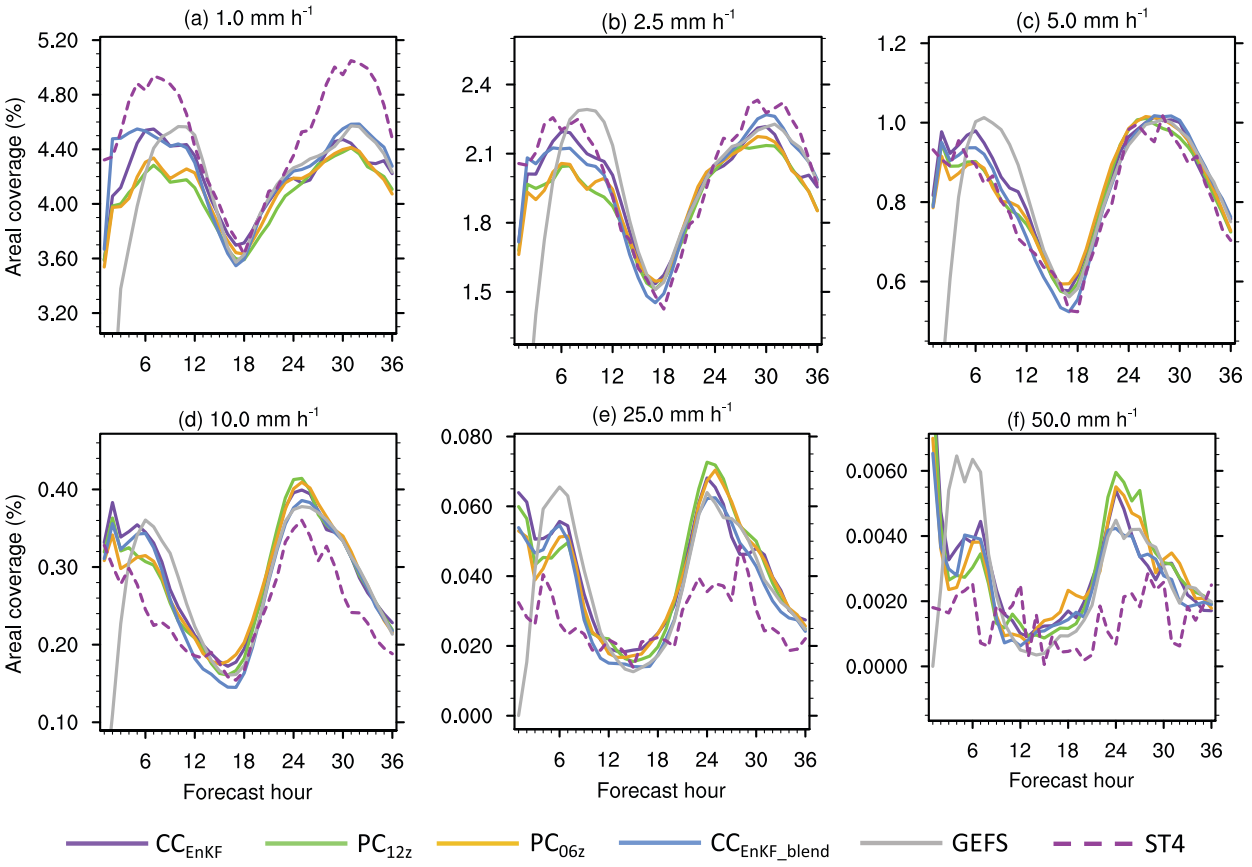

FIG. 14. Fractional areal coverage (%) of 1-h accumulated precipitation meeting or exceeding (a) 1.0, (b) 2.5, (c) 5.0, (d) 10.0, (e) 25.0, and (f) 50.0 mm h^{-1} over the CONUS east of 105°W (Fig. 1), computed on native grids and aggregated over all 26 3-km forecasts as a function of forecast hour. These statistics were computed for all 10 ensemble members, but for readability, only ensemble mean areal coverages are shown. At the earliest forecast hours, mean GEFS areal coverages were nonzero but below the *x* axis for some thresholds. 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.

characteristic (ROC) curve (Mason 1982; Mason and Graham 2002) were calculated. FSSs and ROC areas range between 0 and 1, with higher values indicating more skill. We present FSSs and ROC areas for NEPs computed with $r = 100$ km; overall conclusions were identical when using different neighborhood length scales. Relative differences of FSSs and ROC areas between CAE forecasts with partial and continuously cycling EnKF ICs did not systematically change throughout the experimental period, so we focus on aggregate statistics over all 26 3-km forecasts.

Through ~18 h, GEFS-initialized CAE forecasts were typically worst (Figs. 10, 11) and the ensembles with blended and unblended continuously cycling EnKF ICs (i.e., CC_{EnKF} and CC_{EnKF_blend}) had similar FSSs and ROC areas that were usually comparable to or somewhat higher than those from PC_{06z}, which in turn generally had better scores than PC_{12z}. Although these findings suggest ICs that are more spun up (e.g., Figs. 5, 6) are beneficial for ~1–18-h forecasts, differences between PC_{06z}, PC_{12z}, CC_{EnKF}, and CC_{EnKF_blend} were only occasionally statistically significant. This broad similarity was consistent with the relatively small differences between partial and continuously cycling EnKF ICs at small scales (Figs. 7, 9), which are important for short-term forecast evolution. In sum, FSSs and ROC areas indicated no benefits of partial cycling over continuous cycling for short-term (~1–18-h) CAE forecasts.

Conversely, after ~18 h, when large-scale ICs exert greater forecast impacts, unblended continuously cycling EnKF analyses initialized CAE forecasts that were comparable to or worse than those with GEFS or partial cycling EnKF ICs (Figs. 10, 11). The biggest degradations of CC_{EnKF} relative to ensembles with partial cycling ICs occurred after ~27 h, where some differences were statistically significant (Figs. 10a,b, 11a–d). However, blended ICs yielded next-day (~18–36-h) CAE forecasts that were typically better than those with unblended EnKF ICs and statistically indistinguishable from forecasts with partial cycling EnKF ICs.

These findings indicate limited-area EnKF ICs produced better ~18–36-h forecasts when they had memory of GFS large scales through partial cycling or were explicitly linked to GFS large scales through blending. However, closeness to large-scale GFS analyses alone did not determine next-day forecast quality: for example, blended ICs were much closer

to large-scale GFS analyses than partial cycling EnKF analyses (Fig. 7), yet ~18–36-h forecasts with blended ICs were not systematically better than those with partial cycling EnKF ICs. Thus, although ICs too far from GFS analysis large scales clearly seem to degrade next-day forecasts, controls on ~18–36-h forecast quality appear complex.

2) Reliability statistics

Reliability statistics (e.g., Wilks 2011) computed with r = 100 km aggregated over all 26 3-km forecasts revealed 1–12-h probabilistic precipitation forecasts with blended and unblended continuously cycling EnKF ICs typically had comparable reliabilities to those initialized from partial cycling EnKF analyses (Fig. 12). All EnKF-initialized forecasts were more reliable than GEFS-initialized forecasts over the first 12 h.

For 24–36-h forecasts, while there were a few probability bins where unblended continuously cycling EnKF ICs yielded similar or better reliability compared to partial cycling EnKF ICs (Fig. 13), partial cycling EnKF ICs led to statistically significantly more reliable forecasts than unblended continuously cycling EnKF ICs in many bins, especially for probabilities $< 55\%$ at the 90.0th–97.5th percentiles (Figs. 13a–c). Blended ICs typically provided comparable 24–36-h forecast reliability as ICs from the two partial cycling EnKFs, which usually had similar reliabilities at both forecast ranges (Figs. 12, 13). Differences regarding probabilistic forecast distributions (i.e., sharpness) between the various CAEs were not noteworthy (not shown).

Like FSSs and ROC areas, reliability statistics indicated both that short-term CAE forecasts did not benefit from partial cycling and that ICs with large-scale spectral characteristics with memory of or forced to those of GFS analyses improved 24–36-h forecasts. Reliability statistics computed over just the 24–30- and 30–36-h forecast periods yielded identical conclusions as the 24–36-h aggregate statistics (not shown).

3) Precipitation climatologies

Aggregate areal coverages of 1-h accumulated precipitation meeting or exceeding various thresholds (e.g., 5.0 mm h^{-1}) were calculated to examine precipitation distributions. At all thresholds, ensembles with partial cycling EnKF ICs had lower mean areal coverages than the ensemble with unblended continuously cycling EnKF ICs over the first 12 h (Fig. 14). These lower 1–12-h forecast coverages in PC_{06z} and PC_{12z} were further from observed coverages than CC_{EnKF} for thresholds $\leq$ 2.5 mm h^{-1} (Figs. 14a,b) but closer to observations than CC_{EnKF} at higher thresholds (Figs. 14c–f). Relative to CC_{EnKF}, aside from the first 6 h at the 1.0 mm h^{-1} threshold, blended ICs yielded lower coverages. GEFS-initialized forecasts usually had areal coverages furthest from observations over the first 6 h due to spin up from their coarse (0.5°) ICs.

Commensurate with areal coverages, partial cycling EnKF ICs yielded less domain-total precipitation than blended and unblended continuously cycling EnKF ICs before 12 h, and these lower amounts agreed best with observations (Fig. 15). Lower total precipitation and areal coverages in PC_{06z} and PC_{12z} relative to CC_{EnKF} through 12 h was consistent with drier 0000 UTC PC_{06z} and PC_{12z} states compared to CC_{EnKF} (Figs. 5i,k).

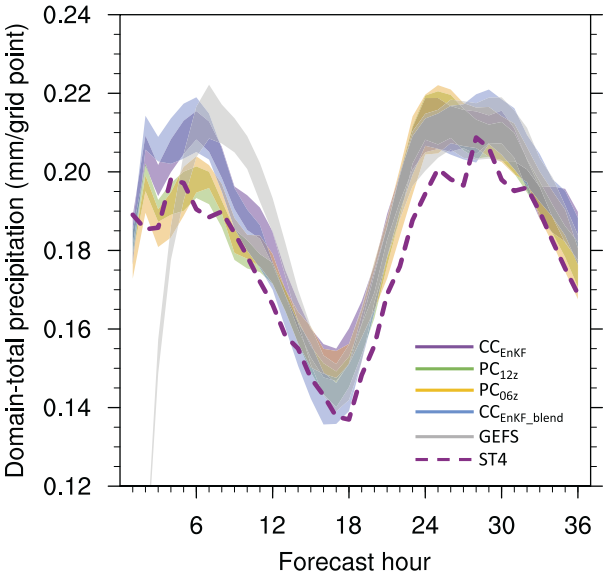

FIG. 15. Average 1-h accumulated precipitation (mm) per grid point over all 26 3-km forecasts and the CONUS east of 105°W (Fig. 1), computed on native grids as a function of forecast hour. Shadings represent envelopes of the 10 members composing the various ensembles indicated in the legend, and darker shadings represent intersections of two or more ensemble envelopes. Values on the x axis represent ending forecast hours of 1-h accumulation periods (e.g., an x-axis value of 24 is for 1-h accumulated precipitation between 23 and 24 h). At the earliest forecast hours, GEFS domain-total precipitation was nonzero but below the x axis.

Differences between the CAEs generally diminished after 12 h, where all ensembles accurately captured timing of the observed diurnal maximum, underpredicted peak coverages for thresholds $\leq$ 2.5 mm h^{-1} (Figs. 14a,b), and overpredicted both areal coverages $\geq$ 10.0 mm h^{-1} and domain-total precipitation (Figs. 14d–f, 15). Overall, considering the entire forecast period, the partial and continuously cycling EnKFs had their strengths and weaknesses, and no ensemble had a clearly superior precipitation climatology.

6. Summary and conclusions

Several EnKF DA experiments with 80 members and 15-km horizontal grid spacing were performed over a computational domain spanning the CONUS for a 4-week period. These EnKFs were configured identically except for cycling procedure: one EnKF employed continuous cycling, while the others used a partial cycling methodology where limited-area analyses were discarded after 12 or 18 h of self-contained cycles and reinitialized from global model fields the next day. Posterior 0000 UTC ensembles from all EnKFs initialized 36-h, 3-km, 10-member CAE forecasts that were evaluated with a focus on precipitation. Additionally, CAE forecasts were initialized from both GEFS ICs and "blended" states constructed by combining small scales from continuously cycling EnKF analyses with large scales from GEFS ICs using a low-pass filter.

Through ~18 h, all EnKF-initialized forecasts outperformed GEFS-initialized forecasts, consistent with S21 and indicating that limited-area EnKFs can initialize better short-term CAE forecasts than global ICs. In addition, ~1–18-h forecasts with blended and unblended continuously cycling EnKF ICs were comparable to or better than those with partial cycling EnKF ICs. These results suggest continuously cycling EnKFs hold promise for short-term CAE forecast applications, for which partial cycling does not obviously represent a superior initialization approach.

Conversely, partial cycling EnKF analyses and GEFS ICs yielded ~18–36-h precipitation forecasts comparable to or better than those with unblended continuously cycling EnKF ICs, although improvements were only sometimes statistically significant. However, blended ICs produced comparable quality ~18–36-h forecasts as partial cycling EnKF ICs. Therefore, blending appears to be a simple way of improving ~18–36-h CAE forecasts initialized from continuously cycling EnKFs, corroborating S21 and suggesting that blending may be a viable alternative to partial cycling initialization for next-day CAE forecast systems. Moreover, there may be opportunities to improve blending methodologies to ameliorate issues regarding balance (e.g., S21) and physical inconsistencies that could potentially arise if corresponding limited-area and global fields greatly differ.

Benefits of ~18–36-h forecasts engendered by partial cycling EnKF and blended ICs were associated with large-scale spectral characteristics of blended and partial cycling EnKF ICs more closely resembling those of GFS analyses than unblended continuously cycling EnKF ICs. These findings suggest that limited-area ICs should strive to emulate large-scale characteristics of global models to initialize the best possible next-day forecasts, which are more influenced by large-scale flows than shorter-term forecasts.

Precisely why the limited-area continuously cycling EnKF had difficulty achieving large-scale characteristics of global analyses is unclear and should be examined in future studies, with the ultimate goal of improving large-scale continuously cycling EnKF analyses such that blending is no longer needed. However, lateral boundaries place an inherent limit on the longest resolvable waves, which may fundamentally constrain ability of limited-area continuously cycling DA systems to accurately depict and predict large-scale features. Insights about this potential limitation may be provided by experimenting with limited-area continuously cycling DA systems over progressively larger domains to assess whether longwave characteristics eventually attain those of global analyses. Furthermore, while the RRFS and other future CAEs over the CONUS will likely have finer-resolution ICs than our 15-km analyses, solely increasing analysis resolution is unlikely to recover large-scale characteristics of global analyses, and we suspect our overall conclusions about partial versus continuous cycling would hold in both higher and lower resolution DA systems with similar domain sizes. Nonetheless, further work is needed to confirm this hypothesis.

Partial cycling EnKFs can likely be improved, perhaps by carefully specifying initial spread on a per-variable basis and tuning DA parameters. Additionally, other partial cycling methodologies might be explored; as opposed to our method of periodically restarting entire ensembles from external (i.e., global) fields, perturbations derived from continuously cycling EnKFs could be periodically recentered about externally provided central initial states (e.g., Schwartz et al. 2020), thus propagating limited-area ensembles indefinitely through time while still introducing external information. Also, our overarching findings suggest an ideal self-contained cycling length for CAE initialization may exist where partial cycling states are sufficiently spun up yet retain sufficiently strong memories of large-scale external model characteristics, and further work could identify this optimum, which likely depends on domain size and external model traits.

However, our findings instead provide justification for devoting resources toward developing and improving continuously cycling EnKFs over the CONUS for CAE initialization, rather than investing in further partial cycling DA developments. In fact, a combination of continuous cycling and blending may altogether obviate the need for partial cycling, as continuously cycling EnKF analyses both initialized short-term CAE forecasts comparable to or better than those initialized from partial cycling EnKF analyses, and, when blended with GEFS ICs, yielded next-day CAE forecasts usually statistically indistinguishable from those with ICs produced through partial cycling. Thus, partial cycling systems can be replaced by continuously cycling DA systems that incorporate blending without sacrificing forecast quality at either short-term or next-day forecast ranges. Accordingly, given that continuously cycling methodologies have numerous advantages compared to partial cycling approaches and can streamline and accelerate model improvement efforts, we suggest NCEP strongly consider adopting continuously cycling DA to initialize future operational limited-area models over the CONUS like the RRFS.

Acknowledgments. This work was partially funded by NCAR's Short-term Explicit Prediction (STEP) program and NOAA/OAR Office of Weather and Air Quality Grants NA17OAR4590182 and NA19OAR4590232. All forecasts were produced on NCAR's Cheyenne supercomputer (Computational and Information Systems Laboratory 2017). Thanks to two anonymous reviewers for their constructive comments. NCAR is sponsored by the National Science Foundation.

Data availability statement. Data supporting the findings of this work are available from the corresponding author. Specifically, raw precipitation forecasts from all numerical experiments are available from the corresponding author, and ST4 observations used for verification are available from https://doi.org/10.5065/D69Z93M3. 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 and blending were saved and are available upon request.

APPENDIX

Partial Cycling EnKF Sensitivity to Land Surface State Initialization

Even though PC_{06z} and PC_{12z} were initialized without soil state spread, they both had more domain-average top-

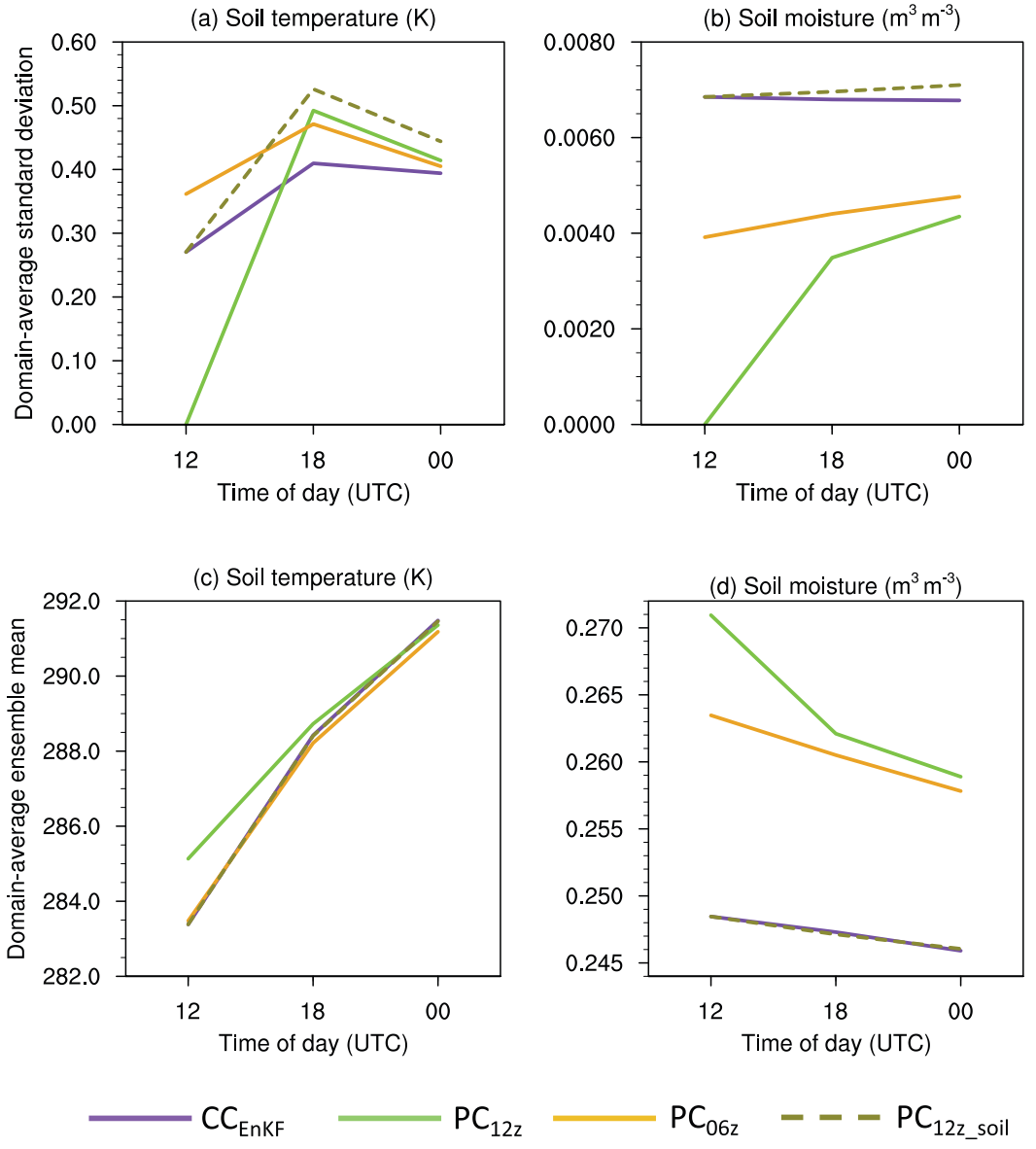

FIG. A1. Standard deviation of top-layer (a) soil temperature (K) and (b) soil moisture (m^3 m^{-3}) averaged over land points within the portion of the 15-km domain collocated with the 3-km domain (Fig. 1) and all posterior ensembles between 1200 UTC 24 Apr and 0000 UTC 20 May 2017 (inclusive) as a function of time of day. (c),(d) As in (a) and (b), respectively, but for domain-average ensemble means.

layer soil temperature spread than CC_{EnKF} at 0000 UTC (Fig. A1a), indicating quick adjustments to diverse atmospheric fields. Additionally, domain-average ensemble mean top-layer soil temperatures in the partial and continuously cycling EnKFs were similar by 0000 UTC (Fig. A1c). Conversely, although partial and continuously cycling EnKF top-layer soil moistures became closer with time, 0000 UTC soil moisture spread was ~50%–75% lower in the partial cycling EnKFs compared to the continuously cycling EnKF (Fig. A1b), and top-layer soil moisture remained wetter in the partial cycling EnKFs (Fig. A1d).

For domain-average ensemble mean top-layer soil temperature and moisture, PC_{12z_soil} paralleled CC_{EnKF} (Figs. A1c,d), which is sensible, as their 1200 UTC soil states were identical. However, PC_{12z_soil} spread quickly deviated from CC_{EnKF} spread and became larger by 0000 UTC (Figs. A1a,b), suggesting top-layer soil state spread is sensitive to low-level atmospheric spread and consistent with Figs. 6a–c, which revealed low-level 1200 UTC atmospheric spread was larger in PC_{12z} (and PC_{12z_soil}) compared to CC_{EnKF}.

To assess whether the soil moisture differences impacted precipitation forecasts, 0000 UTC analyses from PC_{12z_soil}

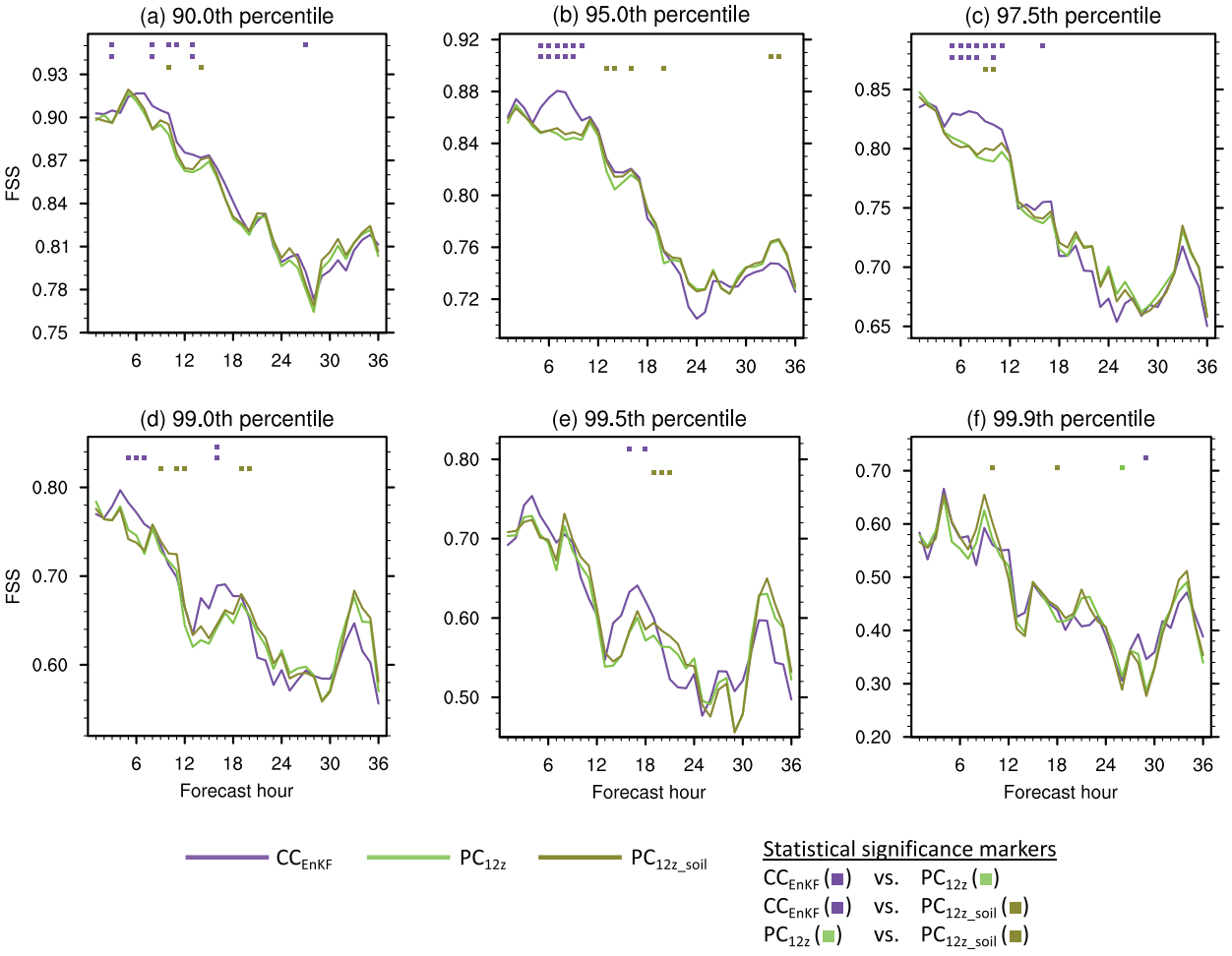

FIG. A2. As in Fig. 10, but for FSSs aggregated over 13 3-km, 10-member ensemble forecasts initialized at 0000 UTC between 25 Apr and 7 May 2017 (inclusive), focusing on experiments designed to assess forecast sensitivity to land surface state initialization in partial cycling EnKFs.

initialized 36-h, 10-member ensemble forecasts over the nested domain (Fig. 1), but these forecasts were only produced between 25 April and 7 May 2017 (inclusive) to save computing resources; differences between PC_{12z} and PC_{12z_soil} were attributable to different 1200 UTC soil states, while differences between PC_{12z_soil} and CC_{EnKF} were attributable to different 1200 UTC atmospheric fields. CAE forecasts were clearly more sensitive to atmospheric ICs than initial soil states, as FSS differences between PC_{12z_soil} and CC_{EnKF} were much larger than those between PC_{12z} and PC_{12z_soil} (Fig. A2). Therefore, differences between CAE forecasts initialized from CC_{EnKF} and PC_{12z} were not due to different soil moistures.

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.

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.

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

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.

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

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*, Amer. Meteor. Soc., 12.8, https://ams.confex.com/ams/101ANNUAL/meetingapp.cgi/Paper/378383 .

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

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

Clark, A. J., 2017: Generation of ensemble mean precipitation forecasts from convection-allowing ensembles. *Wea. Forecasting*, **32**, 1569–1583, https://doi.org/10.1175/WAF-D-16-0199.1.

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

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

Computational and Information Systems Laboratory, 2017: Cheyenne: HPE/SGI ICE XA System (NCAR Community Computing). National Center for Atmospheric Research, accessed 6 May 2021, https://doi.org/10.5065/D6RX99HX.

COSMO, 2021: MeteoSwiss Operational Applications within COSMO. Accessed 6 May 2021, http://www.cosmo-model.org/content/tasks/operational/meteoSwiss/default.htm.

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.

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.

Djalalova, I. V., and Coauthors, 2016: The POWER experiment: Impact of assimilation of a network of coastal wind profiling radars on simulating offshore winds in and above the wind turbine layer. *Wea. Forecasting*, **31**, 1071–1091, https://doi.org/10.1175/WAF-D-15-0104.1.

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.

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

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.

——, and J. A. Weyn, 2016: Thunderstorms do not get butterflies. *Bull. Amer. Meteor. Soc.*, **97**, 237–243, https://doi.org/10.1175/BAMS-D-15-00070.1.

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.

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

——, M. Chen, Y. Li, and J. Zhong, 2021: An implementation of full cycle strategy using dynamic blending for rapid refresh short-range weather forecasting in China. *Adv. Atmos. Sci.*, **38**, 943–956, https://doi.org/10.1007/s00376-021-0316-7.

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

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

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

Gehne, M., T. M. Hamill, G. T. Bates, P. Pegion, and W. Kolczynski, 2019: Land surface parameter and state perturbations in the global ensemble forecast system. *Mon. Wea. Rev.*, **147**, 1319–1340, https://doi.org/10.1175/MWR-D-18-0057.1.

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, 22 pp., http://polar.ncep.noaa.gov/mmab/papers/tn260/MMAB260.pdf.

Gilleland, E., A. S. Hering, T. L. Fowler, and B. G. Brown, 2018: Testing the tests: What are the impacts of incorrect assumptions when applying confidence intervals or hypothesis tests to compare competing forecasts? *Mon. Wea. Rev.*, **146**, 1685–1703, https://doi.org/10.1175/MWR-D-17-0295.1.

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.

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.

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.

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

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

Hu, M., S. G. Benjamin, T. T. Ladwig, D. C. Dowell, S. S. Weygandt, C. R. Alexander, and J. S. Whitaker, 2017: GSI three-dimensional ensemble-variational hybrid data assimilation using a global ensemble for the regional rapid refresh model. *Mon. Wea. Rev.*, **145**, 4205–4225, https://doi.org/10.1175/MWR-D-16-0418.1.

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.

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

——, ——, and B. D. Jamison, 2020: Commercial-aircraft-based observations for NWP: Global coverage, data impacts, and COVID-19. *J. Appl. Meteor. Climatol.*, **59**, 1809–1825, https://doi.org/10.1175/JAMC-D-20-0010.1.

Janjić, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927–945, https://doi.org/10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2.

——, 2002: Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP Meso model. NCEP Office Note 437, 61 pp., http://www.emc.ncep.noaa.gov/officenotes/newernotes/on437.pdf.

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

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

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

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

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

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

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

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

Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP Global Data Assimilation System. *Wea. Forecasting*, **24**, 1691–1705, https://doi.org/10.1175/2009WAF2222201.1.

Kong, F., and Coauthors, 2008: Real-time storm-scale ensemble forecasting during the 2008 Spring Experiment. *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.

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

——, ——, A. H. N. Lim, M. Hu, J. M. Brown, and S. G. Benjamin, 2017b: Radiance preprocessing for assimilation in the hourly updating rapid refresh mesoscale model: A study using AIRS data. *Wea. Forecasting*, **32**, 1781–1800, https://doi.org/10.1175/WAF-D-17-0028.1.

Lin, Y., and K. E. Mitchell, 2005: The NCEP stage II/IV hourly precipitation analyses: Development and applications. *19th*

Conf. on Hydrology, San Diego, CA, Amer. Meteor. Soc., 1.2., http://ams.confex.com/ams/pdfpapers/83847.pdf.

Lorenc, A. C., and Coauthors, 2000: The Met. Office global three-dimensional variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc.*, **126**, 2991–3012, https://doi.org/10.1002/qj.49712657002.

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.

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

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

Politis, D. N., and J. P. Romano, 1992: A circular block-resampling procedure for stationary data. *Exploring the Limits of Bootstrap*, R. LePage and L. Billard, Eds., John Wiley and Sons, 263–270.

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

Poterjoy, J., G. J. Alaka, Jr., and H. R. Winterbottom, 2021: The irreplaceable utility of sequential data assimilation for numerical weather prediction system development: Lessons learned from an experimental HWRF system. *Wea. Forecasting*, **36**, 661–677, https://doi.org/10.1175/WAF-D-20-0204.1.

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.

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

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

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

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

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

Roberts, 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.

Rogers, E., and Coauthors, 2009: The NCEP North American Mesoscale modeling system: Recent changes and future plans. *23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction*, Omaha, NE, Amer. Meteor. Soc., 2A.4., http://ams.confex.com/ams/pdfpapers/154114.pdf.

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.

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

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

Schumacher, R. S., and A. J. Clark, 2014: Evaluation of ensemble configurations for the analysis and prediction of heavy-rain-producing mesoscale convective systems. *Mon. Wea. Rev.*, **142**, 4108–4138, https://doi.org/10.1175/MWR-D-13-00357.1.

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

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

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

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

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

——, G. S. Romine, K. R. Smith, and M. L. Weisman, 2014: Characterizing and optimizing precipitation forecasts from a convection-permitting ensemble initialized by a mesoscale ensemble Kalman filter. *Wea. Forecasting*, **29**, 1295–1318, https://doi.org/10.1175/WAF-D-13-00145.1.

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

——, 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.

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

Skamarock, W. C., 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.

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

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

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

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.

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

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

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

Weyn, J. A., and D. R. Durran, 2017: The dependence of the predictability of mesoscale convective systems on the horizontal scale and amplitude of initial errors in idealized simulations. *J. Atmos. Sci.*, **74**, 2191–2210, https://doi.org/10.1175/JAS-D-17-0006.1.

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

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

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

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

Wilks, D. S., 1997: Resampling hypothesis tests for autocorrelated fields. *J. Climate*, **10**, 65–82, https://doi.org/10.1175/1520-0442(1997)010<0065:RHTFAF>2.0.CO;2.

——, 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.

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

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

Xue, M., and Coauthors, 2007: CAPS real-time storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2007 Spring Experiment. *22nd Conf. on Weather Analysis and Forecasting/18th Conf. on*

Numerical Weather Prediction, Salt Lake City, UT, Amer. Meteor. Soc., 3B.1, http://ams.confex.com/ams/pdfpapers/124587.pdf.

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

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

Zhang, Y., D. J. Stensrud, and F. Zhang, 2019: Simultaneous assimilation of radar and all-sky satellite infrared radiance observations for convection-allowing ensemble analysis and prediction of severe thunderstorms. *Mon. Wea. Rev.*, **147**, 4389–4409, https://doi.org/10.1175/MWR-D-19-0163.1.

Zhao, Q. Y., and F. H. Carr, 1997: A prognostic cloud scheme for operational NWP models. *Mon. Wea. Rev.*, **125**, 1931–1953, https://doi.org/10.1175/1520-0493(1997)125<1931:APCSFO>2.0.CO;2.

Zheng, W., M. Ek, K. Mitchell, H. Wei, and J. Meng, 2017: Improving the s table surface layer in the NCEP Global Forecast System. *Mon. Wea. Rev.*, **145**, 3969–3987, https://doi.org/10.1175/MWR-D-16-0438.1.

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.

Zhu, K., M. Xue, Y. Pan, M. Hu, S. G. Benjamin, S. S. Weygandt, and H. Lin, 2019: The impact of satellite radiance data assimilation within a frequently updated regional forecast system using a GSI-based ensemble Kalman filter. *Adv. Atmos. Sci.*, **36**, 1308–1326, https://doi.org/10.1007/s00376-019-9011-3.

Zou, X., Z. Qin, and F. Weng, 2011: Improved coastal precipitation forecasts with direct assimilation of *GOES-11/12* imager radiances. *Mon. Wea. Rev.*, **139**, 3711–3729, https://doi.org/10.1175/MWR-D-10-05040.1.