ABSTRACT

Title of Dissertation:	AN EVALUATION OF CONVECTION-ALLOWING ENSEMBLE FORECAST SENSITIVITY TO INITIAL CONDITIONS
	Craig Steven Schwartz, Doctor of Philosophy, 2021
Dissertation directed by:	Professor Jonathan Poterjoy, Department of Atmospheric and Oceanic Science

This dissertation aims to advance understanding of initial conditions (ICs) for convection-allowing ensembles (CAEs). To do so, experiments with 80-member limited-area ensemble Kalman filters (EnKFs) were performed over the entire conterminous United States for a 4-week period. The EnKF data assimilation systems differed in terms of their cycling strategies (continuous or partial cycling) and horizontal grid spacings (15- or 3-km horizontal grid spacing). EnKF analyses initialized 36-h, 3-km, 10-member CAE forecasts that were evaluated with a focus on precipitation, providing insights about CAE forecast sensitivity to ICs. Additionally, EnKF analyses were leveraged to isolate CAE forecast sensitivity to resolution of both IC perturbations and central initial states about which IC perturbations were centered. A “blending” approach was also used to produce new sets of CAE ICs by combining small scales from continuously cycling EnKF analyses with large scales from Global Ensemble Forecast System (GEFS) ICs using a low-pass filter.

Key results are as follows:

- CAE forecasts initialized from continuously cycling 3-km EnKF analyses were more skillful and reliable than those initialized from downscaled GEFS and continuously cycling 15-km EnKF ICs through 12–18 and 6–12 h, respectively. Conversely, after 18 h, GEFS-initialized forecasts were better than forecasts initialized from continuously cycling EnKFs. Blended 3-km ICs led to ~18–36-h forecasts possessing comparable quality as GEFS-initialized forecasts while preserving short-term forecast benefits of unblended continuously cycling 3-km EnKF analyses.
- Continuously cycling EnKF analyses initialized ~1–18-h 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.
- It is more important for central initial states than for IC perturbations to possess convection-allowing horizontal grid spacing for short-term CAE forecasting applications.

These collective findings have important implications for model developers working on next-generation CAEs and suggest paths toward potentially saving computing resources, streamlining processes for improving CAE ICs, and unifying short-term and next-day CAE forecasting systems.

AN EVALUATION OF CONVECTION-ALLOWING ENSEMBLE FORECAST SENSITIVITY TO INITIAL CONDITIONS

by

Craig Steven Schwartz

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Advisory Committee:
Professor Jonathan Poterjoy, Chair
Professor Kayo Ide
Professor Jacob Wenegrat
Professor Maria Cameron, Dean's Representative
Dr. David Dowell
Dr. Jacob Carley

Acknowledgements

Thanks to my advisor, Dr. Jonathan Poterjoy, for his collaborative spirit and valuable feedback. I would also like to thank Drs. Kayo Ide, Jacob Wenegrat, Jacob Carley, David Dowell, and Maria Cameron for formally serving on my committee and discussions about my research. Thanks to Jamie Bresch and Dr. Glen Romine from the National Center for Atmospheric Research (NCAR) for their contributions to this work. My dissertation builds on decades of scientific and technical advances by staff at NCAR, who deserve collective recognition, as do the many WRF, GSI, and DART developers both within NCAR and at other organizations. Finally, thanks to my wife, Julia, for her support.

This material is based upon work supported by NCAR, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977. Additional funding was provided by NCAR's Short-term Explicit Prediction (STEP) program and NOAA/OAR Office of Weather and Air Quality Grants NA17OAR4590182, NA19OAR4590232, NA17OAR4590114, and NA17OAR4590122. All computations were performed on NCAR's Cheyenne supercomputer (Computational and Information Systems Laboratory 2017).

Chapters 2 and 4 are reprinted with permission by the American Meteorological Society. The complete citations are:

Schwartz, C. S., 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. © American Meteorological Society. Used with permission.

Schwartz, C. S., J. Poterjoy, J. R. Carley, D. C. Dowell, G. S. Romine, and K. Ide, 2021: Comparing partial and continuously cycling ensemble Kalman filter data assimilation systems for convection-allowing ensemble forecast initialization. *Wea. Forecasting*, In press, https://doi.org/10.1175/WAF-D-21-0069.1. © American Meteorological Society. Used with permission. This preliminary version has been accepted for publication in *Weather and Forecasting* and may be fully cited. The final typeset copyedited article will replace the EOR when it is published.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

3DVAR	Three-dimensional variational
BEC	Background error covariance
CAE	Convection-allowing ensemble
CONUS	Conterminous United States
DA	Data assimilation
DART	Data Assimilation Research Testbed
DCT	Discrete cosine transform
EAKF	Ensemble adjustment Kalman filter
EnKF	Ensemble Kalman filter
EnVar	Ensemble–variational
FSS	Fractions skill score
GDAS	Global Data Assimilation System
GEFS	Global Ensemble Forecast System
GFS	Global Forecast System
GSI	Gridpoint Statistical Interpolation
HREF	High-resolution Ensemble Forecast
HRRR	High-resolution Rapid Refresh
HRRRE	High-resolution Rapid Refresh Ensemble
IC	Initial condition
JW20	Johnson and Wang (2020)
LBC	Lateral boundary condition
NAM	North American Mesoscale
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NEP	Neighborhood ensemble probability
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical weather prediction
RAP	Rapid Refresh
RH	Relative humidity
RMSD	Root-mean-square difference
RMSE	Root-mean-square error
ROC	Relative operating characteristic
RRFS	Rapid Refresh Forecast System
RRTMG	Rapid Radiative Transfer Model for Global Climate Models
S20	Schwartz et al. (2020)
S21	Schwartz et al. (2021)
ST4	Stage IV
WoF	Warn-on-Forecast
WRF	Weather Research and Forecasting

Chapter 1: Introduction

Convection-allowing numerical weather prediction (NWP) models[1] produce better precipitation and severe weather forecasts than coarser resolution models with parameterized convection (e.g., Done et al. 2004; Kain et al. 2006; Weisman et al. 2008; Clark et al. 2009, 2010; Schwartz et al. 2009) and have revolutionized weather forecasting (e.g., Clark et al. 2016). While the first real-time convection-allowing model forecasts[2] were deterministic (e.g., Done et al. 2004), expansion of convection-allowing modeling to ensembles occurred quickly (Xue et al. 2007), and convection-allowing ensembles (CAEs) are now operational at most major NWP centers around the world (e.g., Gebhardt et al. 2011; Peralta et al. 2012; Hagelin et al. 2017; Raynaud and Bouttier 2017; Jirak et al. 2018; Klasa et al. 2018). Although there are many avenues to improve CAEs, like advancing boundary condition perturbation methods, stochastic physics implementations, and multi-physics ensemble configurations, this dissertation focuses solely on CAE initial conditions (ICs), which exert a strong influence on CAE forecasts (e.g., Raynaud and Bouttier 2016; Zhang 2019; Schwartz et al. 2020).

[1] In a convection-allowing NWP model, the deep cumulus parameterization scheme is removed to allow explicit representation of deep convection through model dynamics. Over the United States, cumulus parameterization can be safely removed with approximately 4-km horizontal grid spacing (e.g., Kain et al. 2008).

[2] Convection-allowing models are routinely run to study physical processes, often in idealized scenarios. However, this dissertation does not concern these types of simulations and instead focuses on convection-allowing models explicitly designed for forecasting applications.

At first, CAEs were initialized by downscaling pre-existing operational analyses or short-term forecasts, like those provided by the National Centers for Environmental Prediction's (NCEP's) Short-range Ensemble Forecast System (Du et al. 2015), onto the convection-allowing model grid (Xue et al. 2007; Kong et al. 2008, 2009), and data assimilation (DA) was not explicitly performed during any stage of the initialization process. While this downscaling approach is still sometimes used today to initialize CAEs (e.g., Clark 2017; Jirak et al. 2018; Cafaro et al. 2019; Porson et al. 2019), as computing has increased and ensemble-based DA methods like the ensemble Kalman filter (EnKF; Evensen 1994; Houtekamer and Zhang 2016) have matured, limited-area DA systems have been developed specifically for CAE initialization (e.g., Romine et al. 2014; Schwartz et al. 2014, 2015a; Wheatley et al. 2015; Schraff et al. 2016; Hagelin et al. 2017; Johnson and Wang 2017; Keresturi et al. 2019).

Despite these developments, there is still a need to improve ICs for CAEs, and it has historically been challenging to develop "formally designed" CAEs (e.g., Schwartz et al. 2019) over the conterminous United States (CONUS) that are equally likely (in a mean sense) and outperform CAEs initialized by ad hoc methods (e.g., Roberts et al. 2020; Schwartz et al. 2020). Additionally, there is little consensus about how to best initialize CAEs, with the few papers rigorously devoted to intercomparing CAE ICs painting an incomplete picture (Raynaud and Bouttier 2016; Johnson and Wang 2020; Schwartz et al. 2020).

Therefore, with overarching goals to both improve and advance understanding of CAE initialization over the CONUS, this dissertation examines topics at the intersection of DA and CAE forecasting by performing systematic objective evaluations. The outcomes of several novel investigations presented in chapters 2–4 have highly relevant implications for future operational CAEs currently being developed at NCEP and other meteorological centers around the world. Additionally, the experiments and evaluations provide insights about the current capabilities of limited-area NWP models, demonstrate new possibilities for CAE initialization, and serve as important baselines that other modelers can attempt to replicate. Moreover, some of the findings are applicable to future *global* CAEs that will eventually be developed when computing resources permit.

To perform these investigations, we used the Advanced Research Weather Research and Forecasting (WRF) model. The Advanced Research WRF model equations are exhaustively documented by Skamarock et al. (2008) and not repeated here. Our experiments focus on WRF-based CAE forecasts with 3-km horizontal grid spacing, which was chosen to match the grid spacing of current operational convection-allowing models over the CONUS. Choices for physical parameterizations were guided by almost a decade of experience working with WRF-based CAEs and included the Thompson et al. (2008) microphysics scheme, Rapid Radiative Transfer Model for Global Climate Models (RRTMG) longwave and shortwave radiation parameterizations (Mlawer et al. 1997; Tegen et al. 1997; Iacono et al. 2008), Mellor–Yamada–Janjić planetary boundary layer parameterization (Mellor and Yamada 1982; Janjić 1994, 2002), Noah land surface model (Chen and Dudhia 2001), and, on a 15-km domain that provided lateral boundary conditions for the 3-km forecasts, the Tiedtke cumulus parameterization (Tiedtke 1989; Zhang et al. 2011). The 3-km forecasts did not use cumulus parameterization.

For the DA components of this research, we employed an EnKF, which is a state-of-the-science method that uses flow-dependent background error covariances and seamlessly melds ensemble DA and ensemble forecasting. The EnKF updates a background ($\mathbf{x}_b$) to produce an analysis ($\mathbf{x}_a$). The analysis is found by optimally combining the background (i.e., $\mathbf{x}_b$) and observations ($\mathbf{y}^o$), as expressed by

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y}^o - H(\mathbf{x}_b)), \tag{1.1}$$

where H is the observation operator that transforms model quantities into model-simulated observations and $\mathbf{K}$ is the Kalman Gain, given by

$$\mathbf{K} = \mathbf{BH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1}. \tag{1.2}$$

In Eq. (1.2), $\mathbf{H}$ is the linearization of H about $\mathbf{x}_b$, and $\mathbf{B}$ and $\mathbf{R}$ are the background error covariance and observation error covariance matrices, respectively. The EnKF elegantly uses an N-member ensemble to represent $\mathbf{B}$ as $\mathbf{B} = 1/(N-1)\sum_{i=1}^{N} \delta\mathbf{x}_i^b (\delta\mathbf{x}_i^b)^T$, where $\delta\mathbf{x}_i^b$ is the perturbation of the ith prior (before assimilation) ensemble member with respect to the ensemble mean prior. There are different EnKF flavors (e.g., Tippett et al. 2003), and the particular EnKF used in this research was the ensemble adjustment Kalman filter (EAKF; Anderson 2001, 2003; Anderson and Collins 2007). Eqs. (2.1)–(2.13) of Liu et al. (2007) can be consulted for a gentle description of the specific EAKF algorithm we used.

Using our WRF-based EnKFs, chapter 2 demonstrates the first continuously cycling convection-allowing EnKF over the entire CONUS, assesses CAE forecast sensitivity to EnKF resolution, and shows that a nearly cost-free "blending" approach substantially improves CAE forecast performance after ~18 h. Leveraging the EnKFs described in chapter 2, chapter 3 investigates CAE forecast sensitivity to resolution of both IC perturbations and central initial states about which IC perturbations are centered. In chapter 4, performance of CAEs initialized from EnKFs using both partial and continuously cycling DA methodologies is directly compared. Finally, chapter 5 briefly summarizes key findings and provides some thoughts about how this research could be extended.

Chapter 2: 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

2.1. Introduction

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

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

However, there are obstacles to implementing convection-allowing DA systems over domains large enough to resolve meso-alpha- to synoptic-scale features, especially when using state-of-the-science ensemble-based DA algorithms like the ensemble Kalman filter (EnKF; Evensen 1994; Houtekamer and Zhang 2016), which produces flow-dependent analysis ensembles and has become popular for initializing CAEs (e.g., Jones and Stensrud 2012; Melhauser and Zhang 2012; Schumacher and Clark 2014; Schwartz et al. 2014, 2015a,b, 2019). One challenge is simply computational expense, which grows directly with domain size[3], and accordingly, most convection-allowing EnKFs and their associated CAE forecasts have relatively small domains centered on a single European country (e.g., Schraff et al. 2016; COSMO 2021) or a small portion of the conterminous United States (CONUS). For example, NOAA's experimental "Warn-on-Forecast" (WoF; Stensrud et al. 2009, 2013) system, initialized from 36-member 3-km EnKF analyses, covers less than 1000 km x 1000 km (Wheatley et al. 2015; Jones et al. 2016, 2018, 2020; Skinner et al. 2018).

Fortunately, computing challenges can be overcome with increased resources, and recently, several studies initialized CAE forecasts from 40-member EnKF analyses with 3-km or finer horizontal grid spacing over the entire CONUS (Duda et al. 2019; Gasperoni et al. 2020; Johnson et al. 2020). Similarly, NOAA's real-time, experimental High-Resolution Rapid Refresh Ensemble (HRRRE) is initialized from CONUS-spanning, 3-km, 36-member EnKF analyses (Dowell et al. 2016; Ladwig et

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

al. 2018). However, 36–40-member EnKFs are likely smaller than desirable, considering that operational global EnKFs run by the United States and Canada respectively have 80 and 256 members, and generally, EnKFs benefit from larger ensembles (e.g., Zhang et al. 2013; Houtekamer et al. 2014).

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

Given these scientific and computing challenges, operational convection-allowing continuously cycling EnKFs and attendant CAEs over Europe have small domains (e.g., Schraff et al. 2016; COSMO 2021), while large-domain convection-allowing EnKFs over the CONUS (e.g., Duda et al. 2019; Gasperoni et al. 2020; Johnson et al. 2020; HRRRE) employ "partial cycling" strategies that periodically discard convection-allowing analysis cycles and replace them with coarser-resolution, large-scale external analyses in hopes of tempering bias accumulations (e.g., Hsiao et al. 2012; Benjamin et al. 2016; Wu et al. 2017). This partial cycling approach over the CONUS seems justified, as Schwartz et al. (2020) showed that a limited-area continuously cycling EnKF with convection-parameterizing resolution did not initialize better CAE precipitation forecasts over the CONUS than downscaled global analyses.

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

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

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

2.2. Model configurations, EnKF settings, and experimental design

2.2.1. FORECAST MODEL

All forecasts were produced by version 3.9.1.1 of the Advanced Research Weather Research and Forecasting (WRF) model (Skamarock et al. 2008; Powers et al. 2017) over a nested computational domain (Fig. 2.1a). The horizontal grid spacing was 15 km in the outer domain and 3 km in the nest, and time steps were 60 and 12 s in the 15- and 3-km domains, respectively. Both domains had 51 vertical levels distributed as in the Rapid Refresh model (Benjamin et al. 2016) with a 15-hPa top. Physical parameterizations were identical across the two domains (Table 2.1), except no cumulus parameterization was employed on the convection-allowing 3-km grid, and all ensemble members used common physics and dynamics options.

2.2.2. ENKF DA SYSTEMS

2.2.2.1. EnKF experiments and configurations

Two primary DA experiments with 80-member ensembles were performed using an ensemble adjustment Kalman filter (Anderson 2001, 2003; Anderson and Collins 2007), a type of EnKF, as implemented in the Data Assimilation Research Testbed (DART; Anderson et al. 2009) software. The first EnKF experiment only produced analyses on the 15-km domain (Fig. 2.1a), and the 3-km domain was removed during WRF model advances between EnKF analyses. Conversely, the

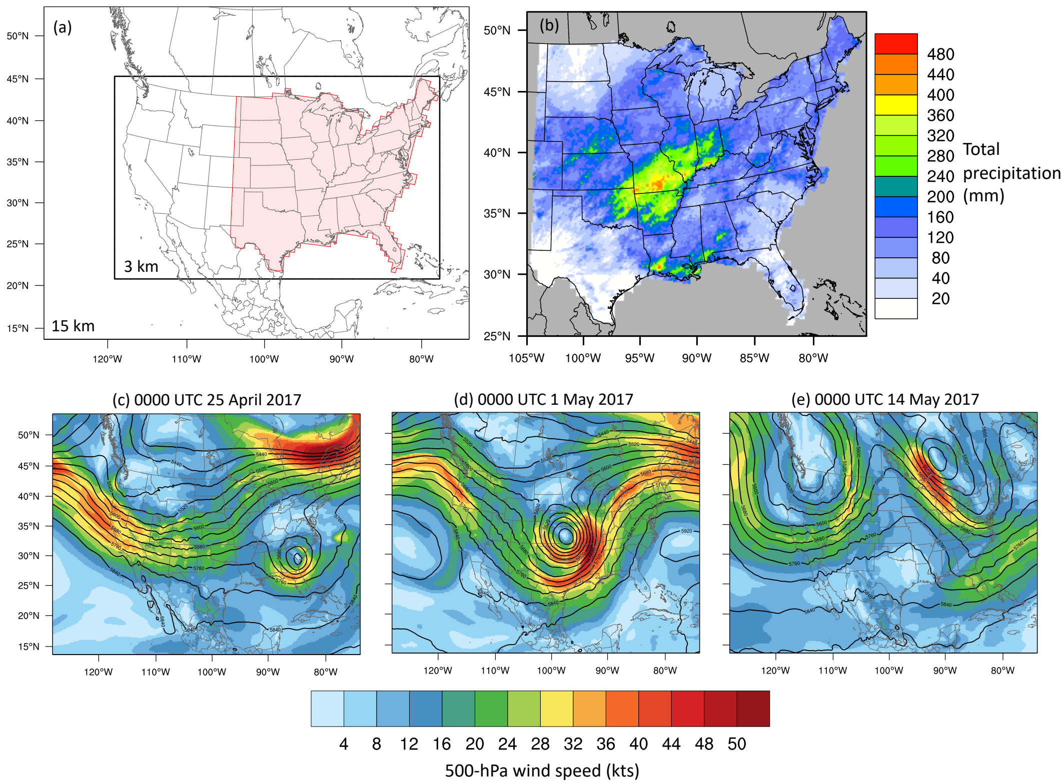

Fig. 2.1. (a) Computational domain. Horizontal grid spacing was 15 km in the outer domain (415 x 325 points) and 3 km in the nest (1581 x 986 points). Objective precipitation verification only occurred over the red shaded region of the 3-km domain (CONUS east of 105°W). (b) Total accumulated Stage IV (ST4) precipitation (mm) over the verification region between 0000 UTC 25 April and 1200 UTC 21 May 2017, which encompasses all possible valid times of the 36-h forecasts. (c)–(e) 500-hPa wind speed (shaded; kts) and height (m; contoured every 40 m) from Global Forecast System analyses valid at 0000 UTC (c) 25 April, (d) 1 May, and (e) 14 May 2017.

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

Table 2.1. Physical parameterizations for all WRF model forecasts. Cumulus parameterization was only used on the 15-km domain.

second EnKF experiment produced separate, independent analyses on *both* the 15- and 3-km domains, with nested WRF model forecasts between EnKF analyses. During these nested forecasts, which were ~45 times more expensive than the single-domain 15-km model advances, one-way feedback was employed such that the 15-km EnKF DA system was unaffected by the 3-km EnKF DA system (i.e., 15-km fields in the nested- and single-domain EnKF DA systems were identical), permitting a clean comparison of analysis and forecast sensitivity to EnKF resolution. The 15- and 3-km EnKFs updated identical state variables (Table 2.2), with hydrometers included in anticipation of experimentation with radar DA (section 2.4.3).

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

Table 2.2. Summary of EnKF configurations.

Initial 80-member ensembles were produced by interpolating the 0000 UTC 23 April 2017 0.25° NCEP Global Forecast System (GFS) analysis onto the 15-km domain and adding random, correlated, Gaussian noise with zero mean (e.g., Barker 2005; Torn et al. 2006) drawn from background error covariances provided by the WRF model's DA system (Barker et al. 2012). The randomly-produced 15-km ensemble was then downscaled onto the 3-km grid to initialize the 3-km EnKF, ensuring initial 15- and 3-km ensembles were identical aside from interpolation errors. These randomly-generated ensembles served as prior (before assimilation) ensembles for the first EnKF analyses, and the posterior (after assimilation) ensembles at 0000 UTC 23 April 2017 initialized 1-h, 80-member ensemble forecasts that became prior ensembles for the next EnKF analyses at 0100 UTC 23 April 2017. Analysis–forecast cycles with a 1-h period continued until 0000 UTC 20 May 2017 (649 total DA cycles). This experimental period (23 April – 20 May 2017) was similar to that in Schwartz (2019), which featured several heavy precipitation episodes primarily driven by strong synoptic forcing, a broad overall precipitation maximum centered in Missouri (Fig. 2.1b), and a variety of flow patterns (Figs. 2.1c–e).

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

Spurious correlations due to sampling error were mitigated with a sampling error correction scheme (Anderson 2012) and covariance localization [Eq. (4.10) of Gaspari and Cohn (1999)]. Vertical localization limited analysis increments to ± 1.0 scale height (in log pressure coordinates) away from an observation in both the 15- and 3-km EnKFs. However, horizontal localizations differed depending on EnKF resolution: 15-km EnKF analysis increments were forced to zero 1280 km from an observation, but to lessen expense and complete 3-km EnKF analyses quickly enough for operational applications, 3-km EnKF analysis increments were forced to zero 640 km from an observation, except rawinsonde observations could produce increments up to 1280 km away (Table 2.2). The vertical and 15-km EnKF horizontal localization distances were guided by previous experiences with DART (e.g., Romine et al. 2013, 2014; Schwartz et al. 2015a,b, 2019), and while our 3-km EnKF horizontal localization distances were similar to Johnson et al. (2015), they were larger than those in many other convection-allowing EnKFs (e.g., Harnisch and Keil 2015; Yussouf et al. 2015, 2016; Degelia et al. 2018; Gasperoni et al. 2020; Jones et al. 2020). However, these studies with smaller localization distances either used partial cycling strategies or only continuously cycled for a short period (days), and we believed that larger localization distances were necessary to provide stronger observational constraints in a large-domain continuously cycling 3-km EnKF.

EnKF spread was maintained by applying covariance inflation to posterior state-space perturbations about the ensemble mean following Whitaker and Hamill (2012)'s "relaxation-to-prior spread" algorithm with an inflation parameter $\alpha = 1.06$ in both the 15- and 3-km EnKFs. As noted by Schwartz and Liu (2014), $\alpha > 1$ meant inflated posterior spread was greater than prior spread, which, while counterintuitive, was necessary to maintain reasonable spread given absence of other spread-inducing methods like multi-physics ensembles, additive inflation, or stochastic physics. Several iterative weeklong trials with 15-km EnKFs were performed to settle on $\alpha = 1.06$, which provided acceptable prior observation-space statistics for the assumed observation errors (section 2.3).

2.2.2.2. Observations

Although DART has observation processing capabilities, we instead used NCEP's operational Gridpoint Statistical Interpolation (GSI) DA system (Kleist et al. 2009; Shao et al. 2016) for observation processing, which, relative to DART, has more sophisticated quality control, observation thinning, and observation error assignment capabilities. In addition, GSI's observation operators were used instead of DART's built-in observation operators to produce model-simulated conventional observations. Initially-specified observation errors were based on the HRRRE and identical in the 15- and 3-km EnKFs (Fig. 2.2; Table 2.3); GSI adjusted these errors to produce "final" observation error standard deviations (σ_o) actually used in the assimilation, as described by several texts (e.g., Schwartz and Liu 2014; Developmental Testbed Center 2016; Johnson and Wang 2017). These adjustments often inflated initially-specified observation errors (Fig. 2.2).

Time windows for the observation platforms varied and were based on Rapid Refresh model (Benjamin et al. 2016) and HRRRE settings, with generally smaller windows for frequently-reporting, stationary platforms, like METAR observations

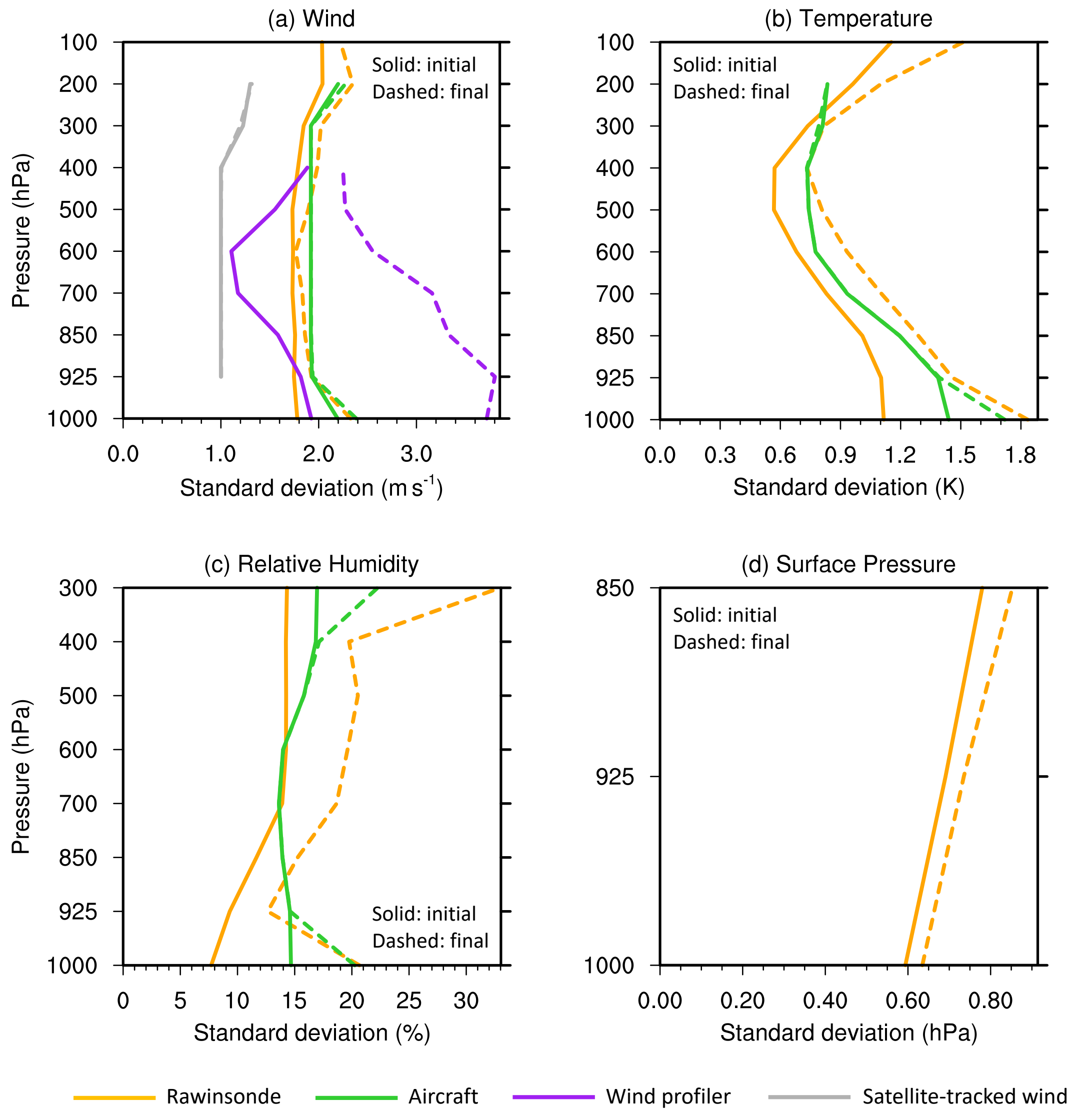

Fig. 2.2. Initially-specified (solid lines) and final (after GSI adjustment; dashed lines) observation error standard deviations as a function of pressure for (a) wind (m s^{-1}), (b) temperature (K), (c) relative humidity (%), and (d) surface pressure (hPa) observations with vertically varying errors averaged over all observations assimilated between 0000 UTC 25 April and 0000 UTC 20 May 2017 (inclusive) by both the 15- and 3-km EnKFs. If a particular observation type was not assimilated at a certain pressure level, no value is plotted.

(Table 2.3), and all observations were assumed valid at the analysis time. Moisture observations were initially processed as specific humidity, but because GSI requires moisture observation errors in terms of relative humidity, moisture observations were ultimately converted to and assimilated as relative humidity using the prior ensemble

Observing platform	Observation type	Initial observation error	Outlier check threshold (*a*)	Time window (h)
Rawinsonde	Surface pressure Temperature Relative humidity Wind	Fig. 2.2d Fig. 2.2b Fig. 2.2c Fig. 2.2a	5 7 7 10	1.5 1.5 1.5 1.5
Aircraft	Temperature Relative humidity Wind	Fig. 2.2b Fig. 2.2c Fig. 2.2a	7 7 10	0.75 0.75 0.75
Wind profiler	Wind	Fig. 2.2a	5	0.4
Global positioning system radio occultation (GPSRO)	Refractivity	1% of observation value	10	3.0
Infrared and water vapor channel satellite-tracked wind	Wind	Fig. 2.2a	2.5	1.5
Ship and buoy	Surface pressure Temperature Relative humidity Wind	0.44 hPa 0.8 K 3.9% 1.45 m s^{-1}	5 7 7 5	1.5 1.5 1.5 1.5
SYNOP and METAR	Surface pressure Temperature Relative humidity Wind	0.54 hPa 2.3 K 3.4% 1.2 m s^{-1}	5 5 7 5	0.25 0.25 0.25 0.25
Oklahoma and West Texas mesonet	Surface pressure Temperature Relative humidity Wind	0.35 hPa 1.5 K 4% 1.1 m s^{-1}	5 5 7 5	0.1 0.1 0.1 0.1

Table 2.3. Conventional observations that were assimilated and their outlier check thresholds, time windows, and initially-specified observation error standard deviations.

mean saturation specific humidity. Satellite-tracked wind and aircraft observations were thinned such that remaining observations were spaced 25 hPa apart vertically and 30 and 15 km apart horizontally in the 15- and 3-km EnKFs, respectively (Table 2.2); these different horizontal thinnings were chosen so the 15- and 3-km EnKFs had equal numbers of satellite-tracked wind and aircraft observations within their respective horizontal localization radii. Radiance observations were not assimilated since they generally yield small impacts over the CONUS (Lin et al. 2017a) given the multitude of available conventional observations. Additionally, the EnKFs did not assimilate radar observations, although an auxiliary experiment was performed where radar observations were assimilated with a 3-km EnKF (section 2.4.3).

Observations were subject to numerous quality control procedures, such as excluding observations from specific aircraft with known biases and applying an "outlier check" to reject observations whose ensemble mean innovations[4] were > $a\sigma_o$, where a varied from 2.5–10 depending on observation type and platform (Table 2.3). These a were generally fairly lenient and allowed most observations to pass the outlier check, which, along with our relatively large localization distances, reflected a philosophy that we wanted observations to heavily constrain the 1-h WRF model forecasts between EnKF analyses. Overall, the EnKFs assimilated 30,000–100,000 conventional observations each cycle, with a relative dearth of overnight observations due to fewer commercial flights and maxima at 0000 and 1200 UTC reflecting the majority of rawinsonde launches (Fig. 2.3). Ultimately, GSI-provided observations, final observation errors, and prior model-simulated observations for each ensemble member were ingested directly into DART for use in EnKF DA.

2.2.2.3. Forecast initialization

EnKF analysis ensembles initialized 36-h 10-member ensemble forecasts over the nested computational domain (Fig. 2.1a) at 0000 UTC between 25 April and 20

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

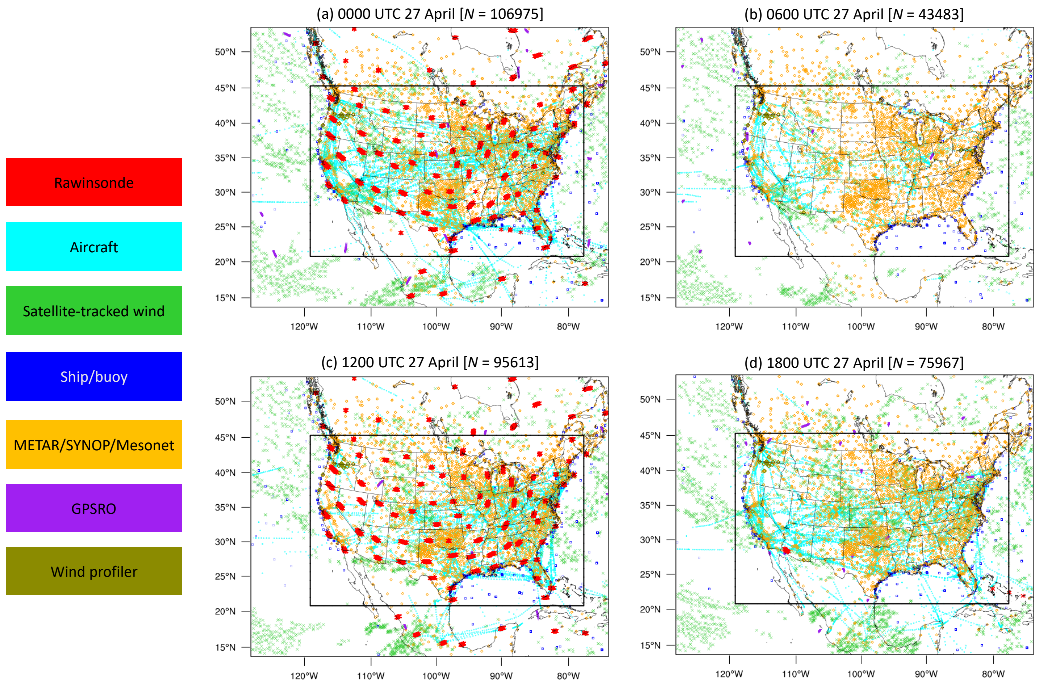

Fig. 2.3. Computational domain overlaid with observations assimilated by the 15-km EnKF during the (a) 0000 UTC, (b) 0600 UTC, (c) 1200 UTC, and (d) 1800 UTC 27 April 2017 analyses. Values of *N* in the headers indicate the number of assimilated observations. The inner box represents bounds of the 3-km domain; most observations located within the 3-km domain were also assimilated by the 3-km EnKF at these times.

May 2017 (inclusive; 26 forecasts). Although 80 EnKF analysis members were available, due to computing constraints, 36-h forecasts were only initialized from members 1–10; 10-member CAEs are sufficient to provide skillful and valuable probabilistic forecasts (e.g., Clark et al. 2011, 2018; Schwartz et al. 2014) and similar in size as the HRRRE and NCEP's operational High-Resolution Ensemble Forecast system (Jirak et al. 2018). Choosing members 1–10 was effectively the same as randomly selecting 10 members since all ensemble members had identical configurations (e.g., Schwartz et al. 2014). In principle, free forecasts could have been initialized every hour, but given finite resources, forecasts were solely initialized at 0000 UTC, which allowed us to focus on both short-term and next-day forecast periods featuring active convection.

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

2.2.3. Benchmark ensemble

To serve as a benchmark for the EnKF-initialized CAE forecasts, 36-h forecasts on the nested grid (Fig. 2.1a) with the configurations in section 2.2.1 were initialized by interpolating 0.5° ICs from perturbation members 1–10 of the GEFS onto the computational domain at 0000 UTC daily between 25 April and 20 May 2017 (inclusive), with LBCs provided by GEFS forecasts identically as in the EnKF-initialized CAEs. As described by Zhou et al. (2017), GEFS ICs were produced by adding 6-h forecast perturbations from a global EnKF DA system (Whitaker and Hamill 2002) to "hybrid" variational-ensemble analyses produced for NCEP's deterministic GFS (e.g., Wang and Lei 2014; Kleist and Ide 2015a,b). Relative to the limited-area EnKF analyses, GEFS ICs were much coarser but reflected assimilation of many more observations, including satellite radiances. Overall, comparison of GEFS- and EnKF-initialized CAE forecasts provides insight about whether the vastly more expensive EnKF initialization procedure was warranted.

2.2.4. BLENDING

Based on performance of the EnKF- and GEFS-initialized CAE forecasts (section 2.4.2), additional ensemble ICs were created by "blending" small scales from EnKF analyses with large scales from GEFS ICs. Blending was solely performed at 0000 UTC between 25 April and 20 May 2017 (inclusive) immediately after EnKF DA and before initializing 36-h CAE forecasts; blending was not employed within the context of continuously cycling EnKF DA, as the blended 0000 UTC fields were not used to initialize 1-h WRF model forecasts that served as priors for the next DA cycle.

Specifically, ICs from corresponding GEFS and EnKF ensemble members were blended on both the 15- and 3-km domains[5] to create new initial ensembles using

$$x^{i}_{\mathrm{blend}} = (EnKF_i - EnKF_{FILT,i}) + GEFS_{FILT,i} \ , \qquad (2.1)$$

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

where x^i_{blend} represents the blended ICs for the ith ensemble member, $EnKF_i$ is the EnKF analysis for the ith member, and $EnKF_{FILT,i}$ and $GEFS_{FILT,i}$ are the low-pass filtered EnKF and GEFS ICs for the ith member, respectively, for i = 1...10. To perform the scale separation, a low-pass, 6th-order implicit tangent filter (e.g., Raymond 1988; Raymond and Garder 1991) as implemented by several studies (e.g., Yang 2005; H. Wang et al. 2014; Hsiao et al. 2015; Feng et al. 2020) and given by

$$H(L) = [1 + \tan^{-6}(\pi\Delta x/L_x)\tan^{6}(\pi\Delta x/L)]^{-1} \tag{2.2}$$

was employed (Fig. 2.4), where Δx is the horizontal grid spacing (either 15 or 3 km), L the wavelength, $H(L)$ the scale-dependent response function, and L_x a specified filter cutoff (km) physically representing the spatial scale (wavelength) where the blended ICs (e.g., x^i_{blend}) had equal contributions from GEFS and EnKF initial states [i.e., when $L = L_x$, $H(L) = 0.5$]. Blending was applied at all 51 vertical levels to zonal and meridional wind components; perturbation geopotential height, potential temperature, and dry surface pressure; and water vapor mixing ratio, and the cutoff length was height- and variable-invariant.

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

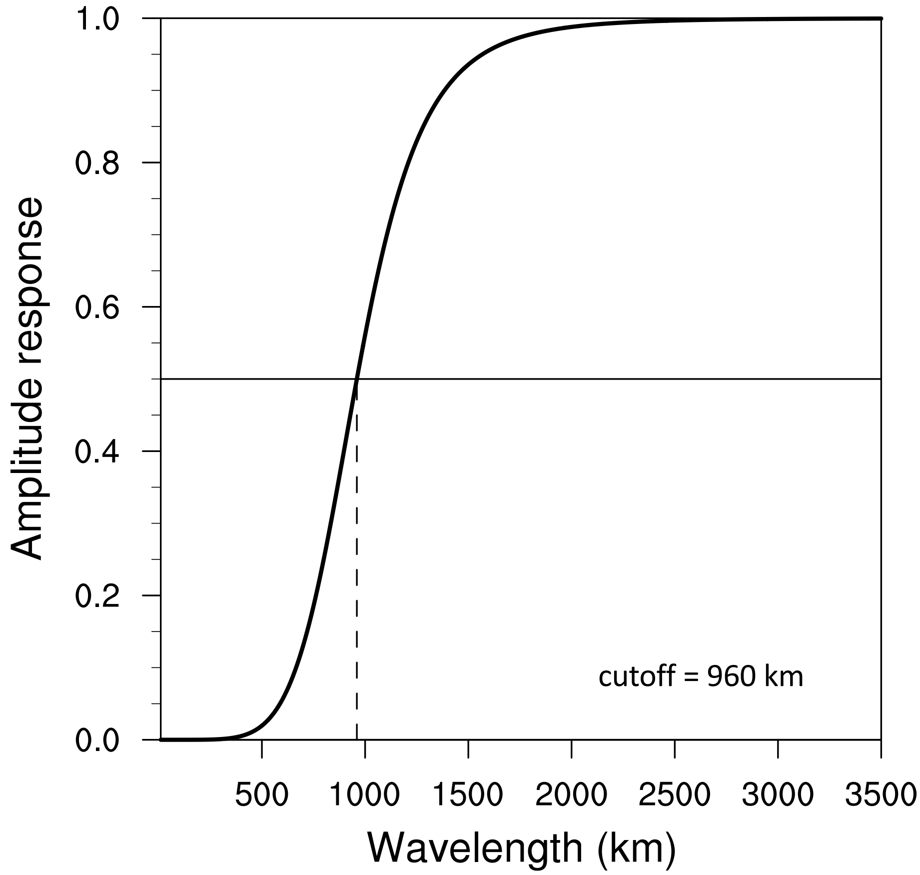

Fig. 2.4. Amplitude response (*y* axis) of a 6th-order implicit tangent filter as a function of wavelength (km) for a specified cutoff length of 960 km. In the context of this study, the curve denotes the contribution of GEFS ICs to blended ICs at a given wavelength (e.g., for wavelengths where the amplitude response is 1, 100% of the blended ICs at those wavelengths were from the GEFS). The dashed vertical and solid horizontal lines illustrate how the amplitude response is 0.5 at the cutoff length.

2.3. EnKF performance

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

Ensemble mean additive biases (model minus observations) and RMSEs aggregated over all prior ensembles (1-h forecasts) between 0000 UTC 25 April and 0000 UTC 20 May 2017 (inclusive) were similar in the 15- and 3-km EnKFs with respect to zonal wind and temperature observations at most levels (Figs. 2.5a,b,d,e),

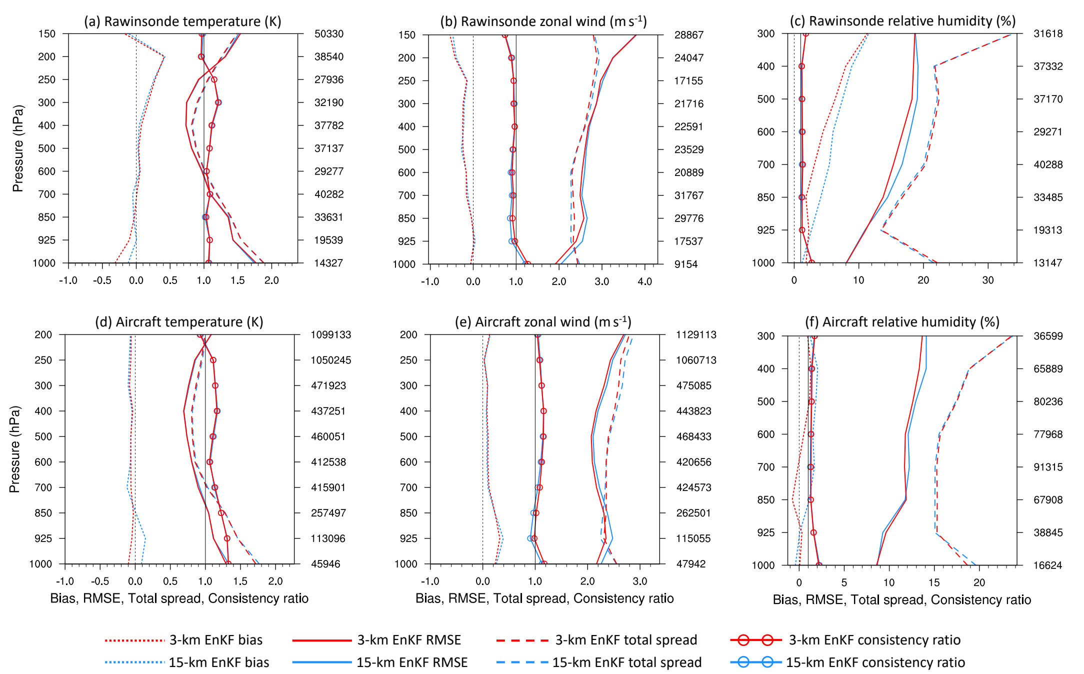

Fig. 2.5. Ensemble mean additive bias (model minus observations; short-dashed lines), ensemble mean RMSE (solid lines), total spread (long-dashed lines), and consistency ratio (CR; solid lines with circles) for (a) rawinsonde temperature (K), (b) rawinsonde zonal wind (m s^{-1}), (c) rawinsonde relative humidity (%), (d) aircraft temperature (K), (e) aircraft zonal wind (m s^{-1}), and (f) aircraft relative humidity (%) observations aggregated over all prior ensembles (1-h forecasts) between 0000 UTC 25 April and 0000 UTC 20 May 2017 (inclusive). These statistics were computed for those observations assimilated by both the 15- and 3-km EnKFs. Sample size at each pressure level is shown at the right of each panel. Vertical lines at *x* = 0 and *x* = 1 are references for biases and CRs, respectively.

while biases and RMSEs for moisture were typically smaller in the 3-km EnKF (Figs. 2.5c,f). Magnitudes of temperature biases were typically < 0.1 K, except near the surface and in the upper troposphere for rawinsonde observations (Fig. 2.5a); the latter is consistent with other continuously cycling EnKFs over the CONUS (e.g., Romine et al. 2013; Schumacher and Clark 2014; Schwartz and Liu 2014; Cavallo et al. 2016; Schwartz 2016) and likely due to closer fits to the more numerous aircraft observations that may have systematically warm biases compared to rawinsonde observations (Ballish and Krishna Kumar 2008). That upper-tropospheric temperature biases relative to aircraft observations (Fig. 2.5d) were smaller than and opposite the sign of temperature biases relative to rawinsonde observations (Fig. 2.5a) further supports this reasoning.

Prior total spreads were similar in both EnKFs (Fig. 2.5) and CRs were usually between 0.8–1.2, although CRs suggest moisture observation errors could potentially be decreased. While more spread may have been expected in the 3-km EnKF because small-scale errors grow rapidly upscale (e.g., Lorenz 1969; Zhang et al. 2003; Hohenegger and Schär 2007), cumulus parameterization in the 15-km DA system may have served as an error source that compensated for missing storm-scale structures, and assimilating copious observations each cycle (Fig. 2.3) with fairly large localization distances highly constrained the 15- and 3-km EnKFs, limiting spread growth during 1-h WRF model integrations between analyses. In balance, these factors potentially contributed to the similar 15- and 3-km prior spreads.

Overall, systematic biases were usually small and EnKF performance appeared acceptable. Moreover, after the first two days, prior total spread and ensemble mean biases were steady throughout the cycles (Fig. 2.6), and observation rejection rates varied little with time (not shown). These results indicate the continuously cycling EnKFs maintained stable climates, which is particularly noteworthy for the 3-km EnKF, as it has not previously been demonstrated that a convection-allowing EnKF can be continuously cycled over a large domain without deleterious consequences like a drifting model climate or filter divergence [see Appendix A of Houtekamer and Zhang (2016) for a succinct summary of filter divergence].

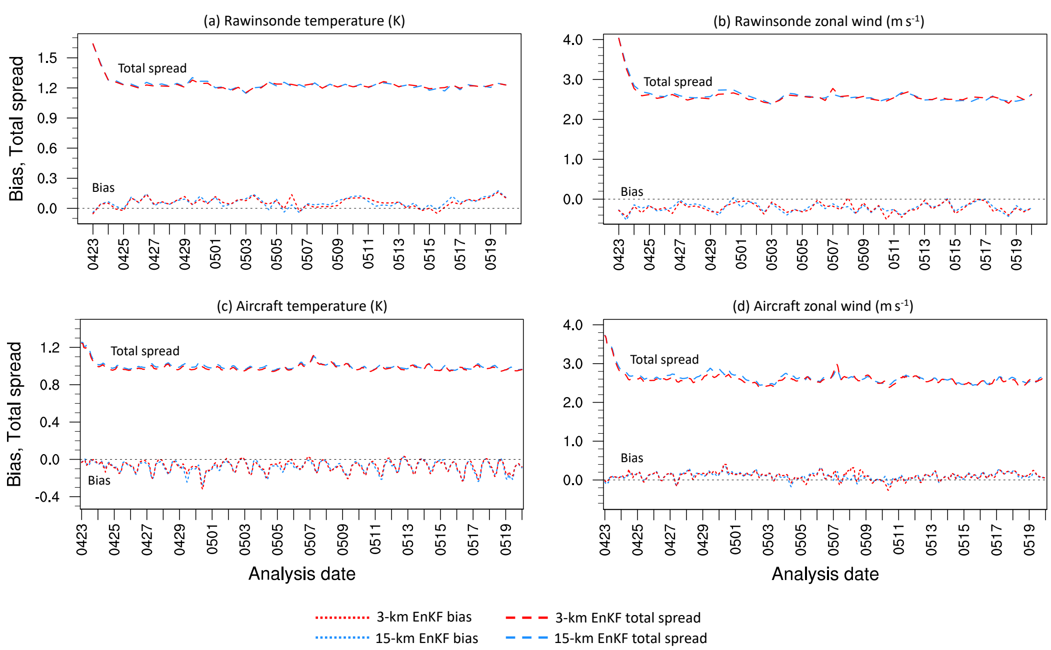

Fig. 2.6. Prior (1-h forecast) total spread (long-dashed lines) and ensemble mean additive bias (model minus observations; short-dashed lines) for (a) rawinsonde temperature (K), (b) rawinsonde zonal wind (m s^{-1}), (c) aircraft temperature (K), and (d) aircraft zonal wind (m s^{-1}) observations between 150–1000 hPa as a function of time. In (c),(d) values are plotted every hour between 0000 UTC 23 April and 0000 UTC 20 May 2017 (inclusive) and smoothed with a 6-h running average, while in (a),(b) values are plotted every 12 h between 0000 UTC 23 April and 0000 UTC 20 May 2017 (inclusive) without smoothing. These statistics were computed for those observations assimilated by both the 15- and 3-km EnKFs. The *x* axis labels represent 0000 UTC for a specific month and day in 2017 (e.g., the marker for "0511" denotes 0000 UTC 11 May 2017). Dashed lines at *y* = 0 are for reference.

2.4. *Precipitation forecast verification*

Hourly-accumulated precipitation forecasts were verified against Stage IV (ST4) analyses (Lin and Mitchell 2005) produced at NCEP considered as "truth". Objective evaluations were performed over the CONUS east of 105°W (hereafter the "verification region"; Fig. 2.1a), where ST4 analyses were most robust (e.g., Nelson et al. 2016). For metrics requiring a common grid for forecasts and observations, we used a budget algorithm (e.g., Accadia et al. 2003) to interpolate forecast precipitation to the ST4 grid (4.763-km horizontal grid spacing). Otherwise, metrics were computed from native grid output.

The following statistics were aggregated over all 26 0000 UTC-initialized 3-km forecasts.

2.4.1. PRECIPITATION CLIMATOLOGIES

To assess precipitation climatologies, aggregate domain-total precipitation per grid point and fractional coverages of 1-h accumulated precipitation meeting or exceeding various accumulation thresholds (e.g., 2.5 mm h^{-1}) were calculated on native grids over the verification region. Additionally, spatial patterns of total precipitation over all 26 forecasts were examined, which were similar in the various ensembles and generally agreed with observations (e.g., Fig. 2.1b), including the SW–NE-oriented maximum across Missouri and adjacent areas. Although magnitudes of these maxima differed across the ensembles, these differences were manifested by the following domain-average statistics, so spatial variations of precipitation climatologies are not discussed further.

2.4.1.1. Impact of analysis resolution

Differences between ensembles were largest over the first 12 h, when GEFS-initialized forecasts were spinning-up precipitation from coarse 0.5° ICs. While this spin-up meant GEFS-initialized forecasts underpredicted total precipitation (Fig. 2.7) and areal coverages (Fig. 2.8) over the first 5 h, ultimately, the spin-up process yielded too much 6–12 h total precipitation and excessive coverages ≥ 2.5 mm h^{-1}. Forecasts initialized from 15-km EnKF analyses also overpredicted total precipitation over the first 12 h, accompanied by excessive coverages for thresholds ≥ 5.0 mm h^{-1}.

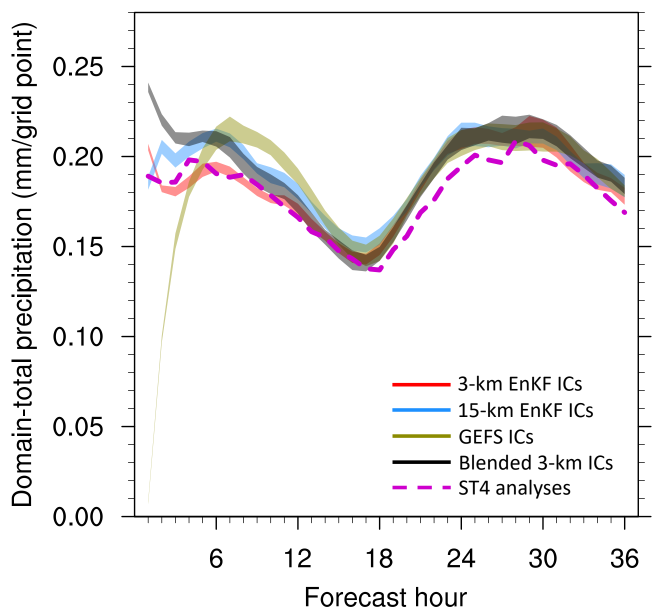

Fig. 2.7. Average 1-h accumulated precipitation (mm) per grid point over all 26 3-km forecasts and the verification region (CONUS east of 105°W) computed on native grids as a function of forecast hour. Red, blue, gold, and black shadings represent envelopes of the 10 members comprising the ensembles with 3-km EnKF ICs, 15-km EnKF ICs, GEFS ICs, and blended 3-km ICs, respectively, and darker shadings indicate intersections of two or more ensemble envelopes. Values on the *x* axis represent ending forecast hours of 1-h accumulation periods (e.g., an *x*-axis value of 24 is for 1-h accumulated precipitation between 23–24 h). ST4 data during the 0–12- and 24–36-h forecast periods were identical except for one day (the former included data between 0000–1200 UTC 25 April – 20 May while the latter instead included data between 0000–1200 UTC 26 April – 21 May), and because domain-total ST4 precipitation between 0000–1200 UTC 21 May was much larger than that between 0000–1200 UTC 25 April, average 24–36-h domain-total ST4 precipitation was greater than average 0–12-h domain-total ST4 precipitation.

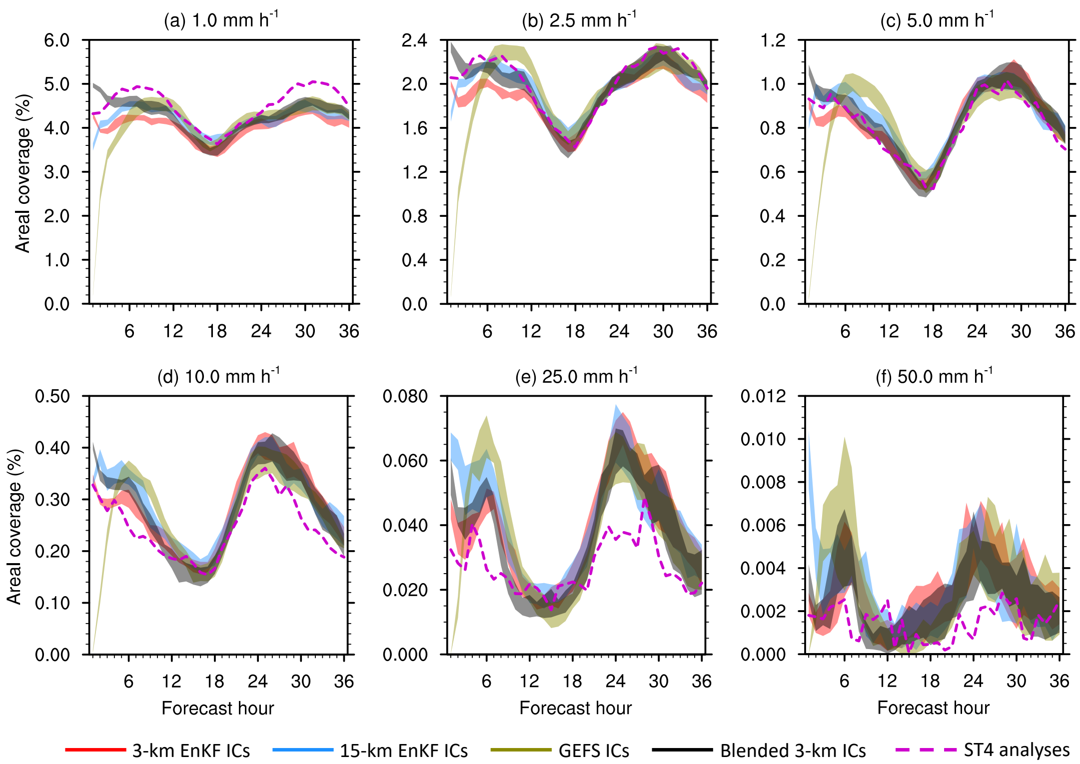

Fig. 2.8. Fractional areal coverage (%) of 1-h accumulated precipitation meeting or exceeding (a) 1.0, (b) 2.5, (c) 5.0, (d) 10.0, (e) 25.0, and (f) 50.0 mm h^{-1} over the verification region (CONUS east of 105°W), computed on native grids and aggregated over all 26 3-km forecasts as a function of forecast hour. Red, blue, gold, and black shadings represent envelopes of the 10 members comprising the ensembles with 3-km EnKF ICs, 15-km EnKF ICs, GEFS ICs, and blended 3-km ICs, respectively, and darker shadings indicate intersections of two or more ensemble envelopes. Values on the *x* axis represent ending forecast hours of 1-h accumulation periods (e.g., an *x*-axis value of 24 is for 1-h accumulated precipitation between 23–24 h).

Overall, forecasts initialized from unblended 3-km EnKF analyses had precipitation climatologies best matching observations through 12 h, but there were shortcomings. For example, although at 1 h, unblended 3-km EnKF analyses produced forecasts with areal coverages closest to observations (Fig. 2.8), coverages rapidly decreased between 2–3 h and were further from those observed between 2–12 h for the 1.0 and 2.5 mm h^{-1} thresholds (Figs. 2.8a,b) compared to forecasts with 15-km or blended 3-km ICs, suggesting poor maintenance of stratiform precipitation regions after initialization. However, forecasts with unblended 3-km ICs had 6–12-h areal coverages at the 5.0 mm h^{-1} threshold well-matching observations (Fig. 2.8c) and 2–6-h coverages at the 10.0–50.0 mm h^{-1} thresholds closer to observations than forecasts with GEFS and 15-km EnKF ICs (Figs. 2.8d–f). Furthermore, 2–12-h domain-total precipitation was clearly best in forecasts with unblended 3-km ICs (Fig. 2.7).

Despite differences between the ensembles through 12 h, domain-total precipitation and areal coverages were broadly similar between 18–36 h, with too much total precipitation (Fig. 2.7) and general underprediction and overprediction of areal coverages at the 1.0 and 10.0–50.0 mm h^{-1} thresholds, respectively (Figs. 2.8a,d–f). Collectively, for precipitation climatologies, these findings suggest benefits of convection-allowing analyses relative to convection-parameterizing analyses are primarily confined to short-term forecasts and heavier rainfall rates.

2.4.1.2. Impact of blending

With respect to forecasts initialized from unblended 3-km EnKF analyses, forecasts with blended 3-km ICs (using a 960-km cutoff) had similar 18–36-h areal coverages and total precipitation but higher domain-total precipitation and areal coverages over the first 6–12 h that typically compared worse to observations through 3 h (Figs. 2.7, 2.8). Examination of individual forecasts indicated blended 3-km ICs mostly enhanced 1–3-h forecast precipitation within and near precipitation entities also predicted by forecasts with unblended 3-km ICs and that widespread spurious features did not cause the overprediction. This behavior is illustrated by the forecast initialized at 0000 UTC 1 May 2017, which had the largest difference of domain-total precipitation (e.g., Fig. 2.7) between member 1 in the CAEs with blended and unblended 3-km ICs across all 26 36-h forecasts (Fig. 2.9). While both 1–3-h precipitation forecasts had similar spatial patterns, blended ICs led to more numerous cells in places with scattered rainfall, and these additional entities were usually erroneous compared to observations (black and gold circles in Fig. 2.9). Additionally, within features, the forecast with blended ICs had heavier rainfall maxima than ST4 observations and the forecast with unblended ICs (red circles in Figs. 2.9b,c,e,f,h,i).

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

2.4.2. Ensemble precipitation verification

As in many studies, we used percentile thresholds to define events (e.g., the 95th percentile, which selects the top 5% of values), which removes bias and permits a thorough assessment of spatial performance given a model's climate (e.g., Roberts and Lean 2008; Mittermaier and Roberts 2010; Mittermaier et al. 2013; Dey et al. 2014; Gowan et al. 2018; Woodhams et al. 2018; Schwartz 2019). Our application of percentile thresholds exactly followed section 5a(1) of Schwartz (2019), where physical thresholds corresponding to percentile thresholds were obtained separately

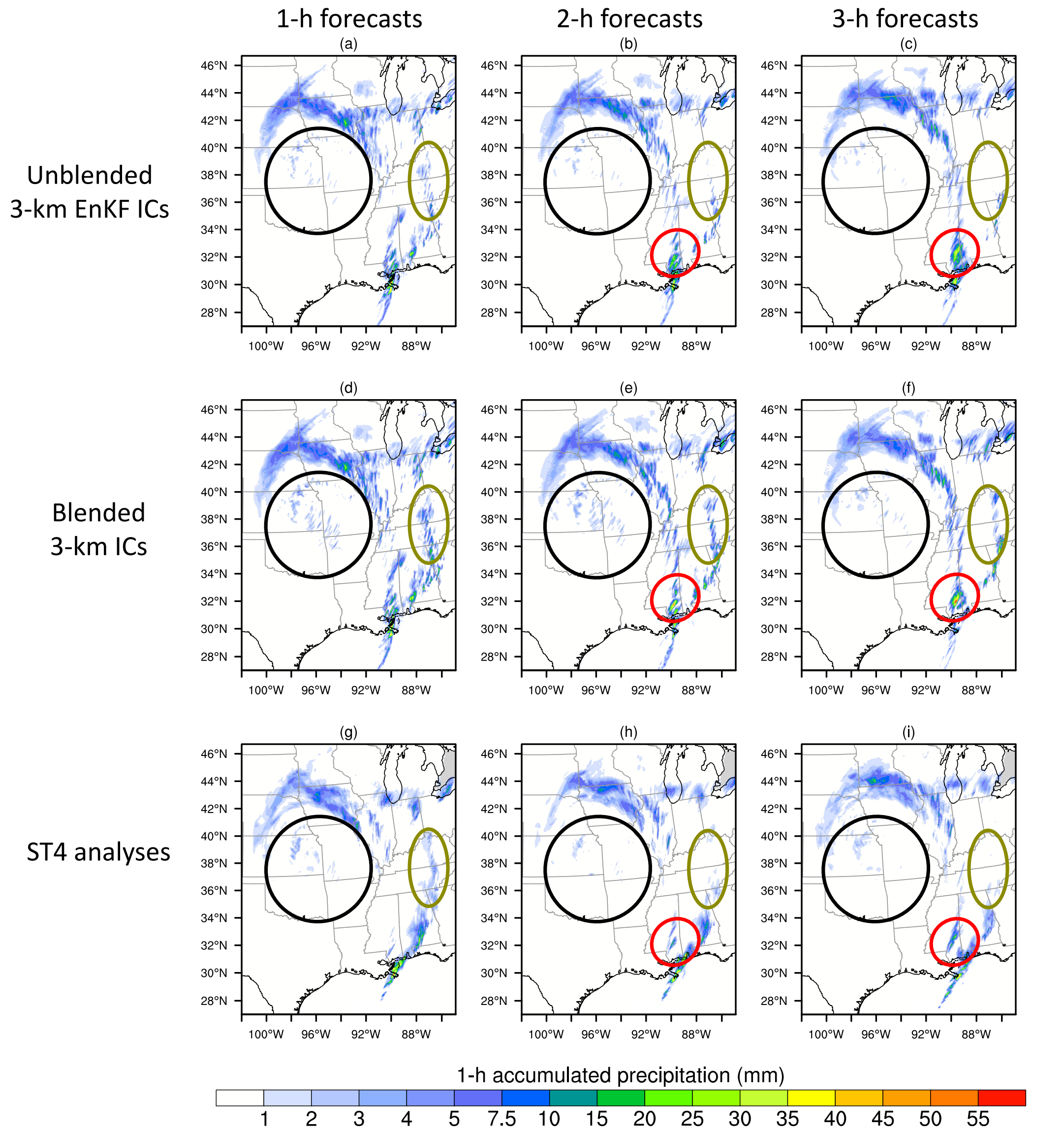

Fig. 2.9. 1-h accumulated precipitation (mm) for (a,d,g) 1-, (b,e,h) 2-, and (c,f,i) 3-h forecasts initialized at 0000 UTC 1 May 2017 from member 1 of the 3-km ensembles with (a)–(c) unblended 3-km EnKF ICs and (d)–(f) blended 3-km ICs (using a 960-km cutoff length). (g)–(i) Corresponding ST4 analyses, with grey-shaded areas denoting no data. Annotated circles correspond to features noted in the text.

for observations and each ensemble member on the ST4 grid for each precipitation accumulation interval. These physical thresholds were ultimately used to determine forecast and observed event occurrence. To help interpret subsequent objective statistics, mean physical thresholds corresponding to specific percentile thresholds are provided in Fig. 2.10. As with areal coverages (Fig. 2.8), the largest differences amongst the ensembles' percentiles were over the first 6–12 h.

After interpolating precipitation forecasts to the ST4 grid, a "neighborhood approach" (e.g., Theis et al. 2005; Ebert 2008, 2009) was used to produce "neighborhood ensemble probabilities" (NEPs; Schwartz et al. 2010; Schwartz and Sobash 2017) that were ultimately verified. In short, NEPs were computed at the *i*th grid point by averaging point-based ensemble probabilities over all grid points within the neighborhood of the *i*th point, which incorporates spatial uncertainty and reflects

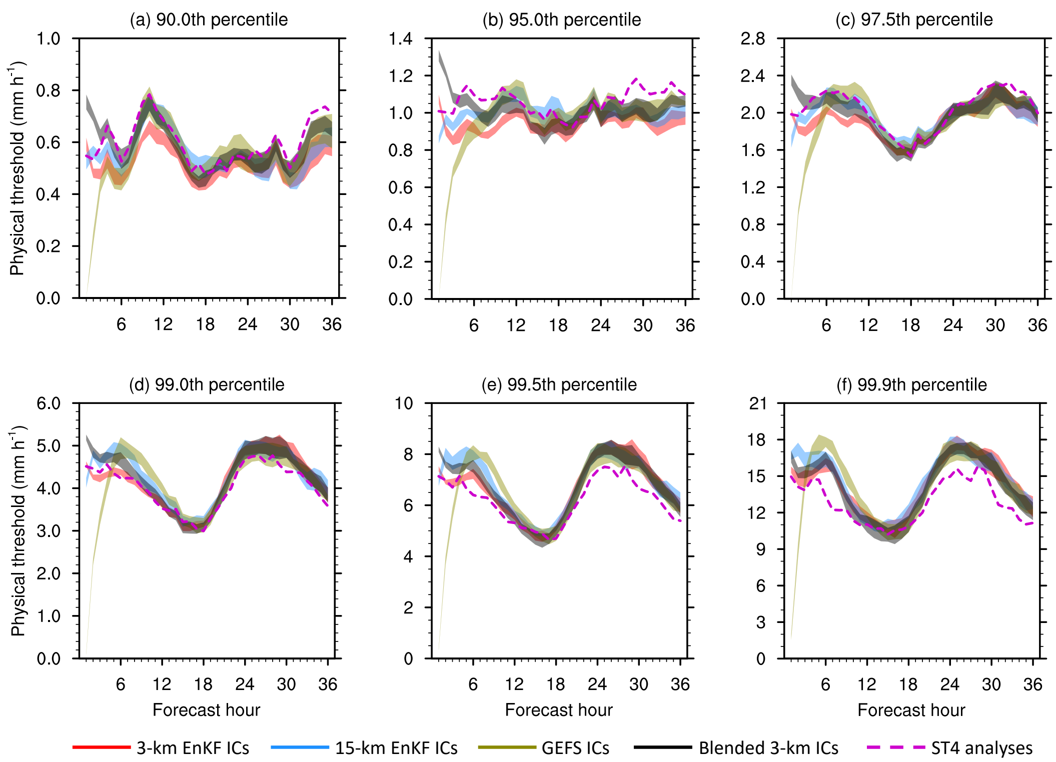

Fig. 2.10. Average physical thresholds (mm h^{-1}) over all 26 3-km forecasts of 1-h accumulated precipitation corresponding to the (a) 90th, (b) 95th, (c) 97.5th, (d) 99th, (e) 99.5th, and (f) 99.9th percentile thresholds as a function of forecast hour. The physical thresholds were computed separately for each day and 1-h forecast period on the ST4 grid over the verification region (CONUS east of 105°W) and averaged to obtain the *y*-axis values. Red, blue, gold, and black shadings represent envelopes of the 10 members comprising the ensembles with 3-km EnKF ICs, 15-km EnKF ICs, GEFS ICs, and blended 3-km ICs, respectively, and darker shadings indicate intersections of two or more ensemble envelopes. Values on the *x* axis represent ending forecast hours of 1-h accumulation periods (e.g., an *x*-axis value of 24 is for 1-h accumulated precipitation between 23–24 h).

the inherent inaccuracy of high-resolution NWP models at individual grid points. We produced NEPs for neighborhood length scales (*r*) between 5 and 200 km, which represented radii of circular neighborhoods. Please see section 2a of Schwartz and Sobash (2017) for more information about constructing and verifying NEPs and Eqs. (1)–(3) in Schwartz (2019), which explicitly describe NEP computation when using percentile thresholds.

Statistical significance testing followed section 5a(3) of Schwartz (2019). Specifically, a pairwise difference bootstrap technique with 10,000 resamples was used to determine whether aggregate differences between two ensembles' statistics were statistically significant at the 95% level (e.g., Hamill 1999; Wolff et al. 2014).

2.4.2.1. Attributes statistics and rank histograms

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

Over the first 12 h for *r* = 90 km, the ensemble initialized from unblended 3-km EnKF analyses was statistically significantly more reliable than the ensembles initialized from GEFS and 15-km EnKF ICs, with the GEFS-initialized ensemble having the worst reliability (Fig. 2.11). Conversely, between 18–36-h for *r* = 125 km,

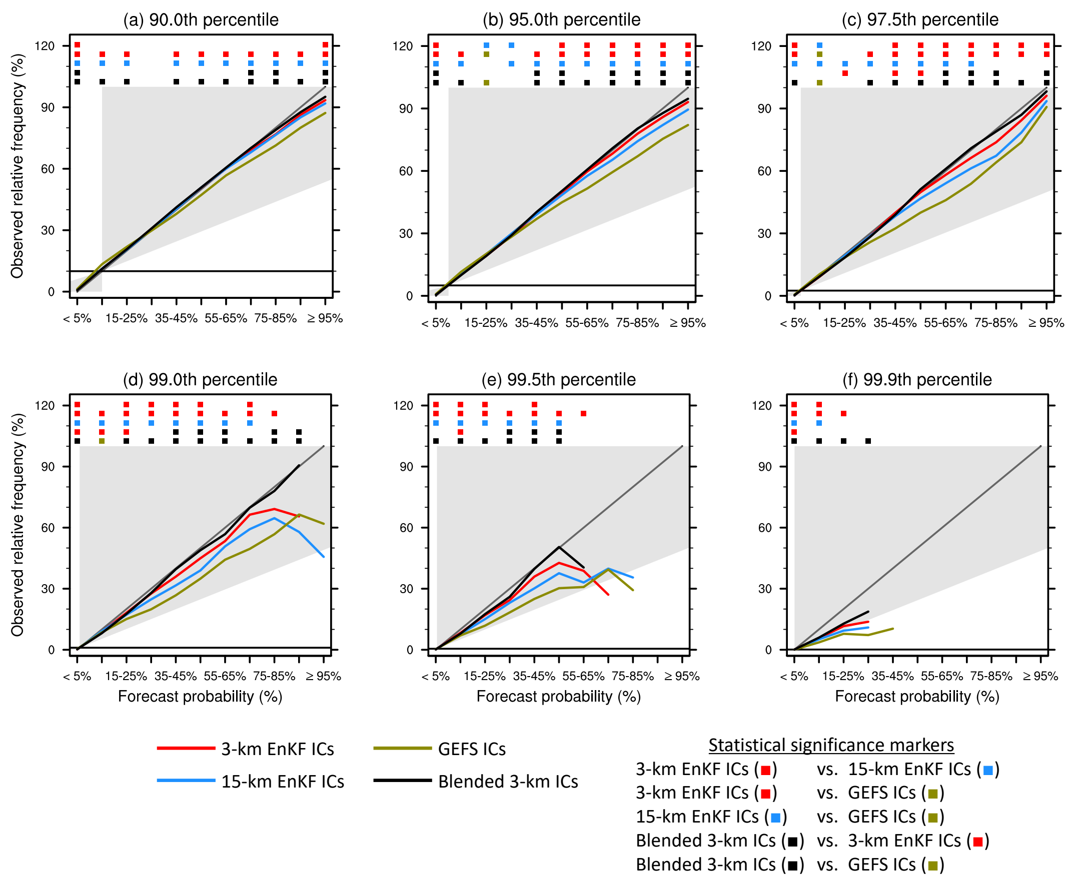

Fig. 2.11. Attributes statistics computed over the verification region (CONUS east of 105°W) with a 90-km neighborhood length scale aggregated over all 26 1–12-h 3-km forecasts of 1-h accumulated precipitation for the (a) 90th, (b) 95th, (c) 97.5th, (d) 99th, (e) 99.5th, and (f) 99.9th percentile thresholds. Horizontal lines near the *x* axis represent observed frequencies of the event (sample climatology) and diagonal lines are lines of perfect reliability. Points lying in grey-shaded regions had skill compared to forecasts of sample climatology as measured by the Brier skill score (Brier 1950; Wilks 2011). Values were not plotted for a particular bin if fewer than 500 grid points had forecast probabilities in that bin over the verification region and all 26 forecasts. Symbols along the top axis denote those probability bins where differences between two ensembles were statistically significant at the 95% level, with the five rows of colored symbols corresponding to the five comparisons in the legend to denote which ensemble was statistically significantly closest to perfect reliability. For example, in the top row, red symbols indicate the ensemble with 3-km EnKF ICs had statistically significantly better reliability than the ensemble with 15-km EnKF ICs, while blue symbols indicate the ensemble with 15-km EnKF ICs had statistically significantly better reliability than the ensemble with 3-km EnKF ICs. Absence of a symbol means the differences were not statistically significant at the 95% level. Note that the attributes diagrams themselves stop at 100%; area above 100% was added to make room for statistical significance markers.

the GEFS-initialized ensemble was regularly statistically significantly more reliable than the ensembles with unblended 15- and 3-km EnKF ICs, and the ensemble with

15-km ICs usually had comparable or better reliability than the ensemble with unblended 3-km ICs (Fig. 2.12). Except for the 99.9% threshold, all ensembles had skill with respect to forecasts of sample climatology.

These findings suggest aspects of GEFS ICs were beneficial for next-day (18–36-h) forecasts, which motivated blending GEFS and EnKF initial states. Indeed, blended 3-km ICs led to 18–36-h forecasts with comparable or better reliability as GEFS-initialized forecasts and statistically significantly better reliability than the ensemble with unblended 3-km ICs (Fig. 2.12). Over the first 12 h, differences between the ensembles with blended and unblended 3-km ICs were also often

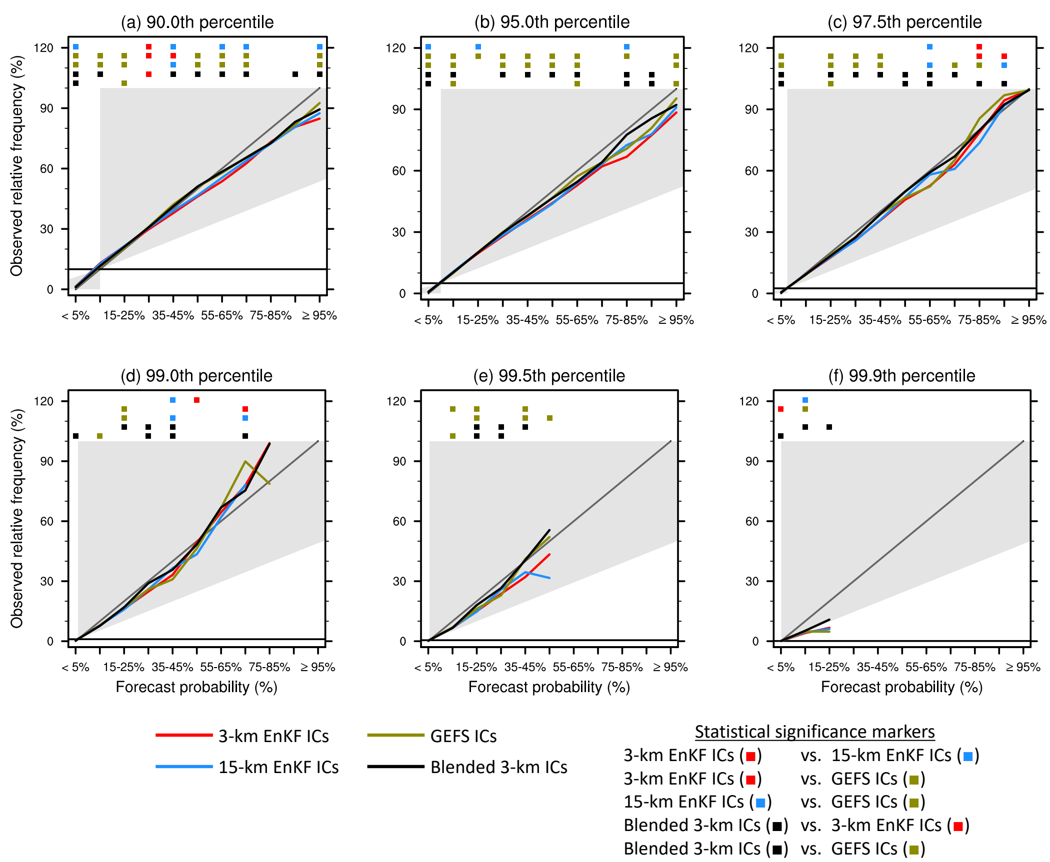

Fig. 2.12. As in Fig. 2.11 except statistics were aggregated over all 26 18–36-h 3-km forecasts of 1-h accumulated precipitation using a 125-km neighborhood length scale.

statistically significant, suggesting that blending can additionally improve short-term forecast reliability (Fig. 2.11).

Rank histograms (e.g., Hamill 2001) based on domain-total precipitation (e.g., Schwartz et al. 2014, 2020) corroborated attributes statistics. Specifically, over the first 12 h, bin counts in the ensemble with unblended 3-km ICs were closer to optimal in most bins compared to those for the ensembles with GEFS and 15-km EnKF ICs (Fig. 2.13a), which was quantified by the smaller-is-better reliability index (RI; Delle Monache et al. 2006). Blended 3-km ICs yielded slightly lower 1–12-h RIs than unblended 3-km ICs, but the difference was small compared to that between 18–36 h (Fig. 2.13b), where rank histograms and RIs indicated more observations fell within the ensemble and dispersion was improved when GEFS initial states were either used as standalone ICs or combined with 3-km EnKF analyses through blending.

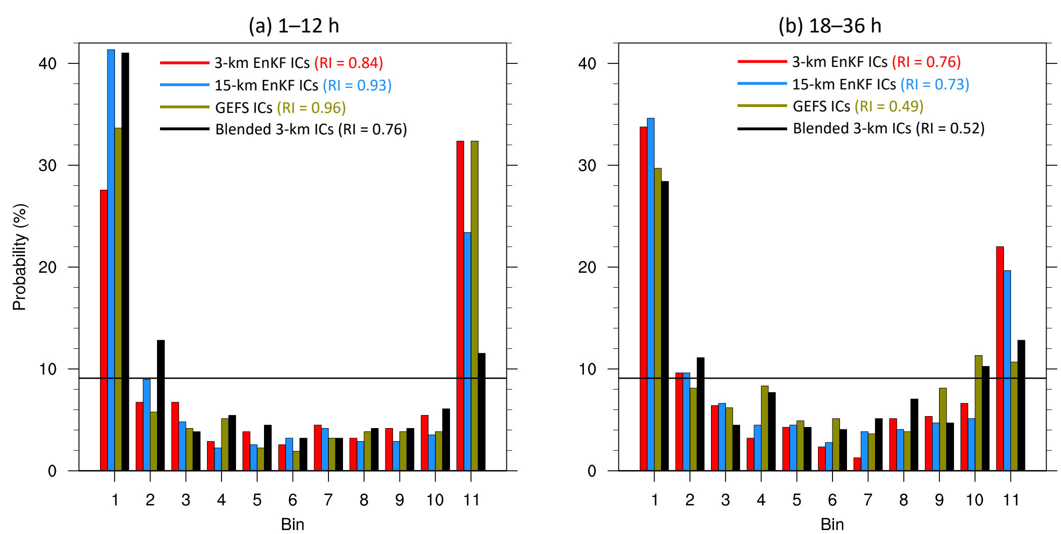

Fig. 2.13. Rank histograms containing all 26 3-km (a) 1–12- and (b) 18–36-h forecasts of domain-total 1-h accumulated precipitation on the ST4 grid over the verification region (CONUS east of 105°W) for the various ensembles. Horizontal lines are optimal values, and the reliability index (RI; Delle Monache et al. 2006) is annotated for each ensemble in the legend; lower values are better and indicate flatter rank histograms.

2.4.2.2. Spread and spectra

Improved reliability and rank histograms engendered by GEFS and blended 3-km ICs was associated with increased ensemble spread. In particular, the ensembles with GEFS and blended 3-km ICs had statistically significantly more 24–30-h precipitation spread compared to the ensembles with unblended EnKF ICs (Fig. 2.14).

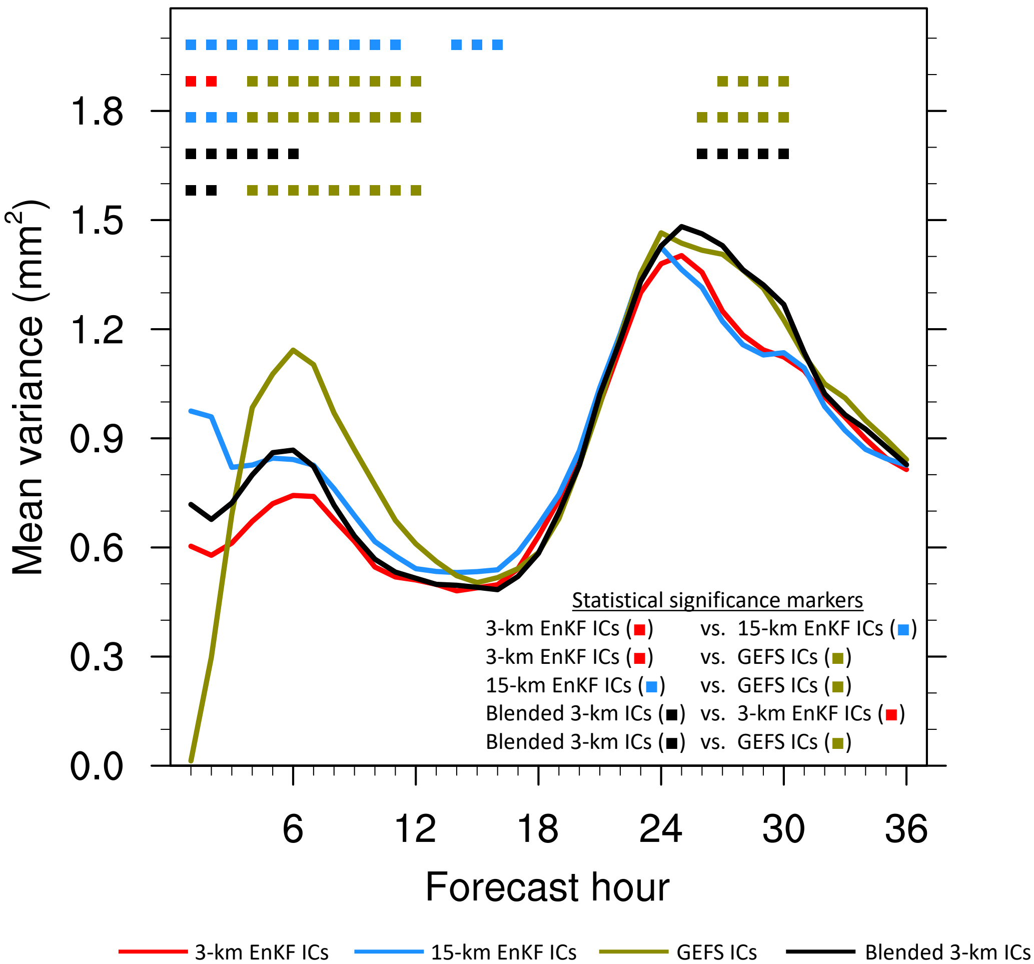

Fig. 2.14. Average ensemble variance (mm^2) over the verification region (CONUS east of 105°W) and all 26 3-km forecasts of 1-h accumulated precipitation as a function of forecast hour. Values on the *x* axis represent ending forecast hours of 1-h accumulation periods (e.g., an *x*-axis value of 24 is for 1-h accumulated precipitation between 23–24 h). Symbols along the top axis indicate forecast hours when differences between two ensembles were statistically significant at the 95% level as in Fig. 2.11 and denote the ensemble with statistically significantly higher variance.

Additionally, blended 3-km ICs led to significantly more spread than unblended 3-km ICs over the first 6 h that may have improved reliability statistics and rank histograms, even though this enhanced spread reflected excessive early precipitation (e.g., Figs. 2.7–2.9). The greater spread through ~18 h in the ensembles with GEFS and 15-km ICs relative to that from the ensemble with unblended 3-km ICs may reflect a substantial contribution from the small, yet intense precipitation entities that were more numerously predicted when forecasts had downscaled, rather than 3-km, ICs (Figs. 2.8c–f, 2.10d–f).

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

At 1 h, 500-hPa perturbation kinetic energy (PKE) in the ensemble with blended 3-km ICs broadly followed PKEs of the GEFS-initialized ensemble at scales > 500 km and the ensemble with unblended 3-km ICs at smaller scales, reflecting the blending procedure (Fig. 2.15a). Compared to the GEFS-initialized ensemble, the ensemble with unblended 3-km ICs had more 1-h forecast PKE at most scales (Fig. 2.15a), with enhanced large-scale power possibly reflecting upscale error growth with time through the continuous 3-km DA cycles. But, PKE in the GEFS-initialized ensemble grew fastest between 3–6 h (Figs. 2.15b,c) and was largest at all scales after 6 h (Figs. 2.15d–f), while unblended 3-km ICs yielded the least 12–36-h PKE at

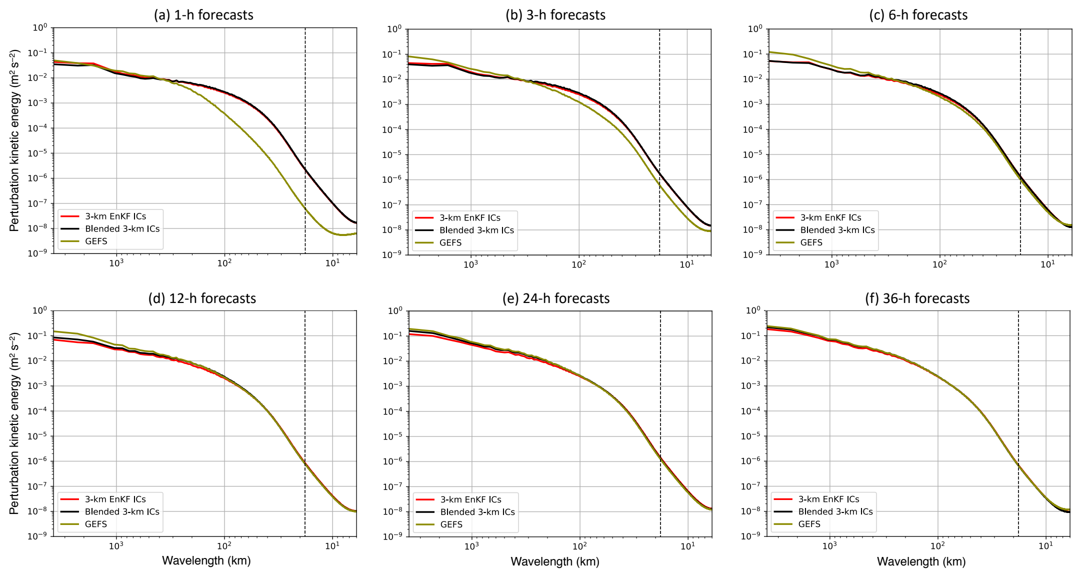

Fig. 2.15. Average 500-hPa perturbation kinetic energy (m^2 s^{-2}) as a function of wavelength (km) computed from all 26 (a) 1-, (b) 3-, (c) 6-, (d) 12-, (e) 24-, and (f) 36-h 3-km, 10-member ensemble forecasts over the entire 3-km domain, excluding the 15 grid points nearest each lateral boundary. Perturbations were computed with respect to the ensemble mean, and the spectra were averaged over all 10 perturbations and 26 forecasts. Dashed vertical lines denote 6 times the horizontal grid spacing (3 km), the approximate effective resolution of the forecasts (Skamarock 2004). The discrete cosine transform was used to perform the spectral analysis and spectral variance binning employed the method of Ricard et al. (2013).

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

2.4.2.3. Fractions skill scores

Forecast skill was further evaluated with the fractions skill score [FSS; Roberts and Lean (2008)], where FSS = 1 indicates a perfect forecast and FSS = 0 means no skill. We present FSSs for r = 100 km, although conclusions were unchanged when FSSs were computed with different neighborhood length scales. Moreover, areas under the relative operating characteristic curve (Mason 1982; Mason and Graham 2002) provided identical conclusions as FSSs and are not discussed further.

Forecasts initialized from unblended 3-km EnKF analyses had higher FSSs than those initialized from downscaled 15-km EnKF analyses through 6–12 h, both when aggregated over all forecasts (Fig. 2.16) and on an hour-by-hour basis (Figs. 2.17a–d), with many instances of significant differences. However, after 6–12 h, the ensembles with unblended 15- and 3-km EnKF ICs usually had statistically indistinguishable FSSs. Compared to the GEFS-initialized ensemble, the unblended EnKF-initialized ensembles had statistically significantly higher aggregate FSSs through 12–18 h but comparable or lower aggregate FSSs thereafter (Fig. 2.16), similar to attributes statistics. These 1–12-h forecast benefits from unblended 3-km EnKF ICs compared to GEFS ICs were evident for most hourly forecasts (Figs. 2.17i–l), while individual 1-h accumulated precipitation forecasts over the 18–36-h period from the ensemble with GEFS ICs were frequently comparable to or better than those from the ensemble with unblended 3-km EnKF ICs (Figs. 2.17m–p).

Blended 3-km ICs led to FSSs mirroring those from unblended 3-km EnKF ICs over the first 12–18 h (Fig. 2.16), indicating blending preserved short-term forecast benefits of increased analysis resolution for spatial placement. Furthermore, after 18–24 h, the ensemble with blended 3-km ICs had higher FSSs than the ensemble with unblended 3-km EnKF ICs both on an hourly basis (Figs. 2.17e–h) and in aggregate that were similar to or higher than FSSs from the GEFS-initialized ensemble.

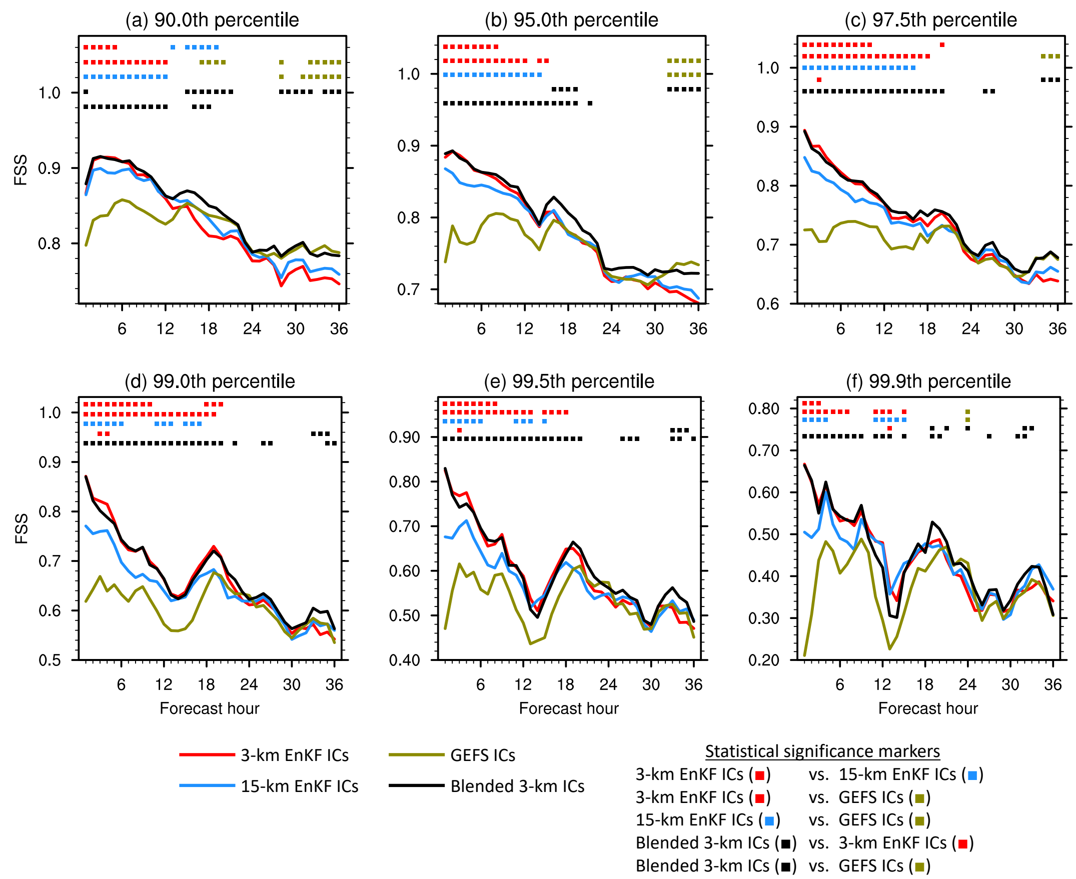

Fig. 2.16. Fractions skill scores (FSSs) over the verification region (CONUS east of 105°W) with a 100-km neighborhood length scale for the (a) 90th, (b) 95th, (c) 97.5th, (d) 99th, (e) 99.5th, and (f) 99.9th percentile thresholds aggregated over all 26 3-km forecasts of 1-h accumulated precipitation 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 23–24 h). The *y* axis scales are different in each panel. Symbols along the top axis indicate forecast hours when differences between two ensembles were statistically significant at the 95% level as in Fig. 2.11 and denote the ensemble with statistically significantly higher FSSs. Note that the maximum FSS is 1.0; area above 1.0 was added to make room for statistical significance markers.

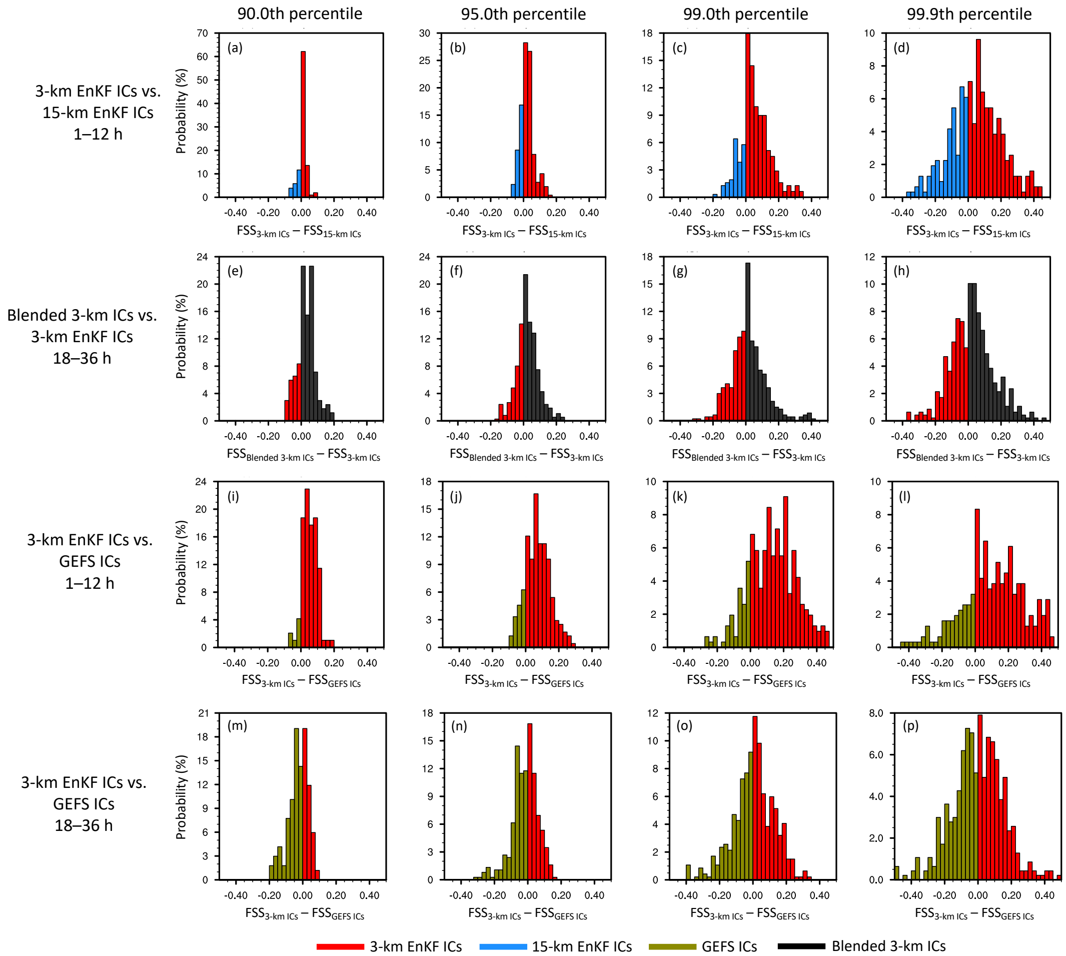

Fig. 2.17. (a)–(d) Histogram [expressed as probabilities (%)] of FSS differences with *r* = 100 km between the ensembles with 3-km EnKF ICs and 15-km EnKF ICs (3-km ICs minus 15-km ICs) computed from all 26 0–1-, 1–2-, …, 10–11-, and 11–12-h 3-km forecasts of 1-h accumulated precipitation for the (a) 90th, (b) 95th, (c) 99th, and (d) 99.9th percentile thresholds. (e)–(h) As in (a)–(d) but for differences from all 26 18–19-, 19–20-, …, 34–35-, and 35–36-h 3-km forecasts of 1-h accumulated precipitation between the ensembles with blended and unblended 3-km ICs (blended 3-km ICs minus unblended 3-km ICs). (i)–(l) and (m)–(p) As in (a)–(d) and (e)–(h) respectively, but for differences between the ensembles with unblended 3-km EnKF and GEFS ICs (3-km ICs minus GEFS ICs). Values on the *x* axis denote the leftmost points of each bin, and bin widths were 0.025 (e.g., the bars with left edges at 0.05 are for bins spanning 0.05–0.075). Colors of the bars correspond to the legend and indicate the experiment with the higher FSS in that bin.

2.4.2.4. Synthesis

FSSs, attributes statistics, and rank histograms revealed clear benefits of convection-allowing analyses compared to convection-parameterizing analyses for 1–12-h precipitation forecasts, consistent with previous work (e.g., Johnson et al. 2015;

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

2.4.3. Impact of hourly radar DA

Because our 3-km EnKF was highly constrained, we wondered whether assimilating radar observations could realize meaningful analysis and forecast improvements. So, to assess the impact of assimilating radar reflectivity observations, another EnKF was configured exactly as the nested 15-/3-km EnKF DA system (section 2.2.2), except reflectivity observations throughout the CONUS were assimilated into 3-km analyses along with conventional observations hourly from 1900–0000 UTC. Although reflectivity observations could easily be assimilated more frequently in our framework, hourly radar DA mimics the HRRRE configuration. Backgrounds for 1900 UTC radar-assimilating EnKF analyses were provided by 1-h forecasts initialized from 1800 UTC posterior ensembles from the nested 15-/3-km EnKF assimilating solely conventional observations. Thus, the impact of assimilating reflectivity observations was confined to a 6-h period each day. This approach was adopted primarily to avoid the expense of continuously cycling another 3-km EnKF over the entire 4-week period. However, assimilating radar observations for just a few hours was methodologically consistent with numerous other high-resolution DA systems, including the WoF system (e.g., Wheatley et al. 2015; Jones et al. 2016; Skinner et al. 2018), and 6 h of assimilating radar observations was more than sufficient to assess the data impact (e.g., Johnson and Wang 2017 and references therein).

Specific radar DA configurations mostly followed Duda et al. (2019) and references therein (Table 2.4), and like the other EnKFs, 0000 UTC analysis ensembles initialized 36-h, 3-km, 10-member CAE forecasts. Furthermore, to examine the interplay of blending and radar DA, we also created a set of ICs by blending GEFS ICs with radar-assimilating 3-km EnKF analyses using a 960-km filter cutoff.

Assimilating reflectivity observations generally improved FSSs over the first 3 h but had small impacts thereafter (Fig. 2.18), similar to other studies finding short-lived benefits of radar DA (e.g., Kain et al. 2010; Johnson et al. 2015; Fabry and Meunier 2020). Within the radar-assimilating experiments, blending boosted FSSs at later times, as with the non-radar DA experiments. Assimilating reflectivity observations negligibly impacted attributes statistics, although assimilation of 100,000–200,000 radar observations each cycle lessened precipitation spread over the first hour (not shown). While more frequent assimilation cycles could potentially realize additional improvements from radar DA, it is unlikely that the small-scale information from radar observations can consistently yield forecast improvements

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

Table 2.4. Settings for assimilation of radar reflectivity observations.

after the shortest forecast ranges, especially in an EnKF highly constrained by other observations. Nonetheless, these experiments suggest feasibility of performing radar-assimilating, WoF-like analyses over large domains in a continuously cycling EnKF DA framework.

2.5. Summary and conclusions

EnKF DA systems with 80 members and 15- and 3-km horizontal grid spacings were continuously cycled with a 1-h period for 4 weeks over a computational domain spanning the entire CONUS. Both the 15- and 3-km EnKFs had stable climates throughout the cycling period and acceptable prior observation-space statistics, demonstrating the viability of a convection-allowing continuously

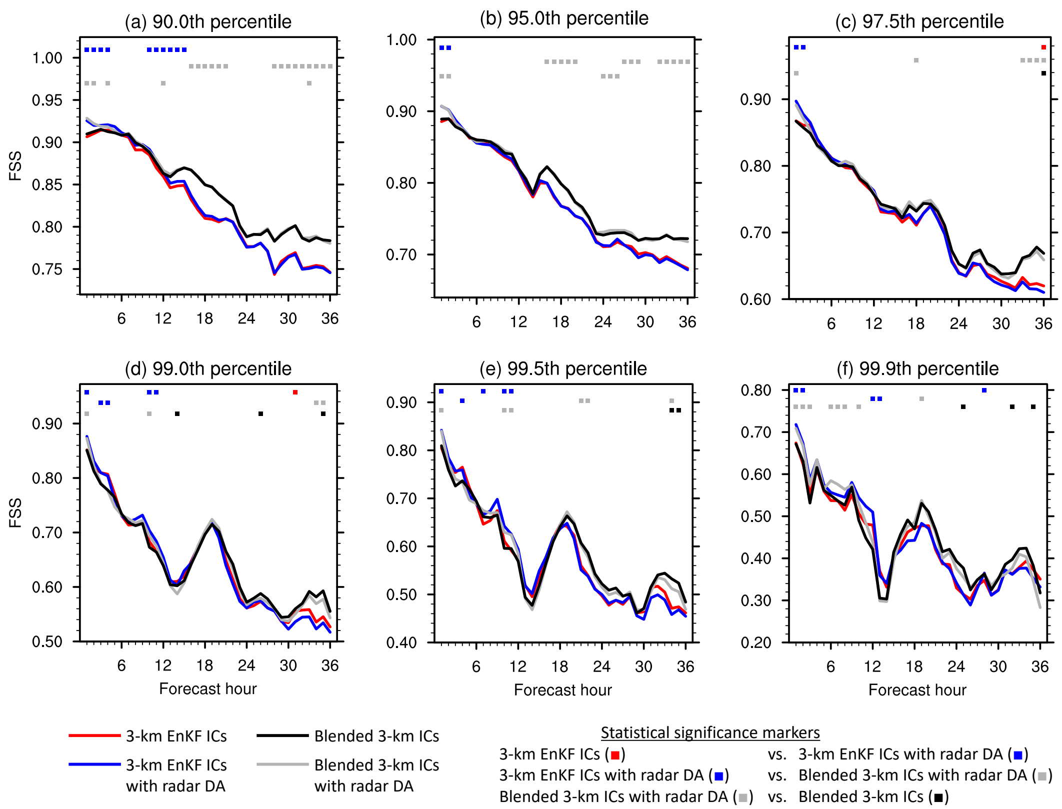

Fig. 2.18. As in Fig. 2.16, but for comparisons focusing on the impact of assimilating radar reflectivity observations and aggregated over 24, rather than 26, forecasts (necessitated due to missing radar observations that precluded radar data assimilation sensitivity experiments for the forecasts initialized at 0000 UTC 13 and 14 May 2017). Grey curves are often beneath the black curves, especially in (a) and (b).

cycling EnKF over the CONUS. However, our EnKFs were highly constrained by observations, and whether convection-allowing EnKFs can be continuously cycled without deleterious consequences over large data-sparse domains is unclear.

At 0000 UTC, EnKF analyses initialized 36-h, 10-member CAE forecasts with 3-km horizontal grid spacing that were evaluated with a focus on precipitation. CAE forecasts were also initialized from NCEP's operational GEFS and "blended" ICs produced by using a low-pass filter to combine large scales from GEFS ICs with small scales from EnKF analysis members. Precipitation forecasts initialized from continuously cycling EnKF analyses outperformed GEFS-initialized forecasts through 12–18 h, and benefits from initializing 3-km forecasts from corresponding 3-km analyses, rather than downscaled 15-km analyses, were realized through 6–12 h. But, after 18 h, GEFS-initialized forecasts were comparable to or better than EnKF-initialized forecasts, indicating limitations of limited-area continuously cycling EnKFs as initialization tools for next-day CAE precipitation forecasts, consistent with Schwartz et al. (2020). Benefits of assimilating radar reflectivity observations into the 3-km EnKF were confined solely to 1–3-h forecasts.

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

There are many avenues for additional research and improvements. For example, while using identical inflation factors and observation errors in the 15- and 3-km EnKFs provided reasonable results, these choices may have been suboptimal. In particular, because observation errors are the sum of measurement and representativeness errors and representativeness errors are resolution-dependent (e.g., Ben Bouallegue et al. 2020), observation errors should arguably be tuned for each domain, which, in turn, might require adjusting inflation factors. Thus, it may be possible to further improve our 3-km EnKF.

Additionally, our blending procedure did not impact the continuously cycling EnKF DA systems, and future work might assess whether incorporating large scales from global analyses into hourly limited-area DA cycles is beneficial. Furthermore, blending could potentially be optimized by dynamically determining the filter cutoff scale (e.g., Feng et al. 2020) or using height- and variable-specific cutoffs (e.g., Zhang et al. 2015), and efforts to mitigate blending-induced initial imbalances tailored for high-resolution models are needed. Moreover, next-day forecast benefits of blending suggest further exploring the value of mixed-resolution ensemble-based DA systems for convective applications may be worthwhile. Also, blending and partial cycling DA approaches should be compared; while both methods introduce external large-scale information into limited-area ICs, whether either method is preferable is unclear. It is also important to note that our blending methodology [Eq. (2.1)] changed the large-scale component of both the IC perturbations and the initial ensemble mean state, differing from an approach of blending *perturbations* derived from two different ensembles without changing the spectral representation of the initial ensemble mean (e.g., Caron 2013). Therefore, we cannot determine whether the 18–36-h forecast improvements from blending were due to altering the large-scale IC perturbations or large-scale initial ensemble mean, and it would be interesting to refine attribution in future work.

Finally, computing availability limited our cycling period to just 4 weeks, and additional experimentation is needed over longer periods, different seasons, and varied geographic regions to further understand large-domain convection-allowing continuously cycling EnKF performance and whether benefits of blending are regime- and location-dependent. Nonetheless, this work suggests a combination of blending and high-resolution EnKF DA may represent a promising pathway toward an operational ensemble-based convection-allowing analysis–forecast system suitable for both nowcasting and next-day prediction over the CONUS.

Chapter 3: Short-term convection-allowing ensemble forecast sensitivity to resolution of initial condition perturbations and central initial states

3.1. Introduction

An ensemble of initial conditions (ICs) can be viewed as a set of IC perturbations added to a deterministic model solution. In this framework, the deterministic solution serves as a central initial state[6] for the IC ensemble. Theoretically, central initial states and IC perturbations can originate from disparate sources with different underlying physics, dynamics, and resolutions, which is common for convection-allowing ensemble (CAE) applications. For example, CAE ICs have regularly been constructed by adding perturbations derived from relatively coarse analyses or short-term forecasts to comparatively higher-resolution deterministic analyses (e.g., Xue et al. 2007; Peralta et al. 2012; Kühnlein et al. 2014; Tennant 2015; Raynaud and Bouttier 2016, 2017; Hagelin et al. 2017; Johnson and Wang 2020).

High-quality central initial states and IC perturbations are both critical for

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

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

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

The answer to this question has important implications for how next-generation CAEs, like NCEP's Rapid Refresh Forecast System (RRFS; Carley et al. 2021), are designed. For example, substantial computational resources can potentially be saved if increasing central initial state resolution to convection-allowing scales dramatically improves CAE forecasts but increasing IC perturbation resolution to convection-allowing scales has comparatively smaller impacts. In this case, RRFS development efforts can primarily be devoted to producing high quality, convection-allowing, deterministic central initial states about which relatively coarse, inexpensive IC perturbations are centered[7]. Conversely, if increasing IC perturbation resolution to convection-allowing scales is more important, a stronger emphasis should be placed on developing a pure ensemble-based convection-allowing data assimilation (DA) system for the RRFS.

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

In addition, Schwartz et al. (2021; hereafter S21) showed that 3-km ensemble

[7] A potential caveat: Even if IC perturbations can be coarsened without harming CAE forecasts, CAEs could be critical for providing flow-dependent background error covariances within convective-scale data assimilation systems. We discuss this possibility more thoroughly in section 4.5.1.

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

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

Thus, to directly assess whether it is more important for IC perturbations or central initial states to possess convection-allowing horizontal grid spacing, we designed, executed, and evaluated a series of CAE forecast experiments based upon S21's EnKFs. Sections 3.2 and 3.3 describe our experiments, while section 3.4 presents results. Although our focus is on IC resolution requirements for CAE forecasting applications, there are parallels between this topic and resolution requirements for mixed-resolution ensemble-based DA systems, which we discuss in section 3.5. Our overall conclusions (section 3.6) provide guidance about how development efforts to improve CAE ICs might best proceed.

3.2. Model and EnKF configurations

Thirty-six-hour (36-h), 10-member CAE forecasts were initialized from nine sets of ICs with different IC perturbations and central initial states. Eight of the IC sets leveraged analyses produced by the continuously cycling EnKF DA systems described by S21, who provided comprehensive details and justifications for specific EnKF DA settings (summarized in Table 3.1). Thus, only brief descriptions of EnKF and NWP model configurations are provided here, with a more thorough discussion reserved for how the nine sets of ICs were constructed (section 3.3).

Specifically, S21 performed two limited-area continuously cycling DA experiments using a square-root form of the EnKF (Anderson 2001) implemented within the Data Assimilation Research Testbed software (Anderson et al. 2009). Both EnKFs used 80 ensemble members and produced analyses every hour between 0000 UTC 23 April and 0000 UTC 20 May 2017 (inclusive; 649 hourly cycles). As noted by S21, this experimental period featured a variety of flow patterns and several heavy precipitation episodes primarily driven by strong synoptic forcing. The EnKFs

Parameter	Setting
EnKF algorithm	Ensemble adjustment Kalman filter (EAKF; Anderson 2001, 2003; Anderson and Collins 2007)
Ensemble size	80 members
Cycling period	1 hour
Updated WRF model variables	Zonal and meridional wind components; perturbation geopotential height, potential temperature, and dry surface pressure; and water vapor, graupel, snow, and rain mixing ratios
Localization function	Eq. (4.10) from Gaspari and Cohn (1999)
Horizontal localization full-width	For 15-km EnKF analyses, 1280 km for all observations. For 3-km EnKF analyses, 640 km, except 1280 km for rawinsonde observations
Vertical localization full-width	1.0 scale height
Sampling error correction	Anderson (2012)
Inflation method	Posterior relaxation-to-prior-spread [RTPS; Whitaker and Hamill (2012)]
Inflation factor	1.06
Lateral boundary condition perturbations	Random perturbations based on Gaussian noise added to GFS analyses and forecasts (e.g., Torn et al. 2006)
Sea surface temperature updates	Daily updates from NCEP's 0.12° analyses (e.g., Gemmill et al. 2007)
Assimilated observations	Rawinsonde, aircraft, wind profiler, satellite-tracked wind, global positioning system radio occultation (GPSRO), and surface observations
Observation errors and time windows	Based on the High-resolution Rapid Refresh Ensemble (HRRRE; Dowell et al. 2016, 2021)
Horizontal thinning for aircraft and satellite-tracked wind observations	30 km for 15-km EnKF analyses 15 km for 3-km EnKF analyses
Vertical thinning for aircraft and satellite-tracked wind observations	25 hPa

Table 3.1. Summary of continuously cycling EnKF configurations. See S21 for justifications for these settings.

assimilated approximately 30,000 – 100,000 observations from conventional measurements each cycle, and the first two days of cycling (i.e., 23 and 24 April) were regarded as spin-up where model solutions moved away from randomly-perturbed initial states prescribed at 0000 UTC 23 April 2017 (see S21 for details).

One EnKF experiment produced analyses solely on a 15-km computational domain (Figs. 3.1, 3.2a), while the second produced separate, parallel analyses on both 15- and 3-km domains (Figs. 3.1, 3.2b). To advance the 80-member ensemble states between hourly analyses, version 3.9.1.1 of the Advanced Research Weather Research and Forecasting (WRF) model (Skamarock et al. 2008; Powers et al. 2017) was used. WRF model physical parameterizations (Table 3.2) were identical across all 80 ensemble members and the two domains, except no cumulus parameterization was employed on the convection-allowing 3-km grid. In the nested 15-/3-km EnKF

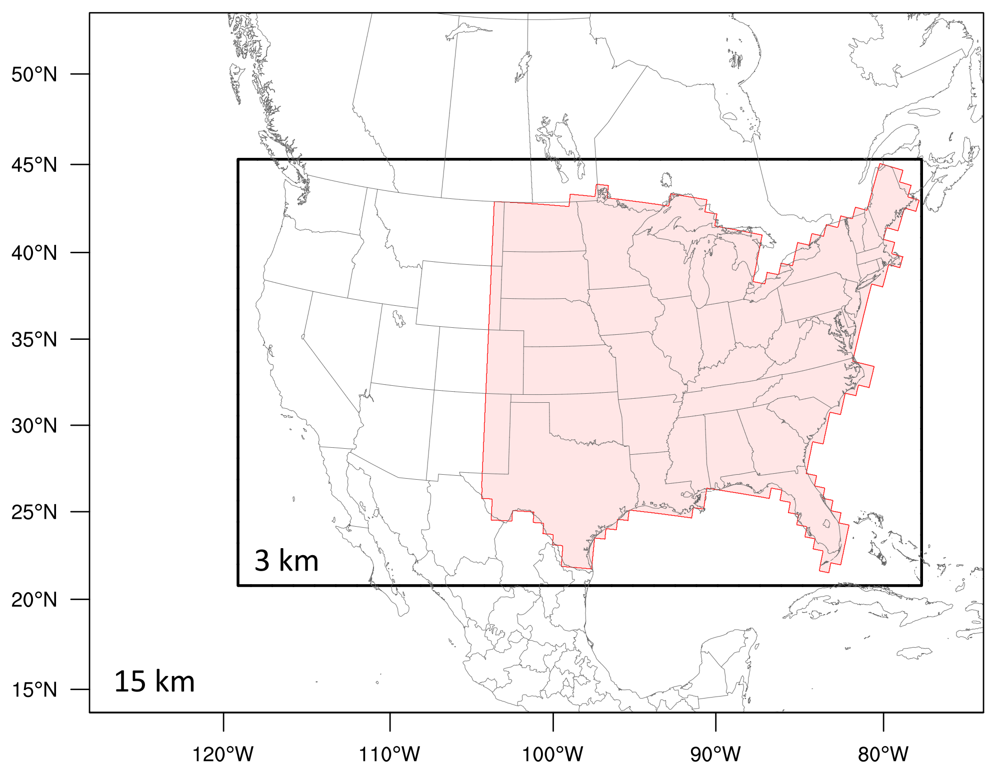

Fig. 3.1. Computational domain. Horizontal grid spacing was 15 km in the outer domain (415 x 325 points) and 3 km in the nest (1581 x 986 points). Lateral boundary conditions (LBCs) provided by global models were applied to the 15-km domain, which in turn provided LBCs for the 3-km domain. Objective precipitation verification only occurred over the red shaded region of the 3-km domain (CONUS east of 105°W).

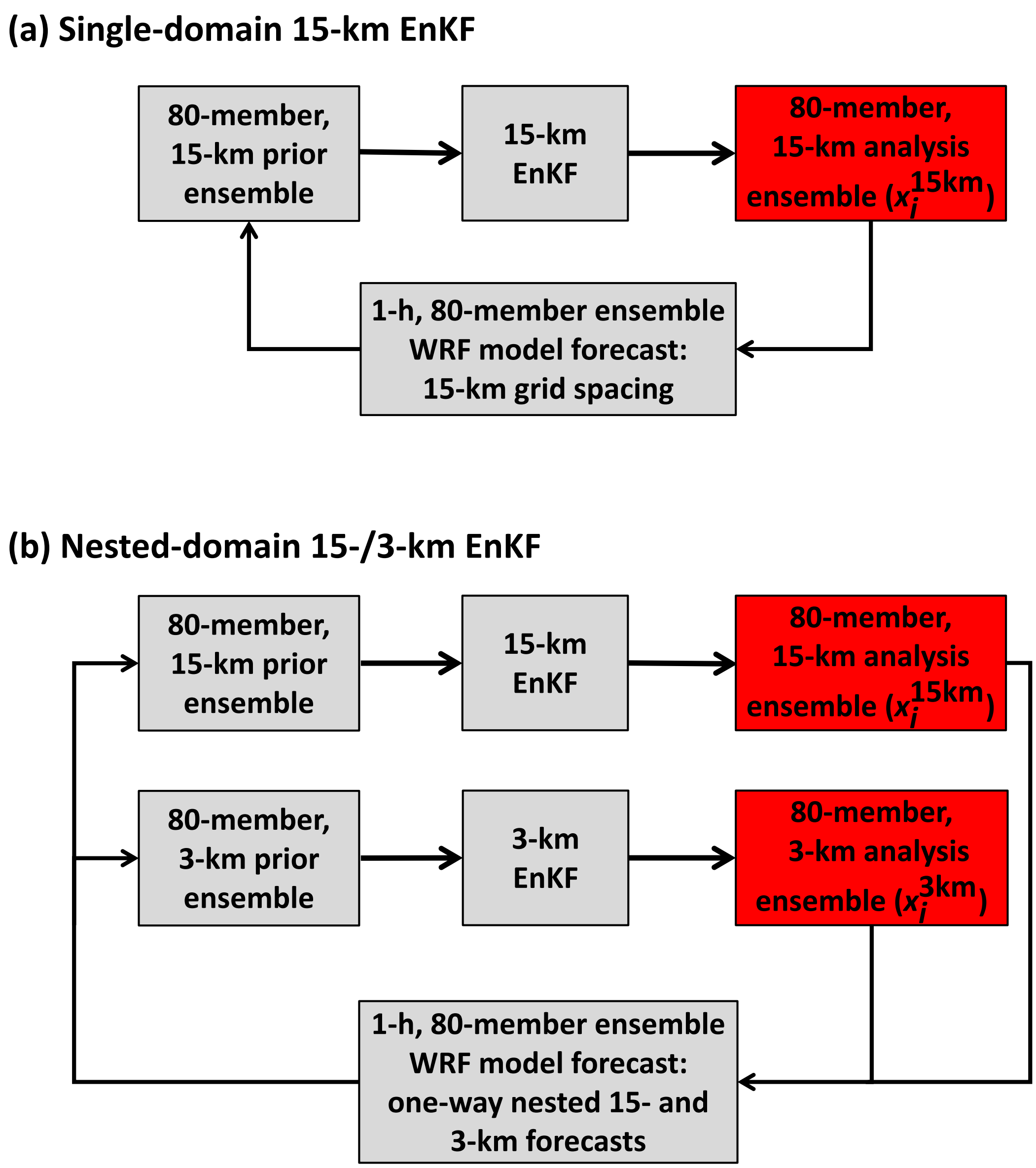

Fig. 3.2. Flowcharts of continuously cycling EnKF data assimilation systems over (a) solely the 15-km computational domain (i.e., outer domain in Fig. 3.1) and (b) both the 15- and 3-km computational domains (i.e., both domains in Fig. 3.1). Posterior ensembles (red shaded boxes) were used to construct initial conditions for CAE forecasts.

DA system (Fig. 3.2b), 1-h WRF model forecasts between analyses were one-way nested, ensuring that the 15-km EnKF was unaffected by the 3-km EnKF (i.e., 15-km fields in the nested- and single-domain EnKF DA systems were identical). Both EnKFs performed well, yielding acceptable spread–error relationships (e.g.,

Parameter	WRF model setting
Model version	Version 3.9.1.1 of the Advanced Research WRF model
Horizontal grid spacing	15 and 3 km in the outer and inner domains, respectively
Time step	60 and 12 s in the 15- and 3-km domains, respectively
Number of vertical levels	51 (based on the Rapid Refresh model; Benjamin et al. 2016)
Model top	15 hPa
Microphysics parameterization	Thompson (Thompson et al. 2008)
Longwave and shortwave radiation parameterizations	Rapid Radiative Transfer Model for Global Climate Models (RRTMG) with ozone and aerosol climatologies (Mlawer et al. 1997; Iacono et al. 2008; Tegen et al. 1997)
Planetary boundary layer parameterization	Mellor–Yamada–Janjić (MYJ) (Mellor and Yamada 1982; Janjić 1994, 2002)
Land surface model	Noah (Chen and Dudhia 2001)
Cumulus parameterization	Tiedtke (15-km domain only; Tiedtke 1989; Zhang et al. 2011)

Table 3.2. WRF model settings used for the EnKFs and CAE forecasts.

Houtekamer et al. 2005) and similar model climates that were stable throughout the 4 weeks of continuous cycling (see Figs. 5 and 6 in S21).

3.3. Experimental design

Analysis ensembles from the 15- and 3-km EnKFs (i.e., red boxes in Fig. 3.2) were used to derive central initial states and IC perturbations for various CAEs. Operational global models were also used as sources of IC perturbations and central initial states.

Ultimately, the forecasts of interest were those produced on the 3-km grid over the conterminous United States (CONUS) east of the Rockies (Fig. 3.1). Employing a nested WRF model configuration provided a computationally affordable way to place this region far from lateral boundaries, which is potentially advantageous (e.g., Warner et al. 1997). However, this choice required ICs to be constructed for both the 15- and 3-km domains, which somewhat complicated the experimental design. While we offer some remarks about initializing the 15-km domain in upcoming subsections, the primary function of the 15-km domain was to provide LBCs for the 3-km domain, and the most important aspect of the experimental design concerned how ICs were produced for the 3-km domain. Note that ICs for the 3-km domain did not have to possess 3-km horizontal grid spacing, as coarser fields could be downscaled onto the 3-km grid.

3.3.1. CENTRAL INITIAL STATES FOR THE 3-KM DOMAIN

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

$$\overline{x^{3\text{km}}} = \frac{1}{80}\sum_{i=1}^{80} x_i^{3\text{km}}, \tag{3.1}$$

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

$$\overline{x^{15\text{km}}} = \frac{1}{80}\sum_{i=1}^{80} x_i^{15\text{km}}, \tag{3.2}$$

where $x_i^{15\text{km}}$ and $x_i^{3\text{km}}$ denote 15- and 3-km EnKF analyses for the *i*th of 80 ensemble members, respectively[8]. Operational Global Forecast System (GFS) analyses (x^{GFS})

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

with 0.25° horizontal grid spacing[9] also served as central initial states for the 3-km domain. From a spectral perspective, all three sources of central initial states were usually similar on mutually resolvable scales, with the largest differences at scales < 100 km that were not resolvable by all analyses (Fig. 3.3).

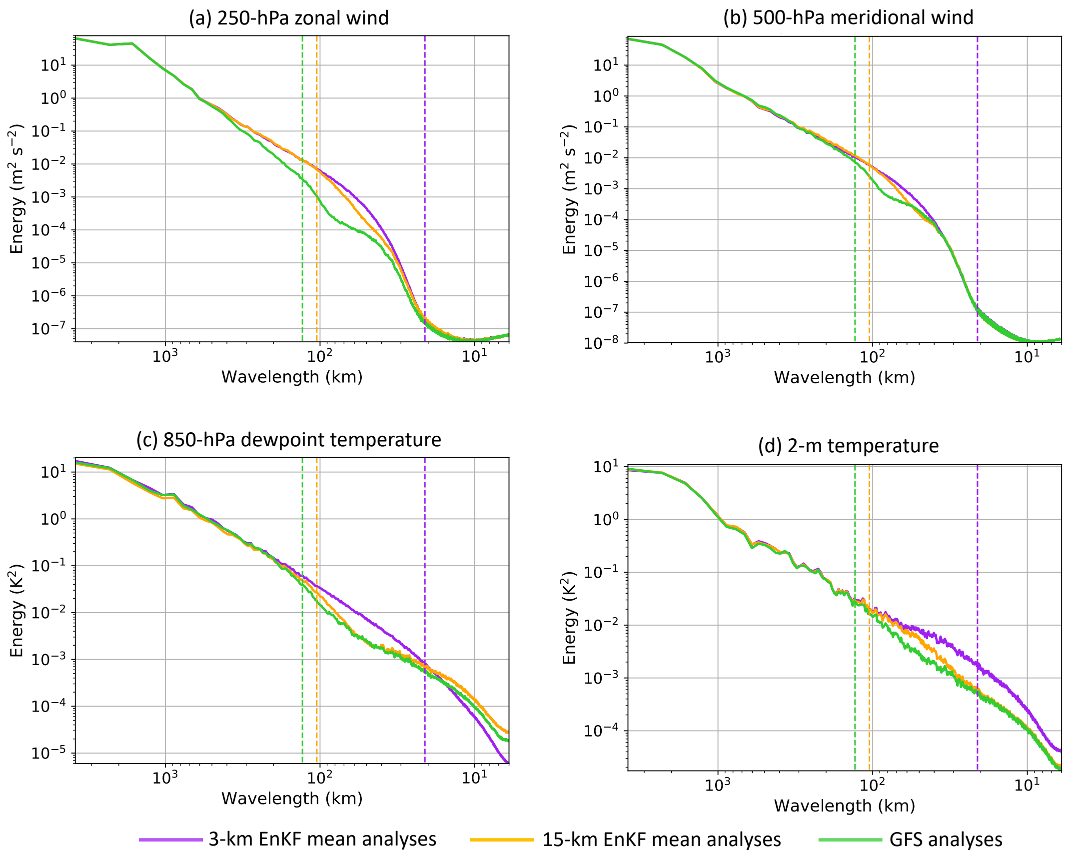

Fig. 3.3. Power spectra of 3-km EnKF mean analyses (purple), 15-km EnKF mean analyses (orange), and GFS analyses (green) as a function of wavelength (km) for (a) 250-hPa zonal wind ($m^2\ s^{-2}$), (b) 500-hPa meridional wind ($m^2\ s^{-2}$), (c) 850-hPa dewpoint temperature (K^2), and (d) 2-m temperature (K^2) averaged over all 0000 UTC analyses between 25 April and 20 May 2017 (inclusive). To compute spectra from the various datasets, all fields were interpolated onto the 3-km domain (Fig. 3.1). Spectra were then computed over the geographic area covered by the 3-km domain, excluding points within 45 km of each lateral boundary. The discrete cosine transform (Denis et al. 2002) was used to compute spectra and spectral variance binning employed the method of Ricard et al. (2013). Dashed purple, orange, and green vertical lines denote effective resolutions of 3-km EnKF mean analyses, 15-km EnKF mean analyses, and GFS analyses, respectively, which were approximately 7 times the horizontal grid spacing for the WRF-based EnKFs (e.g., Skamarock 2004) and approximately 10 times the horizontal grid spacing for the GFS (e.g., Ji et al. 2016). The *y*-axis values are different in each panel.

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

Differences between 15- and 3-km EnKF mean analyses were attributable to differences in horizontal grid spacing and associated representation of convection (parameterized versus explicit). Conversely, because GFS and limited-area EnKF analyses reflected entirely different NWP models and DA systems, differences between GFS and limited-area EnKF mean analyses were potentially due to many factors other than disparities in resolution and associated treatment of convection. Nonetheless, employing GFS analyses as a source of central initial states was useful for examining whether CAE forecast sensitivity to IC perturbations depended on the central initial state and provides an operational baseline for the EnKF-based central initial states.

3.3.2. IC PERTURBATIONS FOR THE 3-KM DOMAIN

IC perturbations for the 3-km domain were derived from three sources, including from both 15- and 3-km EnKF analysis ensembles (e.g., Fig. 3.2), given by

$$\delta x_i^{15\mathrm{km}} = x_i^{15\mathrm{km}} - \overline{x^{15\mathrm{km}}} \tag{3.3}$$

and

$$\delta x_i^{3\mathrm{km}} = x_i^{3\mathrm{km}} - \overline{x^{3\mathrm{km}}}, \tag{3.4}$$

where $\delta x_i^{15\mathrm{km}}$ and $\delta x_i^{3\mathrm{km}}$ respectively denote 15- and 3-km posterior (after assimilation) EnKF perturbations for the ith ensemble member. Although Eqs. (3.3) and (3.4) were valid for $i = 1 \ldots 80$, IC perturbations from just members 1–10 were required because CAE forecasts only had 10 members (as explained in section 3.3.4). Thus, there was some uncertainty about whether EnKF-based IC perturbations should be computed with respect to the mean of all 80 posterior members [as in Eqs. (3.3) and (3.4)] or just the mean of posterior members 1–10. However, testing revealed that CAE forecasts were insensitive to whether IC perturbations were computed with respect to the mean of posterior members 1–80 or 1–10, because averages over members 1–80 and 1–10 were extremely similar (not shown).

ICs from NCEP's operational 0.5° Global Ensemble Forecast System (GEFS; Zhou et al. 2017) provided the third source of IC perturbations for the 3-km domain[10]. Specifically, perturbations for the ith GEFS member ($\delta x_i^{\mathrm{GEFS}}$) were given by

$$\delta x_i^{\mathrm{GEFS}} = x_i^{\mathrm{GEFS}} - \overline{x^{\mathrm{GEFS}}}, \tag{3.5}$$

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

$$\overline{x^{\mathrm{GEFS}}} = \frac{1}{10}\sum_{i=1}^{10} x_i^{\mathrm{GEFS}}. \tag{3.6}$$

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

The 15-km EnKF IC perturbations had more energy than 3-km EnKF IC perturbations at most mutually resolvable scales (Fig. 3.4), possibly due to the 15-km EnKF's use of cumulus parameterization, a well-known error source (e.g., Torn and Davis 2012; Romine et al. 2013; Mahoney 2016; Wong et al. 2020), whereas the 3-

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

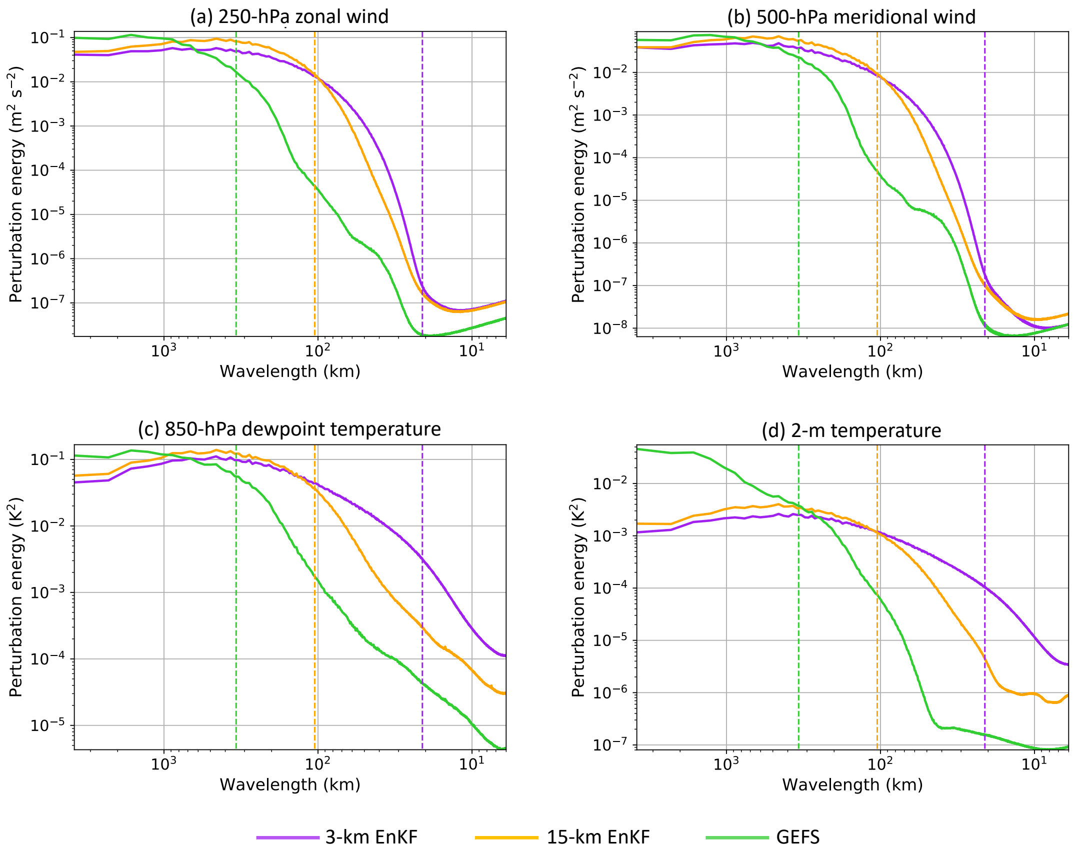

Fig. 3.4. Perturbation power spectra of 3-km EnKF analysis ensembles (purple), 15-km EnKF analysis ensembles (orange), and GEFS initial condition ensembles (green) as a function of wavelength (km) for (a) 250-hPa zonal wind ($m^2\,s^{-2}$), (b) 500-hPa meridional wind ($m^2\,s^{-2}$), (c) 850-hPa dewpoint temperature (K^2), and (d) 2-m temperature (K^2) averaged over all ensemble perturbations and all 0000 UTC analyses between 25 April and 20 May 2017 (inclusive). Perturbations were defined with respect to the ensemble mean. To compute spectra from the various datasets, all fields were interpolated onto the 3-km domain (Fig. 3.1). Spectra were then computed over the geographic area covered by the 3-km domain, excluding points within 45 km of each lateral boundary. The discrete cosine transform (Denis et al. 2002) was used to compute spectra and spectral variance binning employed the method of Ricard et al. (2013). Dashed purple, orange, and green vertical lines denote effective resolutions of the 3-km ensemble, 15-km ensemble, and GEFS, respectively, which were approximately 7 times the horizontal grid spacing for the WRF-based ensembles (e.g., Skamarock 2004) and approximately 10 times the horizontal grid spacing for the GEFS (e.g., Ji et al. 2016). The *y*-axis values are different in each panel.

km EnKF was convection-allowing. However, 3-km IC perturbations had the most power at scales < 100 km. Differences between 15- and 3-km EnKF IC perturbation power spectra were typically small compared to their collective differences with respect to GEFS IC perturbation power spectra, which had the most perturbation energy at scales > 1000 km. Similarly, differences of domain-average spread between 15- and 3-km posterior ensembles were small compared to differences between EnKF and GEFS IC ensembles, the latter of which had uniformly smaller spreads except for low-level temperature[11] (Fig. 3.5).

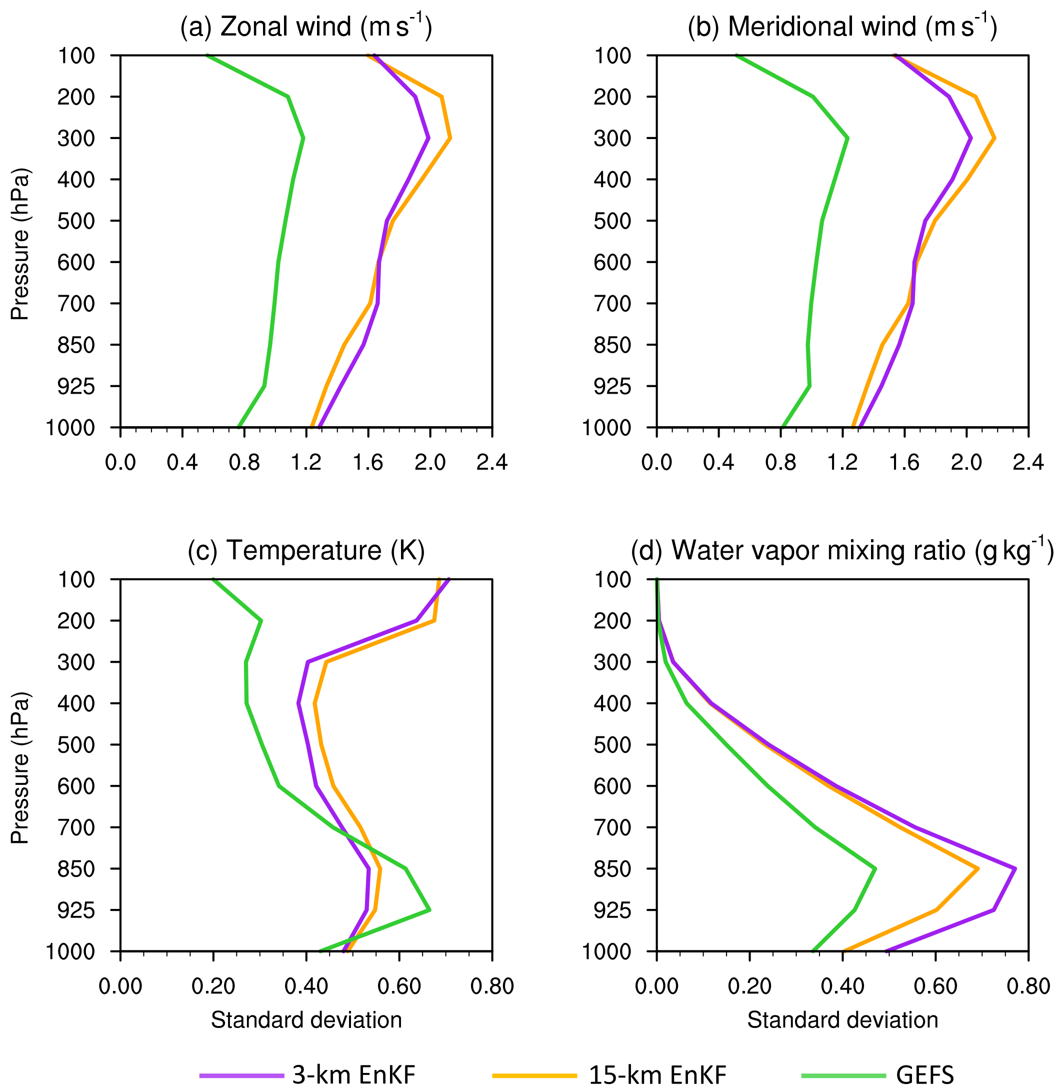

Fig. 3.5. Average standard deviation over the geographic area covered by the 3-km domain (Fig. 3.1) and all 0000 UTC 3-km EnKF analysis ensembles (purple), 15-km EnKF analysis ensembles (orange), and GEFS IC ensembles (green) between 0000 UTC 25 April and 0000 UTC 20 May 2017 (inclusive) for (a) zonal wind (m s^{-1}), (b) meridional wind (m s^{-1}), (c) temperature (K), and (d) water vapor mixing ratio (g kg^{-1}). The *x*-axis values are different in each row.

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

3.3.3. IC CONSTRUCTION

At 0000 UTC daily between 25 April and 20 May 2017 (inclusive), each of the three sets of IC perturbations [Eqs. (3.3)–(3.5)] was re-centered about each of the three sets of central initial states [Eq. (3.1), Eq. (3.2), and GFS analyses (x^{GFS})], yielding nine sets of IC ensembles that differed by their IC perturbations and central initial states (Tables 3.3, 3.4). While re-centering is a common CAE initialization technique (e.g., Xue et al. 2007; Kong et al. 2008, 2009; Peralta et al. 2012; Kühnlein et al. 2014; Tennant 2015; Johnson and Wang 2016; Raynaud and Bouttier 2016, 2017; Hagelin et al. 2017; JW20; S20), an alternative way of assessing forecast sensitivity to IC resolution would be to remove small-scale features from 3-km EnKF posterior ensembles with a low-pass filter to produce coarser central initial states and

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

Table 3.3. Experiment names and their corresponding central initial states (columns) and IC perturbations (rows) for the 3-km domain. All IC sets were constructed at 0000 UTC each day during the experimental period. Bolded experiments denote the four "EnKF-only" experiments, while italicized experiments denote the four "EnKF+Global" experiments.

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

Table 3.4. Mathematical expressions for the *i*th ensemble member's ICs for the 3-km domain in the various experiments. Term *D* represents a downscaling operator within the WRF model that remaps fields with horizontal grid spacing coarser than 3 km (e.g., 15-km EnKF analyses) onto the 3-km computational domain. Terms $x_i^{15\text{km}}$ and $x_i^{3\text{km}}$ denote 15- and 3-km EnKF analyses for the *i*th ensemble member, respectively, and x^{GFS} denotes GFS analyses. All other terms are defined in Eqs. (3.1)–(3.5).

IC perturbations (e.g., Potvin et al. 2017; JW20). Although this method is elegant, we did not apply filtering to construct ICs with varied resolutions because re-centering has greater relevance within operational environments, where output from multiple modeling systems with different resolutions is typically available. Furthermore, re-centering is common within operational DA systems (e.g., Clayton et al. 2013; Wang et al. 2013).

Similar to S20 and Peralta et al. (2012), only zonal and meridional wind, potential temperature, water vapor mixing ratio, and perturbation geopotential and dry surface pressure (U, V, θ, q_v, ϕ, and μ, respectively) were re-centered. All other fields, like hydrometeors, were provided by the *i*th member of either the 15- or 3-km posterior ensemble when the finest-resolution field used to construct a specific set of

ICs had 15- or 3-km horizontal grid spacing, respectively (e.g., for IC ensembles produced by combining 3-km central initial states with GEFS or 15-km IC perturbations, initial microphysics states were provided by the *i*th member of the 3-km posterior ensemble). We used ensemble members for these fields to provide IC diversity, rather than forcing these other variables to common values provided by deterministic central initial states. However, initialization of these auxiliary fields likely had little forecast impact, as S20 suggested that precipitation forecasts over the CONUS east of the Rockies were more sensitive to IC perturbation characteristics of U, V, θ, q_v, ϕ, and μ than initial hydrometeor states.

3.3.3.1. ICs based solely on limited-area EnKF analyses

Four sets of ICs were based solely on 0000 UTC limited-area EnKF analyses and are henceforth collectively referred to as the "EnKF-only" experiments (bold experiments in Table 3.3). Because the nested-domain EnKF employed one-way nesting (section 3.2, Fig. 3.2b), ICs for the 15-km domain across all EnKF-only experiments were identical, namely, 0000 UTC 15-km EnKF analysis ensembles (i.e., $x_i^{15\mathrm{km}}$). Thus, CAE forecast differences among the four EnKF-only experiments were solely due to how the 3-km domain (Fig. 3.1) was initialized at 0000 UTC.

The most straightforward EnKF-based ICs for the 3-km domain were provided by 3-km EnKF analysis members (i.e., $x_i^{3\mathrm{km}}$), which are 3-km EnKF analysis perturbations centered on 3-km EnKF mean analyses ("3kmCent_3kmPert"; Tables 3.3, 3.4). Initial states for the 3-km domain were also provided by downscaling 15-km EnKF analysis members (i.e., $x_i^{15\mathrm{km}}$) onto the 3-km computational domain

("15kmCent_15kmPert"; Tables 3.3, 3.4). The downscaling process does not add detail, so even though downscaled 15-km fields resided on the 3-km grid, the ICs still possessed spatial resolution associated with 15-km horizontal grid spacing.

The final two sets of ICs for the 3-km domain within the EnKF-only experiments had mixed resolutions of IC perturbations and central initial states. In one set, the 3-km domain was initialized by re-centering downscaled 15-km EnKF analysis perturbations [i.e., Eq. (3.3)] about 3-km EnKF mean analyses ("3kmCent_15kmPert"; Tables 3.3, 3.4). The other set was produced by re-centering 3-km EnKF analysis perturbations [i.e., Eq. (3.4)] about downscaled 15-km EnKF mean analyses ("15kmCent_3kmPert"; Tables 3.3, 3.4).

Because 3-km IC ensembles can be averaged to create 3-km central initial states [e.g., Eq. (3.1)], a configuration like 15kmCent_3kmPert would be unlikely within operational environments. However, this experiment was helpful for elucidating whether it is more important for central initial states or IC perturbations to possess convection-allowing horizontal grid spacing, as CAE forecast differences between 15kmCent_3kmPert and 15kmCent_15kmPert were solely attributable to IC perturbation resolution for the 3-km domain and 15kmCent_3kmPert and 3kmCent_3kmPert only differed regarding central initial state resolution.

3.3.3.2. ICs based on a combination of limited-area EnKF analyses and global fields

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

Specifically, another four sets of ICs for the 3-km domain were constructed that relied in part on the GFS or GEFS and are collectively referred to as the "EnKF+Global" experiments (italicized experiments in Table 3.3). Two sets of ICs for the 3-km domain were produced by re-centering downscaled perturbations derived from 0000 UTC GEFS ICs [Eq. (3.5)] about 0000 UTC 15- and 3-km EnKF mean analyses ("3kmCent_GEFSPert" and "15kmCent_GEFSPert"; Tables 3.3, 3.4). The other two sets of ICs for the 3-km domain were produced by re-centering 0000 UTC 15- and 3-km EnKF analysis perturbations [Eqs. (3.3), (3.4)] about downscaled 0000 UTC GFS analyses ("GFSCent_3kmPert" and "GFSCent_15kmPert"; Tables 3.3, 3.4).

Introducing global fields poses somewhat of a dilemma: should GFS and GEFS fields participate in re-centering only for purposes of initializing the 3-km domain or for purposes of initializing *both* the 15- and 3-km domains (Fig. 3.1)? The former would mean all four EnKF+Global experiments have identical ICs for the 15-

km domain that are equal to those of the EnKF-only experiments (i.e., $x_i^{15\mathrm{km}}$). Conversely, the latter would maintain consistency across both domains, which intuitively seems desirable. However, in this latter scenario, 3-km forecasts from the four EnKF+Global experiments could potentially be sensitive to LBCs inherited from different 15-km forecasts. Likewise, 3-km forecasts from the EnKF+Global experiments could potentially differ from those produced by the EnKF-only experiments because of different 15-km forecasts.

To address this conundrum, we performed several exploratory experiments where pairs of experiments solely differed by either their ICs for the 15-km domain or ICs for the 3-km domain. Findings revealed that 3-km precipitation forecasts were far more sensitive to ICs for the 3-km domain than ICs for the 15-km domain. Moreover, S21 arrived at identical conclusions (see their footnote 3). Therefore, ICs for the 15-km domain appeared to have little impact on 3-km forecasts, and differences between various 3-km CAE forecasts were attributed to different ICs for the 3-km domain.

Ultimately, we chose to maintain consistency across both domains, so GEFS IC perturbations and GFS analyses were introduced to ICs for both the 15- and 3-km domains in the EnKF+Global experiments. Expressions for the 15-km domain's ICs are analogous to those for the 3-km domain's ICs given in Table 3.4, except $\overline{x^{3\mathrm{km}}}$ and $\delta x_i^{3\mathrm{km}}$ are replaced by $\overline{x^{15\mathrm{km}}}$ and $\delta x_i^{15\mathrm{km}}$, respectively, GFS and GEFS fields are downscaled onto the 15-km domain, and there is no need to downscale 15-km fields.

3.3.3.3. ICs based solely on global fields

The final set of ICs for the 3-km domain was produced by re-centering downscaled GEFS IC perturbations about downscaled GFS analyses at 0000 UTC ("GFSCent_GEFSPert"; Tables 3.3, 3.4). These ICs were independent of the limited-area EnKFs and were nearly identical to GEFS ICs (i.e., x_i^{GEFS}) on account of the GEFS initialization procedure (Zhou et al. 2017). Additionally, CAEs with these ICs served as benchmarks for the other CAEs whose ICs relied on limited-area EnKF analyses (Tables 3.3, 3.4). Per the above discussion, ICs for the 15-km domain were also produced by re-centering GEFS IC perturbations about GFS analyses.

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

3.3.4. CAE FORECASTS

Members 1–10 from all nine sets of ICs (i.e., section 3.3.3; Tables 3.3, 3.4) initialized 36-h, 10-member ensemble forecasts over the nested domain (Fig. 3.1) at 0000 UTC daily between 25 April and 20 May 2017 (inclusive; 26 CAE forecasts per experiment). These 36-h forecasts employed identical WRF model configurations as the nested 15-/3-km EnKF DA system (Table 3.2).

Although ICs for more than 10 ensemble members were available, computing constraints limited 36-h CAE forecasts to 10 members, which is sufficient to provide skillful and valuable probabilistic precipitation forecasts (e.g., Clark et al. 2011, 2018; Schwartz et al. 2014) and comparable to ensemble sizes of other CAEs that operate regularly over the CONUS (e.g., Dowell et al. 2016, 2021; Roberts et al. 2020).

For all 36-h forecasts, perturbation members 1–10 from the GEFS provided LBCs for the 15-km domain, which in turn provided LBCs for the 3-km nest. The 3-km forecasts were then verified with a focus on precipitation, as described next.

3.4. Precipitation forecast verification

3.4.1. METHODS

As this study builds upon S21, overall methods for precipitation verification closely followed S21, which can be consulted for additional details. Specifically, forecasts of 1-h accumulated precipitation were objectively verified against NCEP's ~4.763-km Stage IV (ST4) analyses (Lin and Mitchell 2005) over the CONUS east of 105°W (Fig. 3.1), where ST4 analyses are most reliable (e.g., Nelson et al. 2016). Precipitation forecasts were interpolated to the ST4 grid with a budget algorithm (Accadia et al. 2003) for comparison to ST4 analyses.

Following S21, we used percentile thresholds (e.g., the 95th percentile), rather than absolute thresholds (e.g., 1.0 mm h^{-1}), to define events. Using percentile thresholds removes bias, permitting a robust assessment of spatial placement within the context of a model's climate (e.g., Roberts and Lean 2008; Mittermaier and Roberts 2010; Dey et al. 2014; Gowan et al. 2018; Cafaro et al. 2021). We used the

95.0th, 97.5th, 99.0th, 99.5th, 99.75th, and 99.9th percentile thresholds to examine a range of precipitation intensities, which roughly corresponded to physical thresholds of 1.1, 2.0, 4.0, 6.2, 8.8, and 12.8 mm h^{-1} (based on ST4 analyses).

Additionally, rather than verifying point-based probabilities, we used a neighborhood approach to derive and verify "neighborhood ensemble probabilities" (NEPs; Schwartz et al. 2010; Schwartz and Sobash 2017), which were computed by spatially averaging point-based probabilities within circular neighborhoods surrounding each grid point. NEPs are more appropriate for verifying convection-allowing model forecasts than point-based probabilities because they recognize that high-resolution NWP models are inaccurate at the grid scale (e.g., Theis et al. 2005; Ebert 2008, 2009). Statistical significance was assessed with a bootstrap resampling approach (with replacement) applied to pairwise differences between two experiments (e.g., Hamill 1999; Wolff et al. 2014) using 10,000 resamples.

3.4.2. RESULTS

Probabilistic precipitation forecasts from the nine CAEs were evaluated with fractions skill scores [FSSs; Roberts and Lean (2008)], attributes diagrams (e.g., Wilks 2011), the "reliability component" of the Brier score decomposition (BS_{REL}; Murphy 1973), and areas under the relative operating characteristic (ROC) curve (Mason 1982; Mason and Graham 2002). Both FSSs and ROC areas range between 0 and 1, with higher values indicating more skill, while perfect reliability is achieved for curves lying on the diagonal of attributes diagrams. Smaller values of BS_{REL} are better.

These metrics were computed for NEPs constructed with neighborhood length scales (r) between 5 and 150 km, but varying r did not change overall conclusions. Thus, we focus on statistics computed with r = 100 km, and to provide a holistic assessment we present statistics aggregated over all 26 forecasts.

3.4.2.1. CAE forecast sensitivity to central initial states

Given identical IC perturbations, CAEs with 3-km central initial states typically had higher aggregate FSSs over the first ~12–21 h than CAEs with 15-km central initial states, which in turn had higher FSSs than CAEs with GFS central initial states (compare purple, orange, and green curves with common line patterns in Fig. 3.6). These relationships held at all thresholds and differences were regularly statistically significant at the 90% confidence level, particularly those between CAEs with 15- and 3-km central initial states and CAEs with 3-km and GFS central initial states (Fig. 3.7). Aggregate ROC areas over the first 12 h yielded identical conclusions as FSSs (Fig. 3.8), and statistical significance between pairs of experiments for ROC areas echoed patterns in Fig. 3.7 (not shown). Moreover, for constant IC perturbations, CAEs with 3-km central initial states usually had better 1–12-h forecast reliabilities than CAEs with 15-km central initial states, and CAEs with GFS central initial states typically had the poorest reliabilities (Fig. 3.9, Table 3.5). All CAEs had comparable sharpness and were typically skillful compared to climatological forecasts (Fig. 3.9). The consistency of FSSs, ROC areas, and reliabilities through 12 h strongly suggests that short-term CAE precipitation forecasts are improved by using convection-allowing analyses as central initial states.

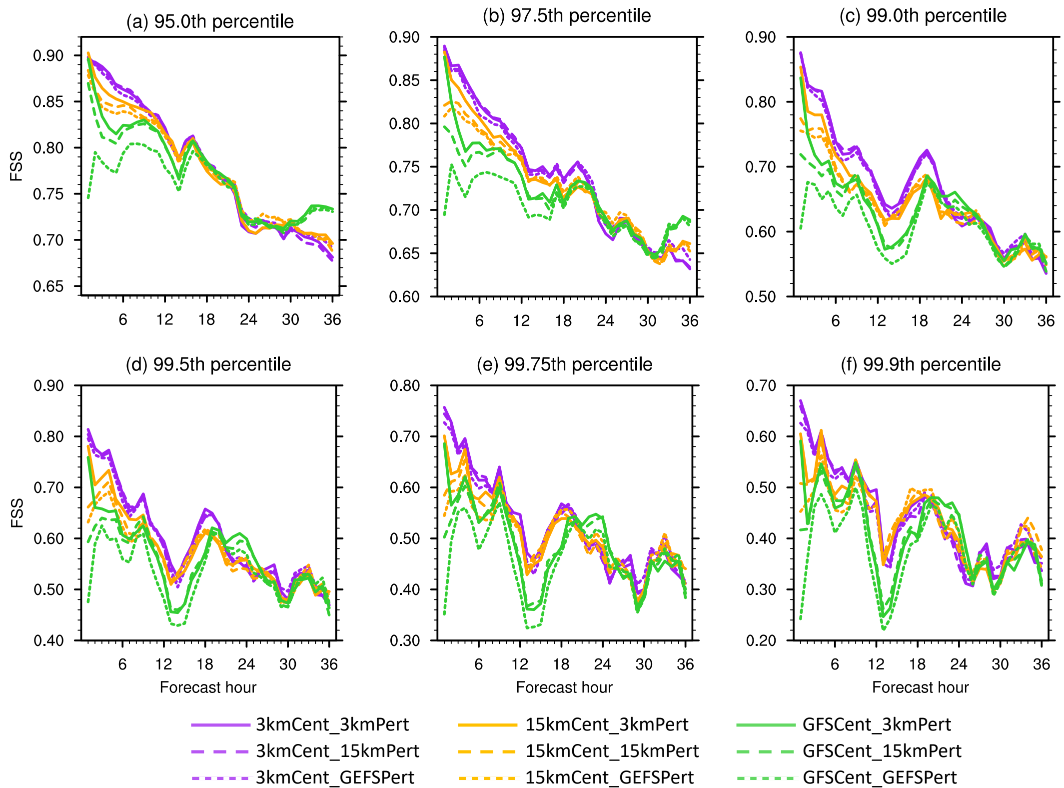

Fig. 3.6. Fractions skill scores (FSSs) over the CONUS east of 105°W (Fig. 3.1) with a 100-km neighborhood length scale for the (a) 95th, (b) 97.5th, (c) 99th, (d) 99.5th, (e) 99.75th, and (f) 99.9th percentile thresholds aggregated over all 26 3-km forecasts of 1-h accumulated precipitation as a function of forecast hour. Values on the *x* axis represent ending forecast hours of 1-h accumulation periods (e.g., an *x*-axis value of 24 is for 1-h accumulated precipitation between 23–24 h). The *y*-axis scales are different in each panel.

Conversely, after ~18–21 h, differences between CAEs were generally smaller, with the largest differences occurring after ~30 h at the 95th and 97.5th percentile thresholds, where CAEs with GFS central initial states were more skillful than those with 15- and 3-km central initial states given fixed IC perturbations (Figs. 3.6a,b). Additionally, at higher thresholds, GFS central initial states typically led to the best ~22–26-h forecasts (Figs. 3.6c–f). ROC areas and attributes statistics after ~18 h yielded similar results as FSSs (not shown).

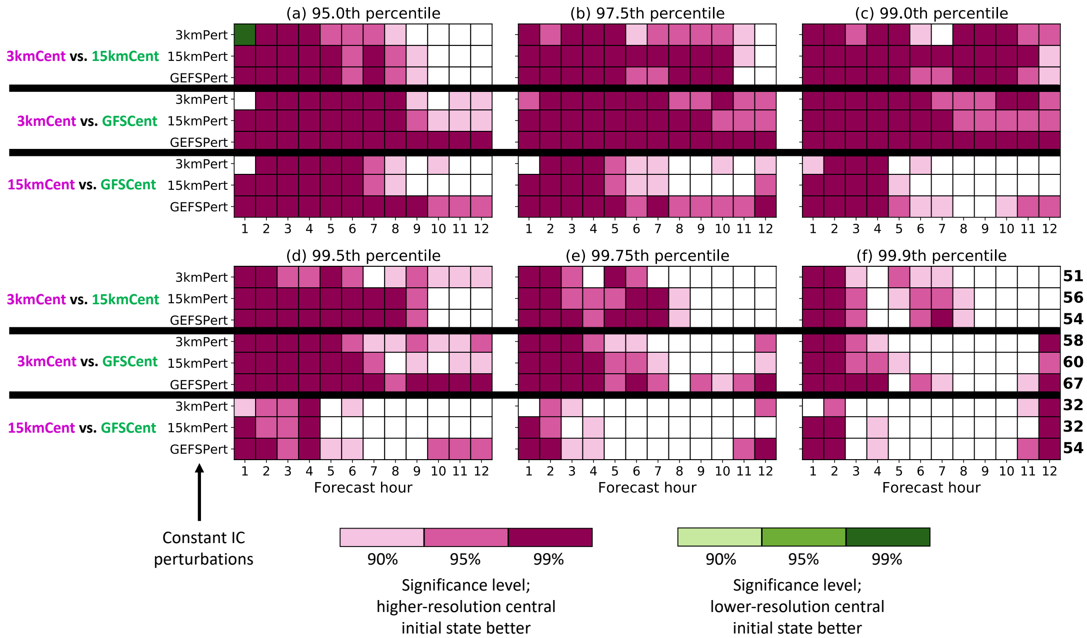

Fig. 3.7. Statistical significance levels of aggregate FSS differences (e.g., Fig. 3.6) between various experiments for the (a) 95th, (b) 97.5th, (c) 99th, (d) 99.5th, (e) 99.75th, and (f) 99.9th percentile thresholds for forecast hours 1–12. These comparisons assess the impact of changing central initial state resolution. Specifically, in each panel, a given row represents a fixed set of IC perturbations; "3kmPert", "15kmPert", and "GEFSPert" refer to IC perturbations provided by 3-km EnKF analysis ensembles, 15-km EnKF analysis ensembles, and GEFS ICs, respectively. Each panel is broken into thirds to represent different comparisons, and pink text denotes the experiment in each comparison with the higher resolution central initial state. The top third compares experiments with 15- and 3-km central initial states ("3kmCent vs. 15kmCent"). The middle third compares experiments with 3-km and 0.25° GFS central initial states ("3kmCent vs. GFSCent"). The bottom third compares experiments with 15-km and 0.25° GFS central initial states ("15kmCent vs. GFSCent"). Pink shadings indicate that higher resolution central initial states led to statistically significantly higher FSSs for the fixed IC perturbations, while green shadings indicate that lower resolution central initial states led to statistically significantly higher FSSs for the fixed IC perturbations. White cells indicate that aggregate FSSs of two experiments with varied central initial states but common IC perturbations were not statistically significantly different at the 90% confidence level or higher. Annotations to the right of (f) represent the number of occurrences where, for the given row, experiments with higher resolution central initial states had statistically significantly higher FSSs than experiments with lower resolution central initial states across all 6 percentile thresholds and all forecast hours (i.e., the total number of pink-shaded boxes in each row across all 6 panels).

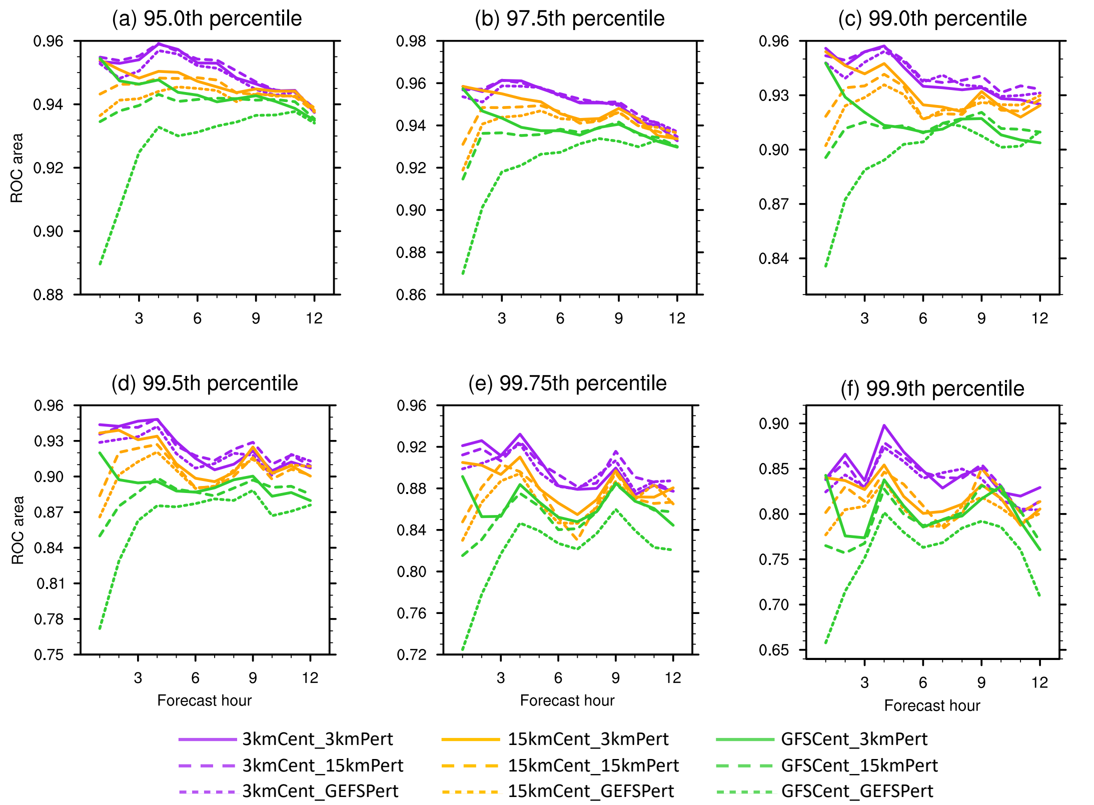

Fig. 3.8. As in Fig. 3.6, except for areas under the relative operating characteristic (ROC) curve computed using decision thresholds of 1%, 2%, 3%, 4%, 5%, 10%, 15%, …, 95%, and 100% and a trapezoidal method. Only forecast hours 1–12 are presented to zoom-in on the period with the largest systematic differences between experiments and to more easily visualize differences between CAEs with identical central initial states but different IC perturbations (which have common line colors but different line patterns). The *y*-axis scales are different in each panel.

Overall, benefits of convection-allowing central initial states were mostly confined to the first ~12 h, consistent with S21, who found that GEFS-initialized CAEs had better ~18–36-h precipitation forecasts than CAEs with 15- and 3-km EnKF ICs due to improved large-scale representation in GEFS ICs compared to limited-area EnKF analyses. We thus presume that large scales, which are more critical for next-day (i.e., ~18–36-h) forecasts than shorter-term forecasts, were better represented in GFS analyses than in limited-area EnKF ICs. Because these general ~18–36-h forecast behaviors were exhaustively discussed by S21, we henceforth focus on ~1–12-h forecasts, where differences among the CAEs were usually largest. However, some thoughts about how our findings enhance S21's conclusions about ~18–36-h forecast quality are provided in section 3.5.4.

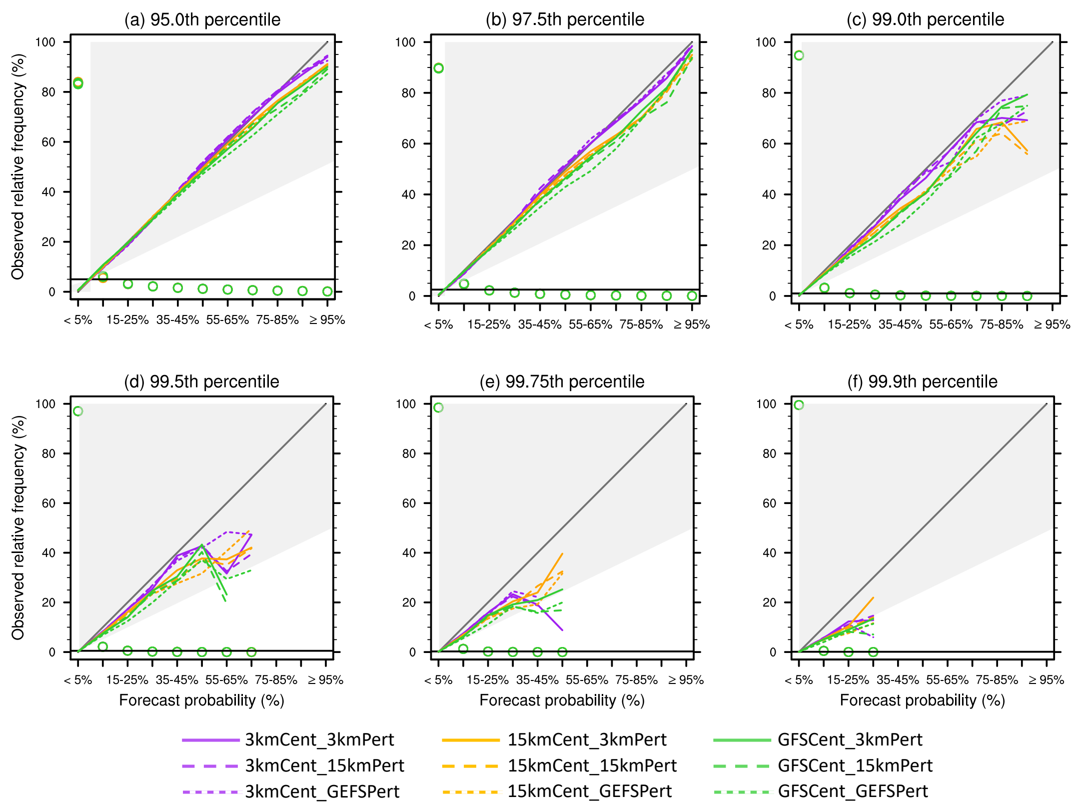

Fig. 3.9. Attributes diagrams computed over the CONUS east of 105°W (Fig. 3.1) with a 100-km neighborhood length scale aggregated over all 26 1–12-h 3-km forecasts of 1-h accumulated precipitation for the (a) 95th, (b) 97.5th, (c) 99th, (d) 99.5th, (e) 99.75th, and (f) 99.9th percentile thresholds. Horizontal lines near the *x* axis represent observed frequencies of the event, diagonal lines are lines of perfect reliability, and forecast frequencies (%) within each probability bin are shown as open circles (all nine CAEs had very similar probability distributions, so the circles lie atop each other). Points lying in grey-shaded regions had skill compared to climatological forecasts as measured by the Brier skill score (Brier 1950; Wilks 2011). Values were not plotted for a particular bin if fewer than 500 grid points had forecast probabilities in that bin over the CONUS east of 105°W and all 26 forecasts.

		Percentile threshold					
		95.0	97.5	99.0	99.5	99.75	99.9
Experiment	3kmCent_3kmPert	0.000342	0.000434	0.000507	0.000551	0.000577	0.000596
	3kmCent_15kmPert	0.000359	0.000442	0.000505	0.000548	0.000575	0.000595
	3kmCent_GEFSPert	0.000351	0.000433	0.000504	0.000543	0.000573	0.000596
	15kmCent_3kmPert	0.000347	0.000445	0.000521	0.000553	0.000577	0.000598
	15kmCent_15kmPert	0.000345	0.000448	0.000526	0.000555	0.000577	0.000597
	15kmCent_GEFSPert	0.000348	0.000446	0.000527	0.000561	0.000582	0.000600
	GFSCent_3kmPert	0.000363	0.000434	0.000524	0.000558	0.000578	0.000597
	GFSCent_15kmPert	0.000364	0.000450	0.000529	0.000562	0.000581	0.000597
	GFSCent_GEFSPert	0.000415	0.000496	0.000565	0.000582	0.000591	0.000602

Table 3.5. Reliability component of the Brier score decomposition (Murphy 1973; smaller is better) aggregated over all 26 1–12-h 3-km forecasts of 1-h accumulated precipitation for various percentile thresholds, computed over the CONUS east of 105°W (Fig. 3.1) with a 100-km neighborhood length scale. These values correspond to the curves in Fig. 3.9.

3.4.2.2. CAE forecast sensitivity to IC perturbations

CAE precipitation forecast sensitivity to IC perturbations somewhat depended on the central initial state. For example, given identical 3-km central initial states, 3-km IC perturbations rarely yielded better CAE forecasts than 15-km IC perturbations over the first 12 h (compare solid and long-dashed purple curves in Figs. 3.8–3.10; also see Table 3.5 and the top row of each panel in Fig. 3.11). But, when the CAEs had common GFS or 15-km central initial states, there were more instances where 3-km IC perturbations led to better forecasts than 15-km IC perturbations (Figs. 3.8–3.10; also see Table 3.5 and the second and third rows from the top of each panel in Fig. 3.11). Overall, any benefits of decreasing IC perturbation horizontal grid spacing from 15 to 3 km were reserved for situations where central initial states had convection-parameterizing horizontal grid spacings that were associated with poorer short-term forecasts. There were no consistent benefits of 3-km IC perturbations

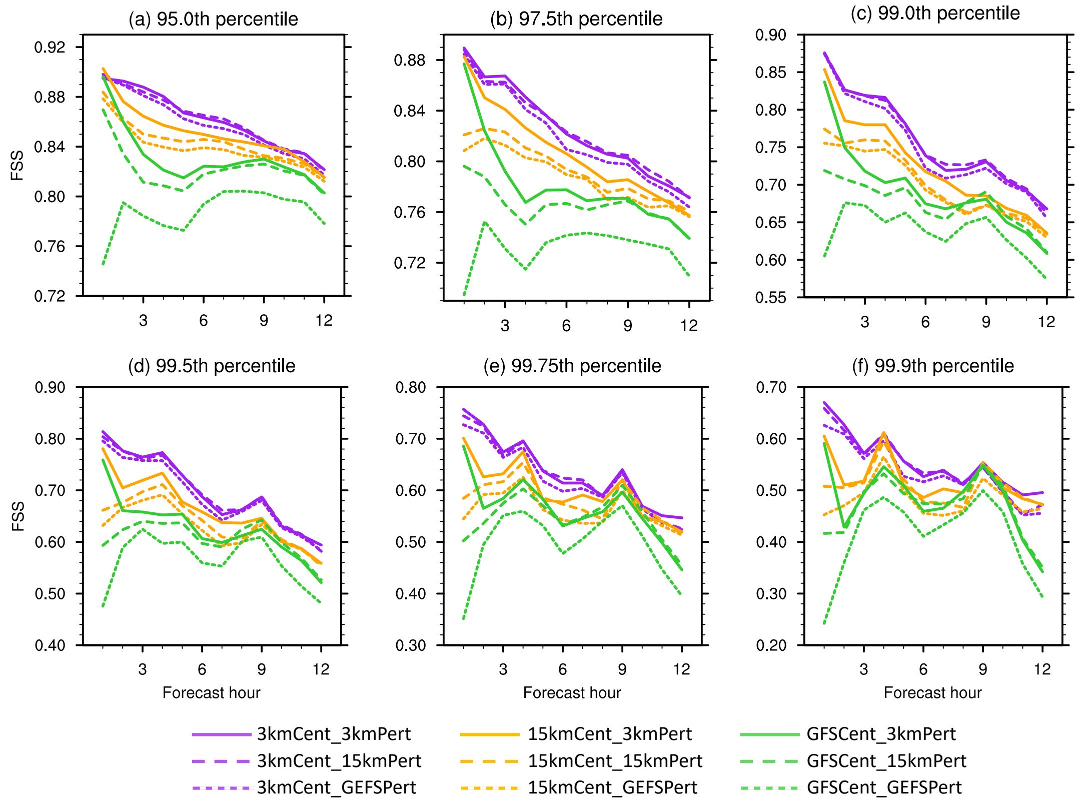

Fig. 3.10. Exactly as in Fig. 3.6, but zoomed-in over the first 12 forecast hours to more easily visualize differences between CAEs with identical central initial states but different IC perturbations (which have common line colors but different line patterns).

relative to 15-km IC perturbations when CAEs had demonstrably preferable 3-km central initial states.

However, 1–12-h forecasts with 3-km IC perturbations were usually better than those with GEFS IC perturbations when holding central initial states constant (e.g., compare solid and short-dashed curves with common colors in Figs. 3.8 and 3.10; also note the middle third of each panel in Fig. 3.11 has more statistically significant differences than the top third). CAEs with 15-km IC perturbations also typically outperformed CAEs with GEFS IC perturbations given common central initial states (Figs. 3.8 and 3.10; also see the bottom third of each panel in Fig. 3.11). The greatest benefit of 15- and 3-km EnKF IC perturbations compared to GEFS IC

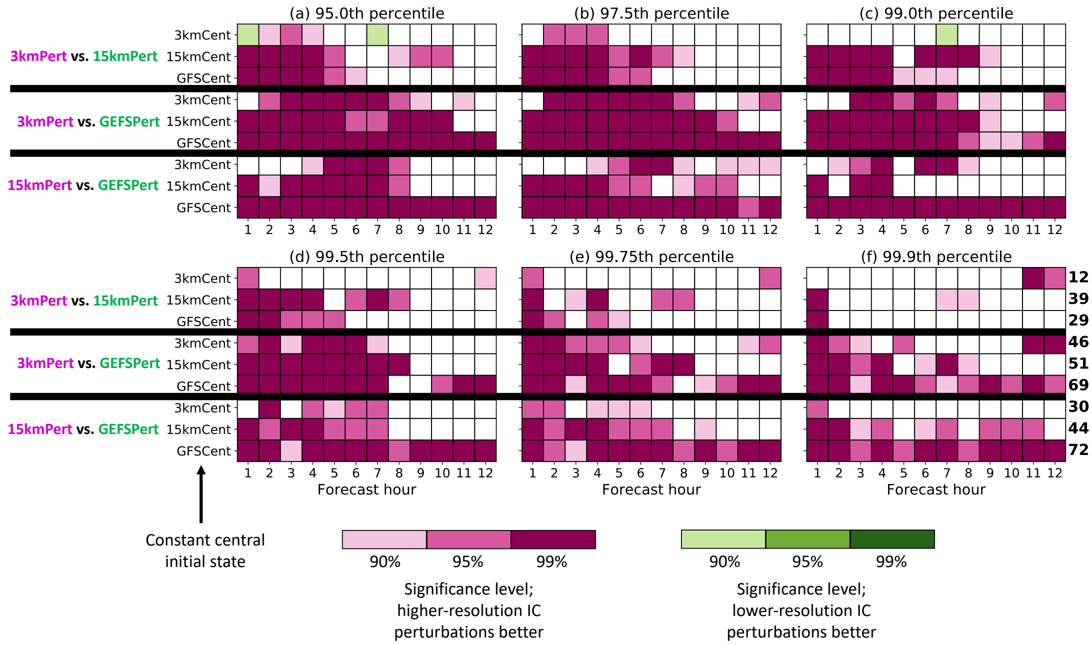

Fig. 3.11. As in Fig. 3.7, but for different comparisons that assess the impact of changing IC perturbation resolution. Specifically, in each panel, a given row represents a fixed set of central initial states; "3kmCent", "15kmCent", and "GFSCent" refer to central initial states provided by 3-km EnKF ensemble mean analyses, 15-km EnKF ensemble mean analyses, and GFS analyses, respectively. Each panel is broken into thirds to represent different comparisons, and pink text denotes the experiment in each comparison with the higher resolution IC perturbations. The top third compares experiments with 15- and 3-km IC perturbations ("3kmPert vs. 15kmPert"). The middle third compares experiments with 3-km and 0.5° GEFS IC perturbations ("3kmPert vs. GEFSPert"). The bottom third compares experiments with 15-km and 0.5° GEFS IC perturbations ("15kmPert vs. GEFSPert"). Pink shadings indicate that higher resolution IC perturbations led to statistically significantly higher FSSs for the fixed central initial state, while green shadings indicate that lower resolution IC perturbations led to statistically significantly higher FSSs for the fixed central initial state. White cells indicate that aggregate FSSs of two experiments with varied IC perturbations but common central initial states were not statistically significantly different at the 90% confidence level or higher. Annotations to the right of (f) represent the number of occurrences where, for the given row, experiments with higher resolution IC perturbations had statistically significantly higher FSSs than experiments with lower resolution IC perturbations across all 6 percentile thresholds and all forecast hours (i.e., the total number of pink-shaded boxes in each row across all 6 panels).

perturbations occurred when GFS analyses provided central initial states (e.g., see the sixth row from the top and bottom row of each panel in Fig. 3.11), illustrating that combining EnKF-based IC perturbations with GFS analyses is preferable to employing purely global ICs. Moreover, differences between CAEs with 15- and 3-km IC perturbations were generally smaller than differences between CAEs with

EnKF- and GEFS-based IC perturbations, consistent with GEFS IC perturbations possessing very different spectral characteristics (Fig. 3.4) and spread (Fig. 3.5) than the two sets of EnKF IC perturbations, which resembled each other in many ways (Figs. 3.4, 3.5).

Short-term precipitation forecast sensitivity to IC perturbations appeared to have some association with forecast evolution of small-scale perturbations, which exert greater control on short-term forecasts than large-scale perturbations. For instance, given 3-km central initial states, the CAE with 15-km IC perturbations quickly spun-up fine-scale structures and had nearly identical perturbation spectra to the CAE with 3-km IC perturbations at scales < 100 km by 3 h (Figs. 3.12a,b). This fast spin-up of small-scale structures from 15-km IC perturbations may be related to why CAEs with 15- and 3-km IC perturbations had similar short-term precipitation forecast skill given common 3-km central initial states (e.g., Figs. 3.8–3.11).

Similarly, the CAE with GEFS IC perturbations and 3-km central initial states also quickly spun-up perturbations at scales < 100 km over the first 3 h. However, some differences between CAEs with GEFS and limited-area EnKF IC perturbations remained through 6 h (Figs. 3.12a–c), consistent with most statistically significant differences regarding precipitation forecast skill between CAEs with EnKF and GEFS IC perturbations occurring before 6 h (see the fourth and seventh rows from the top of each panel in Fig. 3.11). By 12 h, all three CAEs with 3-km central initial states had similar perturbation spectra at scales < 800 km (Fig. 3.12d), consistent with little precipitation forecast sensitivity to IC perturbations after 12 h (e.g., see purple curves

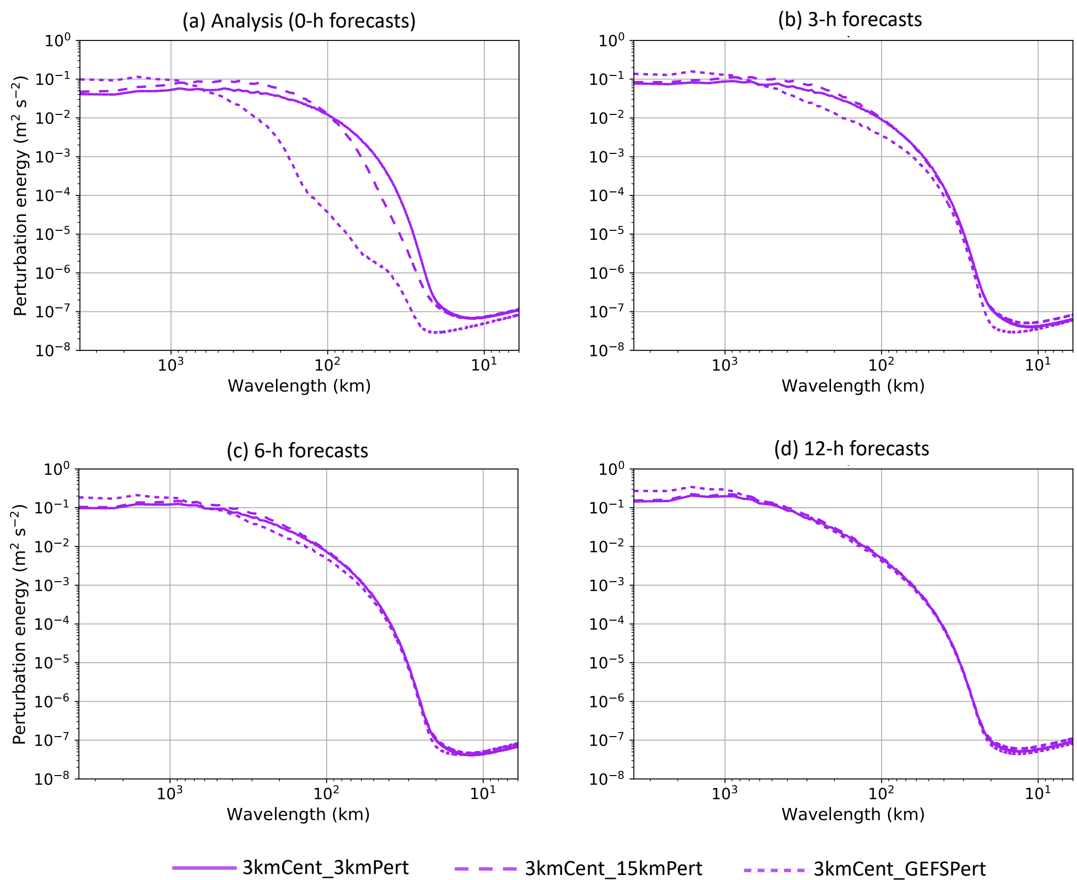

Fig. 3.12. Average 250-hPa zonal wind perturbation energy ($m^2\ s^{-2}$) over all 26 3-km forecasts and all 10 ensemble perturbations as a function of wavelength (km) for the three CAEs with 3-km central initial states for (a) analyses (0-h forecasts) and (b) 3-, (c) 6-, and (d) 12-h forecasts. Perturbations were defined with respect to the ensemble mean, and spectra were computed over the entire 3-km domain (Fig. 3.1), excluding points within 45 km of each lateral boundary. The discrete cosine transform (Denis et al. 2002) was used to compute spectra and spectral variance binning employed the method of Ricard et al. (2013).

in Fig. 3.6). Similar spectral evolutions occurred over the first 12 h for other meteorological variables (not shown).

These results echo previous studies indicating that small-scale (e.g., < 100 km) perturbations quickly develop from relatively coarse ensemble ICs once high-resolution model integration commences (e.g., Harnisch and Keil 2015; Tennant 2015; Johnson and Wang 2016; Raynaud and Bouttier 2016; Potvin et al. 2017; JW20). These findings are also commensurate with hypotheses that large-scale perturbations are more important drivers of error growth than small-scale perturbations (e.g., Durran and Gingrich 2014), as the absence of small-scale perturbations in GEFS ICs (Fig. 3.4) did not seem to fundamentally limit error growth in CAEs with GEFS IC perturbations (Fig. 3.12).

3.4.2.3. Overall sensitivities of short-term CAE forecasts

Considering all nine CAEs, those with 3-km central initial states usually had better 1–12-h precipitation forecasts than CAEs with 15-km central initial states, which in turn were typically better than CAEs with GFS central initial states, *regardless of IC perturbations* (e.g., Figs. 3.8–3.10; Table 3.5; note that all purple curves are usually above all orange curves, which are usually above all green curves in Figs. 3.8 and 3.10). The only systematic exception occurred at forecast hour 1, where the five CAEs with at least one 3-km IC component had the five best forecasts, suggesting that information content at convection-allowing scales provided by either central initial states or IC perturbations is helpful for the shortest forecasts. However, skill in the CAEs with 3-km IC perturbations and GFS or 15-km EnKF central initial states (i.e., 15kmCent_3kmPert and GFSCent_3kmPert) diminished between hours 1–2, sometimes rapidly (Figs. 3.8, 3.10).

Forecast skill characteristics over the first 2 h in 15kmCent_3kmPert and GFSCent_3kmPert appear related to precipitation spin-up. For instance, given GFS and 15-km central initial states, 3-km IC perturbations led to the most domain-total precipitation at forecast hour 1 (Fig. 3.13). Adding 3-km perturbations to relatively coarser GFS and 15-km fields surely led to imbalances; the small-scale perturbations

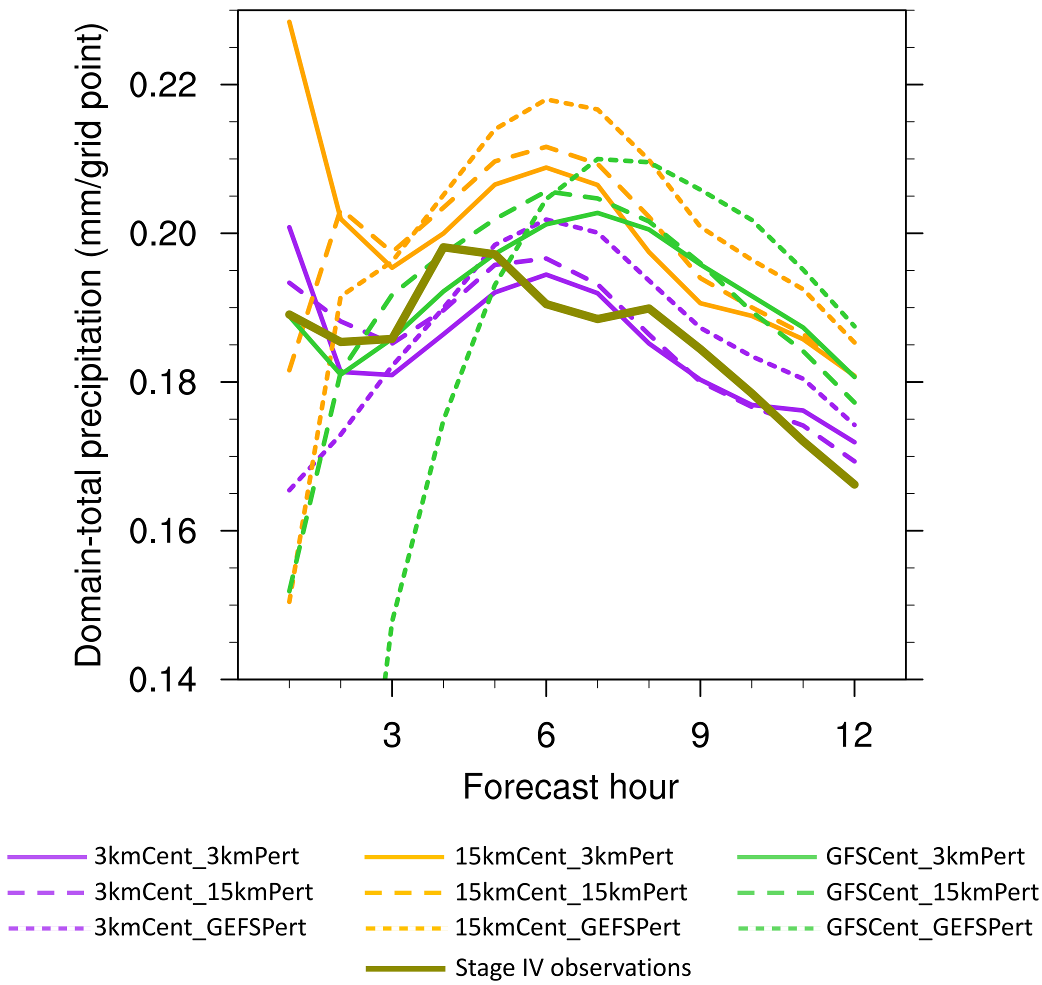

Fig. 3.13. Average 1-h accumulated precipitation (mm) per grid point over all 26 3-km forecasts and the CONUS east of 105°W (Fig. 3.1), computed on native grids as a function of forecast hour. These statistics were computed for all 10 ensemble members, but for readability, only ensemble means are shown. Values on the *x* axis represent ending forecast hours of 1-h accumulation periods (e.g., an *x*-axis value of 9 is for 1-h accumulated precipitation between 8–9 h). At forecast hours 1 and 2, GFSCent_GEFSPert domain-total precipitation was non-zero but below the *x* axis.

likely acted as noise that stimulated precipitation development. Although the greater domain-total precipitation resulting from 3-km IC perturbations did not always agree well with observed domain-total precipitation (e.g., see the solid orange line in Fig. 3.13), these enhanced precipitation elements were often placed correctly, given that spatial skill was greatly improved at forecast hour 1 by adding 3-km IC perturbations, rather than GEFS or 15-km IC perturbations, to GFS and 15-km central initial states

(Figs. 3.8, 3.10). Between 1–2 h, domain-total precipitation in 15kmCent_3kmPert and GFSCent_3kmPert decreased as imbalances were resolved (Fig. 3.13), which is consistent with the sometimes precipitous decline of skill in these two experiments over this period (Figs. 3.8, 3.10) and epitomizes the dominant influence of central initial states for short-term forecast evolution. Overall, after 3–6 h, domain-total precipitation provided similar conclusions as other metrics: for a fixed central initial state GEFS IC perturbations produced forecasts that were typically furthest from observations, 15- and 3-km IC perturbations added to 3-km central initial states yielded comparable performance, and CAEs with 3-km central initial states were usually closest to observations regardless of IC perturbations.

Furthermore, differences between CAEs with identical IC perturbations but different central initial states were statistically significant more often than differences between CAEs with identical central initial states but varied IC perturbations (compare Figs. 3.7 and 3.11). Most importantly, the four EnKF-only experiments (bold experiments in Table 3.3) clearly revealed that CAEs with 3-km central initial states were statistically significantly better than CAEs with 15-km central initial states (given constant IC perturbations) more often than CAEs with 3-km IC perturbations were statistically significantly better than CAEs with 15-km IC perturbations (given constant central initial states; compare the top two rows of each panel in Figs. 3.7 and 3.11).

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

3.5. Discussion

3.5.1. CONNECTION TO MIXED-RESOLUTION ENSEMBLE–VARIATIONAL DA SYSTEMS

Our results are broadly consistent with previous work that examined deterministic forecasts initialized from ensemble–variational (EnVar) DA systems (e.g., Hamill and Snyder 2000; Lorenc 2003; Wang et al. 2008; Wang 2010) with "dual-resolution" configurations, where a comparatively low-resolution ensemble provides background error covariances (BECs) for a relatively higher-resolution deterministic background. Specifically, several studies noted that deterministic forecasts were improved when increasing resolution of the deterministic background while holding ensemble perturbation resolution constant (e.g., Schwartz 2016; Lu et al. 2017; Pan et al. 2018; Wang et al. 2019). Conversely, studies isolating sensitivity to perturbation (i.e., BEC) resolution in dual-resolution EnVar DA systems have yielded mixed results and provide scant overall evidence that BECs coarser than the deterministic background systematically degrade subsequent forecasts (e.g., Schwartz et al. 2015c; Schwartz 2016; Lei and Whitaker 2017; Bédard et al. 2018, 2020; Kay and Wang 2020). Furthermore, some ensemble-based DA systems use simpler procedures to update ensemble perturbations relative to methods for updating deterministic backgrounds, finding few adverse impacts from the simplifications and implicitly acknowledging the overriding importance of central states (e.g., Buehner et al. 2017; Lorenc et al. 2017; Bédard et al. 2018).

Therefore, past research collectively suggests that ensemble perturbation resolution likely has secondary importance relative to resolution of deterministic backgrounds in dual-resolution EnVar DA systems, mirroring our results for short-term CAE forecasts. However, most previous studies assessing sensitivity of dual-resolution EnVar analyses to BEC resolution focused on DA systems at convection-parameterizing scales. The exception is Schwartz (2016), who found that 20- and 4-km BECs yielded 4-km EnVar analyses that initialized similar deterministic forecasts. But, absence of continuous cycling in their comparisons meant potential benefits of 4-km BECs could not accumulate through time.

Moreover, even if BEC resolution is *less important* than resolution of deterministic backgrounds, BEC resolution cannot necessarily be *entirely ignored* for convective-scale DA applications. For instance, there are likely situations where convection-parameterizing BECs cannot provide relevant spatiotemporal details about small-scale features represented in convection-allowing EnVar backgrounds, potentially leading to suboptimal analyses. Thus, while it may be unnecessary to initialize CAE "free forecasts" with convection-allowing IC perturbations, it is conceivable that convection-allowing BECs are in fact critical to producing optimal convection-allowing analyses. Future studies should investigate this topic to better understand whether BEC resolution can be degraded without also degrading convective-scale EnVar analyses.

3.5.2. Theoretical aspects

Our findings are limited by the effectiveness of Gaussian-based DA methodologies, like the EnKF, to effectively represent posterior means and perturbations at various spatial resolutions. For example, in a highly idealized scenario, Posselt and Bishop (2012) suggested that given non-Gaussian priors (states before assimilation), EnKFs reasonably represent posterior means (i.e., central initial states) but poorly estimate posterior covariances (i.e., IC perturbations). This deficiency can directly limit forecast performance of convective-scale NWP systems that use EnKFs (Poterjoy et al. 2017, 2019). Assuming that 3-km EnKF priors were more non-Gaussian than 15-km EnKF priors, our findings are consistent with past research: benefits of decreasing central initial state horizontal grid spacing from 15 to 3 km suggest reliable EnKF updates for the mean at all scales, while lack of benefits from decreasing IC perturbation horizontal grid spacing from 15 to 3 km suggests the 3-km EnKF may not properly represent convective-scale posterior covariances. These concepts are consistent with the discussion in section 3.5.1 and may provide a theoretical basis for using dual-resolution DA systems to initialize CAEs.

3.5.3. Similarities with JW20

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

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

3.5.4. FURTHER INSIGHTS INTO S21

Comparison of the EnKF-only experiments (bold experiments in Table 3.3) indicates that S21's 3-km EnKF initialized better 6–12-h forecasts than their 15-km EnKF due to increased resolution of its central initial state, not because of finer-resolution IC perturbations. Furthermore, S21 noted that ICs produced by "blending" small scales from 3-km EnKF analysis members with large scales from corresponding GEFS IC members led to better ~18–36-h CAE forecasts than ICs provided by unblended 3-km EnKF analysis ensembles (i.e., x_i^{3km}). However, because individual members from 3-km EnKF and GEFS IC ensembles were blended (i.e., x_i^{3km} and x_i^{GEFS} were blended for the ith member to create new ICs), S21 could not assess whether changing the *large-scale central initial state* or the *large-scale IC perturbations* was responsible for the success of blending.

Although we did not perform experiments to explicitly examine the impact of modifying large-scale central initial states and IC perturbations, our experiments nonetheless provide some insights on S21's findings. Specifically, after ~30 h at the 95th and 97.5th percentile thresholds (Figs. 3.6a,b) and between ~22–26 h at higher thresholds (Figs. 3.6c–f), CAEs with GFS central initial states outperformed CAEs with EnKF central initial states regardless of IC perturbation resolution. Moreover, ~18–36-h forecast sensitivity to using global IC perturbations was relatively modest: given 3-km central initial states, while employing GEFS IC perturbations sometimes boosted 30–36-h FSSs compared to using 15- or 3-km IC perturbations, much bigger performance gains were realized by changing central initial states to GFS analyses (Figs. 3.6a,b).

Therefore, ~18–36-h forecasts were improved most by using central initial states provided by a global model. This finding suggests that S21's blended 3-km ICs yielded better next-day forecasts than unblended 3-km ICs because of forcing large-scale central initial states to global model large scales during blending, rather than forcing large-scale IC perturbations to those provided by a global model. It thus seems more critical to accurately depict large-scale central initial states than large-scale IC perturbations for next-day CAE forecasts.

3.6. *Summary and conclusions*

Nine sets of 36-h, 10-member CAE forecasts were produced over the CONUS for a 4-week period spanning April–May 2017. The various CAEs differed solely with regard to their central initial states and IC perturbations (Tables 3.3, 3.4) and were verified with a focus on precipitation east of the Rockies.

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

Overall, considering all nine CAEs, in aggregate, the three with 3-km central initial states produced better short-term precipitation forecasts than CAEs with GFS or 15-km central initial states, *regardless of IC perturbations*. Therefore, while increasing IC perturbation resolution can be helpful in some instances, it is far more important for central initial states than for IC perturbations to possess convection-allowing horizontal grid spacing for short-term CAE forecasting applications; IC perturbation resolution is subordinate to central initial state resolution. Of course, these findings must be interpreted within the context of this study, which focused primarily on strongly forced events and used EnKF-based IC perturbations limited by their various Gaussian assumptions. In weakly forced scenarios and with future advances in non-Gaussian DA methods and NWP models, conclusions could differ.

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

Another consequence of our findings is that deterministic forecasts initialized from central initial states (e.g., EnKF mean analyses) can potentially be used as proxies for CAE forecasts at $1/M$ the cost of an M-member CAE. As evidence of this possibility, comparison of deterministic forecasts initialized from GFS, 3-km EnKF mean, and 15-km EnKF mean analyses over the nested domain (Fig. 3.1) yielded identical conclusions as comparisons of 10-member CAEs solely differing by their central initial states: over the first ~12–21 h, 3-km EnKF mean analyses yielded the best forecasts and GFS analyses the worst, whereas forecasts initialized from GFS analyses were comparable to or better than those initialized from limited-area EnKF analyses after ~18–24 h (Fig. 3.14). Therefore, CAE developers may only need to

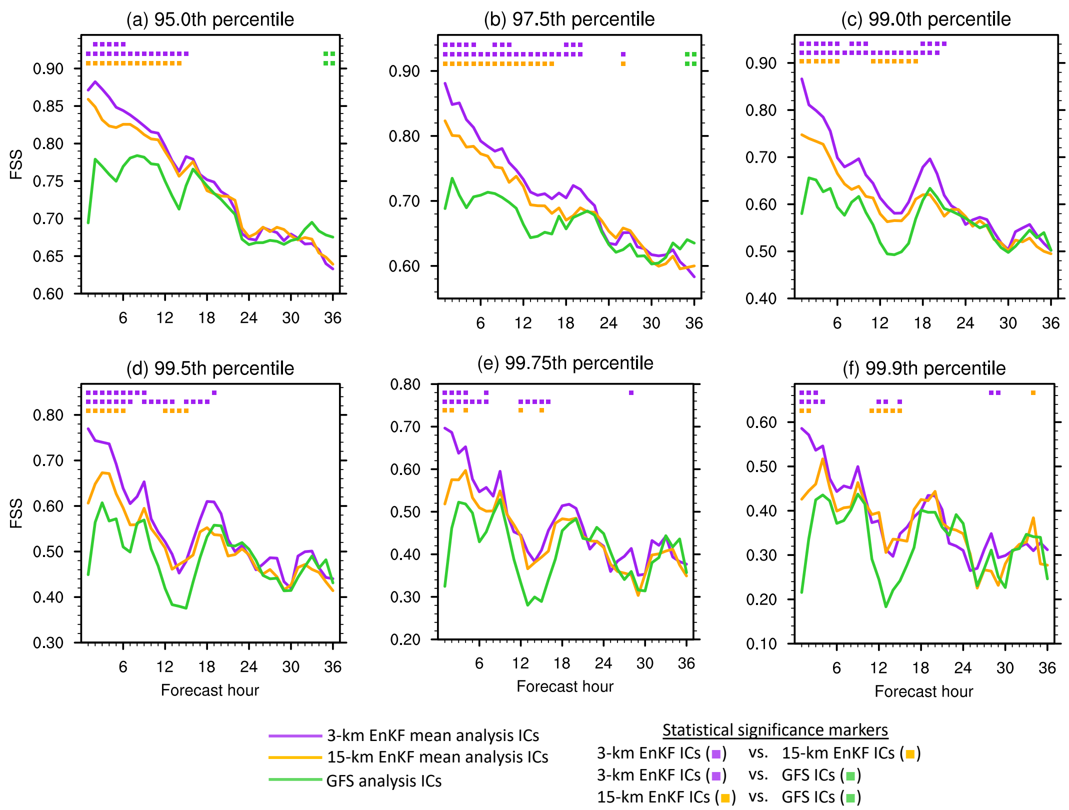

Fig. 3.14. Fractions skill scores (FSSs) over the CONUS east of 105°W (Fig. 3.1) with a 100-km neighborhood length scale aggregated over all 26 forecasts of 1-h accumulated precipitation for deterministic 3-km forecasts initialized from 3-km EnKF mean analyses (purple), 15-km EnKF mean analyses (orange), and GFS analyses (green) for the (a) 95th, (b) 97.5th, (c) 99th, (d) 99.5th, (e) 99.75th, and (f) 99.9th percentile thresholds as a function of forecast hour. Values on the *x* axis represent ending forecast hours of 1-h accumulation periods (e.g., an *x*-axis value of 24 is for 1-h accumulated precipitation between 23–24 h). Symbols along the top axis denote instances where differences between two forecasts were statistically significant at the 95% level, with the three rows of colored symbols corresponding to the three comparisons in the legend to denote which forecast had statistically significantly higher FSSs. For example, in the middle row, purple symbols indicate the forecasts with ICs provided by 3-km EnKF mean analyses had statistically significantly higher FSSs than forecasts with ICs provided by GFS analyses, while green symbols indicate forecasts with ICs provided by GFS analyses had statistically significantly higher FSSs than forecasts with ICs provided by 3-km EnKF mean analyses. Absence of a symbol means the differences were not statistically significant at the 95% level. The *y*-axis scales are different in each panel.

initialize deterministic forecasts from central initial states during portions of experimentation, potentially saving resources and enabling trials over longer time periods.

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

Chapter 4: Comparing partial and continuously cycling ensemble Kalman filter data assimilation systems for convection-allowing ensemble forecast initialization

4.1. 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, 2015a, 2021; Harnisch and Keil 2015; Wheatley et al. 2015; Dowell et al. 2016, 2021; 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 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 (NAM), Rapid Refresh (RAP; Benjamin et al. 2016), and High-resolution Rapid Refresh (HRRR; Benjamin et al. 2016; Dowell et al. 2021) models, as well as NOAA's experimental real-time CAEs, the HRRR-Ensemble (HRRRE; Dowell et al. 2016, 2021) and "Warn-on

Forecast" system (Stensrud et al. 2009, 2013; Wheatley et al. 2015; Jones et al. 2016), all employ partial cycling[12]. 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, "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

---

[12] 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.

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.

4.2. 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.

4.2.1. MODEL CONFIGURATIONS

All EnKF experiments used identical model configurations as S21 (Table 4.1). Specifically, version 3.9.1.1 of the Advanced Research Weather Research and Forecasting (WRF) model (Skamarock et al. 2008; Powers et al. 2017) produced all

Parameter	WRF model setting
Model version	Version 3.9.1.1 of the Advanced Research WRF model
Horizontal grid spacing	15 and 3 km in the outer and inner domains, respectively
Time step	60 and 12 s in the 15- and 3-km domains, respectively
Number of vertical levels	51 (based on the Rapid Refresh model; Benjamin et al. 2016)
Model top	15 hPa
Microphysics parameterization	Thompson (Thompson et al. 2008)
Longwave and shortwave radiation parameterizations	Rapid Radiative Transfer Model for Global Climate Models (RRTMG) with ozone and aerosol climatologies (Mlawer et al. 1997; Iacono et al. 2008; Tegen et al. 1997)
Planetary boundary layer parameterization	Mellor–Yamada–Janjić (MYJ) (Mellor and Yamada 1982; Janjić 1994, 2002)
Land surface model	Noah (Chen and Dudhia 2001)
Cumulus parameterization	Tiedtke (15-km domain only; Tiedtke 1989; Zhang et al. 2011)

Table 4.1. WRF model settings for all experiments.

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. 4.1). The same physics options (Table 4.1) 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.

4.2.2. 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 4.2). 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

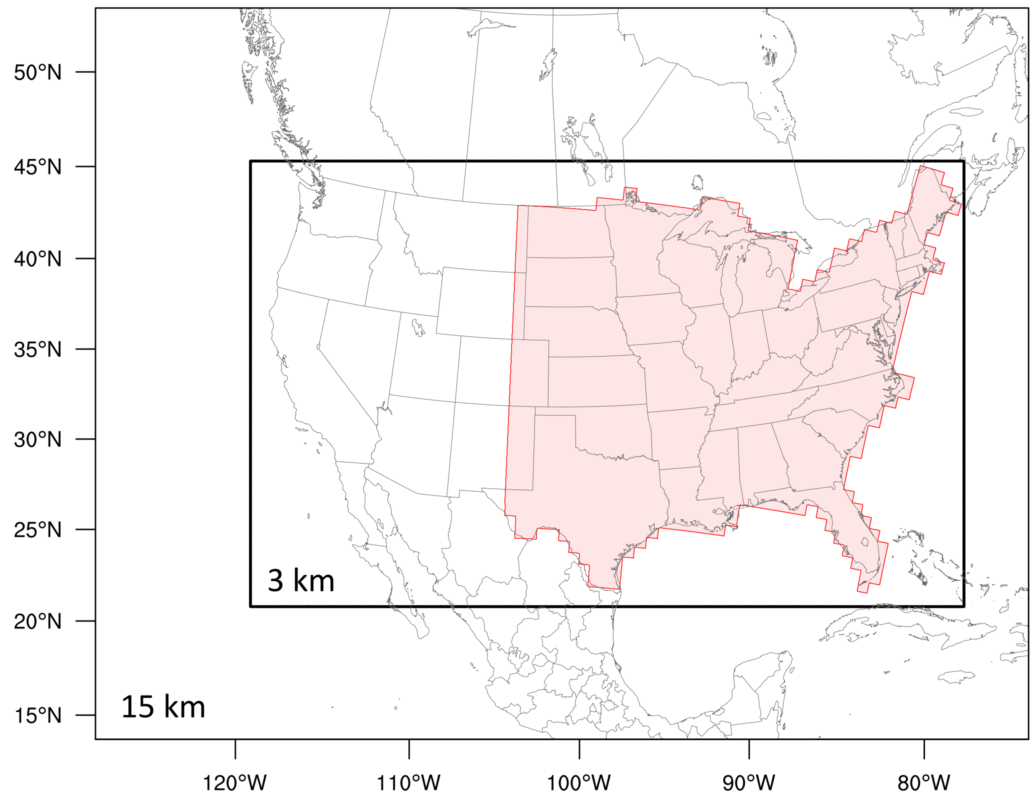

Fig. 4.1. Computational domain. Horizontal grid spacing was 15 km in the outer domain (415 x 325 points) and 3 km in the nest (1581 x 986 points). Objective precipitation verification only occurred over the red shaded region of the 3-km domain (CONUS east of 105°W).

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. 4.1); 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

Parameter	Setting
EnKF algorithm	Ensemble adjustment Kalman filter (EAKF; Anderson 2001, 2003; Anderson and Collins 2007)
Ensemble size	80 members
Cycling period	1 hour
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

Table 4.2. Summary of partial and continuously cycling EnKF configurations. See S21 for more details and justifications for these settings.

provide 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. 2015a, 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 4.3) given finite computing resources. Although future operational CAEs will likely be initialized from convection-allowing DA systems, as we further discuss in section 4.6, 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 4.2). Approximately 30,000 – 100,000 conventional observations were assimilated each cycle (Table 4.2), 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 4.2), 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 4.2) 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.

4.3. 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.

4.3.1. CONTINUOUSLY CYCLING ENKF

The 80-member continuously cycling 15-km EnKF in S21 and used again here ("CC_{EnKF}"; Fig. 4.2; Table 4.3) was initialized by downscaling the 0000 UTC 23 April 2017 0.25° GFS analysis onto the 15-km domain (Fig. 4.1) and adding random, correlated, Gaussian noise with zero mean, akin to the method for generating perturbed LBCs (section 4.2.2). 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.

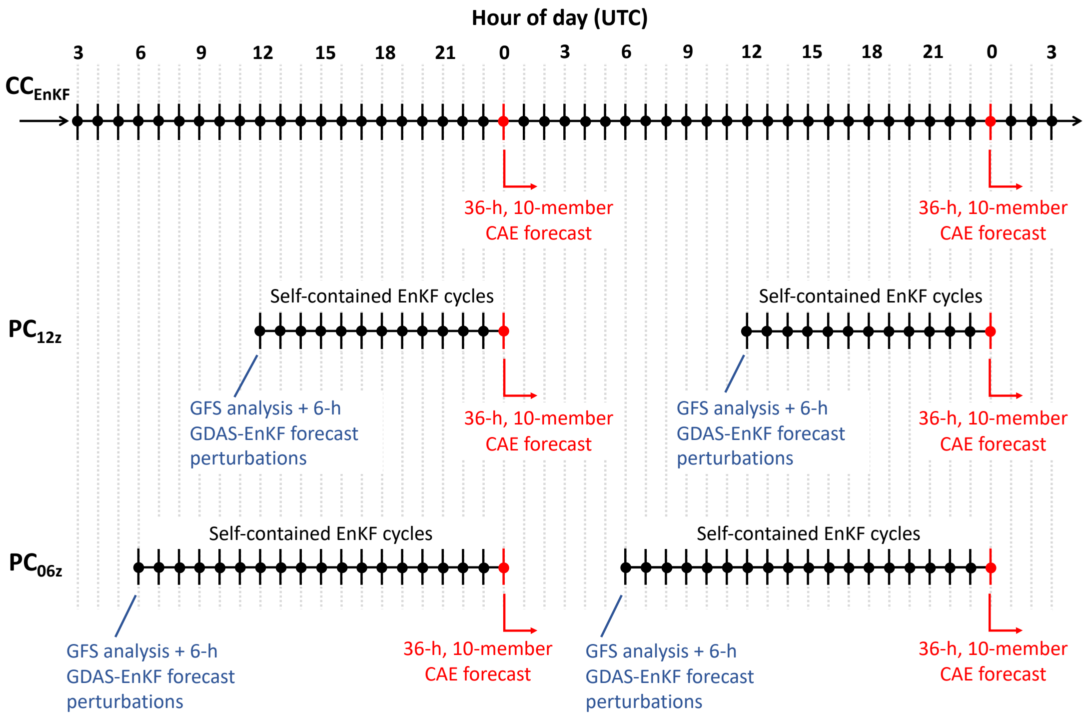

Fig. 4.2. Schematic diagram of CC_{EnKF} (top), PC_{12z} (middle), and PC_{06z} (bottom) cycling methodologies. Solid vertical lines with filled black circles represent EnKF analyses, and red denotes CAE forecast initialization times. Priors for EnKF analyses at 0600 and 1200 UTC in PC_{06z} and PC_{12z}, respectively, were 6-h GDAS-EnKF forecast perturbations re-centered about GFS analyses.

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 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).

Experiment name	Description
CC_{EnKF}	Continuously cycling EnKF initialized at 0000 UTC 23 April 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 April and 19 May 2017 (inclusive) by re-centering perturbations derived from 6-h GDAS-EnKF forecasts about 1200 UTC GFS analyses. The perturbations were inflated according to Fig. 4.3. 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 re-centering 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. 4.4). 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.

Table 4.3. Summary of experiments. Also see Fig. 4.2.

Members 1–10 from 0000 UTC posterior ensembles initialized 36-h forecasts over the nested domain (Fig. 4.1) 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.

4.3.2. PRIMARY PARTIAL CYCLING ENKF

The primary partial cycling EnKF ("PC_{12z}"; Table 4.3; Fig. 4.2) 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. 4.1). 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[13] 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

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

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). 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) given operational constraints sometimes 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 4.2), 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. 4.2). 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).

4.3.3. 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.

4.3.3.1. 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). However, 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 4.3; Fig. 4.2). 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 4.3.2).

4.3.3.2. 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). Thus, following the HRRRE, 6-h GDAS-EnKF forecast perturbations were inflated while initializing all partial cycling EnKFs (Table 4.3), with inflation factors linearly increasing from 1.0 (no inflation) at model level 26 to 2.0 at the lowest model level (Fig. 4.3). 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 4.2). However, the ideal external ensemble spread for partial

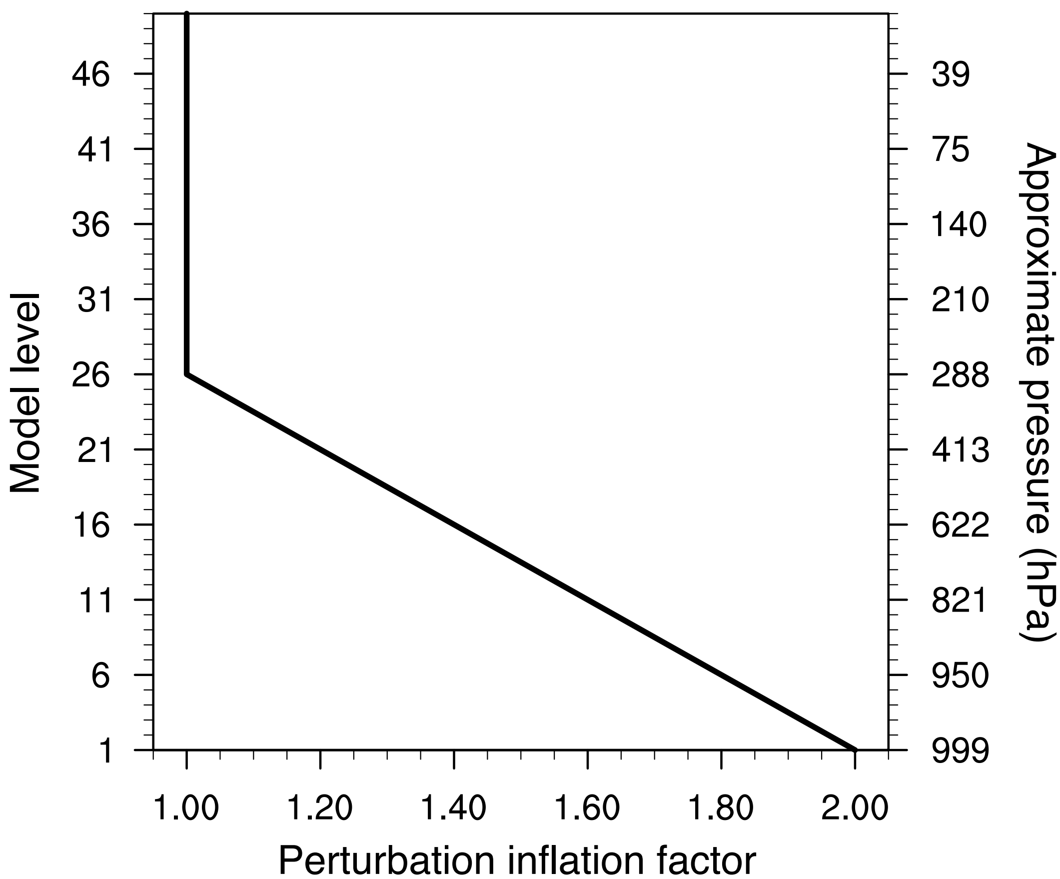

Fig. 4.3. Perturbation inflation factor as a function of model level (level 1 is nearest the ground). These inflation factors were applied to perturbations derived from 6-h GDAS-EnKF forecasts during partial cycling EnKF initialization. The approximate pressure (hPa) at each model level is given on the right axis.

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.

4.3.3.3. 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 4.3). 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.

4.3.3.4. 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 4.1) 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 microphysics fields to rapidly adjust to dynamic and thermodynamic states during the self-contained cycling period.

4.3.4. 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 EnKF analysis ensembles with large scales from corresponding members 1–10 of 0.5° GEFS IC ensembles using a low-pass, 6th-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. 4.4).

The 0000 UTC blended states initialized 36-h, 10-member CAE forecasts like unblended EnKF analysis ensembles ("CC_{EnKF_blend}"; Table 4.3). 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.

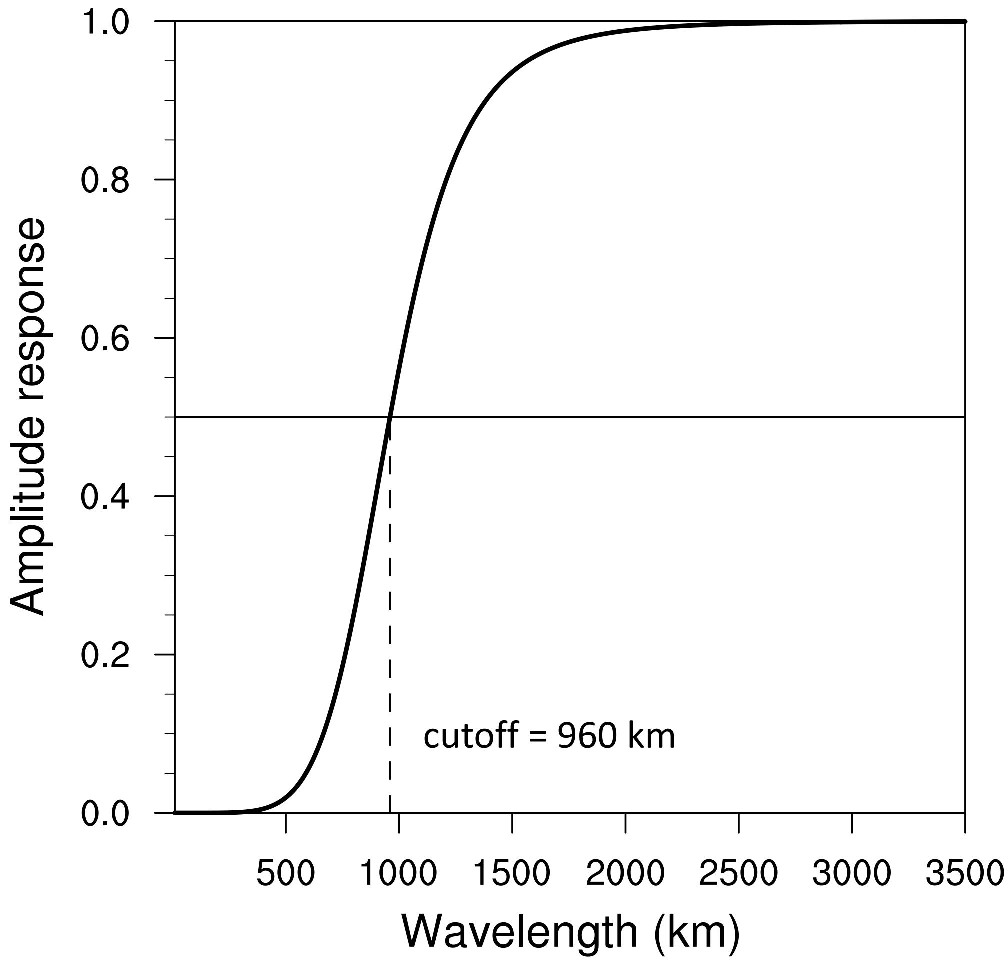

Fig. 4.4. Amplitude response (*y* axis) of a 6th-order implicit tangent filter as a function of wavelength (km) for a specified cutoff length of 960 km. In the context of this study, the curve denotes the contribution of GEFS ICs to blended ICs at a given wavelength (e.g., for wavelengths where the amplitude 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.

4.3.5. 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. 4.1) daily at 0000 UTC between 25 April and 20 May 2017 (inclusive). These GEFS-initialized CAE forecasts (Table 4.3) 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 spin-up was expected.

4.4. 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 4.3.3). As partial cycling EnKF spin-up can largely be controlled through DA parameters like observation errors, the following statistics were purely diagnostic, and ultimately, we hoped to relate differences between 0000 UTC-initialized CAE forecasts to differences between their ICs.

4.4.1. 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 4.4.2). Given the partial cycling initialization procedure (sections 4.3.2, 4.3.3), 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[14] and were drier at most levels (compare orange and purple lines in Figs. 4.5a,g and green and purple lines in Figs. 4.5b,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. 4.5e), but not at 0600 UTC

[14] 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.

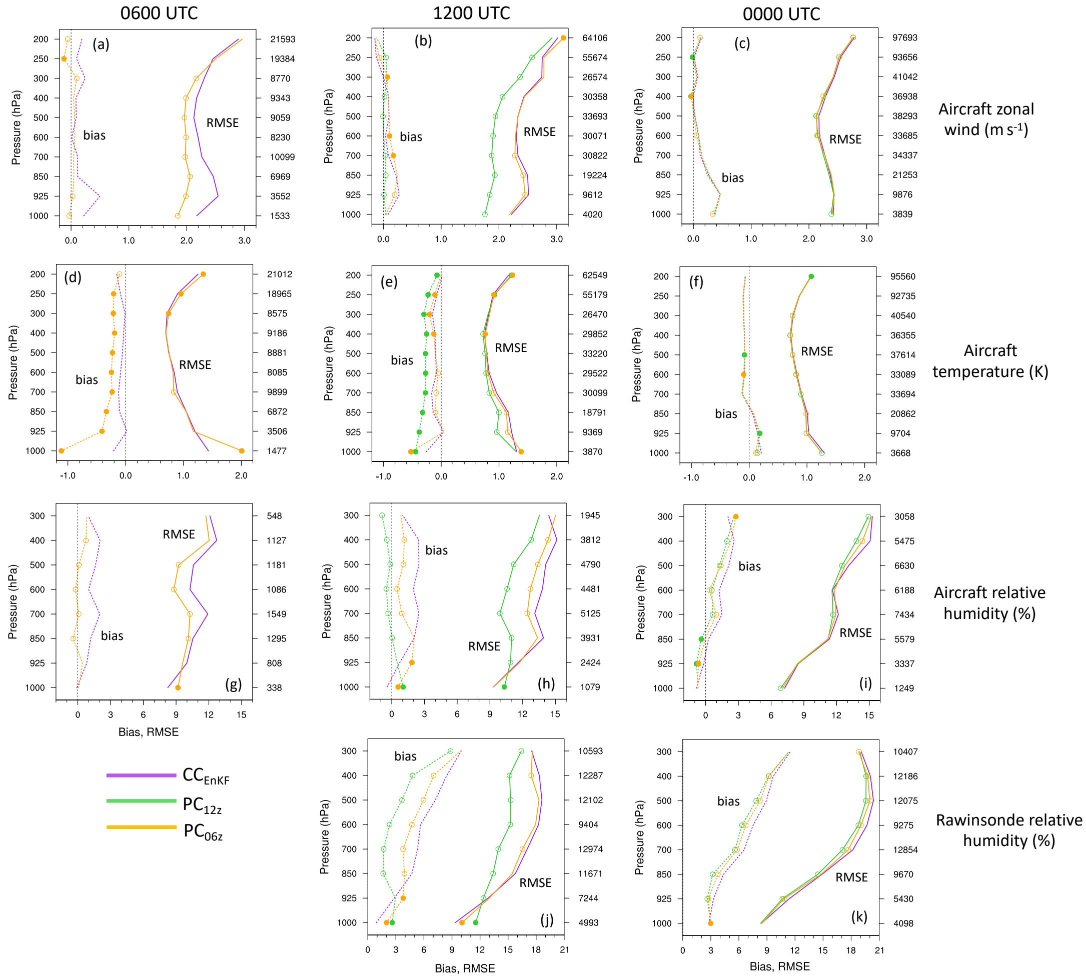

Fig. 4.5. Ensemble mean additive bias (model minus observations; short-dashed lines) and RMSE (solid lines) compared to (a)–(c) aircraft zonal wind (m s^{-1}), (d)–(f) aircraft temperature (K), (g)–(i) aircraft relative humidity (%), and (j),(k) rawinsonde relative humidity (%) observations aggregated over all prior ensembles valid at (a),(d),(g) 0600 UTC, (b),(e),(h),(j) 1200 UTC, and (c),(f),(i),(k) 0000 UTC between 0600 UTC 24 April 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.

(Fig. 4.5d). However, at both 0600 and 1200 UTC, GFS analyses had significant cold biases (Figs. 4.5d,e), possibly due to GFS physics errors (e.g., Zheng et al. 2017).

GFS analyses also had cold biases compared to rawinsonde observations between 925–300 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. 4.5a–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. 4.5i,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.

4.4.2. State-space characteristics

State-space characteristics were also assessed to explore partial cycling EnKF evolution. Regarding ensemble spread, 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[15] than CC_{EnKF} for zonal wind, temperature, and water vapor mixing ratio above 600 hPa and higher standard deviations below (compare purple and circular-marked dashed curves in Figs. 4.6a–c), reflecting initial variance inflation (Fig. 4.3). 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 (compare solid lines in Figs. 4.6a,b), even 18 h of self-contained cycles was not enough for PC_{06z} spread to match the larger CC_{EnKF} spread, suggesting spin-up was not complete. In addition, the partial cycling EnKFs had more moisture spread than CC_{EnKF} below 500 hPa (compare solid lines in Fig. 4.6c). 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 due to the relative scarcity of moisture observations to directly constrain EnKF analyses (see sample sizes on Fig. 4.5), 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. 4.6d–f) and means (Figs. 4.6g–i) of rain, snow, and graupel mixing ratios in PC_{06z} and PC_{12z} were comparable to or

[15] 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.

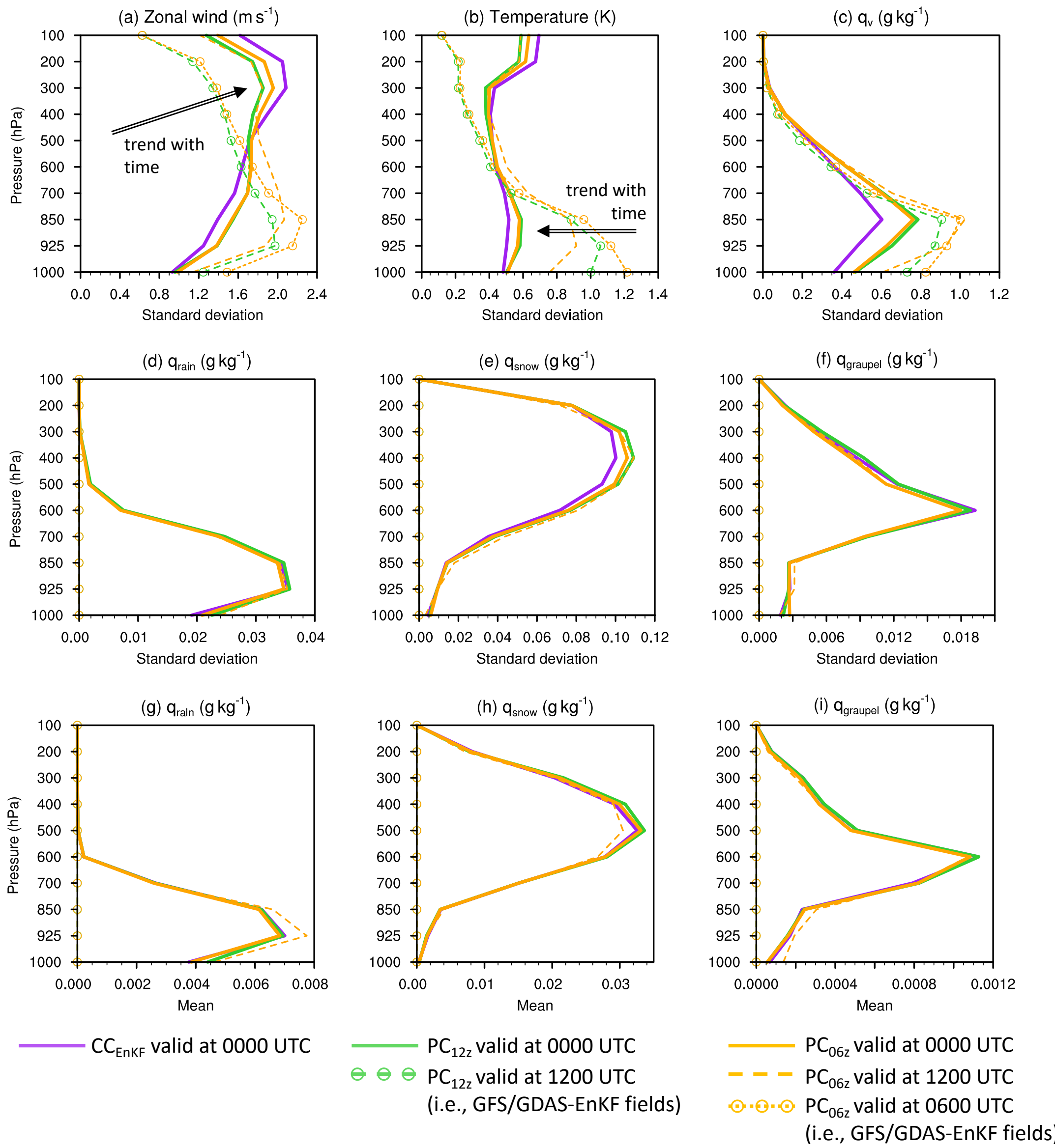

Fig. 4.6. Average standard deviation over land points within the portion of the 15-km domain colocated with the 3-km domain (Fig. 4.1) and all posterior ensembles between 0600 UTC 24 April and 0000 UTC 20 May 2017 (inclusive) at 0000 UTC (solid lines), 1200 UTC (long-dashed lines) and 0600 UTC (short-dashed 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 on (a) and (b) indicate how partial cycling EnKF statistics changed with time. (g)–(i) As in (d)–(f) except 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.

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.

4.4.3. 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. 4.6 (there was typically more 0000 UTC perturbation power below 500 hPa and less above in the partial cycling EnKFs compared to 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. 4.7). Moreover, the gap between difference spectra of the partial and continuously cycling EnKFs typically widened as wavelength increased, especially for temperature and moisture (Figs. 4.7c–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

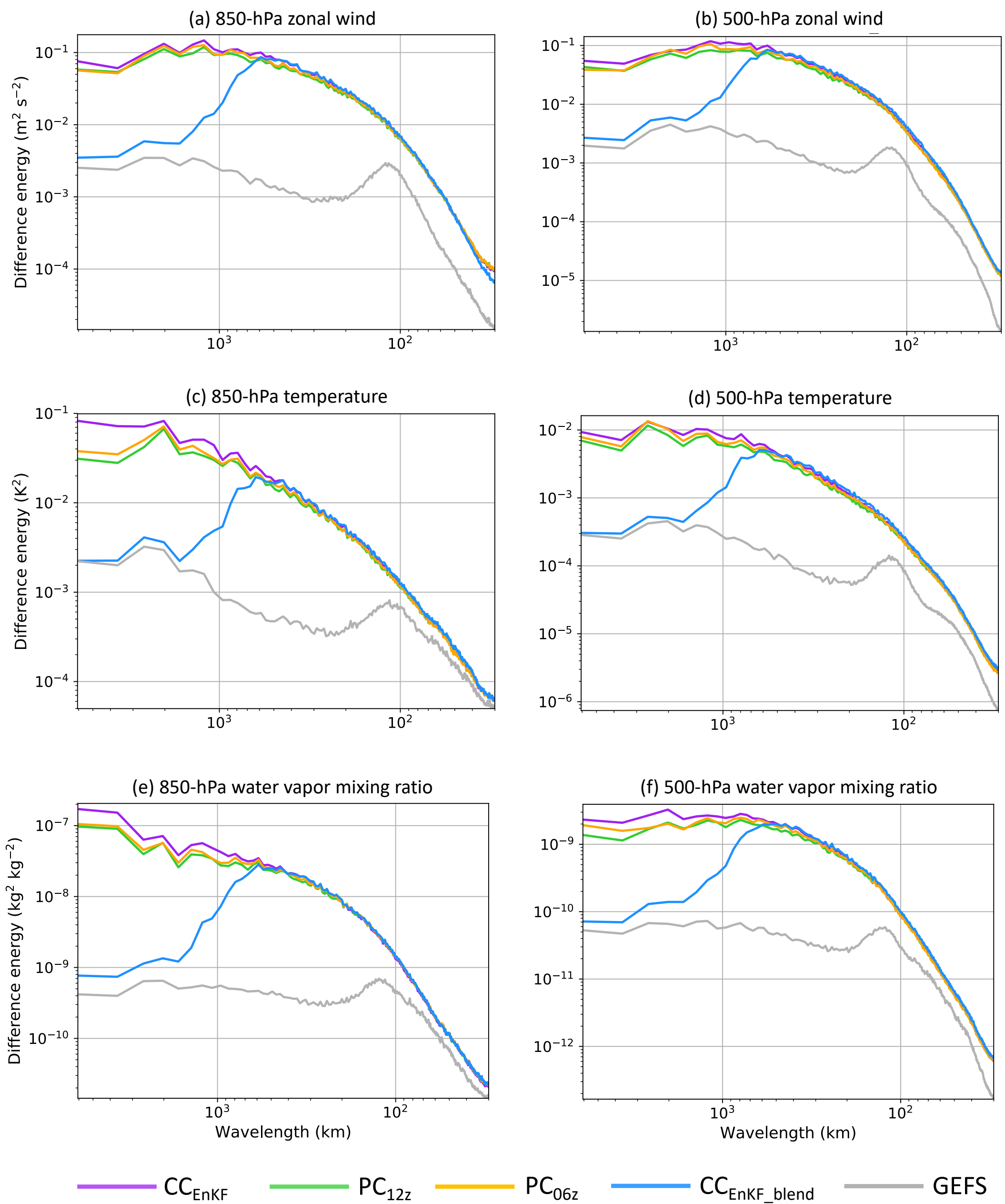

Fig. 4.7. Spectra for differences between GFS analyses and ensemble mean initial states for various experiments (see legend) as a function of wavelength (km) for (a) 850-hPa zonal wind (m^2s^{-2}), (b) 500-hPa zonal wind (m^2s^{-2}), (c) 850-hPa temperature (K^2), (d) 500-hPa temperature (K^2), (e) 850-hPa 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 April 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.

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.

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 band-pass 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

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(GFS_k - \overline{EnKF}_k)^2} \quad , \qquad (4.1)$$

where for the kth of N points, GFS_k is the GFS analysis and $\overline{EnKF}_k$ the EnKF mean analysis for a particular experiment (e.g., Table 4.3). Additionally, normalized reductions of RMSDs between two experiments (D) were computed as

$$D = \frac{RMSD_i - RMSD_j}{RMSD_j} \times 100\% \quad , \qquad (4.2)$$

where $RMSD_i$ and $RMSD_j$ are RMSDs of the ith and jth experiments, respectively (Table 4.3). 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., $RMSD_i < RMSD_j$) and was more similar to GFS analyses than experiment j.

Corroborating Fig. 4.7 and generally consistent with Figs. 4.5c,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. 4.8a–c) and in the 200–500- and 1000–1500-km wavelength bands (Figs. 4.8d–i). RMSDs were smaller in the 1000–1500-km band than the 200–500-km band, indicating EnKF mean and GFS analyses were more alike on larger,

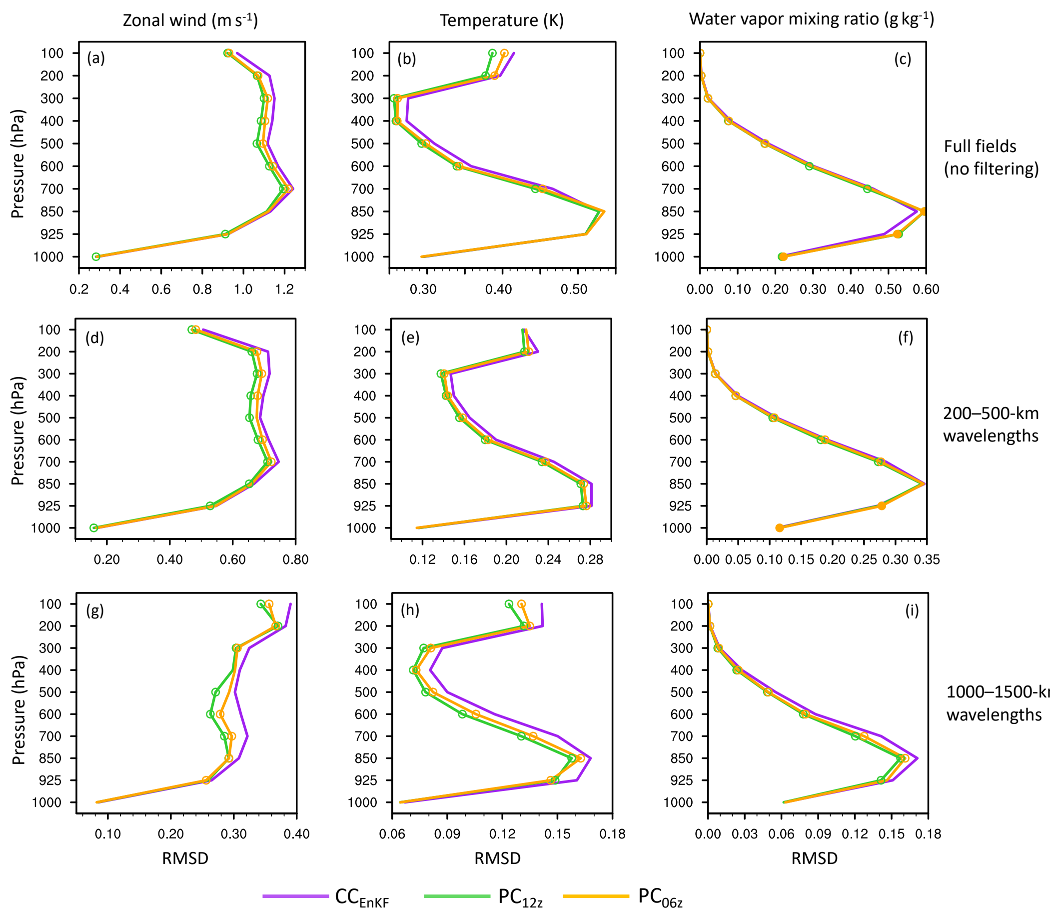

Fig. 4.8. Aggregate RMSDs between GFS and EnKF mean analyses [Eq. (4.1)] for (a),(d),(g) zonal wind (m s^{-1}), (b),(e),(h) temperature (K), and (c),(f),(i) water vapor mixing ratio (g kg^{-1}) over all 0000 UTC analyses between 25 April and 20 May 2017 (inclusive) for (a)–(c) full fields, (d)–(f) band-pass filtered fields for 200–500-km wavelengths, and (g)–(i) band-pass filtered fields for 1000–1500-km wavelengths. These statistics were computed over land points within the portion of the 15-km domain colocated 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. 4.5. Note that *x*-axis values are different in each panel.

more predictable scales. However, when letting $RMSD_i$ and $RMSD_j$ represent aggregate 0000 UTC RMSDs of PC_{06z} and CC_{EnKF}, respectively, D [Eq. (4.2)] typically became more negative as spatial scale increased (Figs. 4.9a–c), especially for temperature and moisture, for which PC_{06z} had ~10–20% smaller RMSDs than

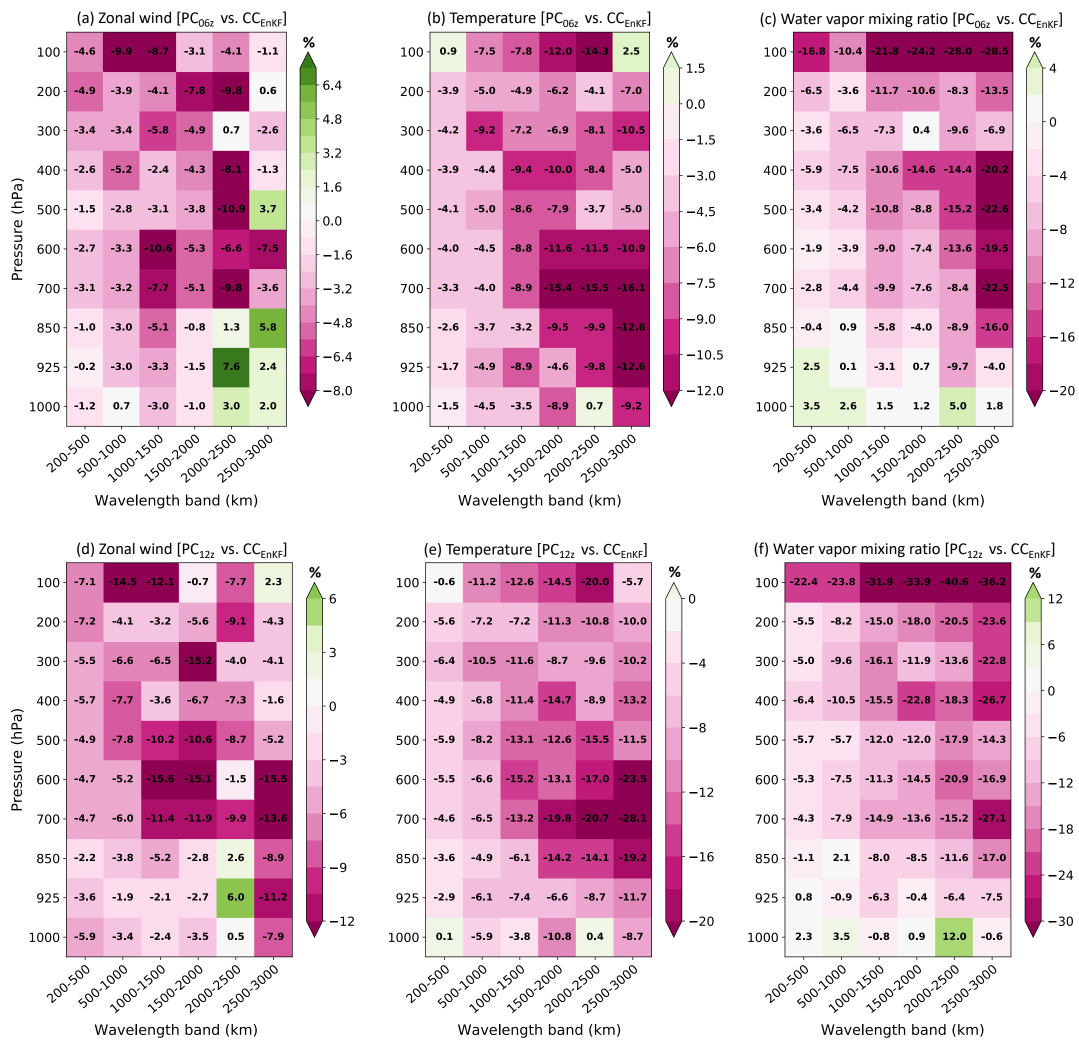

Fig. 4.9. Normalized RMSD reductions [%; Eq. (4.2)] between CC_{EnKF} and PC_{06z} for (a) zonal wind, (b) temperature, and (c) water vapor mixing ratio for various wavelength bands (km) and pressure levels (hPa) aggregated over all 0000 UTC analyses between 25 April and 20 May 2017 (inclusive). (d)–(f) As in (a)–(c) except for normalized RMSD reductions between CC_{EnKF} and PC_{12z}. Negative values indicate RMSDs with respect to GFS analyses [Eq. (4.1)] were smaller in PC_{06z} and PC_{12z} compared to CC_{EnKF}. Colorbars and their ranges are different in each panel. These statistics were computed over land points within the portion of the 15-km domain colocated with the 3-km domain.

CC_{EnKF} for many wavelengths ≥ 1000 km but RMSD reductions of typically only ~5% or less for scales < 1000 km (Figs. 4.9b,c). Therefore, relative differences between PC_{06z} and CC_{EnKF} generally grew as wavelength increased.

Collectively, Figs. 4.7–4.9 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. 4.7, 4.8; also compare Figs. 4.9a–c and Figs. 4.9d–f). The next section shows how the large-scale differences between partial and continuously cycling EnKF ICs impacted subsequent CAE precipitation forecasts.

4.5. Precipitation forecast verification

4.5.1. 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 precipitation forecasts were objectively compared to NCEP's Stage IV (ST4) analyses (Lin and Mitchell 2005) over the CONUS east of 105°W (Fig. 4.1), 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%–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.

4.5.2. Results

4.5.2.1. Fractions skill scores and ROC areas

To assess spatial skill, fractions skill scores [FSSs; Roberts and Lean (2008)] and areas under the relative operating 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. 4.10, 4.11) 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. 4.5, 4.6) 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. 4.7, 4.9), 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.

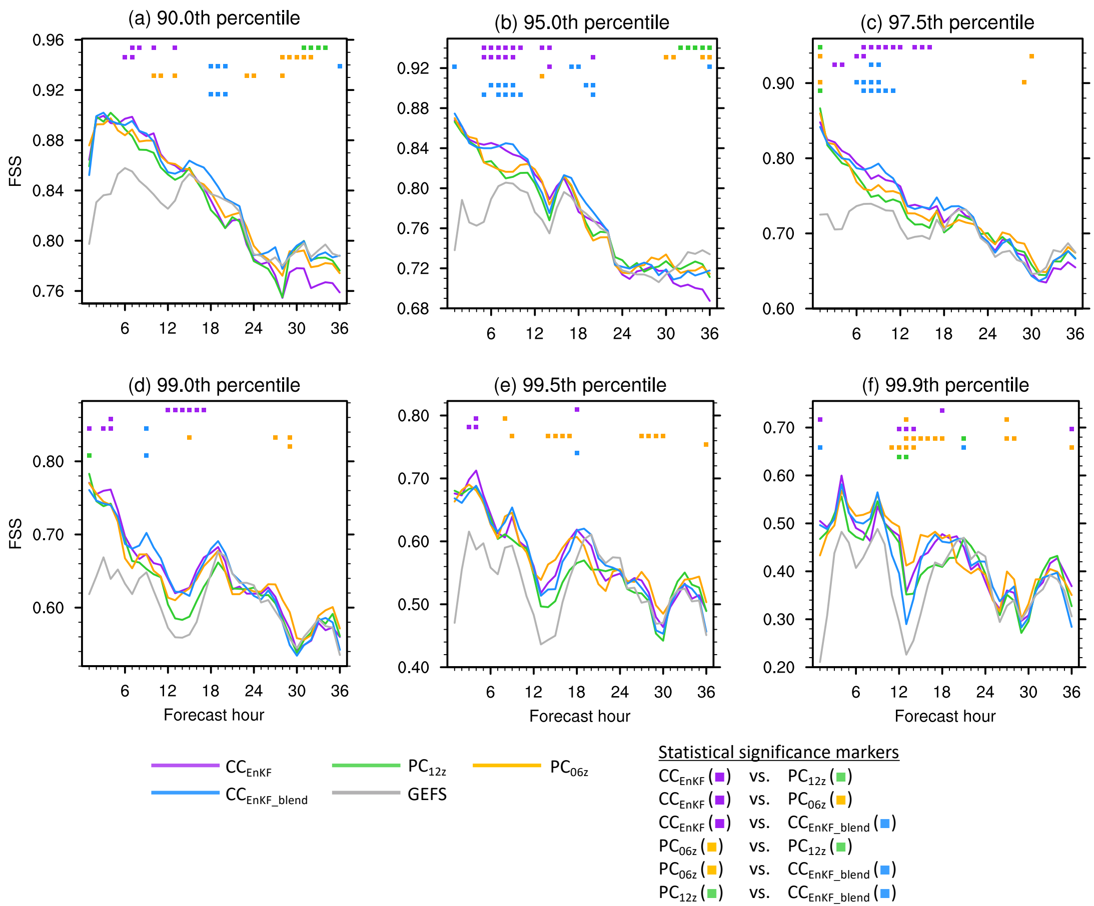

Fig. 4.10. Fractions skill scores (FSSs) over the CONUS east of 105°W (Fig. 4.1) with a 100-km neighborhood length scale for the (a) 90th, (b) 95th, (c) 97.5th, (d) 99th, (e) 99.5th, and (f) 99.9th percentile thresholds aggregated over all 26 3-km forecasts of 1-h accumulated precipitation as a function 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 23–24 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 4.3 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.

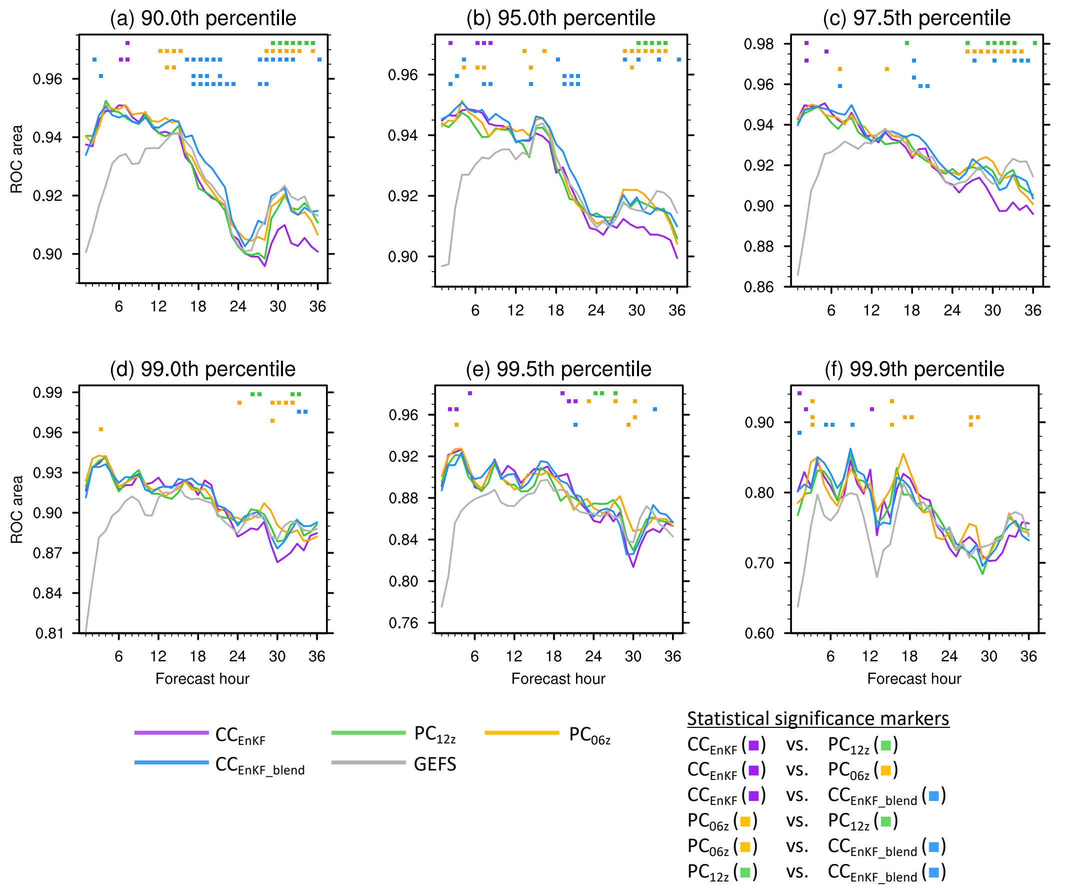

Fig. 4.11. As in Fig. 4.10, except for areas under the relative operating characteristic (ROC) curve computed using decision thresholds of 1%, 2%, 3%, 4%, 5%, 10%, 15%, …, 95%, and 100% and a trapezoidal method. Symbols along the top axis indicate forecast hours when differences between two ensembles were statistically significant at the 95% level as in Fig. 4.10 and denote the ensemble with statistically significantly higher ROC areas. The *y* axis scales are different in each panel.

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. 4.10, 4.11). The biggest degradations of CC_{EnKF} relative to ensembles with partial cycling ICs occurred after ~27 h, where some differences were statistically significant (Figs. 4.10a,b, 4.11a–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. 4.7), 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.

4.5.2.2. 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. 4.12). 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. 4.13), partial cycling EnKF ICs led to statistically significantly more reliable forecasts than unblended continuously cycling

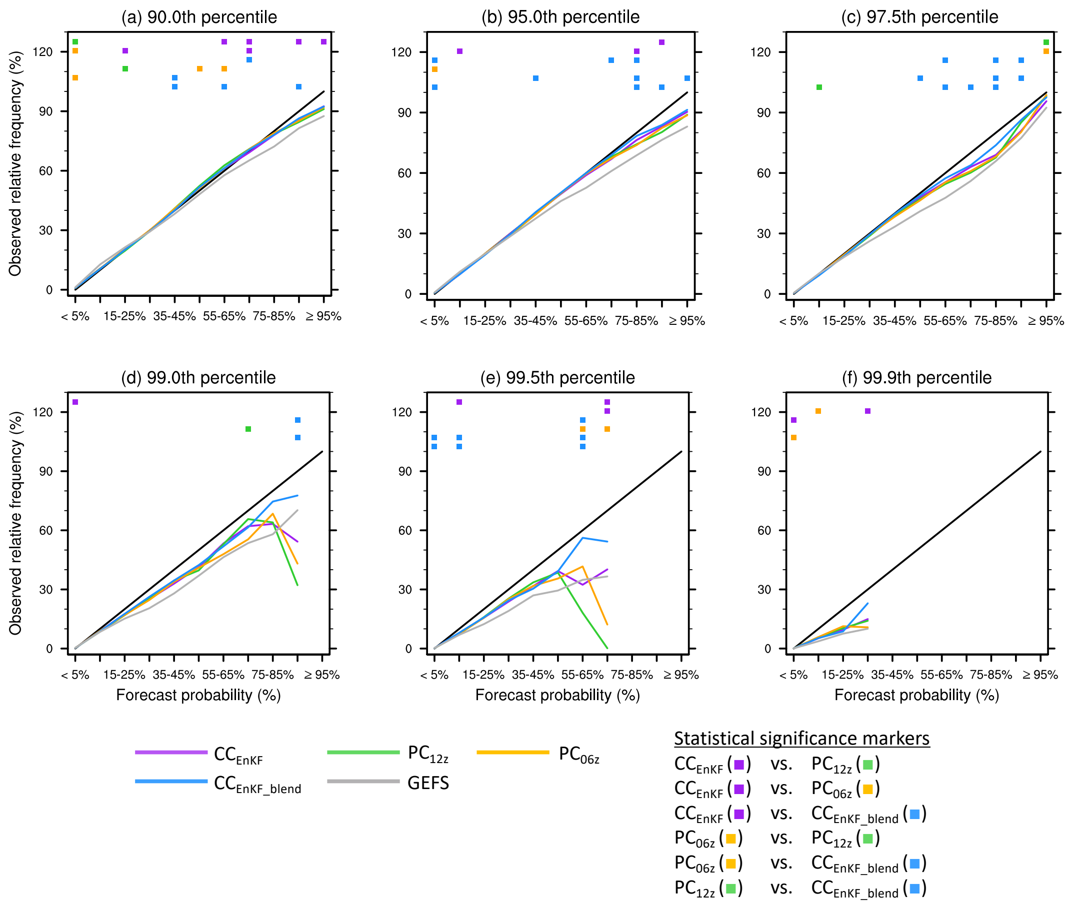

Fig. 4.12. Reliability diagrams computed over the CONUS east of 105°W (Fig. 4.1) with a 100-km neighborhood length scale aggregated over all 26 1–12-h 3-km forecasts of 1-h accumulated precipitation for the (a) 90th, (b) 95th, (c) 97.5th, (d) 99th, (e) 99.5th, and (f) 99.9th percentile thresholds. Diagonal 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. 4.10 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 ICs in many bins, especially for probabilities < 55% at the 90.0th–97.5th percentiles (Figs. 4.13a–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. 4.12, 4.13). Differences regarding

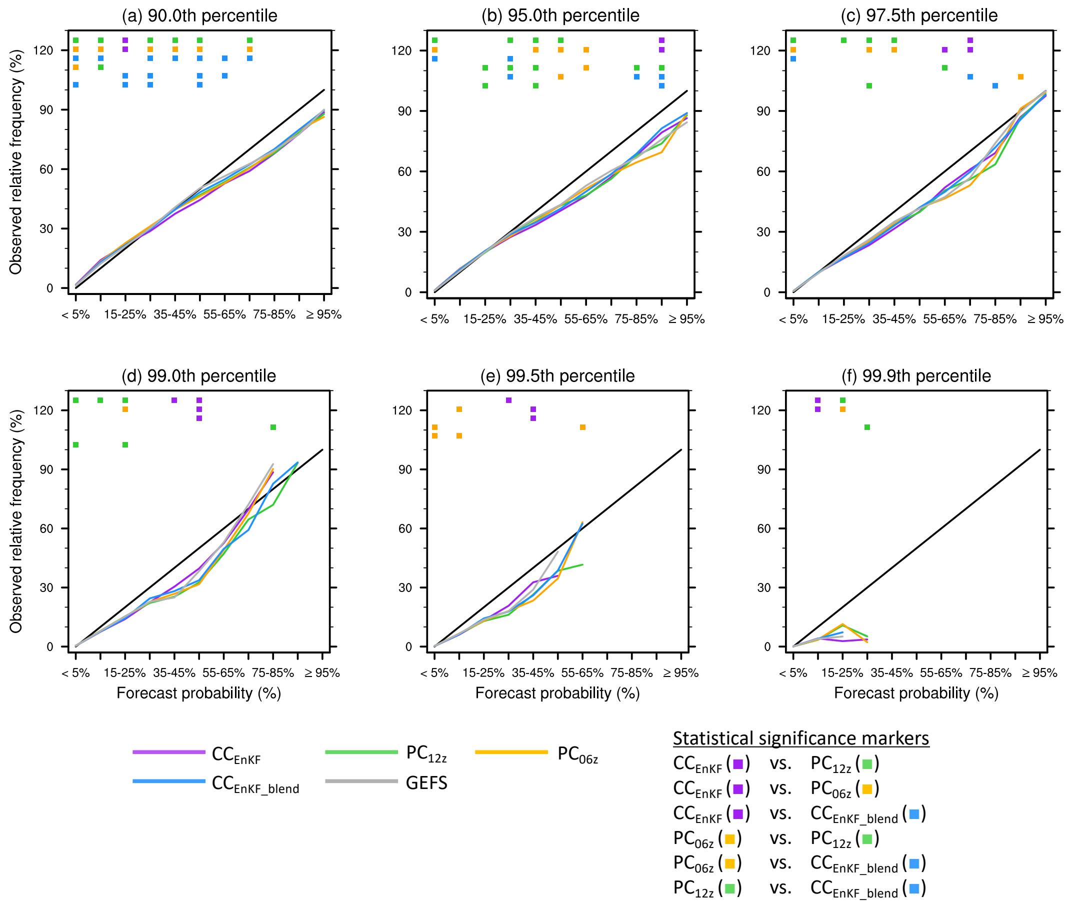

Fig. 4.13. As in Fig. 4.12, except statistics were aggregated over all 26 24–36-h 3-km forecasts of 1-h accumulated precipitation.

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).

4.5.2.3. 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. 4.14). 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. 4.14a,b) but closer to observations than CC_{EnKF} at

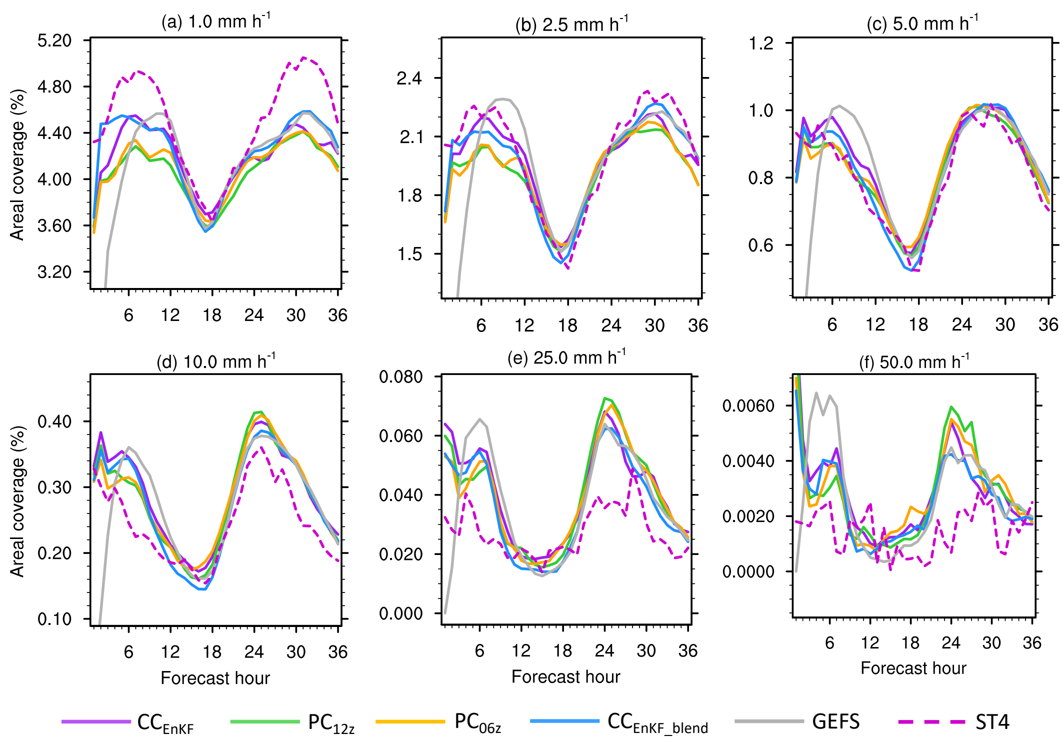

Fig. 4.14. Fractional areal coverage (%) of 1-h accumulated precipitation meeting or exceeding (a) 1.0, (b) 2.5, (c) 5.0, (d) 10.0, (e) 25.0, and (f) 50.0 mm h^{-1} over the CONUS east of 105°W (Fig. 4.1), computed on native grids and aggregated over all 26 3-km forecasts as a function of forecast hour. 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 non-zero 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 23–24 h). The *y* axis scales are different in each panel.

higher thresholds (Figs. 4.14c–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. 4.15).

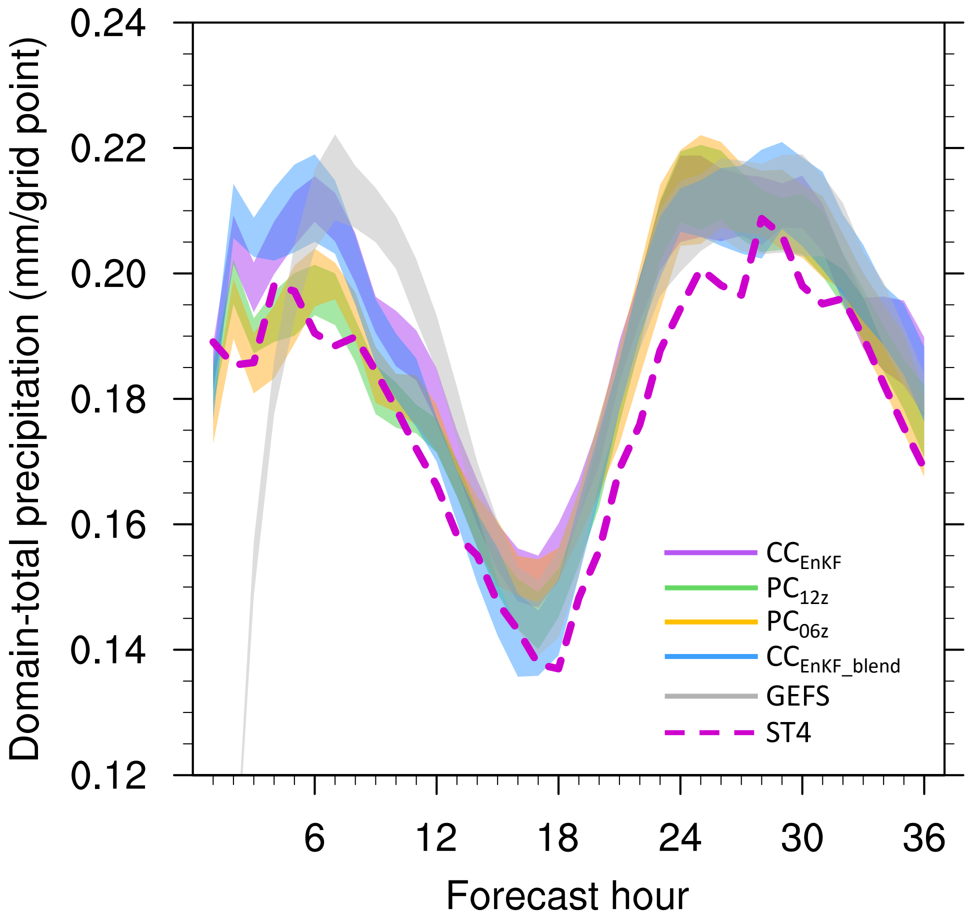

Fig. 4.15. Average 1-h accumulated precipitation (mm) per grid point over all 26 3-km forecasts and the CONUS east of 105°W (Fig. 4.1), computed on native grids as a function of forecast hour. Shadings represent envelopes of the 10 members comprising the various ensembles indicated in the legend, and 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 23–24 h). At the earliest forecast hours, GEFS domain-total precipitation was non-zero but below the *x* axis.

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. 4.5i,k).

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 ≤ 2.5 mm h^{-1} (Figs. 4.14a,b), and overpredicted both areal coverages ≥ 10.0 mm h^{-1} and domain-total precipitation (Figs. 4.14d–f, 4.15). 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.

4.6. 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 re-initialized 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 re-centered 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.

Chapter 5: Summary and conclusions

5.1. Summary and key results

Chapters 2–4 described novel experiments with 80-member limited-area EnKFs over the entire CONUS and adjacent areas. The EnKF DA experiments were performed over 4 weeks with a 1-h cycling period and differed in terms of their cycling strategies (e.g., continuous or partial cycling) and horizontal grid spacings (e.g., 15- or 3-km horizontal grid spacing). EnKF analyses initialized 36-h, 3-km, 10-member CAE forecasts that were evaluated with a focus on precipitation, providing insights about CAE forecast sensitivity to ICs. Additionally, EnKF analyses were leveraged to isolate CAE forecast sensitivity to resolution of both IC perturbations and central initial states about which IC perturbations were centered. A "blending" approach was also used to produce new sets of CAE ICs by combining small scales from EnKF analyses with large scales from 0.5° GEFS ICs using a low-pass filter.

Key results and conclusions are as follows:

- This work, for the first time, demonstrated that a convection-allowing EnKF can be continuously cycled over a large domain without deleterious consequences, as the 3-km continuously cycling EnKF maintained a stable climate and had small biases. This finding should motivate further studies using convection-allowing DA systems in continuously cycling frameworks over large areas like the CONUS.
- CAE forecasts initialized from continuously cycling 3-km EnKF analyses were more skillful and reliable than those initialized from downscaled

GEFS and continuously cycling 15-km EnKF ICs through 12–18 and 6–12 h, respectively. Conversely, after 18 h, GEFS-initialized forecasts were better than forecasts initialized from continuously cycling EnKFs. Blended 3-km ICs led to ~18–36-h forecasts possessing comparable quality as GEFS-initialized forecasts while preserving short-term forecast benefits of employing 3-km ICs produced through continuous cycling. Thus, blending high-resolution EnKF analyses with low-resolution global fields can potentially unify short-term and next-day CAE forecast systems under a common framework, suggesting operational potential of incorporating blending within limited-area analysis–forecast systems over the CONUS.

- 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. Therefore, 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, again highlighting the potential operational use of blending for CONUS-centric modeling systems.

- It is more important for central initial states than for IC perturbations to possess convection-allowing horizontal grid spacing for short-term CAE forecasting applications. These results suggest dual-resolution DA systems should be further explored for CAE initialization. Moreover, these findings potentially enable substantial computational savings and suggest scientists working on initializing future operational CAEs primarily concentrate their energies on producing the best possible high-resolution deterministic analyses that can be used as central initial states for CAEs.

All these conclusions are directly relevant to ongoing efforts at NCEP working toward developing a next-generation CAE slated to become operational in 2023. Our findings also have implications for, and could possibly inspire, modelers at other meteorological centers using various NWP models, dynamic cores, and physics suites. For instance, scientists at meteorological centers outside the United States might view our demonstration of a convection-allowing continuously cycling EnKF over the entire CONUS as an impetus to develop similar systems, even if their countries span large areas. Such efforts can be successful! Additionally, as holistic benefits of continuously cycling are model agnostic, modeling centers around the world should strive to implement and improve limited-area continuously cycling DA systems to accelerate NWP model development. Even if previous efforts with continuous cycling were disappointing—as in the United States—our results suggest that NWP model capabilities and DA techniques have sufficiently progressed to warrant another attempt. Finally, our demonstration that limited-area EnKFs initialized better short-term CAE forecasts than GEFS ICs sets a high bar for future experimental DA systems, which should compare forecasts initialized from their experimental analyses to those initialized from operational analyses to provide context for their developments.

5.2. Future directions

Despite our promising findings and their meaningful implications, there are many avenues for future work and improvements. For instance, although we demonstrated the first convection-allowing continuously cycling EnKF over a domain as large as the CONUS, additional experimentation is needed over longer time periods, different seasons, and varied geographic regions to further understand large-domain convection-allowing continuously cycling EnKF performance. In particular, as our 3-km EnKF was highly constrained by observations, it is unclear whether convection-allowing EnKFs can be successfully continuously cycled over large data-sparse domains like oceanic regions. Moreover, convection-allowing EnKFs can likely be improved by carefully specifying and tuning inflation parameters and observation errors.

Additionally, it would be interesting to further consider assimilation of radar reflectivity observations within large-domain continuously cycling DA systems. While our 3-km EnKF did not assess reflectivity DA within a continuously cycling framework, it would be useful to understand how assimilating radar reflectivity observations in a large-domain continuously cycling DA system impacts the model climate. Intuition suggests the small-scale nature of reflectivity observations may impart little overall impact to the mean state, but experimentation is needed to know for sure.

Furthermore, blending can be incorporated into continuous DA cycles, contrasting our method where blending did not impact the continuously cycling EnKFs and was only used to initialize free forecasts. Such methodology would potentially be attractive in operations and should be tested and compared with traditional partial cycling DA systems. The filter cutoff scale for blending should also be tuned for potential operational systems.

There are also opportunities to study the relative importance of central initial state resolution and IC perturbation resolution within idealized modeling frameworks to better understand reasons for our findings presented in chapter 3. Additionally, it would be interesting to revisit this topic in real-data situations with non-Gaussian DA methodologies, like localized particle filters (e.g., Poterjoy 2016; Poterjoy et al. 2019), to determine whether theoretically better ensemble-based covariance updates translate into greater importance of IC perturbation resolution relative to central initial state resolution. Finally, the ability of mixed-resolution DA systems to produce high-quality central initial states suitable for CAE initialization requires substantial examination.

Ultimately, the community will be working toward global convection-allowing NWP modeling systems for operational weather forecasting purposes. At least some of our results have implications for these future global systems. For instance, our conclusion that it is more important for central initial states to possess convection-allowing horizontal grid spacing than IC perturbations suggests the possibility of only needing to continuously cycle one instance of a global convection-allowing NWP model in a DA system aimed at CAE initialization, possibly saving an immense amount of resources. These resources could then potentially be reinvested into increasing the size of relatively coarse-resolution ensembles used to provide BECs. Of course, this possibility will need to be confirmed by experiments focusing on BEC resolution requirements for global convective-scale DA systems.

Furthermore, the role of blending is unclear in future global convection-allowing modeling systems. If the presence of lateral boundaries indeed inherently limits the ability of limited-area DA systems to accurately depict large scales, there is hope that global convection-allowing analyses—free of LBC influences—may appropriately analyze large scales. On the other hand, if either fine resolution itself or limitations of high-resolution DA (e.g., necessity of small localization distances owing to computing constrains) is responsible for suboptimal large-scale limited-area analyses, then blending coarse- and fine-scale analyses may be necessary even within global modeling systems to produce optimal forecasts beyond ~18 h. These types of experiments will need to be performed once computing allows, and there is much to learn about convection-allowing global NWP and DA in upcoming years and decades.

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-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. But, 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. 4.6a–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} initialized 36-h, 10-member ensemble forecasts over the nested domain (Fig. 4.1), 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

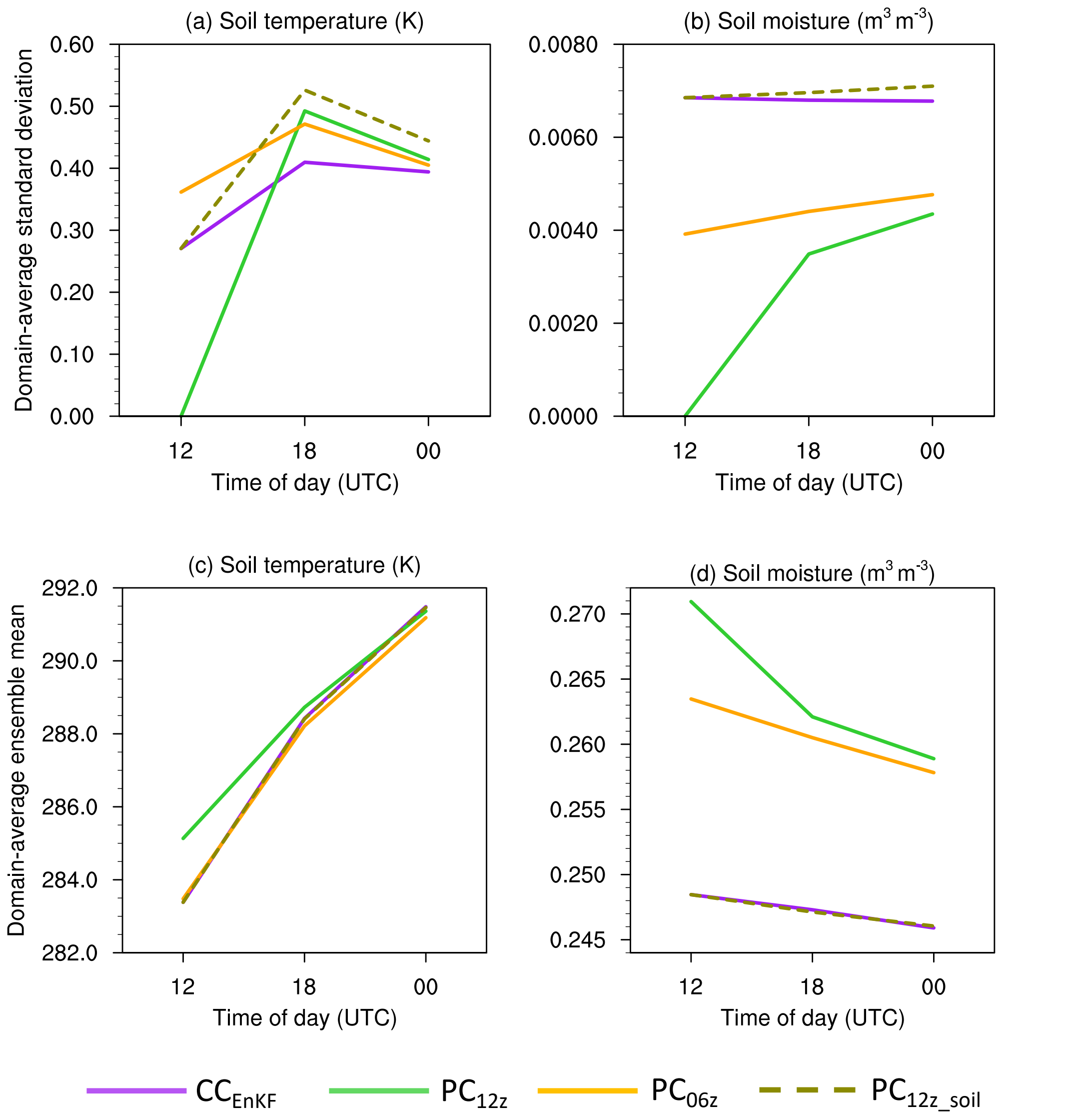

Fig. A1. Standard deviation of top-layer (a) soil temperature (K) and (b) soil moisture (m^3m^{-3}) averaged over land points within the portion of the 15-km domain colocated with the 3-km domain (Fig. 4.1) and all posterior ensembles between 1200 UTC 24 April and 0000 UTC 20 May 2017 (inclusive) as a function of time of day. (c),(d) As in (a) and (b), respectively, but for domain-average ensemble means.

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.

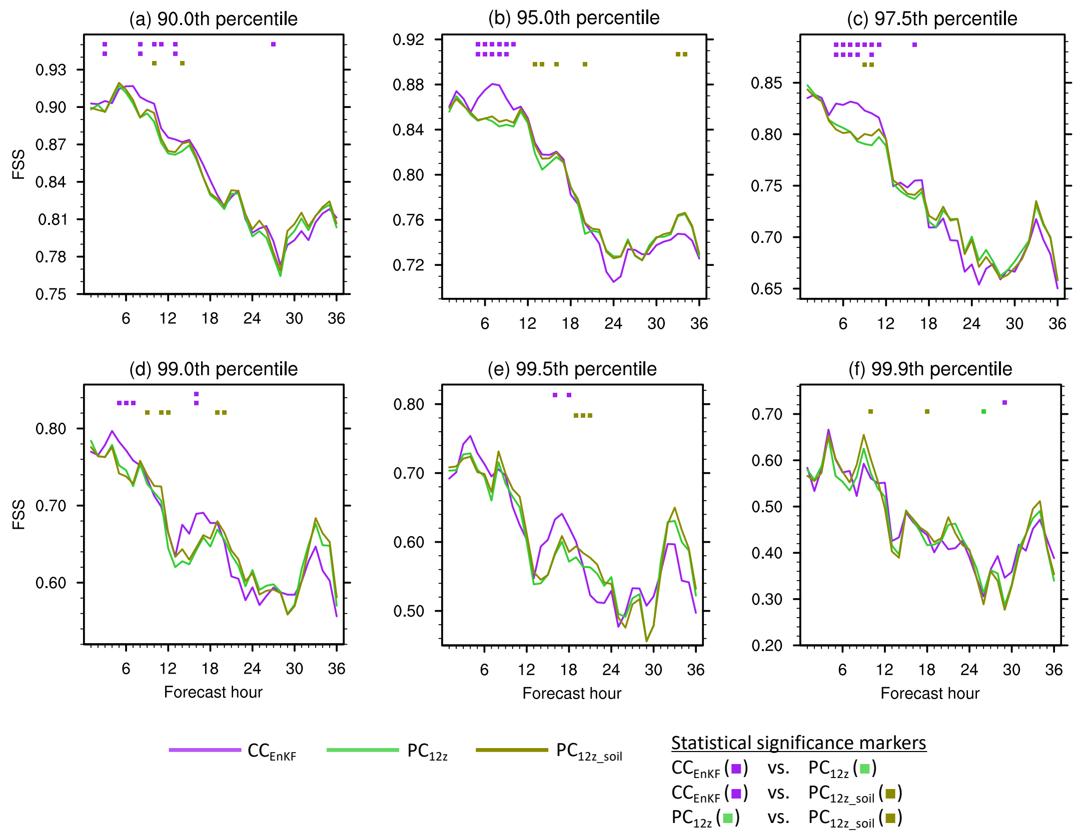

Fig. A2. As in Fig. 4.10, but for FSSs aggregated over the 13 3-km, 10-member ensemble forecasts initialized at 0000 UTC between 25 April and 7 May 2020 (inclusive), focusing on experiments designed to assess forecast sensitivity to land surface state initialization in partial cycling EnKFs.

Bibliography

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

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

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

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

Anderson, J. L., 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.

Anderson, J. L., 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.

Anderson, J. L., 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.

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

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

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

Barker, D. M., and Coauthors, 2012: The Weather Research and Forecasting Model's Community Variational/Ensemble Data Assimilation System: WRFDA. *Bull. Amer. Meteor. Soc.*, **93**, 831–843, https://doi.org/10.1175/BAMS-D-11-00167.1.

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

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

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

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

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

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

Cafaro, C., 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.

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

Carley, J. R., and Coauthors, 2021: Status of NOAA's next generation convection-allowing ensemble: The Rapid Refresh Forecast System. *Special Symposium on Global and Mesoscale Models*, Virtual, Amer. Meteor. Soc., 12.8.

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

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

Chen, F., and J. Dudhia, 2001: Coupling an advanced land-surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model 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.

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

Clark, A. J., W. A. Gallus Jr., and M. L. Weisman, 2010: Neighborhood-based verification of precipitation forecasts from convection-allowing NCAR WRF Model simulations and the operational NAM. *Wea. Forecasting*, **25**, 1495–1509, https://doi.org/10.1175/2010WAF2222404.1.

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

Clark, A. J., and Coauthors, 2012: An overview of the 2010 Hazardous Weather Testbed Experimental Forecast Program Spring Experiment. *Bull. Amer. Meteor. Soc.*, **93**, 55–74, https://doi.org/10.1175/BAMS-D-11-00040.1.

Clark, A. J., 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.

Clark, P., N. Roberts, H. Lean, S. P. Ballard, and C. Charlton-Perez, 2016: Convection-permitting models: A step-change in rainfall forecasting. *Meteor. Appl.*, **23**, 165–181, https://doi.org/10.1002/met.1538.

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

Computational and Information Systems Laboratory, 2017: Cheyenne: HPE/SGI ICE XA System (NCAR Community Computing). Boulder, CO: National Center for Atmospheric Research. 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.

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

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

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

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

Developmental Testbed Center, 2016: Gridpoint Statistical Interpolation Advanced User's Guide Version 3.5.0.0. [Available at https://dtcenter.org/com-GSI/users/docs/users_guide/AdvancedGSIUserGuide_v3.5.0.0.pdf.], 119 pp.

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.

Done, J., C. A. Davis, and M. L. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the Weather Research and Forecasting (WRF) Model. *Atmos. Sci. Lett.*, **5**, 110–117, https://doi.org/10.1002/asl.72.

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

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.

Dowell, D. C., and Coauthors, 2021: The High-Resolution Rapid Refresh (HRRR): An hourly updating convection-allowing forecast model. Part 1: Motivation and system description. Submitted to *Wea. Forecasting*.

Du, J., G. DiMego, B. Zhou, D. Jovic, B. Ferrier, and B. Yang, 2015: Short Range Ensemble Forecast (SREF) system at NCEP: Recent development and future transition. *23rd Conf. on Numerical Weather Prediction/27th Conf. on Weather Analysis and Forecasting*, Chicago, IL, Amer. Meteor. Soc., 2A.5, https://ams.confex.com/ams/27WAF23NWP/webprogram/Paper273421.html.

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

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

Durran, D. R., 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.

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

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

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

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

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162, https://doi.org/10.1029/94JC00572.

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

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

Feng, J., 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.

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

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

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

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

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. [Available online at 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.

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

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

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

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

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

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

Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter–3D variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905–2919, https://doi.org/10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2.

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

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

Hohenegger, C., A. Walser, W. Langhans, and C. Schär, 2008: Cloud-resolving ensemble simulations of the August 2005 Alpine flood. *Quart. J. Roy. Meteor. Soc.*, **134**, 889–904, https://doi.org/10.1002/qj.252.

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

Houtekamer, P. L., 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.

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

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

Hsiao, L.-F., 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.

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

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

James, E. P., S. G. Benjamin, and B. D. Jamison, 2020: Commercial-aircraft-based observations for NWP: Global coverage, data impacts, and COVID-19. *Journal of Applied Meteorology and Climatology*, **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.

Janjić, Z. I., 2002: Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP Meso model. NCEP Office Note 437, 61 pp. [Available online at http://www.emc.ncep.noaa.gov/officenotes/newernotes/on437.pdf.]

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

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

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

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.

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

Johnson, A., X. Wang, 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.

Johnson, A., X. Wang, and S. Degelia, 2017: Design and implementation of a GSI-based convection allowing ensemble-based data assimilation and forecast system for the PECAN field experiment. Part II: Overview and evaluation of a real-time system. *Wea. Forecasting*, **32**, 1227–1251, https://doi.org/10.1175/WAF-D-16-0201.1.

Johnson, A., X. Wang, 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.

Jones, T. A., 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.

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

Jones, T. A., P. Skinner, N. Yussouf, K. Knopfmeier, A. Reinhart, X. Wang, K. Bedka, W. Smith, and R. Palikonda, 2020: Assimilation of GOES-16 radiances and retrievals into the Warn-on-Forecast System. *Mon. Wea. Rev.*, **148**, 1829–1859, https://doi.org/10.1175/MWR-D-19-0379.1.

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

Kain, J. S., S. J. Weiss, J. J. Levit, M. E. Baldwin, and D. R. Bright, 2006: Examination of convection-allowing configurations of the WRF Model for the prediction of severe convective weather: The SPC/NSSL Spring Program 2004. *Wea. Forecasting*, **21**, 167–181, https://doi.org/10.1175/WAF906.1.

Kain, J. S., and Coauthors, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting*, **23**, 931–952, https://doi.org/10.1175/WAF2007106.1.

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

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

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

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

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

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

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. Preprints, *24th Conf. on Severe Local Storms*, Savannah, GA, Amer. Meteor. Soc., 12.3. [Available online at https://ams.confex.com/ams/pdfpapers/141827.pdf.]

Kong, F., and Coauthors, 2009: A real-time storm-scale ensemble forecast system: 2009 Spring Experiment. Preprints, *23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction*, Omaha, NE, Amer. Meteor. Soc., 16A.3. [Available online at https://ams.confex.com/ams/pdfpapers/154118.pdf.]

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

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

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

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

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

Lin, H., S. S. Weygandt, S. G. Benjamin, and M. Hu, 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.

Lin, H., S. S. Weygandt, 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. Preprints, *19th Conf. on Hydrology*, San Diego, CA, Amer. Meteor. Soc., 1.2. [Available online at http://ams.confex.com/ams/pdfpapers/83847.pdf.]

Liu, H., J. L. Anderson, Y.-H. Kuo, and K. Raeder, 2007: Importance of forecast error multivariate correlations in idealized assimilation of GPS radio occultation data with the ensemble adjustment filter. *Mon. Wea. Rev.*, **135**, 173–185, https://doi.org/10.1175/MWR3270.1.

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

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.

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

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

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

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

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

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

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

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

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

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

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

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the long-wave. *J. Geophys. Res.*, **102**, 16 663–16 682, https://doi.org/10.1029/97JD00237.

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

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

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

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

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.

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

Poterjoy, J., 2016: A localized particle filter for high-dimensional nonlinear systems. *Mon. Wea. Rev.*, **144**, 59–76, https://doi.org/10.1175/MWR-D-15-0163.1.

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

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

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.

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

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

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

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.

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

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

Raynaud, L., and F. Bouttier, 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. Preprints, *23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction*, Omaha, NE, Amer. Meteor. Soc., 2A.4. [Available online at http://ams.confex.com/ams/pdfpapers/154114.pdf].

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

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

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

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

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

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

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

Schwartz, C. S., 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.

Schwartz, C. S., 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.

Schwartz, C. S., 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.

Schwartz, C. S., and Coauthors, 2009: Next-day convection-allowing WRF Model guidance: A second look at 2-km versus 4-km grid spacing. *Mon. Wea. Rev.*, **137**, 3351– 3372, https://doi.org/10.1175/2009MWR2924.1.

Schwartz, C. S., 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.

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

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

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

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

Schwartz, C. S., G. S. Romine, R. A. Sobash, K. R. Fossell, and M. L. Weisman, 2019: NCAR's real-time convection-allowing ensemble project. *Bull. Amer. Meteor. Soc.*, **100**, 321–343, https://doi.org/10.1175/BAMS-D-17-0297.1.

Schwartz, C. S., 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.

Schwartz, C. S., 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. © American Meteorological Society. Used with permission.

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

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

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.

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

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

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

Stensrud, D. J., 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.

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

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

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

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

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, https://doi.org/10.1175/2008MWR2387.1.

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

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490, https://doi.org/10.1175/1520-0493(2003)131<1485:ESRF>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.

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

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

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

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

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

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

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

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

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.

Weisman, M. L., C. Davis, W. Wang, K. W. Manning, and J. B. Klemp, 2008: Experiences with 0–36-h explicit convective forecasts with the WRF-ARW Model. *Wea. Forecasting*, **23**, 407–437, https://doi.org/10.1175/2007WAF2007005.1.

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.

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

Whitaker, J. S., T. M. Hamill, 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.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Elsevier, 676 pp.

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

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

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. Preprints, *22nd Conf. on Weather Analysis and Forecasting/18th Conf. on Numerical Weather Prediction*, Salt Lake City, UT, Amer. Meteor. Soc., 3B, http://ams.confex.com/ams/pdfpapers/124587.pdf.

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

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

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

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

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

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

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

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

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

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 stable 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.