

# Interactions between Physics Diversity and Multiscale Initial Condition Perturbations for Storm-Scale Ensemble Forecasting

AARON JOHNSON AND XUGUANG WANG

*School of Meteorology, University of Oklahoma, Norman, Oklahoma*

(Manuscript received 9 April 2020, in final form 20 June 2020)

## ABSTRACT

This study investigates impacts on convection-permitting ensemble forecast performance of different methods of generating the ensemble IC perturbations in the context of simultaneous physics diversity among the ensemble members. A total of 10 convectively active cases are selected for a systematic comparison of different methods of perturbing IC perturbations in 10-member convection-permitting ensembles, both with and without physics diversity. These IC perturbation methods include simple downscaling of coarse perturbations from a global model (LARGE), perturbations generated with ensemble data assimilation directly on the multiscale domain (MULTI), and perturbations generated using each method with small scales filtered out as a control. MULTI was found to be significantly more skillful than LARGE at early lead times in all ensemble physics configurations, with the advantage of MULTI gradually decreasing with increasing forecast lead time. The advantage of MULTI, relative to LARGE, was reduced but not eliminated by the presence of physics diversity because of the extra ensemble spread that the physics diversity provided. The advantage of MULTI, relative to LARGE, was also reduced by filtering the IC perturbations to a commonly resolved spatial scale in both ensembles, which highlights the importance of flow-dependent small-scale ( $< \sim 10$  m) IC perturbations in the ensemble design. The importance of the physics diversity, relative to the IC perturbation method, depended on the spatial scale of interest, forecast lead time, and the meteorological characteristics of the forecast case. Such meteorological characteristics include the strength of synoptic-scale forcing, the role of cold pool interactions, and the occurrence of convective initiation or dissipation.

## 1. Introduction

Convection-allowing model (CAM; i.e., grid spacing  $< \sim 4$  km) forecasts of convective precipitation are sensitive to errors on a broad range of spatial scales. For example, as theorized by Lorenz (1969) and more recently demonstrated in the context of CAM forecasts (e.g., Zhang et al. 2007; Hohenegger and Schar 2007), very small initial condition (IC) errors can rapidly grow in both amplitude and spatial scale. The upscale error growth is a source of forecast uncertainty not only for the evolution of convective systems (e.g., Flora et al. 2018), but also for the mesoscale and synoptic-scale environments in which they occur (e.g., Perkey and Maddox 1985; Zhang et al. 2007). The uncertainty of the ambient environment then feeds back to the convective-scale uncertainty on time scales as short as a few hours (e.g., Hohenegger and Schar 2007; Cintineo and Stensrud 2013; Kerr et al. 2019). The IC uncertainty on both

convective scales and mesoscales contributes to substantial forecast uncertainty even out to lead times of a day or more (Johnson et al. 2014).

The motivation for ensemble forecasting, simply stated, is to account for the predictability limitations resulting from the nonlinear error growth described by Lorenz (1969) through Monte Carlo sampling of a large number of equally plausible forecasts (Ehrendorfer 1997). Ensemble forecasting aims to predict not the exact future state of the atmosphere, but rather the distribution of possible future states of the atmosphere given the best guess of the IC, the model dynamics and physics, and the uncertainty in each. Thus, a key consideration for optimally designing a CAM ensemble is how to choose a finite number of equally plausible model configurations and IC states to adequately sample both the analysis uncertainties that dominate relevant forecast uncertainties, and the uncertainties in the forward integration of the model (e.g., Anderson 1996; Stensrud et al. 2000; Hamill et al. 2000; Romine et al. 2014; Johnson et al. 2014; Johnson and Wang 2016).

---

*Corresponding author:* Dr. Aaron Johnson, ajohns14@ou.edu

DOI: 10.1175/MWR-D-20-0112.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) ([www.ametsoc.org/PUBSReuseLicenses](https://www.ametsoc.org/PUBSReuseLicenses)).

A common method of generating IC perturbations for CAM ensembles has been to interpolate them from a coarser resolution (e.g., global) ensemble (e.g., Hohenegger et al. 2008; Zhang et al. 2010; Xue et al. 2010; Schwartz and Liu 2014). Early studies by Durran and Gingrich (2014) and Durran and Weyn (2016) have shown that the convective-scale predictability at lead times beyond a few hours is more strongly limited by errors on spatial scales of  $\sim 100$  km or larger, rather than the smaller scales. From an ensemble design perspective, this raises the question of whether it is necessary to pay much attention to ensemble IC perturbations on spatial scales smaller than  $\sim 100$  km. As a counterpoint, an early study by Johnson et al. (2014) showed that convective-scale IC perturbations contribute a similar amount of uncertainty to 1-day lead-time CAM forecasts of mesoscale precipitation as contributed by mesoscale IC perturbations. However, both Durran and Gingrich (2014) and Johnson et al. (2014) used random homogeneous IC perturbations rather than flow-dependent IC perturbations that sample the fast-growing modes of IC uncertainty, which are expected to contribute the most to forecast uncertainty.

Johnson and Wang (2016) used observation system simulation experiments (OSSEs) with a “perfect-model” assumption to go a step further and compare CAM ensemble forecasts with IC perturbations generated by a CAM ensemble-based data assimilation (DA) system compared to IC perturbations downscaled from a coarser resolution convection-parameterizing (i.e., 12-km grid spacing) ensemble. It was shown that the mesoscale precipitation forecasts out to at least the 9-h lead time were improved in the experiment with flow-dependent multiscale IC perturbations from the CAM ensemble compared to the flow-dependent coarser resolution IC perturbations from the convection-parameterizing ensemble. There were two factors contributing to the difference. First, the multiscale IC perturbations contained convective-scale structure while the coarser resolution IC perturbations did not. Second, the multiscale IC perturbations were more consistent with the analysis errors than the coarser resolution IC perturbations, even on the commonly resolved scales.

Past studies aimed at understanding optimal methods of sampling the analysis uncertainty in CAM ensembles through the IC perturbations have generally omitted interactions with the model and physics diversity within the ensemble design (e.g., Wang et al. 2014; Johnson and Wang 2016; Keresturi et al. 2019). However, it is well established that including a representation of model and physics uncertainty during ensemble forecast integration is necessary for achieving optimal forecast performance (e.g., Romine et al. 2014; Johnson et al. 2017; Gasperoni et al. 2020). It is unclear if IC perturbation

methods optimized in the context of a fixed-model, fixed-physics ensemble would still be optimal in the presence of other sources of forecast diversity, such as in a multiphysics ensemble. In contrast to the OSSE framework with perfect model assumption of Johnson and Wang (2016), the present study uses real data cases and a multiphysics forecast ensemble. This allows for a more realistic assessment of the impacts on ensemble forecast performance of the different IC perturbation methods in the operationally applicable situation where model and physics errors are also contributing to the forecast uncertainty, and an existing global ensemble system is used for the downscaled IC perturbations. This study uses the operational global ensemble from the National Centers for Environmental Prediction (NCEP) to provide downscaled IC perturbations that more closely replicate what would be available to initialize an operational CAM ensemble in the absence of a CAM ensemble-based DA system. Specifically, we aim to compare and understand ensemble forecast performance when IC perturbations are downscaled from coarser resolution perturbations from the Global Ensemble Forecast System (GEFS) from NCEP to the performance when multiscale IC perturbations are generated using the GSI-based EnVar DA system (Johnson et al. 2015; Wang and Wang 2017) directly on the convection-permitting grid in the presence of model error and ensemble physics diversity. Any advantages of the multiscale IC perturbations for forecast performance would represent an additional benefit to justify the cost of operationally running ensemble-based DA directly on the convection-permitting grid in comparison to other DA approaches.

The organization of this paper is as follows. Section 2 describes the experiment design, including an overview of 10 case studies considered in this study, data assimilation and forecast system configuration, and details of the different IC perturbation methods. Results are then presented in section 3 and a summary and discussion are contained in section 4.

## 2. Experiment design

### *a. Overview of cases*

A total of 10 cases are selected for these experiments (Table 1). The cases were selected because of the presence of widespread, potentially hazardous, deep convection and were also used in Gasperoni et al. (2020). The cases have been subjectively categorized as strongly forced or weakly forced based on the apparent strength of the synoptic-scale forcing, as implied by the strength of the jet stream and the presence or absence of synoptic features such as surface fronts or jet streaks near the

TABLE 1. Cases included in this study, including DA start time, forecast initialization time, and subjective classification of strength of synoptic-scale forcing.

DA start time	Forecast initialization time	Synoptic-scale forcing
1700 UTC 16 May 2015	2300 UTC 16 May 2015	Strong
0700 UTC 25 May 2015	1300 UTC 25 May 2015	Strong
2200 UTC 25 Jun 2015	0400 UTC 26 Jun 2015	Weak
1300 UTC 14 Jul 2015	1900 UTC 14 Jul 2015	Strong
1900 UTC 10 Sep 2015	0100 UTC 11 Sep 2015	Strong
1700 UTC 22 May 2016	2300 UTC 22 May 2016	Strong
1400 UTC 17 Jun 2016	2000 UTC 17 Jun 2016	Weak
1900 UTC 5 Jul 2016	0100 UTC 6 Jul 2016	Weak
1800 UTC 6 Jul 2016	0000 UTC 7 Jul 2016	Weak
2200 UTC 9 Jul 2016	0400 UTC 10 Jul 2016	Weak

convective systems of interest. The mean 250-hPa heights and winds at the analysis time for the strongly forced and weakly forced cases are shown in Figs. 1a and 2a, respectively. Also shown in Figs. 1b–f and 2b–f are the

observed reflectivity at the 6-h forecast time in the strongly forced and weakly forced cases, respectively. While the strongly forced cases are generally characterized by diffluent southwesterly flow of  $>50$  kt ( $\sim 26$  m s $^{-1}$ )

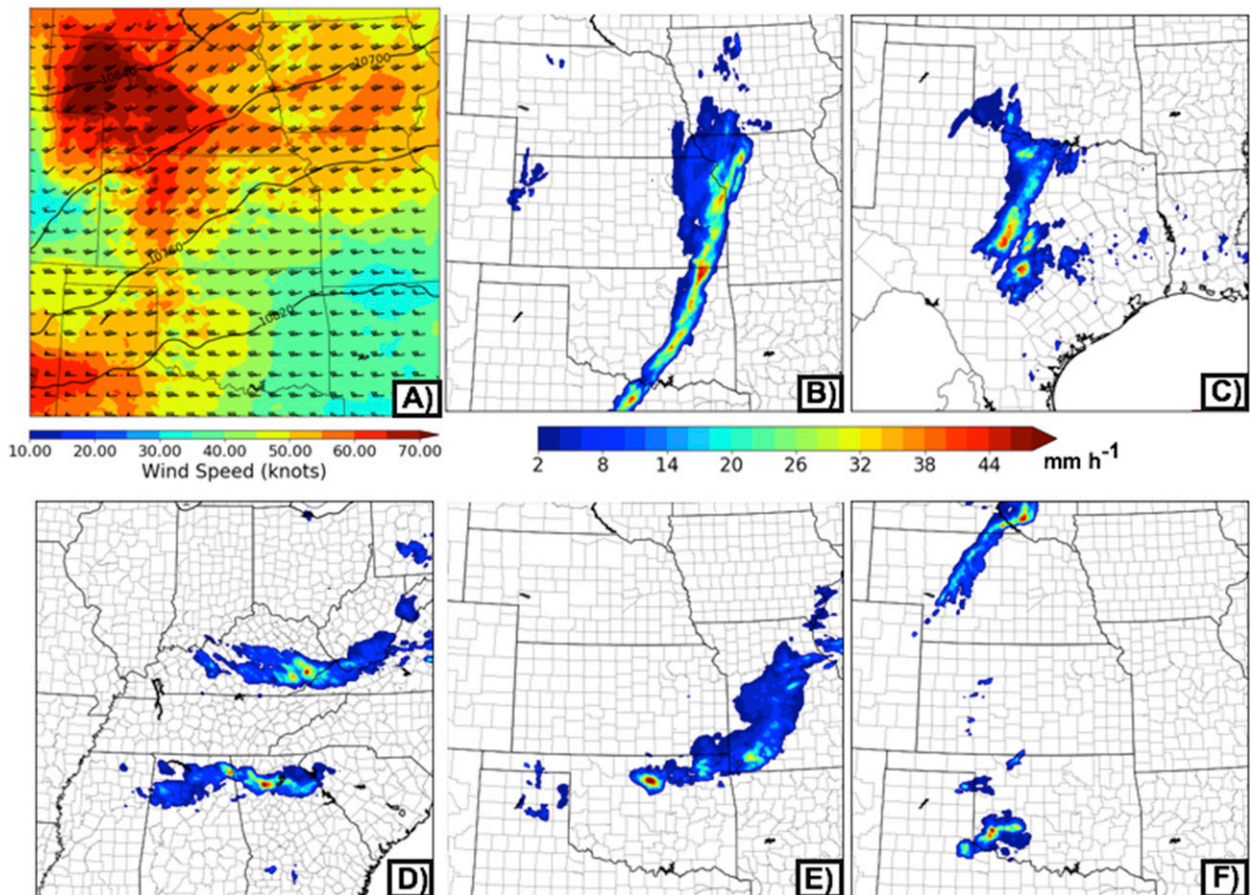


FIG. 1. Summary of cases with “strong” synoptic-scale forcing, including (a) the average 250-hPa heights, wind speeds, and wind barbs, and observed hourly accumulated precipitation at the 6-h forecast time for forecasts initialized at (b) 2300 UTC 16 May 2015, (c) 1300 UTC 25 May 2015, (d) 1900 UTC 14 Jul 2015, (e) 0100 UTC 11 Sep 2015, and (f) 2300 UTC 22 May 2016. The state boundaries in (a) are for size reference only, since the averaged domains have different center points, as can be inferred from (b)–(f). Heights (m; 60-m interval) are contoured and labeled in (a).

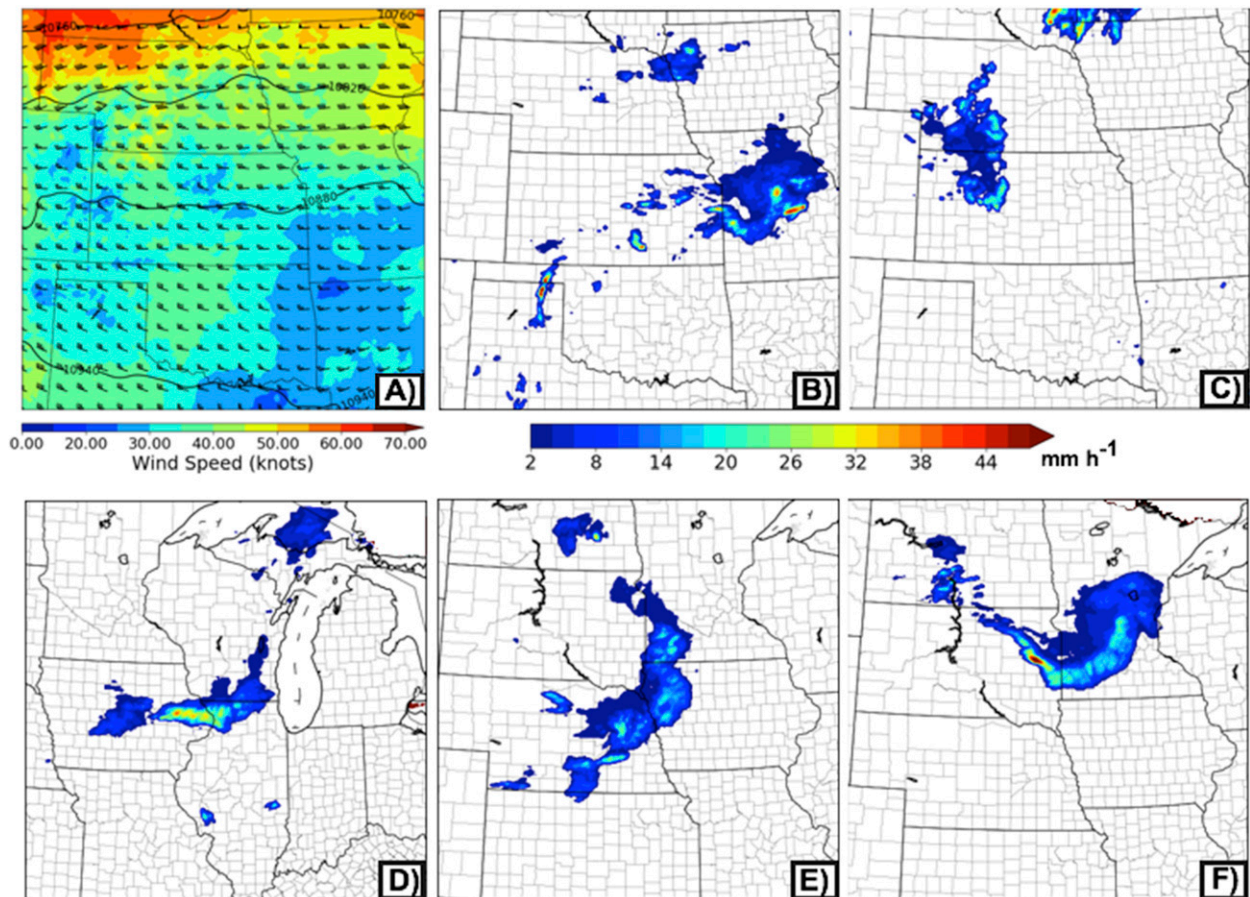


FIG. 2. Summary of cases with “weak” synoptic-scale forcing, including (a) the average 250-hPa heights, wind speeds, and wind barbs, and observed hourly accumulated precipitation at the 6-h forecast time for forecasts initialized at (b) 0400 UTC 26 Jun 2015, (c) 2000 UTC 17 Jun 2016, (d) 0100 UTC 6 Jul 2016, (e) 0000 UTC 7 Jul 2016, and (f) 0400 UTC 10 Jul 2016. The state boundaries in (a) are for size reference only, since the averaged domains have different center points, as can be inferred from (b)–(f). Heights (m; 60-m interval) are contoured and labeled in (a).

at 250 hPa (Fig. 1a), the weakly forced cases are characterized by generally westerly flow of  $<50$  kt over most of the forecast domain (Fig. 2a). The cases are chosen such that a variety of convective systems occur during the 18-h forecast period, although much of the convection dissipates or moves out of the forecast domain by the 18-h forecast lead time.

#### b. Ensemble analysis and forecast system configuration

This study leverages the ensemble analyses produced by Gasperoni et al. (2020). The data assimilation (DA) system adopted is the two-way coupled GSI-based ensemble-variational (EnVar) hybrid system based on the Advanced Research version of the Weather Research and Forecasting (WRF) Model (ARW; version 3.9; Skamarock et al. 2008) that has been used for U.S. operational global numerical weather prediction (e.g., Wang et al. 2013; Wang

and Lei 2014). This system has been extended for mesoscale and convective-scale DA, including the capability to directly assimilate ground-based radar observations (Johnson et al. 2015; Wang and Wang 2017). Conventional surface and upper-air observations from the North American Mesoscale model DA system were assimilated every hour for a 6-h period, followed by a 1-h period of assimilating NEXRAD reflectivity observations every 20 min, following the system configuration further detailed in Gasperoni et al. (2020) and Johnson et al. (2020). In short, the system consists of a 40-member Ensemble Kalman filter (EnKF), coupled to a 3D EnVar that provides a control analysis around which each EnKF ensemble member is recentered after each analysis cycle in order to prevent divergence of the two ensembles. The DA covers the CONUS [see Fig. 1 of Gasperoni et al. (2020)]. The horizontal grid spacing is 3 km, and there are 50 vertical levels using the WRF

TABLE 2. Physics configuration of the forecast ensemble labeled as “phys1.” Microphysics abbreviations are Thom. (Thompson et al. 2008), Morr. (Morrison et al. 2009), MY (Milbrandt and Yau 2005), NSSL2M (National Severe Storms Laboratory double moment; Mansell et al. 2010), NSSL (NSSL single moment; Gilmore et al. 2004; Mansell et al. 2010), and P3 (Predicted Particle Properties; Morrison and Milbrandt 2015). Planetary boundary layer (PBL) abbreviations are MYJ (Mellor–Yamada–Janjic; Janjić 1994, 2001), YSU (Yonsei University; Noh et al. 2003), and MYNN (Mellor–Yamada–Nakanishi–Niino; Nakanishi and Niino 2009). Other physics schemes are the same as the data assimilation configuration described in the text. The source of lateral boundary conditions for each forecast ensemble member are given in the far-right column.

Member	Microphysics scheme	PBL scheme	Lateral boundary condition
001	Thom.	MYJ	GEFS 001
002	Thom.	MYNN	GEFS 002
003	NSSL2M	YSU	GEFS 003
004	NSSL2M	MYNN	GEFS 004
005	Morr.	MYJ	GEFS 005
006	P3	YSU	GEFS 006
007	NSSL	MYJ	GEFS 007
008	Morr.	YSU	GEFS 008
009	P3	MYNN	GEFS 009
010	Thom.	MYNN	GEFS 010

stretched terrain-following coordinate. The relaxation to prior spread (RTPS) method of posterior covariance inflation (Whitaker and Hamill 2012) was employed with a coefficient of 0.95 to maintain ensemble spread during the DA cycles, consistent with the coefficient found to be optimal in several past studies (e.g., Whitaker and Hamill 2012; Harnisch and Keil 2015; Maldonado and Ruiz 2020). The large value of RTPS coefficient that is needed to maintain ensemble consistency likely compensates for undersampled model and physics uncertainty in the DA ensemble design. To reduce the impacts of potentially spurious ensemble covariances, horizontal covariance localization was applied with a cutoff length scale of 300 km for all observations except for the more densely spaced radar reflectivity which used a zero-correlation cutoff length of 15 km. The vertical covariance localization was applied with a cutoff scale of 1.1 and 0.55 for radar and nonradar observations, respectively (Gasparoni et al. 2020). The vertical length scale is a scale height calculated from the difference in the natural logarithm of pressure (hPa) between the observation height and the height of the model grid point being updated.

A fixed physics configuration is used during DA, similar to the operational High-Resolution Rapid Refresh Ensemble (HRRRE; Benjamin et al. 2016; Jankov et al. 2019). These physics schemes include the Mellor–Yamada–Nakanishi–Niino (Nakanishi and Niino 2009) boundary layer parameterization, Thompson

TABLE 3. Physics configuration of the forecast ensemble labeled as “phys2.” See the Table 2 caption for explanations of the abbreviations. An additional microphysics abbreviation here is WDM6 (WRF double-moment 6-class; Lim and Hong 2010). Additional PBL abbreviations here are QNSE (quasi-normal-scale elimination; Sukoriansky et al. 2005) and ACM2 (Asymmetric Convective Model, version 2; Pleim 2007). Other physics schemes are the same as the data assimilation configuration described in the text. Bold-font members 001 and 002 correspond to the “fixed1” and “fixed2” fixed-physics ensemble configurations, respectively.

Member	Microphysics scheme	PBL scheme	Lateral boundary condition
001	<b>Thom.</b>	<b>QNSE</b>	GEFS 001
002	<b>Morr.</b>	<b>MYJ</b>	GEFS 002
003	MY	YSU	GEFS 003
004	WDM6	ACM2	GEFS 004
005	P3	MYNN	GEFS 005
006	Morr.	MYJ	GEFS 006
007	Thom.	QNSE	GEFS 007
008	MY	ACM2	GEFS 008
009	WDM6	MYNN	GEFS 009
010	P3	YSU	GEFS 010

et al. (2008) microphysics parameterization, Rapid Update Cycle (Smirnova et al. 2016) land surface model, and Rapid Radiative Transfer Model (Mlawer et al. 1997) radiation parameterization. The lateral boundary conditions during DA for the first 20 members are driven by GEFS forecasts from NCEP, initialized at the most recent GEFS cycle time before the first DA cycle on each case. For example, for the ensemble forecast initialized at 2300 UTC 16 May 2015 with DA beginning at 1700 UTC 16 May 2015, the GEFS forecasts initialized at 1200 UTC 16 May 2015 are used as boundary conditions for the first 20 background forecasts in the DA

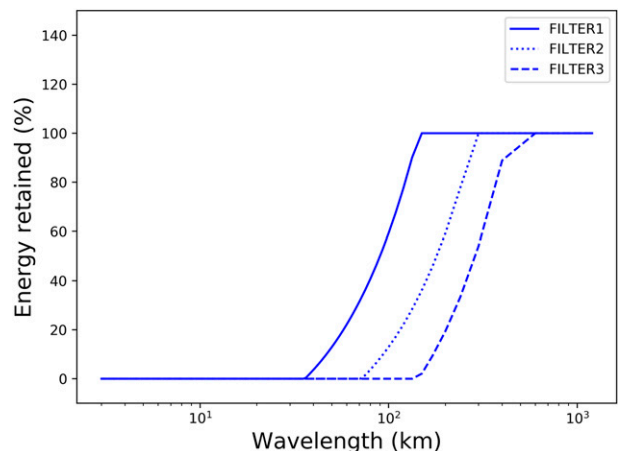


FIG. 3. Filter response function for low-pass filters applied when generating the FILTER1 (solid), FILTER2 (dotted), and FILTER3 (dashed) IC perturbations.

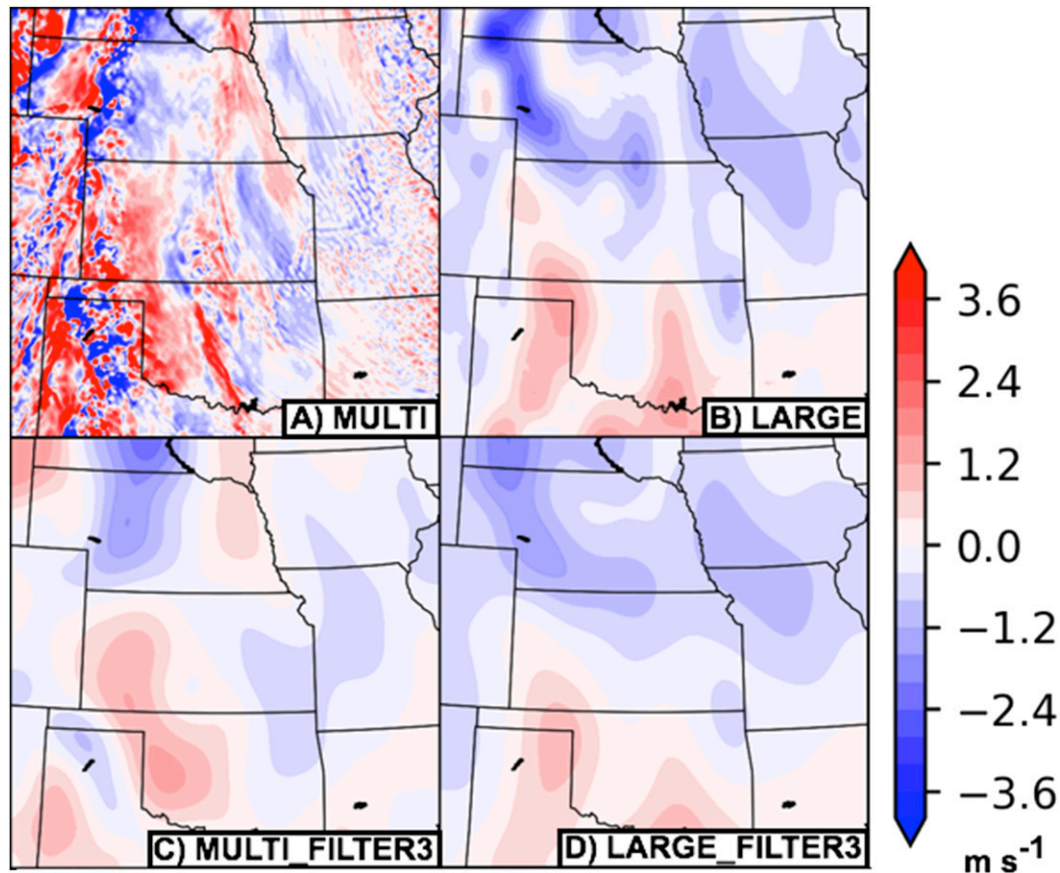


FIG. 4. Example IC perturbations to the  $u$ -wind component at model level 5 ( $\sim 850$  hPa) for the 22 May 2016 case from the same ensemble member, including (a) MULTI, (b) LARGE, (c) MULTI\_FILTER3, and (d) LARGE\_FILTER3.

ensemble. Similarly, the most recent Short-Range Ensemble Forecast (SREF) cycle from NCEP is used to drive members 21–40 in the DA ensemble background forecasts. Two different ensemble systems were combined to drive the lateral boundaries of the 40-member DA ensemble because the ensemble size for GEFS and SREF individually is 21 and 22 members, respectively.

In this study, a smaller domain of  $400 \times 400$  grid points (i.e.,  $1200 \text{ km} \times 1200 \text{ km}$ ; e.g., Figs. 1a and 2a) is used to investigate the interaction between IC and physics perturbation methods. We denote these domains as “forecast domain.” These forecast domains are centered differently for each case to focus on the convective systems of interest (Figs. 1b–f and 2b–f). The forecast domain is smaller than CONUS in order to make the large number of experiments that were conducted computationally tractable. Following the OU MAP real time forecast ensembles from the 2017–19 Hazardous Weather Testbed (HWT) Spring Forecasting Experiments (SFEs), a 10-member forecast ensemble is adopted, using the EnVar control analysis and 9 perturbed

members as defined in section 2c for the different experiments. The lateral boundary conditions for the forecast domains are specified using forecasts from the operational GEFS ensemble members. The LBC tendency is updated using the analysis over the forecast domain before launching the free forecast (Barker et al. 2012).

The forecast ensembles are configured with two different multiphysics configurations (Tables 2 and 3 correspond to “phys1” and “phys2,” respectively), as well as two fixed-physics configurations using each of the members in bold font in Table 3 (members 001 and 002 correspond to “fixed1” and “fixed2,” respectively). These configurations are chosen to emphasize the impact of physics diversity on the differences among IC perturbation experiments by comparing them to a fixed-physics ensemble while also considering the robustness of the results to other plausible configurations of a multi- or fixed-physics ensemble. The forecast ensembles are configured to replicate a current operationally practical ensemble size (10 members), one of which is an EnVar deterministic analysis and the others are perturbed

analyses from the coupled EnKF part of the hybrid EnVar system.

*c. Definition of IC perturbation methods*

All forecast experiments in this study are initialized with the same 10-member ensemble-mean multiscale IC, obtained from the mean of the EnVar control analysis and the EnKF analyses from the first 9 DA ensemble members as used in the real time OU MAP forecast ensembles in the 2017–19 HWT SFEs. Centering the perturbations around the ensemble mean, rather than the control analysis, ensures both a zero-mean of the perturbations and that only the perturbations of each member from the shared ensemble mean are different among the experiments. The experiment that directly uses the analysis perturbations from the DA system described in section 2b is referred to as “MULTI.” An experiment referred to as “LARGE” uses IC perturbations taken from the perturbations of the GEFS members from their own ensemble mean before being added to the MULTI ensemble mean. The GEFS perturbations are obtained from short-term (0–6 h) forecasts from the operational global ensemble at NCEP.

The GEFS spectral model forecasts for these 2015–16 case studies were run with a T574 (~34-km) resolution (Zhou et al. 2017) and were obtained from the National Centers for Environmental Information archive on a 0.5° (~50-km) grid. Although moist convection on scales of ~1–10 km is an important contributor to forecast uncertainty (e.g., Hohenegger and Schar 2007), such processes are not resolved and can only be represented in the LARGE perturbations indirectly through the cumulus parameterization scheme and the stochastic physics schemes applied during the GEFS DA cycles (Zhou et al. 2017). Therefore, we would not expect the LARGE perturbations to accurately reflect the forecast uncertainty in convectively active scenarios (Johnson and Wang 2016).

Finally, experiments referred to as “MULTI\_FILTER” and “LARGE\_FILTER” are similar to MULTI and LARGE, except that a low-pass filter is applied to each perturbation before adding it to the full-resolution ensemble mean, which is the same for every experiment. MULTI\_FILTER and LARGE\_FILTER control for the difference in resolvable spatial scales between MULTI and LARGE. We choose a simple filter that linearly decreases the response function from 1.0 to 0.0 over a range of wavelengths, rather than a spectral truncation, to minimize the introduction of any unphysical artifacts of the filtering process itself. There are three experiments each of MULTI\_FILTER (MULTI\_FILTER1, MULTI\_FILTER2 and MULTI\_FILTER3) and LARGE\_FILTER (LARGE\_FILTER1, LARGE\_FILTER2 and LARGE\_FILTER3) corresponding to low-pass filters with the response

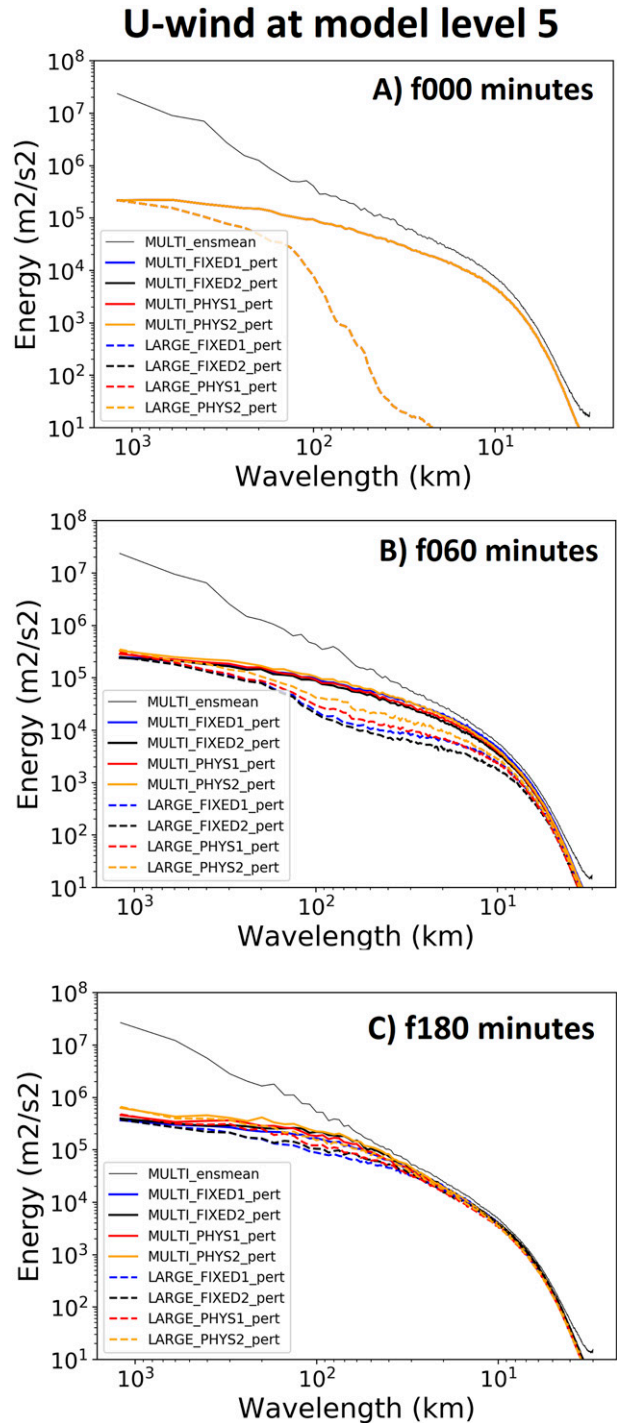


FIG. 5. Perturbation energy spectra of the *u*-wind component at model level 5 (~850 hPa) for different physics configurations of the MULTI and LARGE ensembles at the (a) 0-, (b) 60-, and (c) 180-min forecast lead time. The thin black line is the energy spectrum for the MULTI\_FIXED1 ensemble mean.

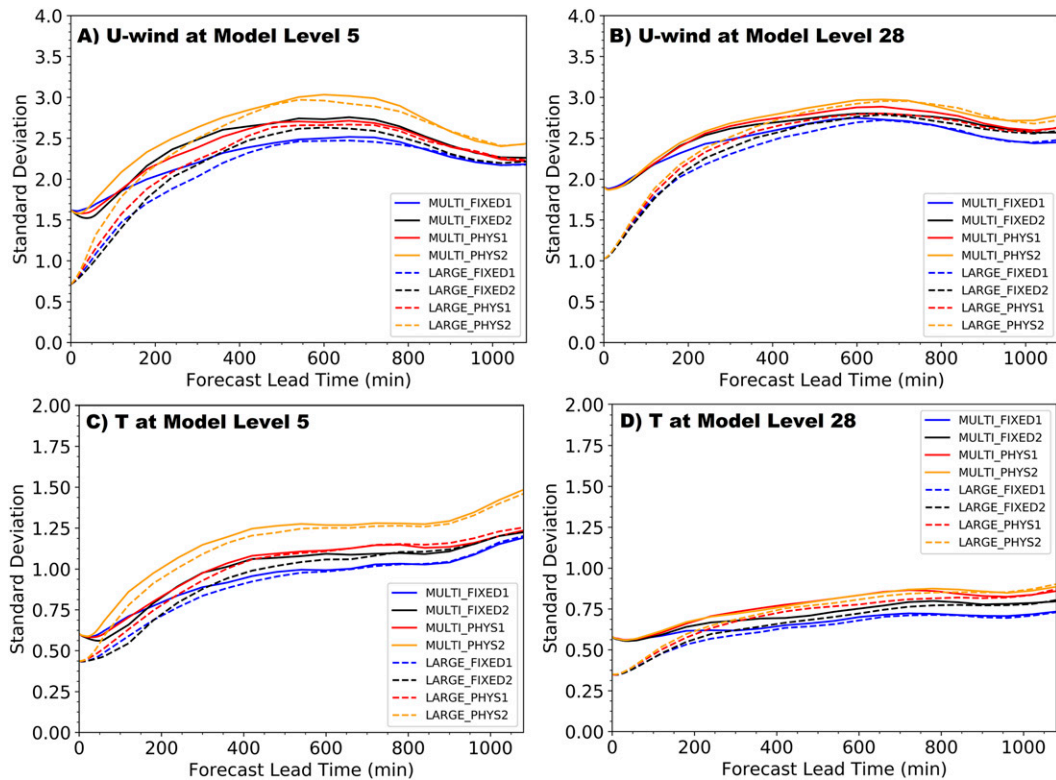


FIG. 6. Ensemble standard deviation of the forecast perturbations of different configurations of the MULTI and LARGE ensembles as a function of forecast lead time for (a) the  $u$ -wind component at model level 5 ( $\sim 850$  hPa), (b) the  $u$ -wind component at model level 28 ( $\sim 250$  hPa), (c) potential temperature at model level 5, and (d) potential temperature at model level 28.

functions shown in Fig. 3. The filtering was performed using the 2D discrete cosine transform (DCT2D; Denis et al. 2002) by converting the perturbations to spectral space using the DCT2D, reducing the amplitude of the spectral coefficient by a factor between 0.0 and 1.0 depending on the wavelength of the spectral coefficient, then converting the modified coefficients back to physical spacing using the inverse DCT2D. Figure 4 illustrates qualitatively the differences among several experiments for a single perturbation of the  $u$  component of wind at model level 5 ( $\sim 850$  hPa). The presence of convective-scale detail in MULTI but not LARGE is clear from comparing Figs. 4a and 4b, respectively. However, while the MULTI\_FILTER3 and LARGE\_FILTER3 perturbations have similar spatial scale, they also have different spatial patterns which sample different modes of IC uncertainty, as seen by comparing Figs. 4c and 4d, respectively.

### 3. Results

#### a. Nonprecipitation variables

The distribution of perturbation energy across spatial scales in the different experiments is first evaluated

using domain-wide perturbation energy spectra of non-precipitation variables. The variables considered include  $u$  and  $v$  wind components and potential temperature, at model levels 5 ( $\sim 850$  hPa), 18 ( $\sim 500$  hPa), and 29 ( $\sim 250$  hPa). Results were generally consistent among the different variables and levels, so Fig. 5 shows  $u$  wind at model level 5 as a representative example. The spectra in Fig. 5 are calculated as the average of all 1D latitudinal spectra (Skamarock 2004). The spectra are calculated using the discrete cosine transform, and averaged over all 10 cases and all 10 ensemble members.

While the initial differences in perturbation energy on small scales between MULTI and LARGE (Fig. 5a) are rapidly reduced, there remains a pronounced difference on meso- $\beta$  scales of  $\sim 30$  to  $\sim 300$  km at the 3-h forecast time (Fig. 5d). The impact of physics diversity dominates over the impact of IC perturbation method on larger scales, as indicated by the primary clustering of the spectra according to the physics configuration on scales greater than  $\sim 400$  km (Fig. 5). However, the impact of IC uncertainty is dominant on the mesoscales for these lead times. In general, as lead time increases the relative impact of the physics configuration dominates on



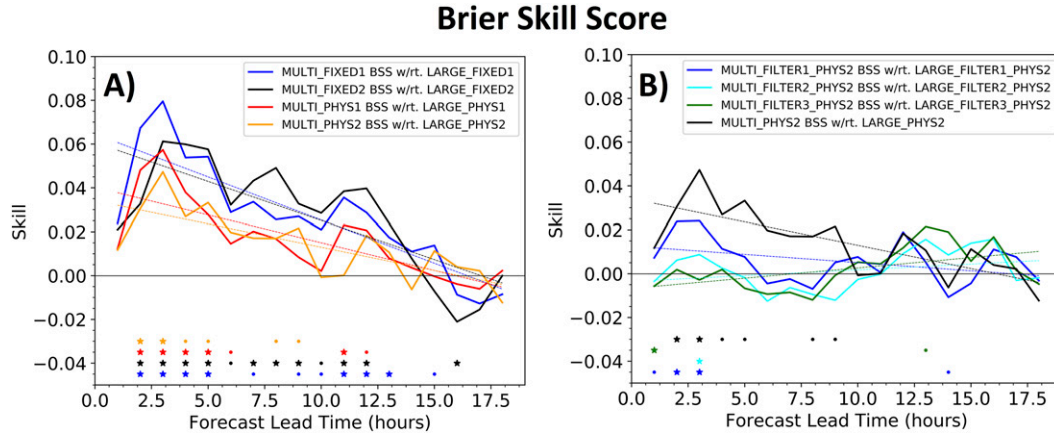


FIG. 7. Brier skill score for neighborhood maximum ensemble probability (NMEP) of hourly precipitation exceeding 6.35 mm (a) for each physics configuration of the MULTI ensemble with respect to the corresponding LARGE ensemble as a reference forecast and (b) for different scales of filtering the IC perturbations in the phys2 ensemble. Thin dashed trend lines are also shown. Statistically significant difference from zero skill is indicated at the 90% confidence level with a star and at the 80% confidence level with a dot along the bottom axis.

increasingly smaller scales. At the 1-h lead time the spectra cluster by IC up to  $\sim 400$  km (Fig. 5b), while at the 2-h (3-h) lead time this clustering is only seen on scales up to about 200 km (100 km) (Figs. 5c,d).

In terms of the total spread of nonprecipitation variables, the differences between IC perturbation methods generally dominate during the first  $\sim 3$  h while the differences among physics configurations generally dominate after the first  $\sim 6$  h (Fig. 6). The relative impact of the IC perturbation method lasts a little longer at upper levels (Figs. 6b,d) than at lower levels (Figs. 6a,c). For each physics configuration, MULTI tends to maintain slightly more spread than the corresponding LARGE ensemble even after the time when the physics configuration dominates the clustering of the lines.

### b. Hourly accumulated precipitation

The experiments are next evaluated in terms of the skill of probabilistic forecasts of hourly accumulated precipitation in mesoscale neighborhoods. Precipitation thresholds of 2.54, 6.35, and 12.7 mm  $\text{h}^{-1}$  were all evaluated and found to provide similar conclusions. Therefore, the representative threshold of 6.35 mm  $\text{h}^{-1}$  is used in the following figures. The verification is based on the neighborhood maximum ensemble probability (NMEP; Schwartz and Sobash 2017) with a neighborhood radius of 15 km. Other neighborhood radii revealed similar qualitative conclusions (not shown). To avoid arbitrarily defining a “reference” forecast, the Brier skill score is used to evaluate the skill of forecasts from experiment A, with respect to the forecasts from experiment B, as follows:

$$\text{BSS}_{A\text{ wrt } B} = 1 - \frac{\text{BS}_A}{\text{BS}_B}, \quad (1)$$

where BS is the Brier score (Brier 1950) of the NMEP forecast from the corresponding experiment. In this context a positive BSS indicates a more skillful forecast with experiment A than experiment B, while a negative BSS indicates a more skillful forecast with experiment B than experiment A. When evaluating skill over all 10 cases, statistical significance is determined using paired sample permutation resampling (Hamill 1999; Johnson and Wang 2012).

MULTI shows a skill advantage over LARGE for all experiments (Fig. 7a). The advantage generally decreases with lead time, although it is not clear how much of the loss of MULTI advantage with forecast lead time is simply a result of convection dissipating or moving out of the model domain at later times for these cases. The MULTI advantage generally corresponds to an improved reliability of the probabilistic forecasts (e.g., at the 1–3-h lead times in Fig. 8). In both the fixed and multiphysics MULTI ensembles, there is less underforecasting at the low (below about 30%) forecast probabilities than all LARGE ensembles (Fig. 8). For each physics configuration, the MULTI ensemble also shows reduced overforecasting at the higher (above about 60%) forecast probabilities than the corresponding LARGE ensemble. The reduced overconfidence of precipitation forecasts in MULTI compared to LARGE at early lead times is consistent with the pronounced increase in spread of nonprecipitation variables at these times in Fig. 6.

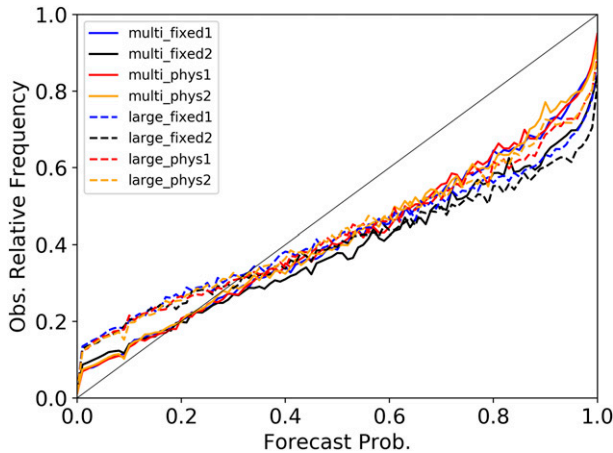


FIG. 8. NMEP forecast reliability diagram during the first three forecast hours for the different physics configurations of the MULTI (solid) and LARGE (dashed) ensembles. Thin black line indicates the line of perfect reliability.

The impact of including the ensemble physics diversity is generally to reduce the skill advantage of MULTI over LARGE (red and orange lines; Fig. 7a). However, even with the physics diversity a clear advantage of MULTI over LARGE is still found. The relative sensitivity of the NMEP forecasts to IC and physics perturbations is further evaluated by calculating at each grid point the absolute value of the difference between the MULTI\_FIXED1 and LARGE\_FIXED1 NMEP forecast, minus the difference between the LARGE\_FIXED1 and LARGE\_PHYS2 NMEP forecast. Positive (negative) values of sensitivity indicate that the LARGE\_FIXED1 forecast is more sensitive to changing the IC (physics) configuration than the physics (IC) configuration (Fig. 9). On average, a dominant impact of IC configuration relative to the physics configuration is greatest at the early lead times and lasts longer for the strongly forced cases than the weakly forced cases (Fig. 9). At later lead times, the greater relative impact of physics configuration for strongly forced cases (blue line; Fig. 9) than weakly forced cases (green line; Fig. 9) is likely due to greater overall precipitation on the strongly forced cases (not shown).

For ease of qualitative interpretation, the sensitivity values are smoothed with a Gaussian convolution with radius 90 km before plotting an illustrative example from the (weakly forced) case of forecasts initialized at 0400 UTC 26 June 2015 in Fig. 10. Figure 10 shows the overall transition from IC configuration dominated probabilistic forecast sensitivity (positive values) to physics configuration dominated sensitivity (negative values), consistent with Fig. 9. However, the transition is not uniform across the domain, and there are some exceptions to the overall trend. For example, a particularly

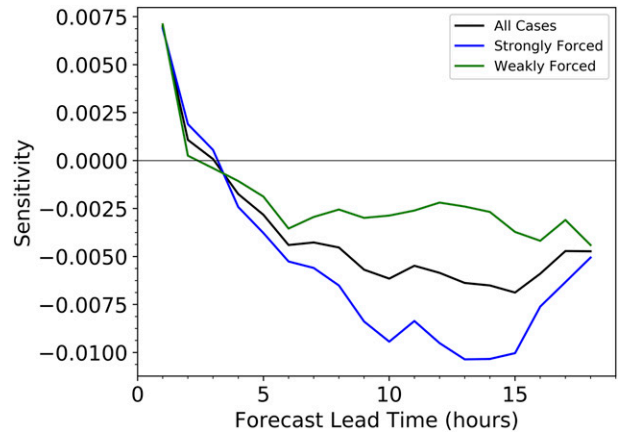


FIG. 9. Time series of the average value of the metric for ensemble IC configuration sensitivity relative to physics configuration sensitivity (defined in section 3b) for all forecast cases (black), cases with strong synoptic-scale forcing (blue), and cases with weak synoptic-scale forcing (green).

large area of physics-dominated sensitivity develops at forecast hours 4–6 in the vicinity of the maturing MCS in northern Missouri, and at forecast hours 9–12 (and beyond) in the vicinity of another maturing MCS in southeast Kansas and southwest Missouri (Fig. 10). Areas of IC-dominated sensitivity also temporarily develop after the first few forecast hours in areas of newly initiating convection, such as in northeast Kansas at forecast hours 7–8 and in northeast to central Iowa at forecast hours 5–12 (Fig. 10). Similar qualitative trends were also observed in other cases, in addition to areas of temporary IC-dominated sensitivity around the dissipation time of an MCS in a few other cases (not shown). Thus, while improved physics diversity can lead to similar skill improvements as the improved IC perturbations from MULTI instead of LARGE, the uncertain processes sampled by the IC and physics perturbations can be quite different. An optimal ensemble configuration therefore should include both multiphysics and multiscale IC perturbations that are optimized together.

### c. Skill contributions from different spatial scales

We now consider the contributions to the MULTI skill advantage from different spatial scales in the context of the best-performing phys2 physics configuration. As shown in Fig. 7b by the generally reduced skill of MULTI\_FILTER1, MULTI\_FILTER2, and MULTI\_FILTER3 with respect to LARGE\_FILTER1, LARGE\_FILTER2, and LARGE\_FILTER3 (blue, cyan, and green lines, respectively; Fig. 7b) compared to the skill of MULTI with respect to LARGE (black line; Fig. 7b) at lead times of ~3–8 h, most of the advantage of MULTI over LARGE during these times comes from

### IC sensitivity relative to physics for 26 June 2015 case

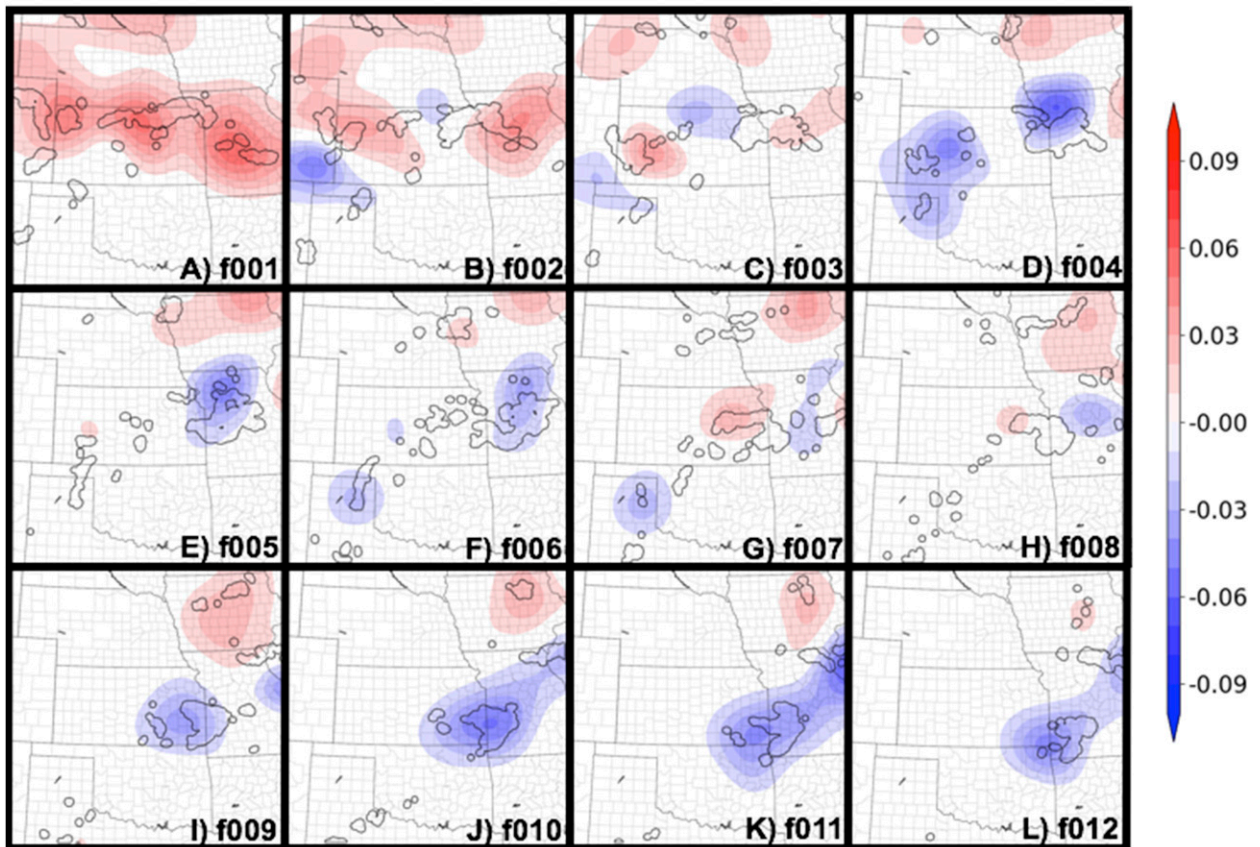


FIG. 10. Physics sensitivity at different forecast hours for the 26 Jun 2015 case, smoothed with a Gaussian convolution with a 90-km radius. The black contour is the 6.35-mm threshold of the observed hourly precipitation.

scales that are not well resolved by the LARGE perturbations. At the later lead times, the skill differences are generally not statistically significant. Although consistent differences among the lines in Fig. 7b are difficult to discern after 10-h lead time, the MULTI\_FILTER1 skill (blue line) is reduced from the MULTI skill (black line) mainly during forecast hours  $\sim 5$ –9. The MULTI\_FILTER2 (cyan line) and MULTI\_FILTER3 (green line) skills are reduced compared to the MULTI skill (black line) even during the first few forecast hours.

For the nonprecipitation variables, upscale growth of the small-scale, flow-dependent IC perturbations is evident at even longer lead times. Figure 11 shows the ensemble spread after smoothing the perturbations with a Gaussian convolution with a radius of 90 km before calculating the spread. Comparing the smoothed spread in MULTI and LARGE (Fig. 11a) to the smoothed spread in MULTI\_FILTER3 and LARGE\_FILTER3 (Fig. 11b) reveals the impact of upscale growth from the filtered small-scale perturbations to the larger-scale spread. Although the relatively large-scale spread is greater

in MULTI than LARGE throughout the forecast period (Fig. 11a), the large-scale spread in LARGE\_FILTER3 is generally greater than in MULTI\_FILTER3 throughout most of the forecast period (Fig. 11b).

The impact of this upscale growth in spread for nonprecipitation variables on the skill of precipitation forecasts was investigated subjectively for several cases and is now described with a representative case as an example. In the 26 June 2015 case, MULTI generally remains more skillful than LARGE throughout most of the forecast period (Fig. 12). The skill advantage of MULTI over LARGE at later lead times is still present on this case when filtering the IC perturbations to a commonly resolved spatial scale (green line; Fig. 12). The skill results from enhanced spread of some convection in Iowa that developed several hours after the IC time (Fig. 13). The enhanced spread can be traced back to the IC perturbations in the vicinity of an upstream upper-level shortwave disturbance (Fig. 14). Although LARGE is able to resolve this feature, the structure of the spread, in terms of variance of the  $v$  component of

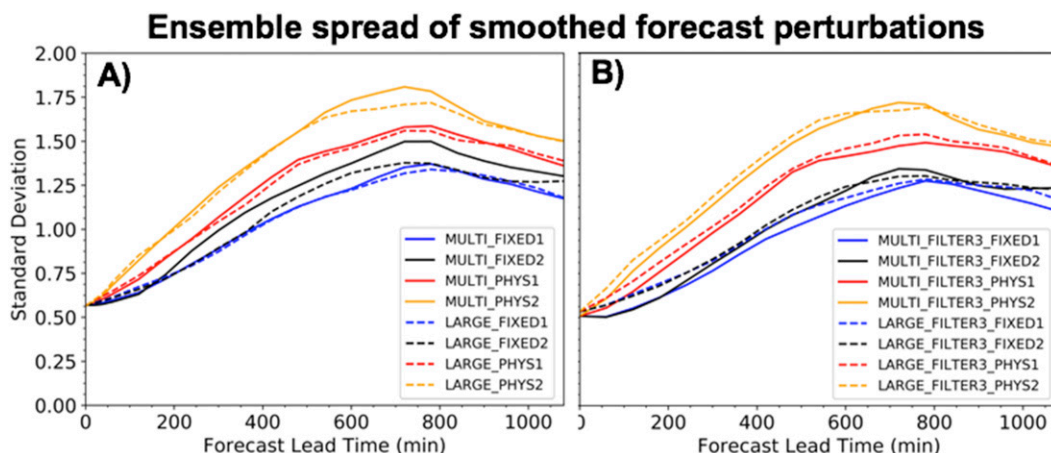


FIG. 11. As in Fig. 6a, but for perturbations that have been smoothed with a Gaussian convolution with 90-km radius before calculating the standard deviation, for different physics configurations of the MULTI and LARGE ensembles (a) without and (b) with filtering of the small-scale component of the IC perturbations.

wind at jet stream level, is different between MULTI and LARGE. While LARGE generally has a larger amplitude of spread here in the first few hours (e.g., Figs. 14b,e), the MULTI perturbations exhibit greater growth such that there is more spread in the location of the shortwave disturbance in MULTI than LARGE by forecast hour 5 when it is supporting the development of new convection in Iowa (Figs. 14c,f). The spread in the shortwave location is implied by the variance of  $v$  wind just ahead of and behind the shortwave trough axis (Figs. 14c,f). This case thus demonstrates an example of better upscale growth of the MULTI perturbations to nonprecipitation variables surrounding the upper-level shortwave disturbance leading to improved ensemble forecasts of the uncertainty in the initiation and evolution of precipitating convection at later lead times, compared to the slower-growing LARGE perturbations to the initial shortwave disturbance.

#### 4. Summary and conclusions

While previous studies have demonstrated advantages of multiscale ensemble IC perturbations from DA on a convection-permitting grid, to the authors' knowledge these studies have not yet been conducted in the context of also having physics diversity in the ensemble configuration. The present study compares ensemble forecasts with IC perturbations generated during multiscale ensemble DA on the convection-permitting grid (MULTI) to forecasts with IC perturbations interpolated from a global ensemble (LARGE) on 10 retrospective cases in the context of two different fixed-physics and two different multiphysics ensembles.

While the small-scale variance in the LARGE ensemble quickly spins up to a similar perturbation energy

spectrum as MULTI during the first few forecast hours, significant advantages of MULTI for probabilistic forecasts of hourly accumulated precipitation are found well beyond that time. The magnitude of this advantage is somewhat reduced, but is still present, when the ensemble members also contain different physics configurations compared to the fixed-physics ensembles. The variance of nonprecipitation variables is mostly determined by the physics configuration on spatial scales larger than  $\sim 400$ ,  $\sim 200$ , and  $\sim 100$  km at the 1-, 2-, and 3-h lead times, respectively. It is mostly determined by the IC configuration on the  $\sim 10$ – $400$ -,  $\sim 30$ – $200$ -, and  $\sim 30$ – $100$ -km spatial scales at the 1-, 2-, and 3-h lead times, respectively. The total variance of nonprecipitation variables is generally larger in MULTI than LARGE, for

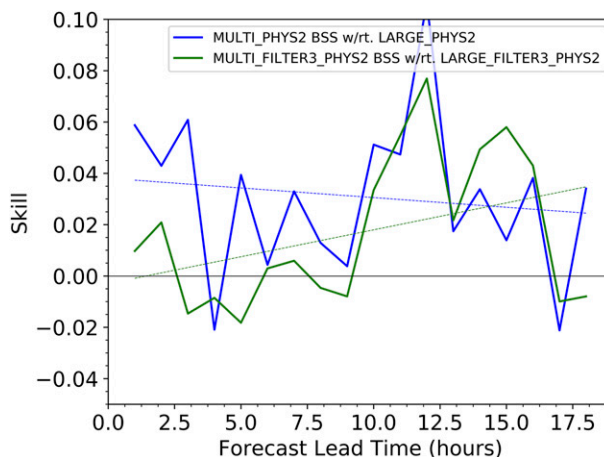


FIG. 12. Brier skill score of MULTI\_PHYS2 with respect to LARGE\_PHYS2 (blue) and MULTI\_FILTER3\_PHYS2 with respect to LARGE\_FILTER3\_PHYS2 (green) for the 26 Jun 2015 forecast case.

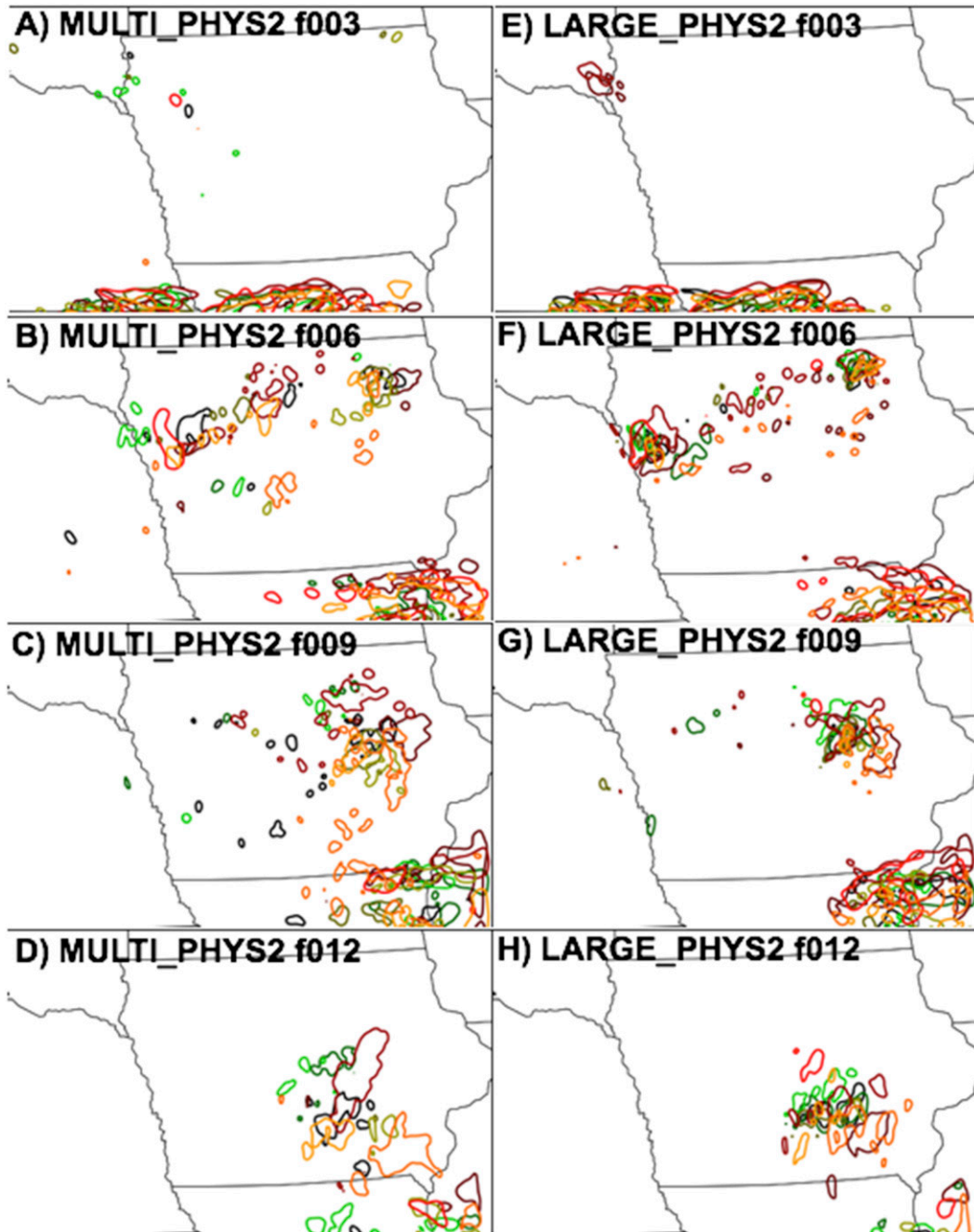


FIG. 13. Spaghetti plots of the  $6.35 \text{ mm h}^{-1}$  contour of hourly accumulated precipitation from all 10 ensemble members overlaid with different colors for the (a)–(d) MULTI\_PHYS2 and (e)–(h) LARGE\_PHYS2 ensembles at the (a),(e) 3-; (b),(f) 6-; (c),(g) 9-; and (d),(h) 12-h forecast lead times of the 26 Jun 2015 case.

any given physics configuration, throughout the 18-h forecast period.

The variance of smoothed forecast perturbations was greater in MULTI than LARGE, but was greater in LARGE\_FILTER1 than MULTI\_FILTER1. This shows the importance of upscale growth of the flow-dependent

small-scale IC perturbations for maintaining forecast spread even out to the 18-h forecast lead time. The advantage of MULTI over LARGE for precipitation forecasts was significantly reduced by filtering the small scales out from the IC perturbations, further emphasizing the importance of the initial small-scale perturbations

### Model Level 28 (~250 hPa) V-wind Ensemble Standard Deviation

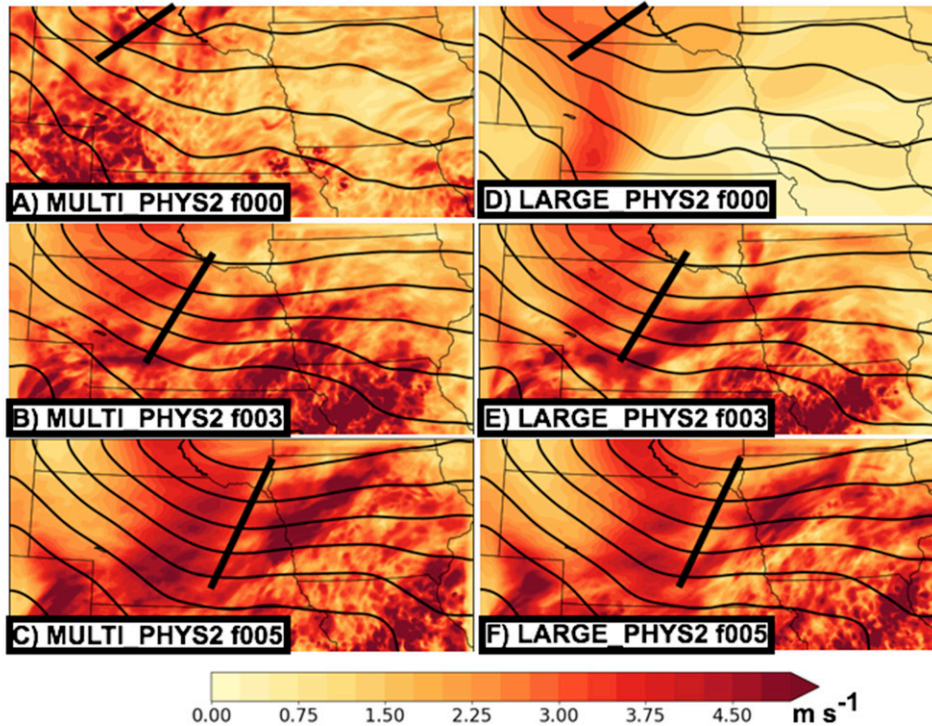


FIG. 14. Ensemble standard deviation of the  $v$ -wind component at model level 28 (~250 hPa) in (a)–(c) MULTI\_PHYS2 and (d)–(f) LARGE\_PHYS2 at the (a),(d) 0-; (b),(e) 3-; and (c),(f) 5-h forecast lead times for the 26 Jun 2015 case.

even after the nonprecipitation variables have spun up similar spread on the small scales. These results suggest that it is only after ~10 h of forecast lead time that the small-scale IC perturbations could potentially be neglected without loss of precipitation forecast skill. However, this also suggests that there is substantial upscale growth from the convective and meso- $\beta$  scales of flow-dependent IC perturbation that affects the quality of the ensemble spread throughout the 18-h forecast period. Thus, it is possible that the importance for precipitation forecasting of sampling the meso- $\beta$ -scale uncertainty with IC perturbations from a multiscale DA system may last even longer in cases containing more active convection at even later lead times than the cases considered in this study. Furthermore, additional cases with active convection at later lead times may also reveal more pronounced advantages of MULTI over LARGE even on the commonly resolved scales.

Qualitative evaluations of the relative impact of the physics and IC perturbation configuration revealed that the potential advantages of optimizing the multiscale IC configuration in a system that also uses physics diversity to maintain ensemble spread can be quite dependent on the physical processes controlling convective evolution.

In particular, cold pool–driven MCS evolution was more strongly determined by the ensemble physics diversity while the IC perturbation configuration was particularly important for cases of newly initiating convection or the timing of MCS decay. Consistent with previous studies such as [Stensrud et al. \(2000\)](#), the relative dominance of the IC perturbation configuration for precipitation forecasts persists longer on strongly forced cases than weakly forced cases.

Previous studies have suggested that, compared to limited area regional models, global-scale models may be better suited for initializing longer-lead-time forecasts than considered in the present study (e.g., [Merkova et al. 2011](#)). In the context of IC perturbation design, it seems plausible that global forecasts cycled at 6-h intervals would better suited to sample the growing modes of IC uncertainty than the more frequently cycled multiscale DA system ([Pena and Kalnay 2004](#)). Therefore, further forecast skill improvements may result from blending the subsynoptic-scale perturbations from the multiscale DA system with the synoptic and larger-scale perturbations from the operational global ensembles. Additional cases and larger forecast domains with longer forecast lead times would be needed to conclusively

test this hypothesis. Such experiments are ongoing and will be reported in a future study.

*Acknowledgments.* The work is primarily supported by Grant NA17OAR4590116. The authors appreciate discussions and model datasets contributed to this work by Nicholas Gasperoni and Yongming Wang in the OU MAP lab. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation Grant ACI-1053575. Some of the computing for this project was also performed at the OU Supercomputing Center for Education and Research (OSCER) at the University of Oklahoma (OU).

**Data availability statement:** All WRF forecast data produced during this study have been archived locally and are available from the authors upon request.

#### REFERENCES

- Anderson, J. L., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate*, **9**, 1518–1530, [https://doi.org/10.1175/1520-0442\(1996\)009<1518:AMFPAE>2.0.CO;2](https://doi.org/10.1175/1520-0442(1996)009<1518:AMFPAE>2.0.CO;2).
- Barker, D. M., and Coauthors, 2012: The Weather Research and Forecast Model's community variational/ensemble data assimilation system. *Bull. Amer. Meteor. Soc.*, **93**, 831–843, <https://doi.org/10.1175/BAMS-D-11-00167.1>.
- 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](https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2).
- Cintineo, R. M., and D. J. Stensrud, 2013: On the predictability of supercell thunderstorm evolution. *J. Atmos. Sci.*, **70**, 1993–2011, <https://doi.org/10.1175/JAS-D-12-0166.1>.
- Denis, B., J. Cote, 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](https://doi.org/10.1175/1520-0493(2002)130<1812:SDOTDA>2.0.CO;2).
- Durran, D. R., and M. Gingrich, 2014: Atmospheric predictability: Why butterflies are not of practical importance. *J. Atmos. Sci.*, **71**, 2476–2488, <https://doi.org/10.1175/JAS-D-14-0007.1>.
- , and J. A. Weyn, 2016: Thunderstorms do not get butterflies. *Bull. Amer. Meteor. Soc.*, **97**, 237–243, <https://doi.org/10.1175/BAMS-D-15-00070.1>.
- Ehrendorfer, M., 1997: Predicting the uncertainty of numerical weather forecasts: A review. *Meteor. Z.*, **6**, 147–183, <https://doi.org/10.1127/metz/6/1997/147>.
- Flora, M. L., C. K. Potvin, and L. J. Wicker, 2018: Practical predictability of supercells: Exploring ensemble forecast sensitivity to initial condition spread. *Mon. Wea. Rev.*, **146**, 2361–2379, <https://doi.org/10.1175/MWR-D-17-0374.1>.
- 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>.
- Gilmore, M. S., J. M. Straka, and E. N. Rasmussen, 2004: Precipitation and evolution sensitivity in simulated deep convective storms: Comparisons between liquid-only and simple ice and liquid phase microphysics. *Mon. Wea. Rev.*, **132**, 1897–1916, [https://doi.org/10.1175/1520-0493\(2004\)132<1897:PAESIS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1897:PAESIS>2.0.CO;2).
- 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](https://doi.org/10.1175/1520-0434(1999)014<0155:HTFENP>2.0.CO;2).
- , C. Snyder, and R. E. Morss, 2000: A comparison of probabilistic forecasts from bred, singular-vector, and perturbed observation ensembles. *Mon. Wea. Rev.*, **128**, 1835–1851, [https://doi.org/10.1175/1520-0493\(2000\)128<1835:ACOPFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<1835:ACOPFF>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. Schar, 2007: Predictability and error growth dynamics in cloud-resolving models. *J. Atmos. Sci.*, **64**, 4467–4478, <https://doi.org/10.1175/2007JAS2143.1>.
- , A. Walser, W. Langhans, and C. Schar, 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>.
- 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](https://doi.org/10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2).
- , 2001: Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP Meso Model. NCEP Office Note 437, 61 pp.
- Jankov, I., J. Beck, J. Wolff, M. Harrold, J. B. Olson, T. Smirnova, C. Alexander, and J. Berner, 2019: Stochastically perturbed parameterizations in an HRRR-based ensemble. *Mon. Wea. Rev.*, **147**, 153–173, <https://doi.org/10.1175/MWR-D-18-0092.1>.
- Johnson, A., and X. Wang, 2012: Verification and calibration of neighborhood and object-based probabilistic precipitation forecasts from a multimodel convection-allowing ensemble. *Mon. Wea. Rev.*, **140**, 3054–3077, <https://doi.org/10.1175/MWR-D-11-00356.1>.
- , and —, 2016: A study of multiscale initial condition perturbation methods for convection-permitting ensemble forecasts. *Mon. Wea. Rev.*, **144**, 2579–2604, <https://doi.org/10.1175/MWR-D-16-0056.1>.
- , and Coauthors, 2014: Multiscale characteristics and evolution of perturbations for warm season convection-allowing precipitation forecasts: Dependence on background flow and method of perturbation. *Mon. Wea. Rev.*, **142**, 1053–1073, <https://doi.org/10.1175/MWR-D-13-00204.1>.
- , X. Wang, J. R. Carley, L. J. Wicker, and C. Karstens, 2015: A comparison of multiscale GSI-based EnKF and 3DVar data assimilation using radar and conventional observations for midlatitude convective-scale precipitation forecasts. *Mon. Wea. Rev.*, **143**, 3087–3108, <https://doi.org/10.1175/MWR-D-14-00345.1>.
- , —, and S. Degelia, 2017: Design and implementation of a GSI-based convection allowing ensemble-based data assimilation

- and forecast system for the PECAN field experiment. Part II: Overview and evaluation of a real-time system. *Wea. Forecasting*, **32**, 1227–1251, <https://doi.org/10.1175/WAF-D-16-0201.1>.
- , 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>.
- 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>.
- Kerr, C., D. J. Stensrud, and X. Wang, 2019: Diagnosing convective dependencies on near-storm environments using ensemble sensitivity analyses. *Mon. Wea. Rev.*, **147**, 495–517, <https://doi.org/10.1175/MWR-D-18-0140.1>.
- Lim, K.-S., and S.-Y. Hong, 2010: Development of an effective double-moment cloud microphysics scheme with prognostic cloud condensation nuclei (CCN) for weather and climate models. *Mon. Wea. Rev.*, **138**, 1587–1612, <https://doi.org/10.1175/2009MWR2968.1>.
- 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>.
- Maldonado, P., and J. Ruiz, 2020: Parameter sensitivity of the WRF-LETKF system for assimilation of radar observations: Imperfect-model observing system simulation experiments. *Wea. Forecasting*, **35**, 1345–1362, <https://doi.org/10.1175/WAF-D-19-0161.1>.
- Mansell, E. R., C. L. Ziegler, and E. C. Bruning, 2010: Simulated electrification of a small thunderstorm with two-moment bulk microphysics. *J. Atmos. Sci.*, **67**, 171–194, <https://doi.org/10.1175/2009JAS2965.1>.
- Merkova, D., I. Szunyogh, and E. Ott, 2011: Strategies for coupling global and limited-area ensemble Kalman filter assimilation. *Nonlinear Processes Geophys.*, **18**, 415–430, <https://doi.org/10.5194/npg-18-415-2011>.
- Milbrandt, J. A., and M. K. Yau, 2005: A multimoment bulk microphysics parameterization. Part I: Analysis of the role of the spectral shape parameter. *J. Atmos. Sci.*, **62**, 3051–3064, <https://doi.org/10.1175/JAS3534.1>.
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682, <https://doi.org/10.1029/97JD00237>.
- Morrison, H., and J. A. Milbrandt, 2015: Parameterization of cloud microphysics based on the prediction of bulk ice particle properties. Part I: Scheme description and idealized tests. *J. Atmos. Sci.*, **72**, 287–311, <https://doi.org/10.1175/JAS-D-14-0065.1>.
- , G. Thompson, and V. Tatarskii, 2009: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Mon. Wea. Rev.*, **137**, 991–1007, <https://doi.org/10.1175/2008MWR2556.1>.
- Nakanishi, M., and H. Niino, 2009: Development of an improved turbulence closure model for the atmospheric boundary layer. *J. Meteor. Soc. Japan*, **87**, 895–912, <https://doi.org/10.2151/jmsj.87.895>.
- Noh, Y., W. G. Cheon, S. Y. Hong, and S. Raasch, 2003: Improvement of the K-profile model for the planetary boundary layer based on large eddy simulation data. *Bound.-Layer Meteor.*, **107**, 401–427, <https://doi.org/10.1023/A:1022146015946>.
- Pena, M., and E. Kalnay, 2004: Separating fast and slow modes in coupled chaotic systems. *Nonlinear Processes Geophys.*, **11**, 319–327, <https://doi.org/10.5194/npg-11-319-2004>.
- Perkey, D. J., and R. A. Maddox, 1985: A numerical investigation of a mesoscale convective system. *Mon. Wea. Rev.*, **113**, 553–566, [https://doi.org/10.1175/1520-0493\(1985\)113<0553:ANIOAM>2.0.CO;2](https://doi.org/10.1175/1520-0493(1985)113<0553:ANIOAM>2.0.CO;2).
- Pleim, J. E., 2007: A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: Model description and testing. *J. Appl. Meteor. Climatol.*, **46**, 1383–1395, <https://doi.org/10.1175/JAM2539.1>.
- 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>.
- 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>.
- , 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>.
- Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev.*, **132**, 3019–3032, <https://doi.org/10.1175/MWR2830.1>.
- , and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., <https://doi.org/10.5065/D68S4MVH>.
- Smirnova, T. G., J. M. Brown, S. G. Benjamin, and J. S. Kenyon, 2016: Modifications to the Rapid Update Cycle Land Surface Model (RUC LSM) available in the Weather Research and Forecasting (WRF) Model. *Mon. Wea. Rev.*, **144**, 1851–1865, <https://doi.org/10.1175/MWR-D-15-0198.1>.
- Stensrud, D. J., J.-W. Bao, and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.*, **128**, 2077–2107, [https://doi.org/10.1175/1520-0493\(2000\)128<2077:UICAMP>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<2077:UICAMP>2.0.CO;2).
- Sukoriansky, S., B. Galperin, and V. Perov, 2005: Application of a new spectral theory of stably stratified turbulence to atmospheric boundary layers over sea ice. *Bound.-Layer Meteor.*, **117**, 231–257, <https://doi.org/10.1007/s10546-004-6848-4>.
- 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>.
- Wang, X., and T. Lei, 2014: GSI-based four-dimensional ensemble-variational (4DensVar) data assimilation: Formulation and single-resolution experiments with real data for NCEP Global Forecast System. *Mon. Wea. Rev.*, **142**, 3303–3325, <https://doi.org/10.1175/MWR-D-13-00303.1>.
- , D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar-based ensemble-variational hybrid data assimilation for NCEP Global Forecast System: Single resolution experiments. *Mon. Wea. Rev.*, **141**, 4098–4117, <https://doi.org/10.1175/MWR-D-12-00141.1>.
- Wang, Y., and X. Wang, 2017: Direct assimilation of radar reflectivity without tangent linear and adjoint of the nonlinear observation



- operator in the GSI-based EnVar system: Methodology and experiment with the 8 May 2003 Oklahoma City tornadic supercell. *Mon. Wea. Rev.*, **145**, 1447–1471, <https://doi.org/10.1175/MWR-D-16-0231.1>.
- , 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>.
- 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>.
- Xue, M., and Coauthors, 2010: CAPS realtime storm scale ensemble and high resolution forecasts for the NOAA Hazardous Weather Testbed 2010 Spring Experiment. *25th Conf. on Severe Local Storms*, Denver, CO, Amer. Meteor. Soc., 7B.3, [https://ams.confex.com/ams/25SLS/techprogram/paper\\_176056.htm](https://ams.confex.com/ams/25SLS/techprogram/paper_176056.htm).
- Zhang, F., N. Bei, R. Rotunno, C. Snyder, and C. C. Epifanio, 2007: Mesoscale predictability of moist baroclinic waves: Convection-permitting experiments and multistage error growth dynamics. *J. Atmos. Sci.*, **64**, 3579–3594, <https://doi.org/10.1175/JAS4028.1>.
- , Y. Weng, Y.-H. Kuo, J. S. Whitaker, and B. Xie, 2010: Predicting typhoon Morakot's catastrophic rainfall with a convection-permitting mesoscale ensemble system. *Wea. Forecasting*, **25**, 1816–1825, <https://doi.org/10.1175/2010WAF2222414.1>.
- Zhou, X., Y. Zhu, D. Hou, Y. Luo, J. Peng, and R. 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>.