

Growth of Spread in Convection-Allowing and Convection-Parameterizing Ensembles

ADAM J. CLARK AND WILLIAM A. GALLUS JR.

Department of Geological and Atmospheric Sciences, Iowa State University, Ames, Iowa

MING XUE

School of Meteorology, and Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma

FANYOU KONG

Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma

(Manuscript received 15 June 2009, in final form 27 October 2009)

ABSTRACT

During the 2007 NOAA Hazardous Weather Testbed Spring Experiment, a 10-member 4-km grid-spacing Storm-Scale Ensemble Forecast (SSEF) system was run in real time to provide experimental severe weather forecasting guidance. Five SSEF system members used perturbed initial and lateral boundary conditions (ICs and LBCs) and mixed physics (ENS4), and five members used only mixed physics (ENS4^{phys}). This ensemble configuration facilitates a comparison of ensemble spread generated by a combination of perturbed ICs/LBCs and mixed physics to that generated by only mixed physics, which is examined herein. In addition, spread growth and spread-error metrics for the two SSEF system configurations are compared to similarly configured 20-km grid-spacing convection-parameterizing ensembles (ENS20 and ENS20^{phys}). Twelve forecast fields are examined for 20 cases.

For most fields, ENS4 mean spread growth rates are higher than ENS20 for ensemble configurations with both sets of perturbations, which is expected as smaller scales of motion are resolved at higher resolution. However, when ensembles with only mixed physics are compared, mass-related fields (i.e., geopotential height and mean sea level pressure) in ENS20^{phys} have slightly higher spread growth rates than ENS4^{phys}, likely resulting from the additional physics uncertainty in ENS20^{phys} from varied cumulus parameterizations that were not used at 4-km grid spacing. For 4- and 20-km configurations, the proportion of spread generated by mixed physics in ENS4 and ENS20 increased with increasing forecast lead time. In addition, low-level fields (e.g., 2-m temperature) had a higher proportion of spread generated by mixed physics than mass-related fields. Spread-error analyses revealed that ensemble variance from the current uncalibrated ensemble systems was not a reliable indicator of forecast uncertainty. Furthermore, ENS4 had better statistical consistency than ENS20 for some mass-related fields, wind-related fields, precipitation, and most unstable convective available potential energy (MUCAPE) with no noticeable differences for low-level temperature and dewpoint fields. The variety of results obtained for the different types of fields examined suggests that future ensemble design should give careful consideration to the specific types of forecasts desired by the user.

1. Introduction

To sufficiently account for model and observational errors so that all possible states of the future atmosphere are simulated, perturbation strategies for recent short-range ensemble forecast (SREF) systems include 1) perturbing the initial conditions (ICs; e.g., Toth and Kalnay

1997; Palmer et al. 1992; Molteni et al. 1996), 2) using different combinations of physical parameterizations (mixed physics; e.g., Houtekamer et al. 1996; Stensrud et al. 2000; Du et al. 2004; Jones et al. 2007), and 3) using different numerical models (e.g., Hou et al. 2001; Wandishin et al. 2001; Du et al. 2004; Eckel and Mass 2005; Jones et al. 2007). In current SREF systems, sensible parameters influenced by small-scale processes that must be parameterized [e.g., planetary boundary layer (PBL) temperature and moisture, convective precipitation] are associated with notably underdispersive

Corresponding author address: Adam J. Clark, Iowa State University, 3010 Agronomy Hall, Ames, IA 50010.
E-mail: clar0614@iastate.edu

forecasts (Fritsch and Carbone 2004; Eckel and Mass 2005). The error growth for these sensible parameters typically contains a much larger contribution from model uncertainty relative to IC uncertainty than for synoptic-scale parameters [e.g., 500-hPa geopotential heights and winds, mean sea level pressure (MSLP); Stensrud et al. 2000; Eckel and Mass (2005)]. The underdispersion may be a result of several deficiencies including 1) inadequate methods of accounting for model error, 2) inability to capture small-scale variability because of insufficient resolution (Eckel and Mass 2005), 3) coarsely resolved and temporally interpolated lateral boundary conditions (LBCs; Nutter et al. 2004), and 4) inadequate sampling of the most important growth directions by the limited-size ensemble.

One method commonly used to gain information about ensemble spread is to isolate the error sources by using different perturbation strategies for a set of forecasts (e.g., Houtekamer et al. 1996; Stensrud et al. 2000; Clark et al. 2008). For example, to isolate model errors, the “perfect analysis” assumption can be used, in which identical sets of ICs/LBCs are used to initialize various ensemble members with mixed physics. Similarly, to isolate IC errors, the “perfect model” assumption can be used in which identically configured ensemble members are initialized with different sets of perturbed ICs. During the 2007 National Oceanic and Atmospheric Administration (NOAA) Hazardous Weather Testbed (HWT) Spring Experiment (SE07; Xue et al. 2007; Kong et al. 2007), a 10-member, 4-km grid-spacing Storm-Scale Ensemble Forecast (SSEF) system was run in real time to provide severe weather forecasting guidance to the SE07 participants. Five of the SSEF members used perturbed ICs/LBCs and mixed physics (ENS4; four perturbed members and one control member), while five members used only mixed physics (ENS4^{phys}) so that the impacts of the different physical parameterization schemes could be isolated. This configuration of the 2007 SSEF system also facilitates an isolation of physics-related model errors because five members use the “perfect analysis” assumption. Unfortunately, because there were not any ensemble subsets with only IC perturbations, the perfect model assumption could not be assessed.

The goal of this paper will be to use the 2007 SSEF system to compare ensemble spread from a mixed-physics-only ensemble to an ensemble with both mixed physics and perturbed IC/LBCs for various fields in a convection-allowing ensemble. In addition, ensemble-spread growth and spread–error relationships associated with the two five-member subsets of the SSEF system will be compared to two similarly configured subsets of a 20-km grid-spacing convection-parameterizing ensemble

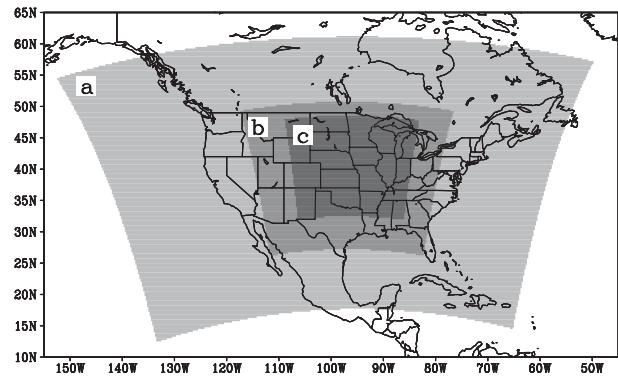


FIG. 1. Domains for (a) NCEP SREF ensemble members, (b) ENS4 and ENS20 ensemble members, and (c) the analyses conducted in this study.

to examine the impacts of horizontal resolution for various forecast fields. This paper is organized as follows: in section 2, a description of the data and methodology is provided, in section 3 the results are examined, and a summary and discussion are provided in section 4.

2. Data and methodology

The 2007 SSEF system was run during April–June 2007 and used version 2.2.0 of the Advanced Research module of the Weather and Research Forecasting model (WRF-ARW; Skamarock et al. 2005). The 10 SSEF members were run by the Center for Analysis and Prediction of Storms (CAPS) of the University of Oklahoma, initialized daily at 2100 UTC, and integrated 33 h over an approximately 3000 km × 2500 km domain covering about two-thirds of the continental United States (Fig. 1). SSEF system ensemble member specifications for ENS4 and ENS4^{phys} are listed in Tables 1 and 2, respectively. Note, because of the 2100 UTC initialization, forecast hours 3 and 27 correspond to 0000 UTC. For the SSEF control member, the 2100 UTC analyses from the National Centers for Environmental Prediction’s (NCEP’s) operational North American Mesoscale (NAM; Janjić 2003) model (at 12-km grid spacing) were used for ICs and the 1800 UTC NAM 12-km forecasts were used for LBCs. For the members with perturbed ICs, perturbations were extracted from the 2100 UTC NCEP SREF system ICs (Du et al. 2004) and added to the 2100 UTC NAM analyses. Corresponding NCEP SREF system forecasts were used for LBCs (3-h updates). Xue et al. (2007) and Kong et al. (2007) provide more details on the configurations.

For a comparison of the five-member SSEF ensemble subsets to a similarly configured convection-parameterizing ensemble, a 30-member 20-km grid-spacing ensemble was generated at Iowa State University, which was also

TABLE 1. ENS4 Ensemble member specifications. NAMA and NAMf indicate NAM analyses and forecasts, respectively; em_pert and nmm_pert are perturbations from different SREF members; and em_n1, em_p1, nmm_n1, and nmm_p1 are different SREF members that are used for LBCs. The remaining table elements are described in the text.

Ensemble member	ICs	LBCs	Microphysics scheme	Surface-layer scheme	Boundary layer scheme
CN	2100UTC NAMA	1800 UTC NAMf	WSM-6	Janjić Eta	MYJ
N1	CN – em_pert	2100 UTC SREF em_n1	Ferrier	Janjić Eta	MYJ
P1	CN + em_pert	2100 UTC SREF em_p1	Thompson	Janjić Eta	MYJ
N2	CN – nmm_pert	2100 UTC SREF nmm_n1	Thompson	Monin–Obukhov	YSU
P2	CN + nmm_pert	2100 UTC SREF nmm_p1	WSM-6	Monin–Obukhov	YSU

composed of WRF-ARW (version 2.2.0) members. Fifteen of the 20-km members have mixed physics and perturbed ICs/LBCs (ENS20), while another 15 members have only mixed physics (ENS20^{phys}). Ensemble member specifications for ENS20 and ENS20^{phys} are provided in Tables 3 and 4, respectively. It should be noted that the ENS20 and ENS20^{phys} ensembles have one more set of varied physics schemes relative to ENS4 and ENS4^{phys}, because in addition to the different PBL, microphysics, and surface-layer schemes, both 20-km ensembles use different cumulus parameterizations (CPs; the 4-km ensembles do not use CP). However, five-member subsets of ENS20 and ENS20^{phys} that use the same CP do have the same number of varied physics schemes as ENS4 and ENS4^{phys}. These five-member subsets will be referred to as ENS20_{cp} and ENS20_{cp}^{phys}, where CP refers to one of the three different CPs used: 1) Kain–Fritsch (KF; Kain and Fritsch 1993), 2) Betts–Miller–Janjić (BMJ; Betts 1986; Betts and Miller 1986; Janjić 1994), or 3) Grell–Devenyi (GD; Grell and Devenyi 2002). For the 20-km ensemble members, different sets of ICs and corresponding LBCs for each member are obtained directly from NCEP SREF members initialized at 2100 UTC.

Both ensembles use the rapid radiative transfer method (RRTM) scheme for shortwave radiation (Mlawer et al. 1997) and the Goddard longwave radiation scheme (Chou and Suarez 1994), along with the Noah land surface model (Ek et al. 2003). Varied PBL schemes include the Mellor–Yamada–Janjić (MYJ; Mellor and Yamada 1982; Janjić 2002) and Yonsei University

(YSU; Noh et al. 2003) schemes. Varied microphysics schemes include Thompson et al. (2004), WRF single-moment six-class method (WSM-6; Hong and Lim 2006), and Ferrier et al. (2002), as well as surface-layer schemes including Monin–Obukhov (Monin and Obukhov 1954; Paulson 1970; Dyer and Hicks 1970; Webb 1970) and the Janjić Eta Model (Janjić 1996, 2002). Note that none of the ensemble members use positive-definite advection of moisture, which may have contributed to high biases in the precipitation forecasts (e.g., Skamarock and Weisman 2009).

The forecasts were examined for 20 cases during April–June 2007 (Fig. 2). These 20 cases are included in the 23 cases in which ENS4 and ENS20 precipitation forecasts were compared in Clark et al. (2009). The other three cases examined in Clark et al. (2009) are excluded from the current study because some of the ENS4^{phys} members were not available. As noted by Clark et al. (2009), the period examined was relatively active with a variety of convective precipitation events.

This study examines the growth of spread (i.e., ensemble variance) and statistical consistency [i.e., correspondence between ensemble variance and mean square error of the ensemble mean (MSE)] for 12 fields: 500-, 700-, and 850-hPa geopotential heights (500Z, 700Z, and 850Z, respectively); MSLP; 2-m temperature (T2); 2-m dewpoint (Td2); 850-hPa wind magnitude (850WMAG), 850-hPa temperature (850T), and 850-hPa dewpoint (850Td); 3-hourly accumulated precipitation (PREC); most unstable convective available potential energy (MUCAPE); and magnitude of the 10-m to

TABLE 2. As in Table 1, but for ENS4^{phys} ensemble member specifications.

Ensemble member	ICs	LBCs	Microphysics scheme	Surface-layer scheme	Boundary layer scheme
PH1	2100 UTC NAMA	1800 UTC NAMf	Thompson	Janjić Eta	MYJ
PH2	2100 UTC NAMA	1800 UTC NAMf	Ferrier	Janjić Eta	MYJ
PH3	2100 UTC NAMA	1800 UTC NAMf	WSM-6	Monin–Obukhov	YSU
PH4	2100 UTC NAMA	1800 UTC NAMf	Thompson	Monin–Obukhov	YSU
PH5	2100 UTC NAMA	1800 UTC NAMf	Ferrier	Monin–Obukhov	YSU

TABLE 3. ENS20 Ensemble member specifications. The members are grouped into five-member subsets that have the same cumulus parameterizations. The ICs/LBCs table elements represent various SREF members and the remaining table elements are described in the text.

Ensemble member	ICs/LBCs	Cumulus scheme	Microphysics	Surface layer	Boundary layer
ENS20 _{BMJ}					
1	em_ctl	BMJ	Thompson	Janjić Eta	MYJ
2	em_p1	BMJ	WSM-6	Janjić Eta	MYJ
3	em_n1	BMJ	WSM-6	Monin–Obukhov	YSU
4	nmm_ctl	BMJ	Thompson	Monin–Obukhov	YSU
5	nmm_p1	BMJ	Ferrier	Monin–Obukhov	YSU
ENS20 _{KF}					
6	nmm_n1	KF	Thompson	Janjić Eta	MYJ
7	eta_ctl1	KF	WSM-6	Janjić Eta	MYJ
8	eta_n1	KF	WSM-6	Monin–Obukhov	YSU
9	eta_n2	KF	Thompson	Monin–Obukhov	YSU
10	eta_n3	KF	Ferrier	Monin–Obukhov	YSU
ENS20 _{GD}					
11	eta_n4	Grell	Thompson	Janjić Eta	MYJ
12	eta_p1	Grell	WSM-6	Janjić Eta	MYJ
13	eta_p2	Grell	WSM-6	Monin–Obukhov	YSU
14	eta_p3	Grell	Thompson	Monin–Obukhov	YSU
15	eta_p4	Grell	Ferrier	Monin–Obukhov	YSU

500-hPa shear vector (WSHR). The 12 fields examined are separated into those that are “mass related,” or heavily dependent on the properties of the atmosphere within a vertical column (500Z, 700Z, 850Z, and MSLP), and “low level” fields that have more dependence on boundary layer processes and, thus, have a noticeable diurnal signal (T2, Td2, 850WMAG, 850T, 850Td, PREC, MUCAPE, and WSHR). In the subsequent analyses of ensemble variance and MSE, comparisons are made between ensemble subsets that have different numbers of members. Thus, it should be kept in mind

that the forecast probability distribution function (PDF) sampled by the smaller ENS4 membership should be less representative of the forecast PDF than the larger ENS20 membership, which would imply a less “certain” estimate of the ensemble variance in ENS4.

For the computation of MSE and the application of a bias-correction procedure in section 3a(4), operational 20-km grid-spacing Rapid Update Cycle (RUC) model analyses provided by NCEP and available at 1-hourly temporal resolution are used as “truth” for nonprecipitation fields. The RUC analyses are generated using hourly

TABLE 4. As in Table 3, but for ENS20^{phys} ensemble member specifications.

Ensemble member	ICs/LBCs	Cumulus scheme	Microphysics	Surface layer	Boundary layer
ENS20 ^{phys} _{BMJ}					
16	eta_ctl2	BMJ	Thompson	Janjić Eta	MYJ
17	eta_ctl2	BMJ	WSM-6	Janjić Eta	MYJ
18	eta_ctl2	BMJ	WSM-6	Monin–Obukhov	YSU
19	eta_ctl2	BMJ	Thompson	Monin–Obukhov	YSU
20	eta_ctl2	BMJ	Ferrier	Monin–Obukhov	YSU
ENS20 ^{phys} _{KF}					
21	eta_ctl2	KF	Thompson	Janjić Eta	MYJ
22	eta_ctl2	KF	WSM-6	Janjić Eta	MYJ
23	eta_ctl2	KF	WSM-6	Monin–Obukhov	YSU
24	eta_ctl2	KF	Thompson	Monin–Obukhov	YSU
25	eta_ctl2	KF	Ferrier	Monin–Obukhov	YSU
ENS20 ^{phys} _{GD}					
26	eta_ctl2	Grell	Thompson	Janjić Eta	MYJ
27	eta_ctl2	Grell	WSM-6	Janjić Eta	MYJ
28	eta_ctl2	Grell	WSM-6	Monin–Obukhov	YSU
29	eta_ctl2	Grell	Thompson	Monin–Obukhov	YSU
30	eta_ctl2	Grell	Ferrier	Monin–Obukhov	YSU

April 2007							May 2007							June 2007						
S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S
1	2	3	4	5	6	7			1	2	3	4	5						1	2
8	9	10	11	12	13	14	6	7	8	9	10	11	12	3	4	5	6	7	8	9
15	16	17	18	19	20	21	13	14	15	16	17	18	19	10	11	12	13	14	15	16
22	23	24	25	26	27	28	20	21	22	23	24	25	26	17	18	19	20	21	22	23
29	30						27	28	29	30	31			24	25	26	27	28	29	30

FIG. 2. Gray-shaded dates indicate when 10-member SSEF system simulations were conducted for SE07 and dark gray shading indicates which cases are used in the analysis for this study.

intermittent three-dimensional variational data assimilation (3DVAR) cycles in which recent observations from various sources (e.g., wind profiler, radar, aircraft, surface aviation routine weather reports (METARs, satellites, etc.) are assimilated using the previous 1-h RUC model forecasts as the background field. Additional information on the RUC model is found in Benjamin et al. (2004a,b). For precipitation fields, stage IV multisensor rainfall estimates (Baldwin and Mitchell 1997) are used. The ensemble variance and MSE are computed according to Eqs. (B7) and (B6), respectively, in Eckel and Mass (2005), which are designed to account for an ensemble with a finite number of members. Finally, to obtain a more equitable comparison between the 4- and 20-km forecast fields, the 4-km fields were remapped to a 20-km grid covering the central United States, which is simply a subdomain of the ENS20 members, using a neighbor-budget interpolation (e.g., Accadia et al. 2003).

3. Results

a. Spread growth

1) ENSEMBLE VARIANCE TIME SERIES

To illustrate the temporal evolution of spread growth during the 33-h forecast period, time series of the average ensemble variance for all 12 fields at 3-hourly intervals with box plots overlaid to show variability are displayed in Fig. 3. To compare 4- and 20-km ensembles with the same types of perturbations, each panel in Fig. 3 displays the ensemble variance for ENS4 and ENS20, or ENS4^{phys} and ENS20^{phys}. Note that the different y-axis scales in Fig. 3 do not allow an easy comparison of the growth rates between ensembles with both IC/LBC perturbations and mixed physics and those with only mixed physics (Phys) for each field; these comparisons are made in the next section. A number of distinct features can be seen in these time series. First, for the mass-related fields (Figs. 3a–h), ENS4 and ENS20 (Figs. 3a, 3c, 3e, and 3g) have a generally linear increase in mean

spread and it appears that ENS4 spread is increasing at a faster rate than ENS20, which is a generally expected pattern of behavior because the smaller scales being resolved in ENS4 should be associated with faster perturbation growth that feeds back to the larger scales (e.g., Lorenz 1969). Also, the biggest differences between ENS4 and ENS20 at each time occur at the higher ends of the variance distributions; that is, differences in the upper part of the box-plot ranges are greater than differences in the lower part, indicating that the distributions are more right skewed in ENS4 relative to ENS20. For the mass-related fields in ENS4^{phys} and ENS20^{phys} (Figs. 3b, 3d, 3f, and 3h), the ENS20^{phys} mean spread increases at a faster rate than in ENS4^{phys}, with the exception of mean MSLP spread (Fig. 3h), which appears to be similar. In addition, the spread increases in ENS4^{phys} and ENS20^{phys} are not linear as they were for ENS4 and ENS20, but instead have a ~6-h period during forecast hours 21–27 (1800–0000 UTC), during which the spread increases at a noticeably faster rate than at the other times. This 6-h period corresponds to when peak insolation occurs and likely corresponds to when the different physics parameterizations are most active and thus result in the most spread increase. For example, in the central United States, the boundary layer typically reaches its maximum depth by early afternoon so that turbulent processes that must be parameterized are occurring over a relatively deep layer. In addition, peak heating and the resulting well-mixed boundary layers also lead to shallow and deep convective clouds, requiring microphysics and cumulus (only for ENS20 and ENS20^{phys}) parameterizations to be more active relative to other times.

For the low-level fields (Figs. 3i–x), ENS4 and ENS20 mean variances (Figs. 3i, 3k, 3m, 3o, 3q, 3s, 3u, and 3w) have clear diurnal signals superimposed on increasing trends. The differences in the mean variances between ENS4 and ENS20 vary among the different fields analyzed. For example, the 850WMAg variances in ENS4 and ENS20 are very similar over the entire forecast

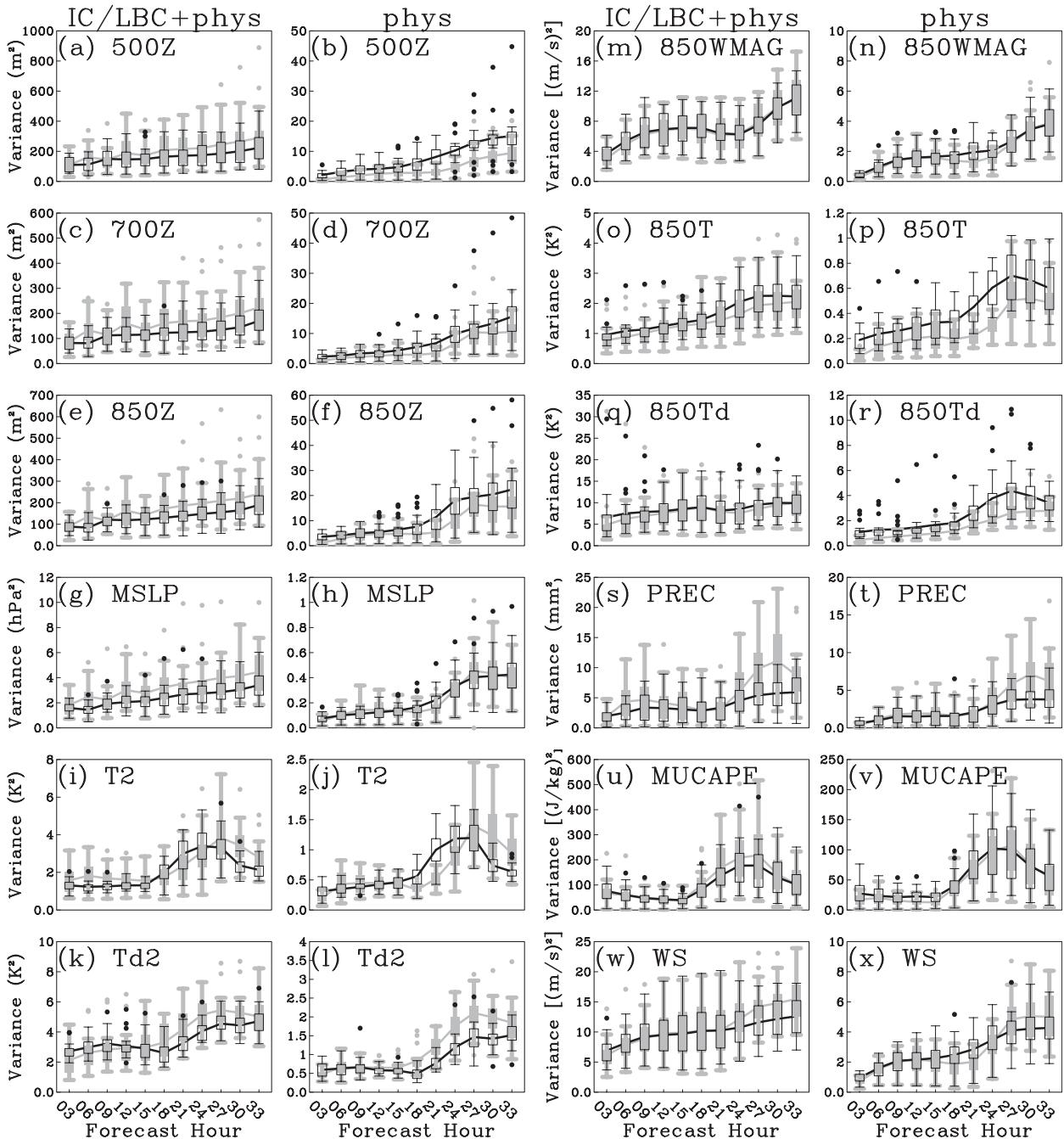


FIG. 3. Time series of mean ensemble variance from ENS4 (gray line) and ENS20 (black line) for the following variables: (a) 500Z, (c) 700Z, (e) 850Z, (g) MSLP, (i) T2, (k) Td2, (m) 850WMAG, (o) 850T, (q) 850Td, (s) PREC, (u) MUCAPE, (w) WSHR. (b),(d),(f),(h),(j),(l),(n),(p),(r),(t),(v),(x) As in (a),(c),(e),(g),(i),(k),(m),(o),(q),(s),(u),(w) but for ENS4^{phys} (gray line) and ENS20^{phys} (black line). Box plots overlay the mean at each time interval. For the box plots, the interquartile range (IQR) is indicated by the area enclosed by a box, outliers defined by values outside of $1.5 \times \text{IQR}$ are marked by dots, and horizontal lines mark the smallest and largest values that are not outliers.

period (Fig. 3m); Td2 variances and variance growth rates are higher in ENS4 relative to ENS20 for most of the forecast period (Fig. 3k), and WSHR variances are similar until forecast hour 21, when there is a marked

increase in ENS4 variances relative to ENS20 (Fig. 3w). The amplitude and phase of the diurnal signal also vary among the different fields. For example, the 850T and 850Td ENS4 and ENS20 variances (Figs. 3o and 3q,

respectively) have smaller amplitudes relative to the other low-level fields, and peak variances occur for 850WMA around 0900–1500 UTC (Fig. 3m), for PREC at 0600–0900 UTC (Fig. 3s), and for MUCAPE at 2100–0000 UTC (Fig. 3u). The peak variances tend to match the time at which the forecasts of the variable considered are maximized. To illustrate this pattern of behavior, time series of mean domain-averaged T2 from all cases for ENS4 and ENS20 members and ensemble means, as well as RUC analyses, are displayed in Fig. 4. The peak T2 values occur around forecast hour 25 when there also appears to be the most spread in domain-averaged T2 among the ensemble members. Also worth noting in Fig. 4 is that all ENS4 members are cooler than the RUC analyses when T2 peaks; however, ENS20 has about equal numbers of members with warm and cool biases resulting in a mean that is very close to the RUC analysis. Coniglio et al. (2009) found similar cool biases for mean 2-m temperatures in convection-allowing WRF model simulations run over the central United States during spring 2008.

2) VARIANCE GROWTH RATES

It was possible to subjectively infer differences in the mean ensemble variance growth rates from the analysis conducted in Fig. 3; however, to compare ENS4 and ENS20 to ENS4^{phys} and ENS20^{phys}, and to better quantify mean variance growth rates, a simple objective method was developed using a standard formula for the growth rate:

$$\frac{(\text{Var}_f - \text{Var}_i)}{\text{Var}_i} \times 100\%, \quad (1)$$

where Var_i and Var_f are the initial and final mean variance, respectively. To reduce the impacts of the diurnal cycle signal on the variance growth rates, mean variances at forecast hours 9 and 33 are used as the initial and final values, respectively, because these forecast hours are separated by 24 h or one complete diurnal cycle. Also, to smooth out some of the higher-frequency variability in the mean variance time series, the Lowess function in the R statistical software package (R Development Core Team 2009), which uses locally weighted polynomial regression, was used as a low-pass filter. For mass-related and low-level fields, the proportions of points influencing the filtered value at each forecast hour, or the “smoother span,” were set to 0.25 and 0.1667, respectively (i.e., ~eight and six points). A slightly larger smoother span was used for the mass-related fields than for low-level fields to adequately filter particularly high-frequency variability during the first 12 h of the forecasts for the mass-related fields. Finally,

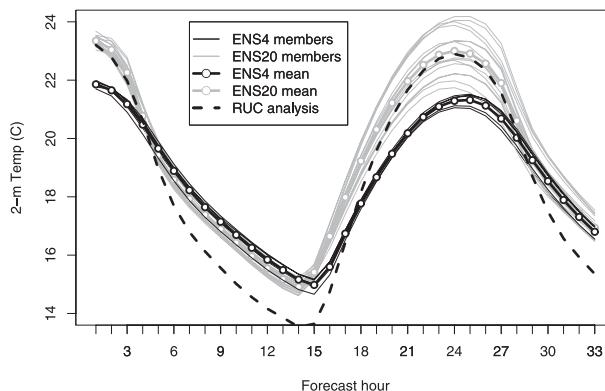


FIG. 4. Time series of mean domain-averaged 2-m temperature from all cases for the ENS4 and ENS20 ensemble members and means, along with the RUC analysis.

the mean variances at all forecast hours from the ENS4^{phys}, ENS20, ENS20^{phys}, and ENS20^{phys} subsets were adjusted by the differences between their variances at forecast hour 9 and that from ENS4 (i.e., the entire time series was shifted by a constant). Thus, the variance growth for all ensemble subsets was computed relative to the same initial mean variance (i.e., the ENS4 mean variance) to allow for comparison between ensemble subsets. The adjusted variance can be expressed as

$$\sigma_{\text{ENS}}^{2*} = \sigma_{\text{ENS}}^2 + [\sigma_{\text{ENS}(\text{thr}9)}^2 - \sigma_{\text{ENS4}(\text{thr}9)}^2], \quad (2)$$

where σ_{ENS}^{2*} is the adjusted variance at any forecast hour for a specified ensemble subset; σ_{ENS}^2 and $\sigma_{\text{ENS}(\text{thr}9)}^2$ are the unadjusted variances for the specified ensemble subset at any forecast hour and at forecast hour 9, respectively; and $\sigma_{\text{ENS4}(\text{thr}9)}^2$ is the unadjusted variance at forecast hour 9 for the ENS4 ensemble. This adjustment step was important because the growth rates are sensitive to the initial variance, which changes according to the field and ensemble subset. Using this procedure to compute the spread growth rate, a 100% growth rate for any field can be interpreted as a 24-h doubling of spread relative to the initial spread in ENS4. Statistical significance tests were performed for growth rate comparisons of ENS4 versus ENS20 and ENS4^{phys} versus ENS20^{phys} for each field using Welch’s *t* test ($\alpha = 0.05$). This test determined whether the differences between the average filtered and adjusted variances of ENS4 and ENS20 (or ENS4^{phys} and ENS20^{phys}) at forecast hour 33¹ were significant.

¹ Note that after variances are adjusted to the same initial value at forecast hour 9, the variance at forecast hour 33 is the value that determines the growth rate; hence, variances at forecast hour 33 are used for the significance tests.

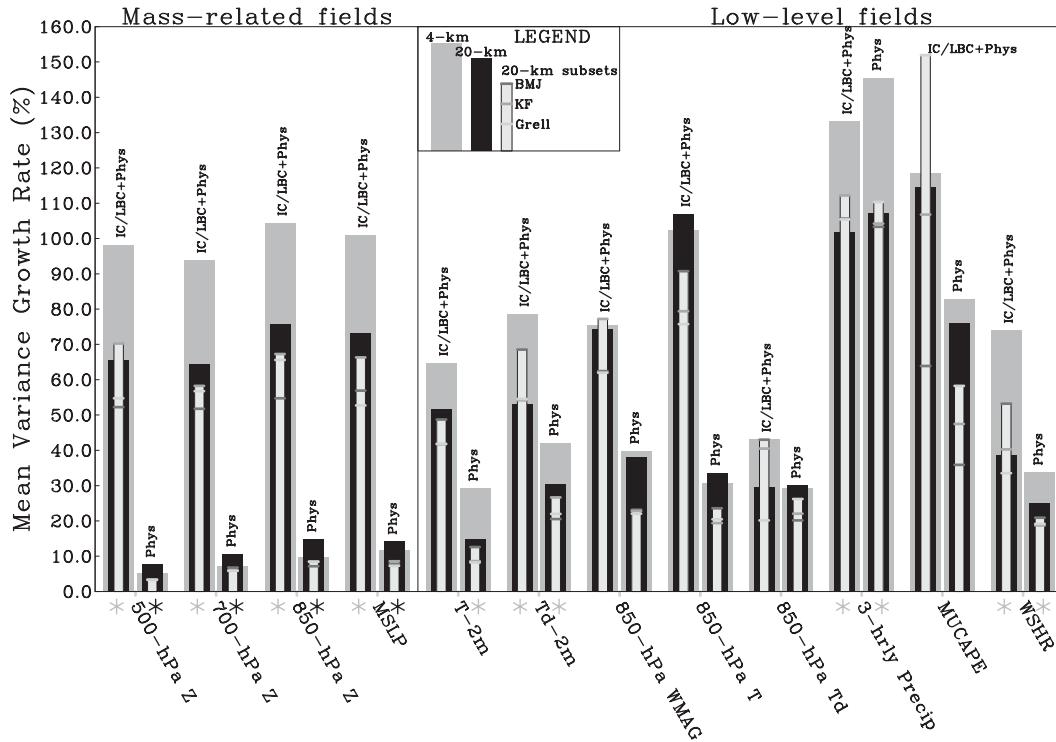


FIG. 5. Mean variance growth rates from the ENS4, ENS4^{phys}, ENS20, and ENS20^{phys} ensembles for fields shown in Fig. 3. Growth rates for five-member subsets of ENS20 and ENS20^{phys} that have the same cumulus parameterization are also shown (marked by the horizontal lines on the lightest gray histogram). The histograms to the left for each variable indicate growth rates for ensembles that have IC/LBC perturbations and mixed physics (IC/LBC+Phys) and the ones to the right are for mixed-physics-only ensembles (Phys). Asterisks below the histogram sets indicate statistically significant differences between ENS4 and ENS20 or ENS4^{phys} and ENS20^{phys} [black (gray) asterisks indicate that ENS4 or ENS4^{phys} (ENS20 or ENS20^{phys}) had larger growth rates]. A legend is provided at the top of the figure.

The mean variance growth rates obtained from this methodology are shown in Fig. 5 (statistically significant differences are indicated by asterisks). For mass-related fields, ENS4 growth rates are ~30% higher than those of ENS20 with differences that are statistically significant. These differences are consistent with faster perturbation growth expected as smaller scales are resolved that feed back to the larger scales (Lorenz 1969; Smagorinsky 1969). The ENS20_{cp} subsets (i.e., five-member subsets with same CPs) tend to have slightly lower growth rates than ENS20, which is consistent with ENS20 having one additional source of model uncertainty relative to the ENS20_{cp} subsets from varied CPs. ENS20 also has one more source of model uncertainty than ENS4, but higher resolution in ENS4 apparently has a greater impact than the additional model uncertainty in ENS20 because ENS4 has much higher growth rates.

For the mass-related fields, the “mixed-physics only” ensemble subsets (denoted Phys in Fig. 5) have growth rates ~90% lower than the subsets with both sets of perturbations (denoted IC/LBC+Phys in Fig. 5). The

much lower growth rates for mass-related fields when using only mixed physics are similar to results found by Kong et al. (2007) using a similar dataset. The differences in growth rates occur because the different physics schemes that parameterize the surface and boundary layer processes mainly influence the PBL, so that mass-related fields dependent on an entire vertical column of the atmosphere exhibit little impact. In addition, the different microphysics and cumulus parameterizations, which can possibly have a more direct influence on layers of the atmosphere above the PBL, are only impacted where the schemes are active, which is usually only over a small fraction of the domain. On the other hand, IC perturbations can directly affect all atmospheric layers and are present over the entire model domain.

All of the growth rates for mass-related fields in ENS20^{phys} are larger than those from ENS4^{phys} with differences that are statistically significant. Thus, unlike the ENS4 versus ENS20 comparison, the impact of one additional source of model uncertainty in ENS20^{phys} (from varied CPs) is greater than the impact of higher

resolution in ENS4^{phys}. Also, similar to the ENS20_{cp} subsets, the ENS20^{phys} subsets have smaller growth rates relative to ENS20^{phys} resulting from having one less source of model uncertainty (no varied CPs).

For low-level fields, the mean spread growth rates are much more variable than in the mass-related fields, and statistically significant differences occur between ENS4 and ENS20 (or ENS4^{phys} and ENS20^{phys}) for T2, Td2, PREC, and WSHR, with ENS4 (or ENS4^{phys}) having the higher growth rates for all the significant differences. This may indicate that, for some variables, higher resolution in ENS4 and ENS4^{phys} results in larger spread growth rates despite the extra source of model uncertainty in ENS20 and ENS20^{phys}, but for other variables, the extra source of model uncertainty in ENS20 and ENS20^{phys} “balances out” the impacts of higher resolution.

It is suspected that, for some of the low-level fields, systematic model biases associated with certain parameterization schemes and combinations of parameterization schemes are impacting the growth rates. These biases are important to consider within the context of an ensemble because systematic biases that increase forecast uncertainty do so “artificially” since the associated errors are not uncertain (Eckel and Mass 2005). Thus, as discussed in recent studies (e.g., Eckel and Mass 2005; Yuan et al. 2007; Hamill and Whitaker 2007; Yussouf and Stensrud 2008), calibration should be performed on raw ensemble output to achieve maximum ensemble utility. Because the small number of cases examined in this study makes it difficult to obtain a useful “training period,” a calibration that could be applied in a real-time forecasting environment is not attempted. The possible influence of bias associated with particular CPs on MUCAPE mean spread growth rates is illustrated by a time series of domain-averaged MUCAPE values for all ENS4 and ENS20 ensemble members (Fig. 6). At forecast hour 33, which is the time used as Var_f in the growth rate calculation [Eq. (1)], ENS20_{GD} members (11–15) tend to have the largest biases in MUCAPE, followed by ENS20_{KF} (6–10) and ENS20_{BMJ} (1–5) members. These biases are consistent with the MUCAPE mean spread growth rates in Fig. 5 (i.e., larger biases in MUCAPE at forecast hour 33 inflate the spread and corresponding growth rates). The impacts of bias on the variance growth rates will be explored in section 3d.

Some other interesting features are revealed from the MUCAPE time series. First, for both the ENS4 and ENS20 ensembles, all members that use the MYJ PBL scheme tend to have much higher MUCAPE values than members using YSU, especially for the second diurnal peak within the forecast period. In addition, the ENS4 members that use the Thompson microphysics

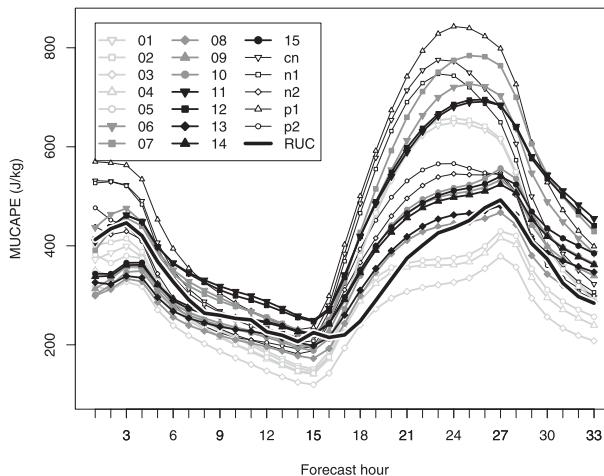


FIG. 6. Time series of mean domain-averaged MUCAPE for the ENS20 ensemble members and RUC analyses. A legend is provided in the upper-right portion of the plot with numbers indicating ENS20 ensemble members corresponding to those listed in Table 3.

with MYJ have higher MUCAPE than members using the WSM-6 or Ferrier schemes with MYJ. However, in the ENS20 ensemble, members that use WSM-6 with MYJ have slightly higher MUCAPE values than members using Thompson. Also, the ENS20 MYJ members tend to have their peak MUCAPE occur about 1–3 h earlier than in the RUC analysis, but all YSU members have the peak occurring at the same time as in the RUC analysis. For the ENS20 ensemble, both MYJ and YSU members have their peak MUCAPE occurring about 3–4 h earlier than in the RUC analysis. The sensitivity of MUCAPE to the different PBL schemes very likely results from systematic temperature and moisture biases associated with each scheme. For example, it has been well documented that the YSU method tends to form boundary layers that are too deep, warm, and dry, while the MYJ approach has a tendency for relatively shallow, cool, and moist boundary layers (e.g., Kain et al. 2005; Weisman et al. 2008).

3) MIXED-PHYSICS ENSEMBLE VARIANCE CONTRIBUTION

To estimate the percent contribution of mixed physics to spread in the IC/LBC+Phys ensembles, the ratio of the mean ensemble variance in the Phys ensembles to that of the corresponding IC/LBC+Phys ensembles {i.e., $[\text{Var}(\text{Phys})/\text{Var}(\text{IC/LBC+Phys})] \times 100\%$ } is computed for all 12 fields at forecast hours 9 and 33 (Fig. 7). Similar to the spread growth rate comparisons, statistical significance tests were performed using Welch’s *t* test ($\alpha = 0.05$) for differences in mixed-physics variance contributions between ENS4 and ENS20. Note the actual contributions to the ensemble spread in the IC/LBC+Phys

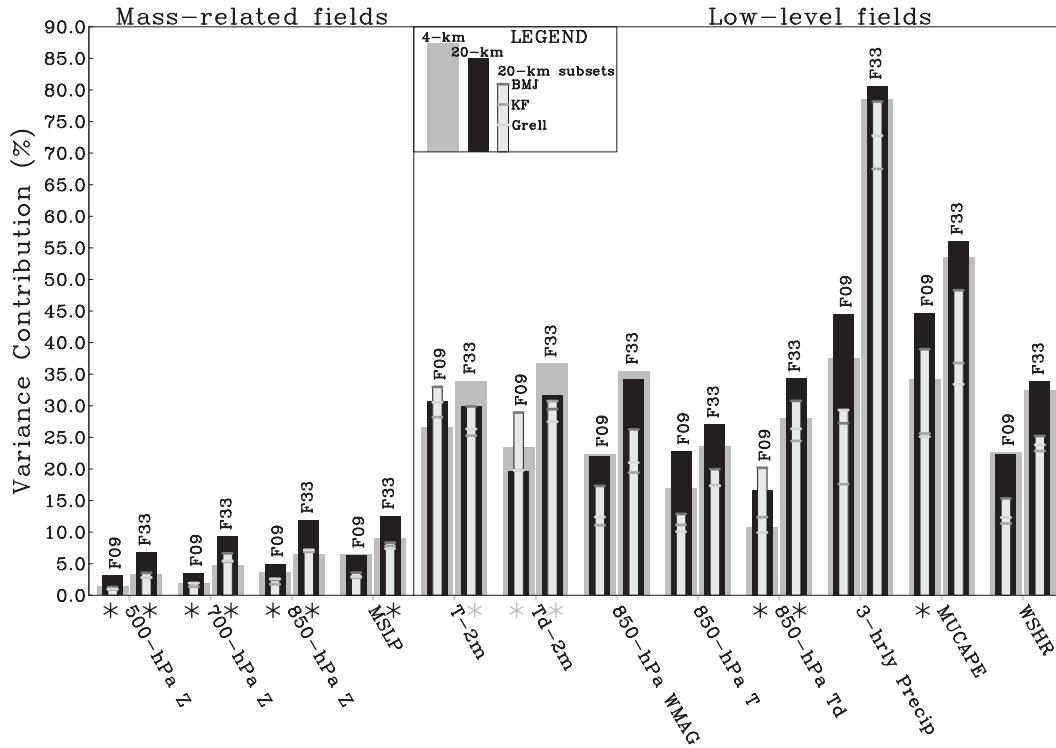


FIG. 7. Mean variance ratio [%; $\text{Var}(\text{ENS4}^{\text{phys}})/\text{Var}(\text{ENS4})$ and $\text{Var}(\text{ENS20}^{\text{phys}})/\text{Var}(\text{ENS20})$] from the 4- and 20-km grid-spacing ensembles and from the five-member 20-km ensemble subsets for the fields in Fig. 3. The histograms to the left (right) for each field are for forecast hour 9 (33). Asterisks below the histogram sets indicate statistically significant differences between ENS4 and ENS20 [black (gray) asterisks indicate ENS4 (ENS20) had larger variance contributions]. A legend is provided at the top of the figure.

ensembles not only result from separate contributions from IC/LBC perturbations and mixed physics, but also from an interaction term (which could be positive or negative) between the two error sources. Because ensembles using only IC/LBC perturbations were not used in this experiment, it is not possible to diagnose this interaction term, and the estimate of spread contribution from mixed physics assumes the interaction term is negligible.

For mass-related fields, the mixed-physics variance contributions in ENS20 are larger than in ENS4 with differences that are statistically significant, consistent with the additional model uncertainty in ENS20 from varied CPs. The ENS20_{cp} subsets generally have contributions similar to those in ENS4. The mixed-physics contributions to ensemble spread for mass-related fields decrease as higher atmospheric levels are examined, which likely occurs because the higher levels are impacted less by the boundary layer where the boundary layer physics have the greatest impact (Fig. 7). Relative to the low-level fields, mixed-physics contributions to ensemble spread are generally much smaller for the mass-related fields, which is consistent with earlier

discussed results (e.g., Figs. 3 and 5). Perhaps the most noticeable feature for the mass-related fields in Fig. 7 is that the mixed-physics variance contributions for all ensemble subsets are higher at forecast hour 33 than 9, implying that the influence of model uncertainty on ensemble spread increases within the forecast period analyzed.

For the low-level fields, similar to the variance growth rates (Fig. 5), there is much more variability in the variance contributions among the different fields examined. Mixed-physics variance contributions range from around 10% for 850Td in ENS4 at forecast hour 9 to around 85% for PREC in ENS4 and ENS20 at forecast hour 33. Furthermore, similar to the mass-related fields, the ensemble subsets for most of the low-level fields have increasing mixed-physics variance contributions with increasing forecast lead time. By far, the highest mixed-physics variance contributions occur with PREC, which is not surprising because, overall, the physics parameterizations are particularly active in association with precipitation and two of the parameterizations (cumulus and microphysics schemes) are directly associated with precipitation production.

4) IMPACTS OF MODEL BIASES ON MEAN VARIANCE GROWTH RATES

To explore the impacts that biases in different fields have on the mean spread growth rates, systematic and nonsystematic biases are removed from each ensemble member for each field at all forecast hours using a procedure based on probability matching (Ebert 2001) described by Clark et al. (2009). The procedure uses probability matching to reassign the distribution of a forecast field with that of the observed field (RUC or stage IV analysis), so that the modified forecast fields have the same spatial patterns as the original forecasts, but have values adjusted so the distribution of their amplitudes exactly matches that of the analyses resulting in zero bias. Thus, the adjusted ensemble member forecasts all have the same distribution of values in the analyses, but with different spatial patterns. Ensemble variances computed from the adjusted forecasts can be interpreted as the variance resulting solely from the placement of “features.”

The mean variance growth rates for “bias corrected” fields along with the differences between the bias-corrected growth rates and raw growth rates are displayed in Fig. 8. For the mass-related fields, the growth rates for the IC/LBC+Phys ensembles are noticeably less relative to the raw growth rates, as shown by the growth rate differences (Fig. 8b). In addition, the differences between the ENS4 and ENS20 bias-corrected growth rates are much less than those for the raw growth rates and are no longer significant, which implies that the higher spread in raw ENS4 mass-related fields relative to ENS20 (Fig. 5) can be attributed to larger differences in forecasts of the amplitude of features as opposed to their placement.

In general, the variance growth rates for the low-level fields change less than for the mass-related fields after the bias-correction procedure is applied (Fig. 8b), which may be related to differences in the evolution of amplitude errors with increasing forecast lead time. For the mass-related fields, IC/LBC perturbations mainly contribute to amplitude errors, which become larger as forecast lead time increases. Thus, applying the bias-correction procedure, which eliminates all amplitude errors, will decrease the ensemble variance more at the later lead times resulting in slower spread growth rates. However, for the low-level fields, the physics perturbations can quickly create large differences in the amplitude of features. However, these differences can quickly saturate (e.g., Stensrud et al. 2000) so that at later forecast lead times the amplitude errors are similar to those at earlier times. Thus, eliminating all amplitude errors will affect the ensemble variance similarly at all forecast lead

times and variance growth rates will not be strongly impacted. The variance growth rates for low-level fields that are strongly impacted by bias correction are likely those that have amplitude errors (bias) that *change* with forecast lead time. For example, the domain-averaged time series of T2 (Fig. 4) implies that T2 biases are similar for corresponding periods within the diurnal cycles for the ENS4 and ENS20 ensemble members. Consistent with these similar biases, the growth rates for bias-corrected T2 are not very different from the raw T2 growth rates. However, domain-averaged time series of MUCAPE (Fig. 6) show that MUCAPE biases during the latter part of the forecast (hours 24–33) are very different than those from the first part of the forecast (hours 0–9), and, consistent with the different biases, some of the growth rates for bias-corrected MUCAPE change dramatically relative to the raw MUCAPE growth rates (Fig. 8b).

To more clearly show how the variance growth rates are affected by the bias-correction procedure, time series of mean differences between raw and bias-corrected ensemble variances for MSLP, T2, and MUCAPE are shown in Fig. 9. As discussed above, for MSLP, the bias correction reduces the ensemble variance more at later forecast lead times (Figs. 9a and 9b). Trends are not as noticeable for the other low-level fields. However, it is clear from Fig. 9f that the impacts of the bias correction are dependent on the CP used; for ENS20_{GD} (ENS20_{BMJ}), the bias-correction procedure results in more (fewer) positive ensemble variances at later forecast lead times. Furthermore, ENS20_{BMJ} has larger mean ensemble variances than ENS4 and ENS20 as well as the other ENS20_{cp} subensembles, which should be investigated further in future work.

Other notable features from the bias-corrected growth rates in Fig. 8 are that ENS4 almost always has higher growth rates than ENS20, which is expected because smaller scales of motion (which have faster growth rates) are resolved in ENS4. However, for many of the variables (e.g., PREC, MUCAPE, and T2), ENS20_{cp} subensembles actually have larger mean spread growth rates than ENS20. This is unexpected behavior because one less source of model uncertainty in ENS20_{cp} should result in slower spread growth. Further analyses, which are beyond the scope of this paper, are needed to explain this unexpected behavior.

b. Spread–skill relationship

Ideally, in a skillful ensemble that accurately accounts for all sources of forecast uncertainty, the ensemble variance should be a reliable predictor of the forecast skill (e.g., Grit and Mass 2007). To quantify the variance–MSE relationship, past works have used linear

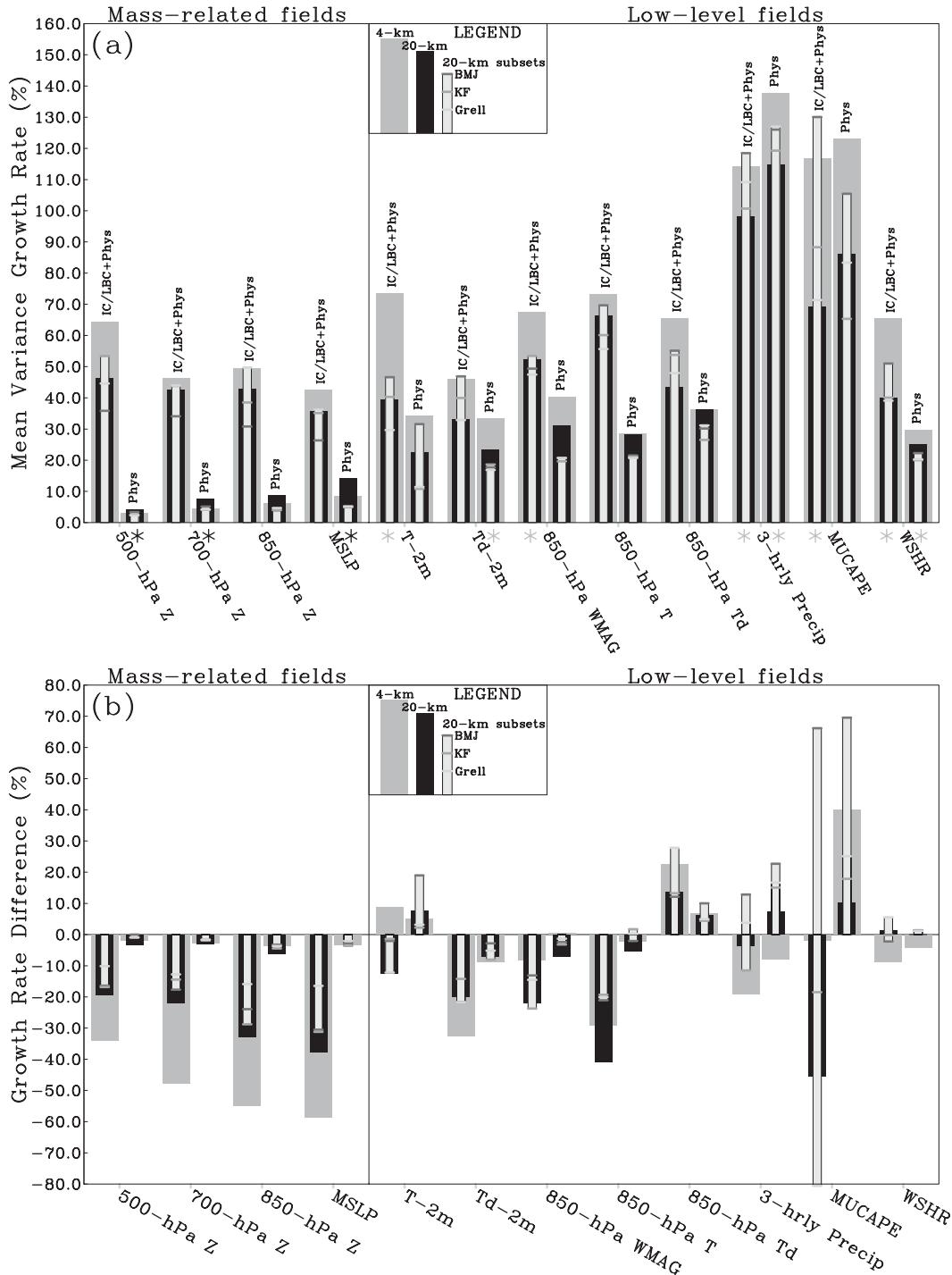


FIG. 8. (a) As in Fig. 5, but for bias-corrected forecasts. (b) Differences between the bias-corrected and raw growth rates (bias-corrected minus raw).

correlations (e.g., Jones et al. 2007). This study also employs variance–MSE linear correlations, but care should be taken when interpreting the correlation coefficients because, as shown by Grimit and Mass (2007), error statistics tend to exhibit increasing variance with

increasing ensemble spread so that the variance–MSE relationship cannot be assumed to be linear. Thus, as noted in a similar analysis conducted by Jones et al. (2007), the linear correlation coefficients only provide an estimate of the predictability of ensemble skill.

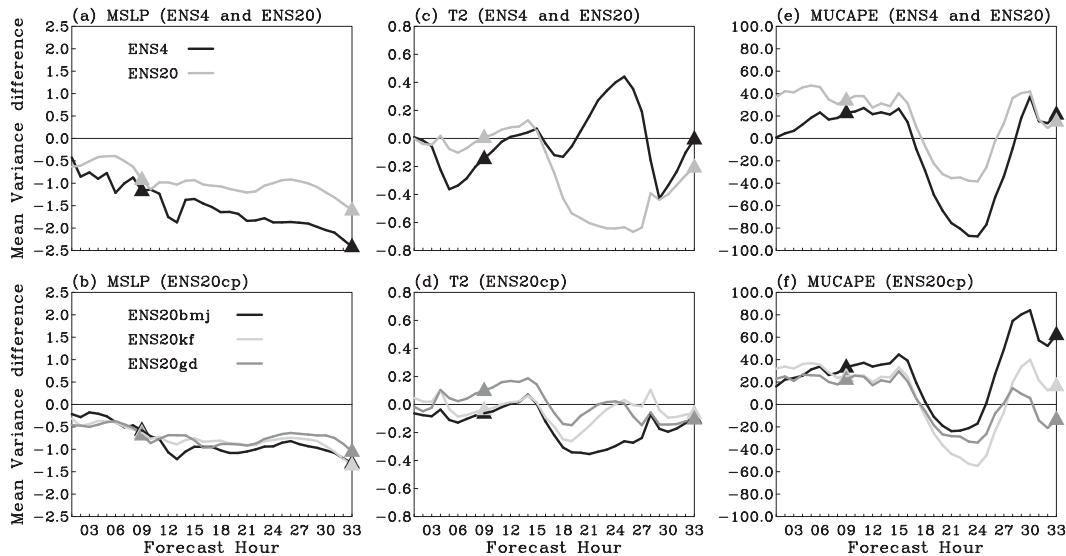


FIG. 9. Differences in mean ensemble variance between the raw and bias-corrected forecasts from ENS4 and ENS20 for the following fields: (a) MSLP (hPa^2), (c) T2 (K^2), (e) MUCAPE [J kg^{-1}]². (b),(d),(f) As in (a),(c),(e) but for the ENS20_{cp} configurations. The triangles denote forecast hours 9 and 33, which are the times used in the computation of variance growth rates.

Statistical consistency describes how well the ensemble variance matches the MSE when averaged over many cases (Talagrand et al. 1999; Eckel and Mass 2005). Thus, unlike the variance–MSE relationship, the amount of correlation is not considered with statistical consistency. A statistical consistency analysis can also provide information on whether an ensemble system is over- or underdispersive. In an underdispersive (overdispersive) ensemble the average MSE is larger (smaller) than the ensemble variance. In this study, the variance–MSE and statistical consistency analyses are used as simple methods for inferring the impacts of the different spread growth rates on the quality of the ensemble forecasts. Note that we only examine spread–skill metrics for the ensembles that have both IC/LBC perturbations and mixed physics because the lack of IC/LBC perturbations in ENS4^{phys} and ENS20^{phys} degrades the statistical consistency for all fields (not shown).

To illustrate the variance–MSE correlation in the ENS4 and ENS20 ensembles for the different fields examined, scatterplots of ensemble variance versus MSE are displayed in Fig. 10. Each panel in Fig. 10 contains variance–MSE points for each case and for each forecast hour (20 cases \times 33 times = 660 points for each panel) and correlation coefficients indicate the degree of correspondence between the ensemble variance and MSE (i.e., the reliability of ensemble variance as a predictor of forecast skill). Rank histograms (e.g., Hamill 2001) provided in Fig. 10 valid at forecast lead times of 9 and 33 h also provide information regarding the representation

of forecast uncertainty: flat rank histograms imply an accurate depiction of forecast uncertainty, U-shaped (n shaped) rank histograms imply underdispersion (overdispersion), and right (left) skewness indicates a tendency for overprediction (underprediction). To allow for a more convenient comparison between ENS4 and ENS20, the 16 bins composing the ENS20 rank histograms were regrouped into 6 bins that each contain an equal portion of the original 16 bins.² This “regrouping” technique has also been used in Clark et al. (2009).

To illustrate the statistical consistency, the average MSE and ensemble variance at forecast hours 9 and 33 for fields in ENS4 and ENS20 are shown in Fig. 11. These forecast hours are chosen because they were the times used to compute the spread growth rates. To estimate whether the variance–MSE differences between ENS4 and ENS20 were statistically significant, a resampling procedure using a 95% confidence interval was applied (Wilks 1995, 145–150). Differences between average randomly resampled variance and MSE for two

² The formulas for regrouped rank histogram bins can be expressed as follows: ENS20_{regroup}(1) = ENS20(1) + ENS20(2) + ENS20(3) \times $\frac{1}{3}$, ENS20_{regroup}(2) = ENS20(3) \times $\frac{1}{3}$ + ENS20(4) + ENS20(5) + ENS20(6) \times $\frac{1}{3}$, ENS20_{regroup}(3) = ENS20(6) \times $\frac{2}{3}$ + ENS20(7) + ENS20(8), ENS20_{regroup}(4) = ENS20(9) + ENS20(10) + ENS20(11) \times $\frac{2}{3}$, ENS20_{regroup}(5) = ENS20(11) \times $\frac{1}{3}$ + ENS20(12) + ENS20(13) + ENS20(14) \times $\frac{1}{3}$, and ENS20_{regroup}(6) = ENS20(14) \times $\frac{2}{3}$ + ENS20(15) + ENS20(16), where ENS20_{regroup}(x) is the value for regrouped rank histogram bins $x = 1, 2, \dots, 6$, and ENS20(x) is the value for raw rank histogram bins $x = 1, 2, \dots, 16$.

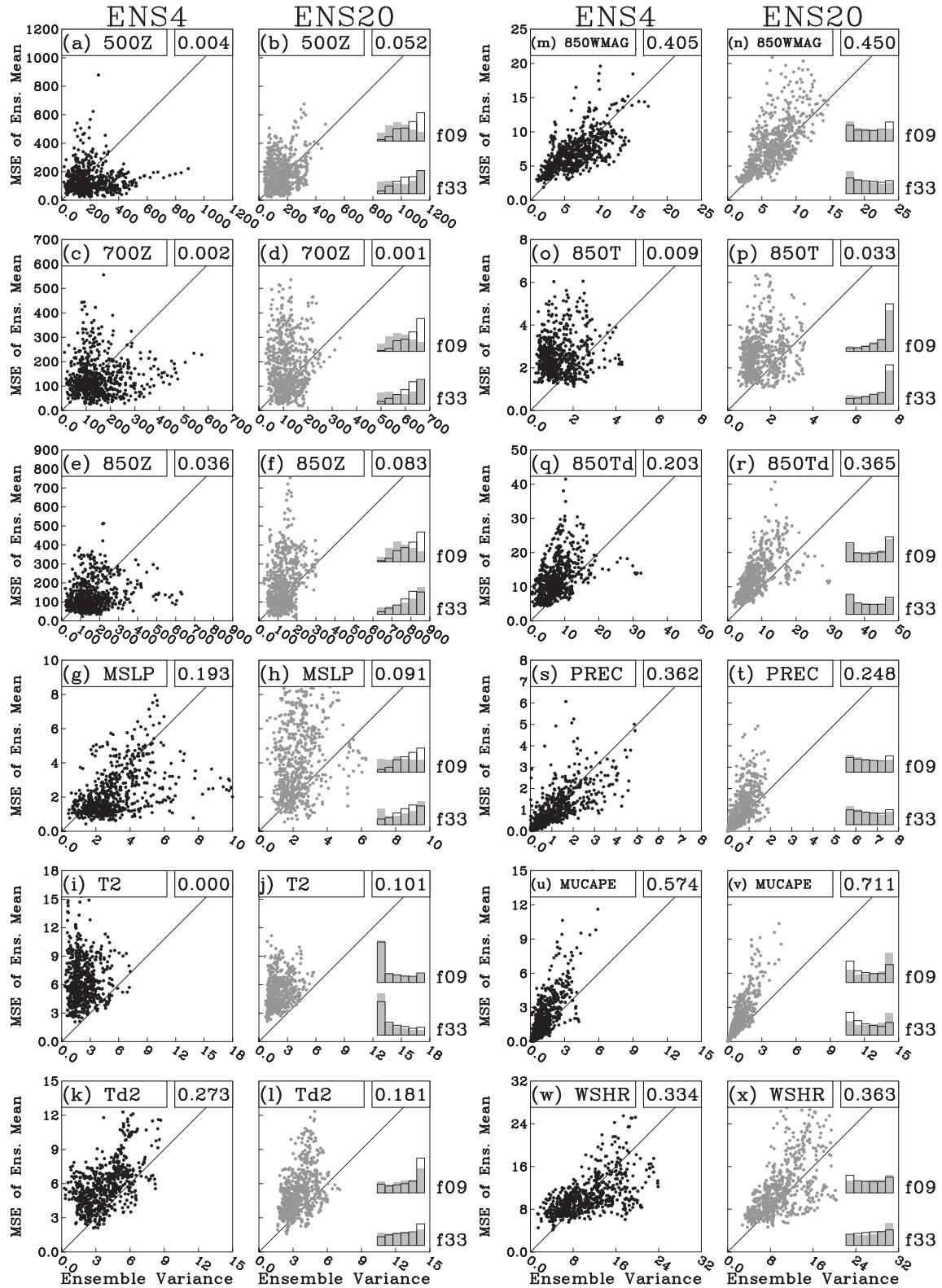


FIG. 10. Scatterplots of ensemble variance vs MSE of the ensemble mean from ENS4 (black dots) for the raw fields: (a) 500Z, (c) 700Z, (e) 850Z, (g) MSLP, (i) T2, (k) Td2, (m) 850WMAG, (o) 850T, (q) 850Td, (s) PREC, (u) MUCAPE, (w) WSHR. (b),(d),(f),(h),(j),(l),(n),(p),(r),(t),(v),(x) As in (a),(c),(e),(g),(i),(k),(m),(o),(q),(s),(u),(w) but for ENS20 (gray dots). Correlation coefficients are provided in the top-right corner of each panel and rank histograms for ENS4 (black outline) and ENS20 (gray shaded) for forecast hours 9 and 33 are displayed in the bottom-right corner of the ENS20 plots.

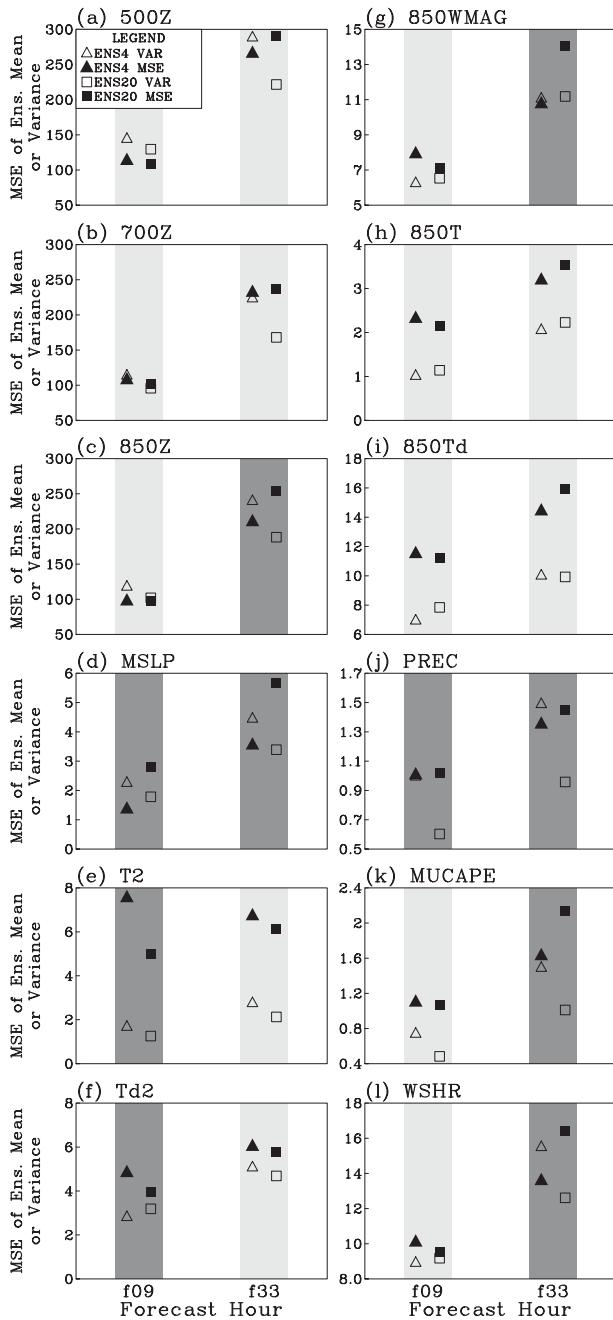


FIG. 11. Average MSEs of the ensemble means and ensemble variances at forecast hours 9 and 33 for the ENS4 and ENS20 ensembles for the fields following fields (a) 500Z (m^2), (b) 700Z (m^2), (c) 850Z (m^2), (d) MSLP (hPa^2), (e) T2 (K^2), (f) Td2 (K^2), (g) 850WMAG [$(m s^{-1})^2$], (h) 850T (K^2), (i) 850Td (K^2), (j) PREC (mm^2), (k) MUCAPE [$J kg^{-1} m^2$], (l) WSHR [$(m s^{-1})^2$]. The times at which differences between the ensemble variance and MSE differences in ENS4 and ENS20 are statistically significant are highlighted with darker shades of gray.

artificial datasets were computed 10 000 times, and the distribution of the differences between the two artificial datasets was used to determine whether the “true” statistic fell outside of the 95% confidence interval, which would suggest statistical significance.

For mass-related fields in Figs. 10a–h, the variance–MSE correlations in ENS4 and ENS20 are very low suggesting that the ensemble variance is not a reliable indicator of forecast skill for these fields. The highest correlations occur for the MSLP forecasts from ENS4 ($R^2 = 0.19$). Considering previous work that has also found small spread–error correlations for fields like midtropospheric geopotential height (e.g., Buizza 1997), these results are not surprising. However, there are noticeable differences in the distribution of variance–MSE points for the mass-related fields: in ENS4 there are more points to the right of the diagonal than in ENS20, indicating that ensemble variance is greater than MSE more frequently in ENS4. Furthermore, the variance–MSE points in ENS20 appear to be positioned in a vertically oriented “plume,” while those in ENS4 veer toward the right (i.e., toward higher values of ensemble variance).

These results for mass-related fields are reflected by the statistical consistency analyses in Figs. 11a–d. At forecast hour 33, differences between ensemble variance and MSE in ENS4 are noticeably less than in ENS20 indicating that ENS4 is more statistically consistent than ENS20 (although, the differences were only found to be significant for 850Z and MSLP). The better statistical consistency in ENS4 results not only from greater spread at forecast hour 33, but also from lower MSEs relative to ENS20 (Figs. 11a–d). Furthermore, while ENS4 forecasts for mass-related fields do not exhibit much change in statistical consistency between forecast hours 9 and 33, ENS20 forecasts become increasingly underdispersive. The increasing underdispersion in ENS20 is also indicated by the change to more “u shaped” rank histograms (Figs. 10b, 10d, 10f, and 10h; gray shaded) from forecast hours 9 to 33. The ENS4 rank histograms (Figs. 10b, 10d, 10f, and 10h; black outline) do not exhibit a noticeable pattern of change in shape and are right skewed at both times, implying underprediction of mass-related fields. This underprediction could possibly result from a general cool bias in the lower part of the troposphere, but further examination is beyond the scope of this study. For mass-related fields, the statistical consistency results imply that the faster spread growth rates in ENS4 are contributing to more reliable forecasts than in ENS20.

The variance–MSE correlations for low-level fields in ENS4 and ENS20 (Figs. 10i–x) are quite variable among the fields examined, with 850WMAG (Figs. 10m and 10n), PREC (Figs. 10s and 10t), MUCAPE (Figs. 10u and 10v), and WSHR (Figs. 10w and 10x) having the

highest values. Furthermore, the rank histograms indicate that ENS4 and ENS20 both suffer from systematic biases and/or underdispersion for most of the low-level fields. For example, warm T2 biases and dry Td2 biases are revealed from the right- and left-skewed rank histograms, respectively, in Figs. 10i–l. Also, the U-shaped rank histograms for 850Td (Fig. 10q) imply underdispersion. The statistical consistency analyses (Figs. 11e–l) show that by forecast hour 33 there is little difference between ENS4 and ENS20 in terms of statistical consistency for the temperature and dewpoint fields (T2, Td2, 850T, and 850Td), while ENS4 has better statistical consistency for the other fields (850WMAG, PREC, MUCAPE, and WSHR, all of which have differences that are statistically significant). Similar to the mass-related fields, the more statistically consistent forecasts for low-level fields in ENS4 appear to result from a combination of lower MSEs and higher ensemble variances relative to ENS20.

4. Summary and discussion

This study compared ensemble spread growth and spread–error relationships for 12 different fields in a 4-km grid-spacing convection-allowing WRF model ensemble to that from a similarly configured but coarser 20-km grid-spacing convection-parameterizing WRF model ensemble. Ensemble subsets that used both IC/LBC perturbations and mixed physics were compared along with subsets that only contained mixed physics. In addition, the contribution of ensemble variance from the mixed physics in the 4- and 20-km ensembles was inferred by comparing the mixed-physics-only ensemble subsets to those that contained both IC/LBC perturbations and mixed physics. A total of 20 cases were examined for a domain centered over the central United States. Our main findings are summarized below. In addition, Table 5 highlights the main differences that were observed for the various ensemble comparisons.

Spread growth rates for mass-related fields were higher in ENS4 than in ENS20 by about 30%. The mixed-physics-only ensemble subsets (Phys) had much smaller spread growth rates than did ensemble subsets with both IC/LBC perturbations and mixed physics (IC/LBC+Phys). For low-level fields, spread growth rates were quite variable among the different fields examined; however, for all of the statistically significant differences between ENS4 and ENS20 (or ENS4^{phys} and ENS20^{phys}), the 4-km ensembles had the higher spread growth rates. The differences between spread growth rates in 4- and 20-km ensemble configurations are summarized in the first two columns of Table 5.

The contributions to spread from mixed physics in the mass-related fields were generally much smaller than

those for the low-level fields, consistent with the differences between the IC/LBC+Phys and Phys spread growth rates. In addition, the contribution to spread from mixed physics increased with increasing forecast lead time. Similar to the spread growth rates, there was much more variability in the spread contributions among the different low-level fields than for the mass-related fields. For most of the statistically significant differences between the ENS4 and ENS20 mixed-physics spread contributions, ENS20 had larger contributions (third and fourth columns of Table 5), which was related to ENS20 having an additional source of model uncertainty in the form of varied cumulus parameterizations.

It was suspected from examination of time series for domain-averaged low-level fields that systematic model biases may be having an impact on the mean ensemble spread, as found by Clark et al. (2009). To explore these potential impacts, biases in all forecast fields were removed by replacing the distributions of values in the forecast fields with the distribution of values in the corresponding RUC or stage IV analyses as described by Clark et al. (2009). It was found that the spread growth rates for the mass-related fields tended to be impacted mostly by the bias-correction procedure, which may be related to differences in how amplitude errors evolve with increasing forecast lead time. The MUCAPE fields in the ENS20_{BMJ} and ENS20_{GD} ensemble subsets were found to be the low-level fields most influenced by the bias-correction procedure because of different systematic biases at later forecast lead times relative to earlier ones.

Finally, to put the spread growth rates for the different fields examined into an appropriate forecasting context, an analysis of the variance–MSE relationship and statistical consistency was conducted. This analysis was important because increased ensemble dispersion does not necessarily imply a better spread–skill relationship. Variance–MSE correlation coefficients indicated that, in general, ensemble variance was not a reliable indicator of forecast uncertainty. Furthermore, at forecast hour 33, ENS4 had better statistical consistency than ENS20 for mass-related fields (850WMAG, PREC, MUCAPE, and WSHR); however, for temperature and dewpoint fields (T2, Td2, 850T, and 850Td), there were no noticeable differences (fifth column of Table 5). It was found that a combination of higher spread and lower MSEs contributed to the improved statistical consistency in ENS4.

Generally, the results from this study could be interpreted as being encouraging for future convection-allowing ensemble systems simply because of the improved statistical consistency for many fields. However, additional

TABLE 5. Summary of results. For each variable, the different columns indicate whether faster spread growth, larger mixed-physics spread contributions, or better statistical consistency was observed for the 4- or 20-km grid-spacing ensemble configurations. Only statistically significant differences are indicated.

	Faster spread growth		Larger mixed-physics spread contribution		Better statistical consistency at forecast hour 33
	IC/LBC+Phys	Phys	Forecast hour 9	Forecast hour 33	
500Z	4 km	20 km	20 km	20 km	—
700Z	4 km	20 km	20 km	20 km	—
850Z	4 km	20 km	20 km	20 km	4 km
MSLP	4 km	20 km	—	20 km	4 km
T2	—	4 km	—	4 km	—
Td2	4 km	4 km	4 km	4 km	—
850WMAG	—	—	—	—	4 km
850T	—	—	—	—	—
850Td	—	—	20 km	20 km	—
PREC	4 km	4 km	—	—	4 km
MUCAPE	—	—	20 km	—	4 km
WSHR	4 km	4 km	—	—	4 km

work is needed to diagnose why the higher resolution of ENS4 did not seem to improve the spread–error metrics for temperature and dewpoint fields. Perhaps the parameterization schemes that the temperature and dewpoint fields depend on are simply not very sensitive to grid spacing, or perhaps a reduction of the large systematic errors in the temperature and dewpoint forecasts would result in better statistical consistency. Further work should also analyze larger sets of cases for different periods and regions. Because this study focuses on the spring season in the central United States, a time period and region characterized by frequent convective systems, examination of other times/regions in which strong convection is not as prevalent may give different results. Furthermore, it would be very useful to assess how much the improved statistical consistency in some fields increases the skill of probabilistic forecasts. Finally, the behavior of the ensemble spread observed in this study should be helpful for designing future ensembles. In particular, the variety of results for different fields suggests that future ensemble design should give careful consideration to the specific types of forecast fields desired by the user. Also, the recognition of systematic model biases should provide motivation for improving the physics parameterizations used with convection-allowing grid spacing.

Acknowledgments. The authors thank Huiling Yuan at the Global Systems Division of the Earth System Research Laboratory (ESRL GSD) for assistance in obtaining SREF data in post–real time. The majority of this particular research was funded by NSF Grant ATM-0537043. Additional support was also provided through a National Research Council Post-doctoral Award for the first author under the guidance of David

Stensrud. The ENS20 simulations were conducted on the 64-processor computing cluster in the meteorology program at ISU. The CAPS real-time 4-km ensemble forecasts were primarily supported by the NOAA CSTAR program. Supplementary support was provided by NSF ITR project LEAD (ATM-0331594). Drs. Kelvin K. Droegemeier, Keith Brewster, John Kain, Steve Weiss, David Bright, Matt Wandishin, Mike Coniglio, Jun Du, Jimmy Dudhia, Morris Weisman, Greg Thompson, and Wei Wang contributed to the ensemble system design and WRF model configuration. Kevin Thomas carried out the real-time runs. The CAPS real-time predictions were performed at the Pittsburgh Supercomputing Center (PSC), which is supported by NSF.

REFERENCES

- Accadia, C., S. Mariani, M. Casaioli, A. Lavagnini, and A. Speranza, 2003: Sensitivity of precipitation forecast skill scores to bilinear interpolation and a simple nearest-neighbor average method on high-resolution verification grids. *Wea. Forecasting*, **18**, 918–932.
- Baldwin, M. E., and K. E. Mitchell, 1997: The NCEP hourly multisensor U.S. precipitation analysis for operations and GCIP research. Preprints, *13th Conf. on Hydrology*, Long Beach, CA, Amer. Meteor. Soc., 54–55.
- Benjamin, S. G., G. A. Grell, J. M. Brown, T. G. Smirnova, and R. Bleck, 2004a: Mesoscale weather prediction with the RUC hybrid isentropic–terrain-following coordinate model. *Mon. Wea. Rev.*, **132**, 473–494.
- , and Coauthors, 2004b: An hourly assimilation–forecast cycle: The RUC. *Mon. Wea. Rev.*, **132**, 495–518.
- Betts, A. K., 1986: A new convective adjustment scheme. Part I: Observational and theoretical basis. *Quart. J. Roy. Meteor. Soc.*, **112**, 677–691.
- , and M. J. Miller, 1986: A new convective adjustment scheme. Part II: Single column tests using GATE wave, BOMEX,

- ATEX and Arctic air-mass data sets. *Quart. J. Roy. Meteor. Soc.*, **112**, 693–709.
- Buizza, R., 1997: Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system. *Mon. Wea. Rev.*, **125**, 99–119.
- Chou, M.-D., and M. J. Suarez, 1994: An efficient thermal infrared radiation parameterization for use in general circulation models. NASA Tech. Memo. 104606, Vol. 3, 85 pp.
- Clark, A. J., W. A. Gallus, and T. C. Chen, 2008: Contributions of mixed physics versus perturbed initial/lateral boundary conditions to ensemble-based precipitation forecast skill. *Mon. Wea. Rev.*, **136**, 2140–2156.
- , 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.
- Coniglio, M., K. L. Elmore, J. S. Kain, S. J. Weiss, M. Xue, and M. Weisman, 2010: Evaluation of WRF model output for severe weather forecasting from the 2008 NOAA Hazardous Weather Testbed Spring Experiment. *Wea. Forecasting*, **25**, 408–427.
- Du, J., and Coauthors, 2004: The NOAA/NWS/NCEP Short Range Ensemble Forecast (SREF) system: Evaluation of an initial condition vs. multiple model physics ensemble approach. Preprints, *20th Conf. on Weather Analysis and Forecasting/16th Conf. on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., 21.3. [Available online at <http://ams.confex.com/ams/pdfpapers/71107.pdf>.]
- Dyer, A. J., and B. B. Hicks, 1970: Flux-gradient relationships in the constant flux layer. *Quart. J. Roy. Meteor. Soc.*, **96**, 715–721.
- Ebert, E. E., 2001: Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon. Wea. Rev.*, **129**, 2461–2480.
- Eckel, F. A., and C. F. Mass, 2005: Aspects of effective mesoscale, short-range ensemble forecasting. *Wea. Forecasting*, **20**, 328–350.
- Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah Land Surface Model advances in the National Centers for Environmental Prediction operational mesoscale Eta Model. *J. Geophys. Res.*, **108**, 8851, doi:10.1029/2002JD003296.
- Ferrier, B. S., Y. Jin, Y. Lin, T. Black, E. Rogers, and G. DiMego, 2002: Implementation of a new grid-scale cloud and rainfall scheme in the NCEP Eta Model. Preprints, *15th Conf. on Numerical Weather Prediction*, San Antonio, TX, Amer. Meteor. Soc., 280–283.
- Fritsch, J. M., and R. E. Carbone, 2004: Improving quantitative precipitation forecasts in the warm season: A USWRP research and development strategy. *Bull. Amer. Meteor. Soc.*, **85**, 955–965.
- Grell, G. A., and D. Devenyi, 2002: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophys. Res. Lett.*, **29**, 1693, doi:10.1029/2002GL015311.
- Grimit, E. P., and C. F. Mass, 2007: Measuring the ensemble spread–error relationship with a probabilistic approach: Stochastic ensemble results. *Mon. Wea. Rev.*, **135**, 203–221.
- Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550–560.
- , and J. S. Whitaker, 2007: Ensemble calibration of 500-hPa geopotential height and 850-hPa and 2-m temperatures using reforecasts. *Mon. Wea. Rev.*, **135**, 3273–3280.
- Hong, S.-Y., and J.-O. J. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Kor. Meteor. Soc.*, **42**, 129–151.
- Hou, D., E. Kalnay, and K. K. Droegemeier, 2001: Objective verification of the SAMEX '98 ensemble forecasts. *Mon. Wea. Rev.*, **129**, 73–91.
- Houtekamer, P. L., L. Lefaire, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225–1242.
- 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.
- , 1996: The surface layer in the NCEP Eta Model. Preprints, *11th Conf. on Numerical Weather Prediction*, Norfolk, VA, Amer. Meteor. Soc., 354–355.
- , 2002: Nonsingular implementation of the Mellor–Yamada level 2.5 scheme in the NCEP Meso model. NCEP Office Note 437, 61 pp.
- , 2003: A nonhydrostatic model based on a new approach. *Meteor. Atmos. Phys.*, **82**, 271–285.
- Jones, M. S., B. A. Colle, and J. S. Tongue, 2007: Evaluation of a mesoscale short-range ensemble forecast system over the northeast United States. *Wea. Forecasting*, **22**, 36–55.
- Kain, J. S., and J. M. Fritsch, 1993: Convective parameterization for mesoscale models: The Kain–Fritsch scheme. *The Representation of Cumulus Convection in Numerical Models*, Meteor. Monogr., No. 46, Amer. Meteor. Soc., 246 pp.
- , S. J. Weiss, M. E. Baldwin, G. W. Carbin, D. A. Bright, J. J. Levit, and J. A. Hart, 2005: Evaluating high-resolution configurations of the WRF model that are used to forecast severe convective weather: The 2005 SPC/NSSL Spring Program. Preprints, *21st Conf. on Weather Analysis and Forecasting/17th Conf. on Numerical Weather Prediction*, Washington, DC, Amer. Meteor. Soc., 2A.5. [Available online at <http://ams.confex.com/ams/pdfpapers/94843.pdf>.]
- Kong, F., and Coauthors, 2007: Preliminary analysis on the real-time storm-scale ensemble forecasts produced as a part of the NOAA Hazardous Weather Testbed 2007 Spring Experiment. Preprints, *22nd Conf. on Weather Analysis and Forecasting/18th Conf. on Numerical Weather Prediction*, Park City, UT, Amer. Meteor. Soc., 3B2. [Available online at <http://ams.confex.com/ams/pdfpapers/124667.pdf>.]
- Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307.
- Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys.*, **20**, 851–875.
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the long-wave. *J. Geophys. Res.*, **102** (D14), 16 663–16 682.
- Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF Ensemble Prediction System: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119.
- Monin, A. S., and A. M. Obukhov, 1954: Basic laws of turbulent mixing in the surface layer of the atmosphere (in Russian). *Contrib. Geophys. Inst. Acad. Sci. USSR*, **151**, 163–187.
- 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.
- Nutter, P., D. Stensrud, and M. Xue, 2004: Effects of coarsely resolved and temporally interpolated lateral boundary conditions on the dispersion of limited-area ensemble forecasts. *Mon. Wea. Rev.*, **132**, 2358–2377.

- Palmer, F., R. Molteni, R. Mureau, P. Buizza, P. Chapelet, and J. Tribbia, 1992: Ensemble prediction. ECMWF Research Dept. Tech. Memo. 188, 45 pp.
- Paulson, C. A., 1970: The mathematical representation of wind speed and temperature profiles in the unstable atmospheric surface layer. *J. Appl. Meteor.*, **9**, 857–861.
- R Development Core Team, cited 2009: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [Available online at <http://www.R-project.org>.]
- Skamarock, W. C., and M. L. Weisman, 2009: The impact of positive-definite moisture transport on NWP precipitation forecasts. *Mon. Wea. Rev.*, **137**, 488–494.
- , J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, 2005: A description of the Advanced Research WRF version 2. NCAR Tech Note NCAR/TN-468+STR, 88 pp. [Available from UCAR Communications, P.O. Box 3000, Boulder, CO, 80307; also online at http://box.mmm.ucar.edu/wrf/users/docs/arw_v2.pdf.]
- Smagorinsky, J., 1969: Problems and promises of deterministic extended range forecasting. *Bull. Amer. Meteor. Soc.*, **50**, 286–311.
- Stensrud, D. J., J. 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.
- Talagrand, O., R. Vautard, and B. Strauss, 1999: Evaluation of probabilistic prediction systems. *Proc. Workshop on Predictability*, Reading, United Kingdom, European Centre for Medium-Range Weather Forecasts, 1–25.
- Thompson, G., R. M. Rasmussen, and K. Manning, 2004: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part I: Description and sensitivity analysis. *Mon. Wea. Rev.*, **132**, 519–542.
- Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319.
- Wandishin, M. S., S. L. Mullen, D. J. Stensrud, and H. E. Brooks, 2001: Evaluation of a short-range multi-model ensemble system. *Mon. Wea. Rev.*, **129**, 729–747.
- Webb, E. K., 1970: Profile relationships: The log-linear range and extension to strong stability. *Quart. J. Roy. Meteor. Soc.*, **96**, 67–90.
- 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.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction*. Academic Press, 467 pp.
- Xue, M., and Coauthors, 2007: CAPS realtime 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*, Park City, UT, Amer. Meteor. Soc., 3B1. [Available online at <http://ams.confex.com/ams/pdfpapers/124587.pdf>.]
- Yuan, H., X. Gao, S. L. Mullen, S. Sorooshian, J. Du, and H. M. H. Juang, 2007: Calibration of probabilistic quantitative precipitation forecasts with an artificial neural network. *Wea. Forecasting*, **22**, 1287–1303.
- Yussouf, N., and D. J. Stensrud, 2008: Reliable probabilistic quantitative precipitation forecasts from a short-range ensemble forecasting system during the 2005/06 cool season. *Mon. Wea. Rev.*, **136**, 2157–2172.