JP3.1 GROWTH OF SPREAD IN CONVECTION-ALLOWING AND CONVECTION-PARAMETERIZING ENSEMBLES

Adam J. Clark*[1], William A. Gallus, Jr.[1], Ming Xue[2,3], and Fanyou Kong[3]

1 Dept. of Geological and Atmospheric Sciences, Iowa State University, Ames, IA.
2 School of Meteorology, and 3 Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, OK

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; Toth and Kalnay 1997; Palmer et al. 1992; Molteni et al. 1996), 2) using different combinations of physical parameterizations (mixed-physics; Houtekamer et al. 1996; Stensrud et al. 2000; Du et al. 2004; Jones et al. 2007), and 3) using different numerical models (e.g., Wandishin et al. 2001; Du et al. 2004; Eckel and Mass 2005). In current SREF systems, model (as opposed to analysis) errors are most difficult to account for. Thus, sensible parameters influenced by small-scale processes that must be parameterized such as low-level temperature/moisture and convective precipitation, are associated with notably underdispersive forecasts (Fritsch and Carbone 2004; Eckel and Mass 2005). The error growth for these sensible parameters contains a much larger contribution from model relative to IC errors than synoptic-scale parameters (e.g., 500-hPa geopotential heights and winds, mean-sea-level pressure; Stensrud et al. 2000; Eckel and Mass 2005), and the underdispersion may be a result of several deficiencies including 1) inadequate methods to account for model error, 2) inability to capture small-scale variability because of insufficient resolution (Eckel and Mass 2005), and 3) coarsely resolved and temporally interpolated lateral boundary conditions (LBCs; Nutter et al. 2003).

One method commonly used to gain information on ensemble spread is to isolate the error sources by using different perturbation strategies for a sample 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 NOAA/Hazardous Weather Testbed (HWT) Spring Experiment (SE07; Xue et al. 2007; Kong et al. 2007; Kain et al. 2008), 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 model errors because five members use the "perfect analysis" assumption.

The goal of this work is to use the 2007 SSEF system to compare ensemble spread contributions from model errors to contributions from a combination of model and analysis errors for various fields in a convection-allowing ensemble. In addition, ensemble spread growth and spread-error relationships associated with the two 5-member subsets of the SSEF system will be compared to two similarly configured subsets of a 20-km grid-spacing convection-parameterizing ensemble to examine the impacts of horizontal resolution for various forecast fields in the two different ensemble configurations.

2. DATA AND METHODOLOGY

The 2007 SSEF system was run during April-June 2007 and used the WRF-ARW (Version 2.2.0; Skamarock et al. 2005) model. 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 hours over an approximately 3000 x 2500 km domain covering much of the central United States (Fig. 1). For the SSEF control member, the 2100 UTC analyses from NCEP's operational North American Mesoscale (NAM; Janjic 2003) model (12-km grid-spacing) are used for ICs and the 1800 UTC NAM 12-km forecasts are used for LBCs. For the members with perturbed ICs, perturbations extracted from the 2100 UTC NAM analyses, and the corresponding SREF forecasts are used for LBCs (3-hr updates). Xue et al. (2007) and Kong et al. (2007) provide more details on the configurations.

For a comparison of the 5-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 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}). It should be noted that the ENS20

* *Corresponding author address:* Adam J. Clark, Iowa State Univ., Dept. of Geol. and Atm. Sciences, Ames, IA; email: clar0614@iastate.edu.

Table 1 ENS4 (pink) and ENS4phys (blue) ensemble member specifications. NAMa and NAMf indicate NAM forecasts and analyses, 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.

Member		ICs		Microphysics		Boundary Layer	
CN	PH1	21Z NAMa	21Z NAMa	WSM-6	Thompson	MYJ	MYJ
N1	PH2	CN-em_pert		Ferrier	Ferrier	MYJ	MYJ
P1	PH3	CN+em_pert		Thompson	WSM-6	MYJ	YSU
N2	PH4	CN-nmm_pert		Thompson	Thompson	YSU	YSU
P2	PH5	CN+nmm_pert		WSM-6	Ferrier	YSU	YSU

Table 2 ENS20 (pink) and ENS20phys (blue) specifications. The ICs/LBCs table elements represent various SREF members.

Member		ICs/LBCs		CP	Microphysics	Boundary Layer
1	16	em_ctl	eta_ctl2	BMJ	Thompson	MYJ
2	17	em_p1		BMJ	WSM-6	MYJ
3	18	em_n1		BMJ	WSM-6	YSU
4	19	nmm_ctl		BMJ	Thompson	YSU
5	20	nmm_p1		BMJ	Ferrier	YSU
6	21	nmm_n1		KF	Thompson	MYJ
7	22	eta_ctl		KF	WSM-6	MYJ
8	23	eta_n1		KF	WSM-6	YSU
9	24	eta_n2		KF	Thompson	YSU
10	25	eta_n3		KF	Ferrier	YSU
11	26	eta_n4		Grell	Thompson	MYJ
12	27	eta_p1		Grell	WSM-6	MYJ
13	28	eta_p2		Grell	WSM-6	YSU
14	29	eta_p3		Grell	Thompson	YSU
15	30	eta_p4		Grell	Ferrier	YSU

and ENS20^{phys} ensembles have one more set of varied physics schemes relative to ENS4 and ENS4^{phys}, because in addition to the different planetary boundary layer, microphysics, and surface layer schemes, both 20-km ensembles use different cumulus parameterizations (the 4-km ensembles don't use cumulus parameterization). However, 5-member subsets of ENS20 and ENS20^{phys} that use the same cumulus parameterization 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$^{phys}_{cp}$ where CP refers to one of the three different cumulus parameterizations used: 1) Kain-Fritsch (KF; Kain and Fritsch 1993), 2) Betts-Miller-Janjic (BMJ; Betts 1986; Betts and Miller 1986; Janjic 1994), and 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. Ensemble member specifications are listed in Tables 1 and 2.

The forecasts were examined for 20 cases during April-June 2007 (Figure 2). Note, these 20 cases are included in the 23 cases in which ENS4 and ENS20

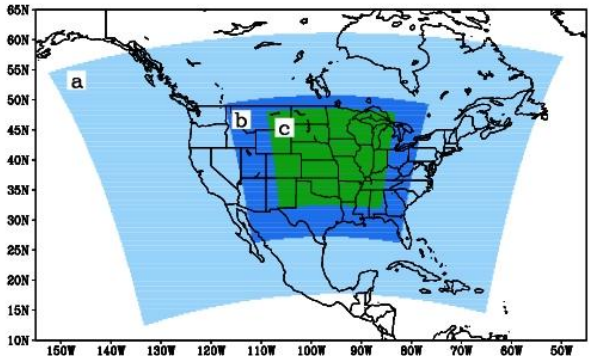

Figure 1 Domains for a) SREF ensemble members b) ENS4 and ENS20 ensemble members, and c) the analyses conducted in this study.

Figure 2 Pink shaded dates indicate when 10-member SSEF system simulations were conducted for SE2007 and dark red shading indicates which cases are used in the analysis for this study.

precipitation forecasts were compared in Clark et al. (2009). 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 occurring.

This study examines growth of spread (i.e. ensemble variance) and statistical consistency [i.e. correspondence between ensemble variance and mean-square-error (MSE) of the ensemble mean] for 12 fields: 500-, 700-, and 850-hPa geopotential height (500Z, 700Z, and 850Z, respectively), mean-sea-level pressure (MSLP), 2-meter temperature (T2), 2-meter dewpoint (Td2), 850-hPa wind magnitude (850WMAG), 850-hPa temperature (850T) and dewpoint (850Td), 3-hrly accumulated precipitation (PREC), most unstable convective available potential energy (MUCAPE), and magnitude of the 10-meter to 500-hPa shear vector (WS). In subsequent analyses the 12 fields examined are separated into those that are "mass-related", or depend 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 WS). Ensemble variance and MSE of the ensemble mean are computed according to Eqs. B7 and B6, respectively, in Eckel and Mass (2005).

3. Results

3.1 Ensemble variance time series

To illustrate the temporal evolution of spread growth during the 33-hour forecast period, time series of average ensemble variance for all 12 fields at 3-hrly intervals with box-plots overlaid to show variability are displayed in Figure 3. In order to compare 4-km and 20-km ensembles with the same types of perturbations, each panel in Fig. 3 displays either ENS4 and ENS20 ensemble variance, or ENS4^{phys} and ENS20^{phys} ensemble variance. Note that the different y-axes in Fig. 3 don't allow a direct comparison of growth rates between ensembles with both sets of perturbations (IC/LBC+Phys) and those with only mixed physics (Phys) for each field, but that 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, c, e, and g) have a generally linear increase in mean spread and it appears that ENS4 spread is increasing at a faster rate than ENS20, which is generally expected behavior because the smaller scales being resolved in ENS4 should lead to faster error growth relative to a model using a coarser grid (e.g., Lorenz 1969).

For the mass-related fields in ENS4^{phys} and ENS20^{phys} (Figs. 3b, d, f, and h), the ENS20^{phys} mean variances increase at a faster rate than those in ENS4^{phys}, with the exception of mean MSLP variances (Fig. 3h) which appear to be similar. In addition, the spread increase in ENS4^{phys} and ENS20^{phys} is not linear as it was for ENS4 and ENS20, but instead has a ~ 6-hr period during forecast hours 21-27 (1800-0000 UTC) during which spread increases at a noticeably faster rate than the other times. This 6-hr period corresponds to when peak solar insolation occurs and likely corresponds to when the different physics parameterizations are most active and thus result in the most error growth. For example, in the central US, 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 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.

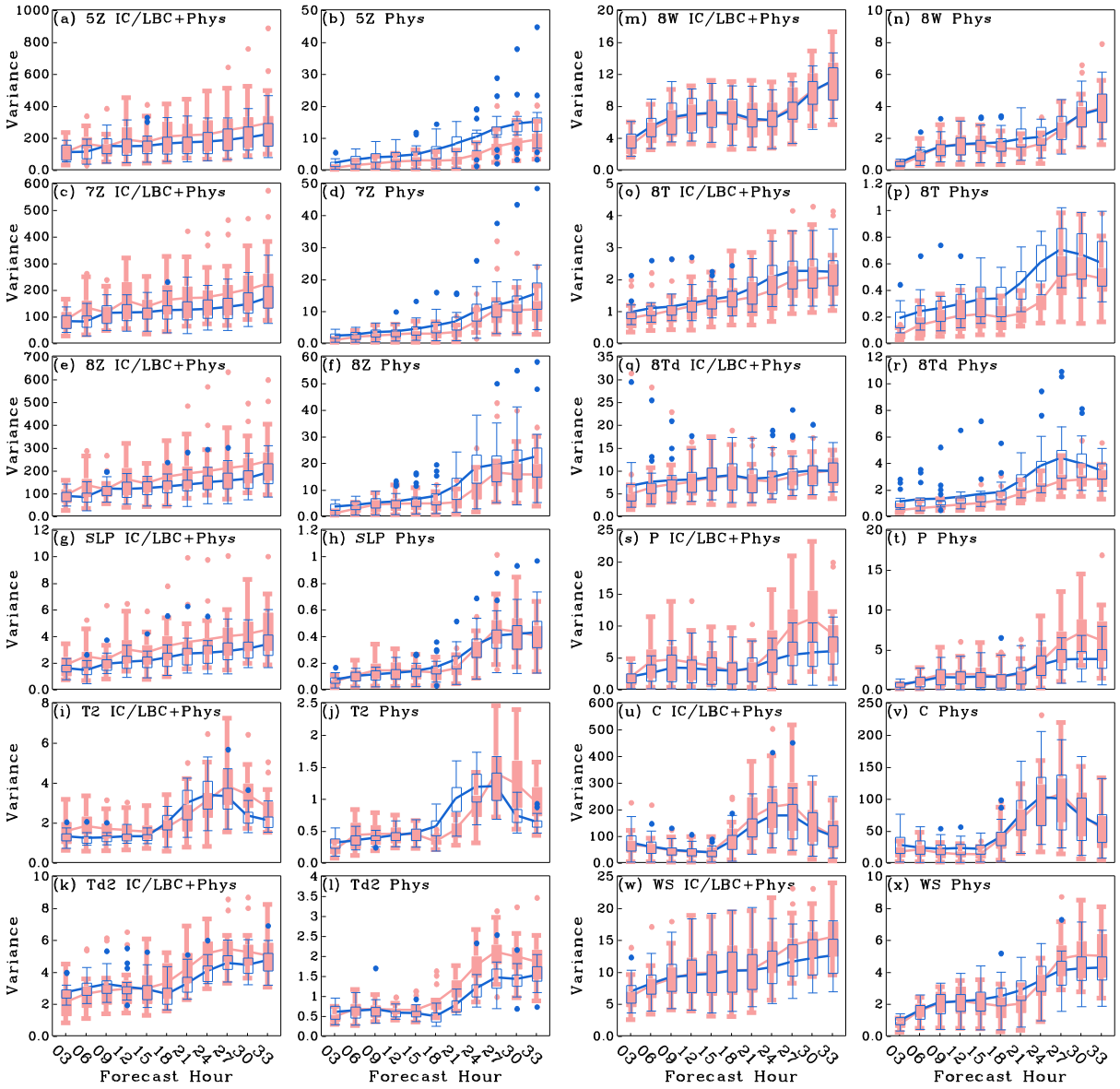

Figure 3 Time series of mean ensemble variance from ENS4 (pink line) and ENS20 (blue line) for the variables a) 500-, c) 700-, and e) 850-hPa geopotential height, g) mean-sea-level pressure, i) 2-meter temperature, k) 2-meter dewpoint, m) 850-hPa wind magnitude, o) 850-hPa temperature, q) 850-hPa dewpoint, s) 3-hrly accumulated precipitation, u) most unstable CAPE, w) magnitude of 500-hPa to 10-m shear vector. b), d), f), h), j), l), n), p), r), t), v), and x) same as a), c), e), g), i), k), m), o), q), s), u), and w) except for ENS4^{phys} and ENS20^{phys}. Boxplots overlay the mean at each time interval.

For the low-level fields (Figs. 3i-x), ENS4 and ENS20 mean variances (Figs. 3i, k, m, o, q, s, u, and w) have clear diurnal signals superimposed on increasing trends. The differences in mean variances and mean variance growth rates between ENS4 and ENS20 are dependent on the variable analyzed. For example, 850WMAG variances in ENS4 and ENS20 are very similar over the entire forecast 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 WS 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 are also dependent on the variable analyzed. For example, 850T and 850Td ENS4 and ENS20 variances (Figs. 3o and q, respectively) have smaller amplitudes relative to the other low-level fields, and peak variances occur for 850WMAG 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.

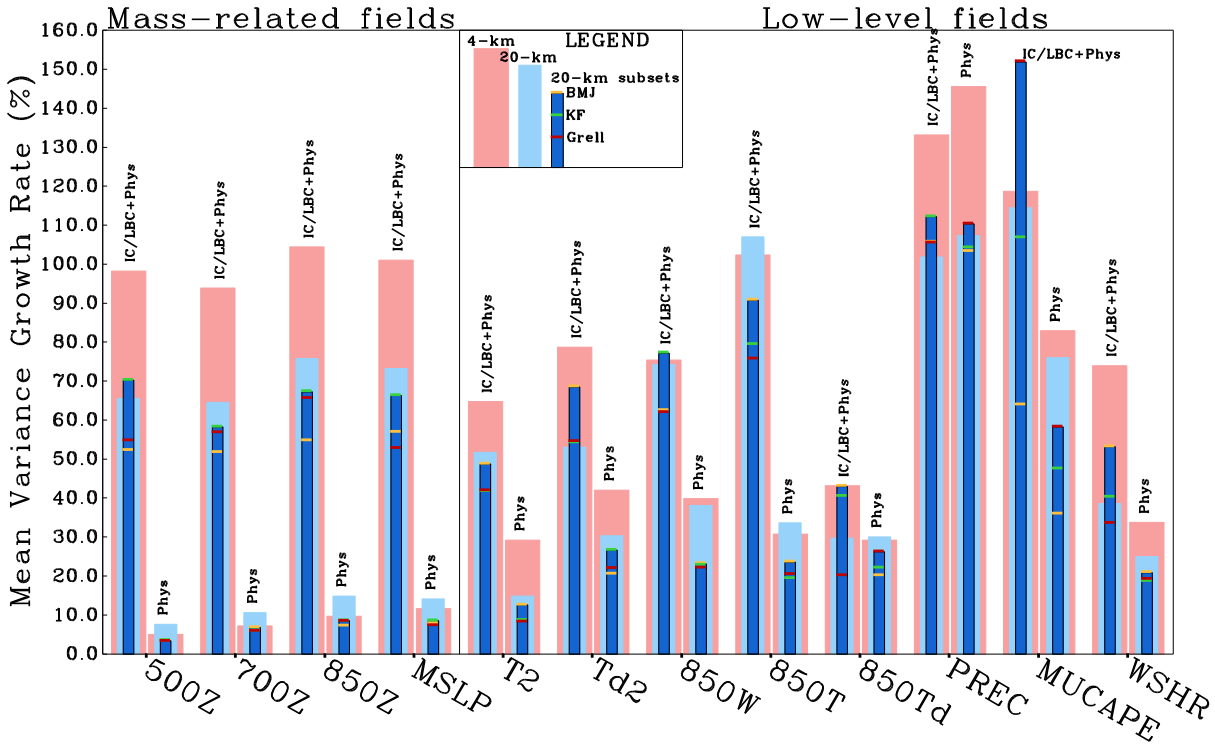

Figure 4 Mean variance growth rates from the ENS4, ENS4^{phys}, ENS20, and ENS20^{phys} ensembles for fields shown in Figure 3. Growth rates for five member subsets of ENS20 and ENS20^{phys} that have the same cumulus parameterization are also shown. The histograms to the left for each variable indicate 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). A legend is provided at the top of the figure.

3.2 Variance growth rates

It was possible to subjectively infer differences in mean ensemble variance growth rates from the analysis conducted in Figure 3; however, to compare ENS4 and ENS20 to ENS4^{phys} and ENS20^{phys} and better quantify mean variance growth rates, a simple objective method was developed using a standard formula for growth rate: $(Var_f-Var_i)/Var_i*100\%$, where Var_i and Var_f are initial and final mean variance, respectively. To reduce the impact 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 hours or one complete diurnal cycle. Also, to smooth out high frequency variability which occurred mainly during the first 12 hours of the forecasts for the mass-related fields, a lowess filter was applied to the 1-hrly mean variance time series using the R statistical software package (R Development Core Team 2007). Finally, the mean variances from ENS4^{phys}, ENS20, ENS20^{phys}, ENS20$^{phys}{}_{kf}$, ENS20$^{phys}{}_{bmj}$, and ENS20$^{phys}{}_{GD}$ were adjusted by the difference between their variance at forecast hour 9 and that from ENS4. 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 mean variance growth rates obtained from this methodology are shown in Figure 4. The growth rates of the mass-related fields exhibit very similar behavior, with ENS4 growth rates for mass-related fields around 100% and those of ENS20 around 70%. Also, the ENS20$_{cp}$ subsets (i.e. 5-member subsets with same cumulus parameterization) have slightly lower growth rates than ENS20, with the exception of 500Z mean variance in ENS20$_{kf}$ which is slightly larger than ENS20. The lower growth rates in ENS20$_{cp}$ subsets relative to ENS20 are expected because ENS20 has one additional source of model uncertainty relative to ENS20$_{cp}$ subsets, and the differences between ENS4 and ENS20 growth rates are consistent with faster error growth expected as smaller scales are resolved (Lorenz 1969; Smagorinsky 1969). Note that ENS20 also has one additional source of model uncertainty relative to ENS4, but that the greater impact of higher resolution in ENS4 relative to having one additional source of model uncertainty in ENS20 results in higher growth rates in ENS4 than in ENS20.

For the mass-related fields, the “mixed-physics only” ensemble subsets (denoted “Phys” in Fig. 4) have much lower growth rates (~ 10%) than the subsets with both sets of perturbations (denoted “IC/LBC+Phys” in Fig. 4) discussed above, which is consistent with results

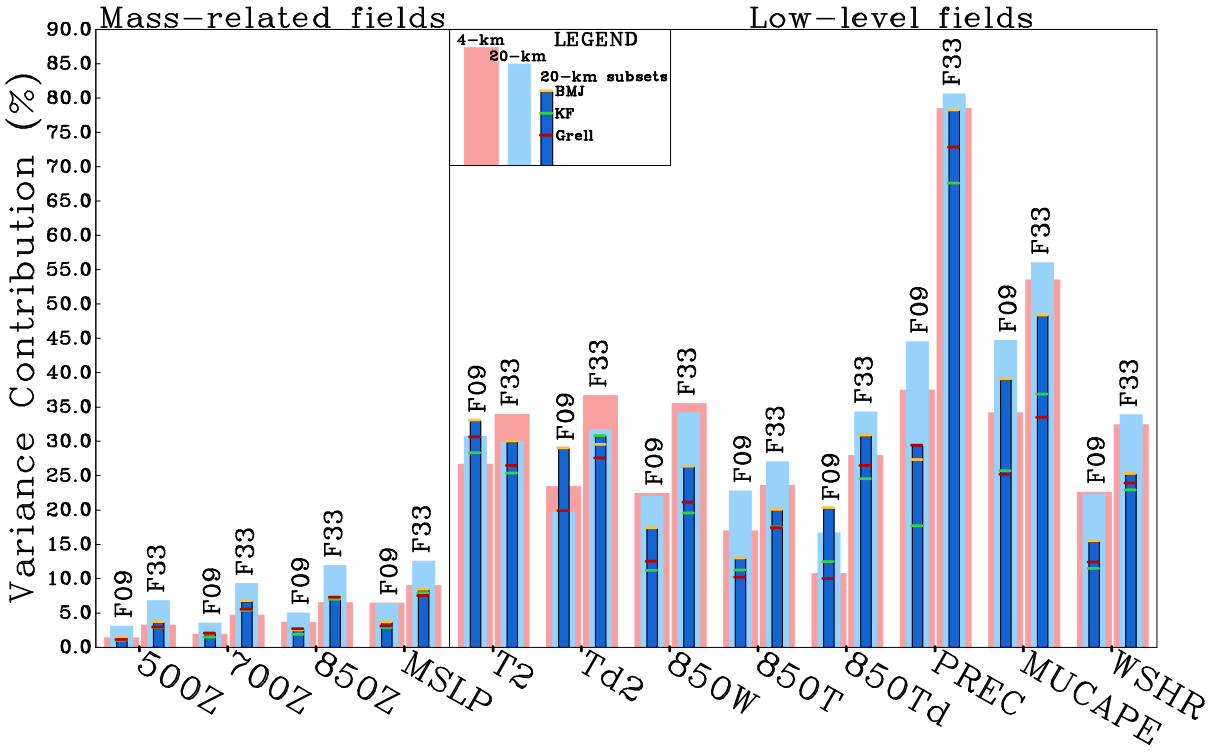

Figure 5 Mean variance ratio [%; Var(ENS4^{phys})/Var(ENS4) and Var(ENS20^{phys})/Var(ENS20)] from the 4-km and 20-km grid-spacing ensembles and from the 5-member 20-km ensemble subsets for the fields in Figure 3. The histograms to the left (right) for each field are for forecast hour 9 (33). A legend is provided at the top of the figure.

found by Kong et al. (2007) using a similar dataset. The differences in growth rates occur because the different physics schemes that parameterize 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. On average, they are only active 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$^{phys.}$ Thus, unlike the ENS4 vs. ENS20 comparison, the impact of one additional source of model uncertainty in ENS20^{phys} is greater than the impact of higher resolution in ENS4^{phys}. Also, similar to the ENS20$_{cp}$ subsets, the ENS20$^{phys}_{cp}$ subsets have smaller growth rates relative to ENS20^{phys} resulting from having one less source of model uncertainty.

The mean variance growth rates among low-level fields are much more variable than for the mass-related fields. However, the two sets of fields are similar in that the IC/LBC+Phys ensemble subsets have faster growth rates than the Phys subsets, with PREC being the only exception. The low-level fields from Phys ensemble subsets all have higher variance growth rates than the mass-related fields, which is consistent with the low-level fields being more dependent on the varied physical parameterization schemes. For the low-level fields T2, Td2, PREC, and WS, the ENS4 and ENS4^{phys} variance growth rates are noticeable larger than those of ENS20 and ENS20^{phys}, respectively, however, for other variables like 850WMAG, 850T, and MUCAPE, the ENS20 and ENS20^{phys} growth rates are similar to or higher than those from ENS4 and ENS4^{phys}, respectively. This may indicate that, for some variables, higher resolution in ENS4 and ENS4^{phys} results in larger error growth 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} has a larger impact than higher resolution. Also note that for many of the low-level variables (e.g., Td2, 850WMAG, 850Td, PREC, MUCAPE, and WS) some of the ENS20$_{cp}$ and ENS20$^{phys}_{cp}$ ensemble subsets have greater mean variance growth rates than ENS20 and ENS20^{phys}, respectively, which is counterintuitive because *smaller* variance growth rates would be expected with one less

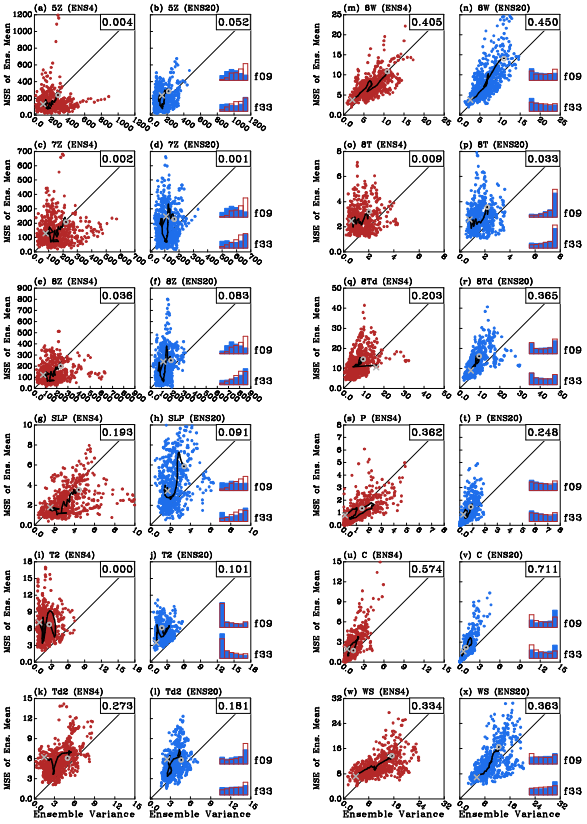

Figure 6 Scatter plots of ensemble variance vs. mean-square-error of the ensemble mean from ENS4 (black dots) for the raw fields a) 500-, c) 700-, and e) 850-hPa geopotential height, g) mean-sea-level pressure, i) 2-meter temperature, k) 2-meter dewpoint, m) 850-hPa wind magnitude, o) 850-hPa temperature, q) 850-hPa dewpoint, s) 1-hrly accumulated precipitation, u) most unstable CAPE, w) magnitude of 500-hPa to 10-m shear vector. b), d), f), h), j), l), n), p), r), t), v), and x) same as a), c), e), g), i), k), m), o), q), s), u), and w) except for ENS20 (grey dots). Correlation coefficients are provided in the right-top corner of each panel and rank histograms for ENS4 (black outline) and ENS20 (grey shaded) for forecast hours 09 and 33 are displayed in the bottom-right corner of the ENS20 plots. The lines overlaying the scatters connect the average variance-MSE points for each forecast hour ("x" indicates hour 1 and "o" indicates hour 33).

source of model uncertainty. In particular, MUCAPE variance growth rates for ENS20$_{GD}$ (~ 150%) are much larger than those of ENS20 (~ 110%), while ENS20$_{kf}$ MUCAPE growth rates (~ 100%) are similar to ENS20 and those from ENS20$_{bmj}$ (~ 60%) are much smaller.

It is suspected that, for all of the low-level fields, systematic model biases associated with certain parameterization schemes and combinations of parameterization schemes are impacting the growth rates for these low-level fields. These biases are important to consider in the context of an ensemble because, as discussed by Eckel and Mass (2005), systematic biases that increase forecast uncertainty do so "artificially" because the associated errors are not uncertain. Further work is planned to examine the impact of these biases.

3.3 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. {Var(Phys)/Var(IC/LBC+Phys)}*100%] is computed for all 12 fields at forecast hours 09 and 33 (Fig. 5). Note that the actual contributions to ensemble spread in the IC/LBC+Phys 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 variance contribution from mixed-physics assumes the interaction term is negligible.

The mixed-physics contributions to ensemble variance were generally much smaller for the mass-related fields relative to the low-level fields, which could be inferred from the time series of mean ensemble variance (Fig. 3; note different y-axis scale) and is also consistent with the differences between the IC/LBC+Phys and Phys variance growth rates (Fig. 4). The contributions for mass-related fields decrease as higher atmospheric levels are examined which likely occurs because the the higher levels are impacted less by the boundary layer where the physics parameterizations have the greatest impact. Also, the mixed-physics variance contributions for mass-related fields in ENS20 were larger than for ENS4, consistent with the extra source of model uncertainty in ENS20, and the ENS20$_{cp}$ subsets had similar contributions to those from ENS4, except for the mixed-physics contributions for MSLP that were smaller in the ENS20$_{cp}$ subsets relative to ENS4. Perhaps the most noticeable feature for the mass-related fields in Figure 5 is that the mixed-physics variance contributions for all ensemble subsets are higher at forecast hour 33 than 09 implying that the influence of model uncertainty on ensemble spread increases with forecast lead time.

For the low-level fields, similar to the variance growth rates (Fig. 4), there is much more variability in variance contributions among the different fields examined and 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. On the other hand, 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. Similar to the growth rates for low-level fields, model biases may be having an impact on the mixed-physics variance contributions.

3.4 Variance-MSE relationship and statistical consistency

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., Grimit and Mass 2007). To quantify the variance-MSE relationship, past works have used linear correlations (e.g., Jones et al. 2007). This study also employs variance-MSE linear correlations, but care should be taken 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. For more details on the spread-error relationship in ensemble prediction systems, Grimit and Mass (2007) provide a thorough literature review.

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. A statistical consistency analysis can also provide information on whether an ensemble system is over- or under-dispersive. In an under-dispersive (over-dispersive) ensemble the average MSE is larger (smaller) than the ensemble variance. In this study, the variance-MSE and statistical consistency analyses are used as a simple method for examining the impacts of the different spread growth rates on the quality of the ensemble forecasts (as inferred from the aforementioned "spread-skill" metrics).

To illustrate the variance-MSE relationship and statistical consistency in the ENS4 and ENS20 ensembles for the different fields examined, scatterplots of ensemble variance vs. MSE are displayed in Figure 6. Each panel in Fig. 6 contains variance-MSE points for each case and for each forecast hour (20 cases x 33 times = 660 points for each panel), correlation coefficients indicate the degree of correspondence between ensemble variance and MSE (i.e. the reliability of ensemble variance as a predictor of forecast skill), and lines overlaying the scatters connect average variance-MSE points for each forecast hour ("x" marks forecast hour 1 and "o" marks forecast hour 33) indicating statistical consistency. The diagonal lines drawn from the bottom left to the upper right of each panel indicate "perfect" statistical consistency. Rank histograms (e.g., Hamill 2001) provided in Figure 6 valid at forecast lead times of 09 and 33 also provide information regarding representation of forecast uncertainty: flat rank histograms imply an accurate depiction of forecast uncertainty, U-shaped (n-shaped) rank histograms imply over-dispersion (under-dispersion), and right (left) skewness indicates a tendency for over-prediction (under-prediction). To allow for a more convenient comparison between ENS4 and ENS20, the 16 bins composing the ENS20 rank histograms were regrouped into 6 bins which each contain an equal portion of the original 16 bins. This "regrouping" technique has also been used in Clark et al. (2009).

For the mass-related fields (Figs. 6a-h), the variance-MSE correlations in ENS4 and ENS20 are very low suggesting that 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 mid-tropospheric 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 towards the right (i.e. higher values of ensemble variance). These results are reflected by the statistical consistency lines in Figs. 6a-h, with ENS4 statistical consistency lines generally oriented along and slightly to the right of the diagonal implying slight over-dispersion, while the ENS20 lines are more erratic and tend to veer toward the left of the diagonal implying more noticeable under-dispersion. Finally, the results for the mass-related fields are also consistent with the higher spread growth rates found in ENS4 relative to ENS20 shown in Figure 3, and imply that the greater ENS4 spread for mass-related fields may result in over-dispersion.

Rank histograms for the mass-related fields (Figs. 6b, d, f, and h) imply that as forecast lead time increases from hour 09 to 33, ENS20 goes from overpredicting to underpredicting forecast uncertainty. Thus, it appears that the perturbations applied to the ENS20 ICs initially encompass the observed atmospheric state, but as lead time increases, the perturbations do not grow (or grow very slowly) and the observed atmospheric state begins to diverge away from the ENS20 members. For ENS4, it is difficult to see the implied depiction of forecast uncertainty from the rank histograms because of their left-skewness which implies underprediction of the mass-related fields. This underprediction is an interesting result and was also verified by an analysis of domain averaged geopotential heights for all cases. The underprediction could possibly result from a general cool bias in the lower part of the troposphere or perhaps stronger synoptic-scale storm systems in ENS4 members; however, a further examination is beyond the scope of this study.

For the low-level fields (Figs. 6i-x), the variance-MSE correlations in ENS4 and ENS20 are quite variable

depending on the field examined with 850W (Figs. 6m-n), PREC (Figs. 6s-t), MUCAPE (Figs. 6u-v), and WSHR (Figs. 6w-x) 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. 6i-l. Also, the U-shaped rank histograms for 850Td (Fig. 6q) imply underdispersion. The statistical consistency lines in ENS4 and ENS20 for the low-level fields are also variable, but exhibit similarities depending on the type of field examined. For example, the temperature and dewpoint fields [T2 (Figs. 6i-j), Td2 (Figs. 6k-l), 850T (Figs. 6o-p), and 850Td (Figs. 6q-r)] behave quite similarly and indicate underdispersion for both ENS4 and ENS20, and also indicate that there are not distinctive differences between ENS4 and ENS20. The variables dependent on wind fields [850WMAG (Figs. 6m-n) and WSHR (Figs. 6w-x)] also behave similarly and have better statistical consistency than the temperature and dewpoint fields. Furthermore, for the wind fields, the ENS20 statistical consistency lines tend to veer to the left of the diagonal with increasing forecast lead time indicating increasing underdispersion, while those from ENS4 tend to veer toward the diagonal indicating improved statistical consistency. For the wind fields, the better statistical consistency in ENS4 relative to ENS20 at the later forecast lead times is reflected in the rank histograms which are flatter at forecast hour 33 in ENS4 (Figs. 6n and x). For PREC and MUCAPE fields in ENS4 and ENS20 (Figs. 6s-t and 6u-v, respectively), under-dispersion is generally indicated except for in the ENS4 PREC field (Fig. 6s) which indicates overdispersion.

Generally, the variance-MSE relationships and statistical consistency lines indicate better statistical consistency in ENS4 relative to ENS20 for the mass-related fields, no noticeable differences in statistical consistency for temperature and dewpoint fields as well as PREC and MUCAPE fields, and better statistical consistency in ENS4 for wind-fields (850WMAG and WSHR) at later forecast lead times.

4. CONCLUSIONS

Generally, the results from this study could be interpreted as encouraging for future convection-allowing ensemble systems simply because statistical consistency analyses indicate that faster spread growth should lead to more reliable forecasts in the convection-allowing ensembles when considering mass-related and wind-related fields. However, the higher resolution of ENS4 did not seem to improve spread-error metrics for temperature and dewpoint fields. However, further work needs to analyze larger sets of cases for different periods and further assess whether the increased dispersion does truly improve probabilistic forecasts. Finally, the behavior of ensemble variance observed in this study should be helpful for future ensemble design, and recognition of systematic model biases should provide motivation for improving the physics parameterizations used with convection-allowing grid-spacing.

5. Acknowledgments

The authors would like to 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. This particular research was funded by NSF Grant ATM-0537043. 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, Jimy 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) supported by NSF

6. References

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, 2004: Mesoscale weather prediction with the RUC hybrid isentropic–terrain-following coordinate model. *Mon. Wea. Rev.*, 132, 473–494.

——, D. Dévényi, S. S. Weygandt, K. J. Brundage, J. M. Brown, G. A. Grell, D. Kim, B. E. Schwartz, T.G. Smirnova, T.L. Smith, and G.S. Manikin, 2004: 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, 3, 85pp.

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.

——, ——, M. Xue, and F. Kong, 2009: A Comparison of Precipitation Forecast Skill between Small Convection-Allowing and Large Convection-Parameterizing Ensembles. *Wea. Forecasting* (In Press).

Coniglio, M., K. L. Elmore, J. S. Kain, S. J. Weiss, M. Xue, and M. Weisman, 2009: Evaluation of WRF model output for severe-weather forecasting from the 2008 NOAA Hazardous Weather Testbed Spring Experiment. *Wea. Forecasting*, (pending review).

Du, J., J. McQueen, G. J. DiMego, T. L. Black, H. Juang, E. Rogers, B. S. Ferrier, B. Zhou, Z. Toth, and S. Tracton, 2004: The NOAA/NWS/NCEP Short Range Ensemble Forecast (SREF) system: Evaluation of an initial condition vs multiple model physics ensemble approach. *Preprints, 20th Conference on Weather Analysis and Forecasting/ 16th Conference on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., CDROM, 21.3.

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.

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(D22), 8851, doi:10.1029/2002JD003296.

Eckel, F.A., and C.F. Mass, 2005: Aspects of effective mesoscale, short-range ensemble forecasting. *Wea. Forecasting*, 20, 328–350.

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.

Hong, S.-Y, and J.-O. J. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Korean Meteor. Soc*., 42, 129-151.

Houtekamer, P. L., L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon Wea. Rev.*, 124, 1225–1242.

Janjic, 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. *11th Conf. on Numerical Weather Prediction*, Norfolk, VA, 19-23 August 1996; Amer. Meteor. Soc., Boston, MA, 354-355.

——, 2002: Nonsingular implementation of the Mellor-Yamada Level 2.5 Scheme in the NCEP Meso model. NCEP Office Note No. 437, NOAA/NWS, 61 pp.

——, 2003: A nonhydrostatic model based on a new approach. *Meteorol. 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-Fritcsh scheme, *The representation of cumulus convection in numerical models*, K. A. Emanuel and D. J. Raymond, Eds., 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 and 17th Conf. on Numerical Weather Prediction*, Washington, DC, Amer. Meteor. Soc., 2A.5. [Available online at http://ams.confex.com/ams/pdfpapers/94843.pdf.]

——, S. J. Weiss, D. R. Bright, M. E. Baldwin, J. J. Levit, G. W. Carbin, C. S. Schwartz, M. L. Weisman, K. K. Droegemeier, D. Weber, and K. W. Thomas, 2008:

Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting* (In Press).

Kong, F., M. Xue, K. K. Droegemeier, D. Bright, M. C. Coniglio, K. W. Thomas, Y. Wang, D. Weber, J. S. Kain, S. J. Weiss, and J. Du, 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 Conference On Weather Analysis and Forecasting/18th Conference on Numerical Weather Prediction*, Park City, UT, Amer. Meteor. Soc., CD-ROM, 3B2.

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), 16663-16682.

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. Contrib. Geophys. Inst. Acad. Sci. USSR, (151), 163-187 (in Russian).

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. *Boundary 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 Department 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 (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Skamarock, W. C., 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 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.

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.

Xue, M., F. Kong, D. Weber, K. W. Thomas, Y. Wang, K. Brewster, K. K. Droegemeier, J. S. Kain, S. J. Weiss, D. Bright, M. Wandishin, M. Coniglio, and J. Du, 2007: CAPS realtime storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2007 Spring Experiment. *Preprints, 22nd Conference On Weather Analysis and Forecasting/18th Conference on Numerical Weather Prediction*, Park City, UT, Amer. Meteor. Soc., CD-ROM, 3B1.