

Hierarchical cluster analysis of a convection-allowing ensemble during the Hazardous Weather Testbed 2009 Spring Experiment. Part II: Ensemble clustering over the whole experiment period

Aaron Johnson

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

Xuguang Wang

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

Ming Xue

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

Fanyou Kong

Center for Analysis and Prediction of Storms, Norman, Oklahoma

Submitted to *Monthly Weather Review*

Jan. 14, 2011

Revised May 7, 2011

Corresponding author address:

Dr. Xuguang Wang
School of Meteorology
University of Oklahoma
120 David L. Boren Blvd.
Norman, OK, 73072
xuguang.wang@ou.edu

Abstract

Twenty-member real-time convection-allowing storm-scale ensemble forecasts with perturbations to model physics, dynamics, initial (IC) and lateral boundary conditions (LBC) during the NOAA Hazardous Weather Testbed Spring Experiments provide a unique opportunity to study the relative impact of different sources of perturbation on convection-allowing ensemble diversity. In part II of this two part study, systematic similarity/dissimilarity of hourly precipitation forecasts among ensemble members from the spring season of 2009 are identified using Hierarchical Cluster Analysis (HCA) with a fuzzy Object-based Threat Score (OTS), developed in Part I. In addition to precipitation, HCA is also performed on ensemble forecasts using the traditional Euclidean distance for wind speed at 10-m and 850 hPa, and temperature at 500 hPa.

At early lead time (3 hours, valid 03 UTC) precipitation forecasts cluster primarily by data assimilation and model dynamic core, indicating a dominating impact of models, with secondary clustering by microphysics. There is an increasing impact of planetary boundary layer (PBL) scheme on clustering relative to microphysics scheme at later lead times. Forecasts of 10-m wind speed cluster primarily by PBL scheme at early lead times, with an increasing impact of LBC at later lead times. Forecasts of mid-tropospheric variables cluster primarily by IC at early lead time and LBC at later lead times. The radar and mesonet data assimilation (DA) show its impact, with members without DA in a distinct cluster, through the 12 hour lead time (valid 12 UTC) for both precipitation and non-precipitation variables. The implication for optimal ensemble design for storm-scale forecasts is also discussed.

1. Introduction

Early studies on the impact of ensemble perturbations beyond initial and lateral boundary conditions (ICs/LBCs) have focused on cumulus-parameterizing (CP) resolution¹ or only limited sampling of sources of forecast uncertainty (e.g. Arribas et al. 2005; Jankov et al. 2005; Gallus and Bresch 2006; Jankov et al. 2007; Kong et al. 2007, Aligo et al. 2007; Clark et al. 2008; Weisman et al. 2008; Palmer et al. 2009; Berner et al. 2011; Hacker et al. 2011). Past studies of the impact of ensemble perturbations found short range mesoscale ensembles with cumulus parameterization to be sensitive to both model and physics uncertainty, in addition to IC uncertainty (Stensrud et al. 2000; Wandishin et al. 2001). Studies have also found that using multiple physics schemes and other methods, such as stochastic energy backscatter, to sample model uncertainty can improve the ensemble forecasts at CP resolution (Palmer et al. 2009; Berner et al. 2011). However, studies based on CP resolution ensembles (Stensrud et al. 2000; Hou et al. 2001; Wandishin et al. 2001; Alhamed et al. 2002; Yussouf et al. 2004; Gallus and Bresch 2006; Aligo et al. 2007; Palmer et al. 2009; Berner et al. 2011; Hacker et al. 2011) are not necessarily applicable to convection-allowing² ensembles. One difference is that cumulus parameterization has been shown to dominate the precipitation forecast uncertainty resulting from model physics in the CP ensembles (Jankov et al. 2005) whereas in the convection-allowing ensemble no cumulus parameterization is applied. Another difference is that growth rates of convective-scale perturbations allowed in the convection-allowing forecasts can be highly non-linear (Hohenegger and Schar 2007). These results motivate investigation of the

¹ CP resolution refers to grid spacing coarser than about 4 km which requires cumulus parameterization schemes to account for sub-grid scale vertical redistribution of heat and moisture resulting from moist convection (Molinari and Dudek 1992).

² Convection-allowing resolution refers to grid spacing less than or equal to 4 km which allows vertical redistribution of heat and moisture to be effectively represented by grid-scale convection (Weisman et al. 1997), making cumulus parameterization unnecessary. The term convection-resolving is avoided because the convective scale details are not necessarily adequately resolved (Bryan et al. 2003; Petch 2006).

impact of different sources of ensemble perturbations at convection-allowing resolution.

Published studies of the impact of different perturbations on ensemble behavior in the context of convection-allowing ensemble forecasting on numerous cases over a period of several weeks, with perturbations that comprehensively sample uncertainty in the ICs, LBCs, model dynamics and multiple physics schemes are scarce.

This paper is the second of a two-part series which takes a step toward understanding the impacts and importance of the sources of uncertainties in model physics, model dynamics, IC, and LBCs for convection-allowing ensemble forecasts. This is done with a Hierarchical Cluster Analysis (HCA, Alhamed et al. 2002; Anderberg 1973) of the storm-scale ensemble forecasts (SSEFs) for the 2009 National Oceanic and Atmospheric Administration Hazardous Weather Testbed (NOAA HWT) Spring Experiment. Some of the key issues for future study of ensemble design and post-processing are also briefly inferred from the results of the HCA. Part I (Johnson et al 2011a) demonstrated that a new object-oriented measure of the dissimilarity (distance measure) of two precipitation forecasts improves automated clustering compared to traditional distance measures for a severe weather forecasting application. The improvement results from the object-oriented distance being based on attributes of discrete objects rather than a point-wise comparison of the forecasts. This paper (Part II) shows composite dendrograms, constructed using the new object-oriented HCA, from cases during the entire NOAA HWT 2009 Spring Experiment³ to explore systematic similarities and dissimilarities among the ensemble members. HCA is also applied here to different lead times and variables beyond precipitation to understand the impact of different sources of perturbations as a function of lead times and/or diurnal cycles and forecast variables.

³ The 2009 Spring Experiment spans from 30 April 2009 to 5 June 2009. Forecasts were only run during the weekdays. A total of 26 forecasts were used in this study after discarding 2 days due to incomplete data and 2 days due to negligible precipitation being predicted.

Johnson et al. 2011a (Part I) describes in detail the HCA algorithm used in this study. HCA consists of initially identifying each forecast as a single-element cluster then iteratively merging two clusters together until all forecasts are in the same cluster. HCA has been often used to study synoptic and larger scale phenomena such as climate regimes (e.g., Kalkstein et al. 1987; Cheng and Wallace 1993; Fovell and Fovell 1993; Weber and Kaufmann 1995). A review of the use of cluster analysis in geophysical research in general is found in Gong and Richman (1995).

HCA has also been applied in an ensemble forecasting context on scales ranging from seasonal to mesoscale (Brankovic et al. 1990; Palmer et al. 1990; Molteni et al. 1996; Alhamed et al. 2002; Nakaegawa and Kanamitsu 2006; Yussouf et al. 2004; Brankovic et al. 2008). These studies have examined the inability of a seasonal forecast ensemble to predict the most likely regime based on cluster membership (Nakaegawa and Kanamitsu 2006), the performance of cluster means relative to overall ensemble mean for a global ensemble (Brankovic et al. 1990; Palmer et al. 1990), and the sensitivity of mesoscale ensemble forecasts to model configuration (Alhamed et al. 2002; Yussouf et al. 2004). Although few studies have examined ensemble behavior through a systematic clustering of forecasts on multiple cases, ensemble cluster analysis on individual cases has been applied both operationally and in a research setting (e.g., Tracton and Kalnay 1993; Atger 1999; Brankovic et al. 2008). For example, cluster analysis has been proposed in operational settings to condense the ensemble data by presenting a manageable subset of forecasts using cluster means (Tracton and Kalnay 1993, Toth et al. 1997) or performing a classification of the forecasts (Atger 1999). A notable exception to the emphasis on individual cases is Yussouf et al. (2004). Yussouf et al. (2004) showed that short range (0-36 hours), mesoscale (20-48 km grid spacings) forecasts with cumulus parameterization

systematically clustered according to models, each of which also had different physics, even when similar ICs were used in different clusters. To the authors' knowledge, HCA has not been systematically applied to convection-allowing forecasts for the purpose of understanding the impact of ensemble perturbations.

This study applies an automated clustering method to examine the systematic impact of ensemble perturbations in a convection-allowing ensemble. Through the use of an object-oriented distance measure, an automated approach is possible, making the results more reproducible and more easily applied to a large number of cases than manual, subjective evaluations. The paper is organized as follows. Section 2 summarizes the 2009 NOAA HWT Spring Experiment, storm-scale ensemble design and two methods of summarizing systematic ensemble clustering. Section 3 presents the HCA results for hourly accumulated precipitation forecasts while section 4 presents the HCA results for other variables. Section 5 is a summary and section 6 discusses implications for convection-allowing ensemble design and post-processing.

2. Description of 2009 NOAA HWT Spring Experiment ensemble and methods of summarizing HCA

a. 2009 NOAA HWT Spring Experiment

The HWT is a collaborative effort between the Storm Prediction Center (SPC), National Severe Storms Laboratory, and the Norman Oklahoma National Weather Service forecast office to facilitate development and transition to operations of new forecast technologies (Weiss et al. 2009). Since 2000 the HWT has hosted an annual Spring Experiment to provide model developers, research scientists, and operational forecasters an opportunity to interact while

evaluating and providing feedback on developing technologies in a simulated operational forecasting environment (Weiss et al. 2009). For the 2009 NOAA HWT Spring Experiment, the Center for Analysis and Prediction of Storms (CAPS) produced an experimental real-time convection-allowing ensemble, 5 days a week for 6 weeks, over a near-CONUS (Continental United States) domain (Kong et al. 2009; Xue et al. 2009).

b. Ensemble overview

The ensemble consists of 20 members, with 10 members from the Weather Research and Forecast (WRF) Advanced Research WRF (ARW; Skamarock et al. 2005), 8 members from the WRF Non-hydrostatic Mesoscale Model (NMM; Janjic 2003), and 2 members from the CAPS Advanced Regional Prediction System (ARPS; Xue et al. 2000, 2001, 2003). Each member has 4 km horizontal grid spacing and does not use cumulus parameterization. For WRF ARW and WRF NMM, 53 vertical levels are adopted. For ARPS, 43 vertical levels are adopted. Besides using multiple models, members are perturbed through the use of different ICs, LBCs, and physics as summarized in Table 1. Microphysics perturbations include Thompson (Thompson et al. 2008), Ferrier (1994), WRF Single Moment 6 class microphysics (WSM6, Hong et al. 2004), and Lin et al. (1983) schemes. Planetary boundary layer (PBL) perturbations include Mellor-Yamada-Janic (MYJ, Janjic' 1994), Yonsei University (YSU, Noh et al. 2003), and a diagnostic Turbulent Kinetic Energy (TKE)-based scheme (Xue et al. 2000). Land surface perturbations include RUC (Rapid Update Cycle, Benjamin et al. 2004) and NOAA (NCEP-Oregon State University-Air Force-NWS Office of Hydrology, Ek et al. 2003) land surface models. Shortwave radiation scheme perturbations include Goddard Space Flight Center (Tao et al. 2003), Dudhia (1989), and Geophysical Fluid Dynamics Laboratory (GFDL, Lacis and Hansen 1974) schemes.

The control members (labeled CN) obtain ICs from the operational National Centers for Environmental Prediction's North American Model (NCEP NAM) 0000 UTC analysis with additional radar and mesoscale observations assimilated using ARPS 3DVAR and cloud analysis package (Xue et al. 2003; Gao et al. 2004; Hu et al. 2006). Radial velocity from over 120 radars in the Weather Surveillance Radar (WSR)-88D network, as well as surface pressure, horizontal wind, potential temperature, and specific humidity from the Oklahoma Mesonet, Surface Aviation Observation, and Wind Profiler networks were assimilated by ARPS 3DVAR. The ARPS cloud analysis package uses radar reflectivity along with Geostationary Operational Environmental Satellite (GOES) visible and infrared channel 4 data to estimate hydrometeor species and adjust in-cloud temperature and moisture (Xue et al. 2009). For more details of the ARPS cloud analysis, please refer to Hu et al. (2006). One member from each of the three models (ARW C0, NMM C0, and ARPS C0) used identical configuration as the corresponding control member with the same model (ARW CN, NMMCN and ARPS CN, respectively) except without assimilating additional radar and mesonet data.

Perturbed ICs were created by adding to the CN IC positive and negative perturbation pairs derived from the three-hour forecasts of the NCEP Short-Range Ensemble Forecast (SREF) members⁴ indicated in Table 1. In Table 1 NAMa and NAMf are the direct NCEP-NAM analysis and forecast, respectively, while the CN IC has additional radar and mesoscale observations assimilated into the NAMa. Perturbations added to CN members to generate the ensemble of ICs, and LBCs for the SREF forecasts are from NCEP SREF (Du et al 2006). SREF members are labeled according to model dynamics: nmm (i.e., Nonhydrostatic Mesoscale Model) members use WRF-NMM, em (i.e., Eulerian Mass core) members use WRF-ARW, etaKF members use Eta model with Kain-Fritsch cumulus parameterization, and etaBMJ use Eta model with Betts-

⁴ NCEP SREF forecasts were initialized at 2100 UTC.

Miller-Janjic cumulus parameterization. Further details on the CAPS 2009 ensemble can be found in Xue et al. (2009) and Kong et al. (2009).

The results presented in sections 3 and 4 emphasize physics perturbations associated with microphysics and PBL scheme. The other physics perturbations do not have a strong enough signal in the HCA results to confidently make any additional inferences about the ensemble design (not shown).

c. Composite dendrograms

Hierarchical clustering is displayed graphically as a dendrogram, showing the step by step merging of clusters. Each forecast is initially a one-element cluster, listed along the bottom of the dendrogram. The distance between (i.e., dissimilarity of) single-forecast clusters is traditionally quantified with the squared Euclidean distance. The distance between multiple-forecast clusters is quantified as the increase in variability, which quantifies the diversity of the cluster, that would result from merging them into a single cluster. The two clusters with smallest distance between them are merged at each step. The merging of forecasts and clusters is depicted as two solid lines joining into one as the clustering proceeds from the bottom to the top of the dendrogram. The vertical axis is a cumulative measure of variability, summed over all clusters at that level. The difference in vertical axis values, $y_i - y_{i-1}$, is therefore the distance between the clusters merged at the i^{th} iteration. Lower level clusters contain more similar forecasts than higher level clusters. For a more detailed description of the clustering algorithm and dendrograms, please refer to Part 1 section 4a.

For each forecast, a separate dendrogram is generated. After creating dendrograms for all forecasts, they are composited into a single dendrogram so that we can analyze systematic

clustering of ensemble members over the Spring Experiment period. To create the composites, a normalized distance, d_N , between ensemble members is defined by subtracting the smallest distance between any two members at the same lead time on the same day, d_{min} , and dividing by the range of distances among all 20 members. The range is defined as the maximum distance minus the minimum distance, $d_{max}-d_{min}$. Mathematically, for an un-normalized distance, d ,

$$d_N = \frac{d - d_{min}}{d_{max} - d_{min}} \quad (1)$$

The normalized distance between each pair of members is then averaged over all forecasts and used as a composite distance measure. The composite distances are used for HCA using the modified Ward's algorithm described in Part I. The effect of the normalization is to give equal consideration to each forecast even though forecast to forecast variation in the distribution of distances is present. The composite dendrograms are intended to focus on systematic forecast similarities, rather than forecast similarities on any given forecast.

For the precipitation forecasts the distance measure, d , is the fuzzy Object-based Threat Score (OTS; defined in Part 1) while for non-precipitation forecasts the distance measure is the traditional squared Euclidean Distance (ED). As described in Part 1, the Method for Object-based Diagnostic Evaluation (MODE; Davis et al. 2006; Johnson et al. 2011a) is a features-based algorithm for identifying and comparing objects in a gridded precipitation field. MODE is used to calculate the fuzzy OTS. For the precipitation forecasts, forecasts with multiple members having no object identified by the MODE algorithm are excluded due to the difficulty of defining a distance between such forecasts. We excluded 2 of 26 days at 3 hour lead time and 6 of 26 days at 12 hour lead time.

Composite dendrograms are also created for forecasts of 10-m wind speed, 850 hPa wind speed, and 500 hPa temperature. Squared Euclidean distance is used as a distance measure for these non-precipitation fields, consistent with the traditional application of Ward's algorithm (Anderberg 1973). Each non-precipitation forecast is first normalized to have zero mean and unit variance, by subtracting from each value the domain average of that forecast and dividing by the domain average standard deviation, as in Alhamed et al. (2002). Non-precipitation composite distances are computed using the average normalized ED.

d. Relative merging height

Results from the composite dendrograms are also shown using an alternative summary measure based on the fraction of total height (hereafter, merging height) on a dendrogram where some characteristic of the clusters first appears.

For each forecast from each member, the merging height where another member with different ICs, LBCs, PBL scheme, model, or microphysics scheme, joins the same cluster as that member is calculated. The merging height that a member without radar and mesoscale data assimilation (i.e., a member labeled C0) joins the same cluster as a member with radar and mesoscale data assimilation is also calculated. The median merging height from the distribution of all members over all forecasts is used to summarize the relative importance of the different types of ensemble perturbations on ensemble diversity.

A lower median merging height for a given type of perturbation (e.g., PBL scheme perturbation) is interpreted as the forecast having a lower sensitivity to that type of perturbation. This is because low merging height indicates that members with that perturbation in common (e.g., members with YSU) are relatively less distinct from members with a different perturbation (e.g. members with MYJ). Likewise, higher values of median merging height indicate an

increased sensitivity of the forecast to that type of perturbation since members with different perturbations (e.g., YSU members vs. MYJ members) are more likely to remain in different clusters until closer to the top of the dendrogram and are therefore relatively more distinct from each other.

Tests using hypothetical dendrograms show that composite dendrograms reflect the perturbation type that produces more complete and cleaner separation of members. In contrast, the median merging heights reflect the perturbation type that more frequently produces clusters based on that type of perturbation, even if clusters based on that type of perturbation are not cleanly separated.

3. HCA for hourly accumulated precipitation

In this section the fuzzy OTS (Object-based Threat Score; defined in Part I) composite dendrograms of forecasts of hourly accumulated precipitation at lead times of 3, 12, and 24 hours (valid at 03 UTC, 12 UTC and 00 UTC respectively) are presented in subsection a. Subsection b presents results using the median merging height.

a. Results of Composite Dendrograms

The composite dendrogram at 3 hour lead time (Fig. 1a) shows that the primary distinction between members is the assimilation of radar and mesoscale data. The C0 members that did not assimilate the radar and mesonet data form a distinct cluster. The remaining members cluster primarily by model dynamic cores with one cluster of all the NMM members and another cluster of all the ARW members. ARPS CN is also included in the ARW cluster.

Within the two main ARW and NMM clusters, all members with common microphysics scheme also cluster together.

The composite dendrogram at 12 hour lead time (Fig. 1b) also shows the WRF members clustered by model dynamic core. The C0 members again form a distinct cluster, but its dissimilarity from the other members is less than at 3 hour lead time. At 12 hour lead time the C0 cluster merges with the ARW cluster before the NMM cluster merges with the ARW cluster. At 12 hour lead time there is not a clear sub-clustering by either microphysics scheme or PBL scheme.

The composite dendrogram at 24 hour lead time (Fig. 1c) has three primary clusters of members with common model dynamic core (ARW, NMM and ARPS). The NMM cluster has two distinct sub-clusters, one containing all the NMM members with MYJ PBL scheme and another containing all the NMM members with YSU PBL scheme. The ARW cluster does not have sub-clusters with a common physics configuration as distinctly as the NMM members.

In summary, the composite dendrograms indicate that systematic clustering of the object-oriented precipitation forecasts is determined by the model dynamic core more than the physics schemes. There is further sub-clustering based on the microphysics schemes at early lead time and an increasing impact of the PBL scheme relative to microphysics scheme at later lead times. The microphysics schemes have the most direct effect on precipitation in the initial hours, especially for the precipitation initialized through radar data assimilation. In contrast, the 24 hour forecast is around the time of peak afternoon precipitation associated with the diurnal cycle (see, e.g., Clark et al. 2009). At such time the development of the convective boundary layer has strong effect on convective initiation and subsequent evolution (e.g., Zhang and Zheng 2004;

Xue and Martin 2006). The composite dendrogram also indicates a decreasing impact of radar and mesonet data assimilation with increasing forecast lead time.

These automated HCA results related to microphysics and PBL schemes are also consistent with manual subjective findings of Weisman et al. (2008) with limited cases. Previous studies also show that the impact of radar and mesoscale data assimilation on precipitation forecasts usually lasts for less than 12 hours (Kong et al. 2009; Xue et al. 2009; Kain et al. 2011), consistent with these composite dendrograms. The sensitivity to model core is also consistent with previous studies at CP resolution (e.g., Yussouf et al. 2004, Gallus and Bresch 2006).

b. Results of median merging height

The median merging heights (Fig. 2) for hourly precipitation forecasts show trends that are consistent with the composite dendrograms (Fig. 1). Recall that higher values of median merging height indicate an increased impact of a type of perturbation on ensemble spread. There are four specific systematic trends illustrated in Fig. 2 that are now discussed.

The first trend is that the dynamic cores have a larger impact than the physics at all lead times. The second trend is an increasing relative impact of the PBL scheme perturbation with increasing lead time. At 3 hour lead time radar data has the largest impact and the microphysics scheme has a larger impact than the PBL scheme. As forecast time increases, the impact of the PBL scheme increases more than the other types of perturbations and becomes more dominant than the microphysics scheme with increasing lead time. The increasing relative impact of the PBL scheme perturbation is consistent with composite dendrograms showing increasingly strong sub-clustering by the PBL scheme, relative to the microphysics scheme, at 3, 12 and 24 hour lead times (Fig. 1). When median merging heights are plotted for only the members with the ARW

and NMM model separately, both models show the PBL scheme becomes increasingly more important with time (not shown). The increasing trend is more pronounced when only considering the NMM members than only the ARW members (not shown). This difference is consistent with the composite dendrogram showing more distinct sub-clusters by the PBL scheme in the NMM members than the ARW members at 24 hour lead time (Fig. 1c). The increasing impact of the PBL scheme with forecast time is due to its increased influence on the mesoscale environment that supports precipitation systems. At the longer forecast ranges these are often newly initiated systems that did not exist at the initial time. While microphysics should continue to influence the precipitation forecasts, the precipitation systems themselves have to be supported by the mesoscale environment in the first place.

The third trend is a diurnal cycle in the impact of the model, PBL and microphysics perturbations (Fig. 2). The median merging height for all of these perturbation types has a peak at 24 hour lead time, which corresponds with the afternoon maximum in the diurnal convective cycle.

The fourth trend illustrated by the median merging heights is the decreasing relative impact of radar and mesoscale data assimilation with increasing forecast time (Fig. 2). This trend is consistent with composite dendrograms which show distinct clusters of members with and without radar and mesonet data assimilation at 3 and 12 hour lead time (Fig. 1a and 1b) but not at 24 hour lead time (Fig. 1c). Median merging heights further reveal that the impact of assimilating radar and mesonet data on forecasts is greater than the model and physics perturbations at early lead time and then becomes less important than the model and physics perturbations at later lead times. This trend is consistent with previous studies (Xue et al. 2009; Kong et al. 2009; Kain et al. 2011).

In summary, the results of median merging heights show more impact of the model dynamic cores than the physics throughout the forecast time, an increasing impact of the PBL scheme with increasing forecast time, a diurnal cycle in the impact of the models and physics on the forecasts, and decreasing impact with time of radar assimilation. These results are consistent with the composite dendrograms, and with physical understanding. The sensitivity of HCA results to the choice of clustering algorithm is examined by comparing the median merging heights using the Unweighted Pair Group Method clustering Algorithm (UPGMA; Jain and Dubes 1988⁵) to the results described in this section. In general, the results from UPGMA (Fig. 3) are consistent with those from Ward's algorithm.

c. Regional HCA

A composite dendrogram is also created based on regional HCA which follows the method described in Part I section 7. Regional sub-domains are of interest because of the localized nature of severe weather forecasts. Here, the 00 UTC verification time (i.e., 24 hour lead time) is emphasized for consistency with the HWT Spring Experiment, which focused on day 1 severe weather forecasts (e.g., Schwartz et al. 2009) and the diurnal maximum of convective instability. The approximate center of regions where widespread intense convection (evaluated subjectively) was either forecast by multiple ensemble members or observed were considered potential regions of interest. This resulted in 49 different regions to be clustered since some forecasts contained multiple non-overlapping regions of interest. Of the subjectively selected 49 potential regions, regions in which more than 3 ensemble members did not have any

⁵ UPGMA defines the distance between two clusters as the average distance between each possible pair of members from the two clusters. This contrasts with Ward's algorithm which defines the distance between clusters as the increase in variability resulting from merging those two clusters.

forecast objects identified by the MODE algorithm were removed from consideration. This is because in such instances, maximum distance of 1.0 are often obtained due to forecasting of no object, which contributed more noise than signal to the overall results. Also eliminated were regions where no severe weather was recorded in the SPC storm reports log within about 300 km of the center of the region within an hour before or after the forecast valid time. These were removed because the focus of this study is on intense convection, such as that of interest to the SPC (reasons for this focus are also discussed in Part I Appendix B). These objective criteria reduced the 49 potential regions, identified using subjective criteria, to 34 regions.

Figure 4 shows that for the regional HCA composite, the primary clustering is again based on models. In addition, the cluster of NMM members is again sub-clustered based on the PBL scheme. This composite dendrogram using regional domains is therefore consistent with the full-domain composite (Fig. 1c).

4. HCA for 10-m wind speed and mid-tropospheric variables

HCA for non-precipitation variables of 10-m wind speed, 850 hPa wind speed and 500 hPa temperature is performed using the traditional implementation of Ward's algorithm.

Composite dendrograms for the non-precipitation variables are discussed in subsection 4a. The results based on median merging heights are discussed in subsection 4b.

a. Results of Composite dendrograms

1) 10-METER WIND SPEED

At 3 hour lead time the 10-m wind speed composite dendrogram shows primary clusters based on the PBL scheme and radar and mesonet data assimilation (i.e., an MYJ cluster, a YSU cluster, and a C0 members cluster) (Fig. 5a)⁶. Secondary clustering is dependent on the PBL scheme that the primary clustering is based on. The MYJ cluster has sub-clusters determined by the model dynamic core (i.e., ARW and NMM sub-clusters). In contrast, for the YSU cluster, the sub-clusters do not show distinct clusters based on models. For example, the two pairs with common IC and LBC perturbation (i.e., sub-clusters of ARWP2 with NMMP2 and ARWP4 with NMMP4) are paired together even though they use different models. Fig. 5a indicates stronger impact of the PBL scheme than the dynamic core on the 10-m wind forecast during the early forecast hours.

At 24 hour lead time the impact of the PBL scheme on the diversity of 10-m wind speed forecasts has diminished compared to the impact of the LBC perturbation (Fig. 5b). Unlike at earlier lead times when primary clusters are based on the PBL scheme, at 24 hour lead time they are clustered primarily by their LBCs. From left to right, the 4 members in the first group, the 3 members in the second group, the 7 members in the third group, and the 4 members in the fourth group share the same NAM forecasts (NAMf in Table 1), the same ARW-member SREF forecasts (emN1), the same ETA-member SREF forecasts (etaKFN1 and etaBMJN1), and the same NMM-member SREF forecasts (nmmN1), respectively, for their LBCs. The secondary clusters at 24 hour lead time (Fig. 5b) also suggest a stronger influence of the synoptic scale IC perturbations from SREF at later lead time than at early lead time since members with common IC are also sub-clustered. The impact of radar and mesonet data assimilation at 24 hour lead time

⁶ The two ARPS members are excluded from the 10-m wind speed dendrograms because 10-m wind was not generated as an output variable in ARPS during the 2009 Spring Experiment.

is minimal since the CN and C0 members cluster together at a low level before merging with any other members.

2) MID-TROPOSPHERIC VARIABLES

Mid-tropospheric variables such as 500 hPa temperature tend to cluster according to the IC at early lead time and the LBC at later lead times (e.g., Fig 6a,b). This clustering indicates a stronger relative impact of the IC and LBC perturbations for mid-tropospheric variables than for near-surface wind speed and precipitation. Temperature forecasts at 500 hPa are representative of the other mid-tropospheric variables such as 850 hPa wind speed.

At 3 hour lead time (Fig. 6a), the primary distinction between members, even for non-precipitation mid-tropospheric variables, is the assimilation of radar and mesonet data. At this time the C0 members form a distinct cluster. This separate cluster is also seen at the 12 hour lead time (not shown). The remaining members tend to cluster based on the ICs. Many pairs of members with the same IC (e.g., P4, N3, N4, P1, and P2) cluster together immediately but none of these pairs then merge together based on common LBC. Recall that the ensemble ICs in 2009 were obtained by adding SREF perturbations to the control analysis. This is done in practice by adding and subtracting half the difference between paired SREF 3-hour forecasts (valid at the initial time 00 UTC) to the control analysis. For example, the ARW N1 analysis is obtained by subtracting from the ARW CN analysis the difference between the SREF em P1 and SREF em N1 forecasts, divided by 2. The resulting perturbations to u and v wind components, potential temperature, and specific humidity are rescaled to have a root mean square value of 1 m/s, 0.5 K, and 0.02 g/kg, respectively.

By the 24 hour forecast time the primary clustering is based on the LBCs. All members using NAM forecast as LBC (ARW CN, ARW C0, NMM CN, NMM C0, ARPS CN, ARPS C0) are in a cluster while all members using SREF em N1 forecast as LBC (ARW N1, ARW P1, NMM P1) are in a cluster and so on. Within the primary clusters, members that also have the same ICs (e.g., ARW P2, NMM P2) form sub-clusters. The composite HCA also revealed that clustering based on the LBCs begins earlier at 500 hPa than at 850 hPa (not shown). The primary clusters in the 24 hour composite suggest a stronger impact of the LBC perturbation than the other types of perturbation for the mid-tropospheric variables at this lead time.

Temperature perturbations from the control forecast at 500 hPa were examined in several forecasts to better understand the primary clustering based on LBC late in the forecasts (Fig. 6b). As a representative example, Figure 7 shows the anomalies of ARW N3, ARW P3 and NMM N3 from the ARW CN control forecast, initialized at 00 UTC 1 May 2009, for 500 hPa temperature over the entire forecast domain. While the control member, ARW CN, obtains LBCs from NAM forecasts, the members ARW N3, ARW P3 and NMM N3 obtain their LBCs from the forecasts of SREF member etaKF N1. ARW N3 and NMM N3 also have the same ICs while ARW N3 and ARW P3 have IC perturbations of opposite sign as discussed above. The members also have different physics configuration as shown in Table 1.

The anomalies resulting from the IC perturbations are still apparent after 3 hours of forecast time while anomalies arising from the LBCs are just beginning to enter the domain (Fig. 7a,b,c). Areas that were inside the domain at the initial time (Northern Plains, Southern Plains and Northern Great Lakes) have anti-correlated anomalies of opposite sign between ARW N3 and ARW P3 (Fig. 7ab) but correlated anomalies of the same sign between ARW N3 and NMM N3 (Fig. 7ac). However, the anomalies (relative to ARW CN) entering the domain from the

LBCs in strong westerly flow (flow pattern is not shown) have both similar shape and similar sign. The similarity between members with common ICs at the 3 hour lead time, ARW N3 and NMM N3 in this example, causes the clustering of such members at the three hour lead time for the mid-tropospheric variables.

After 12 hours of forecast time the anomalies originating from the LBCs already dominate the anomalies of all three members (relative to ARWCN) as the LBCs begin to spread across the interior domain (Fig 7d,e,f). All three members have a large scale cold anomaly of more than 5 degrees C over the northern Rockies, where the flow originated from the western boundary, and a large scale warm anomaly of several degrees C over the northern plains, where the flow originated from the northern boundary. This example further confirms the eventual primary clustering of members based on the LBC forecasts as shown by the 24 hour composite dendrogram (Fig. 6b).

b. Results of median merging height

The impact of different types of perturbations on non-precipitation forecast diversity, as a function of forecast time, is examined in terms of median merging height (Fig. 8). This subsection describes the relative impacts of the different types of perturbation, the change in relative impact at different lead times, the relative impact of the same perturbation type for different variables, and the relative impact of IC versus LBC perturbation with increasing lead times.

The relative impacts of different types of perturbation implied by median merging heights are generally consistent with those implied previously by the composite dendrograms. At early lead times, the 10-m wind speed merging heights indicate the greatest impact is from the PBL scheme. However, the 850 hPa wind speed and 500 hPa temperature merging heights indicate the

greatest impact is from radar data assimilation (Fig. 8). At later lead time the relative impact of the PBL scheme, microphysics scheme and model is not clear from the composite dendrograms (Fig. 5b and 6b). However, the merging heights for both 10-m wind speed and the mid-tropospheric variables indicate a greater impact of the PBL scheme than the microphysics and model (Fig. 8).

In terms of the variation of the impact of different perturbations with forecast lead time, the impact of the PBL scheme decreases at later lead times for 10-m wind speed (Fig. 8a). The impact of the microphysics scheme and model for 10-m wind speed and the impact of the microphysics scheme, PBL scheme and model for the mid-tropospheric variables all have a peak at the 24 hour lead time, valid at 00 UTC around which time the maximum convective activity occurs. The impact of radar data assimilation decreases with lead time for all of the non-precipitation variables (Fig. 8).

The relative impacts of a given perturbation type for the different forecast variables are also consistent with the composite dendrograms. The impact of both the physics and model perturbations is larger for 10-m wind speed than for the other non-precipitation variables (Fig. 8). The impact of model perturbations for all non-precipitation variables is smaller than that for precipitation (Fig. 2). Results are again similar when using UPGMA as a clustering algorithm (not shown).

The impact of IC and LBC perturbation is examined in a similar manner as that used for the model and physics perturbations and is summarized for both precipitation and non-precipitation variables in Fig. 9. Figure 9 shows the median merging height where members with different IC merge together along with the median merging height where members with different LBC merge together, for precipitation and 10-m wind speed. Divergence of these two lines

indicates an increasing impact of the LBC relative to the IC. Figure 9 shows such divergence begins around the 12 hour lead time for 10-m wind speed. This agrees with the timing of the first appearance of clusters with common LBC noted in the composite HCA results. Similar divergence with increasing lead time is observed for the mid-tropospheric variables (not shown). Also noteworthy in Figure 9 is the later onset and smaller amount of divergence of the IC and LBC merging heights, as well as generally smaller values, for precipitation compared to the non-precipitation variables. These differences indicate less impact of the IC and LBC perturbations for precipitation variables than the non-precipitation variables. Also, unlike the mid-tropospheric variables, the LBCs do not dominate the hourly accumulated precipitation forecasts at 24 hour lead time, the peak of the diurnal cycle, which suggests that the 24 hour precipitation features are mostly locally forced (e.g., Weckworth and Parsons 2006).

The results in this section using median merging height are, in general, consistent with the results from the composite dendrograms, except the relative impact of the PBL scheme and LBC perturbation for 10-m wind speed forecasts at 24 hour forecast time. The composite dendrogram indicates greater impact of the LBC than the PBL scheme at 24 hour forecast time, with clean separations of members based on the LBC (Fig 5b). However, the median merging height of the PBL scheme is higher than or comparable to that of the LBC at the 24 hour forecast time (Figs 8a and 9). Examinations of the dendrograms of all forecasts show that this is because both the PBL scheme and LBC are important (not shown). Forecasts dominated by the LBC typically have cleaner and more distinct clusters than forecasts dominated by the PBL scheme, but the PBL scheme dominated forecasts occur more frequently (19 out of 28 forecasts) than the LBC dominated forecasts. As discussed in section 2, the dominance of the LBC is therefore reflected by the composite dendrogram and that of the PBL scheme is reflected by the median

merging height. Subjective examination revealed that the forecasts dominated by the LBC correspond to synoptic scale disturbances entering the domain from the lateral boundaries during the forecast period. This occurred relatively infrequently during the second half of the 2009 Spring Experiment due to a strong blocking pattern.

5. Summary

This paper is the second of a two-part study seeking a better understanding of the impacts and relative importance of different sources of uncertainty on forecast diversity within a convection-allowing ensemble system produced by the CAPS for the 2009 NOAA HWT Spring Experiment. In this paper, an object-oriented HCA is used to identify clusters of forecasts with a focus on the structure, organization and location of intense convection. Traditional HCA is used to identify clusters for non-precipitation variables. The systematic impacts of perturbations are summarized with composite dendrograms and median merging heights of members sharing different perturbations.

The composite dendrograms show that at 3 hour lead time (valid 03 UTC) hourly accumulated precipitation forecasts cluster primarily by assimilation of radar and mesoscale data. Additional sub-clustering then corresponds to common model dynamics followed by common microphysics schemes. At 24 hour lead time (valid at 00 UTC) the clustering is primarily by the model dynamics with secondary clustering by the PBL scheme for the NMM members. At 12 hour lead time (valid at 12 UTC) there is primary clustering based on both the model dynamics and radar and mesoscale data assimilation. Members without assimilation of additional radar and mesonet data form a distinct cluster from members assimilating radar and mesonet data for the first 12 hours of the forecasts.

Median merging height results are consistent with the results of composite dendrograms. Median merging heights for precipitation forecasts further reveal that the model has a larger impact than the physics at all lead times, the microphysics scheme has a larger impact than the PBL scheme at 3 hour lead time, and the PBL scheme has an increasing impact with time that eventually outweighs the impact of the microphysics scheme. The impacts of the model and physics perturbations also follow a diurnal cycle with a maximum during the afternoon when convective activity is often greatest. The impact of radar and mesoscale data assimilation decreases with time and becomes smaller than the model and physics perturbations after the 12~18 hour lead times. The impact of the IC and LBC perturbations, as used in the current ensemble system, did not show up clearly for the precipitation forecasts, relative to the impacts of other perturbations.

For the non-precipitation variables, the composite clusters reveal that forecasts of 10-m wind speed initially (3 hour lead time) cluster primarily by the PBL scheme, with secondary clustering by the model in the MYJ PBL scheme cluster. Eventually (by 24 hour lead time) the clustering is primarily by the LBCs. Forecasts of the mid-tropospheric variables (850 hPa wind speed and 500 hPa temperature) initially cluster by the ICs, and eventually cluster by the LBCs. Radar data assimilation initially results in separate composite clusters even for the non-precipitation variables.

The HCA results from median merging heights are generally consistent with the HCA results from composite dendrograms for the non-precipitation variables as well. For 10-m wind speed and the mid-tropospheric variables, the median merging heights indicate a greater impact of the PBL scheme than the model and microphysics scheme at all lead times. The relative impact of the PBL scheme for 10-m wind speed decreases at later lead times as the LBCs'

impact increases. For the mid-tropospheric variables the median merging heights indicate that the impact of the model and physics peak at 24 hour lead time when the maximum convective activity occurs. For all of the non-precipitation variables, the impact of radar assimilation decreases with forecast time.

The median merging heights also provide a quantitative comparison of the relative impact of different perturbation types among the different forecast variables. For example, the model and physics perturbations have a larger impact on 10-m wind speed than the mid-tropospheric variables. Finally, there is less impact of the IC and LBC perturbations on precipitation forecasts than the non-precipitation forecasts.

More work is needed to further diagnose the physical reasons causing the ensemble to cluster as it does.

6. Discussion: implication for ensemble design, verification and post-processing

In this study a newly developed object-oriented HCA is applied to a convection-allowing ensemble during the 2009 NOAA HWT Spring Experiment. The results of the HCA can have several implications for future research on how to optimally design and appropriately verify, calibrate, and post-process a convection-allowing ensemble. The following only serves to discuss such implication elucidated from the clustering analysis results. Detailed and systematic studies are needed to answer these questions. Studies on quantitatively verifying different sub-groups of the ensemble members are on-going and are planned to be reported in future papers (Johnson et al. 2011b).

a. Ensemble design

Our results suggest that the optimal design of the CAPS 2009 convection-allowing ensemble should depend on the intended use of the ensemble. In this study we focus on the structure and organization of features in hourly accumulated precipitation forecasts by using an object-oriented framework. The HCA results imply that for next day (i.e., 24 hours) forecasts of intense convection, particular attention should be paid to the models and PBL schemes. At earlier lead times, for example for 3 hour forecasts, in addition to the models, more attention should be paid to the microphysics schemes and radar data assimilation.

Users interested in short term forecasts of near surface variables such as 10-m wind speed may find greatest improvements to ensemble design by optimizing the PBL scheme perturbations while the LBC perturbation strategy may be more relevant at later lead times. Users interested in upper level variables may benefit most from an increased emphasis on the LBC perturbations for longer lead time and the IC perturbations at short lead time. Attention should also be paid to the interaction between the model and physics perturbations as Figs. 1b and 5a suggest that sensitivity to the physics schemes can depend on the model dynamics and vice versa. Even for a specific user and a particular modeling system, the effectiveness of the ensemble design can also depend on the large scale flow regime (not shown). Thus, cautions are warranted when extrapolating the results of this study to other applications, seasons and configurations. However, it is worth noting that a composite OTS-HCA of the 2010 CAPS ensemble (Xue et al. 2010), which was configured differently than the 2009 CAPS ensemble, showed the same primary clustering of precipitation forecasts by the model dynamics (not shown).

Another consideration for storm-scale ensemble design is the horizontal scale of both IC and LBC perturbations. The IC/LBC ensembles in this study were generated by simply

downscaling from coarser resolution SREF forecasts. More work is needed to explore how to optimally design the IC and LBC perturbations that include all scales of uncertainty.

Future research should also further identify and quantify the added value of radar data and mesonet observation assimilation. In a composite sense there is a distinct cluster of the members without radar assimilation for at least 12 hours. The impact of assimilating the observations is also likely dependent on the data assimilation method adopted (e.g., Wang et al. 2008ab).

b. Post-processing, calibration and verification

The presence of systematic clusters of ensemble members violates the assumption that each member's forecast is a random (i.e., independent and equally likely) sample of the distribution of possible future states of the atmosphere (Leith 1974). This has implications for appropriate post-processing techniques since methods such as interpreting the percentage of members forecasting an event as the forecast probability are not strictly appropriate. This also implies a need for calibration since different clusters of members can have different systematic behaviors. Such systematic differences should be accounted for before combining the clusters into a single combined Probability Density Function of the ensemble forecast. Future research should seek appropriate post-processing and calibration methods in the presence of unequally likely and/or non-independent members for the purpose of explicit prediction of intense convection and its characteristics.

Forecast verification can be interpreted as quantifying the distance between a forecast and a verification field instead of two forecast fields. Therefore this study is of general relevance as a contribution to understanding the ways that object-oriented methods can be applied to convection-allowing forecasts. This study demonstrates advantages of using object-oriented

methods to measure similarity/dissimilarity of fine resolution precipitation forecasts. By construction, the object-oriented method in general is not able to identify a correct null forecast, which should be kept in mind while interpreting verification results. Future research should explore the use of object-oriented products and verification methods in a probabilistic framework to provide additional insight into convection-allowing ensemble performance.

Acknowledgments.

The authors are grateful to NSSL for the QPE verification data. We thank Nusrat Yussouf for providing tested and documented software for performing hierarchical cluster analyses, Dave Stensrud for helpful discussions of the results, NCAR for making MODE source code available, and two anonymous reviewers whose comments and suggestions greatly improved the manuscript. This research was supported by Science Applications International Corporation (SAIC) as a sponsor of the AMS Graduate Fellowship program and University of Oklahoma faculty start-up award 122-792100 and NSF AGS-1046081. The CAPS real-time forecasts were produced at the Pittsburgh Supercomputing Center (PSC), and the National Institute of Computational Science (NICS) at the University of Tennessee, and were mainly supported by the NOAA CSTAR program (NA17RJ1227). Some of the computing for this project was also performed at the OU Supercomputing Center for Education & Research (OSCER) at the University of Oklahoma (OU). Kevin Thomas, Yunheng Wang, Keith Brewster and Jidong Gao of CAPS are also thanked for the production of the ensemble forecasts.

References

- Alhamed, A., S. Lakshmivarahan, D. J. Stensrud, 2002: Cluster analysis of multimodel ensemble data from SAMEX. *Monthly Weather Review*. **130**. Pp 226-256.
- Aligo, E. A., W. A. Gallus, and M. Segal, 2007: Summer rainfall forecast spread in an ensemble initialized with different soil moisture analyses. *Wea. Forecasting*, **22**, 299–314.
- Anderberg, M.R., 1973: *Cluster Analysis for Applications*. Academic Press, 359 pp.
- Arribas, A., K. B. Robertson, and K. R. Mylne, 2005: Test of a poor man's ensemble prediction system for short-range probability forecasting. *Mon. Wea. Rev.*, **133**, 1825–1839.
- Atger, F., 1999: Tubing: An alternative to clustering for the classification of ensemble forecasts. *Wea. Forecasting*, **14**, 741–757.
- 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.
- Berner, J., S.-Y. Ha, J. P. Hacker, A. Fournier, C. and Snyder, 2011: Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multi-physics representations. *Mon. Wea. Rev.*, In Press.
- Brankovic, C., T. N. Palmer, F. Molteni, S. Tibaldi, and U. Cubasch, 1990: Extended-range predictions with ECMWF models: Timelagged ensemble forecasting. *Quart. J. Roy. Meteor. Soc.*, **116**, 867–912.
- Brankovic, C., B. Matjacic, S. Ivatek-Sahdan, 2008: Downscaling of ECMWF ensemble forecasts for cases of severe weather: Ensemble statistics and cluster analysis. *Mon. Wea. Rev.*, **136**, 3323-3342.

- Bryan, George H., John C. Wyngaard, J. Michael Fritsch, 2003: Resolution requirements for the simulation of deep moist convection. *Mon. Wea. Rev.*, **131**, 2394–2416.
- Cheng, X., and J. M. Wallace, 1993: Cluster analysis of the Northern Hemisphere wintertime 500-hPa height field. *J. Atmos. Sci.*, **50**, 2674–2696.
- 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.
- Clark, A. J., W. A. Gallus, M. Xue, F. Kong, 2009: A comparison of precipitation forecast skill between small convection-allowing and Large convection-parameterizing ensembles. *Wea. Forecasting*, **24**, 1121–1140.
- Davis, C., B. Brown, and R. Bullock, 2006: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784.
- Du, J., J. McQueen, G. DiMego, Z. Toth, D. Jovic, B. Zhou, and H. Chuang, 2006: New dimension of NCEP Short-Range Ensemble Forecasting (SREF) system: Inclusion of WRF members, *Preprint, WMO Expert Team Meeting on Ensemble Prediction System, Exeter, UK, Feb. 6-10, 2006* [available online <http://wwwt.emc.ncep.noaa.gov/mmb/SREF/reference.html>]
- Dudhia, J., 1989: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107.
- Ek, M. B., K. E. Mitchell, Y. Lin, P. Grunmann, E. Rogers, G. Gayno, and V. Koren, 2003: Implementation of the upgraded Noah land-surface model in the NCEP operational mesoscale Eta model. *J. Geophys. Res.*, **108**, 8851.

- Ferrier, B. S., 1994: A double-moment multiple-phase four-class bulk ice scheme. Part I: Description. *J. Atmos. Sci.*, **51**, 249–280.
- Fovell, R. G., and M. Y. C. Fovell, 1993: Climate zones of the conterminous United States defined using cluster analysis. *J. Climate*, **6**, 2103–2135.
- Gallus, W. A., Jr., and J. F. Bresch, 2006: Comparison of impacts of WRF dynamic core, physics package, and initial conditions on warm season rainfall forecasts. *Mon. Wea. Rev.*, **134**, 2632–2641.
- Gao, J.-D., M. Xue, K. Brewster, and K. K. Droegemeier, 2004: A three-dimensional variational data analysis method with recursive filter for Doppler radars. *J. Atmos. Ocean. Tech.*, **21**, 457-469.
- Gong, X. and Richman, M.B., 1995: On the application of cluster analysis to growing season precipitation data in North America east of the Rockies, *J. Climate*, **8**, 897-931.
- Hacker, J. P., S.-Y. Ha, C. Snyder, J. Berner, F. A. Eckel, E. Kuchera, M. Pocerlich, S. Rugg, J. Schramm, and X. Wang, 2011: The U.S. Air Force Weather Agency's mesoscale ensemble: Scientific description and performance results. *Tellus* **63A**, 1-17.
- Hohenegger, Cathy, Christoph Schär, 2007: Predictability and error growth dynamics in cloud-resolving models. *J. Atmos. Sci.*, **64**, 4467–4478.
- Hong, S.-Y., J. Dudhia, and S.-H. Chen, 2004: A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Mon. Wea. Rev.*, **132**, 103–120.
- Hou, Dingchen, Eugenia Kalnay, Kelvin K. Droegemeier, 2001: Objective verification of the SAMEX '98 ensemble forecasts. *Mon. Wea. Rev.*, **129**, 73–91.

- Hu, M., M. Xue, and K. Brewster, 2006: 3DVAR and cloud analysis with WSR-88D level-II data for the prediction of Fort Worth tornadic thunderstorms. Part I: Cloud analysis and its impact. *Mon. Wea. Rev.*, **134**, 675-698.
- Jain, A. J., and R. C. Dubes, 1988: *Algorithms For Clustering Data*. Prentice Hall. Pp 72-80
- 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.
- Janjic', Z. I., 2003: A nonhydrostatic model based on a new approach. *Meteor. Atmos. Phys.*, **82**, 271-285.
- Jankov, I., W.A. Gallus, M. Segal, B. Shaw, and S.E. Koch, 2005: The impact of different WRF model physical parameterizations and their interactions on warm season MCS rainfall. *Wea. Forecasting*, **20**, 1048-1060.
- Jankov, I., W.A. Gallus, M. Segal, and S.E. Koch, 2007: Influence of initial conditions on the WRF-ARW model QPF response to physical parameterization changes. *Wea. Forecasting*, **22**, 501-519.
- Johnson, A., X. Wang, F. Kong, M. Xue, 2011a: Hierarchical cluster analysis of a convection-allowing ensemble during the Hazardous Weather Testbed 2009 Spring Experiment. Part I: Development of object-oriented cluster analysis method for precipitation fields. *Mon. Wea. Rev.*, Submitted.
- Johnson, A., X. Wang, F. Kong, M. Xue, 2011b: Object-oriented clustering analysis of CAPS convective scale ensemble forecasts for the NOAA Hazardous Weather Testbed Spring Experiment: A first step toward optimal ensemble configuration for convective scale

- probabilistic forecasting. *Extended Abstract, 24th Conf. Wea. Anal. Forecasting/20th Conf. Num. Wea. Pred. Amer. Meteor. Soc.*, Paper 9A.3.
- Kalkstein, L., G. Tan, and J. A. Skindlov, 1987: An evaluation of three clustering procedures for use in synoptic climatological classification. *J. Climate Appl. Meteor.*, **26**, 717–730.
- Kain, J. S., M. Xue, M. C. Coniglio, S. J. Weiss, F. Kong, T. L. Jensen, B. G. Brown, J. Gao, K. Brewster, K. W. Thomas, Y. Wang, C. S. Schwartz, and J. J. Levit, 2011: Assessing advances in the assimilation of radar data within a collaborative forecasting-research environment. *Wea. Forecasting*, In press.
- Kong, F., M. Xue, 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. *22nd Conf. Wea. Anal. Forecasting/18th Conf. Num. Wea. Pred.*, Salt Lake City, Utah, Amer. Meteor. Soc., CDROM 3B.2.
- Kong, F., M. Xue, K.W. Thomas, J. Gao, Y. Wang, K. Brewster, K.K. Droegemeier, J. Kain, S. Weiss, D. Bright, M. Coniglio, and J. Du, 2009: A real-time storm-scale ensemble forecast system: 2009 Spring Experiment, *10th WRF Users' Workshop, NCAR Center Green Campus*, Boulder, CO, June 23-26, 2009, Paper 3B.7.
- Lacis, A. A., and J. E. Hansen, 1974: A parameterization for the absorption of solar radiation in the earth's atmosphere. *J. Atmos. Sci.*, **31**, 118–133.
- Leith, C., 1974: Theoretical skill of monte carlo forecasts. *Mon. Wea. Rev.*, **102**, 409–418.
- Lin, Yuh-Lang, Richard D. Farley, Harold D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Climate Appl. Meteor.*, **22**, 1065–1092.

- Molinari, John, Michael Dudek, 1992: Parameterization of convective precipitation in mesoscale numerical models: A critical review. *Mon. Wea. Rev.*, **120**, 326–344.
- Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliaqis, 1996: The ECMWF ensemble prediction system: methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119.
- Nakaegawa, Tosiuyuki, Masao Kanamitsu, 2006: Cluster analysis of the seasonal forecast skill of the NCEP SFM over the Pacific–North America sector. *J. Climate*, **19**, 123–138.
- 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**, 421–427.
- Palmer, T. N., C. Brankovic, F. Molteni, S. Tibaldi, L. Ferranti, A. Hollingsworth, U. Cubasch, E. Klinker, 1990: The European Centre for Medium-Range Weather Forecasts (ECMWF) Program on Extended-Range Prediction. *Bull. Amer. Meteor. Soc.*, **71**, 1317–1330.
- Palmer, T. N., Buizza, R., Doblus-Reyes, F., Jung, T., Leutbecher, M., Shutts, G. J., Steinheimer, M., Weisheimer, A.(2009). Stochastic parametrization and model uncertainty, ECMWF Technical Memorandum, 598, 42.:
<http://www.ecmwf.int/publications/library/do/references/list/14>.
- Petch, J. C., 2006: Sensitivity studies of developing convection in a cloud-resolving model. *Quart. J. Roy. Meteor. Soc.*, **132**, 345–358.
- Schwartz, C., J. Kain, S. Weiss, M. Xue, D. Bright, F. Kong, K. Thomas, J. Levit, and M. Coniglio, 2009: Next-day convection-allowing WRF model guidance: A second look at 2 vs. 4 km grid spacing. *Mon. Wea. Rev.*, **137**, 3351–3372.
- 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.].
- Stensrud, D. J., J. Bao, and T. T. Warner, 2000: Using initial conditions and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.*, **128**, 2077–2107.
- Tao, W.-K., and Coauthors, 2003: Microphysics, radiation, and surface processes in the Goddard Cumulus Ensemble (GCE) model. *Meteor. Atmos. Phys.*, **82**, 97–137.
- Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115.
- Toth, Zoltan, Eugenia Kalnay, Steven M. Tracton, Richard Wobus, Joseph Irwin, 1997: A synoptic evaluation of the NCEP ensemble. *Wea. Forecasting*, **12**, 140–153.
- Tracton, M. S., and E. Kalnay, 1993: Operational ensemble prediction at the National Meteorological Center: Practical aspects. *Wea. Forecasting*, **8**, 379–398.
- Wandishin, M. S., S. L. Mullen, D. J. Stensrud, and H. E. Brooks, 2001: Evaluation of a short-range multimodel ensemble system. *Mon. Wea. Rev.*, **129**, 729–747.
- Wang, X., D. Barker, C. Snyder, T. M. Hamill, 2008a: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: observing system simulation experiment. *Mon. Wea. Rev.*, **136**, 5116–5131.
- Wang, X., D. Barker, C. Snyder, T. M. Hamill, 2008b: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part II: real observation experiments. *Mon. Wea. Rev.*, **136**, 5132–5147.

- Weber, R. O., and P. Kaufmann, 1995: Automated classification scheme for wind fields. *J. Appl. Meteor.*, **34**, 1133–1141.
- Weckwerth, Tammy M., David B. Parsons, 2006: A Review of Convection Initiation and Motivation for IHOP_2002. *Mon. Wea. Rev.*, **134**, 5–22.
- 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.
- Weiss S., J. Kain, M. Coniglio, D. Bright, J. Levit, G. Carbin, R. Sobash, J. Hart, R. Schneider, 2009. NOAA Hazardous Weather Testbed Experimental Forecast Program Spring Experiment 2009: Program Overview and Operations Plan. pg. 49.
http://hwt.nssl.noaa.gov/Spring_2009/
- Xue, M., K. K. Droegemeier, and V. Wong, 2000: The Advanced Regional Prediction System (ARPS)—a multiscale nonhydrostatic atmospheric simulation and prediction tool. Part I: Model dynamics and verification. *Meteor. Atmos. Phys.*, **75**, 161–193.
- Xue, M., and Coauthors, 2001: The Advanced Regional Prediction System (ARPS)—a multi-scale nonhydrostatic atmospheric simulation and prediction tool. Part II: Model physics and applications. *Meteor. Atmos. Phys.*, **76**, 143–166.
- Xue, M., D.-H. Wang, J.-D. Gao, K. Brewster, and K. K. Droegemeier, 2003: The Advanced Regional Prediction System (ARPS), storm-scale numerical weather prediction and data assimilation. *Meteor. Atmos. Physics*, **82**, 139-170.
- Xue, M. and W. J. Martin, 2006: A high-resolution modeling study of the 24 May 2002 case during IHOP. Part I: Numerical simulation and general evolution of the dryline and convection. *Mon. Wea. Rev.*, **134**, 149–171.

- Xue, M., F. Kong, K. W. Thomas, J. Gao, Y. Wang, K. Brewster, K. K. Droegemeier, X. Wang, J. Kain, S. Weiss, D. Bright, M. Coniglio, and J. Du, 2009: CAPS realtime 4-km multi-model convection-allowing ensemble and 1-km convection-resolving forecasts from the NOAA Hazardous Weather Testbed 2009 Spring Experiment. *Extended Abstract, 23rd Conf. Wea. Anal. Forecasting/19th Conf. Num. Wea. Pred. Amer. Meteor. Soc.*, Paper 16A.2.
- Xue, M., F. Kong, K.W. Thomas, Y. Wang, K. Brewster, J. Gao, X. Wang, S.J. Weiss, A.J. Clark, J.S. Kain, M.C. Coniglio, J. Du, T.L. Jensen, and Y.H. Kuo, 2010: CAPS realtime storm scale ensemble and high resolution forecasts for the NOAA Hazardous Weather Testbed 2010 Spring Experiment, Extd Abstracts, 25th Conf. on Severe Local Storms, Denver, CO, Amer. Meteor. Soc., Paper 7B.3., October 11-14, 2010.
- Yussouf N., D. J. Stensrud, S. Lakshmirarahan, 2004: Cluster analysis of multimodel ensemble data over New England. *Mon. Wea. Rev.*, **132**. 2452-2462.
- Zhang, Da-Lin, Wei-Zhong Zheng, 2004: Diurnal cycles of surface winds and temperatures as simulated by five boundary layer parameterizations. *J. Appl. Meteor.*, **43**, 157–169.

List of Figures

- FIG. 1. Composite dendrogram of forecasts for hourly accumulated precipitation over the verification domain at (a) 3 hour lead time, (b) 12 hour lead time, and (c) 24 hour lead time. Please see Table 1 for the symbols denoting different PBL and microphysics schemes and fonts denoting different models.
- FIG. 2. Median height on dendrogram (as ratio of total height) that each member's forecast of hourly accumulated precipitation merged into the same cluster as a member with a different PBL scheme (dash), Microphysics scheme(dotted), model (solid), or radar data assimilation (dash-dotted) option as a function of forecast lead time.
- FIG. 3. Same as Figure 2, except using UPGMA instead of the modified Ward's algorithm.
- FIG. 4. Composite dendrogram of forecasts of hourly accumulated precipitation over 34 cases of regional sub-domains at 24 hour lead time.
- FIG. 5. Dendrogram composited from 28 forecasts of 10-m wind speed over the entire verification domain at (a) 3 hr lead time and (b) 24 hr lead time.
- FIG. 6. Dendrogram composited from season-long forecasts of 500 hPa temperature over the entire verification domain at (a) 3hr lead time and (b) 24 hr lead time.
- FIG. 7. 500 hPa temperature forecasts initialized 00 UTC 1 May 2009, as anomalies from the ARW CN forecast, for (a) ARW N3 at 3 hour lead time, (b) ARW P3 at 3 hour lead time, (c) NMM N3 at 3 hour lead time, (d) ARW N3 at 12 hour lead time, (e) ARW P3 at 12 hour lead time, and (f) NMM N3 at 12 hour lead time.
- FIG. 8. Same as Figure 2, except for (a) 10-m Wind Speed, (b) 850 hPa Wind Speed, and (c) 500 hPa temperature.

FIG. 9. Median merging height that a member joins a cluster with another member using different Initial Condition and Lateral Boundary Condition (IC) or different Lateral Boundary Condition (LBC) for 10-m wind speed and precipitation as a function of forecast lead time.

TABLE 1. Details of ensemble configuration with columns showing the members, Initial Conditions (ICs), Lateral Boundary Conditions (LBCs), whether radar data is assimilated (R), and which Microphysics scheme (MP; Thompson, Ferrier (Ferr.), WRF Single Moment 6-class (WSM6), or Lin microphysics), Planetary Boundary Layer scheme (PBL; Mellor-Yamada-Janjic (MYJ), Yonsei University (YSU) or Turbulent Kinetic Energy (TKE)-based scheme), Shortwave radiation scheme (SW; Goddard, Dudhia or Geophysical Fluid Dynamics Laboratory scheme(GFDL)), and Land Surface Model (LSM; Rapid Update Cycle (RUC) or NCEP-Oregon State University-Air Force-NWS Office of Hydrology (NOAH)) was used with each member. Symbols identifying MP (@, \$, and # for Thompson, Ferrier, and WSM6, respectively) and PBL (^ and & for MYJ and YSU, respectively) schemes in other figures are also included in the brackets. Fonts distinguishing different models are the same as used in other figures where bold indicates an ARW member and italic indicates an NMM member. IC and LBC acronyms are defined in section 2b.

Member	IC	LBC	R	MP	PBL	SW	LSM
ARW CN	CN	NAMf	Y	Thom. (@)	MYJ (^)	Goddard	Noah
ARW C0	NAMa	NAMf	N	Thom. (@)	MYJ (^)	Goddard	Noah
ARW N1	CN – em	em N1	Y	Ferr. (\$)	YSU (&)	Goddard	Noah
ARW N2	CN – nmm	nmm N1	Y	Thom. (@)	MYJ (^)	Dudhia	RUC
ARW N3	CN - etaKF	etaKF N1	Y	Thom. (@)	YSU (&)	Dudhia	Noah
ARW N4	CN - etaBMJ	etaBMJ N1	Y	WSM6 (#)	MYJ (^)	Goddard	Noah
ARW P1	CN + em	em N1	Y	WSM6 (#)	MYJ (^)	Dudhia	Noah
ARW P2	CN + nmm	nmm N1	Y	WSM6 (#)	YSU (&)	Dudhia	Noah
ARW P3	CN + etaKF	etaKF N1	Y	Ferr. (\$)	MYJ (^)	Dudhia	Noah
ARW P4	CN + etaBMJ	etaBMJ N1	Y	Thom. (@)	YSU (&)	Goddard	RUC
<i>NMM CN</i>	CN	NAMf	Y	Ferr. (\$)	MYJ (^)	GFDL	Noah
<i>NMM C0</i>	NAMa	NAMf	N	Ferr. (\$)	MYJ (^)	GFDL	Noah
<i>NMM N2</i>	CN - nmm	nmm N1	Y	Ferr. (\$)	YSU (&)	Dudhia	Noah
<i>NMM N3</i>	CN - etaKF	etaKF N1	Y	WSM6 (#)	YSU (&)	Dudhia	Noah
<i>NMM N4</i>	CN - etaBMJ	etaBMJ N1	Y	WSM6 (#)	MYJ (^)	Dudhia	RUC
<i>NMM P1</i>	CN + em	em N1	Y	WSM6 (#)	MYJ (^)	GFDL	RUC
<i>NMM P2</i>	CN + nmm	nmm N1	Y	Thom. (@)	YSU (&)	GFDL	RUC
<i>NMM P4</i>	CN + etaBMJ	etaBMJ N1	Y	Ferr. (\$)	YSU (&)	Dudhia	RUC
ARPS CN	CN	NAMf	Y	Lin	TKE	2-layer	Noah
ARPS C0	NAMa	NAMf	N	Lin	TKE	2-layer	Noah

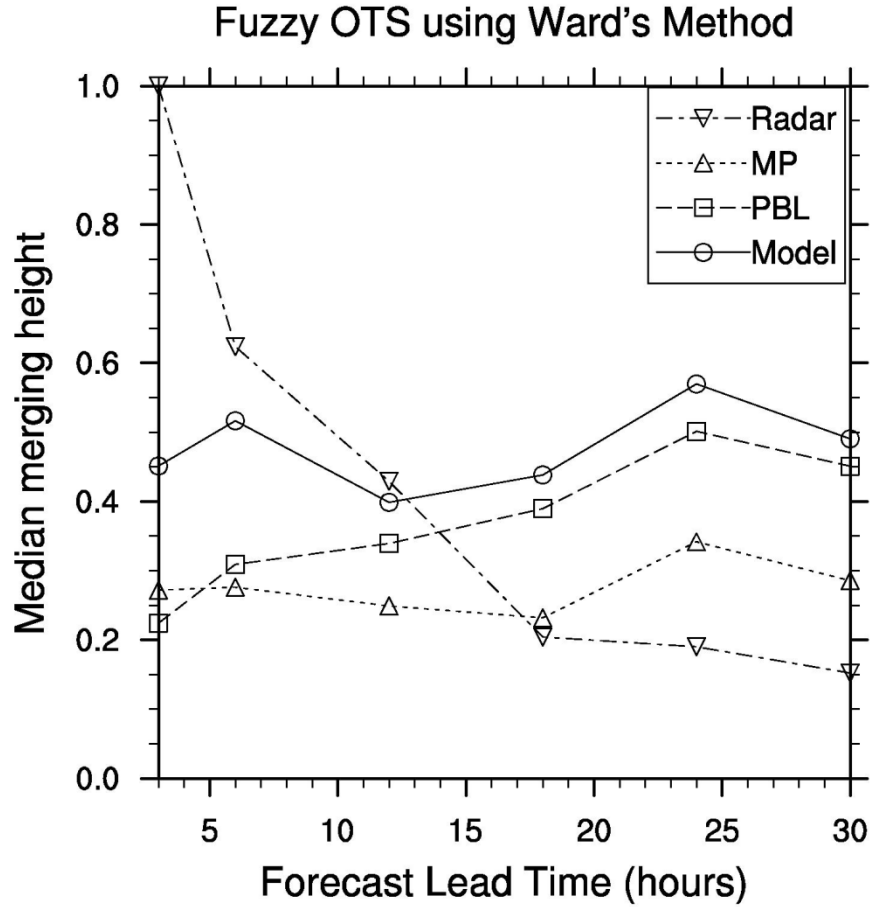


FIG. 2. Median height on dendrogram (as ratio of total height) that each member's forecast of hourly accumulated precipitation merged into the same cluster as a member with a different PBL scheme (dash), Microphysics scheme(dotted), model (solid), or radar data assimilation (dash-dotted) option as a function of forecast lead time.

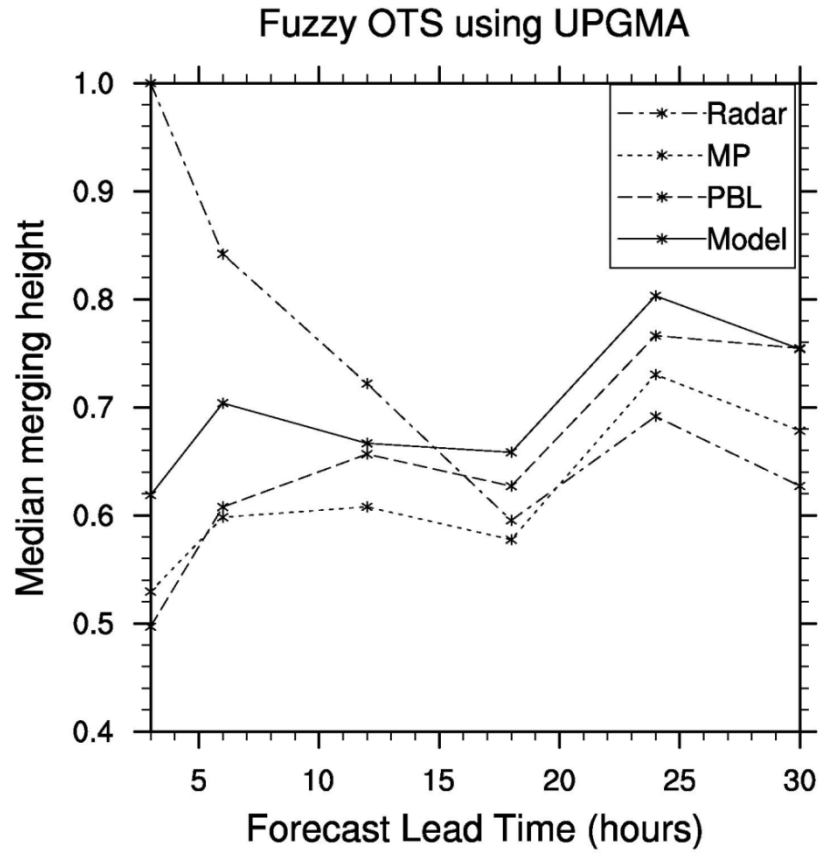


FIG. 3. Same as Figure 2, except using UPGMA instead of the modified Ward's algorithm.

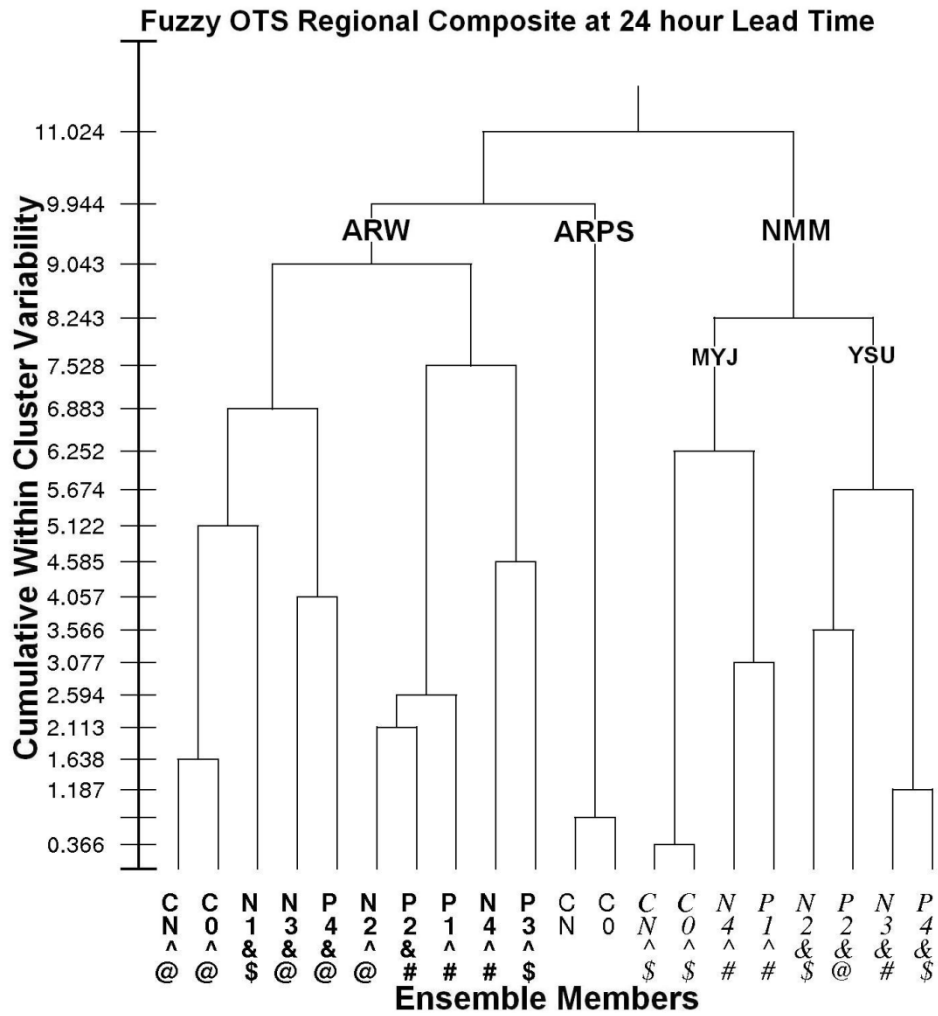


FIG. 4. Composite dendrogram of forecasts of hourly accumulated precipitation over 34 cases of regional sub-domains at 24 hour lead time.

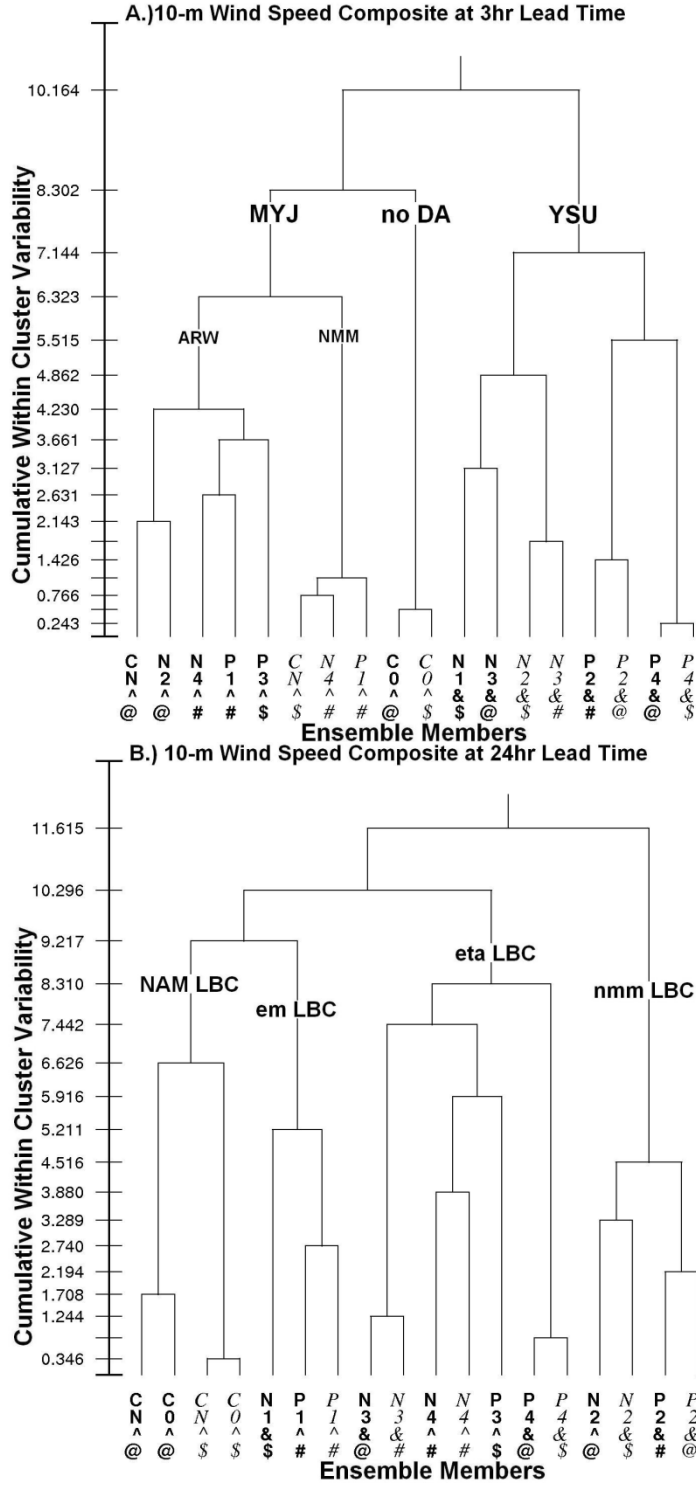


FIG. 5. Dendrogram composited from 28 forecasts of 10-m wind speed over the entire verification domain at (a) 3 hr lead time and (b) 24 hr lead time.

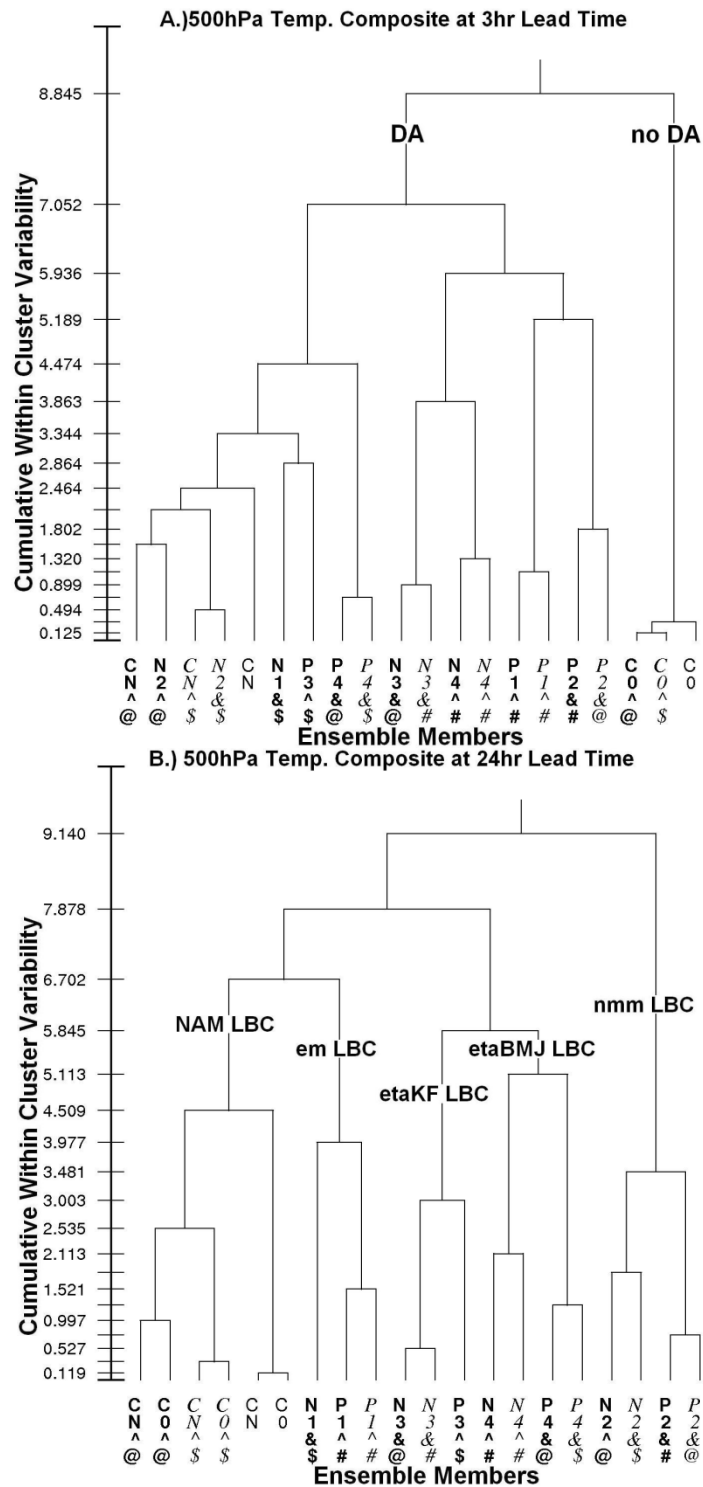


FIG. 6. Dendrogram composited from season-long forecasts of 500 hPa temperature over the entire verification domain at (a) 3hr lead time and (b) 24 hr lead time.

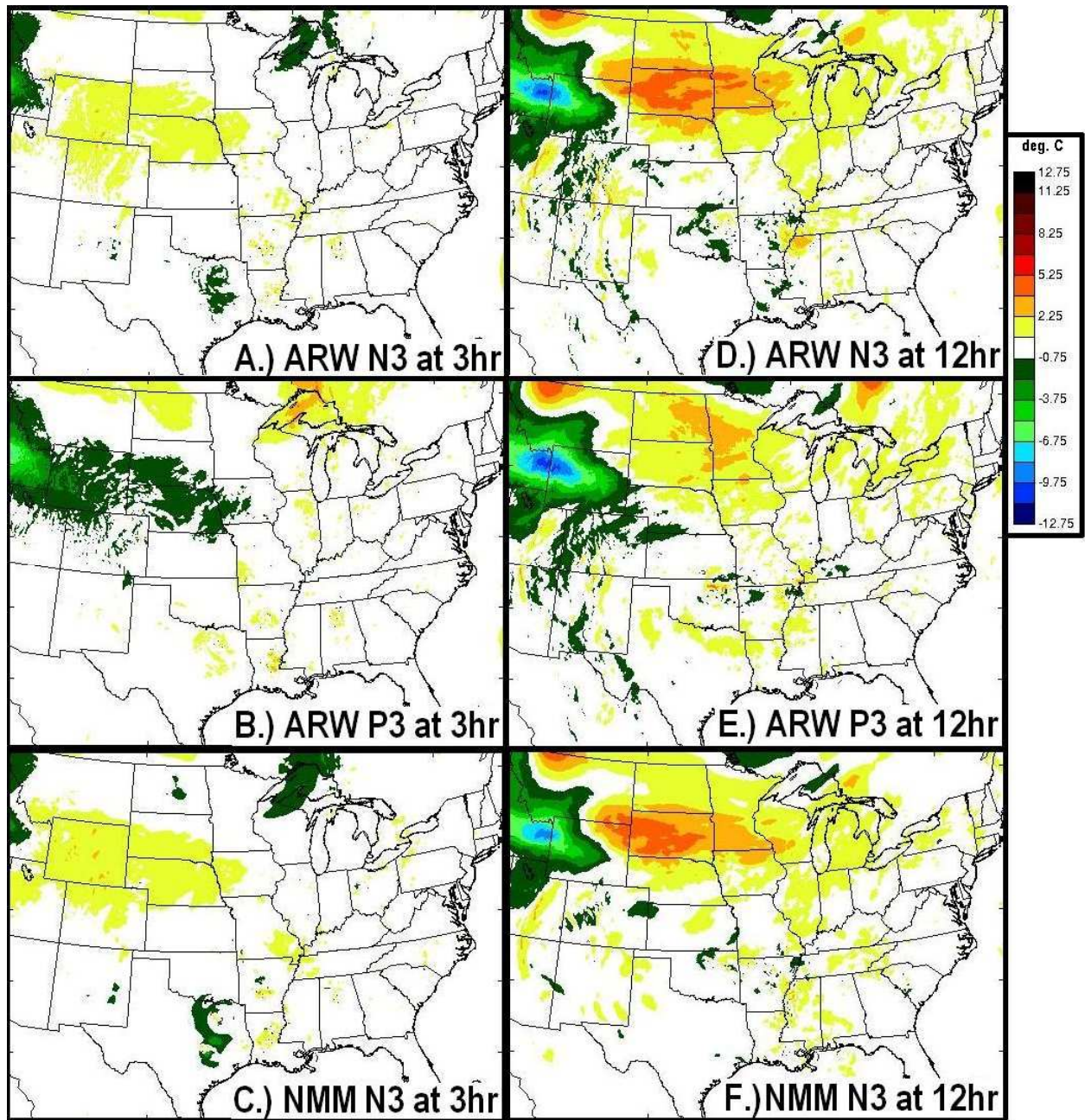


FIG. 7. 500 hPa temperature forecasts initialized 00 UTC 1 May 2009, as anomalies from the ARW CN forecast, for (a) ARW N3 at 3 hour lead time, (b) ARW P3 at 3 hour lead time, (c) NMM N3 at 3 hour lead time, (d) ARW N3 at 12 hour lead time, (e) ARW P3 at 12 hour lead time, and (f) NMM N3 at 12 hour lead time.

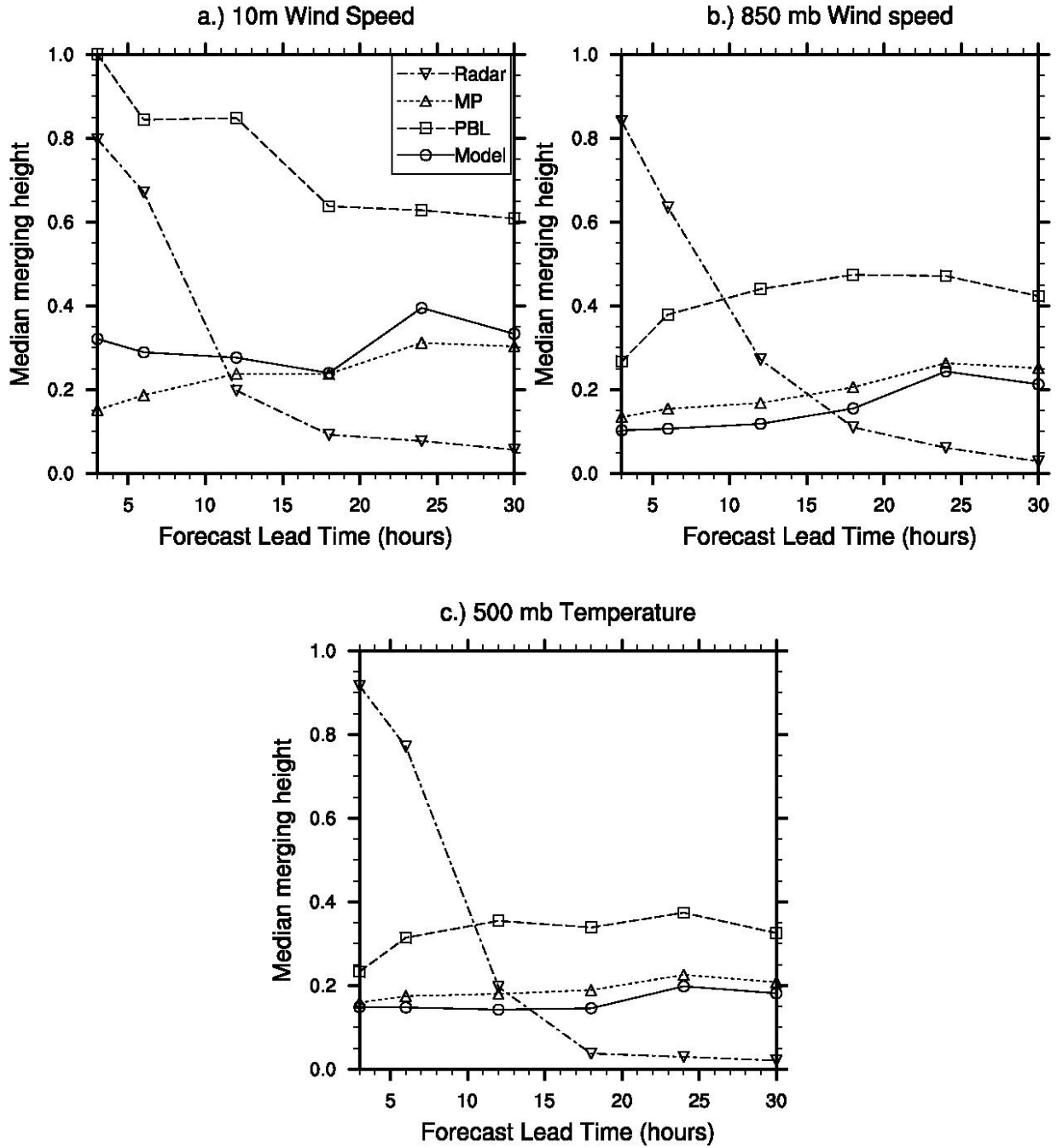


FIG. 8. Same as Figure 2, except for (a) 10-m Wind Speed, (b) 850 hPa Wind Speed, and (c) 500 hPa temperature.

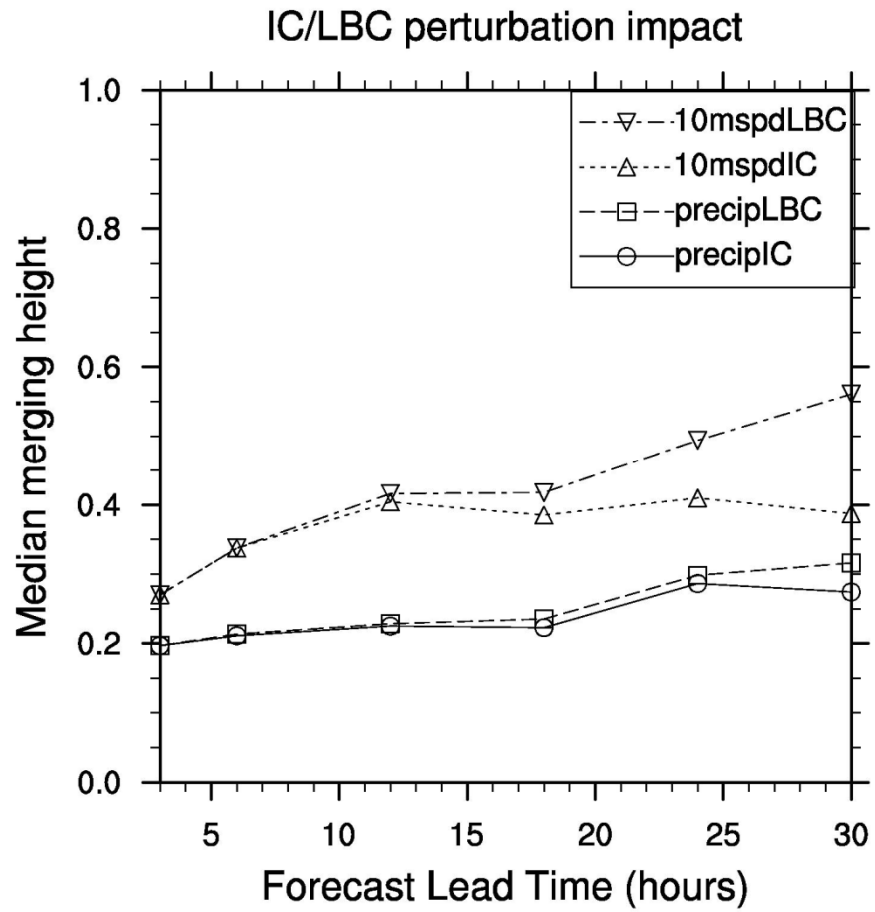


FIG. 9. Median merging height that a member joins a cluster with another member using different Initial Condition and Lateral Boundary Condition (IC) or different Lateral Boundary Condition (LBC) for 10-m wind speed and precipitation as a function of forecast lead time.